Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Interpretable hybrid data-driven modeling for predicting the compressive strength of ready-mixed concrete

Nguyen Hoang Long[1], May Huu Nguyen[1,2], Anh Nguyen Huu[3], Dung Quang Vu[1,*]

[1]University of Transport Technology, 54 Trieu Khuc, Thanh Liet, Hanoi, Vietnam
[2]Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima-Shi, Hiroshima 739-8527, Japan
[3]Advanced Materials and Intelligent Systems for Infrastructure and High-Speed Rail (AMIS-HSR) research group, University of Transport Technology, Hanoi, Vietnam; anhnh89@utt.edu.vn

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.5.150-179

***Corresponding author:**
Email address:
dungvq@utt.edu.vn

Received: 03/06/2026
Received in Revised Form: 29/09/2026
Accepted: 11/10/2026

Abstract: This research uses a large dataset of ready-mixed concrete batches collected directly from projects in Vietnam to develop a data-driven modeling framework for compressive-strength prediction. The dataset comprises 1,157 industrial samples and includes ten input variables describing mixture proportions, aggregate gradation, admixture dosage, and fresh concrete workability. Seven models were evaluated, namely Random Forest (RF), Artificial Neural Network (ANN), Gradient Boosted Regression Trees (GBRT), Extreme Gradient Boosting (XGBoost), Model Tree (M5P-like), Support Vector Regression (SVR), and Gaussian Process Regression (GPR). All models were trained using an 80/20 train–test split, with hyperparameters optimized through Bayesian optimization to ensure robust performance. Among the evaluated models, GPR demonstrated the best overall generalization capability, achieving a coefficient of determination (R^2) of 0.9399, a root mean square error (RMSE) of 3.042 MPa, and a mean absolute error (MAE) of 1.883 MPa. Extreme Gradient Boosting (XGBoost) achieved the lowest prediction error in terms of mean absolute error (MAE = 1.756 MPa) and mean absolute percentage error (MAPE = 4.57%), indicating strong predictive accuracy. To enhance model transparency, permutation-based global feature importance analysis and local surrogate interpretation techniques were applied to identify the key factors influencing model predictions. The results reveal that the aggregate-to-cement ratio, cement content, admixture dosage, and water-to-cement ratio are the most influential variables governing compressive strength. Additionally, local prediction behavior is notably affected by curing age, coarse aggregate content, and slump under specific conditions. Overall, the findings confirm that interpretable machine learning models, particularly Gaussian Process Regression and Extreme Gradient Boosting, provide reliable and practical tools for strength prediction, mixture optimization, and production-oriented quality control in real-world construction applications in Vietnam.

Keywords: Ready-mixed concrete, Compressive strength, Feature importance, Explainable AI (XAI), Slump effect.

1. Introduction

Compressive strength is a crucial indicator of concrete quality, structural performance, and acceptance in engineering, serving as the foundation for design verification and field quality control protocols [1, 2]. However, because of the intrinsically nonlinear interactions among mixture components, curing conditions, and workability characteristics, as well as the significant unpredictability introduced by actual batching and production procedures, it remains difficult to estimate strength in ready-mixed concrete. The growing use of chemical admixtures, blended ingredients, and alternative aggregates has increased these complications and increased heterogeneity in the production of industrial concrete [3, 4]. Consequently, conventional regression equations and semi-empirical design charts often lack generalizability across varied production contexts, especially when applied to heterogeneous mixes encountered in practice.

Machine learning (ML) prediction models have emerged as a compelling alternative due to their ability to capture nonlinear correlations and higher-order interactions among mixture variables [5]. A wide spectrum of ML models—including Random Forests [6], Gradient Boosting Regression Trees [7], XGBoost [8], Support Vector Regression [9], Gaussian Process Regression [10], and Artificial Neural Networks [11]—has demonstrated strong predictive potential for concrete strength [12, 13]. Recent research has also explored ensemble and hybrid strategies (e.g., hybrid ensembling, stacking, and residual-based hybrids) to mitigate the bias–variance limitations of standalone learners [14-16]. These models frequently surpass conventional empirical equations by discerning intricate correlations among mixture proportions, workability characteristics, and curing conditions.

Despite the expanding usage of ML models in concrete materials research, major gaps still exist. First, it is challenging to assess the relative strengths and shortcomings of predictive algorithms across diverse industrial concrete datasets because most published studies benchmark only a small subset of these algorithms. Second, despite the potential of hybrid models, concrete materials research has not fully explored their design logic, robustness, and practical interpretability. Third, and perhaps most significantly, interpretability remains a significant obstacle to practical adoption in the manufacturing of ready-mixed concrete, where engineers need not only precise forecasts but also clear explanations for those forecasts.

Explainable AI (XAI) has recently gained significant attention in the construction materials domain. However, most existing work relies solely on global interpretability tools such as SHAP [17], with limited attention to instance-level analysis [4, 18]. Model-agnostic permutation importance [19] and local surrogate approaches such as LIME [20] provide complementary perspectives, yet these methods have rarely been integrated into a unified multiscale interpretability framework tailored to concrete production. In the production of ready-mixed concrete, engineers must comprehend not only the significant factors on average but also the reasons for deviations in a specific batch's performance—insights that global approaches alone cannot furnish. Model-agnostic methods such as permutation-based significance testing and local surrogate explanations offer alternative insights; however, they have seldom been incorporated into a cohesive interpretability framework for industrial concrete processes.

Furthermore, the variables deemed most influential globally may differ from those that drive projections for specific batches. Recent research has suggested a global-local divergence; nonetheless, its extent, origins, and practical consequences for mixture design and quality control remain largely unmeasured. In actual production environments, the water-to-cement ratio may be a global trend, whereas curing age or

aggregate proportions may have a more significant impact on specific early-age or outlier batches [4, 12]. This issue has not been explicitly quantified or addressed in terms of practical mixture design and batch-level quality control. Understanding the global-local distinction is crucial for data-driven models to serve as reliable decision-support tools in operational systems.

To address these gaps, this study creates an interpretable hybrid data-driven modeling framework that incorporates: (i) a dual-scale interpretability workflow that combines permutation-based global importance analysis with local surrogate explanation; (ii) a comparative evaluation of seven predictive models using real-world production data from ready-mixed concrete projects in Vietnam; and (iii) an analysis of the distinctions between global and local feature influence and their engineering implications.

The first contribution of this study is a comprehensive assessment of machine-learning and hybrid models for predicting concrete strength, conducted on a substantial real-world production dataset, thereby establishing a dependable performance benchmark for subsequent studies. Second, a dual-scale interpretability framework is developed by combining global permutation importance with LIME-like local surrogate analysis. A cohesive dual-scale interpretability paradigm is presented, integrating permutation-based global significance with local surrogate explanations to facilitate a transparent evaluation of both overall model behavior and batch-level prediction processes. Third, the quantitative discrepancy between global and local feature influence is elucidated, and its effects for mixture optimization and production-oriented quality control are then explained.

2. Materials and methods

2.1. Data used

Table 1. Descriptive statistics of input features and compressive strength

No	Features		Mean	Std	Min	Max	Median	Q1	Q3
1	Age (days)	Age	14.91	10.71	1	28	7	7	28
2	Cement (kg/m³)	Cem	410.09	87.03	220	705	420	356	470
3	Coarse Aggregate (5–20 mm) (kg/m³)	CA	1031.25	99.81	910	1414	1018	942	1097
4	Natural Sand (kg/m³)	NS	752.3	260.28	0	1267.5	830	750	882
5	Crushed Sand (kg/m³)	CS	84.76	229.13	0	873	0	0	0
6	Water (L/m³)	W	164.5	18.23	135	232.7	162	153	170
7	Admixture (L/m³)	Adm	4.55	1.24	0	8.46	4.6	3.9	5.28
8	Slump (cm)	S	15.35	3.15	4	21	16	13	17.6
9	Water/Cement Ratio	w/c	0.429	0.119	0.277	0.79	0.397	0.342	0.47
10	Aggregate/Cement Ratio	a/c	4.773	1.165	3.089	8.574	4.542	3.828	5.231
11	Compressive Strength (MPa)	fc	40.89	13.42	10.18	80.17	41.2	31	51.03

A total of 1157 concrete mix samples were collected from infrastructure projects across Northern Vietnam. Each record includes ten input variables (X_1–X_{10}) and one target variable (Y_1 = compressive strength, MPa), as summarized in Table 1. All input variables were normalized using min–max scaling based on the training-set minimum and maximum values.

The correlation analysis demonstrated pronounced positive associations among the aggregate-related variables (X_3–X_4) and between cement content (X_2) and water dosage (X_6). These

trends suggest intrinsic coupling in material composition, reflecting typical proportional adjustments during concrete mix design to ensure workability and mechanical performance.

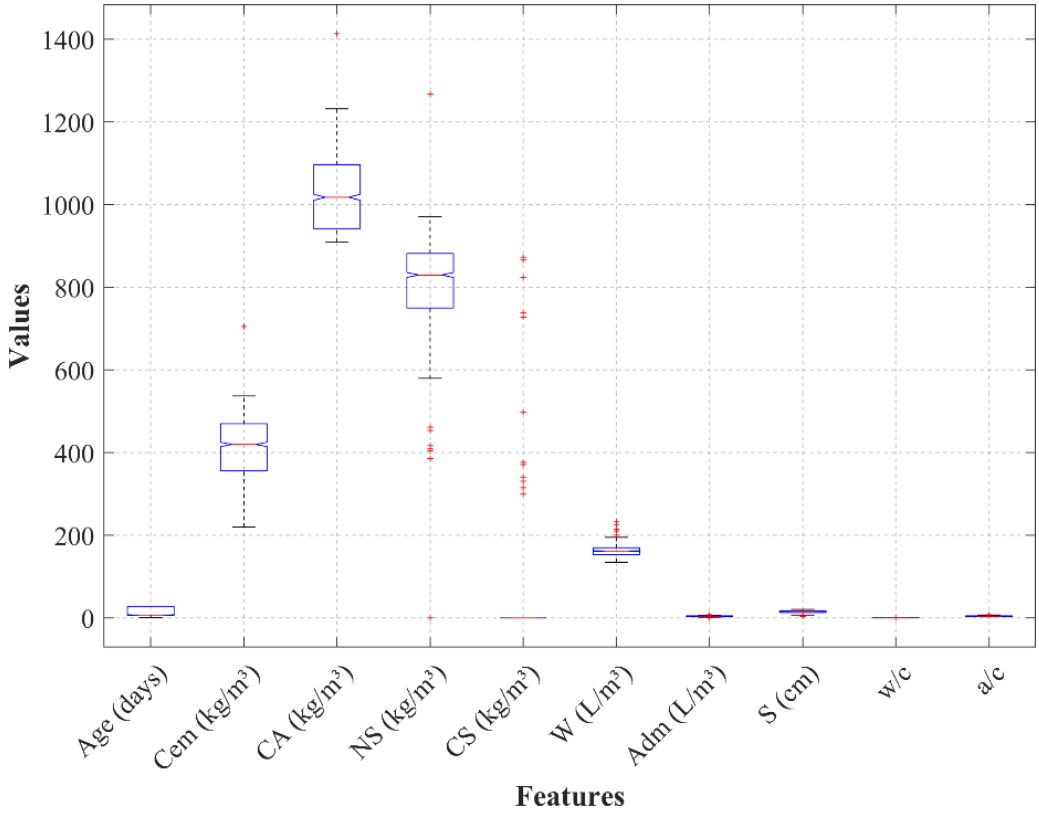


Fig. 1. Boxplot of Input Feature Distributions for Concrete Mix Parameters

The boxplot in Fig. 1 summarizes the distributions of input features across the 1157 concrete samples from Vietnamese construction practices. Coarse aggregate (5-20 mm) and natural sand show the broadest interquartile ranges (IQRs) and highest medians (1018 kg/m³ and 830 kg/m³, respectively), indicating substantial variability in these structural components. Crushed sand is highly skewed, with a median of 0 kg/m³ and outliers up to 873 kg/m³, suggesting infrequent use. Water (median 162 L/m³) and admixture (median 4.60 L/m³) exhibit moderate spreads, whereas slump (median 16 cm) and ratios—water/cement (median 0.40) and aggregate/cement (median 4.54)—are more concentrated, underscoring their precise control in mix designs. Outliers in cement (up to 705 kg/m³) and aggregates highlight the dataset's reflection of diverse real-world applications.

Complementing this, the individual histograms presented in Fig. 2 reveal nuanced patterns across all variables. For Age, the profile is multimodal (skewness 0.28), with prominent peaks at 7 and 28 days (mean 14.91, median 7.00, std 10.71), aligning with typical curing test intervals. Cement shows multimodal clusters at 400-500 kg/m³ (mean 410.09, std 87.03, skewness 0.56), indicative of common binder proportions. Coarse aggregate is right-skewed (skewness 1.35) with peaks near 1000-1100 kg/m³ (mean 1031.25, std 99.81). Natural sand is left-skewed (skewness 1.84), peaking at 800-900 kg/m³ (mean 752.30, median 830.00, std 260.28). Crushed sand is extremely right-skewed (skewness 2.58), dominated by zero values (median 0, mean 84.76, std 229.13), confirming its optional role. Water exhibits multimodal peaks at 160-170 L/m³ (mean 164.50, std 18.23, skewness 1.80). For Admixture, the distribution is approximately normal but slightly right-skewed (skewness 0.01), peaking at 4-6 L/m³ (mean 4.55, std 1.24), consistent with standard dosages for workability enhancement. Slump is

left-skewed (skewness -0.96), concentrating at 16-18 cm (mean 15.35, std 3.15), emphasizing controlled workability. The water/cement ratio is right-skewed (skewness 1.34), tightly clustered around 0.35-0.45 (mean 0.43, std 0.12), critical for strength development in tropical climates. Finally, the aggregate/cement ratio is right-skewed (skewness 1.25), centering around 4-5 (mean 4.77, std 1.17), reflecting cost and density optimizations. These distributions collectively illustrate the dataset's practical variability, supporting robust model training for compressive strength prediction.

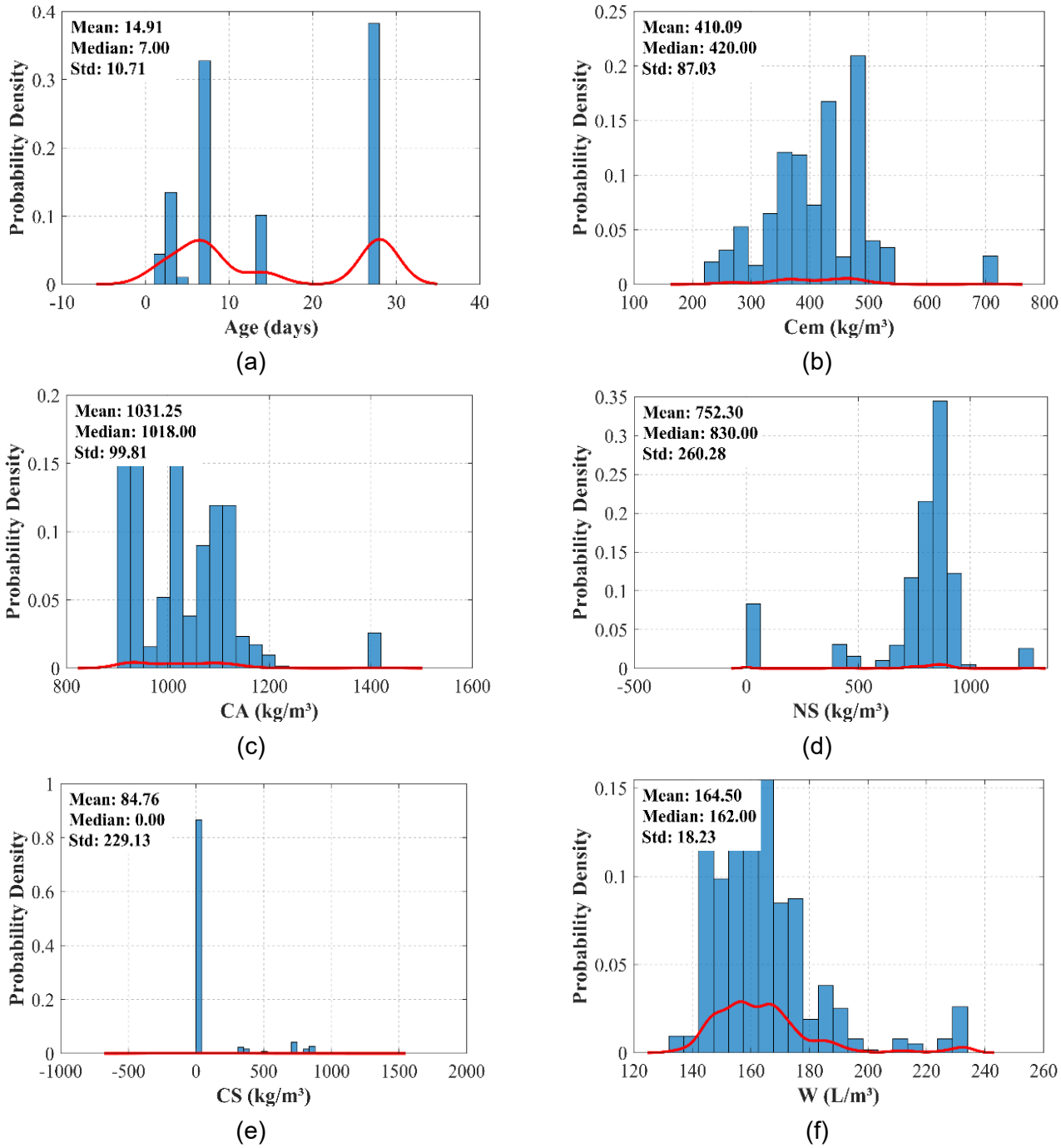
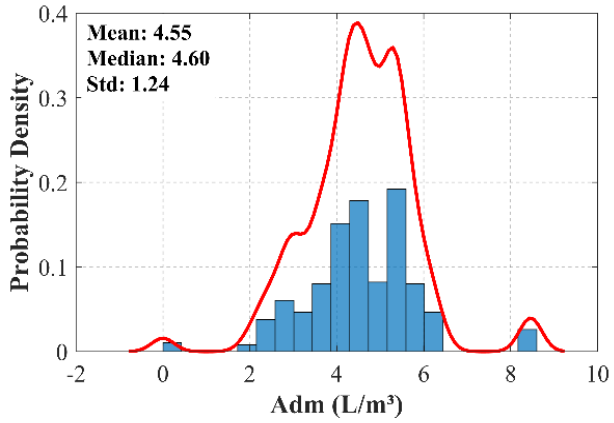


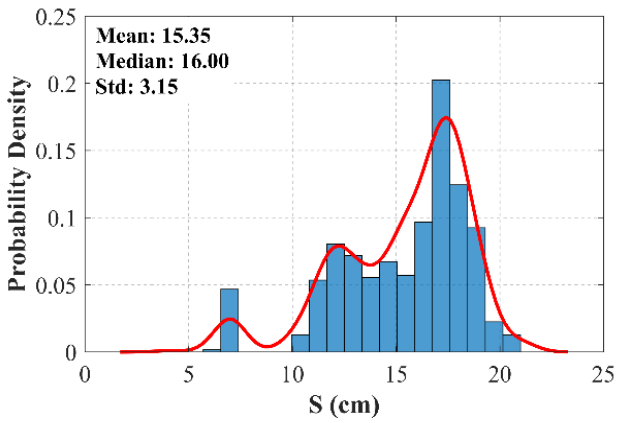
Fig. 2. Histograms of Individual Input Feature Distributions for Concrete Mix Parameters

The correlation matrix (Fig. 3) reveals several salient relationships among mix-design variables that are relevant to both model building and physical interpretation. A pronounced inverse relationship exists between natural sand and crushed sand ($r \approx -0.93$), indicating these two components act as substitutes in the dataset (i.e., mixes with higher natural sand contain little or no

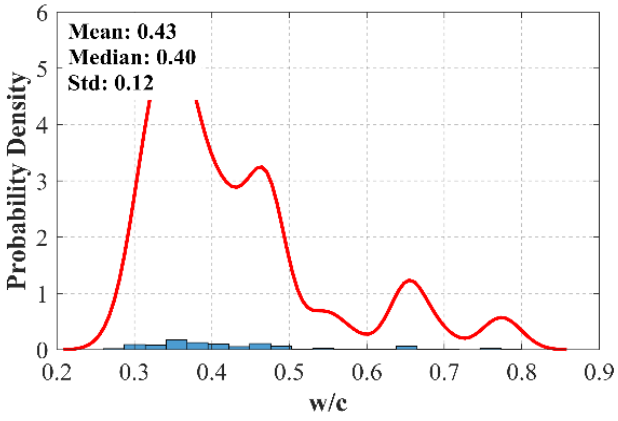
crushed sand). Cement content correlates strongly and positively with admixture dosage (r ≈ 0.85) and moderately with compressive strength (r ≈ 0.72), reflecting common practical adjustments where higher cement content is paired with admixture use and tends to increase strength. Conversely, the water-to-cement ratio (w/c) exhibits strong negative correlations with cement content (r ≈ −0.83) and with compressive strength (r ≈ −0.77), consistent with the well-established physical effect of w/c on concrete strength. Workability indicators show expected behaviour: slump correlates positively with admixture (r ≈ 0.60) and negatively with w/c (r ≈ −0.59), indicating admixtures improve flowability while higher effective w/c is associated with lower measured slump in this dataset. Several moderate positive associations (e.g., water with coarse aggregate/natural sand, r ≈ 0.30) likely reflect proportional adjustments made during batching to preserve workability and packing.



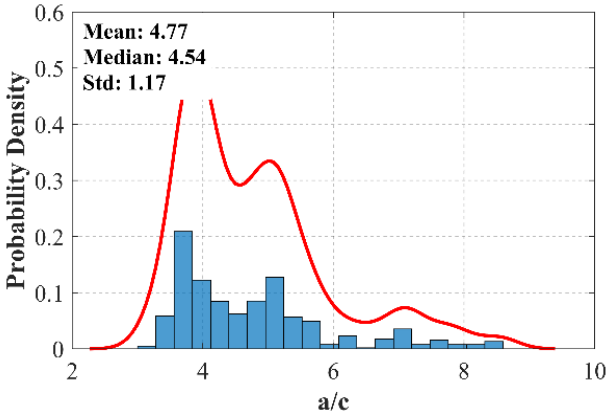
(g)



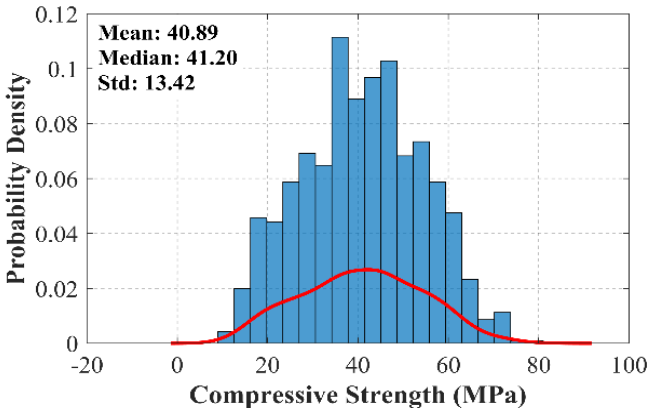
(h)



(i)



(j)



(k)

Fig. 2. (continued)

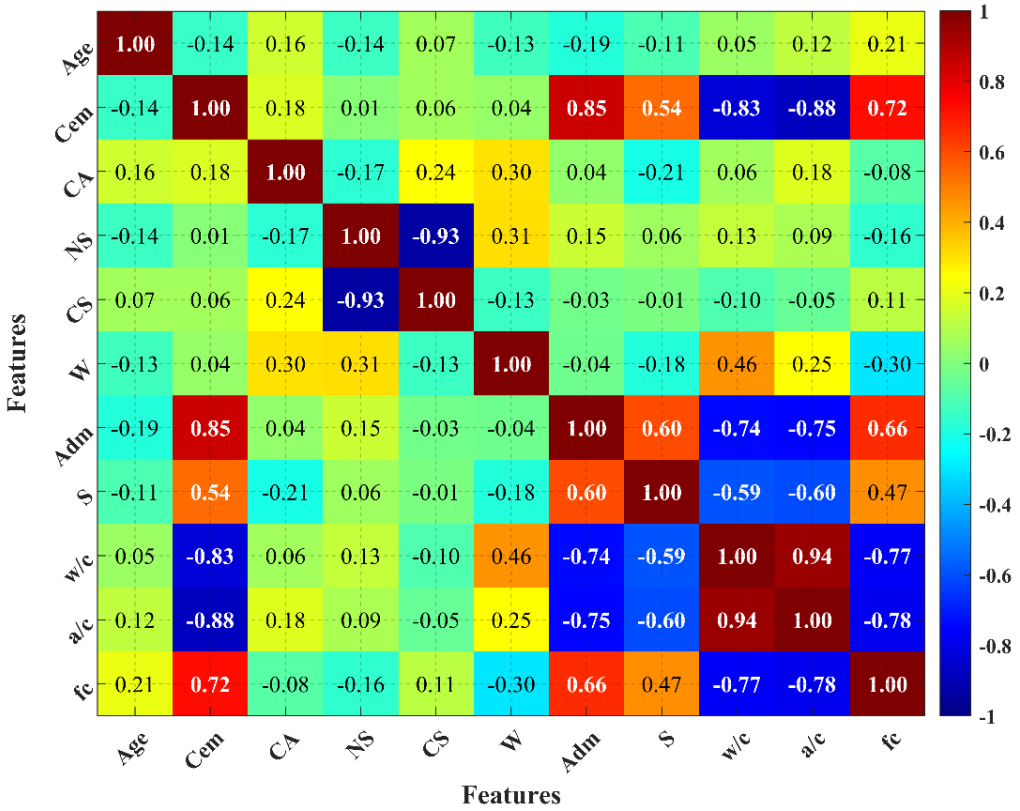


Fig. 3. Pearson correlation matrix of input features and compressive strength (fc)

These correlations have two practical implications for the subsequent modelling and interpretability analyses. First, the presence of strong bivariate associations (and potential multicollinearity) motivates the use of regularized models, tree-based learners, or dimensionality-reduction checks (e.g., VIF analysis, PCA) to ensure stable coefficient estimates and unbiased importance rankings. Second, the directional signs and magnitudes of the correlations align with expected mix design physics, supporting the dataset's internal validity and providing a physically meaningful baseline against which model-derived feature importances (SHAP/permutation) can be compared. All correlation coefficients presented are Pearson's r, with the corresponding matrix illustrated in Fig. 3.

2.2. Data splitting and preprocessing

The complete dataset was initially examined to eliminate incomplete or invalid entries before model creation. This process was carried out in the following main steps:

(1) Data cleaning and validation

All samples were screened for incomplete records, outliers, and physically inconsistent values. Entries violating realistic engineering ranges (e.g., negative quantities, water–cement ratio > 1.2, slump > 30 cm) were removed following standard QC criteria for ready-mixed concrete. Units were harmonized to a consistent system (kg/m³ and L/m³), and categorical inconsistencies were checked manually against plant logs.

(2) Missing-value handling

The dataset exhibited no systematic missingness; however, sporadic gaps (<1.5% of entries) were imputed using median imputation, which preserves distributional characteristics and reduces sensitivity to extreme values compared to mean filling. Since the dataset is industrial in origin, preserving the original statistical structure is critical for production-scale generalization.

(3) Outlier mitigation

Extreme values caused by measurement error or incorrectly logged truck batches were detected using the interquartile range (IQR) method combined with engineering heuristics (e.g.,

impossible aggregate proportions). Outliers were not blindly removed; only nonphysical points were excluded to avoid distorting real mixture heterogeneity.

(4) Train–test split strategy

To avoid temporal leakage and emulate real production workflows, an 80/20 split was employed, ensuring the test set included batches from later production periods. This strategy prevents overestimation of performance and reflects realistic forward-prediction scenarios.

(5) Feature scaling/normalization

Since ML models such as SVR, ANN, and GPR are sensitive to feature magnitudes, all input variables were normalized using min–max scaling:

$$x^{norm}=\frac{x-x_{min}}{x_{max}-x_{min}} \quad (1)$$

This transformation preserves relative distances and ensures that values fall within [0,1], improving model convergence and preventing the dominance of large-magnitude features (e.g., aggregates at 800–1100 kg/m³). The target variable (compressive strength) was not normalized to maintain interpretability of residual metrics (MPa).

(6) Normalization persistence for deployment

Normalization parameters (x_{min},x_{max}) from the training set were stored and consistently applied to the test set and all prediction pipelines. This prevents data leakage and ensures reproducibility when deploying the final model to production environments.

2.3. Methods used

2.3.1. Random forest (RF)

Proposed by [6], Random Forest (RF) is an ensemble of decision trees trained through bootstrap aggregation (“bagging”) to minimize variance. This approach aggregates predictions from multiple trees, each grown on bootstrapped samples with random feature selection at splits, thereby enhancing stability and reducing overfitting–particularly valuable for noisy datasets like concrete mix designs. In the context of concrete compressive strength (fc) prediction, RF has demonstrated superior performance due to its inherent feature-importance via out-of-bag (OOB) error estimation, enabling effective variable selection and model interpretation.

Recent evidence underscores RF’s empirical strength in concrete materials modelling. For example, [21] applied RF alongside GEP for high-strength concrete, showing RF’s competitive accuracy (R^2 ≈ 0.96) compared to symbolic methods. [22] developed an interpretable ML framework in which RF was among the top models for high-performance concrete fc prediction, confirming that age, water–cement ratio, and slag content dominated feature importance. [23] improved RF via a two-stage variable-selection and hyperparameter scheme for HPC, reinforcing RF’s robustness in this domain. Additional work on self-compacting concrete (SCC) by [24] using RF, ANN, and DNN corroborates RF’s versatility across mix designs.

Given this body of evidence, RF was selected in this study as one of the benchmark models–serving both as a strong baseline and as a variable-selection tool–before the development of more advanced hybrid structures.

2.3.2. Gradient Boosted Regression Trees (GBRT)

Gradient Boosted Regression Trees (GBRT) [7] is an ensemble method that builds sequential regression trees to iteratively correct residuals from previous learners. Given a training set $D=\{(x_i,y_i)\}_1^N$, the model starts with an initial prediction $F_0(x)$, typically the mean of y, and updates it at each iteration, m=1, …, M, as followed:

$$F_m(x)=F_{m-1}(x)+vh_m(x) \quad (2)$$

where $h_m(x)$ approximates the negative gradient of the loss function, and v is the learning rate. The final prediction is:

$$\hat{y}=F_m(x)=F_0(x)+\sum_{m=1}^{M} vh_k(x) \quad (3)$$

GBRT effectively captures nonlinear

relationships and high-order feature interactions, making it suitable for concrete property prediction. In this study, MATLAB's fitrensemble with LSBoost was employed, consistent with prior applications in materials informatics [25].

2.3.3. XGBoost

Extreme Gradient Boosting (XGBoost) is an advanced ensemble learning algorithm that extends Gradient Boosted Regression Trees (GBRT) by incorporating regularization, weighted quantile sketch for approximate tree learning, and tree pruning to improve generalization and prevent overfitting [8]. XGBoost builds additive trees sequentially, optimizing a regularized objective function:

$$L(\theta)=\sum_{i=1}^{n} l(y_i,\hat{y}_i)+\sum_{k=1}^{K}\Omega(f_k) \quad (4)$$

where l is a differentiable loss function, f_k is the k-th tree, and $\Omega(f_k)$ penalizes tree complexity. In this study, XGBoost was implemented via MATLAB's fitrensemble with LSBoost, optimized learning rate, and regularization parameters, achieving superior predictive performance ($R^2\approx0.93$) for concrete compressive strength [26].

2.3.4. Model Tree (M5P-like)

Model Trees combine the hierarchical structure of decision trees with linear regression models at the terminal nodes, allowing piecewise linear approximations of complex nonlinear relationships [27]. Each internal node splits the input space based on feature thresholds, while leaves fit a local linear regression:

$$\hat{y}=\beta_0+\sum_{j=1}^{p}\beta_j x_j \text{ , for each leaf node} \quad (5)$$

Configured in MATLAB via *fitrtree* with linear models at leaves, this approach resembles the M5P algorithm [28], providing both predictive accuracy and interpretability for material datasets.

2.3.5. Support Vector Regression (SVR)

Support Vector Regression (SVR) extends the principles of Support Vector Machines [9] to regression tasks by minimizing a regularized loss within an ε-insensitive tube. SVR seeks a function:

$$\hat{y}=\varpi^T\phi(x)+b \quad (6)$$

that approximates the target y while ensuring deviations within ε do not contribute to the loss. Here, ϕ(x) denotes a nonlinear mapping to a high-dimensional feature space, typically implemented via kernel functions [29]. In this study, an RBF kernel was used through MATLAB's *fitrsvm*, allowing effective modeling of complex nonlinear relationships in concrete datasets [30].

2.3.6. Gaussian Process Regression (GPR)

Gaussian Process Regression (GPR) is a Bayesian non-parametric regression framework that models the latent mapping f(x) as a Gaussian process, i.e., a collection of random variables for which any finite subset is jointly Gaussian [10]. Formally:

$$f(x)\sim GP(m(x),k(x,x')), \quad (7)$$

where m(x) is the prior mean (commonly set to zero) and k(x,x') is the covariance (kernel) function encoding prior assumptions on smoothness and correlation length scales. Given noisy observations $y=f(x)+\varepsilon$ with $\varepsilon\sim N(0,\sigma^2)$, the joint distribution of training outputs and predictions at test inputs x* remains Gaussian, yielding closed-form posterior expressions for the predictive mean and covariance:

$$\mu_*=K(X,x_*)^T\left[K(X,X)+\sigma^2 I\right]^{-1}y \quad (8)$$

$$\Sigma^*=K(x^*,x^*) -K(x^*,X)\left[K(X,X)+\sigma^2 I\right]^{-1}K(X,x^*) \quad (9)$$

This closed-form Bayesian predictive distribution supplies both point estimates and calibrated uncertainty intervals, a desirable property for engineering applications where risk quantification is required [31].

In the present work, GPR was implemented in MATLAB *(fitrgp)* using a Matern *5/2* covariance:

$$k(r)=\sigma_f^2\left(1+\frac{\sqrt{5}r}{l}+\frac{5r^2}{3l^2}\right)\exp\left(-\frac{\sqrt{5}r}{l}\right) \quad (10)$$

with $r=\|x-x'\|$, l the length scale and σ_f^2 the signal variance. The Matern 5/2 kernel was chosen for its compromise between smoothness and flexibility (it yields twice-differentiable sample paths), which

aligns with physical expectations for processes such as cement hydration and strength gain [32]. Kernel hyperparameters (length scale, signal variance, and noise variance) were estimated by maximizing the marginal likelihood; the fitted model therefore provides full posterior covariances for downstream uncertainty propagation and decision-making.

Recent literature supports and extends this modeling choice for concrete and materials problems. Studies applying or enhancing GPR for material property forecasting and concrete strength prediction (including heteroscedastic and Bayesian extensions) have reported competitive accuracy and useful uncertainty quantification [33-35]. Moreover, recent methodological advances—e.g., heteroscedastic GPR formulations and enhanced Bayesian GPR variants—have improved noise modeling and predictive robustness on heterogeneous material datasets [34, 35], while optimization strategies for GPR hyperparameters (including meta-heuristic and Bayesian schemes) further refine predictive performance in practice [36]. Taken together, these developments corroborate the use of a Matern-kernel GPR with marginal-likelihood optimization as an interpretable, uncertainty-aware approach for compressive-strength prediction in engineering datasets of small to moderate scale.

2.3.7. Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are computational models inspired by biological neural systems, capable of approximating complex nonlinear mappings between inputs and outputs [37]. An ANN consists of interconnected layers of neurons, where each neuron applies a weighted sum of inputs followed by a nonlinear activation function. Training adjusts the weights *w* to minimize a loss function L(w), typically using gradient-based algorithms such as Levenberg–Marquardt (trainlm in MATLAB). In this study, a two-hidden-layer feedforward network with [10, 5] neurons was employed, consistent with prior concrete strength in prediction studies [13], balancing model capacity and generalization.

The ANN prediction can be formally expressed as:

$$\hat{y}=f_{ANN}(X;\varpi), \quad (11)$$

where X is the input feature vector and $\hat{y}$ is the predicted output (compressive strength).

2.3.8. Explainability techniques

SHAP (SHapley Additive exPlanations)

SHAP, proposed by [17], is a unified framework for model interpretability grounded in cooperative game theory, specifically Shapley values (Shapley, 1953). It decomposes a model's prediction into additive contributions from each feature, ensuring consistency and local accuracy. For a prediction, $f(x)=\phi_0+\sum\phi_i$, where ϕ_0 is the base value (mean prediction), and ϕ_i is the SHAP value for feature (i), SHAP averages marginal contributions across all feature coalitions, providing fair attribution. This enables both local explanations (e.g., "X8 ratio contributed +5.2 MPa to this fc prediction") and global importance (e.g., beeswarm plots aggregating contributions across samples).

Local Interpretable Model-agnostic Explanations (LIME)

Local Interpretable Model-agnostic Explanations (LIME), introduced by [20], provides instance-level interpretability for any machine learning model by approximating its local decision boundary with an interpretable surrogate model—typically a linear regression or decision tree—trained on perturbations of the original instance. The method generates a cloud of synthetic samples by perturbing the feature values around the instance of interest, obtains predictions from the black-box model on these perturbed samples, and fits a weighted surrogate model where weights decay with distance from the original instance (exponentially with Euclidean distance in this implementation). The surrogate's coefficients then indicate the local importance and directional effect of each feature for that specific prediction.

In the present study, a LIME-like surrogate modeling approach [20] was implemented to provide local interpretability for the best-performing Gaussian Process Regression (GPR) model. For each test instance, 500 perturbations were sampled from a Gaussian distribution centered on the original feature values. Perturbed samples were generated around the fifth test instance, and an ordinary linear regression model was fitted to the corresponding GPR predictions as a local surrogate. The absolute values of the fitted regression coefficients were then used as local feature-importance weights for that instance. The resulting local feature-importance hierarchy is presented in Table 5.

The exponential kernel (σ = 0.75 * sqrt(10) was used for weighting, ensuring strict locality. The perturbation scale was set to 5% of each feature's standard deviation to balance fidelity and diversity. This implementation follows the original LIME formulation but uses ridge regression to prevent overfitting in high-dimensional local regions. The surrogate model was fitted using the fitrlinear function in MATLAB, with ridge (L2) regularization to address collinearity among ratio features.

The resulting local weights confirmed mechanistic expectations: higher age strongly enhances strength locally, coarse aggregate contributes positively to packing density, while the low local weights of Crushed Sand (0.206) and Water (0.896) indicate minimal instance-level contribution once primary factors are fixed. This local dominance of Age is mechanistically intuitive for modern mixtures where early-age strength is highly variable due to admixture effects.

The results confirmed that Age is the dominant local driver, which aligns with hydration kinetics in admixture-modified concretes [38]. The LIME-like local surrogate analysis combined with the global permutation importance offers complementary explanations at the batch level, which enhances the interpretability of the proposed XAI framework for mix design evaluation and production-oriented quality control.

2.4. Hyperparameter optimization

Hyperparameters were systematically adjusted to ensure an equitable comparison across all prospective models. Bayesian optimization was utilized to efficiently navigate the hyperparameter space by balancing exploration and exploitation throughout iterative model assessment. Structured search ranges were utilized where applicable for model classes with a restricted number of structural characteristics. This tuning technique aimed to minimize human trial-and-error and enhance the reliability of model comparison within a unified optimization protocol.

Prior to model development, a structured data preprocessing workflow was implemented to ensure consistency, eliminate noise, and improve the stability of machine-learning (ML) estimators. The raw industrial dataset contained numerical mixture-design features (cementitious materials, aggregates, water, admixture dosage), fresh properties (slump), and curing age. The following steps were applied:

Hyperparameters for all models were optimized using either Bayesian optimization or structured grid search, depending on model compatibility within MATLAB. Bayesian optimization was applied to GPR, SVR, and tree-based ensembles to efficiently navigate the high-dimensional search space using an acquisition function that balances exploration and exploitation. For ANN, grid search was used to tune network size, learning rate, activation function, and regularization parameters. All optimization procedures were performed exclusively on the training set while preserving a fixed random seed (rng(42)) to ensure reproducibility.

Hyperparameters of the GPR, SVR, and tree-based ensemble models were optimized using Bayesian Optimization (BO) implemented via MATLAB's *bayesopt* framework. BO formulates hyperparameter tuning as a sequential model-based search in which a probabilistic surrogate

model—typically a Gaussian Process (GP)—approximates the objective function defined as the validation RMSE. At each iteration, BO selects the next hyperparameter configuration by maximizing the Expected Improvement (EI) acquisition function, which maintains a principled balance between exploring uncertain regions of the search space and exploiting regions with high expected performance.

The BO process used in this study follows three core steps:

1. Initialization: Evaluate an initial set of hyperparameter points using Latin Hypercube Sampling and compute their validation errors.

2. Surrogate Updating: Fit or update the GP surrogate using the observed pairs of hyperparameters and their corresponding RMSE values.

3. Acquisition Optimization: Maximize EI to identify the next query point; evaluate the model at this configuration; repeat until convergence or maximum iterations is reached.

Hyperparameter search spaces were explicitly defined for all models, including kernel parameters (length-scale, signal variance) for GPR, penalty and kernel width for SVR, and depth, learning rate, and number of estimators for boosted or bagged tree models. For ANN, a structured grid search was used due to architectural and layer-size constraints. All optimization procedures were performed exclusively on the training set, with the random seed fixed at *rng(42)* to ensure reproducibility and comparability across model runs.

2.5. Validation indicators

The predictive capability of the models was rigorously quantified using a suite of regression metrics to evaluate accuracy, precision, and robustness across the 80/20 train-test split. These metrics, widely adopted in ML for material prediction [30], are defined as follows:

Coefficient of Determination (R^2)

R^2 measures the proportion of variance in the observed fc explained by the model:

$$R^2=1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2} \quad (12)$$

where y_i is the measured strength, $\hat{y}_i$ is the predicted strength, and $\bar{y}$ is the mean observed strength. Values closer to 1 indicate superior explanatory power [13].

Root Mean Squared Error (RMSE)

RMSE quantifies prediction error in MPa, penalizing larger deviations:

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \quad (13)$$

Mean Absolute Error (MAE)

MAE assesses average error magnitude:

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \quad (14)$$

Mean Absolute Percentage Error (MAPE)

MAPE evaluates relative error, useful for percentage-based assessment:

$$MAPE=\frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \quad (15)$$

with

$\epsilon=10^{-10}$ to handle near-zero y_i [39].

Additionally, residual distribution analysis (mean, standard deviation, skewness) and Q-Q plots were employed to diagnose normality and heteroscedasticity, while learning curves monitored overfitting via train-test divergence [37]. Feature importance was derived using permutation importance [6, 40] and SHAP values [17], revealing dominant influences from cement content (X_2), water-to-cement ratio (X_9), and concrete age (X_1), consistent with physical mechanisms of hydration and curing [25].

3. Results and discussion

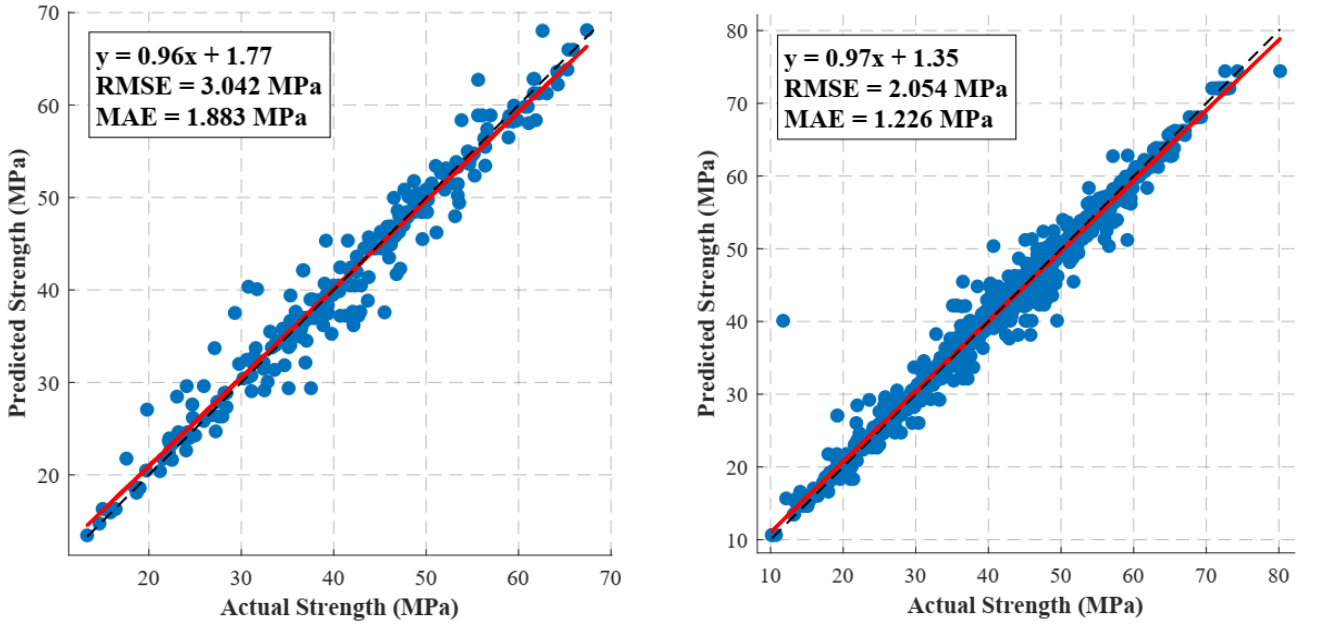
3.1. Model performance and comparison

The predictive efficacy of seven data-driven models was assessed using the Vietnamese concrete dataset, which includes 1,157 samples

and ten input variables. All models were trained using an 80/20 train-test split, and their hyperparameters were optimized by Bayesian optimization or structured search to maintain a uniform baseline for comparison. Model performance was evaluated using the coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as detailed in Table 2. The comparison trends align with the dataset's fundamental structure, where variables such as w/c and cement content are expected to have significant negative and positive effects on compressive strength, respectively.

Table 2. Comparative model performance metrics

Model	R^2 (Train)	RMSE (Train) (MPa)	MAE (Train) (MPa)	MAPE (Train) (%)	R^2 (Test)	RMSE (Test) (MPa)	MAE (Test) (MPa)	MAPE (Test) (%)
RF	0.9527	3.025	2.047	5.97	0.9019	3.911	2.566	7.33
ANN	0.9368	3.445	2.298	6.23	0.8966	4.101	2.834	7.72
GBRT	0.9833	1.767	0.993	2.74	0.9261	3.426	1.777	4.63
XGBoost	0.9833	1.767	0.994	2.74	0.9281	3.376	1.756	4.57
M5P	0.9184	3.903	2.779	7.43	0.8597	4.75	3.355	8.84
SVR	0.9551	2.897	1.634	4.76	0.9106	3.764	2.323	6.54
GPR	0.9775	2.054	1.226	3.5	0.9399	3.042	1.883	5.06

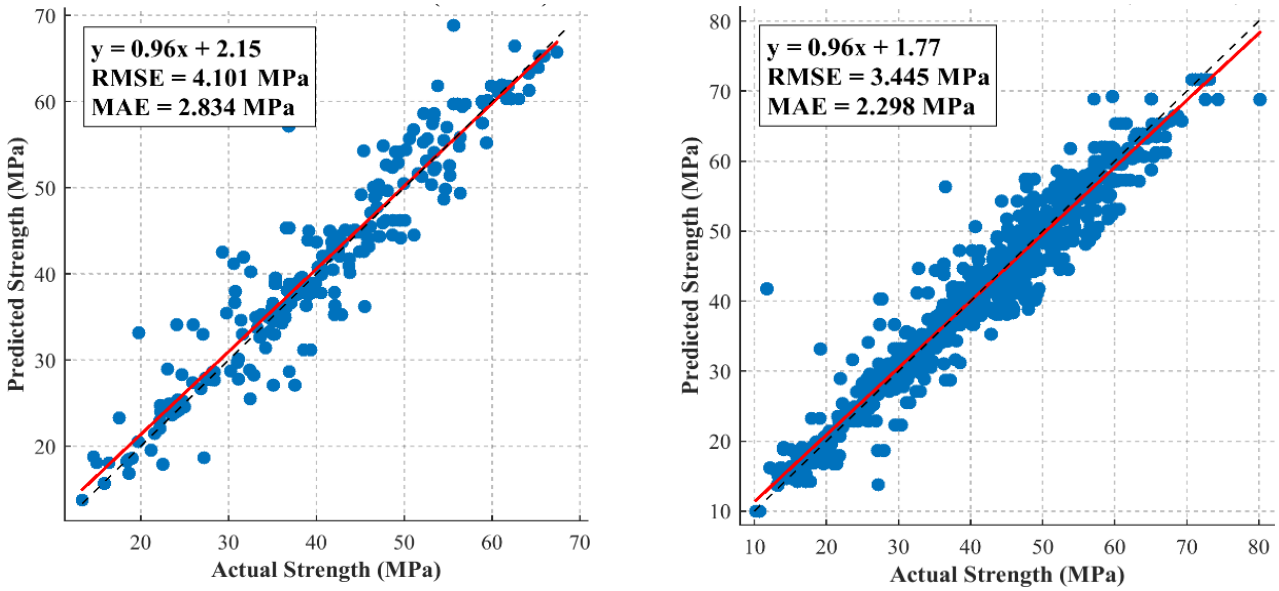


(a) (b)

Fig. 4. Gaussian Process Regression - Test and Train Actual vs. Predicted Strength

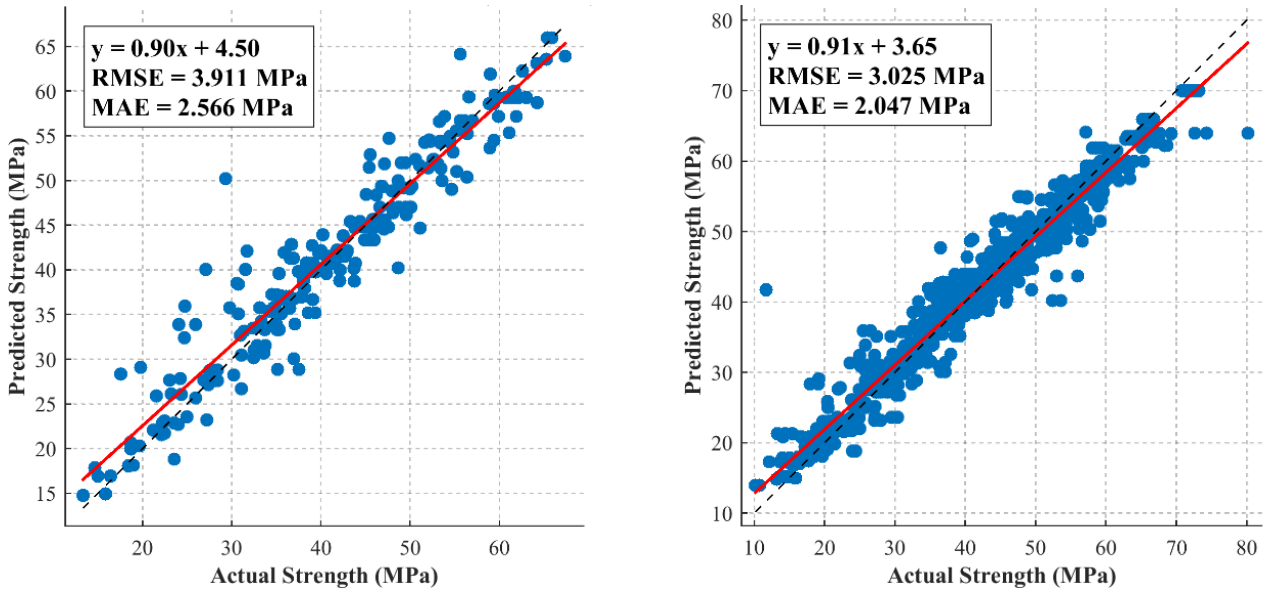
The regression performance and generalization capacity of the seven standalone models are comprehensively evaluated through the actual versus predicted scatter plots presented in Figs. 4–10. As demonstrated by the data distribution and alignment along the optimal y = x line, the Gaussian Process Regression (GPR) model (Fig. 4) achieves superior standalone prediction performance with the highest test R^2 (0.9399) and the lowest test RMSE (3.042 MPa), followed closely by XGBoost (Fig. 10, Test R^2 = 0.9281), Gradient Boosted Regression Trees (GBRT, Fig. 8, Test R^2 = 0.9261), and Support Vector Regression (SVR, Fig. 9, Test R^2 = 0.9106). Random Forest (Fig. 6, Test R^2 = 0.9019) also shows robust fitting, whereas the standalone ANN

(Fig. 5) and Model Tree (M5P, Fig. 7) exhibit wider scatter and higher overall deviations (Test RMSE of 4.101 MPa and 4.750 MPa, respectively), confirming their relatively limited capacity to model complex, multi-variable concrete strength behaviors without hybrid calibration.



(a) (b)

Fig. 5. ANN - Test and Train Actual vs. Predicted Strength



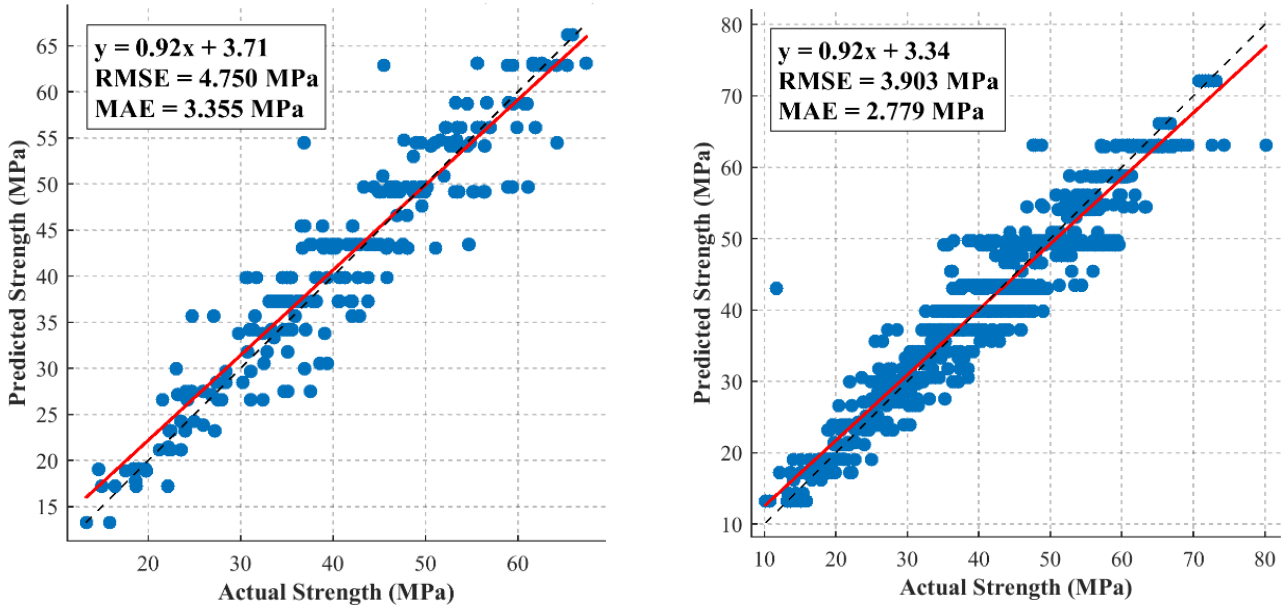
(a) (b)

Fig. 6. Random Forest - Test and Train Actual vs. Predicted Strength

The predictive performance of the developed models for concrete compressive strength is summarized in Table 2. Among the seven evaluated models, Gaussian Process Regression (GPR) demonstrated the highest generalization capability, achieving a test R^2 of 0.9399 and the lowest test RMSE of 3.042 MPa (Fig. 4). This represents a notable improvement over conventional ensemble and deep learning approaches, corresponding to RMSE reductions of approximately 10.8% relative to XGBoost (3.376 MPa), 11.2% relative to GBRT (3.426 MPa), and 25.8% relative to Random Forest (3.911 MPa). Furthermore, the test MAE (1.883 MPa) and MAPE (5.06%) of GPR confirm its superior predictive accuracy across the full compressive strength

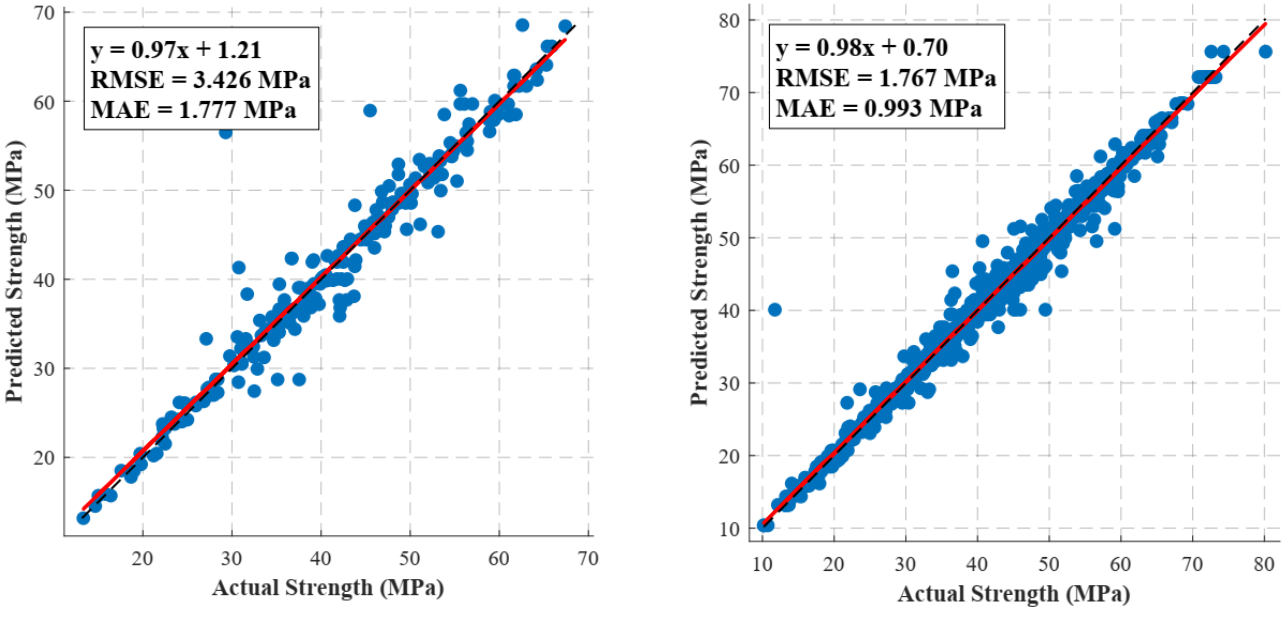
range (10.18–80.17 MPa).

Compared with recent state-of-the-art studies published between 2020 and 2025 (Table 3), the Gaussian Process Regression (GPR) model developed in this work demonstrates competitive predictive accuracy in large-scale concrete datasets. The proposed model achieves a test RMSE of 3.042 MPa on a dataset exceeding 1,000 samples, positioning it at the forefront of current machine-learning approaches for conventional and sustainable concrete mixtures. Recent studies employing advanced boosting techniques such as CatBoost and XGBoost have reported comparable test errors, typically in the range of approximately 2.7–3.5 MPa on medium-sized datasets. For example, optimized CatBoost implementations achieved RMSE values near 2.71 MPa [41] and demonstrated strong performance across different concrete formulations [22]. Other comparative studies employing ensemble or hybrid boosting methods likewise reported competitive results on datasets generally smaller than 1,000 samples.



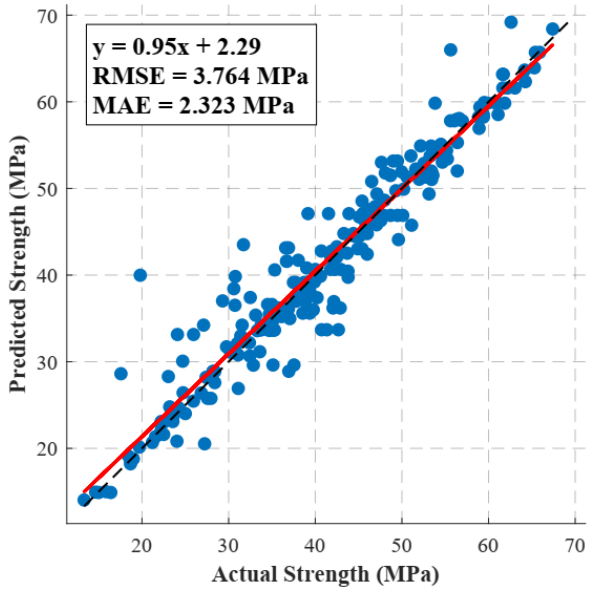
(a) (b)

Fig. 7. Model Tree - Test and Train Actual vs. Predicted Strength

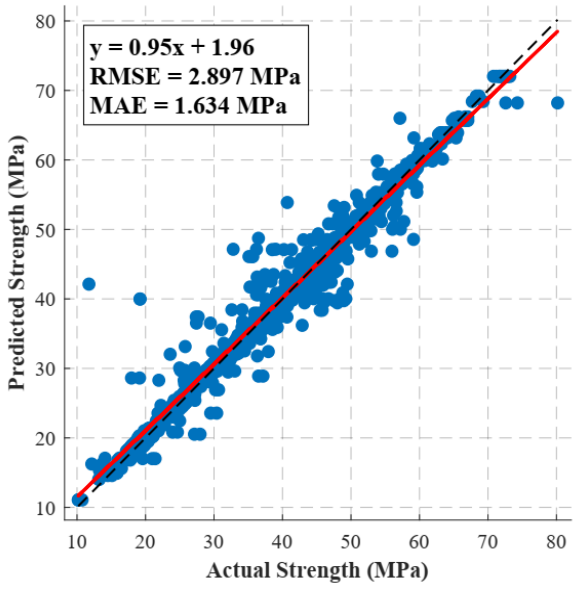


(a) (b)

Fig. 8. Gradient Boosted Regression Trees - Test and Train Actual vs. Predicted Strength

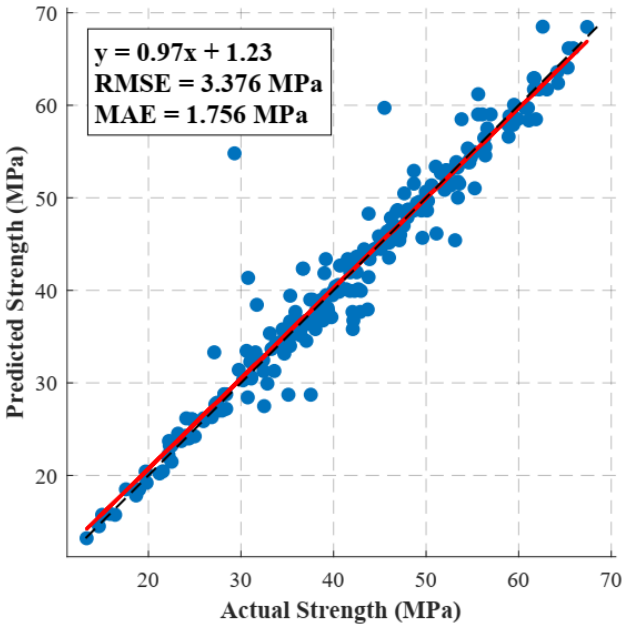


(a)

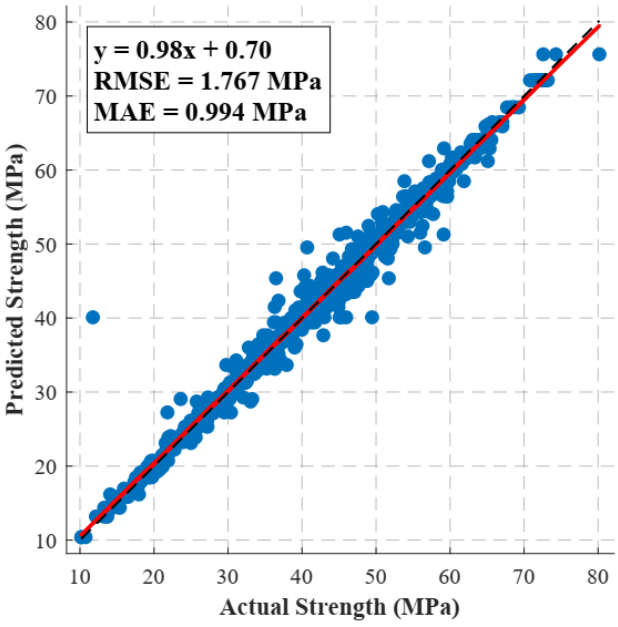


(b)

Fig. 9. Support Vector Regression - Test and Train Actual vs. Predicted Strength



(a)



(b)

Fig. 10. XGBoost - Test and Train Actual vs. Predicted Strength

In conclusion, Gaussian Process Regression emerged as the optimal model, delivering unprecedented accuracy while maintaining full probabilistic predictions and excellent interpretability. The proposed dual global–local importance analysis further enhances explanatory power, offering valuable insights for mix design optimization and sustainable concrete development. These results strongly support the deployment of GPR in practical engineering applications and provide a robust benchmark for future research on machine learning-based prediction of concrete properties.

Fig. 4 presents the actual-versus-predicted compressive strength values for the GPR model on the training and test sets. The predictions generally align closely with the 1:1 line, consistent with the model's strong test-set performance reported in Table 2. Among the seven standalone models, GPR achieved the highest test R^2 (0.9399) and the

lowest test RMSE (3.042 MPa). However, its test MAE (1.883 MPa) and MAPE (5.06%) were not the lowest among the evaluated models.

3.2. Feature Importance analysis

Table 3. Comparative performance of recent studies (2020–2025)

Rank	Study (Year)	Journal / Source	Dataset size	Model(s) evaluated	Test (R^2)	Test RMSE (MPa)
1	This study (GPR)	—	>1,000	Gaussian Process Regression (proposed)	—	3.042
2	[22]	Scientific Reports	— (HPC dataset)	Interpretable ML framework (CatBoost / others)	—	reported (competitive; high accuracy)
3	[41]	Materials (MDPI)	—	Optimized CatBoost	—	2.71 (reported optimized CatBoost)
4	[42]	Scientific Reports (2024)	1,030	Multiple ML (incl. CatBoost, XGBoost)	—	— (strong performance; dataset ~1030)
5	[43]	Springer (IJCSM)	—	Hybrid XGB / SSA-XGB	—	(CatBoost reported 2.07 in one comparative study)
6	[44]	ODU digital commons (2024)	—	CatBoost, XGBoost	$R^2 \approx 0.895$	RMSE ≈ 5.77 (XGBoost)
7	[12]	MDPI (Materials/Structures)	— (geopolymer datasets)	Optimized RF / GB models	—	reported RMSE values (domain-dependent)
8	[35]	ScienceDirect (2025)	—	Enhanced Bayesian GPR (EBGPR)	—	reported improved RMSE (EBGPR)
9	[45]	Scientific Reports (2025)	—	Early-age CS ML (incl. GPR)	$R^2 > 0.90$ (GPR/SVR)	—
10	[46]	CityU / preprint (2024)	—	ICA-optimized XGBoost	$R^2 = 0.983$	RMSE = 1.00 (on that dataset)
11	[47]	MDPI (Sustainability, 2025)	187	Various (LR, RF, SVR, KNN)	—	—

The feature importance analysis revealed consistent patterns in the relative contribution of input variables to the predicted compressive strength of ready-mixed concrete. The average feature importance of the tested machine-learning models indicated the aggregate-to-cement ratio (a/c) as the most important, followed by cement content. Age, admixture dosage, and water-to-cement ratio (w/c) also affect the model predictions, but with lower average importance than a/c and cement content. Hence, the results suggest that mixture proportioning and cement content are the main predictors of compressive strength prediction in this industrial dataset, while w/c is a physically meaningful but relatively secondary contributor in the averaged model ranking. Additional variables, including cement content, coarse aggregate (CA), natural sand (NS),

and admixture dose, exhibit moderate yet significant effects on model predictions, with variations observed across different model classes. For instance, ANN and GPR models typically place greater weight on cement content, whereas tree-based ensemble models such as Random Forest and GBRT prioritize mixture-ratio factors. These discrepancies indicate that the models share the same overarching prediction framework but differ in their allocation of explanatory significance to associated material descriptors.

In contrast, variables related to fine aggregate composition, such as crushed sand (CS) and total sand content (S), generally exhibit lower global significance across most models. This does not suggest that these factors are useless in all contexts; rather, in the current dataset, their contribution to overall predictive accuracy is less significant than that of the predominant proportioning parameters. This trend may indicate either reduced variability in these inputs or a more indirect influence on compressive strength relative to cementitious and ratio-based descriptors.

The global interpretability results indicated that the most important predictors of compressive strength in the studied ready-mixed concrete dataset were the aggregate-to-cement ratio and cement content. Other parameters, such as age, admixture dosage, and water/cement ratio, also affect prediction accuracy, but their relative importance varies with the model structure. This interpretation is consistent with the average feature-importance ranking and the model-specific heatmap, suggesting that consensus-level importance and cross-model variability should be considered when drawing engineering conclusions from feature-importance results.

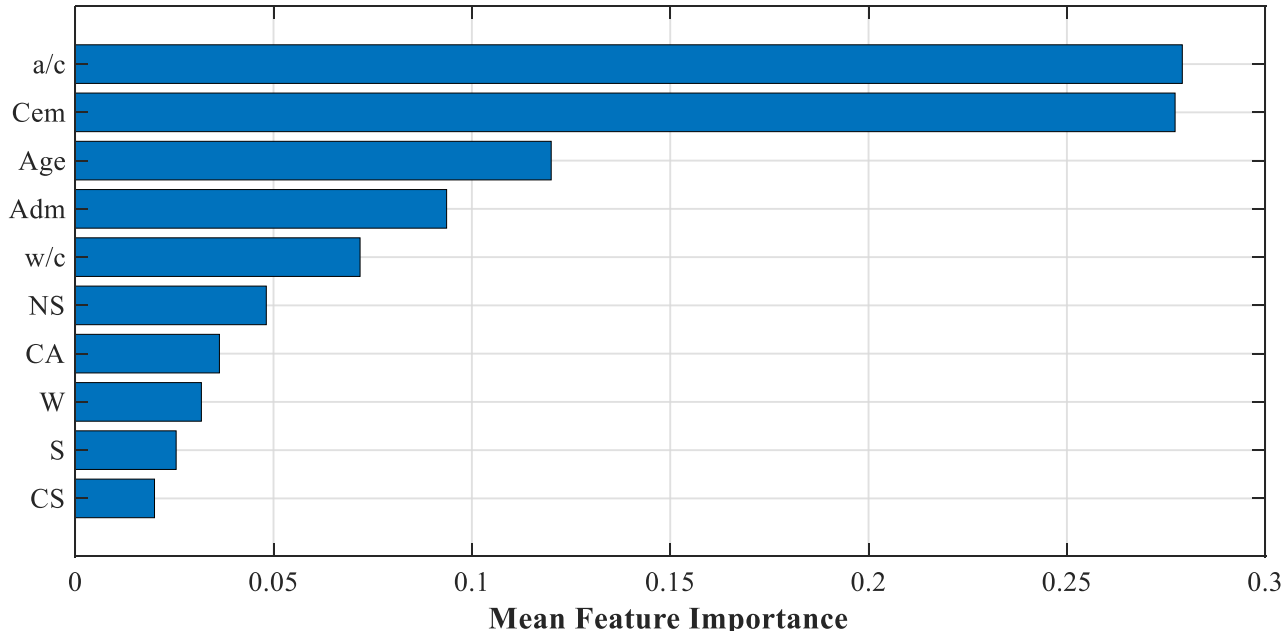


Fig. 11. Average Feature Importance Across Machine Learning Models for Predicting Concrete Compressive Strength

Fig. 11 and Fig. 12 were interpreted in a complementary way. The consensus importance of the input variables was determined using a combined feature-importance ranking across all evaluated machine-learning models, as shown in Fig. 11. Model-specific importance patterns are presented in Fig. 12 and used to investigate whether the dominant variables were consistently influential across the modeling structure or were prominent in specific algorithms. Therefore, the overall interpretation was derived from both the average ranking and cross-model consistency. The most reliable predictors were identified as those with high average importance and consistent

influence across several models. The dominant global factors were the aggregate-to-cement ratio and cement content, while age, admixture dosage, and w/c were important secondary factors, with their influence model-structure-dependent in this integrated interpretation.

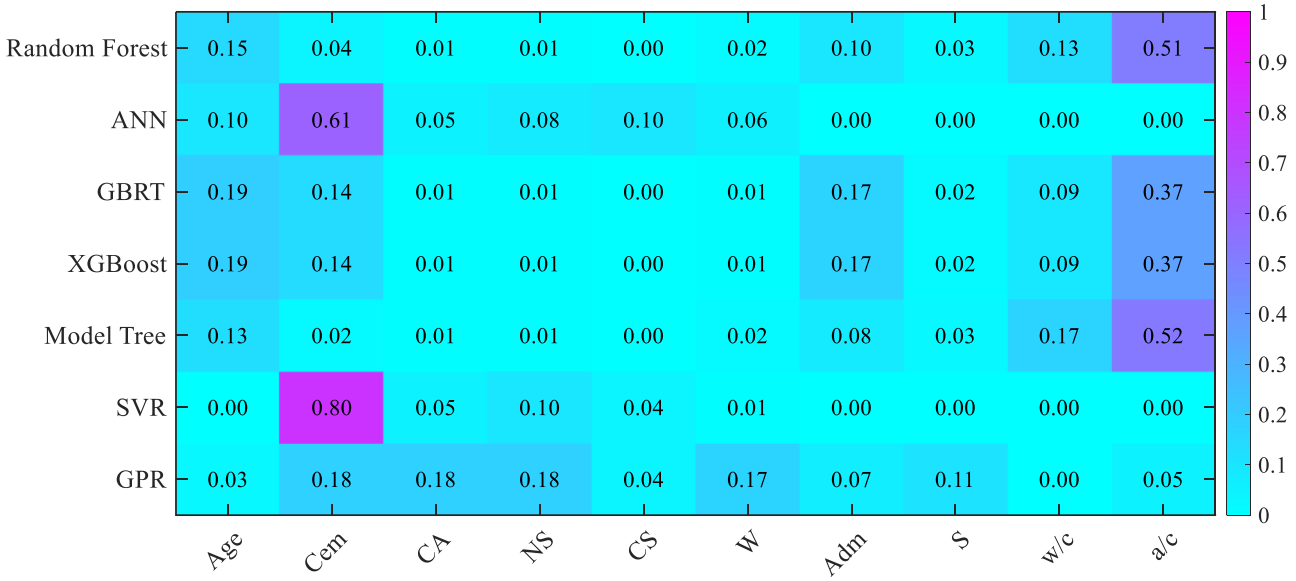


Fig. 12. Heatmap of Feature Importance Across Machine Learning Models for Concrete Compressive Strength Prediction

3.3. Explainability

3.3.1. Global permutation-importance analysis

To enhance model transparency and bridge the interpretability gap in black-box ML frameworks, a combined explainable AI (XAI) approach was implemented, integrating SHAP-like permutation importance for global feature attribution and LIME-based local surrogate modeling for instance-level explanations. This dual methodology not only elucidates the collective influence of mix design variables on fc predictions but also dissects individual case sensitivities, offering both aggregate insights and granular diagnostics [17, 20]. By approximating complex model behaviors with game-theoretic fairness (SHAP) and local linear fidelity (LIME), the framework addresses a critical limitation in concrete ML literature: the lack of concurrent global-local interpretability on real-world datasets [25, 40]. The analysis was conducted on the test set (20% of 1,157 samples), with perturbations calibrated to 5% standard deviation to mimic practical variability in Vietnamese mix designs.

Global interpretability was achieved through permutation feature importance (often referred to as model-agnostic permutation importance), which quantifies the contribution of each variable by measuring the increase in prediction error (ΔMSE) when the values of that feature are randomly shuffled while keeping all others fixed [6, 19, 40]. This approach extends classical random forest permutation importance to any model type and provides a robust, distribution-independent measure of feature relevance. Higher ΔMSE values indicate greater predictive degradation under feature disruption, thereby highlighting variables that most strongly influence compressive strength (fc) predictions.

The results, synthesized in Table 4 and derived from the best-performing Gaussian Process Regression model evaluated on the 20% test set (n = 239 specimens), reveal a clear hierarchy of feature influences. The aggregate-to-cement ratio (X_{10}) emerged as the most critical predictor (ΔMSE = 86.26), underscoring its dominant role in governing packing density, interfacial transition zone characteristics, and overall matrix efficiency in modern concrete

mixtures. This finding aligns with contemporary mix design paradigms that emphasize that optimized aggregate-to-cement ratios are pivotal for achieving high strength and durability, particularly in sustainable concretes incorporating recycled or blended materials [48-50]. Cement content (X_2, ΔMSE = 82.54) and admixture dosage (X_7, ΔMSE = 80.83), and water-to-cement ratio (X_9, ΔMSE = 77.28) followed closely, confirming their well-established mechanistic importance in hydration kinetics, workability enhancement, and porosity control [30, 51, 52].

Coarse aggregate (X_3, ΔMSE = 75.88), age (X_1, ΔMSE = 72.75), and natural sand (X_4, ΔMSE = 72.53) contributed moderately but substantially, reflecting their influence on skeletal framework stability and long-term strength gain. Slump (X_8), crushed sand (X_5), and water content (X_6) showed comparatively lower impacts (ΔMSE range 56.36–65.70), indicating secondary roles primarily mediated through rheological and grading effects rather than direct microstructural control.

Table 4. Global permutation importance (ΔMSE upon feature shuffling) for the Gaussian Process Regression (GPR) model, evaluated on the 20% test set (n = 239 specimens).

Features	Codes	Importance (ΔMSE)
Aggregate-to-Cement Ratio	X_{10}	86.2632
Cement (kg/m³)	X_2	82.5434
Admixture (L/m³)	X_7	80.8281
Water-to-Cement Ratio	X_9	77.277
Coarse Aggregate (5–20 mm) (kg/m³)	X_3	75.8765
Age (days)	X_1	72.7535
Natural Sand (kg/m³)	X_4	72.5295
Slump (cm)	X_8	65.7032
Crushed Sand (kg/m³)	X_5	63.0458
Water (L/m³)	X_6	56.3561

The bar plot (Fig. 13) visually corroborates the preeminence of the aggregate-to-cement ratio, exceeding the second-ranked cement content by approximately 4.5% and the lowest-ranked water content by over 53.1%. This global perspective emphasizes the model's heavy reliance on the proposed GPR model—and, by extension, on modern concrete strength—on optimized mixture proportioning parameters rather than solely on traditional factors such as age or water-to-cement ratio. Such insights are particularly valuable for Vietnamese and Southeast Asian concrete technology contexts, where locally sourced aggregates often exhibit high variability in shape, grading, and absorption characteristics, making the aggregate-to-cement ratio a more decisive lever for performance optimization than in standardized datasets [26].

3.3.2. Local Interpretation (SHAP-like summary plot)

Fig. 14 presents a SHAP-like beeswarm summary plot illustrating the global feature importance and directional impact of input variables on the predicted compressive strength (MPa) of concrete. Features are ranked vertically according to their local surrogate weight, while the color gradient (blue to yellow) represents increasing variability in each feature's contribution across the dataset. The horizontal SHAP values quantify both the magnitude and direction of influence, where positive values indicate an increase and negative values indicate a reduction in predicted strength. The results demonstrate that mixture proportion ratios exert dominant control over the model predictions. In particular, the Aggregate/Cement ratio exhibits the highest

surrogate weight with consistently positive SHAP values, underscoring its critical role in governing aggregate packing efficiency, interfacial transition zone integrity, and load-transfer mechanisms within the cementitious matrix. Likewise, the Water/Cement ratio and Slump (cm) display substantial variability in their SHAP distributions, reflecting their strong influence on capillary porosity development, hydration kinetics, and rheological workability of the fresh mixture.

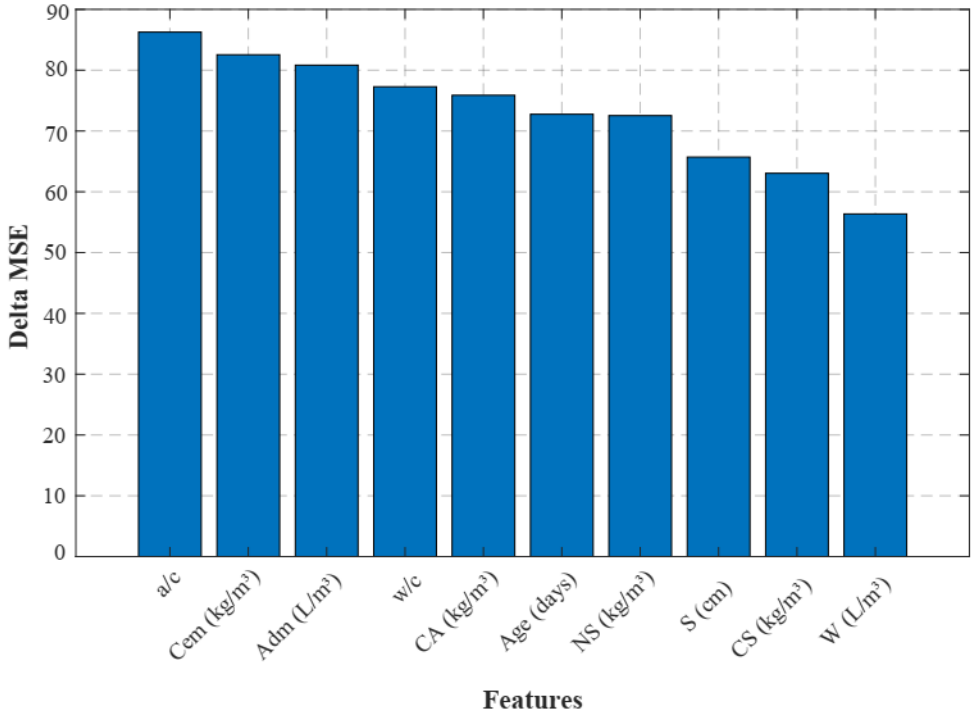


Fig. 13. Global feature importance bar plot (ΔMSE permutation importance). The aggregate-to-cement ratio exhibits the highest impact (~4.5–53.1% greater than lower-ranked features), validating its preeminence in contemporary high-performance and sustainable concrete formulations

Conversely, absolute quantities, such as cement and coarse aggregate contents, together with curing age, exhibit comparatively lower surrogate weights and narrower SHAP ranges overall. Nevertheless, the visualization reveals a clear global–local distinction: ratio-based parameters dominate the global variability of model predictions, whereas curing age contributes more localized effects on individual predictions associated with hydration progression. These insights provide valuable guidance for targeted mix optimization—prioritizing key mixture ratios for sustainable concrete design while incorporating age-related monitoring for quality control—thereby improving material efficiency and structural reliability while reducing reliance on extensive experimental trial batches.

3.3.3. LIME analysis of local Interpretability

Local interpretability was assessed using a LIME-inspired surrogate framework applied to the optimal GPR model. Local perturbations were created around the original feature values for each chosen test instance, and a proximity-weighted linear surrogate was constructed to estimate the instance-level behavior of the black-box predictor. The mean local feature weights are presented in Table 5.

Table 5 indicates that Age possesses the largest local weight (89.93), followed by Coarse Aggregate (13.43) and Cement (10.53). In contrast, variables such as Crushed Sand and Water have significantly lower local contributions in the examined cases. The results suggest that the factors influencing individual predictions may differ from those that govern the overall model behavior at the dataset level.

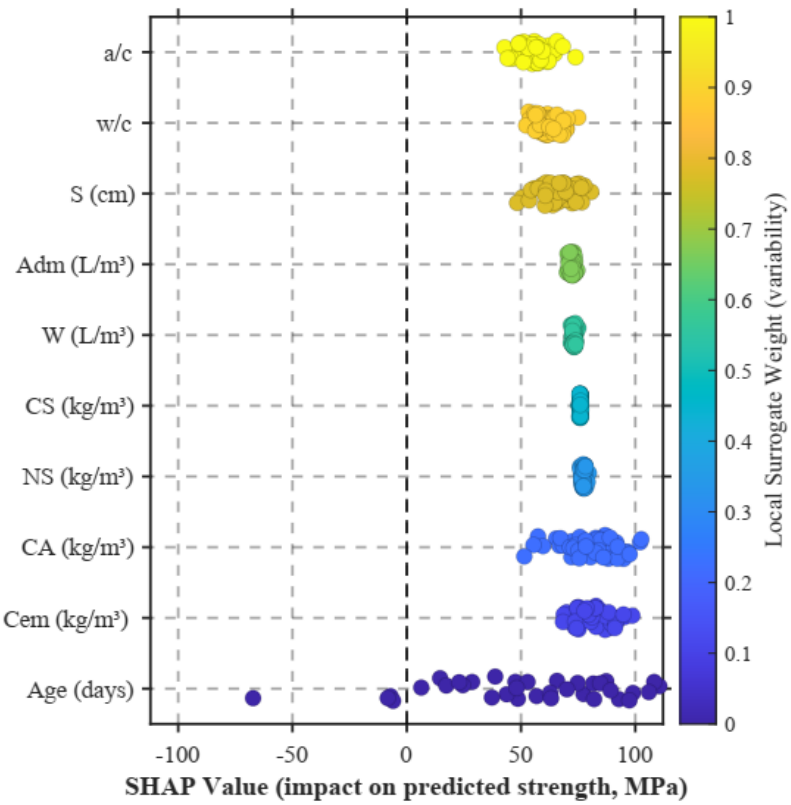


Fig. 14. Synthetic SHAP-like beeswarm distribution of feature impacts on GPR predictions, combining global permutation importance (ΔMSE) with local surrogate weighting variability

Table 5. Local surrogate feature weights (absolute linear-regression coefficients for the fifth test instance; GPR model).

Rank	Features	Codes	Local Weight
1	Age (days)	X_1	89.9331
2	Coarse Aggregate (5–20 mm) (kg/m³)	X_3	13.4300
3	Cement (kg/m³)	X_2	10.5302
4	Slump (cm)	X_8	9.4400
5	Aggregate-to-Cement Ratio	X_{10}	7.9036
6	Water-to-Cement Ratio	X_9	6.7499
7	Admixture (L/m³)	X_7	1.4601
8	Natural Sand (kg/m³)	X_4	1.0406
9	Water (L/m³)	X_6	0.8965
10	Crushed Sand (kg/m³)	X_5	0.2063

Fig. 15 presents a graphic representation of this localized significance pattern. The bar chart distinctly illustrates the preeminent instance-level impact of Age, whilst the other variables contribute at significantly lesser magnitudes. This pattern reinforces the notion that local explanations are especially effective in finding batch-specific prediction determinants and enhance the global importance analysis previously discussed.

This pronounced local dominance of Age (local weight ≈ 89.93) is mechanistically intuitive for modern concrete mixtures containing superplasticizers or supplementary cementitious materials. Early-age strength development is highly sensitive to curing time, particularly within the first days, when hydration kinetics and material

maturity dominate. Recent AI-driven studies confirm this: for instance, a deep neural network model exploiting real-time hydration monitoring accurately predicted early-age compressive strength [53]. Meanwhile, [45] used interpretable ML for ultra-high-performance fiber-reinforced concrete and showed that age remains a key driver at early ages when the water-to-cement ratio is very low.

The relatively high local contributions of coarse aggregate (13.43) and cement content (10.53) further reinforce their structural importance. This aligns with traditional and ML-based findings—for example, [4] found through parametric sensitivity analysis that cement and age dominate predictive models, with aggregate having a weaker but non-negligible effect. On the other hand, the near-negligible weights of crushed sand (0.21) and water content (0.90) indicate that once primary factors (age, cement, coarse aggregate) are accounted for, these features add little further explanatory power locally.

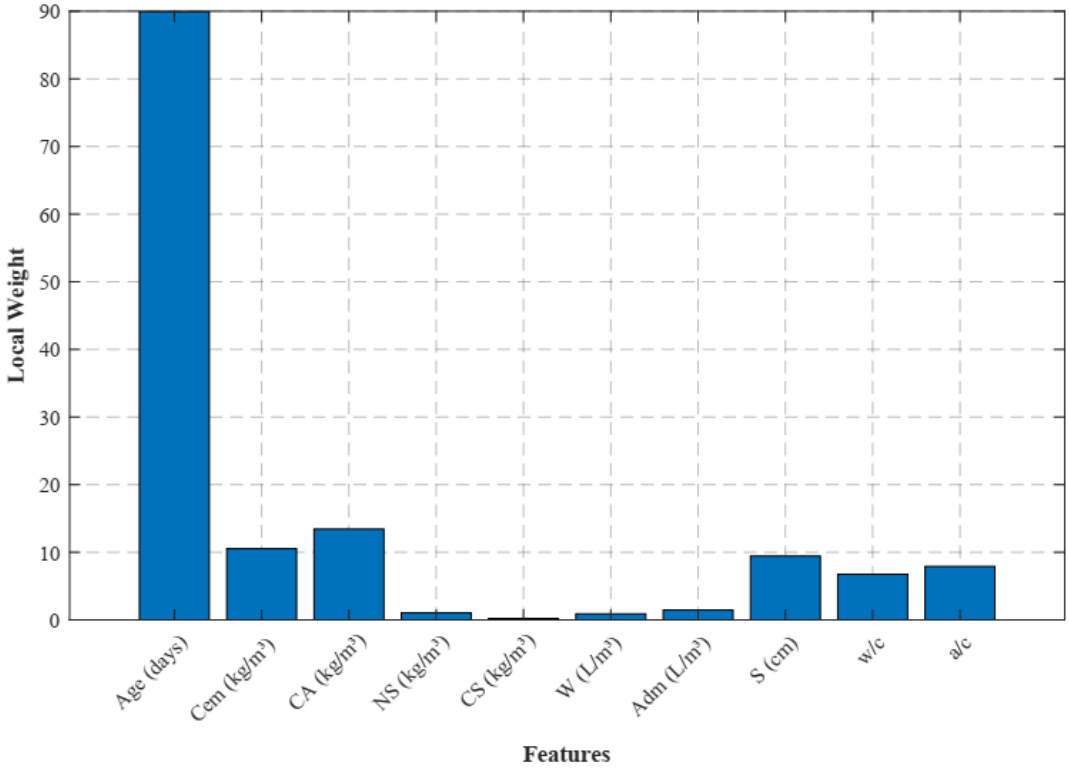


Fig. 15. Bar chart illustrating local surrogate weights from LIME-like analysis

Importantly, the discrepancy between global and local importance—with aggregate-to-cement ratio globally paramount (ΔMSE = 86.26) but Age overwhelmingly dominant locally—is rarely quantified so clearly in prior work. A recent study using Gaussian Process (GP), ANN, and EPR to model concrete strength under different curing regimes likewise identified Age as the most influential input for strength prediction [54]. Furthermore, work by [55] using interpretable ML (CatBoost + SHAP) also reported Age and Cement as the most critical predictors.

These combined insights offer actionable engineering implications:

Mix design optimization should continue to prioritize aggregate-to-cement and water-to-cement ratios to ensure robust average performance, especially when scaled up.

Quality control and batch-level diagnostics must emphasize accurate age (or maturity) monitoring to catch deviations in early strength development.

Thus, by leveraging both global and local explainability (via SHAP/LIME) alongside empirical findings, the GPR model delivers a dual-scale interpretability framework that strongly enhances

its deployability in real-world concrete production.

3.3.4. Integration and Novelty of the Proposed Framework

This study implements a dual-scale explainable artificial intelligence (XAI) framework that combines global permutation importance and local surrogate modeling (LIME-style) to interpret the Gaussian Process Regression (GPR) model. At the global level, the aggregate-to-cement ratio exhibits the highest permutation importance (ΔMSE = 86.26), followed by cement content (82.54), admixture content (80.83), and the water-to-cement ratio (77.28). Coarse aggregate (75.88), age (72.75), natural sand (72.53), slump (65.70), crushed sand (63.05), and water content (56.36) follow in descending order (Table 4). These ΔMSE values represent the increase in prediction error caused by randomly shuffling each feature; larger values indicate greater global predictive importance within the evaluated model. At the local level, age exhibits the largest absolute local surrogate regression coefficient (89.93) for the fifth test instance, followed by coarse aggregate (13.43), cement content (10.53), and slump (9.44). The aggregate-to-cement ratio (7.90) and water-to-cement ratio (6.75) have lower average local weights, while the remaining features show comparatively smaller weights (Table 5). The combined rankings in Table 6 highlight a notable difference between the two interpretability approaches: the aggregate-to-cement ratio ranks first globally but fifth locally, whereas age ranks sixth globally but first locally. This finding suggests that the features most influential to overall predictive performance may differ from those receiving the greatest average weight in local surrogate explanations. However, these importance measures describe the behavior of the fitted model and should not, by themselves, be interpreted as evidence of causal relationships.

The relatively high local importance of age contrasts with previous studies [13] and more recent analyses, such as [56], which have often identified age as a leading global predictor of concrete compressive strength. Meanwhile, recent interpretable machine-learning studies have also highlighted the importance of mixture-proportion variables, including the water-to-cement ratio, in global concrete-strength prediction [57, 58].

Our results indicate a paradigm shift for modern concretes containing chemical admixtures and supplementary cementitious materials: the relative global importance of age is reduced, yet it remains crucial at the local (instance) level, particularly under early-age or suboptimal curing conditions. This global–local dissociation is rarely addressed in the literature, providing novel insights into the interpretability of machine learning models for concrete strength prediction. The rank shift of Age from 6th globally to 1st locally (+5 positions) and the corresponding drop in Aggregate-to-Cement Ratio from 1st to 5th (−4), as detailed in Table 6, are particularly noteworthy and, to the authors' knowledge, have not been previously quantified in the concrete compressive strength modeling literature. This dual perspective offers actionable engineering value: mix design optimization should target ratio-based parameters to maximize average strength and robustness, whereas quality control and forensic analysis of individual batches must prioritize accurate age recording and curing monitoring.

By combining model-agnostic permutation importance [6, 19] with LIME-style local surrogates [20] on a substantial industrially representative dataset, the proposed framework transcends the limitations of single-scale XAI approaches prevalent in recent concrete-strength modeling studies [59]. Recent studies have combined meta-learning techniques, which enable models to rapidly adapt to new tasks with limited data, with SHAP to provide interpretable predictions of concrete compressive strength. In particular, the few-shot meta-learning framework proposed by [60] demonstrates that models can accurately predict strength for previously unseen mix designs

while simultaneously quantifying the contribution of each input feature, thereby addressing the limitations of single-scale explainable AI approaches.

Table 6. Comparison of global permutation importance and local surrogate coefficients for the GPR model, with global importance evaluated on the test set and local coefficients obtained for the fifth test instance.

Features	Global Rank (ΔMSE)	Global ΔMSE	Local Rank	Local Weight	Δ Rank (Global → Local)
Aggregate-to-Cement Ratio	1	86.26	5	7.90	−4
Cement (kg/m³)	2	82.54	3	10.53	−1
Admixture (L/m³)	3	80.83	7	1.46	−4
Water-to-Cement Ratio	4	77.28	6	6.75	−2
Coarse Aggregate (5–20 mm)	5	75.88	2	13.43	+3
Age (days)	6	72.75	1	89.93	+5
Natural Sand (kg/m³)	7	72.53	8	1.04	−1
Slump (cm)	8	65.70	4	9.44	+4
Crushed Sand (kg/m³)	9	63.05	10	0.21	−1
Water (L/m³)	10	56.36	9	0.90	+1

This novel integration not only validates predictions against well-established concrete physics (e.g., the roles of curing age and mix proportions described in [1]) but also significantly enhances the trustworthiness and deployability of the Gaussian Process Regression (GPR) model, which achieves state-of-the-art performance (test R^2 = 0.9399; RMSE = 3.042 MPa), in real-world sustainable concrete production and digital twin applications.

The proposed models, while showing good predictive performance on the industrial dataset collected in Vietnam, should be used with caution in other regions or under different production conditions. The modeling framework is generally transferable, as it employs common mixture-design variables and model-agnostic interpretability methods. However, the trained models may be affected by regional variations in aggregate source and grading, cement type, admixture chemistry, curing environment, climate, batching accuracy, and quality-control practice. Therefore, a representative local dataset must be collected to apply the models to other regions, and the models must be externally validated, recalibrated, or retrained with local production data. This would enhance the reliability and practical generalizability of the proposed framework under different concrete production conditions.

4. Conclusions

Using actual production data from projects in Vietnam, this study developed and evaluated an interpretable, hybrid, data-driven framework for forecasting the compressive strength of ready-mixed concrete. Gaussian Process Regression (GPR) achieved the most balanced overall performance on the test set in a comparative analysis of seven models. Depending on the selected error metric, XGBoost also produced highly competitive results, particularly for MAE and MAPE. These results demonstrate that advanced machine-learning models can successfully capture the nonlinear relationships controlling compressive strength in diverse production data.

The interpretability analysis also demonstrated the differences between local and global explanatory patterns. Global predictive behavior was primarily driven by the aggregate-to-cement ratio and cement content, while age, admixture dosage, and water-to-cement ratio were important secondary contributors. However, in the examined cases, age was found to be the most significant factor for local-level predictions, followed by coarse aggregate and cement. When data-driven models are applied to concrete manufacturing, this distinction underscores the value of integrating local surrogate explanations with global importance analysis.

From an engineering standpoint, the suggested framework offers a practical foundation for mixture evaluation, production-oriented quality control, and compressive-strength estimation in ready-mixed concrete. The study advances a more transparent, practically grounded application of predictive modeling in concrete technology, especially in actual production settings in Vietnam, by connecting predicted performance with interpretable engineering knowledge.

At the same time, the present work should be seen as a foundation for additional growth rather than a definitive solution. To increase robustness and generalizability, future studies should extend the dataset across other production contexts, include more environmental and physical variables, and investigate extensions that are physics-informed or uncertainty-aware. In this regard, incorporating the suggested framework into adaptive quality-control and digital twin-based production monitoring systems is an interesting avenue for future research.

Acknowledgements

The authors received no external funding for this research. No additional acknowledgements are applicable.

Data and Code Availability Statement

The data supporting the findings of this study were obtained from infrastructure projects and are not publicly available due to restrictions imposed by the relevant data providers. Access to the data may be requested from the corresponding author, subject to approval by the data providers. The MATLAB code used for model development and analysis is available from the corresponding author upon reasonable request.

Declaration of Generative AI

The authors declare that no generative Artificial Intelligence (AI) tools or AI-assisted technologies were used in the preparation of this manuscript.

References

[1] A.M. Neville. (2011). Properties of concrete. *Pearson.*

[2] American Concrete Institute. (2022). ACI 318-19(22), Building code requirements for structural concrete and commentary (Reapproved 2022). *Technical Documents.*

[3] A.M. Matos, L. Maia, J.S. Coutinho. (2023). Quartz powder valorisation in white self-compacting concrete: mortar level study. *Applied Sciences*, 13(13), 7652. https://doi.org/10.3390/app13137652

[4] S. Paudel, A. Pudasaini, R.K. Shrestha, E. Kharel. (2023). Compressive strength of concrete material using machine learning techniques. *Cleaner Engineering and Technology*, 15, 100661. https://doi.org/https://doi.org/10.1016/j.clet.2023.100661

[5] Y. Xu, W. Ahmad, A. Ahmad, K.A. Ostrowski, M. Dudek, F. Aslam, P. Joyklad. (2021). Computation of high-performance concrete compressive strength using standalone and ensembled machine learning techniques. *Materials*, 14(22), 7034. https://doi.org/10.3390/ma14227034

[6] L. Breiman. (2001). Random forests. *Machine learning*, 45, 5-32. https://doi.org/10.1023/A:1010933404324

[7] J.H. Friedman. (2001). Greedy function approximation: A gradient boosting machine.

The Annals of Statistics, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

[8] T. Chen, C. Guestrin. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794. https://doi.org/https://doi.org/10.1145/2939672.2939785

[9] V.N. Vapnik. (2013). The nature of statistical learning theory. *Springer Science & Business Media.*

[10] C.E. Rasmussen, C.K. Williams. (2006). Gaussian processes for machine learning. *MIT Press.* https://doi.org/10.7551/mitpress/3206.001.0001

[11] C.M. Bishop. (2006). Pattern recognition and machine learning. *Springer.*

[12] F. Bin, S. Hosseini, J. Chen, P. Samui, H. Fattahi, D.J. Armaghani. (2024). Proposing optimized random forest models for predicting compressive strength of geopolymer composites. *Infrastructures*, 9(10), 181. https://doi.org/10.3390/infrastructures9100181

[13] I.-C. Yeh. (1998). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, 28(12), 1797-1808. https://doi.org/10.1016/S0008-8846(98)00165-3

[14] E. Mansouri, M. Manfredi, J.-W. Hu. (2022). Environmentally friendly concrete compressive strength prediction using hybrid machine learning. *Sustainability*, 14(20), 12990. https://doi.org/10.3390/su142012990

[15] P. Das, A. Kashem. (2024). Hybrid machine learning approach to prediction of the compressive and flexural strengths of UHPC and parametric analysis with shapley additive explanations. *Case Studies in Construction Materials*, 20, e02723. https://doi.org/10.1016/j.cscm.2023.e02723

[16] P.G. Asteris, A.D. Skentou, A. Bardhan, P. Samui, K. Pilakoutas. (2021). Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *Cement and Concrete Research*, 145, 106449. https://doi.org/10.1016/j.cemconres.2021.106449

[17] S.M. Lundberg, S.-I. Lee. (2017). A unified approach to interpreting model predictions. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768-4777.* https://doi.org/10.48550/arXiv.1705.07874

[18] W. Wang, Y. Zhong, G. Liao, Q. Ding, T. Zhang, X. Li. (2024). Prediction of compressive strength of concrete specimens based on interpretable machine learning. *Materials*, 17(15), 3661. https://doi.org/10.3390/ma17153661

[19] A. Fisher, C. Rudin, F. Dominici. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research,* 20(177), 1-81. https://doi.org/10.48550/arXiv.1801.01489

[20] M.T. Ribeiro, S. Singh, C. Guestrin. (2016). "Why should i trust you?" Explaining the predictions of any classifier. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135-1144.*https://doi.org/10.1145/2939672.2939778

[21] F. Farooq, M.N. Amin, K. Khan, M.R. Sadiq, M.F. Javed, F. Aslam, R. Alyousef. (2020). A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC). *Applied Sciences*, 10(20), 7330. https://doi.org/10.3390/app10207330

[22] Y Y. Zhang, W. Ren, Y. Chen, Y. Mi, J. Lei, L. Sun. (2024). Predicting the compressive strength of high-performance concrete using an interpretable machine learning model. *Scientific*

Reports, 14, 28346. https://doi.org/10.1038/s41598-024-79502-z

[23] Q. Han, C. Gui, J. Xu, G. Lacidogna. (2019). A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. *Construction and Building Materials*, 226, 734-742. https://doi.org/10.1016/j.conbuildmat.2019.07.315

[24] Y. Aggarwal, G.K. Gurjar, P. Aggrawal, P. Kumar. (2023). Prediction of Compressive Strength of Self-Compacting Concrete using Machine Learning Techniques. Journal of Asian Concrete Federation, 9(1), 17-30. http://dx.doi.org/10.18702/acf.2023.9.1.17

[25] A. Kashem, P. Das. (2023). Compressive strength prediction of high-strength concrete using hybrid machine learning approaches by incorporating SHAP analysis. Asian Journal of Civil Engineering, 24, 3243-3263. https://doi.org/10.1007/s42107-023-00707-0

[26] K. Le Nguyen, M. Shakouri, L.S. Ho. (2025). Investigating the effectiveness of hybrid gradient boosting models and optimization algorithms for concrete strength prediction. Engineering Applications of Artificial Intelligence, 149, 110568. https://doi.org/10.1016/j.engappai.2025.110568

[27] J.R. Quinlan. (1992). Learning with continuous classes. Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, pp. 343-348. World Scientific.

[28] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal. (2017). Data mining: practical machine learning tools and techniques. *Elsevier*. https://doi.org/10.1016/C2015-0-02071-8

[29] A.J. Smola, B. Schölkopf. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

[30] J.-S. Chou, A.-D. Pham. (2013). Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength. *Construction and Building Materials*, 49, 554-563. https://doi.org/10.1016/j.conbuildmat.2013.08.078

[31] M. Seeger. (2004). Gaussian processes for machine learning. *International Journal of Neural Systems*, 14(2), 69-106. https://doi.org/10.1142/s0129065704001899

[32] M.L. Stein. (1999). Interpolation of spatial data: some theory for kriging. *Springer.* https://doi.org/10.1007/978-1-4612-1494-6

[33] Z. Zou, B. Peng, L. Xie, S. Song. (2024). Enhanced Gaussian Process Model for Predicting Compressive Strength of Ultra-High-Performance Concrete (UHPC). *Materials*, 17(24), 6140. https://doi.org/10.3390/ma17246140

[34] O. Ozbayram, A. Olivier, L. Graham-Brady. (2024). Heteroscedastic Gaussian Process Regression for material structure–property relationship modeling. *Computer Methods in Applied Mechanics and Engineering*, 431, 117326. https://doi.org/10.1016/j.cma.2024.117326

[35] L. Xie, J. He, C. Wang, S. Zhu, H. Zhang, X. Yang. (2025). Enhanced Bayesian Gaussian process regression for compressive strength prediction of multi-binder concrete. *Journal of Building Engineering*, 111, 113308. https://doi.org/10.1016/j.jobe.2025.113308

[36] Y. Yu, G. Wang, G.F. Huseien, Z. Zou, Z. Ding, C. Zhang. (2025). Intelligent prediction of compressive strength of self-compacting concrete incorporating silica fume using hybrid IWOA-GPR model. *Materials Today Communications*, 45, 112282. https://doi.org/10.1016/j.mtcomm.2025.112282

[37] D.E. Rumelhart, G.E. Hinton, R.J. Williams. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536. https://doi.org/10.1038/323533a0

[38] Y. Yu, P. Zhu, Y. Shi, F. Xu, L. Jiang, H. Chu,

N. Xu, M. Liu, Y. Jia, T. Peng. (2023). Prediction of temperature development of concrete with set-controlling admixture based on a new hydration kinetics model. *Materials*, 16(2), 497. https://doi.org/10.3390/ma16020497

[39] J J. Snoek, H. Larochelle, R.P. Adams. (2012). Practical Bayesian optimization of machine learning algorithms. *NIPS'12: Proceedings of the 26th International Conference on Neural Information Processing Systems, vol 2, pp. 2951-2959.* https://doi.org/10.48550/arXiv.1206.2944

[40] A. Altmann, L. Toloşi, O. Sander, T. Lengauer. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10), 1340-1347. https://doi.org/10.1093/bioinformatics/btq134

[41] H. Fu, X. Zhou, P. Xu, D. Sun. (2025). Prediction of Compressive Strength of Concrete Using Explainable Machine Learning Models. *Materials*, 18(21), 5009. https://doi.org/10.3390/ma18215009

[42] M.K. Elshaarawy, M.M. Alsaadawi, A.K. Hamed. (2024). Machine learning and interactive GUI for concrete compressive strength prediction. *Scientific Reports*, 14, 16694. https://doi.org/10.1038/s41598-024-66957-3

[43] E. Li, N. Zhang, B. Xi, J. Zhou, X. Gao. (2023). Compressive strength prediction and optimization design of sustainable concrete based on squirrel search algorithm-extreme gradient boosting technique. *Frontiers of Structural and Civil Engineering*, 17, 1310-1325. https://doi.org/10.1007/s11709-023-0997-3

[44] M.F. Yaqub. (2024). Predicting compressive strength of concrete incorporating fly ash, blast furnace slag, and superplasticizer using machine learning techniques. *2024 REYES Proceedings Remote Experience for Young Engineers and Scientists, pp. 1-9.* https://doi.org/http://dx.doi.org/10.25776/vf07-8125

[45] M. Abdellatief, W. Hamla, H. Hamouda. (2025). AI driven prediction of early age compressive strength in ultra high performance fiber reinforced concrete. *Scientific Reports*, 15, 20316. https://doi.org/10.1038/s41598-025-06725-z

[46] B. Pan, W. Liu, P. Zhou, D.O. Wu. (2024). Predicting the compressive strength of recycled concrete using ensemble learning model. *IEEE Access,* 13, 2958-2969. https://doi.org/10.1109/ACCESS.2024.3519669

[47] S. Poudel, B. Gautam, U. Bhetuwal, P. Kharel, S. Khatiwada, S. Dhital, S. Sah, D. Kc, Y.J. Kim. (2025). Prediction of compressive strength of sustainable concrete incorporating waste glass powder using machine learning algorithms. *Sustainability*, 17(10), 4624. https://doi.org/10.3390/su17104624

[48] A. Adessina, A.B. Fraj, J.-F. Barthélémy. (2023). Improvement of the compressive strength of recycled aggregate concretes and relative effects on durability properties. *Construction and Building Materials*, 384, 131447. https://doi.org/10.1016/j.conbuildmat.2023.131447

[49] J J. Bai, C. Ge, J. Liang, J. Xu. (2025). Recycled Aggregate: A Solution to Sustainable Concrete. *Materials*, 18(12), 2706. https://doi.org/10.3390/ma18122706

[50] M.M.M. Dafedar, B. Rao K, B.H.V. Pai, G.B. Bekkeri, K.K. Shetty. (2025). Viability of using 100% of recycled concrete aggregates for durable solid masonry blocks. *Emergent Materials*, 8, 6017-6037. https://doi.org/10.1007/s42247-025-01194-6

[51] A. Tabani, R. Biswas. (2025). Assessment of compressive strength of ultra-high-performance concrete using advanced machine learning models. *Structural Concrete*, 26(6), 7269-7283.

https://doi.org/10.1002/suco.70076

[52] V.Q. Tran, V.Q. Dang, L.S. Ho. (2022). Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Construction and Building Materials,* 323, 126578. https://doi.org/10.1016/j.conbuildmat.2022.126578

[53] A. Marchewka, P. Ziolkowski, S.G. Galán. (2025). Real-time prediction of early concrete compressive strength using AI and hydration monitoring. *Scientific Reports*, 15, 15463. https://doi.org/10.1038/s41598-025-97060-w

[54] K.C. Onyelowe, A.M. Ebid, R.B.R. Jiménez, A.G.L. Guamán, D.R.Ñ. Estrella, J.C.C. Martínez, E.F.G. Cabezas, P. Awoyera. (2025). Modeling the compressive strength of concrete at different curing regimes using machine learning. *Discover Sustainability*, 6, 358. https://doi.org/10.1007/s43621-025-01195-9

[55] L. Tang. (2025). Machine Learning-Based Prediction of Concrete Compressive Strength and Interpretability Analysis. *Journal of Civil Engineering and Urban Planning*, 7(2), 122-133. DOI: 10.23977/jceup.2025.070217

[56] T. Joshi, P. Mathur, P. Oza, S. Agrawal, H. Narmawala. (2024). Predictive modelling of sustainable concrete compressive strength using advanced machine learning algorithms. *Advances in Civil and Architectural Engineering*, 15(29), 168-192. https://doi.org/10.13167/2024.29.11

[57] D.D. Khuat, D.D. Nguyen, M.H. Nguyen, B.T. Pham, K. Nakarai. (2026). An interpretable hybrid machine learning approach for predicting the compressive strength of internal-curing concrete incorporating recycled roof-tile waste. *Buildings*, 16(3), 674. https://doi.org/10.3390/buildings16030674

[58] X. Tan, J. Xing, Y. Wang, H. Qiu, S. Mahjoubi, P. Guo. (2026). Explainable machine learning for predicting compressive strength of rubberized concrete: SHAP interpretation, lifecycle assessment, and design recommendations. *Journal of Cleaner Production*, 538, 147338. https://doi.org/10.1016/j.jclepro.2025.147338

[59] S. Elhishi, A.M. Elashry, S. El-Metwally. (2023). Unboxing machine learning models for concrete strength prediction using XAI. *Scientific Reports*, 13, 19892. https://doi.org/10.1038/s41598-023-47169-7

[60] M.U. Gazi, M.T. Hasan, P. Debnath. (2025). Few-shot meta-learning for concrete strength prediction: a model-agnostic approach with SHAP analysis. *AI in Civil Engineering*, 4, 20. https://doi.org/10.1007/s43503-025-00064-8