Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 1, 2026

EXPLAINABLE CATBOOST-BASED PREDICTION OF CONCRETE COMPRESSIVE STRENGTH USING MIX PROPORTIONS AND CONCRETE PROPERTIES

Dr. M. Adil Khan[*1], Saad Hanif[2], Muazzam Nawaz[3]

[*1]*Resident Engineer, NESPAK*
[2]*Military College of Engineering Department of civil engineering National University of Sciences & Technology Risalpur*
[3]*Department of civil engineering UET Lahore*

[*1]adee.uol@gmail.com [*2]saadhanif107@tamu.edu [3]muazzam8471542@gmail.com

DOI: https://doi.org/10.5281/zenodo.18373631

Keywords
Concrete compressive strength; CatBoost regression; Machine learning; SHAP analysis; Explainable artificial intelligence; Sustainable concrete

Article History
Received: 27 November 2025
Accepted: 11 January 2026
Published: 26 January 2026

Corresponding Author: *
Dr. M. Adil Khan

Abstract
Accurate prediction of concrete compressive strength is essential for ensuring structural safety, optimizing mix design, and reducing experimental time and cost in construction engineering. Traditional empirical models often struggle to capture the nonlinear interactions among concrete constituents, particularly in mixes incorporating supplementary cementitious materials. This study presents an explainable machine learning framework based on the CatBoost regression algorithm to predict concrete compressive strength using material composition and curing age. A dataset comprising 1,133 concrete samples was employed, including cement, blast-furnace slag, fly ash, water, super-plasticizer, coarse aggregate, fine aggregate, and age of testing as input variables. Comprehensive exploratory data analysis was conducted using statistical characterization, correlation assessment, and distribution analysis to understand feature behavior. The CatBoost model was developed through systematic training, testing, and validation with optimized hyperparameters to ensure robust generalization. Model performance was evaluated using multiple statistical metrics, including variance-based, absolute, relative, and normalized error measures, along with residual analysis to assess bias and error distribution. Furthermore, Shapley Additive Explanations (SHAP) were integrated to interpret feature contributions and enhance model transparency. The results demonstrate high predictive accuracy and consistent performance across datasets, while SHAP analysis identifies curing age and cement content as dominant contributors to strength development, followed by water content and supplementary cementitious materials. The proposed framework combines strong predictive capability with explainability, offering a reliable and interpretable decision-support tool for concrete mix design and performance prediction.

INTRODUCTION

Concrete is the most widely used construction material worldwide due to its versatility, durability, and cost-effectiveness; however, its mechanical performance, particularly compressive strength, is highly dependent on complex interactions among constituent materials, mix proportions, and curing conditions(Emad et al., 2022; Wani & Suthar, 2024; Zhang et al., 2024). Accurate prediction of concrete compressive strength is therefore a critical task in civil engineering, as it directly influences structural safety, material optimization, quality control, and construction scheduling. Traditionally, compressive strength assessment relies on laboratory testing of cast specimens at specified curing ages, which is time-consuming, labor-intensive, and costly. Moreover, experimental approaches offer limited flexibility in exploring the wide range of possible mix combinations encountered in practice(A. Q. Khan et al., 2023; Tran et al., 2020; Yeh, 1998). These limitations have motivated the development of predictive models capable of estimating compressive strength using readily available mix design parameters.

Conventional empirical and statistical models, such as regression-based approaches, have been extensively applied to strength prediction. While these models provide basic insights, they often fail to capture the inherent nonlinearity and complex interactions among concrete constituents, particularly when supplementary cementitious materials such as blast-furnace slag and fly ash are incorporated(Abdellatief et al., 2025; Abuodeh et al., 2020; Chithra et al., 2016; L. Li et al., 2024). The growing emphasis on sustainable construction has further increased the variability of concrete mix designs, as partial replacement of cement with industrial by-products is now common practice. This variability introduces additional challenges for traditional modeling techniques, which typically assume linear relationships and limited interaction effects(Emad et al., 2022; Marani et al., 2020; Wani & Suthar, 2024; Xu et al., 2025; Zhang et al., 2024).

In recent years, machine learning (ML) techniques have emerged as powerful tools for modeling complex, nonlinear systems in civil engineering applications. Algorithms such as artificial neural networks, support vector machines, and ensemble-based methods have demonstrated improved predictive accuracy compared to classical approaches. Among these, gradient boosting methods have gained particular attention due to their ability to iteratively improve model performance by minimizing prediction errors. However, many ML models operate as "black boxes," offering limited interpretability, which restricts their acceptance in engineering practice where transparency and physical understanding are essential(Amjad et al., 2022; Boldini et al., 2023; S. Lee et al., 2022a; Rathakrishnan et al., 2022a; Wu et al., 2025).

CatBoost, a gradient boosting decision tree algorithm, has recently shown superior performance in structured data modeling due to its ordered boosting strategy and effective handling of feature interactions. Unlike conventional boosting algorithms, CatBoost reduces prediction bias and overfitting while maintaining high accuracy with minimal preprocessing. These characteristics make it particularly suitable for concrete strength prediction, where datasets often include heterogeneous features, varying scales, and nonlinear dependencies (Ahmad et al., 2021; Alyami et al., 2023; Farooq et al., 2021; Feng et al., 2020; P. Li et al., 2024; Reddy et al., 2024). Despite its advantages, the application of CatBoost in concrete engineering remains limited, and systematic studies combining performance evaluation with detailed interpretability analysis are still scarce.

Another critical limitation of many ML-based concrete strength studies is the lack of comprehensive model evaluation across independent datasets. Several studies report high accuracy on training data but fail to demonstrate consistent performance on testing and validation sets, raising concerns regarding model generalization. Additionally, residual behavior is often overlooked, despite its importance in identifying bias, error distribution, and prediction reliability. Without residual assessment, it is difficult to determine whether a model's predictions are uniformly accurate across the entire strength range or biased toward specific mix conditions(Alomari & Andó, 2024; Kitani & Iwata, 2023; Y.-G. Lee et al., 2022; Lin & Gao, 2022; Ponce-Bobadilla et al., 2024).

Furthermore, while predictive accuracy is essential, understanding the influence of individual input parameters on model output is equally important for engineering decision-making. Explainable artificial intelligence (XAI) techniques, particularly Shapley Additive Explanations (SHAP), provide a robust framework for interpreting complex ML models by quantifying the contribution of each feature to individual predictions. SHAP analysis enables both global and local interpretability, allowing engineers to verify whether the learned relationships align with established material science principles. However, the integration of SHAP-based interpretation with advanced boosting algorithms such as CatBoost remains underexplored in concrete compressive strength prediction(Gali et al., 2023; W. Khan et al., 2025; Mujtaba et al., 2023; Nawaz et al., 2025; Shah et al., 2023).

In this context, the present study aims to develop a robust and interpretable CatBoost regression model for predicting concrete compressive strength using a comprehensive dataset consisting of 1,133 samples. The dataset includes key material parameters such as cement, blast-furnace slag, fly ash, water, super-plasticizer, coarse aggregate, fine aggregate, and age of testing. These variables collectively represent modern concrete mix designs incorporating both traditional and supplementary cementitious materials. Prior to model development, extensive exploratory data analysis was conducted, including statistical characterization, correlation analysis, and distribution assessment using violin plots, to understand data behavior and feature interactions. The CatBoost model was systematically developed through training, testing, and validation stages, with careful hyperparameter selection to balance accuracy and generalization. Model performance was evaluated using multiple metrics, including variance-based, absolute, relative, and normalized error measures, ensuring a holistic assessment of predictive capability. Residual analysis was performed to examine error distribution, bias, and robustness across all datasets, providing deeper insight into model reliability beyond standard accuracy metrics.

To address the interpretability challenge, SHAP-based explanatory analysis was employed to quantify the contribution of each input variable to compressive strength predictions. This approach enables direct comparison between data-driven insights and established concrete behavior, thereby enhancing trust and transparency in the model. By integrating high-performance prediction with explainable analysis, this study bridges the gap between advanced machine learning techniques and practical engineering applicability.

This research contributes to the field by presenting a comprehensive, interpretable, and well-validated CatBoost-based framework for concrete compressive strength prediction. The findings provide valuable insights into the relative importance of material constituents and curing age, offering practical guidance for mix design optimization and quality control in sustainable concrete construction.

Methodology

Methodological framework

The methodology adopted in this study follows a structured, stepwise framework to develop and interpret a CatBoost-based predictive model for concrete compressive strength. In Step 1, a comprehensive literature review was conducted to examine existing experimental and machine learning approaches, enabling the identification of research gaps related to prediction accuracy and model interpretability. Step 2 involved data collection and exploratory data analysis, where statistical measures, correlation analysis, and violin plots were used to understand data distribution, variability, and inter-feature relationships. In Step 3, the CatBoost machine learning model was developed, including data splitting into training and testing sets, model training and validation, hyperparameter optimization, and residual analysis to evaluate prediction errors and model robustness. Finally, Step 4 focused on explanatory analysis using Shapley Additive Explanations (SHAP) to interpret feature contributions and quantify the influence of each input variable on compressive strength predictions.

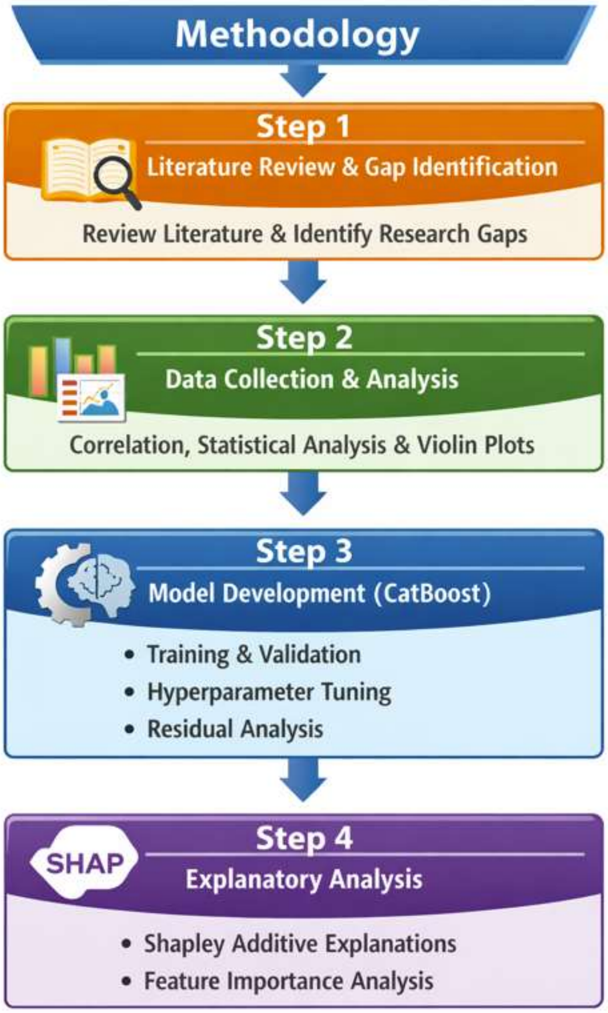


Figure 1 Flowchart illustrating the overall research methodology, including literature review and gap identification, data collection and exploratory analysis, CatBoost model development with training, validation and hyperparameter tuning, and SHAP-based explanatory analysis for model interpretability.

Data characteristics

The dataset used in this study comprises 1,133 observations and captures the material composition, curing age, and resulting compressive strength of concrete mixes. The primary input variables include cement, blast-furnace slag, fly ash, water, super-plasticizer, coarse aggregate, fine aggregate, and age of testing, all measured in standard engineering units. Cement content ranges from 102 to 540 kg/m³, with a mean of 276.5 kg/m³ and moderate variability (standard deviation of 103.5 kg/m³), indicating a wide spectrum of mix designs. Supplementary cementitious materials exhibit substantial dispersion and zero inflation: blast-furnace slag (mean 74.3 kg/m³) and fly ash (mean 62.8 kg/m³) both show medians close to zero, reflecting that many mixes do not contain these components, while others use them in large quantities.

Table 1 Statistical summary of concrete mix constituents, curing age, and compressive strength, including central tendency, dispersion, skewness, and kurtosis measures.

Feature	count	mean	std	min	25%	50%	75%	max	skewness	kurtosis
Cement(kg/m3)	1133.0	276.5	103.5	102.0	190.0	266.0	342.0	540.0	0.5	-0.5
Blast-furnace Slag(kg/m3)	1133.0	74.3	84.2	0.0	0.0	26.0	141.3	359.4	0.8	-0.5
Fly Ash(kg/m3)	1133.0	62.8	71.6	0.0	0.0	0.0	122.0	260.0	0.6	-0.9
Water(kg/m3)	1133.0	183.0	21.7	121.8	167.0	185.7	193.8	247.0	0.1	0.1
Super-plasticizer(kg/m3)	1133.0	6.4	5.8	0.0	0.0	6.7	10.2	32.2	0.8	1.5
Coarse Aggregate (kg/m3)	1133.0	964.8	82.8	708.0	919.0	966.8	1026.6	1145.0	-0.2	-0.4
Fine Aggregate(kg/m3)	1133.0	770.5	79.4	594.0	720.0	777.5	821.0	992.6	-0.2	-0.2
Age of testing(day)	1133.0	44.1	60.4	1.0	14.0	28.0	28.0	365.0	3.5	13.8
Concrete compressive strength(MPa)	1133.0	35.8	16.1	2.3	24.4	34.7	44.9	82.6	0.4	-0.2

Water content is comparatively stable, with a mean of 183.0 kg/m³ and low skewness, suggesting controlled water usage across mixtures. Super-plasticizer content has a low mean (6.4 kg/m³) but high relative variability and positive skewness, indicating selective use in high-performance mixes. Aggregate contents are the most consistent variables: coarse aggregate averages 964.8 kg/m³ and fine aggregate 770.5 kg/m³, both exhibiting low skewness and kurtosis, which reflects standardized aggregate proportions in concrete production. The age of testing shows extreme right skewness (skewness = 3.5, kurtosis = 13.8), ranging from 1 to 365 days, highlighting the inclusion of both early-age and long-term strength measurements. The target variable, concrete compressive strength, varies from 2.3 to 82.6 MPa, with a mean of 35.8 MPa and moderate variability, representing a broad spectrum from low- to high-strength concrete. The correlation diagram illustrates the linear relationships among input variables and compressive strength, highlighting key positive and negative associations between material components and strength development. The violin plots depict the distribution, spread, and skewness of each feature, revealing variability and the presence of outliers across concrete mix parameters.

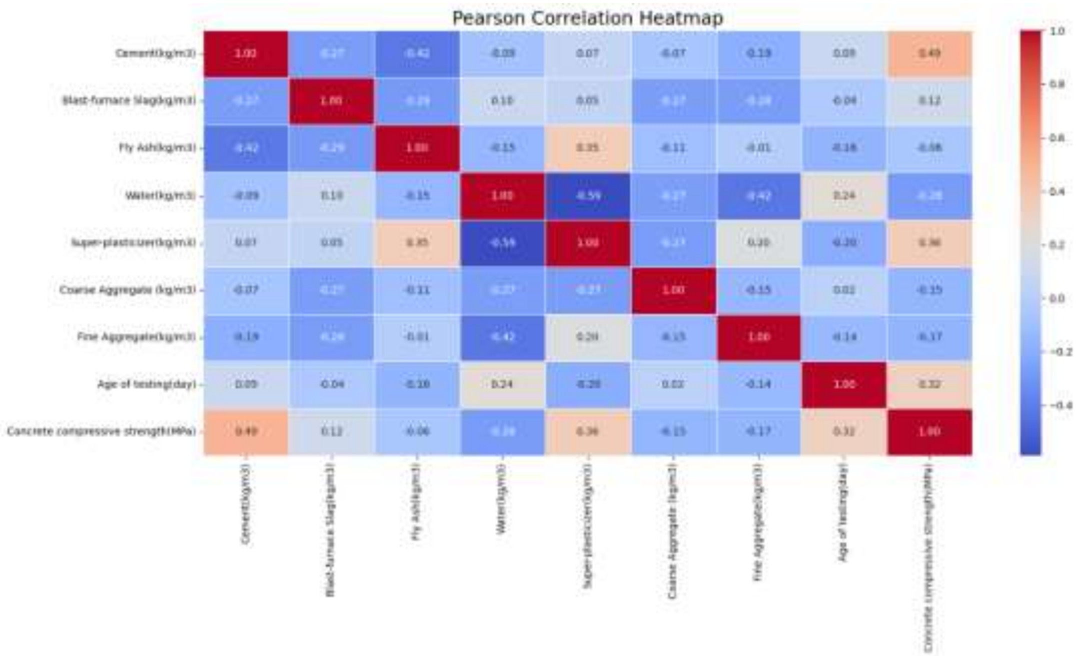
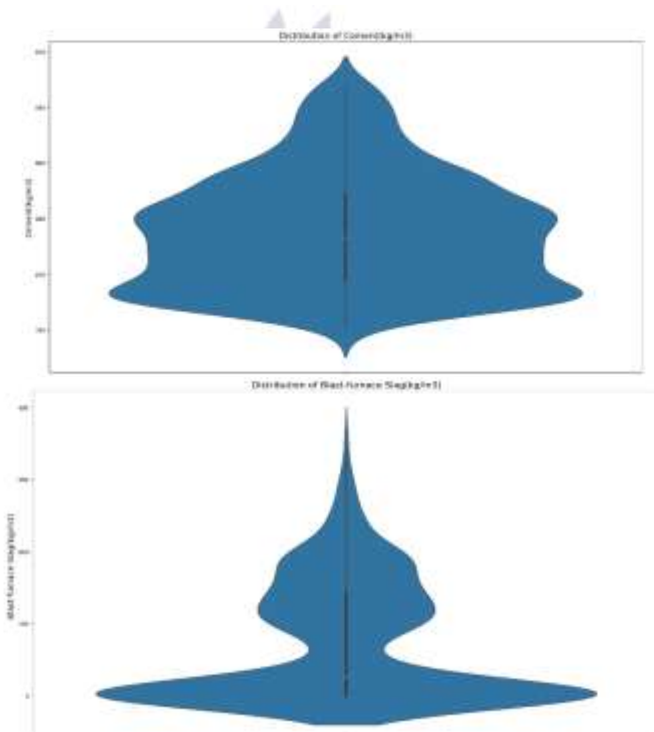
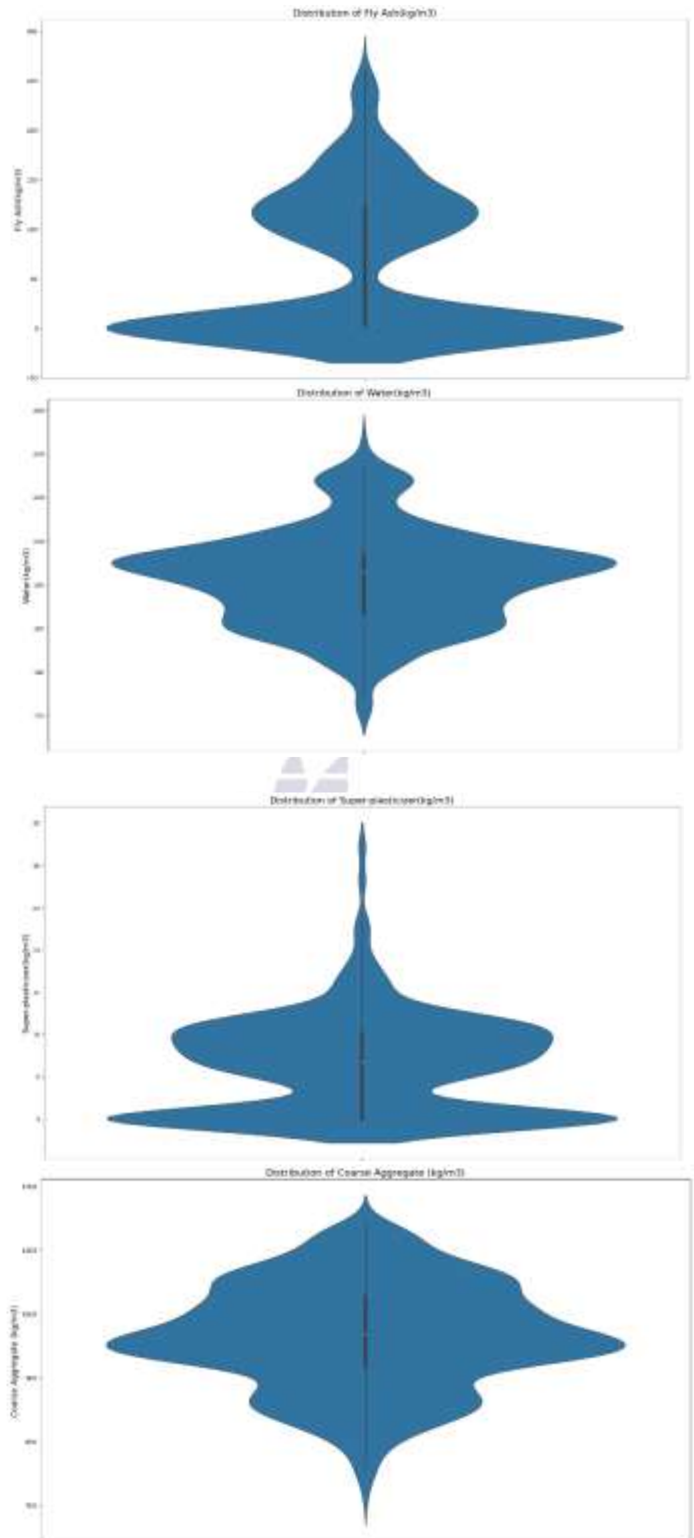


Figure 2 Correlation heatmap illustrating the strength and direction of linear relationships among concrete mix variables and compressive strength.





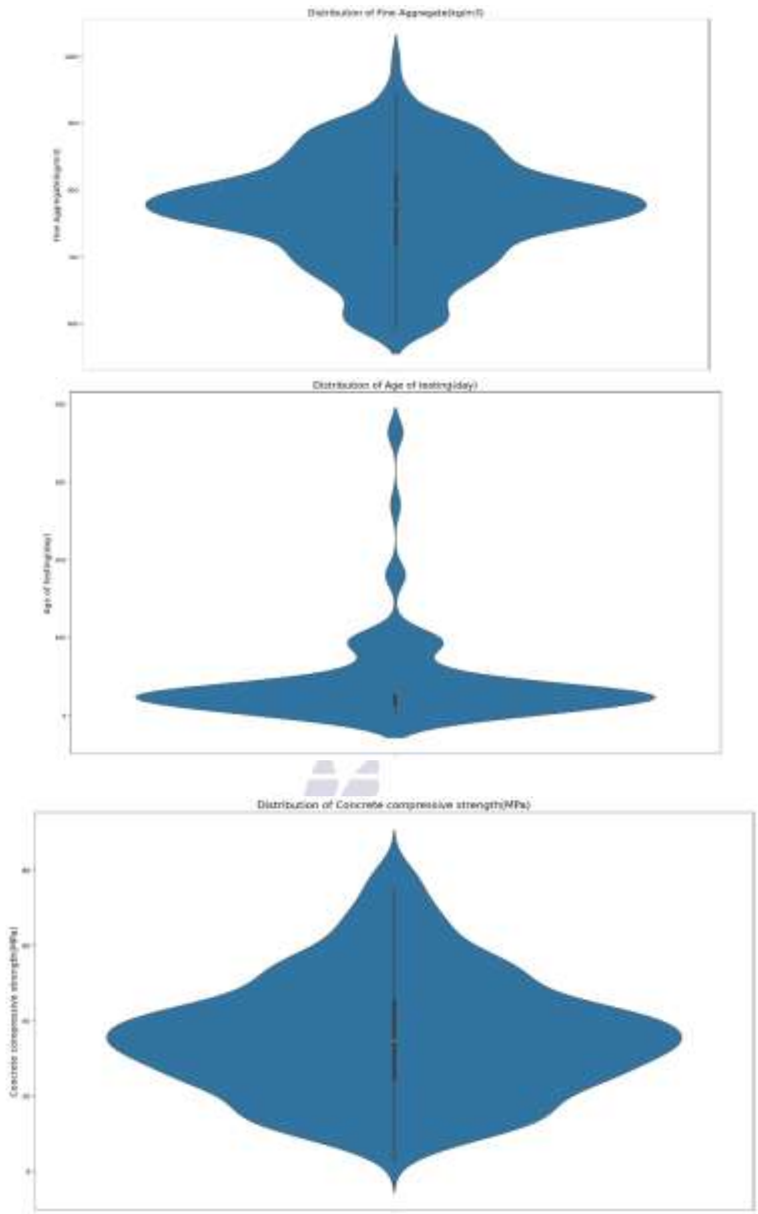


Figure 3 Violin plots showing the distribution, variability, and skewness of individual concrete mix features and compressive strength.

Model development cat boost model

CatBoost Model

CatBoost (Categorical Boosting) is a gradient boosting machine learning algorithm developed to efficiently handle complex, nonlinear relationships in structured datasets. It builds an ensemble of decision trees sequentially, where each new tree corrects the errors of the previous ones. CatBoost is particularly robust against overfitting due to its ordered boosting strategy and symmetric tree structure. In this study, CatBoost is employed as a regression model to predict concrete compressive strength because of its high predictive accuracy, ability to model nonlinear interactions among material components, and minimal requirement for extensive data preprocessing.

Parameter setting

For model development, a CatBoost Regressor was implemented with carefully selected hyperparameters to balance prediction accuracy and generalization performance. The number of trees (iterations) was set to 1,000, providing sufficient learning capacity to capture complex patterns without excessive computational cost. A learning rate of 0.05 was chosen to ensure gradual model convergence, reducing the risk of overfitting while allowing stable optimization. The tree depth was fixed at 8, enabling the model to learn higher-order feature interactions among cementitious materials, aggregates, water content, and curing age. To further control overfitting, an L2 leaf regularization value of 3 was applied, penalizing large leaf weights and improving model robustness. The loss function was defined as Root Mean Squared Error (RMSE), which is appropriate for continuous strength prediction and emphasizes larger prediction errors. Additionally, subsampling was implicitly controlled through CatBoost's stochastic gradient boosting mechanism, enhancing model generalization. Early stopping rounds were set to 50, allowing training to terminate automatically when validation performance ceased to improve. The model was trained using an 80:20 train–test split, with the training set internally divided for validation. Residual analysis was subsequently performed to assess error distribution and bias, confirming the adequacy of the selected hyperparameters.

Table 2 Hyperparameter configuration of the CatBoost regression model used for predicting concrete compressive strength.

Parameter	Selected Value
Model Type	CatBoost Regressor
Loss Function	RMSE
Number of Trees (Iterations)	1000
Learning Rate	0.05
Tree Depth	8
L2 Leaf Regularization	3
Boosting Type	Ordered Boosting
Early Stopping Rounds	50
Train–Test Split	80%–20%
Random Seed	Fixed

Model assessment

The performance of the developed CatBoost regression model was assessed using multiple statistical evaluation metrics to comprehensively examine accuracy, error magnitude, and generalization capability. These metrics were selected to capture both absolute and relative prediction performance and are widely accepted in machine learning applications for civil engineering problems(Chicco et al., 2021).

The coefficient of determination (R^2) was used to evaluate the proportion of variance in the experimental compressive strength explained by the model. It is defined as:

$$R^2 = 1 - \left[\frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}\right]$$

where y_i represents the experimental value, $\hat{y}_i$ the predicted value, and $\bar{y}$ the mean of experimental values. A higher R^2 indicates stronger predictive capability and goodness of fit.

Error-based metrics were employed to quantify prediction deviations. The Mean Squared Error (MSE) measures the average squared difference between predicted and observed values and is expressed as:

$$MSE = \left(\frac{1}{n}\right)\sum(y_i - \hat{y}_i)^2$$

The Root Mean Squared Error (RMSE), calculated as the square root of MSE, provides error magnitude in the same unit as the target variable and is given by:

$$RMSE = \sqrt{[\left(\frac{1}{n}\right)\sum(y_i - \hat{y}_i)^2]}$$

The Mean Absolute Error (MAE) evaluates the average absolute deviation and is defined as:

$$MAE = \left(\frac{1}{n}\right)\sum|y_i - \hat{y}_i|$$

To assess relative prediction accuracy, the Mean Absolute Percentage Error (MAPE) was used:

$$\text{MAPE} = \left(\frac{100}{n}\right)\sum\left|\frac{(y_i - \hat{y}_i)}{y_i}\right|$$

Additionally, the Coefficient of Variation of RMSE (CVRMSE) was applied to normalize RMSE with respect to the mean of experimental values:

$$\text{CVRMSE} = \left(\frac{\text{RMSE}}{\bar{y}}\right)\times 100$$

Model assessment was performed separately for training, testing, and validation datasets to evaluate learning efficiency, generalization performance, and robustness. The combined use of these metrics ensures a balanced evaluation by capturing variance explanation, absolute error, percentage-based error, and normalized performance, thereby providing a reliable framework for assessing the predictive effectiveness of the CatBoost model.

Results

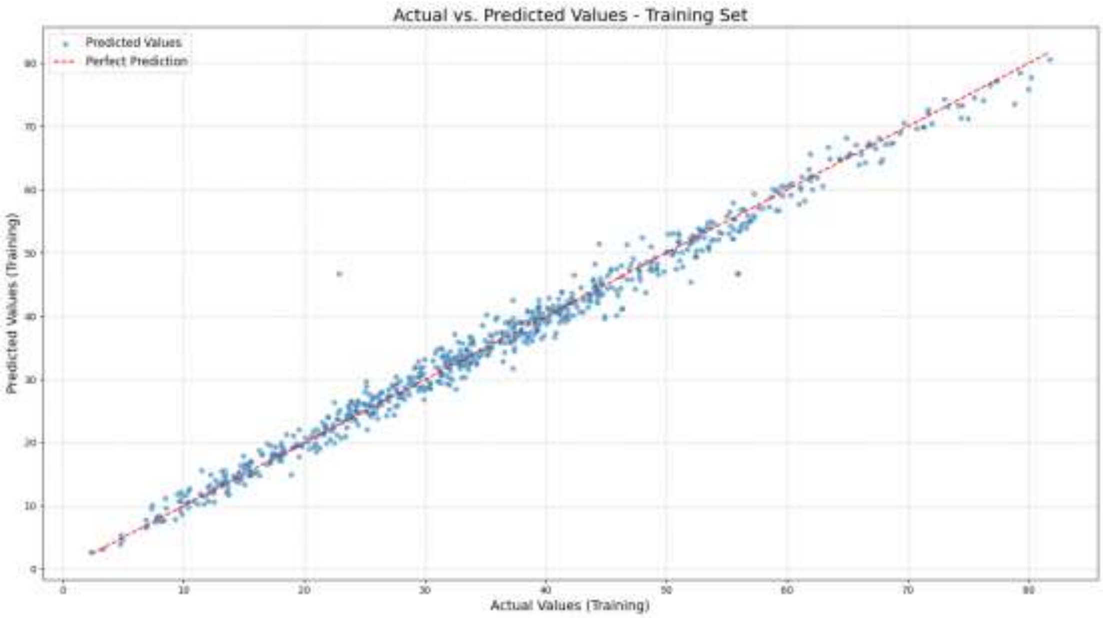
CatBoost model

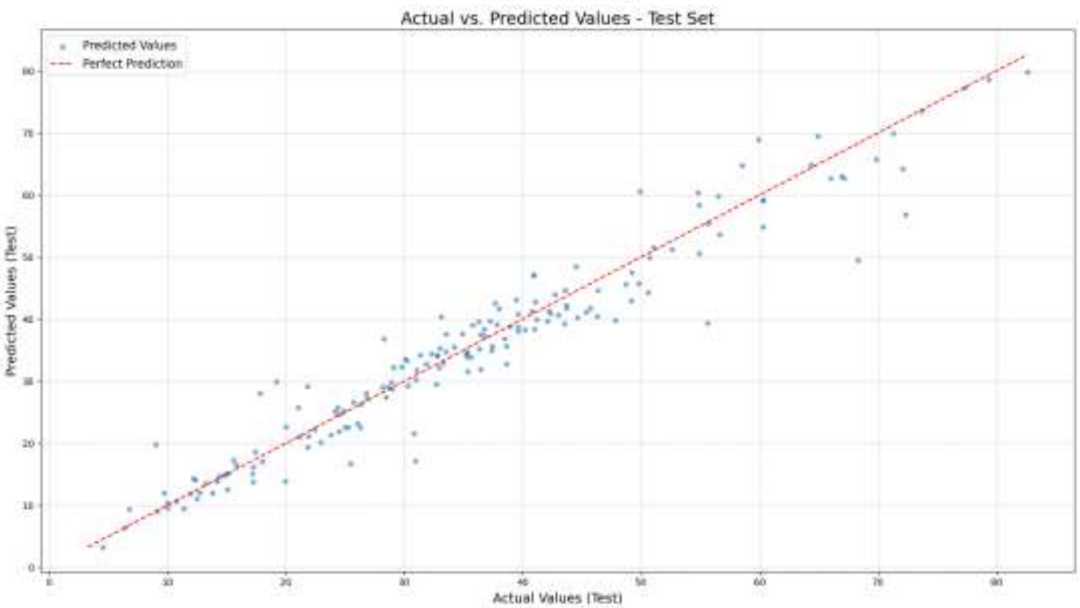
The performance of the developed CatBoost regression model was evaluated using multiple statistical metrics across training, testing, and validation datasets to ensure both predictive accuracy and generalization capability. The coefficient of determination (R^2) values of 0.9852 for training, 0.9348 for testing, and 0.9326 for validation indicate that the model successfully explains more than 93% of the variance in concrete compressive strength for unseen data. The slight reduction in R^2 from training to testing and validation reflects expected generalization behavior and confirms that the model is not overfitted.

Table 3 Performance evaluation metrics of the CatBoost model for training, testing, and validation datasets.

Metric	Training	Test	Validation
R2 Score	0.9852	0.9348	0.9326
MSE	3.7523	18.3707	17.5209
RMSE	1.9371	4.2861	4.1858
MAE	1.3356	2.9019	2.7589
MAPE	4.5400	9.4200	8.7600
CVRMSE	5.4400	12.0700	11.2200

Error-based metrics further support the robustness of the model. The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values are notably low for the training dataset (MSE = 3.75, RMSE = 1.94 MPa), demonstrating an excellent fit during learning. For the testing and validation datasets, RMSE values of 4.29 MPa and 4.19 MPa, respectively, remain within acceptable engineering limits, considering the wide strength range of the dataset (2.3–82.6 MPa). This indicates that the model maintains stable predictive accuracy when exposed to new data. The Mean Absolute Error (MAE) values of 1.34 MPa (training), 2.90 MPa (testing), and 2.76 MPa (validation) show that the average prediction deviation is relatively small and practically meaningful for concrete strength estimation. Additionally, the Mean Absolute Percentage Error (MAPE) values below 10% for both testing (9.42%) and validation (8.76%) datasets demonstrate strong relative accuracy, confirming that prediction errors are proportionally low across varying strength levels.


Actual vs. Predicted Values - Training Set
Predicted Values
Perfect Prediction
Predicted Values (Training)
Actual Values (Training)


Actual vs. Predicted Values - Test Set
Predicted Values
Perfect Prediction
Predicted Values (Test)
Actual Values (Test)

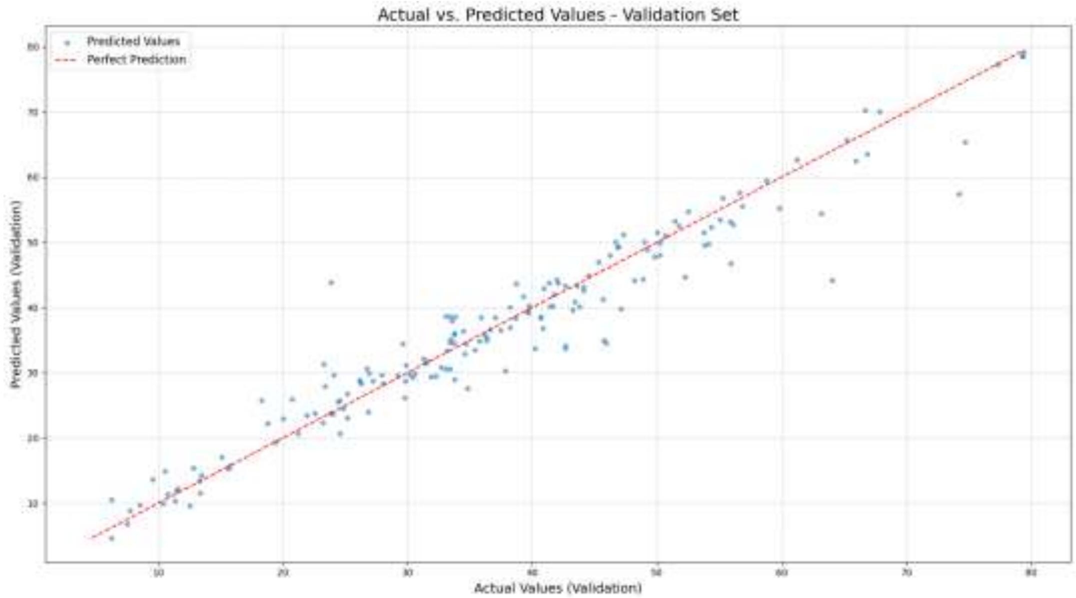


Figure 4 Scatter plots comparing predicted and experimental concrete compressive strength for (a) training, (b) testing, and (c) validation dataset.

The Coefficient of Variation of RMSE (CVRMSE) further validates model reliability, with values of 12.07% and 11.22% for testing and validation datasets, respectively. These values fall within acceptable thresholds for predictive modeling in civil engineering applications, indicating consistent performance across samples. Scatter plots of predicted versus experimental compressive strength for training, testing, and validation datasets visually corroborate the numerical results. Most data points cluster closely around the 45-degree reference line, signifying strong agreement between predicted and actual values. Minor dispersion observed at higher strength ranges suggests increased complexity in high-performance concrete behavior, yet without systematic bias.

Error analysis

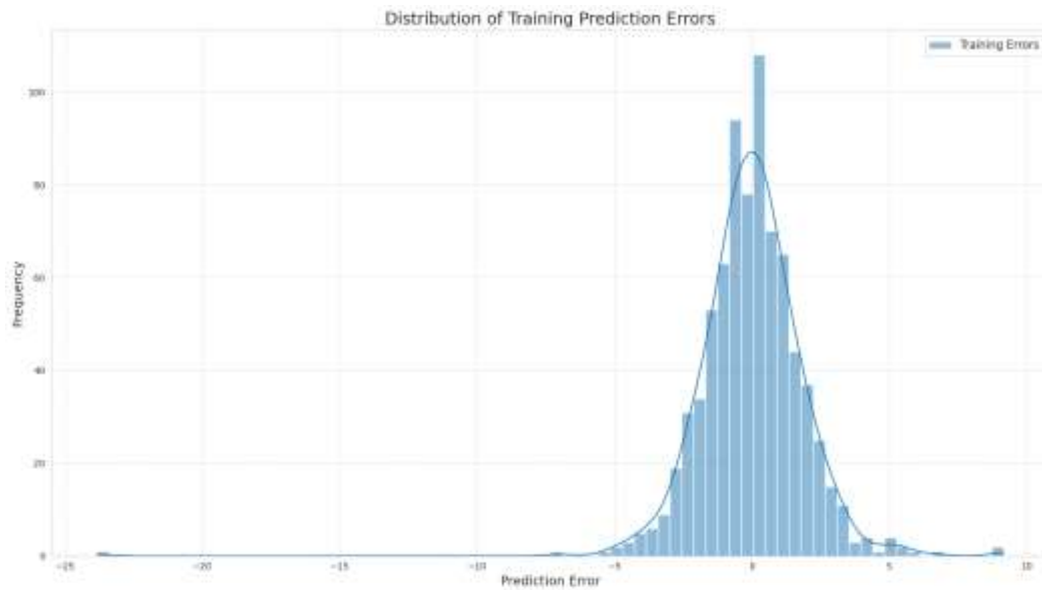
Residual analysis was conducted to further evaluate the predictive reliability, bias, and generalization behavior of the CatBoost model across training, testing, and validation datasets. Residuals, defined as the difference between experimental and predicted concrete compressive strength, provide critical insight into model error distribution and systematic deviations.

Table 4 Statistical summary of residual errors for training, testing, and validation datasets obtained from the Cat Boost model

	Training Errors	Test Errors	Validation Errors
count	793.000000	170.000000	170.000000
mean	0.000862	0.452243	0.410080
std	1.938315	4.274767	4.177963
min	-23.831060	-10.847140	-19.972629
25%	-1.001704	-1.357678	-1.737896
50%	0.045497	0.266399	0.050395
75%	1.018442	2.422239	1.972291
max	9.167261	18.839718	19.869426

ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 1, 2026

For the training dataset, the mean residual is approximately zero (0.0009 MPa), indicating an unbiased model fit with no systematic over- or under-prediction during learning. The standard deviation of 1.94 MPa reflects a tight concentration of errors around zero, consistent with the low RMSE observed in training performance. The interquartile range (−1.00 to 1.02 MPa) further confirms that the majority of predictions deviate minimally from actual values. Although a few extreme residuals are present (minimum −23.83 MPa and maximum 9.17 MPa), these are limited in number and likely correspond to atypical mix designs or high-strength concretes with complex nonlinear behavior. In the testing dataset, the mean residual increases slightly to 0.45 MPa, suggesting a minor tendency toward underprediction; however, this bias remains negligible in practical terms. The standard deviation of 4.27 MPa indicates an expected increase in variability when the model is applied to unseen data. The median residual of 0.27 MPa and balanced quartile range (−1.36 to 2.42 MPa) demonstrate that prediction errors remain symmetrically distributed around zero, reinforcing the model's stability. The presence of larger positive and negative residuals (up to ±18 MPa) reflects the inherent uncertainty associated with diverse concrete compositions and curing ages rather than model instability.



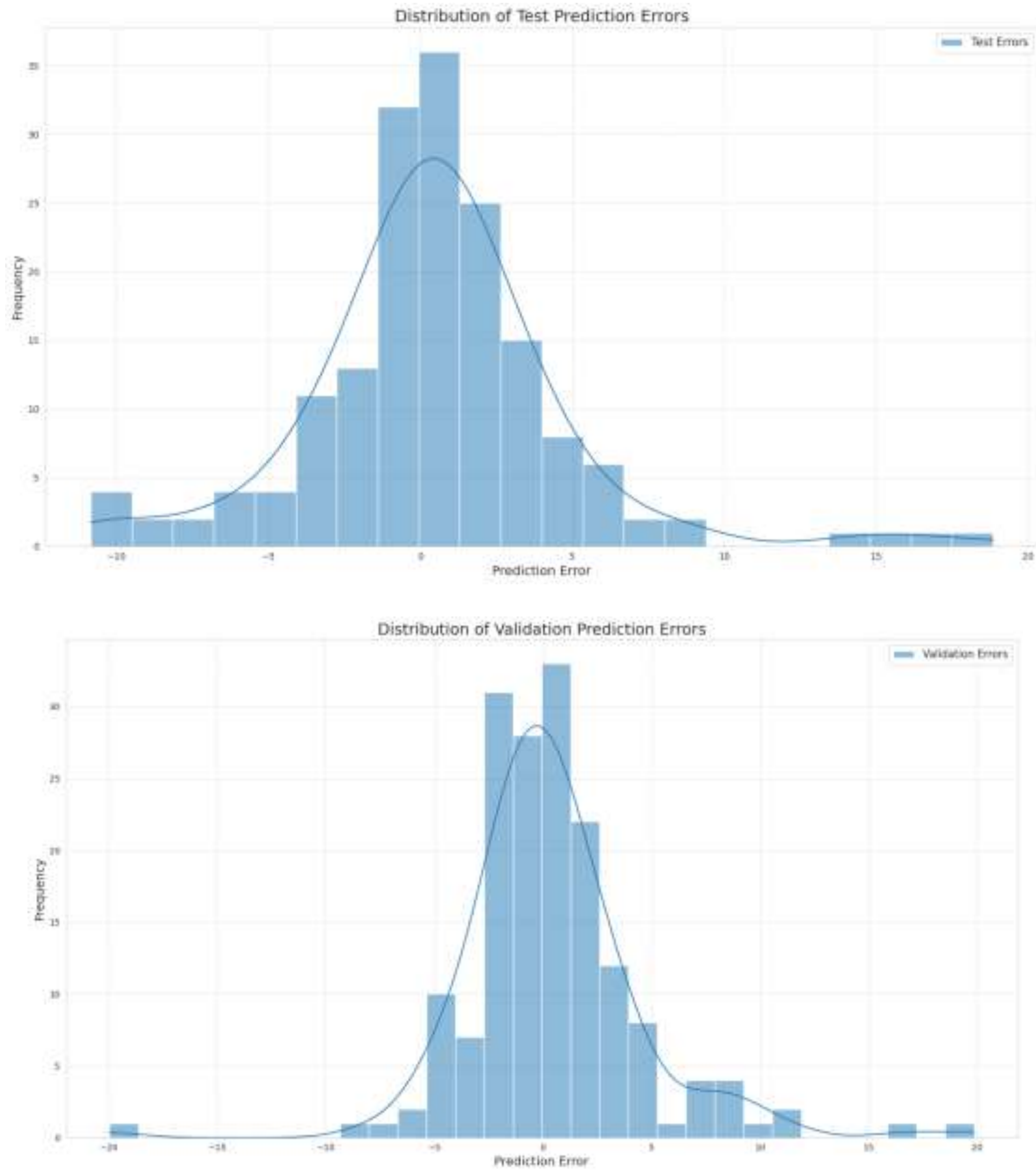


Figure 5 Residual assessment plots for the CatBoost model showing error distributions for (a) training, (b) testing, and (c) validation datasets.

Similarly, the validation dataset exhibits a mean residual of 0.41 MPa and a median close to zero (0.05 MPa), confirming consistency between testing and validation performance. The standard deviation of 4.18 MPa closely matches that of the testing dataset, indicating reliable generalization. The quartile spread (−1.74 to 1.97 MPa) suggests that most predictions fall within an acceptable error margin. While the validation set shows slightly larger extreme residuals (minimum −19.97 MPa and maximum 19.87 MPa), their limited frequency indicates no systematic error pattern. The residual distributions across all datasets are centred near zero with comparable dispersion in testing and validation sets, confirming that the CatBoost model achieves a strong balance between accuracy and generalization. The absence of pronounced skewness or bias in residuals supports the robustness of the selected hyperparameters and validates the model's suitability for predicting concrete compressive strength.

Ranking of feature in prediction

The SHAP summary and feature importance results provide a detailed, instance-level interpretation of how individual input variables influence the CatBoost model's predictions of concrete compressive strength. The SHAP summary plot ranks features based on their overall impact, with the age of testing emerging as the most influential parameter. Higher curing ages (red points) consistently contribute positive SHAP values, indicating a strong increase in predicted compressive strength as hydration progresses, while lower ages (blue points) are associated with negative impacts, reflecting immature strength development. Cement content is the second most influential feature, where higher cement quantities produce positive SHAP values and significantly enhance strength, whereas lower cement contents reduce predicted strength. Water content exhibits an inverse relationship; higher water values predominantly correspond to negative SHAP values, confirming that excessive water content reduces strength due to increased porosity, while lower water levels contribute positively. Blast-furnace slag shows a mixed but generally positive influence at higher contents, suggesting its beneficial contribution to long-term strength through pozzolanic reactions. Fine aggregate displays relatively clustered SHAP values around zero, indicating a moderate and stable influence, where both low and high values produce limited variation in strength predictions.

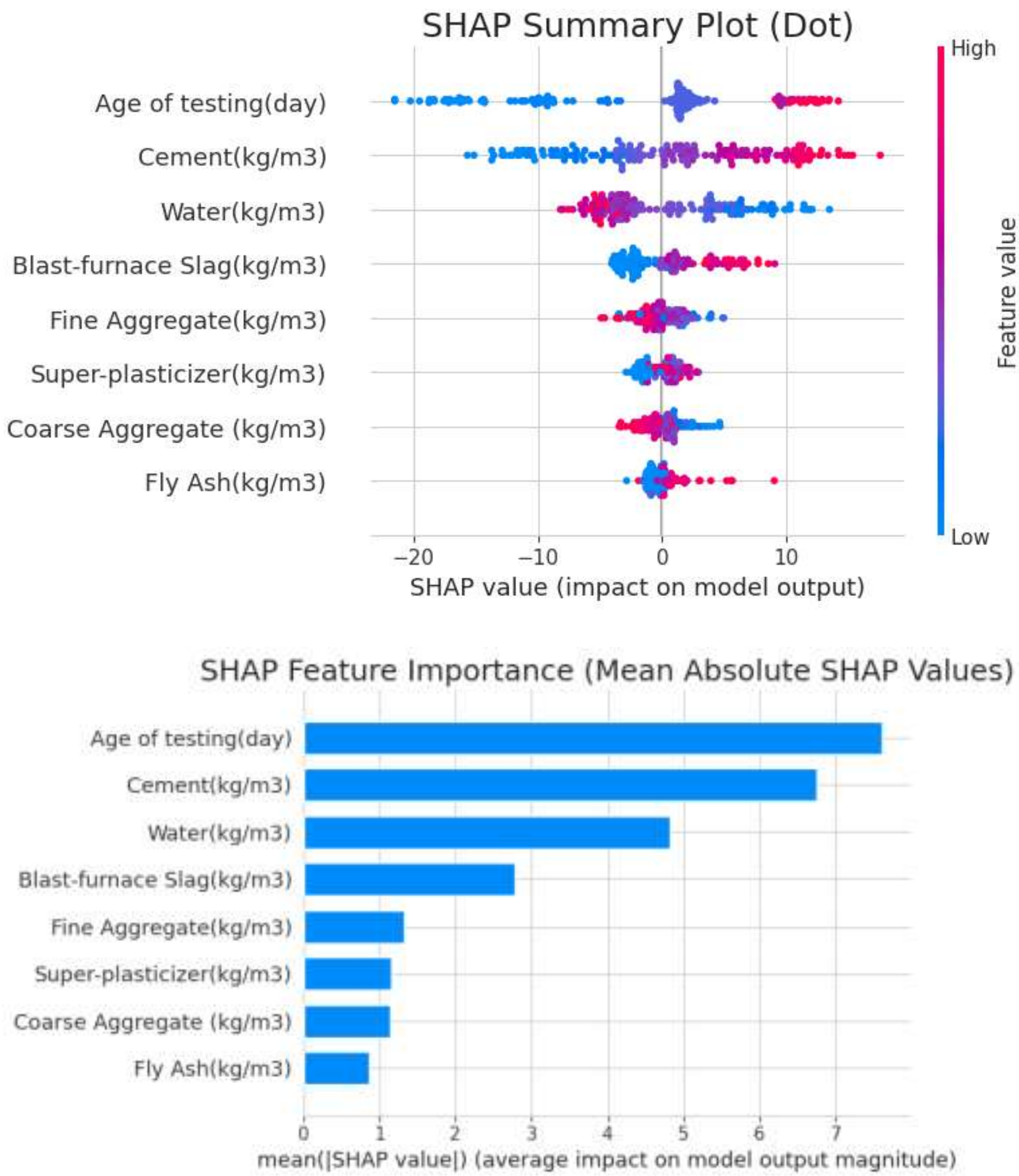


Figure 6 SHAP-based model interpretation showing (a) SHAP summary plot illustrating feature-wise impact and direction on concrete compressive strength predictions and (b) mean absolute SHAP value plot indicating relative feature importance in the CatBoost model.

Super-plasticizer content shows a narrow SHAP distribution with slight positive contributions at higher dosages, reflecting its indirect role in improving workability and enabling lower water–cement ratios. Coarse aggregate content exhibits both positive and negative SHAP values with limited spread, implying a secondary role in strength prediction compared to cementitious components. Fly ash has the lowest overall impact, with most SHAP values close to zero; however, higher fly ash contents occasionally show positive contributions, particularly at later curing ages, indicating delayed strength gain effects. The mean absolute SHAP value plot quantitatively confirms these observations, ranking age of testing and cement content as dominant predictors, followed by water content and blast-furnace slag. Fine aggregate, super-plasticizer, and coarse aggregate contribute marginally, while fly ash has the least influence. The consistency between SHAP distribution patterns and established concrete material behavior validates both the physical relevance of the model and the robustness of the learned relationships.

Discussion

The CatBoost model developed for predicting concrete compressive strength exhibits high predictive capability and generalization across training, testing, and validation phases. R^2 values above 0.93 for unseen data, combined with low RMSE (≈4.2 MPa) and MAPE below 10%, place this model within or better than the accuracy range reported for CatBoost-based concrete strength predictors in recent literature, where R^2 commonly lies between 0.91 and 0.98 and MAPE between about 2–8% (Beskopylny et al., 2024; S. Lee et al., 2022b). Error magnitudes are comparable to or slightly higher than models trained on more homogeneous or narrower-strength datasets (e.g., UHPC or vibrocentrifuged concretes with RMSE ≈2–3 MPa and MAPE ≈2–6%) (Ahmed et al., 2024b; Fu et al., 2025; Islam et al., 2024), which is reasonable given the broader strength range (2.3–82.6 MPa) in the present work.

Residual analysis reinforces the absence of serious overfitting. The near-zero mean residual in training and small positive means in testing and validation reflect negligible bias, consistent with other CatBoost applications where residuals remain symmetrically distributed around zero and large errors are confined to a few complex mixes(Aswal et al., 2024; Pal et al., 2023; Rathakrishnan et al., 2022b; Tang, 2025). The comparable standard deviations and interquartile ranges of residuals for testing and validation indicate stable behavior on unseen data, aligning with studies that regard RMSE-to-range and CVRMSE values in the 10–15% band as acceptable for civil engineering prediction tasks. Extreme residuals, particularly at high strengths, mirror observations in UHPC and RAC models where mix heterogeneity, fibers, or high slag/fly ash contents introduce nonlinearities that are hard to capture fully, even for advanced boosting ensembles (Elshaarawy et al., 2024; Gao et al., 2023; Katlav & Ergen, 2024; Khodadadi et al., 2024; Naciri et al., 2025)

From a mechanistic standpoint, the SHAP-based feature ranking shows strong agreement with independent explainable-ML studies on concrete. Age and cement content being the dominant positive contributors, followed by the penalizing effect of higher water content, reproduce patterns consistently reported for ordinary, high-performance, recycled, and confined concretes (Beskopylny et al., 2022; Shi et al., 2024). The beneficial but sometimes delayed contribution of blast-furnace slag and fly ash at higher dosages is also widely documented, with SHAP analyses in multiple works highlighting curing age × SCM (supplementary cementitious material) interactions as key nonlinear drivers of strength. The relatively modest influence of fine and coarse aggregates and superplasticizer content, compared with binder-related variables and age, matches broader findings that aggregate and admixture roles are more indirect, operating mainly through packing density and workable low water–binder ratios (Ahmed et al., 2024a; Alsaadawi et al., 2025)

In the broader context of model choice, the reported CatBoost performance is in line with or superior to many alternative ensembles such as Random Forest, GBM, and XGBoost when applied to similar multicomponent concrete datasets. Several comparative studies conclude that CatBoost either ranks first or is statistically indistinguishable from the best competitor, especially when combined with careful hyperparameter optimization and SHAP-based feature scrutiny. This consistency across different concrete types normal, high-performance, recycled, fiber-reinforced, and UHPC supports the reliability of using CatBoost as a general framework for compressive strength prediction and mix design support in engineering practice (Beskopylny et al., 2022; Khodadadi et al., 2024; Mustapha et al., 2024; Tabani & Biswas, 2025)

Conclusion and recommendation

Conclusion

- The CatBoost-based predictive framework demonstrated strong capability in modeling the nonlinear relationship between concrete mix constituents, curing age, and compressive strength, achieving high accuracy and stable generalization across training, testing, and validation datasets.
- The comprehensive performance evaluation using variance-based, absolute, relative, and normalized error metrics confirmed the robustness and reliability of the developed model for concrete strength prediction.
- Residual assessment revealed error distributions centred near zero with comparable dispersion for unseen data, indicating minimal bias and effective control of overfitting.
- SHAP-based interpretability analysis provided transparent insights into model behavior, identifying age of testing and cement content as the most influential parameters, followed by water content and supplementary cementitious materials, in agreement with established concrete behavior.
- The consistency between statistical performance, graphical analysis, and domain knowledge validates the suitability of the CatBoost model as a reliable decision-support tool for concrete mix design and performance prediction.

Recommendations

- Future studies should incorporate additional durability-related parameters, such as curing conditions, temperature, and chemical exposure, to extend the applicability of the model beyond compressive strength prediction.
- Hybrid modeling approaches combining CatBoost with optimization techniques may be explored to directly recommend optimal mix proportions for targeted strength and sustainability objectives.
- Expanding the dataset with field-scale and site-specific concrete data could further improve model robustness and practical relevance for real-world construction applications.
- Comparative studies involving deep learning and physics-informed machine learning models are recommended to benchmark performance and interpretability against the proposed approach.
- The integration of explainable AI tools, such as SHAP, should be adopted as a standard practice in civil engineering machine learning studies to enhance trust, transparency, and engineering acceptance of predictive models.

References

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

Abuodeh, O., Abdalla, J., & Hawileh, R. (2020). Assessment of compressive strength of Ultra-high Performance Concrete using deep machine learning techniques. *Appl. Soft Comput.*, *95*, 106552. https://doi.org/10.1016/j.asoc.2020.106552

Ahmad, M., Hu, J., Ahmad, F., Tang, X., Amjad, M., Iqbal, M. J., Asim, M., & Farooq, A. (2021). Supervised Learning Methods for Modeling Concrete Compressive Strength Prediction at High Temperature. *Materials*, *14*. https://doi.org/10.3390/ma14081983

Ahmed, A. H. A., Jin, W., & Ali, M. A. H. (2024a). Comparative analysis of intelligent models for predicting compressive strength in recycled aggregate concrete. *Modeling Earth Systems and Environment*, *10*, 5273–5291. https://doi.org/10.1007/s40808-024-02063-7

Ahmed, A. H. A., Jin, W., & Ali, M. A. H. (2024b). Prediction of compressive strength of recycled concrete using gradient boosting models. *Ain Shams Engineering Journal*. https://doi.org/10.1016/j.asej.2024.102975

Alomari, Y., & Andó, M. (2024). SHAP-based insights for aerospace PHM: Temporal feature importance, dependencies, robustness, and interaction analysis. *Results in Engineering*. https://doi.org/10.1016/j.rineng.2024.101834

Alsaadawi, M., Elshaarawy, M., & Hamed, A. (2025). Concrete compressive strength classification using hybrid machine learning models and interactive GUI. *Innovative Infrastructure Solutions*, *10*. https://doi.org/10.1007/s41062-025-01983-2

Alyami, M., Khan, M., Fawad, M., Nawaz, R., Hammad, A. W., Najeh, T., & Gamil, Y. (2023). Predictive Modeling for Compressive Strength of 3D Printed Fiber-Reinforced Concrete Using Machine Learning Algorithms. *Case Studies in Construction Materials*. https://doi.org/10.1016/j.cscm.2023.e02728

Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P., & Amjad, U. (2022). Prediction of Pile Bearing Capacity Using XGBoost Algorithm: Modeling and Performance Evaluation. *Applied Sciences*. https://doi.org/10.3390/app12042126

Aswal, V. S., Singh, B., & Maheshwari, R. (2024). Machine learning-based model for prediction of concrete strength. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, *8*. https://doi.org/10.1007/s41939-024-00609-x

Beskopylny, A., Stel'makh, S., Shcherban', E., Mailyan, L., Meskhi, B., Razveeva, I., Chernil'nik, A., & Beskopylny, N. (2022). Concrete Strength Prediction Using Machine Learning Methods CatBoost, k-Nearest Neighbors, Support Vector Regression. *Applied Sciences*. https://doi.org/10.3390/app122110864

Beskopylny, A., Stel'makh, S., Shcherban', E., Mailyan, L., Meskhi, B., Razveeva, I., Kozhakin, A., Pembek, A., Elshaeva, D., Chernil'nik, A., & Beskopylny, N. (2024). Prediction of the Compressive Strength of Vibrocentrifuged Concrete Using Machine Learning Methods. *Buildings*. https://doi.org/10.3390/buildings14020377

Boldini, D., Grisoni, F., Kuhn, D., Friedrich, L., & Sieber, S. (2023). Practical guidelines for the use of gradient boosting for molecular property prediction. *Journal of Cheminformatics*, *15*. https://doi.org/10.1186/s13321-023-00743-7

Chicco, D., Warrens, M., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*. https://doi.org/10.7717/peerj-cs.623

Chithra, S., Chinnaraju, K., & Ashmita, F. (2016). A comparative study on the compressive strength prediction models for High Performance Concrete containing nano silica and copper slag using regression analysis and Artificial Neural Networks. *Construction and Building Materials*, *114*, 528–535. https://doi.org/10.1016/j.conbuildmat.2016.03.214

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

Emad, W., Mohammed, A. S., Kurda, R., Ghafor, K., Cavaleri, L., M.A.Qaidi, S., Hassan, A., & Asteris, P. (2022). Prediction of concrete materials compressive strength using surrogate models. *Structures*. https://doi.org/10.1016/j.istruc.2022.11.002

Farooq, F., Ahmed, W., Akbar, A., Aslam, F., & Alyousef, R. (2021). Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *Journal of Cleaner Production*, *292*, 126032. https://doi.org/10.1016/j.jclepro.2021.126032

Feng, D., Liu, Z.-T., Wang, X.-D., Chen, Y., Chang, J., Wei, D., & Jiang, Z.-M. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*. https://doi.org/10.1016/j.conbuildmat.2019.117000

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

Gali, G., Khan, A., khan, S., khalil, S., Shah, Q., & Khan, M. (2023). Experimental Study on the Effects of Freeze-Thaw Progressions and Performance of Soil with Non-Toxic Bio-Enzyme. *Journal of ICT, Design, Engineering and Technological Science*, 7(1), 12–20. https://doi.org/10.33150/JITDETS-7.1.3

Gao, Y., Li, Z., Li, Y., Zhu, Z., & Zhu, J. (2023). Development of chemistry-informed interpretable model for predicting compressive strength of recycled aggregate concrete containing supplementary cementitious materials. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2023.138733

Islam, M. M., Das, P., Rahman, M. M., Naz, F., Kashem, A., Nishat, M. H., & Tabassum, N. (2024). Prediction of compressive strength of high-performance concrete using optimization machine learning approaches with SHAP analysis. *Journal of Building Pathology and Rehabilitation*, *9*. https://doi.org/10.1007/s41024-024-00445-z

Katlav, M., & Ergen, F. (2024). Improved forecasting of the compressive strength of ultra-high-performance concrete (UHPC) via the CatBoost model optimized with different algorithms. *Structural Concrete*, *26*, 212–235. https://doi.org/10.1002/suco.202400163

Khan, A. Q., Awan, H. A., Rasul, M., Siddiqi, Z. A., & Pimanmas, A. (2023). Optimized Artificial Neural Network Model for Accurate Prediction of Compressive Strength of Normal and High Strength Concrete. *Cleaner Materials*. https://doi.org/10.1016/j.clema.2023.100211

Khan, W., Khan, G., Munawar, M., Shah, Q., Khan, M., Daud, K., & Ali, U. (2025). *Transforming Expansive Clays FROM Waste TO Value: Dual Ash Stabilization Strategy FOR Augmenting Black Cotton Subgrade Soil Performance*. *11*, 5065–5073. https://doi.org/10.64252/cz07k049

Khodadadi, N., Roghani, H., De Caso, F., El-Kenawy, E.-S., Yesha, Y., & Nanni, A. (2024). Data-Driven PSO-CatBoost Machine Learning Model to Predict the Compressive Strength of CFRP- Confined Circular Concrete Specimens. *Thin-Walled Structures*. https://doi.org/10.1016/j.tws.2024.111763

Kitani, R., & Iwata, S. (2023). Verification of Interpretability of Phase-Resolved Partial Discharge Using a CNN With SHAP. *IEEE Access*, *11*, 4752–4762. https://doi.org/10.1109/ACCESS.2023.3236315

Lee, S., Nguyen, N. H., Karamanli, A., Lee, J., & Vo, T. (2022a). Super learner machine-learning algorithms for compressive strength prediction of high performance concrete. *Structural Concrete*, *24*, 2208–2228. https://doi.org/10.1002/suco.202200424

Lee, S., Nguyen, N. H., Karamanli, A., Lee, J., & Vo, T. (2022b). Super learner machine-learning algorithms for compressive strength prediction of high performance concrete. *Structural Concrete*, *24*, 2208–2228. https://doi.org/10.1002/suco.202200424

Lee, Y.-G., Oh, J., Kim, D., & Kim, G. (2022). SHAP Value-Based Feature Importance Analysis for Short-Term Load Forecasting. *Journal of Electrical Engineering & Technology*, *18*, 579–588. https://doi.org/10.1007/s42835-022-01161-9

Li, L., Gao, Y., Dong, X., & Han, Y. (2024). Artificial Neural Network Model for Predicting Mechanical Strengths of Economical Ultra-High-Performance Concrete Containing Coarse Aggregates: Development and Parametric Analysis. *Materials*, *17*. https://doi.org/10.3390/ma17163908

Li, P., Zhang, Z., & Gu, J. (2024). Prediction of Concrete Compressive Strength Based on ISSA-BPNN-AdaBoost. *Materials*, *17*. https://doi.org/10.3390/ma17235727

Lin, K., & Gao, Y. (2022). Model interpretability of financial fraud detection by group SHAP. *Expert Syst. Appl.*, *210*, 118354. https://doi.org/10.1016/j.eswa.2022.118354

Marani, A., Jamali, A., & Nehdi, M. (2020). Predicting Ultra-High-Performance Concrete Compressive Strength Using Tabular Generative Adversarial Networks. *Materials*, *13*. https://doi.org/10.3390/ma13214757

Mujtaba, G., Muhammad, N., Shah, Q., Khan, M., & Abbood, N. K. (2023). Water Pollution Hazards and Toxicity Caused by Textile Industries Effluent. *Journal of ICT, Design, Engineering and Technological Science*, 7(2), 22–27. https://doi.org/10.33150/JITDETS-7.2.3

Mustapha, I., Abdulkareem, M., Jassam, T., Alateah, A., Al-Sodani, K. A., Al-Tholaia, M., Nabus, H., Alih, S., Abdulkareem, Z., & Ganiyu, A. (2024). Comparative Analysis of Gradient-Boosting Ensembles for Estimation of Compressive Strength of Quaternary Blend Concrete. *International Journal of Concrete Structures and Materials*, *18*, 1–24. https://doi.org/10.1186/s40069-023-00653-w

Naciri, H., Alaoui, O., Agouni, M., & Xu, J. (2025). Leveraging PSO-Optimized CatBoost, Extra Trees and HistGradientBoosting for Accurate Concrete Strength Prediction. *Cognizance Journal of Multidisciplinary Studies*. https://doi.org/10.47760/cognizance.2025.v05i07.005

Nawaz, S., Buledi, K., Khan, M., Javed, A., Shah, Q., Khan, M., Shah, S., & Ali, U. (2025). *Flexural Performance Evaluation Of Glass Fiber Reinforced Polymer Bars (GFRP) In Doubly Reinforced Beams: An Experimental Approach.*

Pal, A., Ahmed, K., Hossain, F., & Alam, M. (2023). Machine learning models for predicting compressive strength of fiber-reinforced concrete containing waste rubber and recycled aggregate. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2023.138673

Ponce-Bobadilla, A., Schmitt, V., Maier, C., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science*, *17*. https://doi.org/10.1111/cts.70056

Rathakrishnan, V., Beddu, S. Bt., & Ahmed, A. (2022a). Predicting compressive strength of high-performance concrete with high volume ground granulated blast-furnace slag replacement using boosting machine learning algorithms. *Scientific Reports*, *12*. https://doi.org/10.1038/s41598-022-12890-2

Rathakrishnan, V., Beddu, S. Bt., & Ahmed, A. (2022b). Predicting compressive strength of high-performance concrete with high volume ground granulated blast-furnace slag replacement using boosting machine learning algorithms. *Scientific Reports*, *12*. https://doi.org/10.1038/s41598-022-12890-2

Reddy, M., Lomada, R. R., C., V., R, K., Sergei, S., Vatin, N. I., & Joshi, A. (2024). ML prediction and ANN-PSO based optimization for compressive strength of blended concrete. *Cogent Engineering*, *11*. https://doi.org/10.1080/23311916.2024.2380347

Shah, Q., Byemba, K., Gali, G., Muhammad, A., Khan, A., & Khan, M. (2023). To Stabilize Shear Strength Properties of an Unwanted Subgrade Soil Utilizing Rock Dust. *Journal of ICT, Design, Engineering and Technological Science*, 7(2), 1–6. https://doi.org/10.33150/JITDETS-7.2.1

Shi, C., Xi, B., Shen, L., & Liu, C. (2024). Concrete compressive strength prediction model based on RS-Catboost algorithm. *2024 4th International Symposium on Computer Technology and Information Science (ISCTIS)*, 428–431. https://doi.org/10.1109/isctis63324.2024.10698947

Tabani, A., & Biswas, R. (2025). Assessment of compressive strength of ultra-high-performance concrete using advanced machine learning models. *Structural Concrete*. https://doi.org/10.1002/suco.70076

Tang, L. (2025). Machine Learning-Based Prediction of Concrete Compressive Strength and Interpretability Analysis. *Journal of Civil Engineering and Urban Planning*. https://doi.org/10.23977/jceup.2025.070217

Tran, V.-L., Thai, D., & Nguyen, D. (2020). Practical artificial neural network tool for predicting the axial compression capacity of circular concrete-filled steel tube columns with ultra-high-strength concrete. *Thin-Walled Structures*. https://doi.org/10.1016/j.tws.2020.106720

Wani, S. R., & Suthar, M. (2024). Using machine learning approaches for predicting the compressive strength of ultra-high-performance concrete with SHAP analysis. *Asian Journal of Civil Engineering*. https://doi.org/10.1007/s42107-024-01195-6

Wu, Y., Xu, W., Chen, J., Liu, J., & Wu, F. (2025). Prediction of the Shear Strengths of New–Old Interfaces of Concrete Based on Data-Driven Methods Through Machine Learning. *Buildings*. https://doi.org/10.3390/buildings15173137

Xu, L., Yu, X.-L., Zhu, C., Wang, L., & Yang, J. (2025). Prediction of Ultra-High-Performance Concrete (UHPC) Compressive Strength Based on Convolutional Neural Networks. *Materials*, *18*. https://doi.org/10.3390/ma18122851

Yeh, I. (1998). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, *28*, 1797–1808. https://doi.org/10.1016/s0008-8846(98)00165-3

Zhang, Y., Ren, W., Chen, Y., Mi, Y., Lei, J., & Sun, L. (2024). Predicting the compressive strength of high-performance concrete using an interpretable machine learning model. *Scientific Reports*, *14*. https://doi.org/10.1038/s41598-024-79502-z