Scientific Journal of Informatics
Vol. 13, No. 2, May 2026

p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Interpretable XGBoost-Based Explainable AI Model for Concrete Compressive Strength Prediction

Nadia Annisa Maori[1], Nur Aeni Widiastuti[2*], Nor Hidayati[3]
[1, 2]Department of Informatic Engineering, Universitas Islam Nahdlatul Ulama Jepara, Indonesia
[3]Department of Civil Engineering, Universitas Islam Nahdlatul Ulama Jepara, Indonesia

Abstract.
Purpose: The objective of this research is to provide more accurate predictions and transparency in the analysis process, making it easier for engineering professionals to comprehensively interpret, validate, and evaluate the model results.
Methods: The algorithm used is XGBoost with the XAI-SHAP approach, which serves to interpret the contribution of each input feature, like the water-cement ratio, cement content, and concrete age, to the predicted output. This approach ensures that the model is not only accurate but also open and accountable to engineering practitioners. The research methods are data collection, model training, SHAP visualization, and performance evaluation using RMSE, MAE, and R^2 metrics. The dataset used is the Concrete Compressive Strength dataset from the UCI Repository, which consists of 1,030 records.
Result: Subsequent to model training using the XGBoost with optimal hyperparameters identified via Random Search, model performance was evaluated across three scenarios: training data, testing data, and 5-fold cross-validation. The results indicate that the proposed XGBoost exhibits show high performance with an RMSE 4.59 ± 0.63 and R^2 of 93% as well as stability, which is evidenced by consistent performance metrics, without notable signs of overfitting.
Novelty: The novelty of this study lies in the combination of XGBoost with the XAI-SHAP approach for predicting concrete compressive strength, providing both accuracy and interpretability that can work with nonlinear data, as well as its ability to produce results that are easily understood by construction professionals, making it a favorable choice for predicting concrete compressive strength.

Keywords: Compressive strength of concrete, XGBoost, Explainable AI, Machine learning
Received January 2026 / **Revised** February 2026 / **Accepted** March 2026

INTRODUCTION

The rapid integration of Artificial Intelligence (AI) into civil engineering has highlighted the importance of employing advanced construction material, especially for tasks such as predicting concrete compressive strength. Modern high-performance materials are characterized by enhanced durability, ductility, and resistance to environmental influences [1]. Since concrete remains the most widely used material in construction, reliable estimation of concrete strength under compression is crucial for ensuring safety and efficiency in construction projects [2], [3], [4]. Moreover, accurate strength prediction provides a critical foundation for infrastructure design, enabling engineers to optimize structural performance [5][6]. Predicting the compressive strength of concrete presents two major challenges [7]. First, conventional approaches such as linear regression fail to adequately represent the nonlinear interactions among concrete constituents, including cement, water, aggregates and curing age [8], [9], [10], [11]. Second, many machine learning models lack transparency, making it difficult for practitioners to interpret their outputs. Since the physical behavior of concrete is inherently nonlinear, reliance on linear models often results in inaccurate estimations, as highlighted in several comparative investigations [12], [13], [14], [15]. Inaccurate prediction of concrete strength can lead to two critical issues. An *under-designed* structure may lack sufficient strength, thereby compromising safety, while an *over-designed* structure consumes excessive materials and increases construction costs. Consequently, infrastructure may become either unsafe or economically inefficient. Many existing studies emphasize performance metrics or algorithmic comparisons, yet they often neglect the need for interpretability, leaving practitioners unable to fully validate or comprehend model outputs. This lack of transparency undermines confidence in AI-based predictions, particularly in safety-critical applications, and conceals the role of key variables influencing concrete strength. Hence,

*Corresponding author.
Email addresses: nuraeniwidiastuti@unisnu.ac.id (Widiastuti)
DOI: [10.15294/sji.v13i2.41825](https://doi.org/10.15294/sji.v13i2.41825)

there remains a significant gap between the predictive power of AI models and the construction industry's demand for reliability, accountability, and clear validation.

Achieving both accuracy and transparency in predictive modeling is crucial for civil engineering applications. Among various machine learning techniques, XGBoost has proven effective in handling large-scale datasets, managing class imbalance, and capturing complex nonlinear relationships [16], [17], [18]. Studies consistently highlight XGBoost as one of the most reliable algorithms for predictive tasks. To address the challenge of interpretability, Explainable AI (XAI) methods—particularly SHAP (SHapley Additive exPlanations)—provide a valuable framework. It allows researchers and practitioners to identify the contribution of each input variable to the model's output, thereby enhancing trust, enabling validation, and supporting informed decision-making in infrastructure design [5], [19], [20].

With advancements in AI and big data analytics, some of these challenges are being addressed. Algorithms for machine learning, including ensemble methods like XGBoost Regressor, AdaBoost, Gradient Boosting Decision Trees (GBDTs), Bayesian models, Random Forest, SHAP, and XAI, are becoming increasingly widespread. For example, Rana et al. [21] researched the comparison between XGBoost and LR, KNN, RF, DT, SVM, and MLP, and found that XG Boost is more accurate than other algorithms with R^2 88%. However, it has limitations in visualizing its prediction results. Thapa [22] using the XGBoost machine learning algorithm to predicted the compressive strength of concrete, achieved the highest R^2 value of 93%, but the visualization of influential parameters was not presented. Zhang et al. [8] identified the significant impact of curing age, water-cement ratio, and aggregate composition on high-performance concrete. Li et al. [23] apply four ensamble learning models XGBoost, AdaBoost, Gradient Boosting Decision Trees (GBDTs), and Random Forest showed a good performance. Liu et al. [13] predict the concrete compressive strength and explain the contribution of mix ratio factors on the compressive strength using an explainable boosting machine with baysian model. Shahrokhishahraki et al. [24] applied XAI to ultra-high-performance concrete, facilitating result interpretation through feature visibility. Then, when Kashem [19] performed a hybrid machine learning algorithm using the SHAP method, the SHAP investigation revealed the impact and relationship between input variables and output variables. The SHAP analysis results indicated that the input parameters, drying age and steel fiber content, had the highest positive impact on the compressive and flexural strength of UHPC.

To address these challenges, this study proposes an XGBoost based explainable AI Model, complemented by SHAP visualizations that can assess the compressive strength of concrete. The suggested model is helpful in integrating interpretability into the predictive process. This approach empowers civil engineering professionals to comprehensively evaluate model outputs, aligning with the modern construction industry's demands for safety, efficiency, and accountability.

METHODS

This study uses a systematic methodology and adheres to a structured framework for forecasting concrete compressive strength through the integration of the XGBoost algorithm and the XAI-SHAP technique. The process was designed to ensure both predictive accuracy and model robustness. The workflow consists of several sequential phases: data acquisition, preprocessing, model development with XGBoost, performance evaluation, SHAP-based visualization, and subsequent analysis and interpretation. An overview of these stages is illustrated in Figure 1.

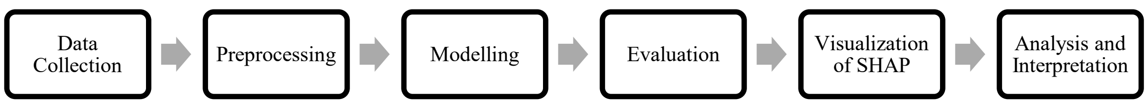

Figure 1. Research method

Data collection

The first stage of this research is data collection. This dataset from UCI Machine Learning Respository in total, 1.030 dataset. The dataset used contains concrete mixture variables such as: cement, slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, age, and concrete compressive strength (MPa).

Table 1. Sample dataset

Cement (component 1) (kg in a m^3 mixture)	Blast Furnace Slag (component 2) (kg in a m^3 mixture)	Fly Ash (component 3) (kg in a m^3 mixture)	Water (component 4) (kg in a m^3 mixture)	Superplasticizer (component 5) (kg in a m^3 mixture)	Coarse Aggregate (component 6) (kg in a m^3 mixture)	Fine Aggregate (component 7) (kg in a m^3 mixture)	Age (day)	Concrete compressive strength (MPa, megapascals)
540,0	0,0	0,0	162,0	2,5	1040,0	676,0	28	79,99
540,0	0,0	0,0	162,0	2,5	1055,0	676,0	28	61,89
332,5	142,5	0,0	228,0	0,0	932,0	594,0	270	40,27
332,5	142,5	0,0	228,0	0,0	932,0	594,0	365	41,05
198,6	132,4	0,0	192,0	0,0	978,4	825,5	360	44,30
266,0	114,0	0,0	228,0	0,0	932,0	670,0	90	47,03
380,0	95,0	0,0	228,0	0,0	932,0	594,0	365	43,70
380,0	95,0	0,0	228,0	0,0	932,0	594,0	28	36,45
266,0	114,0	0,0	228,0	0,0	932,0	670,0	28	45,85
475,0	0,0	0,0	228,0	0,0	932,0	594,0	28	39,29
198,6	132,4	0,0	192,0	0,0	978,4	825,5	90	38,07
198,6	132,4	0,0	192,0	0,0	978,4	825,5	28	28,02
427,5	47,5	0,0	228,0	0,0	932,0	594,0	270	43,01
190,0	190,0	0,0	228,0	0,0	932,0	670,0	90	42,33
304,0	76,0	0,0	228,0	0,0	932,0	670,0	28	47,81
380,0	0,0	0,0	228,0	0,0	932,0	670,0	90	52,91
139,6	209,4	0,0	192,0	0,0	1047,0	806,9	90	39,36
342,0	38,0	0,0	228,0	0,0	932,0	670,0	365	56,14
380,0	95,0	0,0	228,0	0,0	932,0	594,0	90	40,56
475,0	0,0	0,0	228,0	0,0	932,0	594,0	180	42,62

Preprocessing

Before model development, the dataset underwent a series of preprocessing procedures. These steps involve detecting and removing outliers, handling missing or duplicate entries, and applying normalization to ensure consistency across variables. Such preprocessing is crucial for improving data quality and enhancing the reliability of subsequent modeling results.

XGBoost modelling

The predictive modeling stage employed the XGBoost algorithm, a widely recognized ensemble learning technique known for its accuracy and efficiency in both research and practical applications. XGBoost constructs models by iteratively combining decision trees, where each subsequent tree corrects the errors of the previous ones. The algorithm is based on gradient boosting, which optimizes the loss function by leveraging the gradient of prediction errors [8]. In this study, XGBoost was implemented using Python, with specific attention to reducing overfitting and managing imbalanced data distributions, thereby ensuring robust and generalizable performance. The formula for the sample predicted value shown below.

$$\hat{}y(n)i = f1(xi) + \cdots + ft - 1(xi) + ft(xi) = \hat{}y(n-1)i + ft(xi) \quad (1)$$

Evaluation

To evaluate the capability of the model proposed in this research, three indicators were considered, including the RMSE, MAE and R^2. These evaluation metrics are calculated as follows:

1) *Root Mean Square Error* (RMSE)
   Measures how far the model's predictions are from the true value with larger penalties for large errors.

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

2) *Mean Absolute Error* (MAE)
   Measures the average absolute difference between the predicted value and the actual value.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}| \quad (3)$$

3) *Coefficient of Determination* (R^2)
Measures how well the model explains the variability of the actual data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

Visualization of shapley additive explanations (SHAP)

SHAP (SHapley Additive exPlanations) is an approach based on theory that helps explain and show the results of machine learning models [25]. It explains a prediction by measuring the contribution of each feature to that specific prediction. SHAP is a common way to explain the outcomes of machine learning models. It gives each feature a contribution value that shows how much it affects the prediction, based on the Shapley value from cooperative game theory [26]. Complex models that are difficult for experts to understand, such as ensemble models or deep learning, often achieve the highest accuracy for modern data, creating a tension between accuracy and interpretability. Interpretative models can provide engineers, researchers, and stakeholders with an understanding of how they can trust the model's results and comprehend the reasons behind those predictions. SHAP memberikan setiap atribut nilai yang penting untuk prediksi. Tujuan dari metode ini adalah untuk menghitung variabel dan menjelaskan prediksi global dan lokal dengan menghitung nilai Shapley [5].

$$\phi j = \sum |S|!\,(M - (|S| - 1)!\,M!\,s \subseteq M(j)(v(S \cup \{j\}) - v(S)) \tag{5}$$

Certain advantage of SHAP [7] because:

1) It takes into account how features interact with one other, which helps us understand better how numerous features affect predictions.
2) It can explain both individual forecasts and the importance of all features, so it has both local and global interpretability.
3) It makes sure that things are fair and consistent since Shapley values follow basic rules like additivity and symmetry. This makes them more dependable than importance measurements that are based on heuristics.

At this point, the SHAP library is used by the visualization to look at how features affect the output and make global feature importance, dependency, and power charts.

Analysis and interpretation

The terms "interpretability" and "explainability" are often used collectively by researchers; however, there is no concrete mathematical definition for "interpretability" or "explainability," nor has it been measured using any metric. Doshi-Velez and Kim define interpretability as "the ability to explain or present in terms that are easily understood by humans." Based on this definition, interpretability is largely related to the intuition behind a model's output; the idea is that the easier a machine learning system is to interpret, the easier it is to identify cause-and-effect relationships within it [27]. The results of the analysis are used to explain which features most influence the strength of the concrete.

RESULT AND DISCUSSION

This study uses the Concrete Compressive Strength dataset from the UCI Machine Learning Repository, which consists of 1,030 samples with 8 parameters and 1 output target. The parameters are cement, blast furnace slag, water, fly ash, superplasticizer, coarse aggregate, fine aggregate, age, and concrete compressive strength [28], [29], [30]. For an explanation, please refer to Table 2 the dataset includes experimental data from various concrete mix designs with variations in material composition and curing time (age).

Table 2. Description of feature

Feature	Unit	Description	Role
Cement	kg/m³	The main material in concrete production that acts as a binder between fine and coarse aggregates when reacting with water.	Higher cement content generally results in stronger compressive strength.
Blast Furnace Slag	kg/m³	A by-product of steel production (blast furnace slag) used as a partial replacement for cement.	Improves concrete durability, reduces hydration heat, and lowers cement usage.
Fly Ash	kg/m³	Residual ash from coal combustion in steam power plants, used as a supplementary cementitious material (pozzolan).	Enhances workability and crack resistance.
Water	kg/m³	A crucial component that reacts with cement to form binding paste and enables concrete hardening.	Must be balanced: excess reduces strength, deficiency hinders hydration.
Superplasticizer	kg/m³	A chemical additive that increases concrete workability without adding more water.	Makes the mix more fluid and easier to pour, aiding in high-strength concrete production.
Coarse Aggregate	kg/m³	Coarse materials like gravel or crushed stone that form the backbone of concrete structure.	Provides strength and stability, influencing the concrete's composition.
Fine Aggregate	kg/m³	Fine materials like sand that fill gaps between coarse aggregates and reinforce the concrete structure.	Helps produce a dense and compact concrete surface.
Age	Day	The duration concrete is allowed to harden before its strength is tested.	Longer age generally leads to higher compressive strength.
Compressive Strength	MPa	A measure of concrete's ability to withstand compressive load per unit area before failure or cracking.	Represents the target variable (Y) used in machine learning models.

After the data is obtained, the next step is data preprocessing, which includes cleaning missing values and duplicate data. From this step, it was found that all attributes in the dataset do not have missing values, so there is no need for data removal processes as shown in the following Figure 2.

```
df.isnull().sum()

Cement (component 1)(kg in a m^3 mixture)              0
Blast Furnace Slag (component 2)(kg in a m^3 mixture)  0
Fly Ash (component 3)(kg in a m^3 mixture)             0
Water  (component 4)(kg in a m^3 mixture)              0
Superplasticizer (component 5)(kg in a m^3 mixture)    0
Coarse Aggregate  (component 6)(kg in a m^3 mixture)   0
Fine Aggregate (component 7)(kg in a m^3 mixture)      0
Age (day)                                              0
Concrete compressive strength(MPa, megapascals)        0
dtype: int64
```

Figure 2. Missing value examination results

As for cleaning duplicate data, which are rows containing identical values in all columns, 25 identical data points were found. Cleaning needs to be done so that there is no duplicate data that could affect the analysis results.

To maintain data quality, these duplicates are removed to make the results of the analysis or predictive model more accurate. After removing duplicates from the dataset, the number of data points was reduced to 1,005 from the original 1,030. After the next stage of data cleaning comes outlier detection and handling. An outlier is a data value that deviates significantly from most other data. Outliers can occur due to recording errors, input errors, or simply the natural variation in the data. Based on the outlier detection results, it can be concluded that most variables have a small number of outliers, such as blast furnace slag, superplasticizer, and fine aggregate. The variable age has the highest number of outliers, with 59 data points. This is due to the highly diverse testing times for the concrete. The variables water and superplasticizer also have a number of outliers that need to be addressed because they can affect the results of predicting concrete compressive strength. To address the outliers, rows of data containing extreme outliers were removed, particularly for the variables Age, Water, and Superplasticizer. The Interquartile Range (IQR) method detected and got rid of values that were too high or too low. Values that were below Q1 - 1.5 IQR or above Q3 + 1.5 IQR were not counted since they were outliers. So, 59 occurrences were taken out of "Age," 15 from "Water," 10 from "Superplasticizer," and some other properties were adjusted a little bit. This cut the total number of samples from 1,005 to 911.The outlier cleaning technique cut the number of data points down to 911 datasets, but this can make the dataset better. The cleaned dataset (df_clean) has a more even distribution and is ready to be used in the next steps of pre-processing, like normalization and splitting the data into training and testing sets. The model does not experience severe overfitting because the difference is not very big (the R^2 value is 0.9955 on the training data and 0.9418 on the test data) and the 5-fold cross-validation results are consistent (the average R^2 value is 0.9119 with a variation of ±0.0266). Using regularization approaches (such gamma = 0.1) and stopping early during training also helps reduce the danger of overfitting, which is in keeping with best practices for training models.

During the normalization phase, the StandardScaler technique is applied to standardize the scale of all features in the dataset, ensuring that each variable contributes equally to the learning process. This procedure is crucial for improving the model's predictive ability in estimating concrete compressive strength. After normalization, the dataset, consisting of 911 records, was divided into training and testing subsets, with 80% (728 samples) allocated for training and 20% (183 samples) reserved for testing. This division allows the model to learn from the majority of the data while the remaining portion is used to objectively evaluate its performance, thereby reducing the risk of overfitting.

XGBoost modelling

The hyperparameter optimization process was performed using the Randomized Search Cross Validation (RandomizedSearchCV) method on the XGBoost model. This method is used to find the best parameter combination that provides optimal performance on the training data, with the evaluation metrics being the coefficient of determination (R^2) and 5-fold cross validation.

```
Algorithm: RandomizedSearchCV for XGBoost Hyperparameter Optimization

Input:
- Dataset: X (features), y (target: concrete compressive strength)
- Parameter_space: Dictionary of hyperparameters to search, e.g.,
  - n_estimators: range [100, 200, 300, ..., 1000]
  - max_depth: range [3, 4, 5, ..., 10]
  - learning_rate: range [0.01, 0.05, 0.1, ..., 0.3]
  - subsample: range [0.6, 0.7, 0.8, 0.9, 1.0]
  - colsample_bytree: range [0.6, 0.7, 0.8, 0.9, 1.0]
  - gamma: range [0, 0.1, 0.2, ..., 0.5]
- n_iterations: Number of random combinations to try (e.g., 50)
- cv_folds: Number of cross-validation folds (e.g., 5)
- scoring_metric: Metric to optimize (e.g., R²)

Output:
- Best_params: The best hyperparameter combination
- Best_score: The highest score (e.g., average R² from CV)

Procedure:

1. Initialize best_score = -infinity  // Or 0 for metrics like R²
2. Initialize best_params = empty dictionary

3. For i from 1 to n_iterations:
   a. Randomly sample a combination params from Parameter_space
   b. Initialize model = XGBoostRegressor with params
   c. Perform cross-validation:
      - Split dataset into cv_folds folds
      - For each fold:
        - Train model on training fold
        - Predict on validation fold
        - Compute fold_score using scoring_metric (e.g., R² between y_true and y_pred)
      - Compute cv_score = average of fold_scores
   d. If cv_score > best_score:
      - Update best_score = cv_score
      - Update best_params = params

4. Return best_params and best_score
```

Figure 3. Pseudocode of random search

We used RandomizedSearchCV with 50 iterations and 5-fold cross-validation to tune the hyperparameters. The search range is: n_estimators [100, 1000], learning_rate [0.01, 0.2], max_depth [3, 10], min_child_weight [1, 10], gamma [0, 5], subsample [0.5, 1.0], and colsample_bytree [0.5, 1.0]. The optimum values for the parameters are n_estimators=500, max_depth=5, learning rate=0.1, subsample=0.8, colsample_bytree=0.8, and Gamma=0.1. Based on the results of 50 random parameter combinations, the best configuration obtained can be seen in Table 3.

Table 3. The best parameters

Parameter	Nilai
n-estimators	500
Max_depth	5
Learning_rate	0.1
Subsample	0.8
Colsample_bytree	0.8
Gamma	0.1

In Table 3, the most effective parameter values are produced with an R^2 score of 0.928 during the cross-validation process. This value indicates that the model is able to explain approximately 93% of the target data variation based on the features used in the training data.

After training using XGBoost with the best hyperparameters from Random Search, the model's performance was evaluated using 3 models: training data, testing data, and 5-fold cross-validation. The evaluation results can be seen in Table 4.

Table 4. Performance metric of XGBoost

Metrik	Training Set	Testing Set	5-fold Cross Validation
R^2	0.9955	0.9418	0.9119 ± 0.0266
RMSE (MPa)	1.06	3.89	4.59 ± 0.63
MAE (MPa)	0.47	2.66	3.06 ± 0.27
MAPE (%)	1.7	10.8	10.6 ± 0.6

Based on the test results in Table 4, the R^2 value of 0.9955 on the training data indicates that the model is able to explain 99.55% of the variation in the target data. Meanwhile, the R^2 value on the test data of 0.9418 indicates that the model has adequate generalization ability to new data. Cross-validation results also support these findings with an average R^2 value of 0.9119 ± 0.0266, indicating stable model performance across various data subsets.

The prediction error metrics, specifically Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), exhibited acceptable results. The RMSE of 3.89 MPa and MAE of 2.66 MPa, observed on the test dataset, suggest a minor discrepancy between the model's predictions and the observed data. Moreover, the cross-validation findings revealed consistent performance, with an RMSE of 4.59 ± 0.63 MPa and an MAE of 3.06 ± 0.27 MPa.

In terms of relative error, the Mean Absolute Percentage Error (MAPE) of 10.8% in the test data indicates that the average model prediction error is only about 10% of the actual value. The MAPE value in cross-validation of 10.6 ± 0.6% also confirms that the model has a low and stable error rate across various data subsets.

If compared to other algorithms, the performance of each algorithm in prediction can be seen in Table 5.

Table 5. The proposed model was compared with other algorithms

Model	R^2	RMSE(MPa)	MAE(MPa)
Linear Regression	0.78	8.95	6.12
Random Forest	0.88	5.44	3.75
SVM	0.86	7.01	5.34
XGBoost	0.93	4.59	3.06

The comparative analysis presented in Table 5 highlights the superior performance of the proposed XGBoost model compared to other widely used regression algorithms. While traditional models like Linear Regression only achieve moderate predictive capability ($R^2 = 0.78$, RMSE = 8.95 MPa), ensemble methods like Random Forest and SVM improve performance but still fall short in terms of error reduction. the proposed XGBoost Regressor consistently outperformed all benchmarks with $R^2 = 0.93$ and RMSE = 4.59 MPa.

These findings confirm that gradient boosting techniques are highly effective in capturing nonlinear interactions among concrete mix variables. Additionally, the integration of SHAP analysis provides interpretability that is often lacking in other high-performing models. For example, although RF/SVM can achieve relatively high accuracy, their "black box" nature limits practical adoption in engineering contexts where transparency is crucial.

Finally, the benchmarking results show that the proposed XGBoost + SHAP framework offers both predictive superiority and transparent interpretability, positioning it as a practical tool for construction professionals. By combining high performance with clarity, this model addresses two challenges identified in the literature: achieving accuracy while maintaining accountability in safety-critical applications.

Feature importance analysis using the SHAP method in XAI

Global feature importance analysis using the XGBoost algorithm was conducted to determine the extent to which each input variable influences the prediction of concrete compressive strength. The gain value for each feature indicates its relative contribution to improving model accuracy. The higher the gain value, the greater the feature's influence on the prediction results.

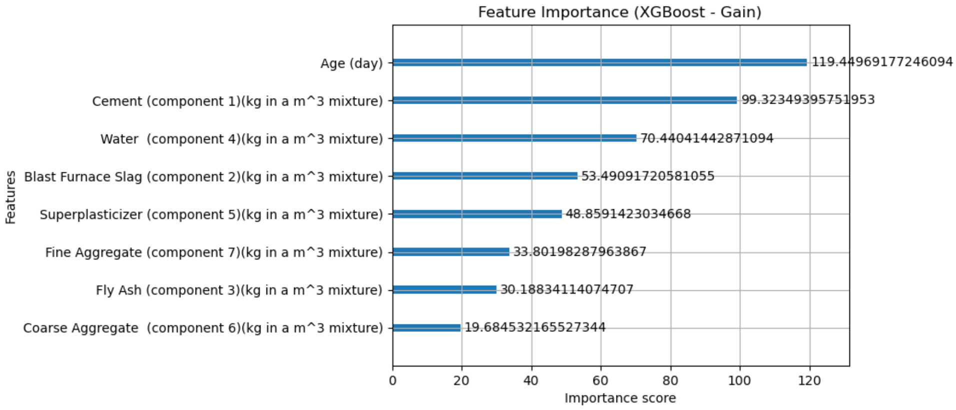

Figure 4. Result global feature importance analysis

The analysis results in Figure 4 show that age (in days) is the most important feature, with a gain value of 119.45, and it is the most important factor in lowering the model's prediction error. This indicates that the variation in the age feature can explain a larger portion of the variance in predicting concrete compressive strength compared to other features. This is in line with the idea that the longer the compaction process takes, the stronger the bond between the cement particles becomes since the hydration reaction keeps going

Cement (component 1) ranks second with a gain value of 99.32, indicating that the amount of cement in the mixture significantly determines concrete strength because it acts as the primary binding agent. This is followed by water (component 4) with a gain value of 70.44. These results emphasize the importance of adjusting the water-cement ratio to achieve optimal strength.

Other variables, such as Blast Furnace Slag (component 2) and Superplasticizer (component 5), also have a moderate influence, with gain values of 53.49 and 48.86, respectively. Both play a role in increasing concrete durability and workability without sacrificing strength. Meanwhile, fine aggregate, fly ash, and coarse aggregate have a relatively small influence but still play an important role in maintaining the stability and homogeneity of the mixture.

Overall, these results confirm that concrete age and cement content are the primary factors determining compressive strength, while other variables play a role in supporting the physical and mechanical characteristics of concrete. The XGBoost model is not only effective in predicting concrete strength but also provides a clear interpretation of the relationships between its constituent variables.

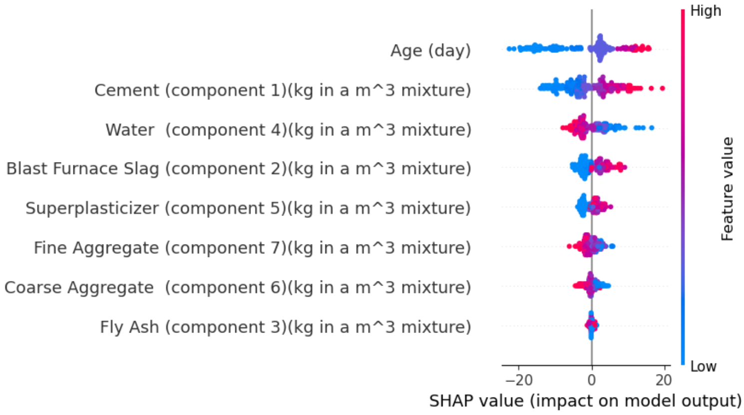

Figure 5. SHAP value (impact on model output)

Based on the SHAP analysis results in Figure 5, we obtained an overview of the direction of influence of each feature on the prediction of concrete compressive strength. A positive SHAP value indicates that increasing the value of a feature will improve the predicted compressive strength, while a negative SHAP value indicates the opposite.

The Age (days) feature shows a dominant positive influence, where high age values contribute positively to increasing concrete compressive strength. This aligns with the characteristics of the cement hydration process, which continues over time, resulting in increased concrete strength.
The Cement feature (component 1) also has a strong positive influence. Increasing the amount of cement in the concrete mix contributes directly to increasing compressive strength because it serves as the primary binding agent. Conversely, the water feature (component 4) shows a negative influence, meaning that increasing the water content in the mix tends to decrease the concrete's compressive strength. This decline is caused by increased porosity due to an excessively high water-cement ratio.

The Blast Furnace Slag (component 2) and Superplasticizer (component 5) features exhibit a moderate positive effect on the prediction results, especially at high feature values. These additives play a role in increasing the durability and workability of concrete without sacrificing its strength. Meanwhile, the fine aggregate, coarse aggregate, and fly ash features exhibit relatively small and variable effects on the model output.

Overall, the SHAP analysis results indicate that concrete age and cement content are factors that consistently increase compressive strength, while excess water content has the opposite effect. This finding aligns with basic theory regarding the relationship between water-cement ratio, concrete age, and compressive strength.

SHAP Dependence Plots are used to understand the relationship between specific feature values (X-axis) and their impact on the prediction of concrete compressive strength (Y-axis). These plots can show interactions between variables and capture complex nonlinear patterns learned by the XGBoost model. An analysis of several key features that most influence the model can be seen in Figure 6.

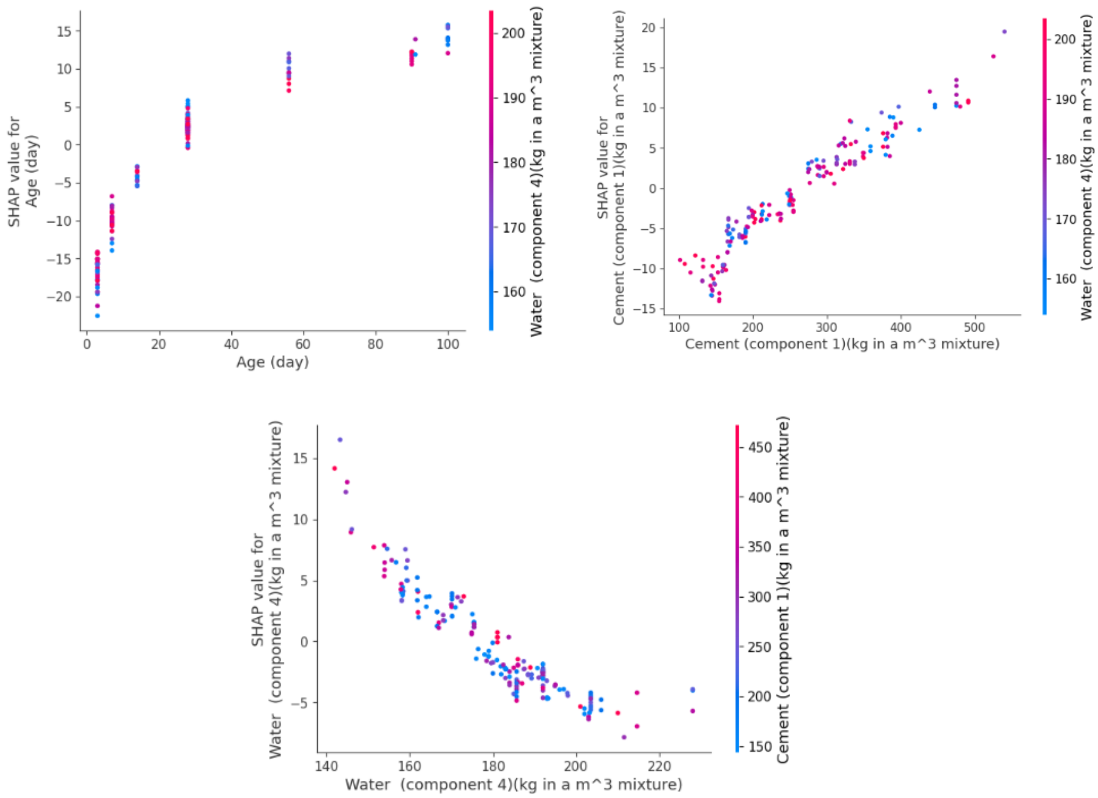

Figure 6. SHAP dependence plot by ages, cement, and water

Based on the results of the tests and interpretive visualization analysis that have been conducted, this study successfully developed a concrete compressive strength prediction model that is not only accurate but also transparent (explainable).

The XGBoost algorithm demonstrated excellent performance with an R^2 value of 0.93. This is almost the same as his research findings [22]. The advantage of using XAI-SHAP is its ability to model non-linear patterns, as shown in the dependence plot for the Age feature in Figure 6, proving its superiority over conventional linear regression methods. This model is capable of detecting the "saturation" phase in the concrete hardening process after 28 days, an important physical phenomenon that is often not captured by simple linear models. The SHAP dependence plot for age reaches a plateau after about 28 days, in accordance with standard concrete curing practices where 90% of strength is achieved at 28 days.

SHAP offers an interpretative framework by aggregating contributions across each feature, hence enabling the disaggregation of the XGBoost model's predictions into the impact of individual variables. This method makes it possible to get interpretability in both Local Explanation and Global Explanation.

Based on previous research [6], [21], it shows that the performance of XGBoost in predicting concrete compressive strength is very good, but the visualization of the influencing parameters is not presented. Therefore, in the research, an approach using the XGBoost algorithm combined with the XAI-SHAP technique was employed to represent the parameters of concrete compressive strength. The XGBoost algorithm can model the complex and multivariate characteristics of concrete, and the performance of XAI-SHAP indicates that the most important parameters/features in determining concrete compressive strength are age, cement, and water, as the top three features. This model can be used to optimize the composition of concrete mixtures (reducing cement while maintaining strength), which has the potential to lower production costs and carbon emissions while still ensuring structural safety standards.

CONCLUSION

This study demonstrates that the integration of the XGBoost algorithm with an explainable AI framework based on SHAP provides high predictive accuracy and interpretability in evaluating concrete compressive strength. The proposed model not only achieves reliable performance in various evaluation scenarios but also offers transparency by revealing the contribution of each input variable to the prediction results. The experimental results show high performance with an RMSE metric of 4.59 ± 0.63, which is considered low, and an R^2 value of 93%, which is considered good, while maintaining stable performance. Such clarity is very important for civil engineering professionals, as it allows for the validation of model results and supports informed decision-making in infrastructure design. However, this research only examines compressive strength. Further research can be developed by expanding the dataset using field data (real-world) or external variations to make the model more representative of real conditions. The addition of new parameters, such as temperature, humidity, or additional types of materials, as well as sensitivity analysis and interaction between variables, can enhance the depth of interpretation and enable the development of multi-output models. Researchers are also advised to compare performance with other AI models or hybrid approaches to obtain a more comprehensive benchmark.

REFERENCES

[1] A. Ahmed, W. Song, Y. Zhang, M. A. Haque, and X. Liu, "Hybrid BO-XGBoost and BO-RF Models for the Strength Prediction of Self-Compacting Mortars with Parametric Analysis," *Materials*, vol. 16, no. 12, 2023, doi: 10.3390/ma16124366.

[2] W. Wang, Y. Zhong, G. Liao, Q. Ding, T. Zhang, and X. Li, "Prediction of Compressive Strength of Concrete Specimens Based on Interpretable Machine Learning," *Materials*, vol. 17, no. 15, Aug. 2024, doi: 10.3390/ma17153661.

[3] A. Song, X. Yu, N. Khodadadi, Y. Guo, and A. Nanni, "Explainable Machine Learning Framework with Experimental Validation for Strength Prediction of Magnesium Phosphate Cement," *Int. J. Concr. Struct. Mater.*, vol. 19, no. 1, p. 117, Nov. 2025, doi: 10.1186/s40069-025-00856-3.

[4] G. Chen, S. A. Suhail, A. Bahrami, M. Sufian, and M. Azab, "Machine learning-based evaluation of parameters of high-strength concrete and raw material interaction at elevated temperatures," *Front. Mater.*, vol. Volume 10-2023, 2023, doi: 10.3389/fmats.2023.1187094.

[5] N. Aqmar, H. Wijayanto, and F. Mochamad Afendi, "Performance Analysis of Machine Learning Models using RFE Feature Selection and Bayesian Optimization in Imbalanced Data Classification with Shap-Based Explanations," *Scientific Journal of Informatics*, vol. 12, no. 3, pp. 539–554, Oct. 2025, doi: 10.15294/sji.v12i3.31459.

[6] J. Thapa, "Machine Learning Approaches for Predicting Concrete Compressive Strength," *Journal of Advanced Research in Civil and Environmental Engineering*, vol. 11, pp. 9–20, May 2024, doi: 10.24321/2393.8307.202404.

[7] S. Talatahari, F. Chen, and A. H. Gandomi, "Developing a robust machine learning framework for predicting the behavior of large-scale structure," *Journal of Building Engineering*, vol. 105, Jul. 2025, doi: 10.1016/j.jobe.2025.112204.

[8] Y. Zhang, W. Ren, Y. Chen, Y. Mi, J. Lei, and L. Sun, "Predicting the compressive strength of high-performance concrete using an interpretable machine learning model," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-79502-z.

[9] M. T. Tiza, O. N. Mogbo, and E. Akande, "Compressive Strength Prediction Using Linear Regression Method," *Journal Int.Environment Application & Science*, vol. 18, p. 100106, Sep. 2023, [Online]. Available: https://www.researchgate.net/publication/374117424

[10] I. Nunez, A. Marani, M. Flah, and M. L. Nehdi, "Estimating compressive strength of modern concrete mixtures using computational intelligence: A systematic review," *Constr. Build. Mater.*, vol. 310, p. 125279, 2021, doi: https://doi.org/10.1016/j.conbuildmat.2021.125279.

[11] H.-T. Thai, "Machine learning for structural engineering: A state-of-the-art review," *Structures*, vol. 38, pp. 448–491, 2022, doi: https://doi.org/10.1016/j.istruc.2022.02.003.

[12] H. Aamir, K. Aamir, and M. F. Javed, "Linear and Non-Linear Regression Analysis on the Prediction of Compressive Strength of Sodium Hydroxide Pre-Treated Crumb Rubber Concrete †," *Engineering Proceedings*, vol. 44, no. 1, 2023, doi: 10.3390/engproc2023044005.

[13] G. Liu and B. Sun, "Concrete compressive strength prediction using an explainable boosting machine model," *Case Studies in Construction Materials*, vol. 18, Jul. 2023, doi: 10.1016/j.cscm.2023.e01845.

[14] H. Nguyen, T. Vu, T. P. Vo, and H. T. Thai, "Efficient machine learning models for prediction of concrete strengths," *Constr. Build. Mater.*, vol. 266, Jan. 2021, doi: 10.1016/j.conbuildmat.2020.120950.

[15] S. Charhate, M. Subhedar, and N. Adsul, "Prediction of Concrete Properties Using Multiple Linear Regression and Artificial Neural Network," *Journal of Soft Computing in Civil Engineering*, vol. 2, no. 3, pp. 27–38, Jul. 2018, doi: 10.22115/SCCE.2018.112140.1041.

[16] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, "Boosting-Based Machine Learning Applications in Polymer Science: A Review," *Polymers (Basel).*, vol. 17, no. 4, Feb. 2025, doi: 10.3390/polym17040499.

[17] D. Siswara, A. M. Soleh, and A. Hamim Wigena, "Classification Modeling with RNN-based, Random Forest, and XGBoost for Imbalanced Data: A Case of Early Crash Detection in ASEAN-5 Stock Markets," *Scientific Journal of Informatics*, vol. 11, no. 3, pp. 569–582, Aug. 2024, doi: 10.15294/sji.v11i3.4067.

[18] Z. T. Pear, T. Anil, H. B. Kibria, N. B. Ghulam Kibria, and N. Islam, "Explainable Stacking-Based Regression Model for Concrete Strength Prediction: PSO vs. Bayesian Optimization," in *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, IEEE, Jul. 2025, pp. 1–6. doi: 10.1109/QPAIN66474.2025.11171696.

[19] P. Das and A. Kashem, "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*, vol. 20, Jul. 2024, doi: 10.1016/j.cscm.2023.e02723.

[20] G. Angelucci, G. Quaranta, F. Mollaioli, and S. K. Kunnath, "Interpretable machine learning models for displacement demand prediction in reinforced concrete buildings under pulse-like earthquakes," *Journal of Building Engineering*, vol. 95, Oct. 2024, doi: 10.1016/j.jobe.2024.110124.

[21] M. S. Rana, M. M. Hossain, and F. Li, "Comparative analysis of machine learning models for predicting the compressive strength of ultra-high-performance steel fiber reinforced concrete," *Journal of Engineering Research (Kuwait)*, 2025, doi: 10.1016/j.jer.2025.01.004.

[22] J. Thapa, "Concrete Compressive Strength Prediction by Ensemble Machine Learning Approach," *Journal of Engineering and Sciences*, vol. 3, no. 1, pp. 66–73, Jun. 2024, doi: 10.3126/jes2.v3i1.66249.

[23] Q.-F. Li and Z.-M. Song, "High-performance concrete strength prediction based on ensemble learning," *Constr. Build. Mater.*, vol. 324, p. 126694, Mar. 2022, doi: 10.1016/j.conbuildmat.2022.126694.

[24] M. Shahrokhishahraki, M. Malekpour, S. Mirvalad, and G. Faraone, "Machine learning predictions for optimal cement content in sustainable concrete constructions," *Journal of Building Engineering*, vol. 82, Apr. 2024, doi: 10.1016/j.jobe.2023.108160.

[25] M. Mersha, K. Lam, J. Wood, A. K. AlShami, and J. Kalita, "Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction," *Neurocomputing*, vol. 599, p. 128111, Sep. 2024, doi: 10.1016/j.neucom.2024.128111.

[26] J. Mukherjee, M. Kar, A. Chakrabarti, and S. Das, "7 - Automated risk prediction of solitary pulmonary nodules," in *Application of Artificial Intelligence in Early Detection of Lung Cancer*, J. Mukherjee, M. Kar, A. Chakrabarti, and S. Das, Eds., Academic Press, 2024, pp. 199–221. doi: https://doi.org/10.1016/B978-0-323-95245-3.00007-X.

[27] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," Jan. 01, 2021, *MDPI AG*. doi: 10.3390/e23010018.

[28] V. Mehta, "Machine learning approach for predicting concrete compressive, splitting tensile, and flexural strength with waste foundry sand," *Journal of Building Engineering*, vol. 70, p. 106363, 2023, doi: https://doi.org/10.1016/j.jobe.2023.106363.

[29] M. Minaz Hossain, M. Nasir Uddin, and M. Abu Sayed Hossain, "Prediction of compressive strength ultra-high steel fiber reinforced concrete (UHSFRC) using artificial neural networks (ANNs)," *Mater. Today Proc.*, 2023, doi: https://doi.org/10.1016/j.matpr.2023.02.409.

[30] P. G. Asteris, A. D. Skentou, A. Bardhan, P. Samui, and K. Pilakoutas, "Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models," *Cem. Concr. Res.*, vol. 145, p. 106449, 2021, doi: https://doi.org/10.1016/j.cemconres.2021.106449.