EPH-International Journal of Applied Science

Volume - 12, Issue - 03
July - 2026
DOI - https://doi.org/10.69980/yepzkz95

DATA-DRIVEN PREDICTION OF CONCRETE COMPRESSIVE STRENGTH

Dr.M.Leena Chandrika

Assistant Professor (Selection Grade), Department: English, Sri Ramakrishna Engineering College,
E-mail id : leena.mahalingam@ srec.ac.in, Orcid ID: 0000-0003-3303-4680

Received:- 30-05-2026, Revised:- 01-07-2026, Accepted:- 08-07-2026, Published:- 15-07-2026

Abstract

Concrete compressive strength is a critical indicator of structural performance and material quality, but conventional laboratory testing is time-consuming and unsuitable for rapid preliminary assessment. This study developed an interpretable data-driven model for predicting concrete compressive strength. The UCI Concrete Compressive Strength Dataset initially contained 1,030 observations; after removing 25 duplicate records, 1,005 unique observations were analysed. Descriptive, correlation, and multiple regression analyses were conducted, followed by an 80:20 training-testing validation. The model was statistically significant and explained 60.4% of the variation in compressive strength. Cement was the strongest positive predictor, followed by blast-furnace slag, curing age, and fly ash, whereas water showed a significant negative effect. The testing model achieved an R2of 0.580, a mean absolute error of 8.90 MPa, and a root mean squared error of 11.19 MPa. The findings demonstrate that concrete strength can be estimated with moderate accuracy using readily available mixture variables. However, the remaining unexplained variation indicates that curing conditions, cement grade, aggregate characteristics, and testing factors should be incorporated in future models. The proposed approach can support preliminary mix evaluation but should complement rather than replace laboratory testing.

Keywords: *Concrete compressive strength; data-driven prediction; multiple regression; concrete mix design; model validation*

Introduction

Concrete is a widely used building material due to its versatility, good availability, durability and resistance to high compression forces. It plays a key role in mechanical performance, especially compressive strength, in structural design, quality control, material selection and evaluation of service-life. Compressive strength is often referred to as an index of the overall quality of hardened concrete since it has been observed to represent the overall effect of the mixture proportions, hydration, curing, aggregate properties and age of the concrete. Accurate strength estimation is thus critical for the development of concrete mixes that meet the engineering specifications before their application in structures. Relationships between the concrete ingredients and strength development, however, are complex and conventional empirical methods necessarily reflect these relationships sufficiently.

The strength of concrete is influenced by several mixing and curing elements. These components are not standalone. For instance, higher cement content can contribute to higher strength, but this must do so, among other things, with the availability of water, the amount of binder replaced, curing conditions, and the proportion of aggregates. The use of supplementary cementitious materials, such as fly ash and slag, can result in lower early-age strength, but can enhance later-age strength from ongoing pozzolanic and hydraulic reactions. These relationships may be of a nonlinear nature and may involve interactions between the variables, thus limiting or even being inconsistent in predicting the behaviour of a mixture in each type of conventional equation. Concrete-property prediction methods have been critically reviewed, and it was demonstrated that several concrete properties influence each other simultaneously in some cases, in which case data-driven methods are particularly useful (Ben Chaabene et al., 2020).

The traditional method for compressive strength testing involves creating concrete specimens, curing them for defined time periods and then testing the strength at a specific time. Laboratory testing is the most accurate way of determining the strength, but it is slow, expensive and impractical for quick preliminary testing. When more than one curing age is required for strength determination and/or many different mixture proportions are being tested, it is especially important to make sure that there are no delays. A complementary method is prediction using data, in which the data of past mixtures and strengths are used to predict performance without having to make a new mix for each initial comparison. Recent studies suggest that the application of predictive analytics in concrete engineering has grown as it can help minimise trial and error and aid in the optimisation of concrete mixture screening (Moein et al., 2023).

With the increasing availability of experimental concrete data, the researchers have been prompted to explore statistical, machine-learning and deep-learning methods of estimating the compressive strength. These methods detect trends in observed data and develop mathematical models for relating the inputs to a mixture to the measured outcomes. The extensive research interest in digitalisation and intelligent decision support in civil engineering is demonstrated through the bibliometric evidence of the rapid growth in research on machine and deep learning for predicting concrete strength (Kumar et al., 2024). However, the prediction accuracy is influenced by the nature of the data, pre-processing, model structure, and range of mixtures used.

In recent years, several studies have investigated the various predictive algorithms for identifying the methods that yield the most stable and accurate estimates. Comparative studies have shown a significant difference in the performance of the models when applied to different datasets, which indicates that no model is always better for all concrete types and conditions of the concrete mix (Loureiro & Stefani, 2024). Optimisation methods have also been used to select the model and optimise the parameters, especially for highly nonlinear relationships between the mixture variables and strength (Lan, 2024). While there are methods that can be more advanced to increase accuracy, these methods can also decrease the interpretability of the solution and may require more computational resources. Thus, it is still useful to have transparent models when the primary task is to estimate the direction and contribution of each variable in the mixture.

High-performance concrete has been of special interest because this concrete type usually uses multiple binders, chemical admixtures, and precise amounts of water. When there are complex interactions, ensemble models and neural networks have yielded better results for estimating the

strength of high-performance concrete (Muhammad et al., 2024). The same conclusion has been drawn for blended cements composed of two or more cementitious materials, and the ensemble models can better describe the joint effect of cement replacement and curing age than the empirical models (Murthy et al., 2024). The usefulness of a model, however, is not just whether it is accurate in predicting the output of interest, but whether it is clear, reproducible, and interpretable in terms of the effect of each input value on the output value.
Strength prediction is further complicated because of the various types of concrete. For example, fly-ash-blended pervious concrete has a different material behaviour, density and pore structure from conventional concrete, thus necessitating prediction models that reflect its special characteristics (Sathiparan et al., 2023). The binder content, aggregate proportions, superplasticiser dosage, and workability also exhibit unique relationships in SSCC, facilitating the application of artificial intelligence (AI) methods for strength estimation (Terlumun et al., 2024). Another challenge arises from the use of ultra-high performance concrete due to its low water/binder ratio, high powder content and special admixtures. Optimised decision tree models have been used successfully to boost the prediction performance of these complex mixtures in recent studies (Zhou et al., 2024).
Despite these advances, there is a need for models that can be interpreted to clearly show the effect of the actual components of the mixtures and the age at which they were cured. There are several studies that provide systematic evidence that these computational intelligence models can also be very effective; however, the comparison between them is difficult due to the differences in the datasets, concrete types, and the procedures used to validate the models (Núñez et al., 2021). Furthermore, optimisation-oriented studies reveal that, apart from accuracy, other features of predictive improvement, such as robustness and practical applicability, should be considered (Mandal et al., 2024). A transparent regression-based analysis could thus give valuable information about the contribution of each predictor while offering a benchmark for comparison to more sophisticated analyses in the future.
Based on this, the present study is conducted to predict the compressive strength of concrete by cement, blast furnace slag, fly ash, water, superplasticiser, coarse aggregate, fine aggregate and curing age. The study analyses descriptive patterns, relationships between variables and the joint predictive power of the eight inputs in the experimental database available in the public domain. It also compares the actual and predicted strength values on traditional performance indicators. The work aims at providing a more interpretable model based on data and identifying the factors that affect the strength of the preliminary mix, and to offer a reproducible framework for the application of these data to future studies with more sophisticated prediction techniques.

2. Methodology

2.1 Research Design

The research design used in this study was quantitative and secondary data collecting design to predict the compressive strength of concrete based on the constituent materials and curing age of the concrete. The study was conducted in a predictive and correlational manner since the relationship between the concrete mixture components and the resulting compressive strength values were explored and the strength values was estimated by an equation. The analysis was performed on the data of concrete mix observations that have been recorded experimentally and are readily available from a public dataset. There was no primary survey or new laboratory experiment carried out. It was a suitable design since the study's objective was to find measurable patterns, quantify the effects of the various concrete ingredients and evaluate the reliability of data-based strength prediction.

2.2 Dataset Source

The Concrete Compressive Strength Dataset, which was obtained from the UCI Machine Learning Repository, was used for the study. The data presented was originally obtained from experimental investigations carried out on concrete mixes that were made with various proportions of binding materials, water, aggregates, chemical admixtures and curing durations. It has 1030 observations; each observation is a mixture and the corresponding compressive strength that was measured in the lab. The selection of the data was made due to the large range of concrete composition, curing time and strength from the data set. Other reasons for its suitability were a clear definition of variables and measurement units.

2.3 Study Variables

Concrete compressive strength (MPa) was the dependent variable. Eight factors—cement, blast-furnace slag, fly ash, water, superplasticiser, coarse aggregate, fine aggregate, and concrete age were considered independent variables. The age was expressed in days, and the amount of cement, slag, fly ash, water, superplasticiser, coarse aggregate, and fine aggregate was expressed in kg/m^3.
The main binding material that is responsible for the strength development in concrete is cement. Blast furnace slag and fly ash were included in the category of supplementary cementitious material, which can have an impact on both early and long-term strength. Water was added since it affects hydration, workability, porosity and water-to-binder ratio. The use of superplasticiser was also considered as it enhances workability and can help to reduce the amount of water used. The coarse and fine aggregates were added as these have been found to influence the density, internal structure and load-bearing capacity of concrete. The age was added as it is generally accepted that the strength increases with a longer curing period.

2.4 Data Screening and Preparation

The data set was explored to make sure that the observations were complete, consistent, and appropriate for predictions before the main analysis was done. Validity of variable names, data type, units of measurement, and number of observations was checked. The data set was examined for missing values, blank data, duplicate data and data that did not make any sense. Duplication of observations was examined to see if it was a real repeated experimental mixture or an accidental repetition.
The distributions of all the variables were also checked for the presence of any extreme values. The automatic exclusion of extreme values was not performed since materials can be present in a concrete in high (or low) quantities, which is a legitimate composition. Rather, such observations were compared to the general distribution of data and the actual features of concrete making. The validity of experimental variation was preserved in this procedure, and there was less chance of data-entry errors affecting the prediction results.

2.5 Descriptive and Relationship Analysis

Descriptive analysis was performed to summarise the general features of the data set. For each predictor and for concrete compressive strength, the mean, standard deviation, minimum and maximum values were calculated. These measures were employed to characterise the central tendency, the variation and the range of the concrete mixture properties. For this purpose, the distribution of compressive strength was also investigated to determine the proportion of low, medium and high strength concrete in the data set.
A correlation analysis between the independent variables and compressive strength was performed to establish the direction and strength of association of each independent variable with the compressive strength. Positive, negative, strong, moderate or weak relationships were determined using correlation coefficients. Relationships between the independent variables were also investigated since the amounts of ingredients used for the concrete production can be dependent on each other. For instance, the mixtures may have fewer water molecules and more superplasticiser, or some of the cement can be substituted with slag or fly ash. These relationships were evaluated to help identify the common predictors and to aid in the interpretation of the prediction model.

2.6 Prediction Model Development

A multiple regression approach was used to develop the concrete compressive strength prediction model. This was chosen because it is a method of estimating the effect of multiple independent variables on a continuous dependent variable. In total, there were eight predictors included to see if they were able to explain the predictive value of compressive strength either individually or in combination. The general prediction equation was given by:

$$CS = \beta_0 + \beta_1 C + \beta_2 BFS + \beta_3 FA + \beta_4 W + \beta_5 SP + \beta_6 CA + \beta_7 FIA + \beta_8 A$$

In this equation, CS represents concrete compressive strength, C represents cement, BFS represents blast-furnace slag, FA represents fly ash, W represents water, SP represents a superplasticiser, CA represents coarse aggregate, FIA represents fine aggregate, and Arepresents concrete age. The term β0 represents the intercept, while β1 to β8 represent the estimated regression coefficients.

The direction of each coefficient was interpreted as the direction of the effect on the compressive strength of the change in the individual predictor; the magnitude of each was interpreted as indicating the strength of the effect of the change in the individual predictor on the compressive strength, with the other factors remaining constant. Predictors were selected based on technical relevance, statistical contribution and proven relevance in the concrete strength development.

2.7 Model Validation and Performance Assessment

The reliability of the developed model was tested by comparing the compressive strength values predicted by the developed model with the experimentally measured values. Prediction errors were obtained by subtracting the actual strength value from the estimated. Variation in compressive strength attributable to predictor variables was represented using the coefficient of determination, R2.

The average absolute difference between the observed and predicted strength values was calculated by using Mean Absolute Error. The Mean Squared Error gave more weight to larger prediction errors, whereas the Root Mean Squared Error measured the prediction error in the same units used for the compressive strength. Additionally, percentage prediction error was used to examine the size of the errors relative to the level of strength. The values of R2 were higher, and the error values were lower, which meant that the predictive performance was better. Using the residual values, it was also investigated if the model was systematically over- or under-predicting the strength of the concrete.

3. Results

3.1 Dataset Screening

The original Concrete Compressive Strength Dataset had a total of 1,030 observations and nine numerical variables. The concrete mixture components and curing age were represented by eight variables and can be seen as independent variables, while the concrete compressive strength is a dependent variable. There was no missing data in the data set. But, while screening the data, 25 identical observations were identified. Duplicate data was deleted to avoid the effects of repeated observations on the results of the analysis. So, to complete the analysis, 1005 distinct observations were used.

The data consisted of cement, blast-furnace slag, fly ash, water, superplasticiser, coarse aggregate, fine aggregate, concrete age, and compressive strength. All variables were quantitative and appropriate for descriptive, correlational and predictive analysis.

3.2 Distribution of Concrete Compressive Strength

The range of concrete compressive strength values in the data set was quite large. The strength values were from 2.33 MPa to 82.60 MPa with a mean value of 35.25 MPa and a standard deviation of 16.29 MPa. This large spread showed that the values were for real concrete mixtures of widely varying strengths. The distribution of concrete compressive strength values is shown in Figure 1.

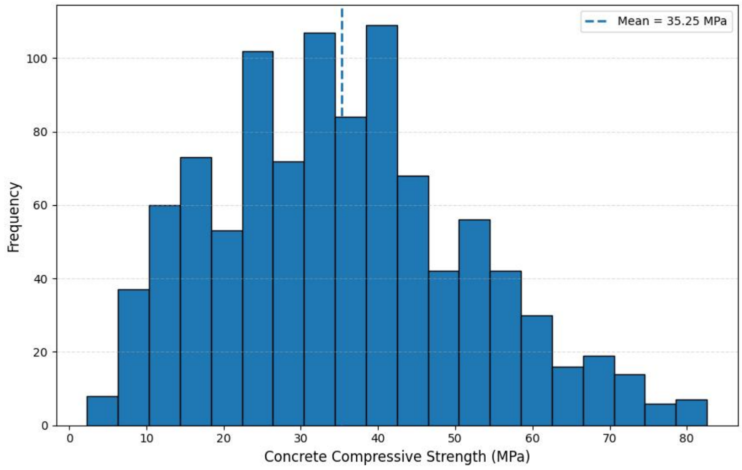

Figure 1. Distribution of concrete compressive strength values

Of the 1,005 unique observations, 196 mixtures had compressive strength values less than 20 MPa. The highest strength category was 20 MPa and less than 40 MPa, with 449 observations. Moreover, the compressive strength of the 281 mixtures fell in the range from 40 MPa to less than 60 MPa, and the compressive strength of the 79 observations was 60 MPa or higher. The data set, therefore, contained low, medium, and high-strength concrete mixes, but medium-strength concrete mixes were generally the most common.

3.3 Descriptive Results of Binding Materials

The average cement content was 278.63 kg/m³, with values ranging from 102.00 kg/m³ to 540.00 kg/m³. The wide range suggested that the data set represented a range of mixtures from low to high cement proportions.

The content of blast-furnace slag ranged from 0.00 kg/m³ to 359.40 kg/m³ with an average value of 72.04 kg/m³. Some of the mixtures did not have any slag, while others included a significant amount of slag as a partial cement replacement. Similar trends were observed for fly ash, the values of which ranged from 0.00 kg/m³ to 200.10 kg/m³ and averaged 55.54 kg/m³. The fact that there were zero values indicated that slag and fly ash were not added to all the concrete mixtures.

3.4 Descriptive Results of Water, Admixture, Aggregates, and Age

Water content ranged from 121.75 kg/m³ to 247.00 kg/m³, with a mean value of 182.07 kg/m³. This variation was due to the workability of the mixtures and the amount of water to binder ratio. The average recorded superplasticiser content is 6.03 kg/m3, and the minimum content is 0.00 kg/m3, and the maximum is 32.20 kg/m3. In some mixtures, no superplasticiser was used, while in others, high proportions of superplasticiser were used.

The average material quantity of coarse aggregate was the highest, with a mean of 974.38 kg/m³ and a standard deviation of 77.58 kg/m³. Fine aggregate ranged from 594.00 kg/m³ to 992.60 kg/m³, with an average of 772.69 kg/m³. The age of concrete varied from 1 day to 365 days, with a mean age of 45.86 days and a standard deviation of 63.74 days. A wide range of curing ages ensured the data span covered early age and long-term concrete strength development. Table 1 shows the descriptive statistics of all the concrete mixture components, the curing age and the compressive strength.

Table 1. Descriptive statistics of the concrete mixture components, curing age, and compressive strength used in the study

Dataset characteristic	Result
Initial number of observations	1,030
Total variables	9
Independent variables	8
Dependent variable	Concrete compressive

	strength
Missing values	0
Exact duplicate observations	25
Final unique observations analysed	1,005
Training observations	804
Testing observations	201

3.5 Correlation Between Predictors and Compressive Strength

Based on the correlation analysis, it was found that the best positive correlation with the concrete compressive strength was with cement (r = 0.488). This result showed that the higher the content of cement, the higher the compressive strength of the mixtures.

The strength was also positively related to Superplasticiser (r = 0.344) and to Concrete Age (r = 0.337). From these results, it was observed that longer curing and optimum use of superplasticiser generally resulted in higher strength.

The correlation coefficient between the blast-furnace slag and compressive strength was r = 0.103, which was weak and positive. Water had the highest negative correlation, r = −0.270, meaning that the more water added to the mixture, the lower the strength. A weak negative correlation of r = −0.187 and r = −0.145 was found for fine aggregate and coarse aggregate, respectively. Fly ash had the least negative direct correlation coefficient (r = −0.081). These correlation values were for individual associations only and did not consider the simultaneous effects of the other concrete components. Table 2 shows the correlations between the predictor variables and the concrete compressive strength.

Table 2. Pearson correlation coefficients between the predictor variables and concrete compressive strength

Variable	Unit	Mean	Standard deviation	Minimum	Maximum
Cement	kg/m³	278.63	104.35	102.00	540.00
Blast-furnace slag	kg/m³	72.04	86.17	0.00	359.40
Fly ash	kg/m³	55.54	64.21	0.00	200.10
Water	kg/m³	182.07	21.34	121.75	247.00
Superplasticizer	kg/m³	6.03	5.92	0.00	32.20
Coarse aggregate	kg/m³	974.38	77.58	801.00	1,145.00
Fine aggregate	kg/m³	772.69	80.34	594.00	992.60
Concrete age	Days	45.86	63.73	1.00	365.00
Compressive strength	MPa	35.25	16.28	2.33	82.60

3.6 Multiple Regression Model Results

Cement, blast furnace slag, fly ash, water, superplasticiser, coarse aggregate, fine aggregate and age were used as predictor variables in the multiple regression model. The overall model was statistically significant, F = 189.75, $p < 0.001$. An R2 of 0.604 and an Adjusted R^2 value of 0.601 were obtained from the model. So, the total variance of the concrete compressive strength was explained by the eight predictors, which accounted for about 60.4% of the variance.

The developed prediction equation was:

$$CS = -17.748 + 0.117(C) + 0.099(BFS) + 0.086(FA) - 0.153(W) + 0.283(SP) + 0.016(CA) + 0.018(FIA) + 0.112(A)$$

In this equation, CS represents concrete compressive strength, C represents cement, BFS represents blast-furnace slag, FA represents fly ash, Wrepresents water, SP represents a superplasticiser, CA represents coarse aggregate, FIA represents fine aggregate, and Arepresents concrete age.

3.7 Contribution of Individual Predictors

The coefficient of cement was 0.117, which was statistically significant and showed a positive correlation ($p < 0.001$). This finding suggested that, with other factors being equal, the estimated increase in the cement content was accompanied by an increase in compressive strength of 0.117 MPa.

The blast furnace slag also proved very positive, B = 0.099, $p < 0.001$. Fly ash increased after removing the effect of other components of the mixture, B = 0.086, $p < 0.001$. The increase in water content was shown to have a significant negative effect (B = −0.153, $p < 0.001$).

Regarding strength, the positive effect of superplasticiser was observed with B = 0.283 and $p = 0.002$. Concrete age was also highly significant as a positive predictor (B = 0.112, $p < 0.001$), indicating that strength increases with the passage of time. However, the coarse aggregate was not statistically significant at 5% level ($p = 0.094$). The p-value of the fine aggregate was also not significant (0.087). Hence, the most meaningful contributors to the prediction model were the cement, blast-furnace slag, fly ash, water, superplasticiser and age. The direction and relative importance of the material strength predictors are presented in Figure 2.

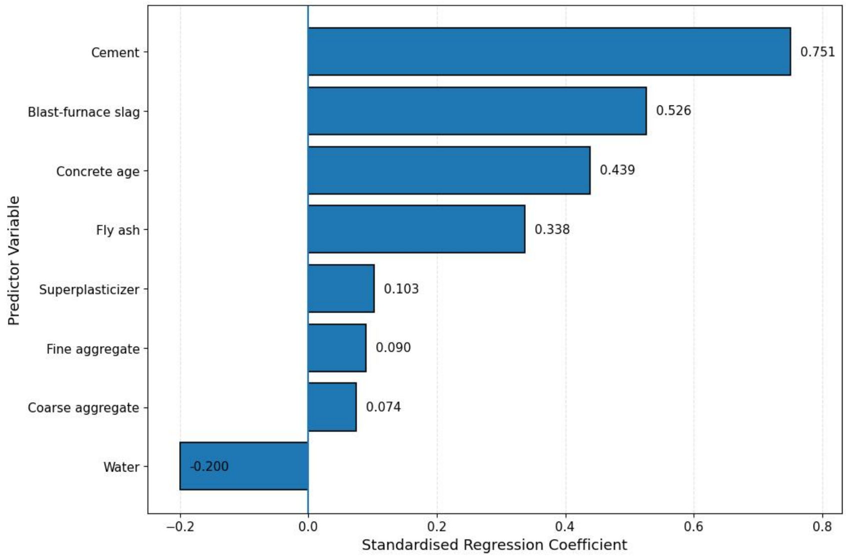

Figure 2. Standardised regression coefficients showing the direction and relative contribution of concrete mixture variables to compressive strength prediction

The regression coefficients and statistical significance of the individual predictors are presented in Table 3.

Table 3. Multiple linear regression coefficients for predicting concrete compressive strength from the study variables

Compressive-strength category	Strength range	Frequency	Percentage
Low-strength concrete	Below 20 MPa	196	19.5%
Medium-strength concrete	20 to <40 MPa	449	44.7%
High-strength concrete	40 to <60 MPa	281	28.0%
Very high-strength concrete	≥60 MPa	79	7.9%
Total		1,005	100.0%

3.8 Model Validation and Prediction Accuracy

To validate the model, 1,005 observations were split into 804 training observations and 201 testing observations. The training model had an R^2 of 0.610 with 61.0% of the variance explained by the training data. The Mean Absolute Error of training was 7.96 MPa, and the Root Mean Squared Error of training was 10.00 MPa.

The model was validated on the test data set, and the R^2 value was 0.580. The Mean Absolute Error was 8.90 MPa, the Mean Squared Error was 125.27, and the Root Mean Squared Error was 11.19

MPa. Good positive agreement (r = 0.764) was found between the measured and predicted compressive strength. The correlation between the measured and predicted relationship of the compressive strength of samples in the test set is shown in Figure 3.

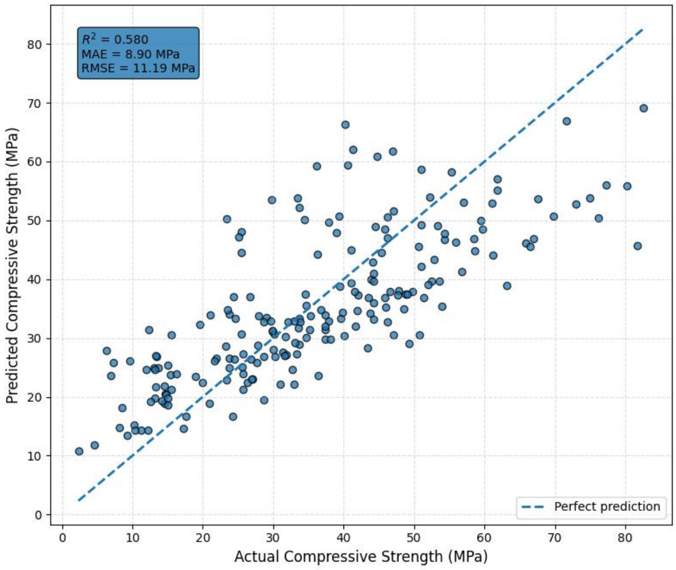

Figure 3. Comparison of actual and predicted concrete compressive strength values

The mean residual of the testing observations was only 0.70 MPa, which indicated little overall prediction bias. However, there were variations in individual prediction errors. Some observations were underestimated by about 26.95 MPa, and some were overestimated by about 36.03 MPa by the model. In general, the predictive accuracy of the model was moderate, and the observed variation in the strength of concrete was almost 60% explained. The training and test results of the Concrete Compressive Strength Prediction model are shown in Table 4.

Table 4. Performance evaluation of the concrete compressive strength prediction model using training and testing datasets

Predictor variable	Correlation with compressive strength (r)	Direction	Relative strength
Cement	0.488	Positive	Moderate
Blast-furnace slag	0.103	Positive	Weak
Fly ash	−0.081	Negative	Very weak
Water	−0.270	Negative	Weak
Superplasticizer	0.344	Positive	Moderate
Coarse aggregate	−0.145	Negative	Weak
Fine aggregate	−0.186	Negative	Weak
Concrete age	0.337	Positive	Moderate

The remaining unexplained variation suggested that additional factors, including curing temperature, humidity, cement grade, aggregate characteristics, and testing conditions, may further influence concrete strength.

4. Discussion

The present study shows that moderate accuracy can be achieved in predicting the concrete compressive strength if the following input variables are used, namely cement, blast-furnace slag, fly ash, water, superplasticiser, coarse aggregate, fine aggregate and curing age. The regression model accounted for 60.4% of the variation in compressive strength, and the testing model yielded an R2 value of 0.580, an MAE of 8.90 MPa and an RMSE of 11.19 MPa. The results showed that the mixture composition and age are important and contribute significantly to the prediction of strength development, but they do not explain most of the variations in the development of strength. The predictive performance observed is comparable to that reported in recent studies, which indicate that there are complex and interdependent material properties that do not seem to be fully represented by a linear structure (Altuncı, 2024).

The cement was the greatest positive predictor of compressive strength and had the highest standardised regression coefficient. This is a technical sound finding because the main binding phases which lead to the development of the load-bearing and dense concrete matrix are created during the cement hydration. It is also shown that the result retained the dominant influence of the cement even after a control was made on the water content, the aggregate content, the supplementary cementitious materials content, the superplasticiser content and the curing age. Recent comparative modelling studies also suggest that the amount of cement or total binder is among the most important variables for a compressive-strength prediction (Sah & Hong, 2024). The other factors were adjusted for, and blast-furnace slag had a strong positive contribution. It was not directly correlated with the strength, but there was a positive regression coefficient for slag, indicating it contributed more when used in combination with cement, water and age. Such a pattern could be a result of its latent hydraulic nature and the progressive development of its concrete microstructure with time. Recent studies have also demonstrated that supplementary cementitious material can offer valuable predictive information if modelled in conjunction with the full mix composition (Nikoopayan Tak et al., 2025).

For fly ash, there was a weak negative direct correlation but a positive and statistically significant regression coefficient with compressive strength. This change indicated that other factors, particularly the percentage of cement replacement, water content and curing time, had a significant influence on its effect. Early age strength can be decreased by fly ash in some mixtures, while fly ash has pozzolanic activity, increasing the strength of the mixture at later ages. Hence, it seems that its relationship with strength is isolated and weak or negative, but when the other factors are adjusted, its relationship with strength becomes positive. Concrete containing supplementary cementitious materials (SCMs) has been shown to have interaction-dependent behaviour in advanced predictive models (Li et al., 2024). The compressive strength of the concrete, on the other hand, clearly decreased because of the water. The negative coefficient indicates that with an increase in water content, strength is generally reduced due to an increase in capillary porosity and a decrease in the strength of the hardened matrix. Water is crucial for hydration and workability of concrete, but excess amounts can cause the density and mechanical properties to decline. Water-related variables have also been found to be important negative factors in strength outcomes in explainable prediction research (Liu & Sun, 2023).

The use of superplasticiser was statistically significant and beneficial to the predicted compressive strength. This relationship can be accounted for by the effect of the superplasticisers in reducing the quantity of mixing water needed and increasing the workability. Consequently, mixtures can be prepared to be more particle dispersed with less porosity while maintaining placement properties. A moderate positive correlation was seen in the present study, which helps to support the fact that chemical admixtures can be used to improve strength with a suitable proportion of binder and water. Predictive research in recent years has also identified the content of superplasticiser as a viable parameter for concrete strength and workability models (Vargas et al., 2024). The other significant positive predictor was the age of the concrete, agreeing that as hydration and secondary reactions progress, strength generally increases. The curing period range was 1–365 days, thus representing both early and long-term behaviour. The curing age has also been found to be one of the most significant factors in the prediction of high-performance concrete in recent studies (Xiao et al., 2025). But gains over time are not always linear, with strength gains generally being greater at younger ages and smaller at older ages. This may be a reason why the linear regression model was not highly accurate for the data. It has been observed that modelling with nonlinear models can more effectively capture the age-related growth of strength in complex concrete datasets (Khan et al., 2025).

The compressive strength had a weak direct correlation with coarse and fine aggregates, and they were not statistically significant in the final regression model. It is not to be taken as a sign of the unimportance of aggregates. The effect of their influence on strength is more dependent on their mass proportions as well as on other factors like grading, particle shape, surface texture, mineral composition, moisture condition and quality of the interfacial transition zone. Studies of recycled and SCC concrete in recent years have indicated that the effects of recycled aggregates are frequently complex and material-specific, with many factors beyond just quantity influencing the material properties (Yang et al., 2024). The limited contribution of aggregates in the current model could thus mean that these are missing in the detailed descriptors, like aggregate type, absorption, strength and particle-size distribution. For concrete systems with special requirements, models have been developed to show that richer aggregate information can lead to better prediction representation (Yu et al., 2024).

The difference between the R^2 value for the training and testing sets is relatively small, indicating that there was no significant overfitting. However, this RMSE of the test predicted values (11.19 MPa) shows that some individual values were significantly different from the observed strength values. Such errors may be linked with the linearity of the model and the exclusion of environmental, material-quality and production-related variables. Some recent soft-computing works have suggested that the ensemble and the nonlinear approaches are better than the classical regression for the prediction of concrete strength, as they can incorporate threshold effects and complex interactions among the input variables (Kumar et al., 2023). Comparative studies have also shown that multiple machine-learning techniques can be used to enhance the prediction when relationships between mixture variables are highly nonlinear (Malhotra et al., 2023). Increased computation complexity may also have the potential to enhance model generalisation and prediction stability (Kim & Lee, 2025), but this is not a direct relationship between computation complexity and meaningful or practically relevant improvement (Ziółkowski, 2025).

In general, the results indicate that data-driven models could be used effectively for initial strength prediction, screening mix-design, and pre-emptively identifying potentially problematic mixtures. Such models should, however, complement and not supplant laboratory testing. More future studies are recommended, including ensemble or deep learning methods and adding more variables like curing temperature, humidity, cement grade, aggregate properties, specimen dimensions and testing conditions. In recent times, there have been indications that, if more detailed and larger sets of data are at hand, advanced learning techniques can lead to better predictive accuracy (Shrivastava & Shrivastava, 2024).

5. Conclusion

This study developed an interpretable data-driven model for predicting concrete compressive strength from eight commonly recorded mixture and curing variables. After removing 25 duplicate observations, the analysis was based on 1,005 unique concrete mixtures. The regression model was statistically significant and explained 60.4% of the variation in compressive strength, while the independent testing dataset produced an R2of 0.580, an MAE of 8.90 MPa, and an RMSE of 11.19 MPa. These findings indicate that the model provides moderate predictive accuracy and reasonable generalisation. Cement was identified as the strongest positive predictor, followed by blast-furnace slag, curing age, and fly ash. Superplasticiser also contributed positively, whereas water had a significant negative effect on predicted strength. Coarse and fine aggregates were not statistically significant after the influence of the remaining variables was controlled, suggesting that aggregate quantity alone may not adequately represent their effects on concrete performance. The findings confirm that mixture composition and curing age provide useful information for preliminary strength estimation, mix-design screening, and comparison of alternative concrete proportions. Nevertheless, the model should not replace standard laboratory compressive-strength testing, particularly for structural design and quality assurance. Future research should include curing temperature, humidity, cement grade, aggregate type, grading, absorption, specimen characteristics, and testing conditions. More advanced nonlinear and ensemble models may also be evaluated to improve predictive accuracy while maintaining interpretability and practical relevance.

References

1. Altuncı, Y. T. (2024). A comprehensive study on the estimation of concrete compressive strength

using machine learning models. Buildings, 14(12), 3851.
2. Ben Chaabene, W., Flah, M., & Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. Construction and Building Materials, 260, 119889.
3. Khan, A. Q., Muhammad, S. G., Raza, A., Chaimahawan, P., & Pimanmas, A. (2025). Advanced machine learning techniques for predicting compressive strength of ultra-high-performance concrete. Frontiers of Structural and Civil Engineering, 19(4), 503–523.
4. Kim, J., & Lee, D. (2025). Comparative study on hyperparameter tuning for predicting concrete compressive strength. Buildings, 15(13), 2173.
5. Kumar, M., Biswas, R., Kumar, D. R., Samui, P., Kaloop, M. R., & Eldessouki, M. (2023). Soft computing-based prediction models for compressive strength of concrete. Case Studies in Construction Materials, 19, e02321.
6. Kumar, R., Althaqafi, E., Patro, S. G. K., Simic, V., Babbar, A., Pamucar, D., Singh, S. K., & Verma, A. (2024). Machine and deep learning methods for concrete strength prediction: A bibliometric and content analysis review of research trends and future directions. Applied Soft Computing, 164, 111956.
7. Lan, W. (2024). Accurate compressive strength prediction using machine learning algorithms and optimization techniques. Journal of Engineering and Applied Science, 71, Article 1.
8. Li, P., Zhang, Z., & Gu, J. (2024). Prediction of concrete compressive strength based on ISSA-BPNN-AdaBoost. Materials, 17(23), 5727.
9. Liu, G., & Sun, B. (2023). Concrete compressive strength prediction using an explainable boosting machine model. Case Studies in Construction Materials, 18, e01845.
10. Loureiro, A. A. B., & Stefani, R. (2024). Comparing the performance of machine learning models for predicting the compressive strength of concrete. Discover Civil Engineering, 1, Article 19.
11. Malhotra, K., Mishra, D., & Tumrate, C. S. (2023). Prediction of concrete compressive strength employing machine learning techniques. Materials Today: Proceedings. Advance online publication.
12. Mandal, S., Shiuly, A., Sau, D., Mondal, A. K., & Sarkar, K. (2024). Study on the use of different machine learning techniques for prediction of concrete properties from their mixture proportions with their deterministic and robust optimisation. AI in Civil Engineering, 3, Article 7.
13. Moein, M. M., Saradar, A., Rahmati, K., Ghasemzadeh Mousavinejad, S. H., Bristow, J., Aramali, V., & Karakouzian, M. (2023). Predictive models for concrete properties using machine learning and deep learning approaches: A review. Journal of Building Engineering, 63, 105444.
14. Muhammad, U. J., Aminu, I. I., Mahmoud, I. A., Aliyu, U. U., Usman, A. G., Jibril, M. M., Malami, S. I., & Abba, S. I. (2024). An improved prediction of high-performance concrete compressive strength using ensemble models and neural networks. AI in Civil Engineering, 3(1), Article 21.
15. Murthy, M. N., Amruth, S. K., Marulasiddappa, S. B., & Naganna, S. R. (2024). Modeling the compressive strength of binary and ternary blended high-performance concrete mixtures using ensemble machine learning models. Soft Computing, 28, 6683–6693.
16. Nikoopayan Tak, M. S., Feng, Y., & Mahgoub, M. (2025). Advanced machine learning techniques for predicting concrete compressive strength. Infrastructures, 10(2), 26.
17. Núñez, I., Marani, A., Flah, M., & Nehdi, M. L. (2021). Estimating compressive strength of modern concrete mixtures using computational intelligence: A systematic review. Construction and Building Materials, 310, 125279.
18. Sah, A. K., & Hong, Y.-M. (2024). Performance comparison of machine learning models for concrete compressive strength prediction. Materials, 17(9), 2075.
19. Sathiparan, N., Jeyananthan, P., & Subramaniam, D. N. (2023). Prediction of compressive strength of fly ash blended pervious concrete: A machine learning approach. International Journal of Pavement Engineering, 24(2), Article 2287146.
20. Shrivastava, S., & Shrivastava, T. (2024). Prediction of concrete’s compressive strength using machine learning algorithms. Materials Today: Proceedings, 103, 183–189.
21. Terlumun, S., Onyia, M. E., & Okafor, F. O. (2024). Predicting the compressive strength of self-compacting concrete using artificial intelligence techniques: A review. AI in Civil Engineering, 3, Article 14.

22. Vargas, J. F., Oviedo, A. I., Ortega, N. A., Orozco, E., Gómez, A., & Londoño, J. M. (2024). Machine-learning-based predictive models for compressive strength, flexural strength, and slump of concrete. Applied Sciences, 14(11), 4426.
23. Xiao, K., Zhang, H., Wei, S., Zhu, C., He, J., Zhu, S., & Yang, X. (2025). Prediction of compressive strength of high-performance concrete based on multiple machine learning models. Frontiers in Materials, 12, 1698248.
24. Yang, S., Sun, J., & Zhifeng, X. (2024). Prediction on compressive strength of recycled aggregate self-compacting concrete by machine learning method. Journal of Building Engineering, 88, 109055.
25. Yu, G., Zhu, S., & Xiang, Z. (2024). The prediction of pervious concrete compressive strength based on a convolutional neural network. Buildings, 14(4), 907.
26. Zhou, R., Tang, Y., Li, H., & Liu, Z. (2024). Predicting the compressive strength of ultra-high-performance concrete using a decision tree machine learning model enhanced by the integration of two optimization meta-heuristic algorithms. Journal of Engineering and Applied Science, 71, Article 43.
27. Ziółkowski, P. (2025). Influence of optimization algorithms and computational complexity on concrete compressive strength prediction machine learning models for concrete mix design. Materials, 18(6), 1386.