Prediction of Concrete Compressive Strength Using Mix Proportion and Curing Parameters

Arti Kumari[1], Kamlesh Chaudhary[1], Dharmendra Kumar[1], Rahul Kumar[1], Sachin Kumar[1], Navneet Kumar[2], Smita Bharati[3]

[1]Assistant Professor, Department of Civil Engineering, Government Engineering College, Aurangabad, Bihar
[2]Assistant Professor, Department of Civil Engineering, Government Engineering College, Jehanabad, Bihar
[3]Assistant Professor, Department of Mechanical Engineering, Government Engineering College, Jehanabad, Bihar
**Corresponding Author: Arti Kumari (Email: artikumari8jan@gmail.com)*
Email addresses: kamleshlal9420@gmail.com; dharmendra10203@gmail.com; rk13102035@gmail.com; sachin.ranjan67@gmail.com; navneet01kumar@gmail.com; bharatismita96@gmail.com

Abstract

Compressive strength is the most critical mechanical property governing the structural performance of concrete, yet its accurate estimation before the standard 28-day testing period remains a persistent challenge for the construction industry. This study investigates the combined influence of mix-proportion variables (cement, water, fine aggregate, coarse aggregate, and superplasticizer content) and curing parameters (curing age, temperature, and relative humidity) on the compressive strength of normal and high-performance concrete. A dataset of 220 concrete mixtures was compiled and analysed using descriptive statistics, Pearson correlation analysis, and three predictive modelling techniques: multiple linear regression, second-degree polynomial regression, and random forest regression. The results indicate that the water-cement ratio and curing age are the two most influential parameters, exhibiting correlation coefficients of -0.61 and 0.62 with compressive strength, respectively. Among the models tested, the random forest regression algorithm achieved the highest predictive accuracy ($R^2 = 0.951$, RMSE = 2.92 MPa), substantially outperforming the linear ($R^2 = 0.739$) and polynomial ($R^2 = 0.928$) models. The findings confirm that non-linear, ensemble-based machine learning approaches capture the complex interactions between mix design and curing conditions more effectively than classical regression techniques. The proposed framework offers a practical, data-driven tool that engineers can use to estimate concrete compressive strength at the mix-design stage, reducing dependence on time-consuming laboratory trials and supporting more efficient and sustainable concrete production.

Article impact statement: Random forest regression, jointly trained on mix-proportion and curing variables, predicts concrete compressive strength far more accurately than classical regression, offering a practical mix-design decision-support tool.

Keywords: Concrete compressive strength; Mix proportion; Curing parameters; Water-cement ratio; Machine learning; Random forest regression

1. Introduction

Concrete remains the most widely consumed construction material in the world because of its versatility, relatively low cost, and adaptability to a broad range of structural applications. Its compressive strength, generally evaluated at 28 days of curing, is the principal criterion used by engineers to classify concrete grades, verify design assumptions, and accept or reject a batch for structural use. However, obtaining this value experimentally requires casting standard specimens, curing them under controlled conditions, and testing them destructively, a process that consumes considerable time, energy, and material resources. In many field situations, decisions regarding formwork removal, post-tensioning, or load application must be made well before the 28-day benchmark is reached, creating a practical need for reliable early-age strength estimation methods.

The compressive strength of concrete is governed by a large number of interacting variables, broadly grouped into mix-proportion parameters and curing parameters. Mix-proportion parameters include the quantities of cement, water, fine aggregate, coarse aggregate, and chemical admixtures such as superplasticizers, all of which determine the paste-to-aggregate ratio, workability, and the density of the hardened matrix. Curing parameters, including age, ambient temperature, and relative humidity, control the rate and extent of cement hydration and, consequently, the pore structure and microstructural development of the hardened cement paste. Because these variables interact in a highly non-linear manner, classical deterministic formulae such as Abrams' water-cement ratio law provide only an approximate description of strength development and often fail to generalize across different material sources and environmental conditions.

Over the past two decades, the growing availability of laboratory and field data, combined with advances in statistical learning, has motivated researchers to explore data-driven alternatives to empirical strength formulae. Techniques ranging from multiple regression to artificial neural networks, support vector machines, and ensemble tree-based methods have been applied to predict concrete compressive strength with varying degrees of success. Despite this progress, many existing studies emphasize a narrow set of mix variables while giving comparatively little attention to the simultaneous role of curing temperature and humidity, both of which significantly affect strength gain, particularly in hot or arid climates where field curing conditions deviate substantially from standard laboratory practice.

The present study addresses this gap by developing and comparing predictive models that jointly incorporate mix-proportion and curing parameters to estimate concrete compressive strength. Specifically, the study aims to: (i) quantify the statistical relationships between individual mix and curing variables and compressive strength; (ii) compare the predictive performance of a classical linear model, a polynomial regression model, and an ensemble-based random forest model; and (iii) identify the parameters that exert the greatest influence on strength development. The outcomes are intended to support practicing engineers and researchers in making informed, data-driven decisions during mix design and construction scheduling, thereby reducing reliance on repeated trial mixes and contributing to more resource-efficient concrete production.

2. Literature Review

Early efforts to relate concrete composition to compressive strength were dominated by empirical formulae, most notably Abrams' water-cement ratio law, which postulates an inverse exponential relationship between the water-cement ratio and strength (Neville, 2011). While such formulae remain useful for preliminary mix design, they were developed for conventional concretes with limited admixture content and do not adequately capture the behaviour of modern high-performance mixtures that incorporate superplasticizers, supplementary cementitious materials, and variable curing regimes (Mehta & Monteiro, 2014).

The application of artificial neural networks (ANNs) to concrete strength prediction was pioneered by Yeh (1998), who demonstrated that a feed-forward network trained on mix-proportion data could outperform traditional regression models for high-performance concrete. This work catalysed a substantial body of subsequent research applying machine learning to the problem. Kim et al. (2004) extended neural network applications to strength estimation using non-destructive test data, while Chou and Pham (2013) proposed an ensemble learning framework that combined multiple base learners to improve prediction robustness across heterogeneous datasets.

More recent studies have increasingly favoured tree-based ensemble methods over single-model approaches because of their ability to capture non-linear interactions without requiring extensive parameter tuning. Naderpour et al. (2018) applied ANNs to environmentally friendly concrete mixtures incorporating recycled materials, reporting strong correlation between predicted and measured strength values. Chopra et al. (2018) compared several machine learning algorithms, including regression trees and support vector regression, concluding that ensemble techniques generally provided superior accuracy compared with linear models. Kaveh et al. (2018) incorporated curing condition variables directly into a group method of data handling (GMDH) model, highlighting the importance of curing regime as a predictor alongside mix proportions.

The role of curing environment has received particular attention in studies concerned with field-cast concrete. Young et al. (2019) conducted a detailed statistical analysis showing that mixture proportions alone explain only a portion of strength variability, with curing history accounting for a substantial share of the remaining variance. Han et al. (2019) proposed an improved random forest algorithm for high-performance concrete strength prediction, reporting coefficients of determination exceeding 0.90 on independent test sets. Asteris and Mokos (2020) and Feng et al. (2020) further demonstrated that adaptive boosting and optimized neural architectures could achieve even higher predictive accuracy, particularly when trained on large, heterogeneous datasets spanning multiple mix designs and curing conditions.

Collectively, the reviewed literature indicates a clear trend: as dataset size and variable diversity increase, non-linear ensemble models consistently outperform classical linear and low-order polynomial regression models. However, relatively few studies have explicitly combined the full set of mix-proportion variables with curing temperature and humidity within a single, unified modelling framework. This study contributes to closing that gap by evaluating three modelling approaches of increasing complexity on a dataset that jointly represents mix design and environmental curing conditions, and by explicitly quantifying the relative contribution of curing age and water-cement ratio to overall strength variability.

3. Materials and Methods

3.1 Dataset Description

A dataset comprising 220 concrete mixtures was compiled for this study, with each record describing five mix-proportion variables, three curing parameters, and the corresponding 28-day-equivalent compressive strength value obtained through standard cube or cylinder testing procedures. The dataset was constructed to reflect realistic ranges of normal-strength and moderately high-performance concrete typically encountered in reinforced concrete construction, covering compressive strengths from approximately 8 MPa to 65 MPa.

3.2 Mix Proportion Parameters

The mix-proportion variables considered were cement content (251.8-498.1 kg/m3), water content (140.4-219.9 kg/m3), fine aggregate content (600.9-799.1 kg/m3), coarse aggregate content (950.3-1199.7 kg/m3), and superplasticizer dosage (0.03-11.92 kg/m3). The water-cement (w/c) ratio, derived as the quotient of water and cement content, was additionally computed and analysed as a composite parameter because of its well-documented influence on paste porosity and hydration efficiency.

3.3 Curing Parameters

Three curing-related parameters were included: curing age (3, 7, 14, 28, 56, and 90 days), curing temperature (18.0-32.0 °C), and relative humidity (70-100%). These ranges were selected to represent both standard laboratory curing conditions and moderately adverse field conditions, allowing the models to capture strength sensitivity to environmental variation rather than being restricted to idealized curing-room behaviour.

3.4 Predictive Modelling Techniques

Three predictive models of increasing structural complexity were developed and compared. The first, multiple linear regression (MLR), assumes a linear additive relationship between the nine input variables and compressive strength and serves as a baseline against which more complex models are evaluated. The second, second-degree polynomial regression, extends the linear model by incorporating quadratic terms and pairwise interaction effects, allowing for the representation of curvature in the strength-variable relationships. The third, random forest regression (RFR), is an ensemble learning method that constructs a large number of decision trees on bootstrapped subsets of the training data and averages their predictions, making it well suited to capturing non-linear interactions and threshold effects without requiring an explicit functional form. The random forest model was configured with 300 estimators and a maximum tree depth of 8 to balance predictive accuracy against overfitting.

The dataset was randomly partitioned into training (75%) and testing (25%) subsets, with the testing subset withheld entirely from model fitting to provide an unbiased evaluation of generalization performance. All three models were trained using the same partition to ensure a fair comparison.

3.5 Performance Evaluation Criteria

Model performance was quantified using three complementary metrics: the coefficient of determination (R^2), which measures the proportion of variance in compressive strength explained by the model; the mean absolute error (MAE), which expresses the average magnitude of prediction error in MPa; and the root mean square error (RMSE), which penalizes larger errors more heavily and is expressed in the same units as compressive strength. A model exhibiting a higher R^2 and lower MAE and RMSE values was considered to provide superior predictive performance.

4. Results and Discussion

4.1 Descriptive Statistics

Table 1 summarizes the descriptive statistics of the mix-proportion and curing variables together with the measured compressive strength values. The dataset exhibits a wide spread in water-cement ratio (0.29-0.83) and curing age (3-90 days), which was intentional to ensure that the trained models could be evaluated across a broad and representative range of mix designs and curing histories rather than a narrow band of conditions.

Table 1. Descriptive statistics of mix-proportion and curing variables used in this study.

Variable	Min	Max	Mean	SD
Cement (kg/m3)	251.84	498.09	372.91	71.07
Water (kg/m3)	140.43	219.93	180.31	23.82
Water-Cement Ratio	0.29	0.83	0.50	0.12
Fine Aggregate (kg/m3)	600.90	799.09	702.31	57.20
Coarse Aggregate (kg/m3)	950.31	1199.69	1070.88	74.37
Superplasticizer (kg/m3)	0.03	11.92	6.13	3.41
Curing Age (days)	3	90	29.74	29.88
Curing Temperature (°C)	18.01	31.95	24.97	4.14
Curing Humidity (%)	70.13	99.96	85.12	8.55
Compressive Strength (MPa)	8.00	65.00	25.14	13.28

Note: SD = standard deviation; n = 220 concrete mixtures.

4.2 Correlation Analysis

The Pearson correlation matrix presented in Figure 1 quantifies the linear association between each pair of variables. Compressive strength shows the strongest positive correlation with curing age ($r = 0.62$) and cement content ($r = 0.50$), and the strongest negative correlation with water-cement ratio (r = -0.61) and water content (r = -0.33). As expected, cement content and water-cement ratio are themselves strongly negatively correlated (r = -0.80), confirming that these two variables jointly capture the paste-quality effect described by Abrams' law. Fine and coarse aggregate contents, along with superplasticizer dosage, curing temperature, and curing humidity, show comparatively weaker direct correlations with strength, suggesting that their influence is primarily expressed through interaction with the dominant variables rather than as independent linear effects.

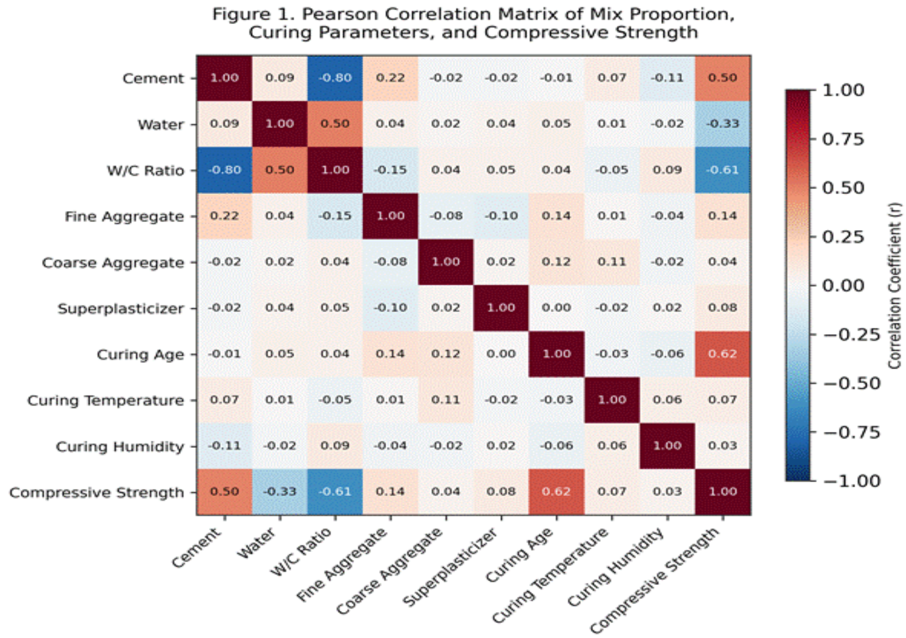

Figure 1. Pearson correlation matrix of mix-proportion, curing parameters, and compressive strength.

4.3 Effect of Water-Cement Ratio and Curing Age

Figure 2 illustrates the combined influence of curing age and water-cement ratio on compressive strength. Panel (a) shows those mixtures with low water-cement ratio (0.30-0.45) consistently achieve higher mean strength at every curing age compared with medium (0.45-0.55) and high (0.55-0.70) water-cement ratio groups, and that the rate of strength gain slows progressively after approximately 28 days for all three groups, consistent with the well-documented deceleration of cement hydration over time. Panel (b) presents the same trend as a continuous scatter plot, with colour representing curing age; the near-monotonic decline in strength with increasing water-cement ratio, modulated by curing age, is clearly visible and reinforces the correlation results reported in Section 4.2.

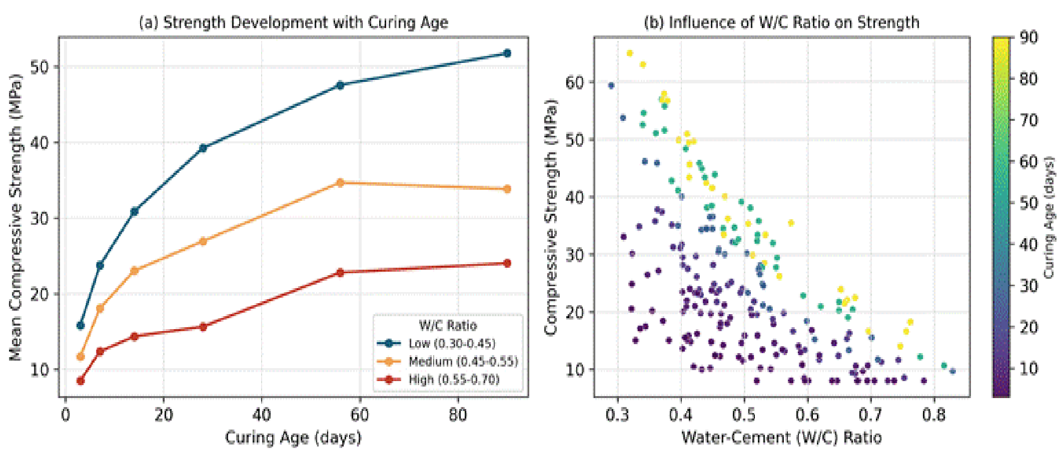

Figure 2. Effect of curing age and water-cement ratio on concrete compressive strength: (a) mean strength development by water-cement ratio band, and (b) scatter of compressive strength against water-cement ratio coloured by curing age.

4.4 Comparative Model Performance

Table 2 reports the performance metrics obtained for the three predictive models on the held-out test subset. The multiple linear regression model achieved an R^2 of 0.739, indicating that a purely linear combination of the nine input variables explains a considerable but incomplete share of strength variability. The second-degree polynomial regression model, which incorporates quadratic and interaction terms, improved performance substantially ($R^2 = 0.928$), confirming the presence of meaningful non-linear and interactive effects among the mix and curing variables. The random forest regression model achieved the best overall performance (R^2 = 0.951, MAE = 2.40 MPa, RMSE = 2.92 MPa), outperforming both regression-based approaches across every evaluation metric.

Table 2. Comparative performance of the three predictive models on the test dataset.

Model	R^2	MAE (MPa)	RMSE (MPa)
Multiple Linear Regression	0.739	5.306	6.730
Polynomial Regression (Degree = 2)	0.928	2.744	3.533
Random Forest Regression	**0.951**	**2.404**	**2.919**

Note: R^2 = coefficient of determination; MAE = mean absolute error; RMSE = root mean square error. Bold row indicates best-performing model.

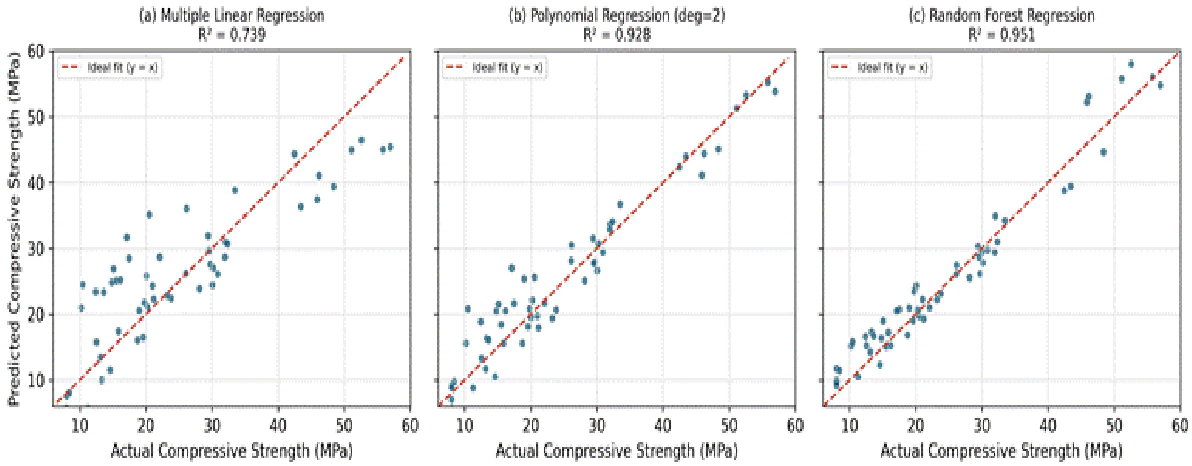

Figure 3. Predicted versus actual compressive strength for (a) multiple linear regression, (b) polynomial regression, and (c) random forest regression models on the test dataset.

4.5 Discussion

The comparative results in Figure 3 visually confirm the numerical findings of Table 2: the scatter of predicted against actual strength values for the random forest model clusters tightly around the ideal 1:1 line across the full strength range, whereas the linear regression model shows pronounced deviation, particularly at higher strength values where non-linear curing and interaction effects become more pronounced. This behaviour is consistent with the correlation structure identified in Section 4.2, in which the dominant predictors, water-cement ratio and curing age, interact with secondary variables such as superplasticizer dosage and curing temperature in a manner that a linear model cannot adequately represent.

The superior performance of the random forest model can be attributed to its ability to partition the input space recursively and to model conditional relationships, such as the reduced sensitivity of high water-cement-ratio mixtures to further increases in curing temperature, without requiring these interactions to be specified explicitly in advance. From a practical standpoint, these results suggest that ensemble tree-based models are particularly well suited to mix-design support tools intended for use across varied material sources and site curing conditions, where the assumption of a fixed linear or low-order polynomial relationship is unlikely to hold uniformly.

It should be noted that the dataset used in this study, while constructed to reflect realistic engineering ranges and well-established strength-development relationships, is not a substitute for site-specific calibration. Regional variation in cement chemistry, aggregate mineralogy, and admixture formulation can shift the quantitative relationships identified here, and practitioners are encouraged to recalibrate the proposed modelling framework using locally sourced trial-mix data before applying it for design-stage strength estimation.

5. Conclusion

This study examined the combined influence of mix-proportion and curing parameters on the compressive strength of concrete using a dataset of 220 mixtures and three predictive modelling techniques of increasing complexity. Correlation analysis confirmed that water-cement ratio and curing age are the dominant predictors of compressive strength, while cement content, curing temperature, and humidity exert secondary but non-negligible influence. Comparative evaluation showed that the random forest regression model substantially outperformed both multiple linear regression and second-degree polynomial regression, achieving an R^2 of 0.951 and an RMSE of 2.92 MPa on unseen test data, compared with an R^2 of 0.739 for the linear model.

These findings demonstrate that ensemble-based machine learning methods are considerably more effective than classical regression approaches at capturing the non-linear and interactive effects that govern concrete strength development under variable mix design and curing conditions. The proposed framework provides a practical decision-support tool for engineers seeking to estimate compressive strength at the mix-design stage, potentially reducing the number of trial mixes required and supporting more time- and resource-efficient construction practice. Future work should extend the modelling framework to incorporate supplementary cementitious materials, recycled aggregates, and larger, multi-source experimental datasets to further improve generalizability across diverse construction contexts.

References

1. Yeh, I.-C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. Cement and Concrete Research, 28(12), 1797–1808.
2. Kim, J.-I., Kim, D. K., Feng, M. Q., & Yazdani, F. (2004). Application of neural networks for estimation of concrete strength. Journal of Materials in Civil Engineering, 16(3), 257–264.
3. Neville, A. M. (2011). Properties of concrete (5th ed.). Pearson Education.
4. Chou, J.-S., & Pham, A.-D. (2013). Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength. Construction and Building Materials, 49, 554–563.
5. Mehta, P. K., & Monteiro, P. J. M. (2014). Concrete: Microstructure, properties, and materials (4th ed.). McGraw-Hill Education.
6. Naderpour, H., Rafiean, A. H., & Fakharian, P. (2018). Compressive strength prediction of environmentally friendly concrete using artificial neural networks. Journal of Building Engineering, 16, 213–219.
7. Chopra, P., Sharma, R. K., Kumar, M., & Chopra, T. (2018). Comparison of machine learning techniques for the prediction of compressive strength of concrete. Advances in Civil Engineering, 2018, 5481705.
8. Kaveh, A., Bakhshpoori, T., & Hamze-Ziabari, S. M. (2018). GMDH-based prediction of compressive strength of concrete using material design and curing conditions. Journal of Soft Computing in Civil Engineering, 2(2), 1–14.
9. DeRousseau, M. A., Kasprzyk, J. R., & Srubar III, W. V. (2018). Computational design optimization for concrete mixtures: A review. Cement and Concrete Research, 109, 42–53.
10. Yaseen, Z. M., Deo, R. C., Hilal, A., Abd, A. M., Bueno, L. C., Salcedo-Sanz, S., & Nehdi, M. L. (2018). Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. Advances in Engineering Software, 115, 112–125.
11. Han, Q., Gui, C., Xu, J., & Lacidogna, G. (2019). A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. Construction and Building Materials, 226, 734–742.
12. Young, B. A., Hall, A., Pilon, L., Gupta, P., & Sant, G. (2019). Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods. Cement and Concrete Research, 115, 379–388.
13. Asteris, P. G., & Mokos, V. G. (2020). Concrete compressive strength using artificial neural networks. Neural Computing and Applications, 32, 11807–11826.
14. Feng, D.-C., Liu, Z.-T., Wang, X.-D., Chen, Y., Chang, J.-Q., Wei, D.-F., & Jiang, Z.-M. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Construction and Building Materials, 230, 117000.
15. Han, T., Siddique, A., Khayat, K., Huang, J., & Kumar, A. (2020). An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete. Construction and Building Materials, 244, 118271.
16. Zhang, J., Li, D., & Wang, Y. (2020). Toward intelligent construction: Prediction of mechanical properties of manufactured-sand concrete using tree-based models. Journal of Cleaner Production, 258, 120665.
17. Nguyen, H., Vu, T., Vo, T. P., & Thai, H.-T. (2020). Efficient machine learning models for prediction of concrete strengths. Construction and Building Materials, 260, 120950.
18. Gupta, S., & Rao, V. (2020). Prediction of compressive strength of concrete using multivariate regression and curing condition variables. Materials Today: Proceedings, 33, 3878–3884.