Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 12, 2025

ACCURATE CONCRETE COMPRESSIVE STRENGTH PREDICTION USING ADABOOST ON MIX PROPORTION DATA

Dr. M. Adil Khan[*1], Hamza Farooq[2], Wasim Khan[3], Shah Khalid[4]

[*1]*Resident Engineer, NESPAK*
[2,3,4]*Department of Civil Engineering, University of Engineering and technology Peshawar Pakistan*

[*1]adee.uol@gmail.com,[2]hamzafarooqshah0@gmail.com,[3]wasim.khan.utmanzai@gmail.com,
[4]sk7190518@gmail.com

DOI: https://doi.org/10.5281/zenodo.17988833

Keywords
AdaBoost regression; Concrete compressive strength; Machine learning; Mix proportion; Curing age; Ensemble learning

Article History
Received: 17 October 2025
Accepted: 30 November 2025
Published: 17 December 2025

Copyright @Author
Corresponding Author: *
Dr. M. Adil Khan

Abstract

Accurate prediction of concrete compressive strength is essential for structural design, quality control, and mix optimization, yet laboratory testing is time-consuming and costly, particularly when diverse binder systems and curing ages must be evaluated. This study presents a data-driven framework for predicting concrete compressive strength using an AdaBoost regression model trained on standard mix-proportion variables and age. A dataset consisting of 1134 concrete mix records was used, including cement, blast-furnace slag, fly ash, water, super-plasticizer, coarse aggregate, fine aggregate, and age of testing as predictors, with compressive strength (MPa) as the response. Exploratory data analysis confirmed substantial heterogeneity and non-normality in several predictors, especially supplementary cementitious materials and curing age, with evidence of skewness, heavy tails, and regime-like behavior (e.g., frequent zero/low SCM contents). Correlation analysis indicated realistic mix-design trade-offs, particularly a strong negative relationship between water and super-plasticizer, while overall multicollinearity remained limited, suggesting that nonlinear interaction effects are more important than simple linear dependence. For model development, AdaBoost was implemented with a decision-tree base estimator, 100 estimators, a learning rate of 1.0, and linear loss for regression. The dataset was partitioned into training and testing subsets using an 80/20 split to provide an unbiased generalization assessment. Model performance was evaluated using multiple complementary metrics, including coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of variation of RMSE (CVRMSE). The trained AdaBoost model achieved strong predictive accuracy on the test set with $R^2 = 0.882$, indicating that a substantial portion of strength variability is captured from mix ingredients and age. Error magnitudes were moderate, with MSE = 30.468, RMSE = 5.520 MPa, and MAE = 4.165 MPa, while MAPE = 17.035% and CVRMSE = 15.402% suggested acceptable relative performance across the strength range. The predicted-versus-observed analysis showed close alignment with the 1:1 line and increased dispersion at higher strengths, highlighting the challenge of generalizing extreme-strength mixes. Feature ranking using univariate regression and RReliefF indicated that cement,

age, water, and super-plasticizer were among the most influential predictors, consistent with concrete strength development mechanisms. Overall, the results demonstrate that AdaBoost provides an effective and interpretable ensemble-learning approach for rapid compressive strength estimation, with practical utility for preliminary mix design decisions and quality-control support.

INTRODUCTION

Concrete compressive strength is a governing property in structural design, quality control, and durability-related decision making. In practice, strength is influenced by a coupled set of mixture proportions (cement, supplementary cementitious materials, water, aggregates, and chemical admixtures) and by curing time, making its behavior highly nonlinear and sensitive to interaction effects. Although standardized laboratory testing remains the accepted basis for certification, experimental programs are time-consuming, resource-intensive, and often impractical when rapid mix adjustments are required in the field or when sustainable binder substitutions introduce additional variability. These constraints have encouraged the adoption of data-driven approaches capable of learning strength patterns directly from historical mix data to support faster decision cycles while reducing testing burden, particularly in contexts where mix optimization, sustainability targets, and production control must be balanced.

Recent studies have shown that machine-learning regression models can provide accurate estimates of compressive strength for conventional, blended, and alternative binder systems, and ensemble learning has emerged as a particularly effective family of methods. Among these, AdaBoost has repeatedly demonstrated strong predictive performance across different concrete types, with reported coefficients of determination commonly in the high range (e.g., approximately 0.89–0.93), indicating that a large proportion of strength variability can be explained when suitable input variables are available (A. Ahmad, Ahmad, Chaiyasarn, et al., 2021; A. Ahmad, Ostrowski, Maślak, et al., 2021; Shen et al., 2022; Kashem et al., 2024; T.-G. Nguyen et al., 2025; Pham, 2025). In addition to high R^2, several investigations have reported engineering-acceptable magnitudes of RMSE and MAE, supporting the argument that predictive models can be practically useful despite inherent material heterogeneity and experimental scatter (Abdullah et al., 2024; Amin et al., 2023; Gao et al., 2023, 2024; Huang et al., 2023; D.-L. Nguyen & Phan, 2024; K. Zhang et al., 2025). However, the literature also emphasizes that performance can deteriorate for ultra-high-strength or highly heterogeneous mixtures and at the extremes of the strength spectrum, where data scarcity and strong nonlinear coupling (e.g., low water/binder ratios combined with high admixture dosages and SCM effects) make generalization more difficult (Cui et al., 2021; Latif, 2021; Lee et al., 2022; Chao et al., 2024; P. Li et al., 2024; Fang et al., 2024; Tak et al., 2025; Khan et al., 2025). This motivates careful, dataset-specific evaluation rather than assuming uniform model behavior across all mix regimes.

In comparative terms, AdaBoost is frequently benchmarked against other ensemble methods such as Random Forest, Gradient Boosting Regression, and modern gradient-boosting variants (e.g., XGBoost). Multiple studies report that these alternatives can surpass AdaBoost under certain conditions, sometimes achieving very high $R^2$and lower errors, particularly when complex high-order interactions dominate or when robustness to noise and outliers is critical (Xu et al., 2021; Hu et al., 2022; Jibril, Malami, et al., 2023; Taffese & Espinosa-Leal, 2023; Kalabarige et al., 2024; Katlav & Ergen, 2024; Ellappan et al., 2025). The reasons are typically attributed to stronger interaction modeling capacity, regularization strategies, and reduced sensitivity to noisy samples relative to classic boosting configurations (Dao et al., 2020; Rathakrishnan et al., 2022; A. Kumar et al., 2022; Wang et al., 2023; Arora et al., 2024; Assegie et al., 2022; L. Chen et al., 2024; H. Li et al., 2024; Logeswaran et al., 2024; Loureiro & Stefani, 2024; Bentegri et al., 2025). Nonetheless, AdaBoost remains attractive because it is conceptually simple,

computationally efficient, and often interpretable when built on decision-tree weak learners, making it a suitable baseline model for engineering datasets where transparency and repeatability are important.

Accordingly, the present work develops an AdaBoost regression model to predict concrete compressive strength from standard mix-proportion variables and curing age using a dataset of 1134 observations. The study emphasizes (i) a structured exploratory analysis of data characteristics and inter-variable relationships, (ii) transparent model development using a decision-tree base estimator under a controlled train/test split, and (iii) multi-metric performance assessment to capture both variance explanation and practical error magnitudes. In addition, feature ranking is used to examine whether the learned importance patterns align with established material understanding where cement content, age, water, and admixture dosage are typically dominant drivers, while SCMs and aggregates often exert influence through interactions and mix-design trade-offs (H. Nguyen et al., 2021; Shishegaran et al., 2020; Tipu et al., 2023; Ahmed et al., 2024a, 2024b; Aswal et al., 2024; Beskopylny et al., 2024; Fei et al., 2023; Harith et al., 2024; Jibril, Zayyan, et al., 2023; Pan et al., 2025; Paudel et al., 2023; Phan, 2023). By positioning AdaBoost within the established literature and providing a critical evaluation of its strengths and limitations, the work aims to support reliable, data-driven strength estimation for mix design support and quality-control applications, while also indicating where more advanced ensembles or hybrid strategies may be warranted (R. Kumar et al., 2024; Liu et al., 2023; Sami et al., 2023).

Methodology

Data characteristics

The statistical characteristics confirm substantial heterogeneity and non-normality across several predictors. Cement shows moderate right skew (0.5) and slightly platykurtic behavior (−0.5), reflecting wide but not extreme dispersion. Blast-furnace slag and fly ash exhibit higher relative variability and positive skewness (0.8 and 0.6), indicating nonuniform utilization across mixtures; the fly ash median absolute deviation of zero strongly suggests that the median is repeated across many records, consistent with frequent absence of fly ash and a distribution dominated by low or zero entries. Water remains near-symmetric (skew 0.1) with low robust coefficient of variation (0.1), supporting its role as a relatively controlled process variable. Super-plasticizer is positively skewed (0.8) and leptokurtic (1.5), indicating occasional high-dosage outliers that can disproportionately influence learning. Coarse and fine aggregates show slight negative skew (−0.2) and low robust variation (0.1), reflecting practical bounds in batching. Age is the most extreme variable (skew 3.5, kurtosis 13.8), demonstrating a heavily tailed distribution that warrants careful evaluation to avoid bias toward the dominant early-age region.

Table 1 Summary statistics describing dataset characteristics for each predictor, including standard error of mean, variance, skewness, kurtosis, interquartile range (Q3–Q1), median absolute deviation, and robust coefficient of variation.

Name	SE of mean	Variance	Skewness	Kurtosis	Interquartile Range (Q3 - Q1)	Median Absolute Deviation	Robust Coefficient of Variation
Cement	3.1	10706.0	0.5	-0.5	152.0	76.0	0.4
Blast-furnace Slag	2.5	7097.5	0.8	-0.5	141.3	26.0	1.5
Fly Ash	2.1	5124.1	0.6	-0.9	122.0	0.0	-
Water	0.6	471.5	0.1	0.1	26.8	14.2	0.1
Super-plasticizer	0.2	33.6	0.8	1.5	10.2	4.9	1.1
Coarse Aggregate	2.5	6853.9	-0.2	-0.4	107.6	56.0	0.1
Fine Aggregate	2.4	6300.2	-0.2	-0.2	101.0	47.5	0.1
Age of testing	1.8	3653.2	3.5	13.8	14.0	14.0	0.7

The descriptive comparison of mean, median, and standard deviation indicates that the predictor set spans markedly different numerical scales, reflecting the physical nature of concrete constituents. Coarse and fine aggregates exhibit the largest central values, followed by cement and water, whereas super-plasticizer remains numerically small. The proximity of mean and median for water and both aggregate fractions suggests relatively stable central tendency and limited distortion from extreme values. In contrast, blast-furnace slag and fly ash show pronounced mean–median separation, implying a mixture of zero/low-content observations combined with a smaller portion of high-replacement mixes; this pattern is consistent with right-skewed distributions and nonuniform usage of supplementary cementitious materials. Age of testing also exhibits a noticeable mean–median difference, indicating an imbalanced concentration of early-age tests with fewer late-age records. Overall, these contrasts highlight that the dataset contains both continuous, well-centred variables (e.g., water) and regime-like variables (e.g., SCM contents), which is expected to promote nonlinear behavior suitable for ensemble learning.

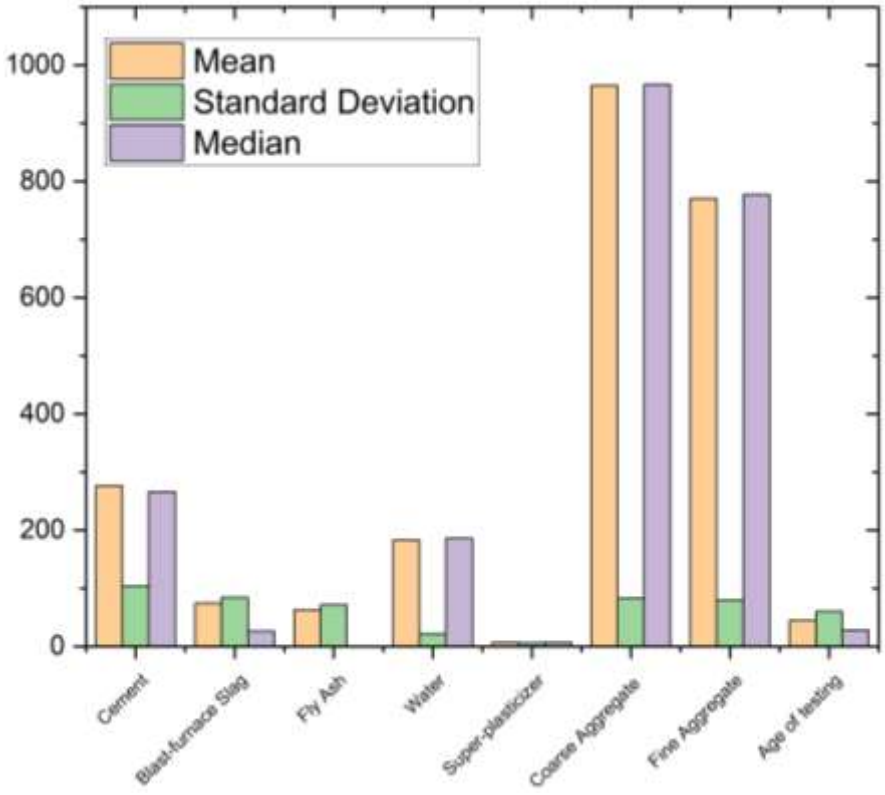


Figure 1 Comparison of mean, median, and standard deviation for the input variables (cement, blast-furnace slag, fly ash, water, super-plasticizer, coarse aggregate, fine aggregate, and age of testing) across the 1134-sample dataset.

The distributional profiles for cement, blast-furnace slag, fly ash, and water demonstrate distinct dispersion and symmetry characteristics. Cement presents a broad spread ($\mu \approx 276.5$, $\sigma \approx 103.5$), indicating that the dataset covers a wide range of binder dosages rather than a narrow design band. Blast-furnace slag ($\mu \approx 74.3$, $\sigma \approx 84.2$) and fly ash ($\mu \approx 62.8$, $\sigma \approx 71.6$) display large standard deviations relative to their means, suggesting substantial heterogeneity and the likely presence of many low or

zero values alongside a long upper tail typical of datasets combining conventional and blended binder systems. Water shows comparatively tight variability (μ≈183.0, σ≈21.7) and a near-symmetric shape, reflecting stronger batching control and a smaller operational range than cementitious substitutions. These contrasting distribution shapes indicate that strength prediction will depend not only on magnitude changes but also on the occurrence of specific mix-design regimes (e.g., SCM-present versus SCM-absent), reinforcing the need for nonlinear modeling.

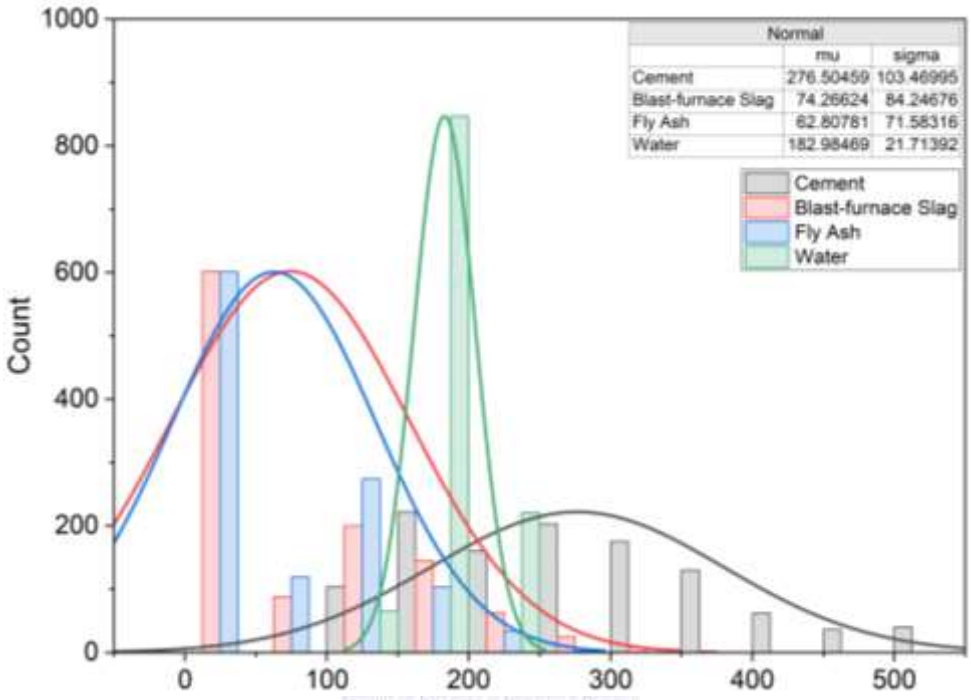


Figure 2 Distribution of cement, blast-furnace slag, fly ash, and water contents shown as histograms with fitted normal curves; the inset table reports the estimated mean (μ) and standard deviation (σ) for each variable.

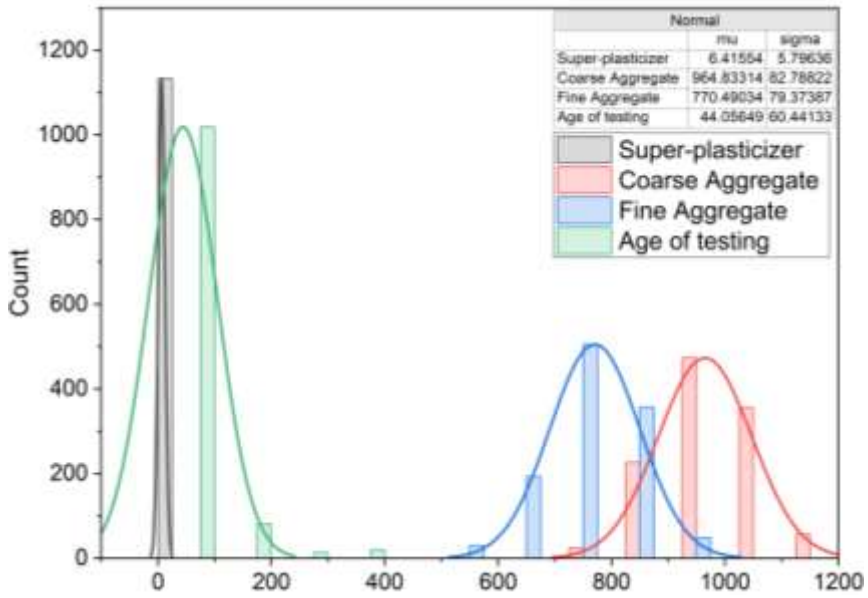


Figure 3 Distribution of super-plasticizer dosage, coarse aggregate, fine aggregate, and age of testing shown as histograms with fitted normal curves; the inset table reports the estimated mean (μ) and standard deviation (σ) for each variable.

The Pearson correlation matrix suggests that the input variables are not dominated by pervasive multicollinearity, yet several relationships reflect realistic mix-design trade-offs. The strongest dependency occurs between water and super-plasticizer ($r \approx -0.588$), indicating that higher admixture dosage is commonly associated with reduced water content, consistent with workability control. Water also correlates negatively with fine aggregate ($r \approx -0.425$) and moderately with coarse aggregate ($r \approx -0.271$), implying that aggregate proportions and water demand are coupled through grading and paste requirement. Cement shows negative correlations with fly ash ($r \approx -0.420$) and slag ($r \approx -0.273$), supporting the interpretation that SCMs act primarily as partial replacements rather than independent additions. Fly ash correlates positively with super-plasticizer ($r \approx 0.353$), suggesting increased admixture demand in mixes containing fly ash. Age demonstrates weak linear association with most ingredients, indicating that it functions as an experimental factor not strongly confounded with specific batching choices. Collectively, the structure points to interaction-driven behavior rather than purely linear dependence.

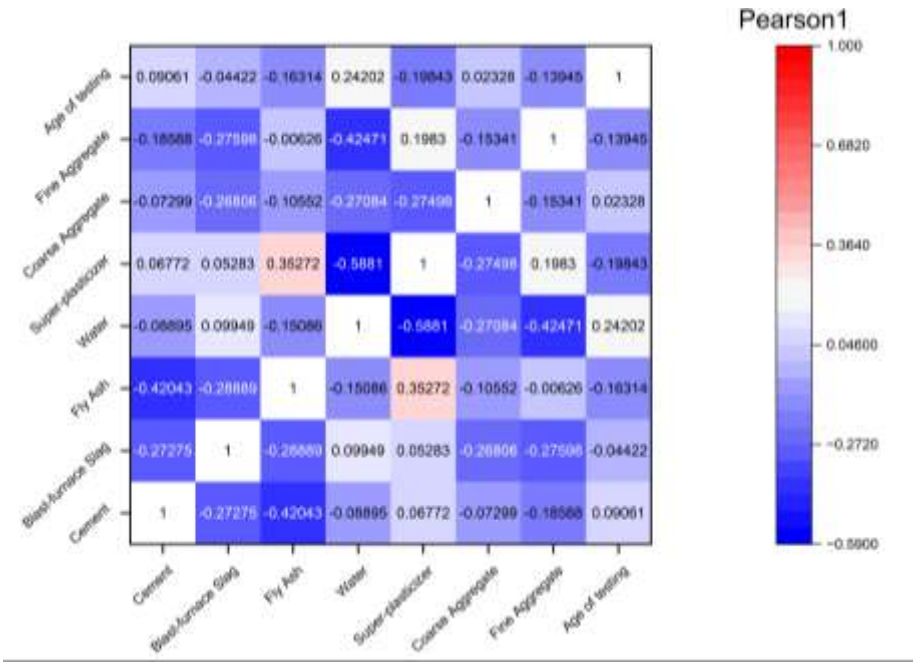


Figure 4 Pearson correlation heatmap of the input variables illustrating linear interdependencies and mix-design trade-offs (red: positive correlation, blue: negative correlation).

Model development

The AdaBoost model was developed as a boosted ensemble of decision-tree weak learners to map the eight mix-design variables and curing age to compressive strength. Using the Orange implementation, the base estimator was set to **Tree**, with **100 estimators** and a **learning rate of 1.0**, while the regression loss was configured as **Linear**. In this framework, training begins by assigning equal weights to all Nobservations ($w_i = 1/N$). At each boosting iteration m, a weak tree $h_m$is fitted on the **weighted** dataset so that samples currently predicted poorly exert greater influence on the next learner. The algorithm then computes the weighted error $\varepsilon_m$and assigns the learner an importance coefficient α_m, after which sample weights are updated to emphasize "hard" cases and de-emphasize well-predicted cases, followed by normalization. The final regressor is therefore not a single tree but an additive combination of many sequentially trained trees, yielding a model capable of capturing strong nonlinearities and interaction effects typical of concrete strength behavior (e.g., binder–water–admixture coupling and age-dependent gain).

Critically, selecting 100 estimators with a learning rate of 1.0 increases representational power but also raises the risk of overfitting noisy or extreme observations (notably the heavy-tailed age distribution), making careful validation on the 20% holdout essential.

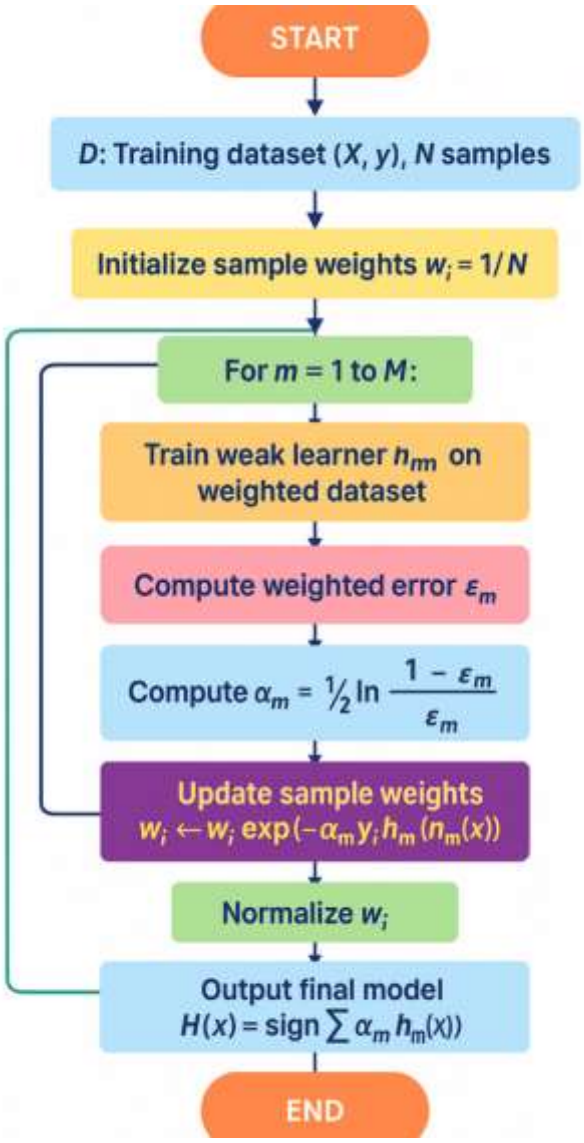


Figure 5 Flowchart of the AdaBoost regression model development used in this study: the training set $D(X, y)$ is initialized with equal sample weights, a sequence of M weak decision-tree learners is trained iteratively on the weighted data, the weighted error ε_m and learner weight α_m are computed at each iteration, sample weights are updated and normalized to emphasize hard-to-predict observations, and the final prediction is produced by the weighted ensemble of all weak learners.

Model assessment

Model performance was evaluated on the 20% holdout set using a complementary set of goodness-of-fit and error-based indicators: R^2, MSE, RMSE, MAE, MAPE, and CVRMSE. The coefficient of determination (R^2) was used to quantify the proportion of variance in compressive strength explained by the AdaBoost regressor, providing an intuitive measure of explanatory power. However, because R^2 alone may hide systematic bias, absolute and squared error metrics were also reported. MSE and RMSE penalize large residuals more strongly and are therefore sensitive to occasional high-error cases, which is critical for concrete datasets where extreme ages or unusual mix designs can generate outliers. RMSE was emphasized because it retains the same unit as the target (MPa), enabling direct engineering interpretation. MAE was included as a robust indicator of typical error magnitude, being less influenced by extreme residuals than RMSE.

Percentage-based measures (MAPE) were computed to express error in relative terms; this supports comparability across different strength ranges, though it can be unstable when measured strengths approach zero. Finally, CVRMSE was reported to normalize RMSE by the mean strength, giving a scale-free measure of dispersion in prediction error and enabling more objective comparison across models and datasets.

Model Assessment Metrics (Equations)

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

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

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

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

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

$$CVRMSE = \frac{RMSE}{\bar{y}} \times 100$$

where y_i is observed, $\hat{y}_i$ is predicted, $\bar{y}$ is mean of y, n is samples.

Figure 6 Mathematical formulations of the performance metrics used to assess the AdaBoost regression model on the test set, including the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of variation of RMSE (CVRMSE).

Results

Adaptive boost analysis

The AdaBoost regressor demonstrates strong predictive capability for concrete compressive strength, as evidenced by the high coefficient of determination on the test set ($R^2 = 0.882$). The predicted-versus-observed scatter aligns closely with the 1:1 reference line, indicating that the ensemble captures the dominant nonlinear relationships between mix constituents, curing age, and strength. Nevertheless, the dispersion widens at higher strengths, suggesting that extreme-strength mixtures are harder to generalize likely due to fewer high-strength samples and stronger interaction effects at low water/binder ratios and higher admixture dosages.

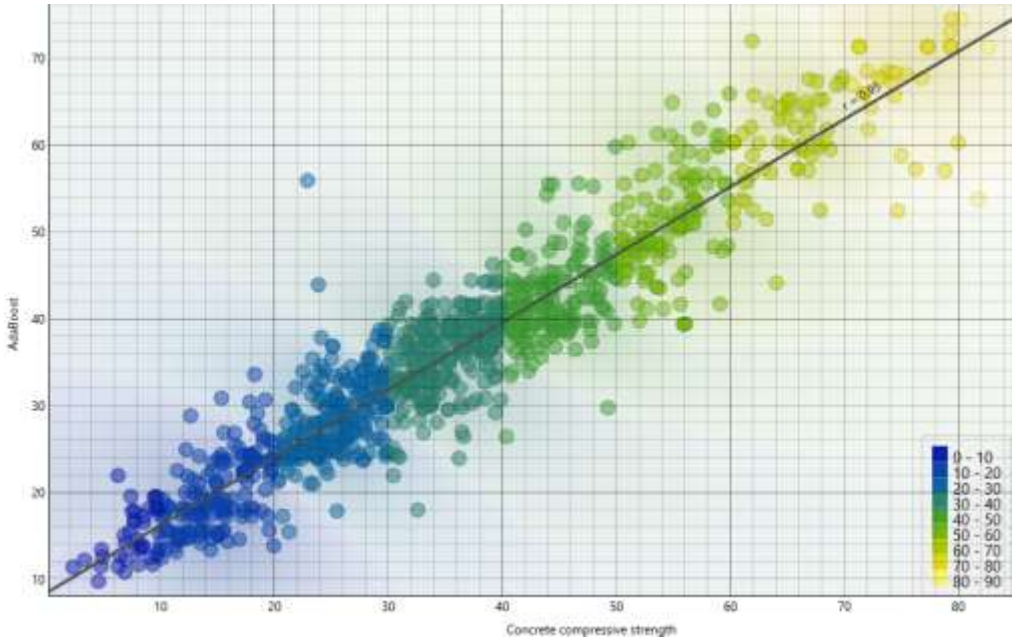


Figure 7 Predicted versus observed concrete compressive strength for the AdaBoost regression model on the test set, with the 1:1 reference line indicating perfect agreement; point colours represent strength ranges, illustrating overall fit and the increased dispersion at higher-strength levels.

Error-based metrics further quantify performance. The test **MSE = 30.468** corresponds to a **RMSE = 5.520 MPa**, meaning the model's typical squared-error magnitude is moderate and the average deviation is on the order of ~5–6 MPa. The **MAE = 4.165 MPa** being lower than RMSE implies that most predictions are reasonably close, while a smaller number of larger residuals inflate RMSE consistent with the few visible outliers in the plot. The **MAPE = 17.035%** indicates moderate relative error; this is expected to be higher at low-strength values where percentage errors amplify. Finally, **CVRMSE = 15.402%** (RMSE normalized by mean strength) confirms a solid, scale-independent fit. However, the reported training error **(Train = 216.694)** should be interpreted carefully and verified for consistency with the chosen metric definition.

Error analysis

The sequence plot of the **experimental compressive strength** across row numbers shows a highly irregular pattern with frequent sharp fluctuations, confirming that the dataset represents a wide variety of mix designs and curing ages rather than a smooth, ordered series. This strong variability is important because it implies that neighbouring rows are not physically related, so any model must generalize across heterogeneous concrete regimes (low-strength, medium-strength, and high-strength mixes). The absence of long monotonic trends suggests limited risk of order-driven leakage, but it also indicates that extreme values occur throughout the dataset and can influence error statistics.

The corresponding plot of **AdaBoost predictions** follows the same overall amplitude and range, indicating that the ensemble captures the dominant structure of the target distribution. However, the predicted signal appears slightly smoother than the experimental series, which is expected for an ensemble average; this smoothing can reduce noise sensitivity but may also attenuate some peaks and troughs, particularly in very high-strength cases.

The **percentage error** plot is largely centred around zero, suggesting no strong global bias (systematic over- or underestimation). Most errors remain within

a moderate band, yet several spikes reach large positive and negative values, indicating localized failures on specific mixes. These outliers are consistent with complex interaction effects (e.g., unusual SCM–water–admixture combinations or extreme ages) and justify reporting both MAE (typical error) and RMSE (outlier-sensitive error).

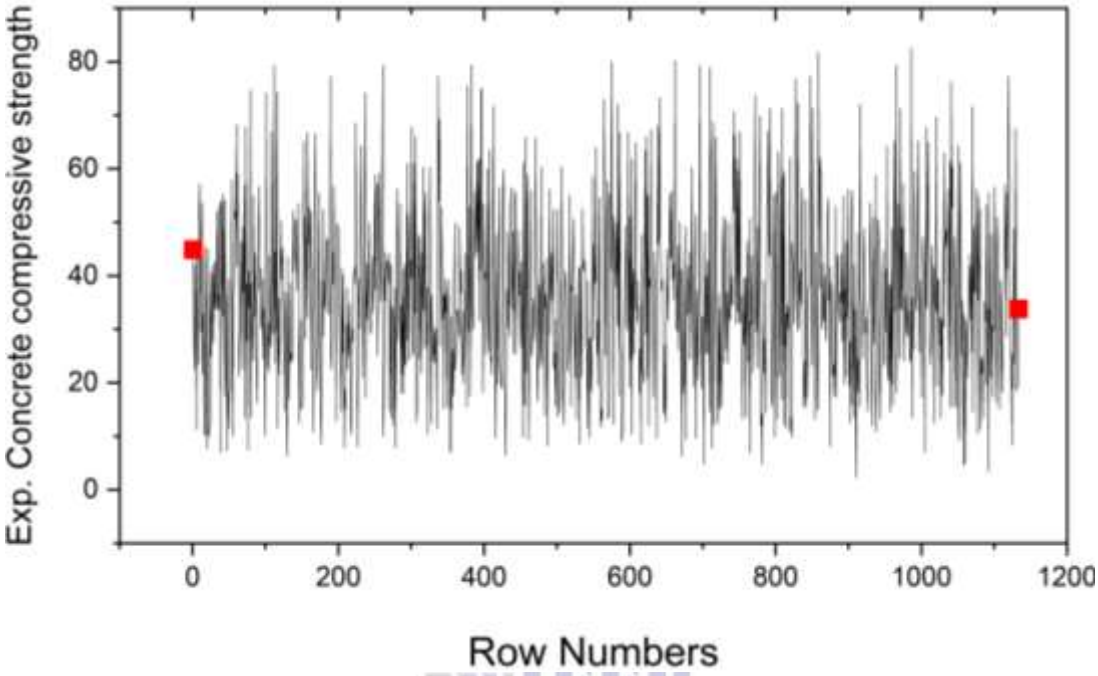


Figure 8 Experimental concrete compressive strength values plotted against dataset row number, illustrating the wide variability and irregular distribution of strength across samples.

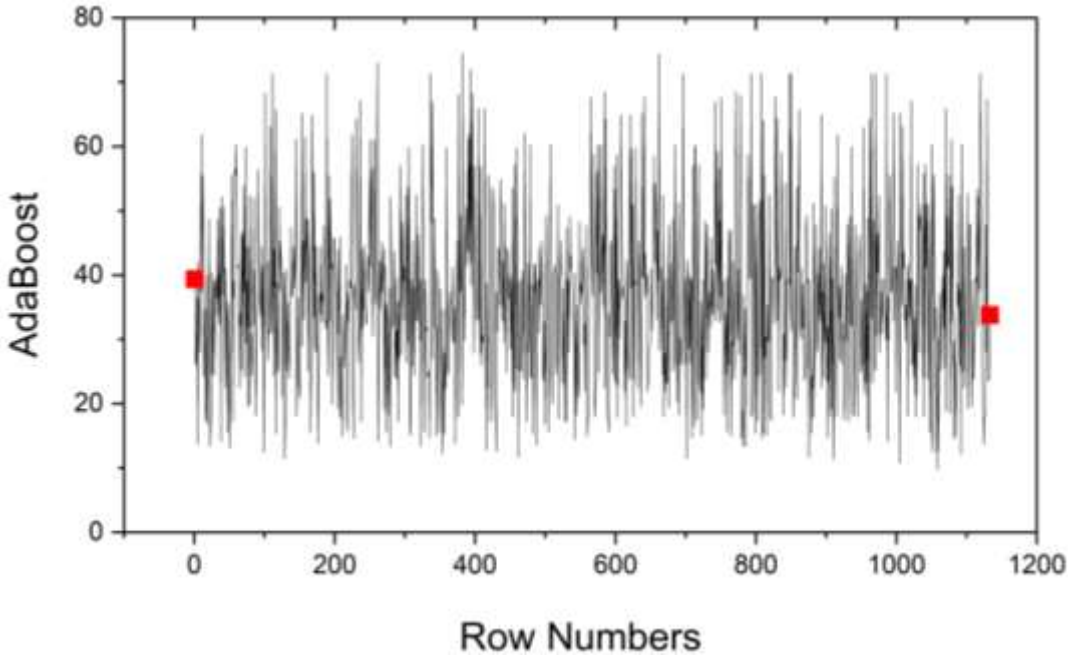


Figure 9 AdaBoost-predicted compressive strength plotted against dataset row number, showing the model's ability to track the overall range and fluctuations of the experimental values.

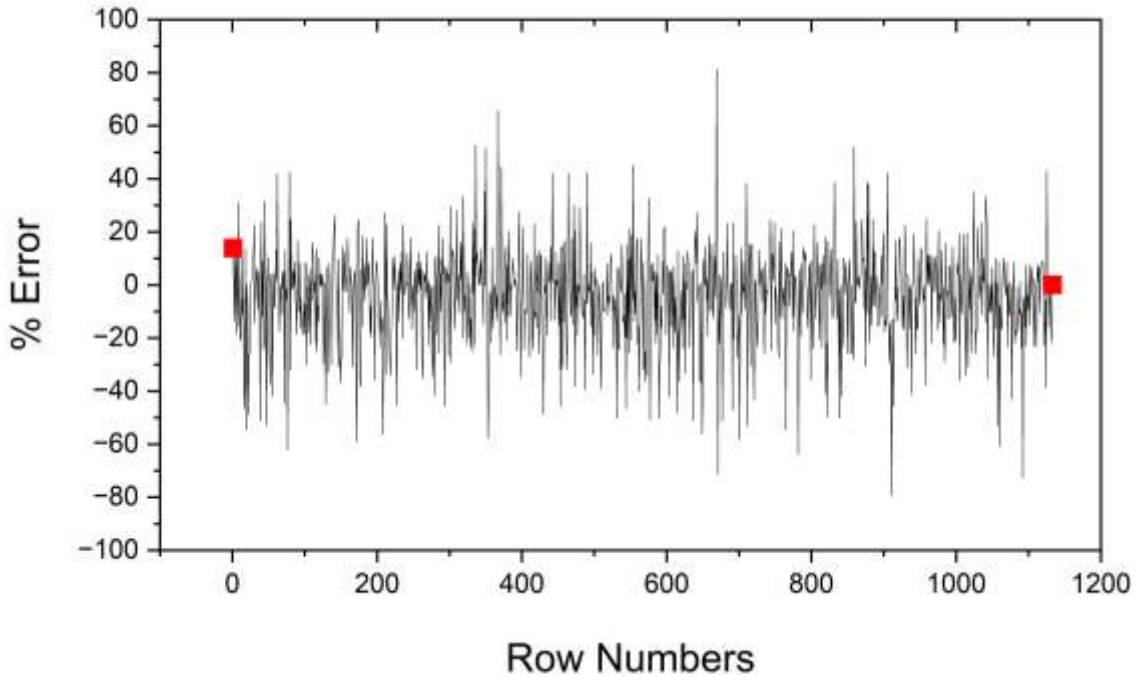


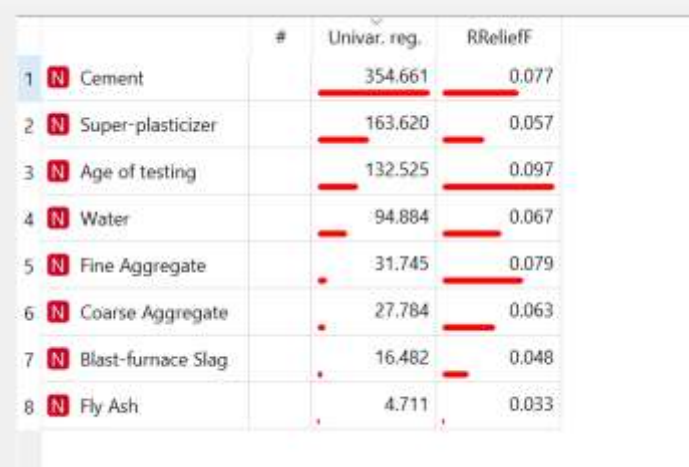
Figure 10 Percentage prediction error (%)versus dataset row number for the AdaBoost model, highlighting near-zero-centered residuals with occasional high-magnitude error spikes.

Ranking of feature in prediction

The feature-ranking results highlight which inputs contribute most to predicting compressive strength, using two complementary criteria: **Univariate regression** (marginal, one-variable-at-a-time association with strength) and **RReliefF** (relevance based on how a feature helps distinguish similar samples, capturing interaction effects more effectively).

Under **Univariate regression**, **Cement** is the dominant predictor **(354.661)**, followed by **Super-plasticizer (163.620)** and **Age of testing (132.525)**, confirming that binder content, admixture dosage, and curing time have strong standalone relationships with strength. **Water (94.884)** ranks next, while aggregates show weaker marginal influence **(Fine aggregate 31.745; Coarse aggregate 27.784)**. The SCMs are lowest in pure marginal contribution **(Blast-furnace slag 16.482; Fly ash 4.711)**, which is consistent with their "often-zero, sometimes-high" distributions and their dependence on mixture context.

However, **RReliefF** shifts the emphasis: **Age of testing becomes the most influential (0.097)**, reflecting nonlinear strength-gain kinetics and strong interactions with binder composition. **Fine aggregate (0.079)** and **Cement (0.077)** remain important, while **Water (0.067)** and **Coarse aggregate (0.063)** contribute moderately. **Fly ash (0.033)** and **slag (0.048)** remain least influential overall. Critically, the mismatch between rankings indicates that strength prediction is **interaction-driven**, so high univariate scores should not be interpreted as causal importance.



	#	Univar. reg.	RReliefF
1 N Cement		354.661	0.077
2 N Super-plasticizer		163.620	0.057
3 N Age of testing		132.525	0.097
4 N Water		94.884	0.067
5 N Fine Aggregate		31.745	0.079
6 N Coarse Aggregate		27.784	0.063
7 N Blast-furnace Slag		16.482	0.048
8 N Fly Ash		4.711	0.033

Figure 11 Ranks of features

Discussion

Predictive Performance and Quantitative Metrics

AdaBoost regression models have demonstrated robust predictive capabilities for concrete compressive strength across a variety of concrete types, including high-performance, geopolymer, and blended concretes. Reported coefficients of determination (R^2) for AdaBoost typically range from 0.89 to 0.93, indicating that the model can explain a substantial proportion of the variance in compressive strength outcomes(Ahmad, Ahmad, Chaiyasarn, et al., 2021; Ahmad, Ostrowski, Maślak, et al., 2021; Shen et al., 2022; Kashem et al., 2024; Nguyen, Tran and Nguyen, 2025; Pham, 2025). For example, in the prediction of fly ash-based geopolymer concrete, AdaBoost achieved an R^2 of 0.93, with mean absolute error (MAE) and root mean square error (RMSE) values of 2.13 MPa and 3.06 MPa, respectively. These error magnitudes are generally considered acceptable for practical engineering applications, where experimental variability and material heterogeneity are inherent challenges(Amin et al., 2023; Gao et al., 2023, 2024; Huang et al., 2023; Abdullah et al., 2024; Nguyen and Phan, 2024; Zhang et al., 2025).

However, AdaBoost's performance is not uniform across all datasets and concrete types. In the context of ultra-high-strength or highly heterogeneous mixtures, the model's predictive accuracy can diminish, particularly at the extremes of the strength spectrum. This is evidenced by increased dispersion in predicted-versus-observed plots at higher strength levels, a phenomenon attributed to the scarcity of high-strength samples and the complex, nonlinear interactions among mix constituents at low water/binder ratios and high admixture dosages(Cui et al., 2021; Latif, 2021; Lee et al., 2022; Chao et al., 2024; Fang et al., 2024; Li, Zhang and Gu, 2024; M. S. Khan et al., 2025; Tak, Feng and Mahgoub, 2025).

Comparative Analysis with Other Ensemble Methods

While AdaBoost is a strong performer, it is often outperformed by other ensemble methods such as Gradient Boosting Regression (GBR), XGBoost, and Random Forest. For instance, GBR models have achieved R^2 values as high as 0.99 and RMSE values as low as 0.93 MPa, substantially outperforming AdaBoost in some studies. In a direct comparison, Random Forest regression achieved an MAE of 0.69 MPa and an R^2 of 0.99, while AdaBoost lagged with higher error metrics and lower R^2. Similarly, XGBoost has been shown to deliver superior accuracy, with R^2 values up to 0.95 and RMSE values as low as 3.06 MPa(Xu et al., 2021; Hu et al., 2022; Jibril, Malami, et al., 2023; Taffese and Espinosa-Leal, 2023; Kalabarige et al., 2024; Katlav and Ergen, 2024; Ellappan et al., 2025).

The reasons for this performance gap are multifaceted. Gradient-based boosting algorithms and Random Forests are generally more adept at capturing complex, high-order interactions and are less sensitive to outliers and noise in the data. AdaBoost, by contrast, can be more susceptible to overfitting, especially when the base learners are not sufficiently regularized or when the dataset contains a significant number of noisy or extreme-value samples(Dao et al., 2020; Ahmad, Hu, Ahmad, et al.,

2021; Assegie, Salau and Badrudeen, 2022; Jia et al., 2022; Kumar et al., 2022; Rathakrishnan, Beddu and Ahmed, 2022; Wang, Zhao and Gao, 2023; Arora, Kumar and Singh, 2024; Chen et al., 2024; Li et al., 2024; Logeswaran et al., 2024; Loureiro and Stefani, 2024; Bentegri et al., 2025).

Model Interpretability and Feature Importance

A notable advantage of AdaBoost and other tree-based ensemble methods is their capacity for interpretability. Feature importance analyses consistently identify curing age, cement content, and water-to-cement ratio as the most influential predictors of compressive strength. For example, sensitivity analyses have shown that the age of concrete, cement content, and water content have the highest information gain values, underscoring their critical roles in strength development(Shishegaran et al., 2020; Nguyen et al., 2021; Fei et al., 2023; Jibril, Zayyan, et al., 2023; Paudel et al., 2023; Phan, 2023; Tipu, Suman and Batra, 2023; Ahmed, Jin and Ali, 2024a, 2024b; Aswal, Singh and Maheshwari, 2024; Beskopylny et al., 2024; Harith et al., 2024; Pan et al., 2025). The use of interpretability tools such as SHapley Additive exPlanations (SHAP) further enhances the transparency of AdaBoost models, allowing practitioners to understand the contribution of each input variable to the final prediction(Ahmad, Ahmad, Aslam, et al., 2021; Song et al., 2021; Chen et al., 2022; Amruthamol et al., 2023; Aslam and Shahab, 2023; Li et al., 2023; Liu et al., 2023; Phoeuk and Kwon, 2023; Alyami et al., 2024; Elshaarawy, Alsaadawi and Hamed, 2024; Islam et al., 2024; Khandagale et al., 2024; Punitha et al., 2025).

Practical Implications and Limitations

The practical utility of AdaBoost regression in concrete mix design and quality control is significant. By providing rapid, data-driven predictions of compressive strength, AdaBoost models can reduce the reliance on time-consuming and costly experimental testing. This is particularly valuable in the context of sustainable construction, where the incorporation of supplementary cementitious materials (e.g., fly ash, slag, rice husk ash) introduces additional variability and complexity(Ekanayake, Meddage and Rathnayake, 2022; Zhang et al., 2023; Habib and Okayli, 2024; Mustapha et al., 2024; Zhao, 2024; Shaaban et al., 2025). Nevertheless, several limitations warrant consideration. AdaBoost's predictive accuracy is contingent on the quality and representativeness of the training data. In cases where the dataset is imbalanced or lacks sufficient coverage of high-strength or specialty mixtures, the model's generalizability may be compromised . Moreover, while AdaBoost is computationally efficient and relatively easy to implement, it may not always be the optimal choice when the highest possible accuracy is required, especially given the superior performance of gradient boosting and random forest algorithms in many studies.

Future Directions

Recent research has explored hybrid and optimized AdaBoost models, such as those incorporating neural networks or advanced optimization algorithms, to enhance predictive performance. For example, integrating AdaBoost with a back-propagation neural network optimized by a sparrow search algorithm has yielded improved accuracy over standard AdaBoost implementations. Additionally, the development of super learner models that combine multiple ensemble algorithms has shown promise in further reducing prediction errors and increasing robustness. AdaBoost regression is a valuable tool for predicting concrete compressive strength, offering a balance of accuracy, interpretability, and computational efficiency. However, its performance must be critically evaluated against alternative ensemble methods, and its limitations addressed through careful data curation and model selection(Liu et al., 2023; Sami et al., 2023; Kumar et al., 2024).

Conclusion and recommendation

The study demonstrates that AdaBoost regression can predict concrete compressive strength with strong accuracy using only mix-proportion variables and curing age, achieving a high test-set fit ($R^2 = 0.882$) and moderate error magnitudes (MSE = 30.468, RMSE = 5.520 MPa, MAE = 4.165 MPa). The combination of squared-error and absolute-error measures confirms that most predictions are close to the experimental values, while a smaller number of

difficult samples produce larger deviations that inflate RMSE relative to MAE. The predicted–observed scatter and the row-wise comparison further indicate that the model captures the overall trend and range of strength development, although dispersion increases at higher-strength levels, where fewer samples and stronger nonlinear interactions are expected. The percentage-error signal is centered near zero, suggesting no strong global bias, but the presence of spikes confirms localized failures for specific mix regimes. Feature ranking supports the physical plausibility of the learned relationships: cement, age, water, and super-plasticizer consistently appear among the most influential inputs, while slag and fly ash show lower marginal contribution but remain important through interactions captured by RReliefF. Overall, the outcomes validate AdaBoost as an effective baseline ensemble model for rapid strength estimation in mix design and quality-control contexts, particularly when experimental testing is limited or time-constrained.

Future work should strengthen generalization by adopting repeated k-fold cross-validation in addition to the single 80/20 split, and by reporting confidence intervals for key metrics to quantify uncertainty. Hyperparameter optimization is recommended, particularly tuning the number of estimators, learning rate, and tree complexity, because the current configuration (100 estimators, learning rate = 1.0) may overfit outliers in heavy-tailed variables such as age. To reduce the impact of extreme errors observed in the residual plots, robust strategies such as outlier screening, winsorization, or Huber-type losses (if supported) should be evaluated. Since MAPE can become unstable at low-strength values, complementary relative metrics (e.g., sMAPE or MAE normalized by mean/median strength) may be included for fairer interpretation. Model interpretability should be extended beyond ranking by using SHAP or partial dependence analysis to reveal how key variables interact (e.g., water–super-plasticizer trade-offs, SCM replacement effects, and age-dependent gain). Expanding the dataset with more high-strength and late-age samples would likely reduce dispersion at the upper end and improve robustness. Finally, benchmarking against stronger ensembles such as Gradient Boosting, Random Forest, and XGBoost on the same split and metrics is recommended to determine whether AdaBoost offers the best balance of accuracy, stability, and interpretability for the specific mix-design domain addressed in this work.

REFERENCES

Abdullah, G. et al. (2024) "Boosting-based ensemble machine learning models for predicting unconfined compressive strength of geopolymer stabilized clayey soil," Scientific Reports, 14. Available at: https://doi.org/10.1038/s41598-024-52825-7.

Ahmad, A., Ostrowski, K., Maślak, M., et al. (2021) "Comparative Study of Supervised Machine Learning Algorithms for Predicting the Compressive Strength of Concrete at High Temperature," Materials, 14. Available at: https://doi.org/10.3390/ma14154222.

Ahmad, A., Ahmad, W., Aslam, F., et al. (2021) "Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques," Case Studies in Construction Materials [Preprint]. Available at: https://doi.org/10.1016/j.cscm.2021.e00840.

Ahmad, A., Ahmad, W., Chaiyasarn, K., et al. (2021) "Prediction of Geopolymer Concrete Compressive Strength Using Novel Machine Learning Algorithms," Polymers, 13. Available at: https://doi.org/10.3390/polym13193389.

Ahmad, M., Hu, J., Ahmad, F., et al. (2021) "Supervised Learning Methods for Modeling Concrete Compressive Strength Prediction at High Temperature," Materials, 14. Available at: https://doi.org/10.3390/ma14081983.

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

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

Alyami, M. et al. (2024) "Application of metaheuristic optimization algorithms in predicting the compressive strength of 3D-printed fiber-reinforced concrete," Developments in the Built Environment [Preprint]. Available at: https://doi.org/10.1016/j.dibe.2023.100307.

Amin, M.N. et al. (2023) "Prediction model for rice husk ash concrete using AI approach: Boosting and bagging algorithms," Structures [Preprint]. Available at: https://doi.org/10.1016/j.istruc.2023.02.080.

Amruthamol, N. et al. (2023) "Bayesian learning approach to foresee compressive strength of concrete," Materials Today: Proceedings [Preprint]. Available at: https://doi.org/10.1016/j.matpr.2023.07.263.

Arora, G., Kumar, D. and Singh, B. (2024) "Tree based Regression Models for Predicting the Compressive Strength of Concrete at High Temperature," IOP Conference Series: Earth and Environmental Science, 1327. Available at: https://doi.org/10.1088/1755-1315/1327/1/012015.

Aslam, F. and Shahab, M. (2023) "Supplementary Cementitious Materials in Blended Cement Concrete: Advancements in Predicting Compressive Strength through Machine Learning," Materials Today Communications [Preprint]. Available at: https://doi.org/10.1016/j.mtcomm.2023.107725.

Assegie, T.A., Salau, A.O. and Badrudeen, T.U. (2022) "Estimation of concrete compression using regression models," Bulletin of Electrical Engineering and Informatics [Preprint]. Available at: https://doi.org/10.11591/eei.v11i5.4210.

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

Bentegri, H. et al. (2025) "Assessment of compressive strength of eco-concrete reinforced using machine learning tools," Scientific Reports, 15. Available at: https://doi.org/10.1038/s41598-025-89530-y.

Beskopylny, A. et al. (2024) "Prediction of the Compressive Strength of Vibrocentrifuged Concrete Using Machine Learning Methods," Buildings [Preprint]. Available at: https://doi.org/10.3390/buildings14020377.

Chao, Z. et al. (2024) "Estimating compressive strength of coral sand aggregate concrete in marine environment by combining physical experiments and machine learning-based techniques," Ocean Engineering [Preprint]. Available at: https://doi.org/10.1016/j.oceaneng.2024.118320.

Chen, H. et al. (2022) "Compressive Strength Prediction of High-Strength Concrete Using Long Short-Term Memory and Machine Learning Algorithms," Buildings [Preprint]. Available at: https://doi.org/10.3390/buildings12030302.

Chen, L. et al. (2024) "Optimizing compressive strength prediction in eco-friendly recycled concrete via artificial intelligence models," Multiscale and Multidisciplinary Modeling, Experiments and Design [Preprint]. Available at: https://doi.org/10.1007/s41939-024-00641-x.

Cui, X. et al. (2021) "Machine learning prediction of concrete compressive strength with data enhancement," J. Intell. Fuzzy Syst., 41, pp. 7219–7228. Available at: https://doi.org/10.3233/jifs-211088.

Dao, D. et al. (2020) "A Sensitivity and Robustness Analysis of GPR and ANN for High-Performance Concrete Compressive Strength Prediction Using a Monte Carlo Simulation," Sustainability [Preprint]. Available at: https://doi.org/10.3390/su12030830.

Ekanayake, I., Meddage, D. and Rathnayake, U. (2022) "A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP)," Case Studies in Construction Materials [Preprint]. Available at: https://doi.org/10.1016/j.cscm.2022.e01059.

Ellappan, P. et al. (2025) "Improving forecasting of concrete strength using advanced machine learning methods," Matéria (Rio de Janeiro) [Preprint]. Available at: https://doi.org/10.1590/1517-7076-rmat-2024-0789.

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

Fang, G.-H. et al. (2024) "Optimized Machine Learning Model for Predicting Compressive Strength of Alkali-Activated Concrete Through Multi-Faceted Comparative Analysis," Materials, 17. Available at: https://doi.org/10.3390/ma17205086.

Fei, Z. et al. (2023) "Ensemble Machine-Learning-Based Prediction Models for the Compressive Strength of Recycled Powder Mortar," Materials, 16. Available at: https://doi.org/10.3390/ma16020583.

Gali, G. et al. (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), pp. 12–20. Available at: https://doi.org/10.33150/JITDETS-7.1.3.

Gao, Y. et al. (2023) "Development of chemistry-informed interpretable model for predicting compressive strength of recycled aggregate concrete containing supplementary cementitious materials," Journal of Cleaner Production [Preprint]. Available at: https://doi.org/10.1016/j.jclepro.2023.138733.

Gao, Y. et al. (2024) "Using Stacking Machine Learning Models to Predict High-Performance Concrete Compressive Strength," Proceedings of the 2024 4th International Conference on Artificial Intelligence, Big Data and Algorithms [Preprint]. Available at: https://doi.org/10.1145/3690407.3690420.

Habib, M. and Okayli, M. (2024) "Evaluating the Sensitivity of Machine Learning Models to Data Preprocessing Technique in Concrete Compressive Strength Estimation," Arabian Journal for Science and Engineering [Preprint]. Available at: https://doi.org/10.1007/s13369-024-08776-2.

Harith, I.K. et al. (2024) "Prediction of high-performance concrete strength using machine learning with hierarchical regression," Multiscale and Multidisciplinary Modeling, Experiments and Design, 7, pp. 4911–4922. Available at: https://doi.org/10.1007/s41939-024-00467-7.

Hu, X. et al. (2022) "Predicting triaxial compressive strength of high-temperature treated rock using machine learning techniques," Journal of Rock Mechanics and Geotechnical Engineering [Preprint]. Available at: https://doi.org/10.1016/j.jrmge.2022.10.014.

Huang, Y. et al. (2023) "Prediction of compressive strength of rice husk ash concrete: A comparison of different metaheuristic algorithms for optimizing support vector regression," Case Studies in Construction Materials [Preprint]. Available at: https://doi.org/10.1016/j.cscm.2023.e02201.

Islam, M.M. et al. (2024) "Prediction of compressive strength of high-performance concrete using optimization machine learning approaches with SHAP analysis," Journal of Building Pathology and Rehabilitation [Preprint]. Available at: https://doi.org/10.1007/s41024-024-00445-z.

Jia, J.-F. et al. (2022) "An interpretable ensemble learning method to predict the compressive strength of concrete," Structures [Preprint]. Available at: https://doi.org/10.1016/j.istruc.2022.10.056.

Jibril, M., Malami, S., et al. (2023) "High strength concrete compressive strength prediction using an evolutionary computational intelligence algorithm," Asian Journal of Civil Engineering, 24, pp. 3727–3741. Available at: https://doi.org/10.1007/s42107-023-00746-7.

Jibril, M., Zayyan, M., et al. (2023) "Implementation of Nonlinear Computing Models and Classical Regression for Predicting Compressive Strength of High-Performance Concrete," Applications in Engineering Science [Preprint]. Available at: https://doi.org/10.1016/j.apples.2023.100133.

Kalabarige, L. et al. (2024) "Machine Learning Modeling Integrating Experimental Analysis for Predicting Compressive Strength of Concrete Containing Different Industrial Byproducts," Advances in Civil Engineering [Preprint]. Available at: https://doi.org/10.1155/2024/7844854.

Kashem, A. et al. (2024) "Hybrid data-driven approaches to predicting the compressive strength of ultra-high-performance concrete using SHAP and PDP analyses," Case Studies in Construction Materials [Preprint]. Available at: https://doi.org/10.1016/j.cscm.2024.e02991.

Katlav, M. and 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, pp. 212–235. Available at: https://doi.org/10.1002/suco.202400163.

Khan, M.S. et al. (2025) "Explainable AutoML models for predicting the strength of high-performance concrete using Optuna, SHAP and ensemble learning," Frontiers in Materials [Preprint]. Available at: https://doi.org/10.3389/fmats.2025.1542655.

Khan, W. et al. (2025) "Transforming Expansive Clays FROM Waste TO Value: Dual Ash Stabilization Strategy FOR Augmenting Black Cotton Subgrade Soil Performance," 11, pp. 5065–5073. Available at: https://doi.org/10.64252/cz07k049.

Khandagale, H. et al. (2024) "Optimizing Concrete Compressive Strength Prediction: A Comparative Study of Machine Learning Algorithms," 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT), 1, pp. 1–6. Available at: https://doi.org/10.1109/iceect61758.2024.10738970.

Kumar, A. et al. (2022) "Compressive Strength Prediction of Lightweight Concrete: Machine Learning Models," Sustainability [Preprint]. Available at: https://doi.org/10.3390/su14042404.

Kumar, R. et al. (2024) "Development of a prediction tool for the compressive strength of ternary blended ultra-high performance concrete using machine learning techniques," Journal of Structural Integrity and Maintenance, 9. Available at: https://doi.org/10.1080/24705314.2024.2385206.

Latif, S. (2021) "Developing a boosted decision tree regression prediction model as a sustainable tool for compressive strength of environmentally friendly concrete," Environmental Science and Pollution Research, 28, pp. 65935–65944. Available at: https://doi.org/10.1007/s11356-021-15662-z.

Lee, S. et al. (2022) "Super learner machine-learning algorithms for compressive strength prediction of high performance concrete," Structural Concrete, 24, pp. 2208–2228. Available at: https://doi.org/10.1002/suco.202200424.

Li, D. et al. (2023) "Machine Learning-Based Method for Predicting Compressive Strength of Concrete," Processes [Preprint]. Available at: https://doi.org/10.3390/pr11020390.

Li, H. et al. (2024) "Compressive Strength Prediction of Fly Ash-Based Concrete Using Single and Hybrid Machine Learning Models," Buildings [Preprint]. Available at: https://doi.org/10.3390/buildings14103299.

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

Liu, L. et al. (2023) "Prediction of compressive strength of high-performance concrete via coupled meta-heuristic random forest regression techniques," Multiscale and Multidisciplinary Modeling, Experiments and Design, pp. 1–15. Available at: https://doi.org/10.1007/s41939-023-00256-8.

Logeswaran, K. et al. (2024) "Machine Learning Model for Concrete Strength Prediction," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–6. Available at: https://doi.org/10.1109/icccnt61001.2024.10724333.

Loureiro, A.A.B. and Stefani, R. (2024) "Comparing the performance of machine learning models for predicting the compressive strength of concrete," Discover Civil Engineering [Preprint]. Available at: https://doi.org/10.1007/s44290-024-00022-w.

Mujtaba, G. et al. (2023) "Water Pollution Hazards and Toxicity Caused by Textile Industries Effluent," Journal of ICT, Design, Engineering and Technological Science, 7(2), pp. 22–27. Available at: https://doi.org/10.33150/JITDETS-7.2.3.

Mustapha, I. et al. (2024) "Comparative Analysis of Gradient-Boosting Ensembles for Estimation of Compressive Strength of Quaternary Blend Concrete," International Journal of Concrete Structures and Materials, 18, pp. 1–24. Available at: https://doi.org/10.1186/s40069-023-00653-w.

Nawaz, S. et al. (2025) "Flexural Performance Evaluation Of Glass Fiber Reinforced Polymer Bars (GFRP) In Doubly Reinforced Beams: An Experimental Approach."

Nguyen, D.-L. and Phan, T.-D. (2024) "Predicting the compressive strength of ultra-high-performance concrete: an ensemble machine learning approach and actual application," Asian Journal of Civil Engineering [Preprint]. Available at: https://doi.org/10.1007/s42107-023-00984-9.

Nguyen, H. et al. (2021) "Efficient machine learning models for prediction of concrete strengths," Construction and Building Materials [Preprint]. Available at: https://doi.org/10.1016/j.conbuildmat.2020.120950.

Nguyen, T.-G., Tran, T.D. and Nguyen, P.D. (2025) "Prediction of Concrete Compressive Strength Using Artificial Neural Network with Bayesian Optimization," Materials and Emerging Technologies for Sustainability [Preprint]. Available at: https://doi.org/10.1142/s3060932125500074.

Pan, B. et al. (2025) "Predicting the Compressive Strength of Recycled Concrete Using Ensemble Learning Model," IEEE Access, 13, pp. 2958–2969. Available at: https://doi.org/10.1109/access.2024.3519669.

Paudel, S. et al. (2023) "Compressive strength of concrete material using machine learning techniques," Cleaner Engineering and Technology [Preprint]. Available at: https://doi.org/10.1016/j.clet.2023.100661.

Pham, B.T. (2025) "BGG-REPT and ROF-REPT: ensemble machine learning models for the prediction of compressive strength of concrete," Innovative Infrastructure Solutions [Preprint]. Available at: https://doi.org/10.1007/s41062-025-01869-3.

Phan, T.-D. (2023) "Fast prediction of the compressive strength of high-performance concrete through a k-nearest neighbor approach," Asian Journal of Civil Engineering, pp. 1–16. Available at: https://doi.org/10.1007/s42107-023-00756-5.

Phoeuk, M. and Kwon, M. (2023) "Accuracy Prediction of Compressive Strength of Concrete Incorporating Recycled Aggregate Using Ensemble Learning Algorithms: Multinational Dataset," Advances in Civil Engineering [Preprint]. Available at: https://doi.org/10.1155/2023/5076429.

Punitha, A. et al. (2025) "Prediction of compressive strength of blended concrete with Alccofine and GGBFS by applying ensemble machine learning algorithms," Journal of Structural Integrity and Maintenance, 10. Available at: https://doi.org/10.1080/24705314.2025.2489858.

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

Sami, B.H.Z. et al. (2023) "Feasibility Analysis for Predicting the Compressive and Tensile Strength of Concrete using Machine Learning Algorithms," Case Studies in Construction Materials [Preprint]. Available at: https://doi.org/10.1016/j.cscm.2023.e01893.

Shaaban, M. et al. (2025) "Machine learning approaches for forecasting compressive strength of high-strength concrete," Scientific Reports, 15. Available at: https://doi.org/10.1038/s41598-025-10342-1.

Shah, Q. et al. (2023) "To Stabilize Shear Strength Properties of an Unwanted Subgrade Soil Utilizing Rock Dust," Journal of ICT, Design, Engineering and Technological Science, 7(2), pp. 1–6. Available at: https://doi.org/10.33150/JITDETS-7.2.1.

Shen, Z. et al. (2022) "Compressive Strength Evaluation of Ultra-High-Strength Concrete by Machine Learning," Materials, 15. Available at: https://doi.org/10.3390/ma15103523.

Shishegaran, A. et al. (2020) "High correlated variables creator machine: Prediction of the compressive strength of concrete," ArXiv, abs/2009.0. Available at: https://doi.org/10.1016/j.compstruc.2021.106479.

Song, Y. et al. (2021) "Prediction of Compressive Strength of Fly-Ash-Based Concrete Using Ensemble and Non-Ensemble Supervised Machine-Learning Approaches," Applied Sciences [Preprint]. Available at: https://doi.org/10.3390/app12010361.

Taffese, W. and Espinosa-Leal, L. (2023) "Multitarget regression models for predicting compressive strength and chloride resistance of concrete," Journal of Building Engineering [Preprint]. Available at: https://doi.org/10.1016/j.jobe.2023.106523.

Tak, M.S.N., Feng, Y. and Mahgoub, M. (2025) "Advanced Machine Learning Techniques for Predicting Concrete Compressive Strength," Infrastructures [Preprint]. Available at: https://doi.org/10.3390/infrastructures10020026.

Tipu, R.K., Suman and Batra, V. (2023) "Enhancing prediction accuracy of workability and compressive strength of high-performance concrete through extended dataset and improved machine learning models," Asian Journal of Civil Engineering, pp. 1–22. Available at: https://doi.org/10.1007/s42107-023-00768-1.

Wang, L., Zhao, J. and Gao, S. (2023) "Strength property estimation of high-performance concrete with novel admixtures using optimal regression-based approaches," J. Intell. Fuzzy Syst., 45, pp. 6655–6666. Available at: https://doi.org/10.3233/jifs-230132.

Xu, Y. et al. (2021) "Computation of High-Performance Concrete Compressive Strength Using Standalone and Ensembled Machine Learning Techniques," Materials, 14. Available at: https://doi.org/10.3390/ma14227034.

Zhang, K. et al. (2025) "A Bio-Inspired Adaptive Probability IVYPSO Algorithm with Adaptive Strategy for Backpropagation Neural Network Optimization in Predicting High-Performance Concrete Strength," Biomimetics [Preprint]. Available at: https://doi.org/10.3390/biomimetics10080515.

Zhang, X. et al. (2023) "Prediction of compressive strength of recycled aggregate concrete using machine learning and Bayesian optimization methods," 11. Available at: https://doi.org/10.3389/feart.2023.1112105.

Zhao, Z. (2024) "Predicting compressive strength of ultra-high-performance concrete using Naive Bayes regression in novel approaches," Multiscale and Multidisciplinary Modeling, Experiments and Design [Preprint]. Available at: https://doi.org/10.1007/s41939-024-00511-6.