EPH–International Journal of Science & Engineering
Published by: E-Pub Solutions Private Limited
ISSN (Online): **2454-2016**
Volume 12 Issue 1 April 2026

COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR PREDICTING CONCRETE COMPRESSIVE STRENGTH

S. Babu*

*School of Marine Engineering and Technology, Indian Maritime University Kolkata,
Kolkata, 700088, India

Article History:

Article Type: **Research**
Received Date: **28/04/2026**
Revised Date: **11/04/2026**
Accepted Date: **19/04/2026**
Published Date: **29/04/2026**

Keywords: Concrete Compressive Strength, Machine Learning, Random Forest, Regression Models, Predictive Modeling, Civil Engineering.

ABSTRACT

Concrete compressive strength is a basic parameter in civil engineering that relates to the quality and longevity of structures. Traditional techniques of measuring compressive strength are based on laboratory tests which are usually time and resource consuming. The paper examines how machine learning methods can be used to estimate compressive strength of concrete in terms of mixture composition and curing age. They used a dataset of 1030 samples reflecting eight input variables; cement, water, aggregates and age. Three machine learning models, all regressions, were implemented: Linear Regression, Decision Tree Regressor, and Random Forest Regressor. The data was split into 80:20 training and testing data. Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and coefficient of determination (R^2) were used to gauge model performance. The Regressor model that performed better was the Random Forest Regressor which had a R^2 of 0.882 with an RMSE of 5.510. The analysis of importance of features showed that the most important factors in the compressive strength prediction are the curing age and cement content. The findings denote that the machine learning models, especially the ensemble-based models, may be useful in forecasting the compressive strength of concrete and provide an effective alternative to conventional experimental techniques.

1. Introduction

Concrete is a construction material that is highly used because of its versatility, durability and also cost-effectiveness. It is important in the construction of infrastructure including buildings, bridges and pavements. Concrete structures rely mostly on its mechanical properties which include compressive strength which is the most important parameter. Compressive strength is the capability of concrete to withstand the imposed loads and not to fail and it is as such important in terms of structural safety and reliability.

Conventionally, compressive strength of concrete has been established using standard laboratory tests. These techniques include the specimen casting and curing them over set periods of time and then testing them under controlled conditions. Even though such procedures give correct information, they are time consuming and involve a lot of effort, materials and cost. Curing can require up to 28 days, which extends construction decision-making and extends project timelines. Consequently, alternative methods capable of yielding quicker and accurate forecasts of concrete strength are increasingly needed.

The past few years have seen data-driven approaches, especially machine learning, garner much attention in the engineering of civil engineering. Machine learning algorithms have the ability to learn the relationships amongst many variables and the responses in a complex manner. This renders them one of the appropriate in forecasting concrete compressive strength concerning mixture proportions and curing conditions. (Sah & Hong, 2024) showed that machine learning models are able to make high-accuracy estimates of concrete strength. (Elshaarawy et al., 2024) also demonstrated that these models are able to decrease reliance on experimental testing but retain their effectiveness in prediction.

A number of machine learning methods have been utilized to forecast concrete strength. (Zhang et al., 2024) introduced a deep forest-based model that can capture the nonlinear correlations in concrete data. (DeRousseau et al., 2019) performed a comparative study and emphasized the usefulness of machine learning methods to predict compressive strength. (Feng et al., 2020) proposed an adaptive boosting model that improves the accuracy of predictions by pooling together a number of learners. Also, (Li & Song, 2022) have shown that ensemble learning methods are more effective in predicting the strength of high-performance concrete than traditional regression models. (Gupta & Sihag, 2022) focused on the significance of the comparison of various predictive models to find the most appropriate one to be used in the engineering practice.

With these developments, research on simple, interpretable and efficient models that can be practically applied is still needed. A lot of current methods are based on the complicated hybrid methods, which are not always practical in real-life situations. Thus, a comparative study of regression models that are frequently used is evident with the aim of quantifying their capability in predicting concrete compressive strength.

The current paper fulfils this requirement by building and testing machine learning-based models with the help of easily accessible parameters of concrete mixtures. This study has the following objectives:

- To develop predictive models for concrete compressive strength using Linear Regression, Decision Tree Regressor, and Random Forest Regressor.
- To evaluate the performance of these models using statistical metrics such as MAE, MSE, RMSE, and R^2.
- To compare the effectiveness of the selected models and identify the most accurate and reliable approach for prediction.

2. Literature Review

Prediction of the concrete compressive strength is a significant field of research in the field of civil engineering since the strength development is influenced by a variety of interacting variables such as cement content, water content, aggregate proportion, mineral admixtures, chemical admixtures, and curing age. The relationship between the proportions of mixtures and compressive strength has been known using traditional statistical methods but these methods have had a difficult time in nonlinear behaviour in concrete mixtures. The study by (Young et al., 2019) investigated the possibility of concrete compressive strength prediction based on the mixture proportions, and demonstrated that the variables of the mixture design still have valuable predictive information, but sophisticated analytical tools are needed to address the complex relationships.

Machine learning has gained great significance in predicting mechanical properties of concrete since it is able to detect concealed trends in huge data volumes. The review of machine learning applications in predicting concrete mechanical properties by (Ben Chaabene et al., 2020) indicated that data-driven models could be used more than traditional empirical methods to predict mechanical properties. Their survey also emphasized that the choice of models, the quality of data sets, and feature representation have a significant impact on prediction performance.

A number of researchers have used machine learning algorithms on concrete mixes with additional cementitious materials. After machine learning algorithms were employed to predict compressive strength of admixture fly ash concrete, (Song et al., 2021) discovered that machine learning methods can be used to successfully model the influence of material composition on the development of its strength. Their research justifies the application of data-based approaches to concrete mixtures in which the traditional equations might not be adequate to describe the effect of admixtures.

Ensemble models have become the focus of attention among machine learning methods because of their capability to enhance predictive performance and minimize model variation. (Han et al., 2019) suggested a better Random Forest algorithm to predict the compressive strength of high-performance concrete and the authors claimed a good predictive ability. The discovery is especially pertinent to the current research since Random Forest is also among the primary models that are tested within the scope of the current research. It has also been observed through comparative studies that it is useful to test many more algorithms as opposed to using only one predictive model. (Silva et al., 2020)explored the machine learning method in concrete compressive strength prediction and showed that varying algorithms can generate varying degrees of accuracy based on the dataset and model setting. This justifies the necessity of comparing models in choosing an appropriate prediction method to apply in civil engineering.

Machine learning models based on Support Vector Machines have been considered in prediction of concrete strength as well. To predict compressive strength of concrete in marine conditions, (Ling et al., 2019) used Support Vector Machine along with K-fold cross-validation. They found that cross-validation can enhance the reliability of models by minimizing the reliance on one train-test split, particularly with environmental conditions influencing concrete behaviour.

Artificial Neural Networks have been extensively used due to its capacity to describe nonlinear relationships. (Naderpour et al., 2018) based their investigation on predicting the compressive strength of environmentally friendly concrete through artificial neural networks and demonstrated that ANN models could be effective in giving the correct results in complex concrete mixtures. Nonetheless, ANN models can be less interpretable than simple regression or tree-based models, and they might need tuning.

Concrete strength prediction has also been done using hybrid and symbolic modelling. (Al-Hashem et al., 2022) estimated the compressive strength of the concrete with fly ash and rice husk ash using ANN and Gene Expression Programming models. Their results revealed that the higher order modelling methods can effectively estimate the concrete strength, but these techniques are more sophisticated to work with the simple regression models available in the market.

Non-destructive testing data has also been tested using machine learning. Ansari et al. made a comparative analysis of machine learning models that predict concrete compressive strength based on non-destructive testing methods. Their work showed that machine learning can be used to estimate the strength not only by taking a mixture composition but also by taking a field-based testing indication to increase its use in construction quality(Ansari et al., 2024).

The increasing significance of computational intelligence in concrete strength prediction has been proven by systematic reviews. Nunez et al. conducted a review of computational intelligence approaches to predicting compressive strength of contemporary concrete mixtures and discovered that machine learning models typically give good predictive performance. Nevertheless, they also stressed the importance of adequate evaluation and comparison of models to be sure of valid conclusions(Nunez et al., 2021).

Machine learning methods have also been used to study high-performance concrete. Liu used machine learning to predict the strength of high-performance concrete and found that the data-driven models can be efficient in estimating compressive strength with the proper input variables. This aids in the greater generality of machine learning in other types of concrete(Liu, 2022).

In general, the literature reviewed has shown that machine learning techniques offer a good alternative to conventional laboratory-based strength estimation. It has been demonstrated in previous research that ANN, SVM, Random Forest, boosting, ensemble learning and other supervised models can be used to predict concrete compressive strength with utility. Most available literature however concentrates on advanced, hybrid or highly complex models, which might not necessarily be simple to adopt in basic engineering applications. Thus, a research gap is still there regarding straightforward and easy comparative research involving the simple and readily available regression-based models. The current paper fills this gap by comparing Linear Regression, Decision Tree Regressor and Random Forest Regressor in predicting concrete compressive strength with the input variables as the mixture composition and curing age.

3. Methodology

This section reports on the dataset used in the study, data preprocessing procedures, machine learning models used and evaluation metrics used to measure the performance of the model. The methodology will be structured in such a way that it provides a systematic and reproducible method of predicting concrete compressive strength with the help of regression-based machine learning models.

3.1 Dataset Description

The dataset that has been used in this research is one of the most commonly used benchmarks dataset to predict compressive strength of concrete in the products of civil engineering. It has 1030 observations, each of which is a distinct concrete mix design and its compressive strength. The dataset has a total of nine variables, eight of which are input features and one that is the target variable. The data in this paper was collected in the UCI machine learning repository and it comprises of 1030 samples of concrete mixtures and the compressive strength values of these mixtures (Yeh, 2007).

The input variables are the concrete mixture composition and the condition of the curing. These are the amounts of cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate and the days of age of the concrete. All these factors contribute greatly towards the strength properties of concrete. The dependent variable in the regression problem is the compressive strength of concrete in megapascals (MPa) which is the output variable.

The dataset can be used in machine learning because it includes both the composition of materials and the time-dependent effects, which can be learned by the models in order to form complex correlations between the input parameters and the compressive strength.

3.2 Data Preprocessing

Before the development of the model, the dataset was pre-processed to achieve consistency and analysis appropriateness. To begin with, the names of the columns were made simpler and uniform in order to enhance ease of reading and simpler implementation in the coded environment. Then the dataset was analyzed in terms of missing data and discrepancies. It was noted that no missing or null values were detected in the dataset, which did not require data imputation or dropping of missing or incomplete records.

The data were divided into input variables and the target variable. The independent variables were represented by the input matrix of eight variables and the compressive strength values were represented in the output vector. The dataset was split into training and testing subsets to assess the performance of the machine learning models. A 80:20 split was taken with 80 percent of the data being used to train and 20 percent to test. This gave 824 training and 206 test samples. To guarantee reproducibility of the results, a fixed random state of 42 was employed.

This study did not use any feature scaling or normalization methods. The rationale behind this decision was the fact that it included tree-based models like Decision Tree and Random Forest that will not be sensitive to the size of the input features. Also, Linear Regression worked well without the need of normalization of this data.

3.3 Machine Learning Models

This paper has used three machine learning models based on regression to forecast the compressive strength of concrete. These models are: Linear Regression, Decision Tree Regressor and the Random Forest Regressor. The models have been chosen depending on their simplicity, interpretability, and ability to deal with regression problems.

Linear Regression is a basic statistical method, which determines a linear relationship between independent variables and the dependent variable. It is based on the assumption that the output variable is a weighted sum of the input features. It is a simple model but can be useful as a baseline of performance of more complex models.

Decision Tree Regressor is a non-linear model which subdivides the data into smaller subsets with respect to the features values. It builds a hierarchical tree where the nodes on the tree are the decision rules (internal nodes) and the predicted values (leaf nodes). This model can represent nonlinear relationships between the variables but can be overfitted when the tree is overgrown.

The Random Forest Regressor is an ensemble learning algorithm, which constructs several decision trees and sums up their results to enhance the degree of prediction. Random Forest does not use just one tree; it uses multiple trees and thereby minimizes variance and increases the generalization performance. The Random Forest model in this study was run with 100 decision trees, and a random state of 42. This will make it more stable and less prone to overfitting than a single decision tree model.

3.4 Evaluation Metrics

Four standard regression measures were used to assess the performance of the machine learning models. These measures give a holistic view of how well prediction is made and the error distribution.

The Mean Absolute Error (MAE) is used to measure the size of the errors in the predicted and actual values without taking into account the direction of the errors. It gives a direct interpretation of the model accuracy in the identical units with the target variable.

Mean Squared Error (MSE) is calculated to obtain the average of the squared error between the predicted and actual values. The measure is sensitive to outliers since larger errors are heavily penalized.

Root Mean Squared Error (RMSE) is the square root of MSE and it is the standard deviation of errors. It is widely applied since it gives the measure of error in the same unit as the output variable.

The Coefficient of Determination, also known as the R 2 score, is a measure that shows the extent to which the model can account in the variation of the target variable. The higher the value of R 2, the greater the model performance with a value approaching one being a stronger predictive ability.

The three machine learning models were evaluated using these metrics to compare their performance and to determine which model best predicts compressive strength of concrete and is most reliable.

4. Results

This section presents the performance evaluation and comparative analysis of the machine learning models developed in this study. The results are discussed using both quantitative metrics and graphical representations to provide a comprehensive understanding of model performance.

4.1 Model Performance Evaluation

The predictive performance of the implemented machine learning models was evaluated using standard regression metrics. These metrics provide a quantitative basis for comparing how accurately each model predicts concrete compressive strength. The performance values obtained from the testing dataset are summarized in Table 1.

Table 1. Performance Comparison of Machine Learning Models

Model	MAE	MSE	RMSE	R^2 Score
Linear Regression	7.745	95.975	9.797	0.628
Decision Tree Regressor	4.591	53.673	7.326	0.792
Random Forest Regressor	3.755	30.358	5.510	0.882

This table presents the performance metrics of all three machine learning models evaluated in this study, showing that the Random Forest Regressor achieves the highest prediction accuracy.

Based on Table 1, it is evident that the Random Forest Regressor is superior to the rest of the models in all aspects of evaluation. It has the least values of MAE, MSE and RMSE and the highest value of R2 of 0.882. The Decision Tree Regressor exhibits intermediate performance whereas Linear Regression has the lowest performance because it is unable to represent nonlinear relationships that occur in the data.

4.2 Comparative Analysis of Model Accuracy

To better visualize the differences in model performance, a graphical comparison of R^2 scores is presented. The R^2 score is a key metric that indicates how well the model explains the variability in the target variable. The comparison of R^2 values for all models is illustrated in Figure 1.

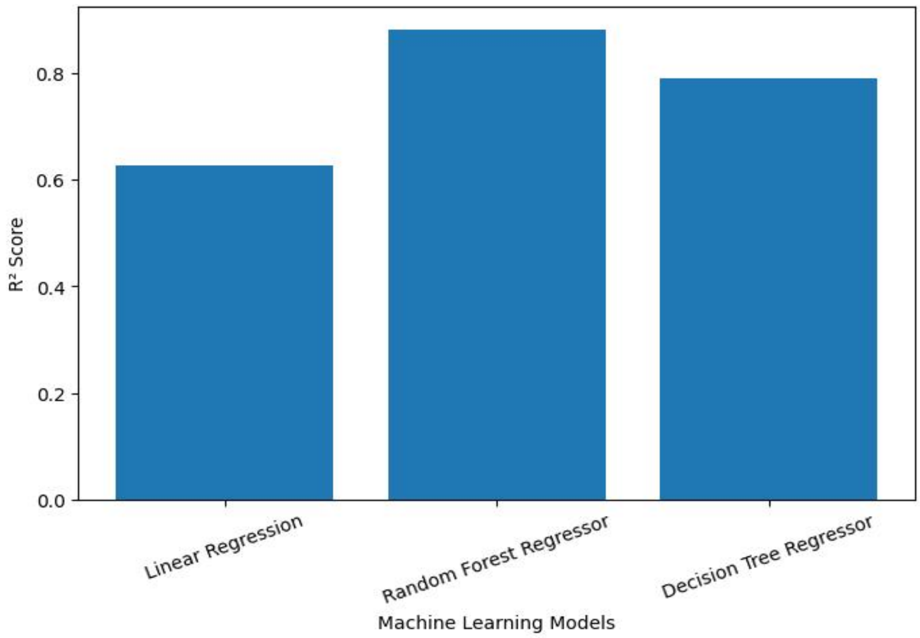

Figure 1. Comparison of R² Scores for Machine Learning Models

The following figure compares the R 2 scores of the Linear Regression, Decision Tree Regressor and the Random Forest Regressor. The best score is that of the Random Forest model and this implies that it has good predictive performance.

The plotting of the data shows clearly that the Random Forest Regressor offers the most suitable fit to the data followed by the Decision Tree Regressor. Linear Regression has the lowest R 2 value which validates its weak ability to model complex relationships between input features and compressive strength.

4.3 Actual vs Predicted Value Analysis

Besides numerical assessment, visual analysis of observed and estimated values are also used to determine the extent to which each model is similar to real performance. The relationship between actual compressive strength values and model predictions are illustrated by using scatter plots. Figures 2, 3, and 4 show prediction behaviour of each model.

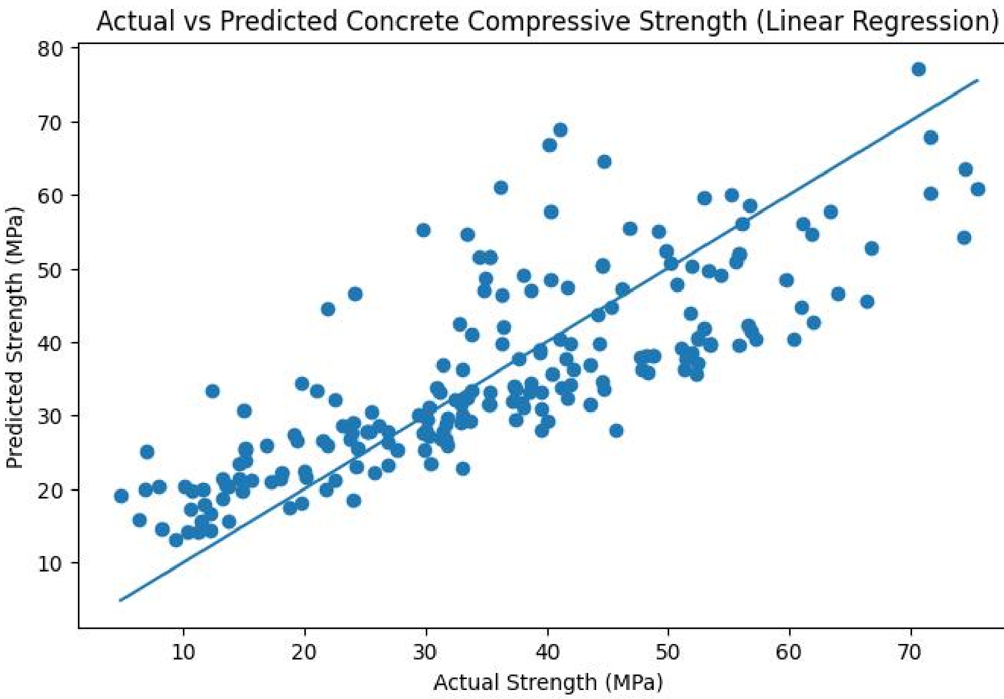

Figure 2. Actual vs Predicted Values for Linear Regression

This figure shows the correlation between the real and the predicted values using Linear Regression. The large standard deviation of the points suggest lower accuracy of prediction. The Linear Regression model has a visible distribution of points off of the optimal diagonal line, which means that the predictions are significantly different than the real values.

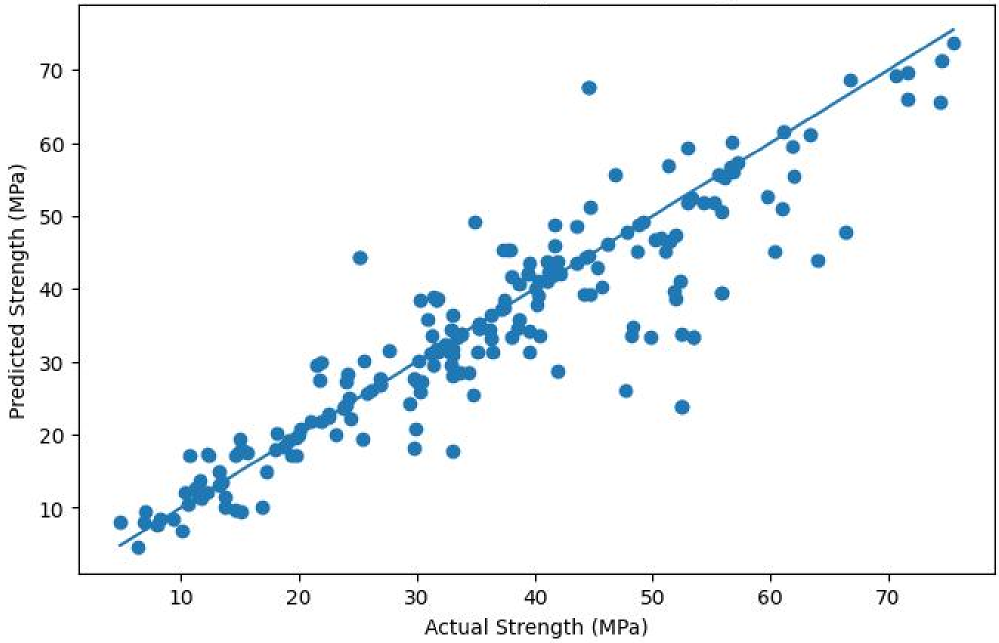

Figure 3 Actual vs Predicted Values for Decision Tree Regressor

This figure shows the comparison between actual and predicted values for the Decision Tree Regressor. The alignment improves compared to Linear Regression, but some variability remains.
The Decision Tree model demonstrates improved prediction accuracy compared to Linear Regression, with data points more closely aligned to the diagonal line. However, some deviations still exist, indicating moderate prediction performance.

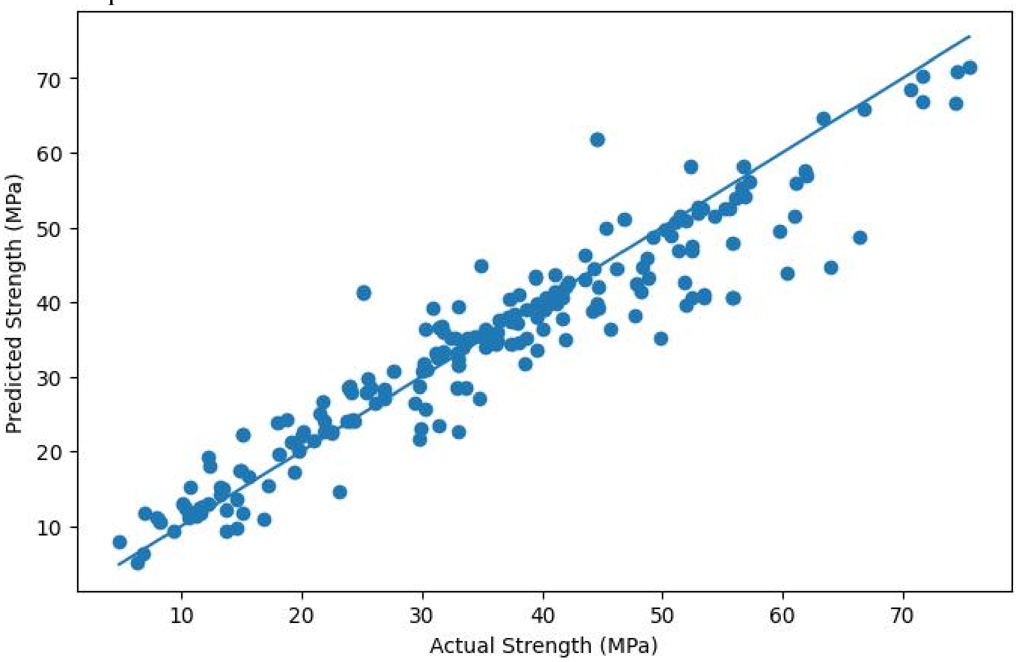

Figure 4 Actual vs Predicted Values for Random Forest Regressor

This figure compares the real and estimated numbers of the Random Forest Regressor. Close clustering of the points on the diagonal line is an indicator of high accuracy of prediction. The Random Forest Regressor records a high level of agreement between the forecasted and the actual values. The majority of the data points are near to the ideal line which proves its strong predictive power and consistency.

4.4 Feature Importance and Correlation Analysis

The feature importance analysis and correlation analysis were conducted to better comprehend the behavior of the dataset and how input variables affect the compressive strength prediction. These analyses give us an understanding of the role of the various variables in predicting and their relationship with each other. The analysis of feature importance was performed with help of Random Forest Regressor, which provided an importance score to each input variable according to its contribution to the prediction process. Figure 5 shows the results of this analysis.

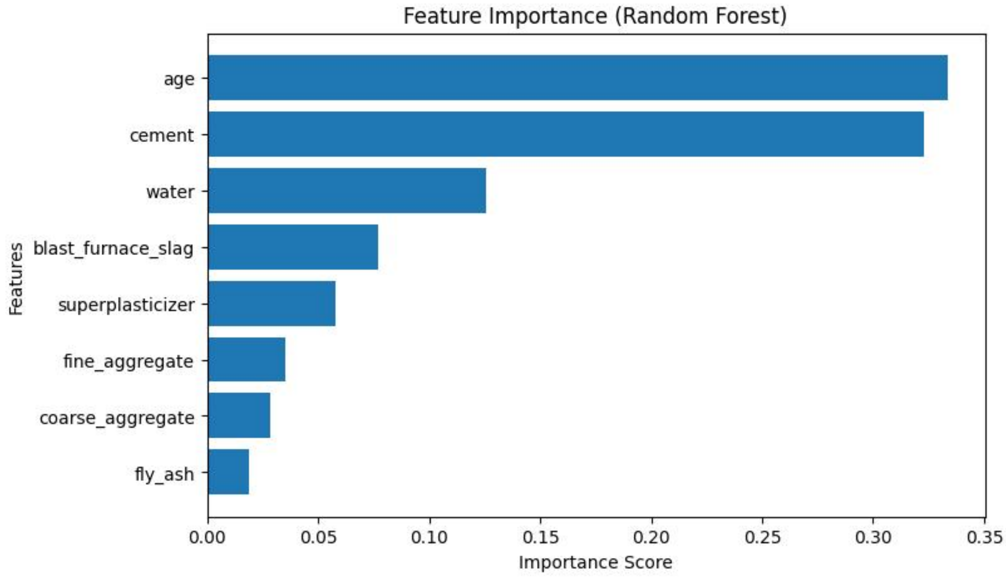

Figure 5. Feature Importance Based on Random Forest Regressor

This figure shows the relative value of the input variables in predicting concrete compressive strength. The most significant features are distinguished as age and cement.

Based on Figure 5, age and cement content are the most important variables with the highest scores meaning that they have the greatest influence when it comes to compressive strength. The water content has also been found to have moderate significance as it affects the water-cement ratio which is also a major factor in the development of concrete strength. The slag of blast furnace, superplasticizer, and aggregates are also the other variables that are involved in the prediction although their contribution is relatively lower.

Besides feature importance, correlation analysis was also done to test the relationships between input variables and the target variable. These relationships were then calculated in the form of a correlation heatmap to visualize the strength and direction of the relationships as presented in Figure 6.

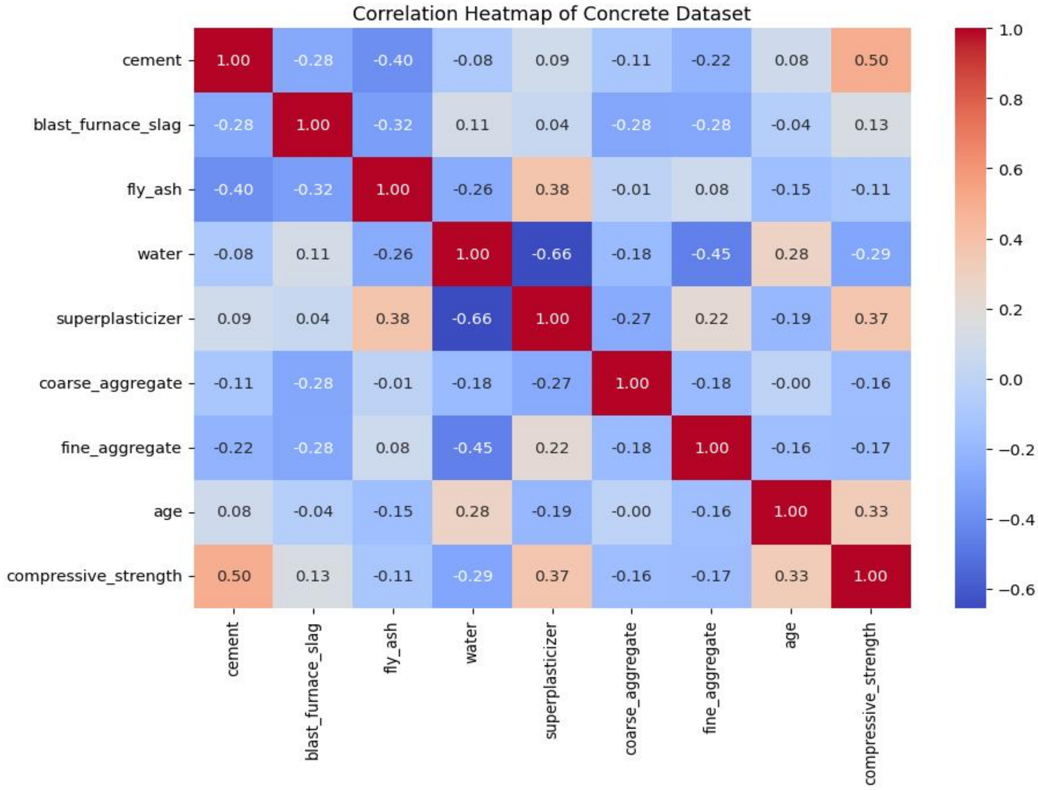

Figure 6. Correlation Heatmap of Concrete Dataset

This number reflects the correlation table of all of the variables in the data. There are positive and negative correlations between input features and compressive strength that are emphasized.

The correlation heatmap indicates that compressive strength is positively correlated with cement content, whereas water content is negatively correlated, suggesting that an increase in water content usually decreases the strength. Age also exhibits a positive correlation with compressive strength which is in line with curing of concrete. The rest of the variables are moderately to weakly related, indicating that their effects are more complicated and could be associated with other characteristics.

Altogether, the analysis of the feature importance and correlation allow obtaining a useful understanding of the structure of the dataset and corroborate the results of the machine learning models. These analyses affirm the fact that the main variables like the age, cement and water are very important to determine the compressive strength of concrete, which is consistent with the stipulated requirements of concrete technology.

4.5 Key Observations

The findings of the numerical and graphical analysis reveal that the performance of machine learning models differs greatly based on the capability of the models to capture nonlinear relationship. Linear Regression is a simple linear model and is not able to capture the complexity of the data.

The Decision Tree Regressor is better at predicting the data than the logistic regression model because it can process nonlinear patterns, but can be prone to overfitting and instability. Conversely, the Random Forest Regressor makes the best and correct predictions since it is an ensemble learning model because it uses several decision trees to improve the generalization and minimization of variance.

All in all, it can be concluded that the Random Forest Regressor model is the most successful in predicting concrete compressive strength in this work. These findings indicate that ensemble-based machine learning techniques are very practical in this kind of regression problem and can be of use in practice in civil engineering applications.

5. Discussion

The findings achieved in this research give a clear insight into the efficiency of various machine learning models in forecasting concrete compressive strength. The comparative analysis reveals that model performance differs considerably with the capacity of every algorithm to reflect intricate relationships regarding input variables as well as the intended output. Linear Regression had the weakest predictive performance whereas the Decision Tree Regressor had moderate improvement. The Random Forest Regressor, however, proved to be the most accurate and this implies that it can be used in this kind of regression problem.

The nature of Linear Regression to assume linear relationship between input features and output could explain the relatively poor performance of the technique. The problem of concrete mixtures is that the dependence of compressive strength on material composition is nonlinear by nature as a result of interdependences between such variables as cement content, water ratio, and curing age. Consequently, Linear Regression will not be able to reflect these complexities, resulting in an increase in the error values and a reduced predictive accuracy.

The Decision Tree Regressor overcomes this drawback by estimating nonlinear relationships by recursively partitioning the data. It is capable of determining patterns that cannot be described by linear models, and that is why it is more effective than Linear Regression. Decision trees however are subject to overfitting, especially in situations where the tree gets to be too complex. This may create a sense of instability and less generalization when used on invisible data.

The superior performance of the Random Forest Regressor can be explained by its ensemble learning approach. Instead of relying on a single decision tree, Random Forest constructs multiple trees and aggregates their predictions. This process reduces variance and enhances model stability. By averaging the results of several trees, Random Forest minimizes the risk of overfitting while maintaining the ability to capture nonlinear relationships. Similar findings have been reported by (Li & Song, 2022), who demonstrated that ensemble learning models outperform traditional regression methods in predicting high-performance concrete strength. In addition, (Han et al., 2019) showed that Random Forest-based approaches provide strong predictive capability for concrete compressive strength due to their robustness and ability to handle complex data structures.

Another important insight from this study is derived from the feature importance analysis performed using the Random Forest model. The results indicate that curing age and cement content are the most influential variables affecting compressive strength. This aligns with fundamental principles of concrete technology, where strength development is highly dependent on hydration processes and material composition. The influence of water content and supplementary materials such as slag and fly ash was also observed, although to a lesser extent. These findings are consistent with previous research, where mixture proportions and curing conditions have been identified as key determinants of compressive strength (Young et al., 2019).

The results of this study are also in agreement with broader trends observed in the literature. For instance, (Sah & Hong, 2024) reported that machine learning models, particularly ensemble methods, provide higher prediction accuracy compared to traditional approaches. Similarly, (DeRousseau et al., 2019) emphasized the importance of comparing different machine learning techniques to determine the most effective model for concrete strength prediction. These studies support the conclusion that ensemble-based models such as Random Forest are well-suited for this application.

From a practical perspective, the use of machine learning models for predicting concrete compressive strength offers significant advantages in construction engineering. By providing rapid and accurate predictions, these models can reduce reliance on time-consuming laboratory testing. This enables faster decision-making during the design and construction phases, improving efficiency and reducing project costs. Engineers can use such models to optimize mix designs, evaluate material properties, and ensure quality control without extensive experimental procedures.

Despite the promising results, this study has certain limitations that should be acknowledged. First, the dataset used, although widely accepted, is limited to 1030 samples and may not fully represent all possible variations in concrete mixtures. A larger and more diverse dataset could further improve model generalization. Second, the study focuses on only three machine learning models. While these models provide valuable insights, the inclusion of additional algorithms such as Support Vector Regression, Artificial Neural Networks, or gradient boosting methods could offer a more comprehensive comparison.

Overall, the findings highlight the effectiveness of Random Forest Regressor as a reliable and accurate model for predicting concrete compressive strength. The study demonstrates that ensemble learning techniques are particularly well-suited for handling complex, nonlinear relationships in engineering datasets. These results contribute to the growing adoption of machine learning in civil engineering and support its application in practical construction scenarios.

6. Conclusion

This paper examined the use of machine learning to predict the compressive strength of concrete given the mixture composition and curing age, as the input variables. Three regression-based models namely the Linear Regression, Decision Tree Regressor and Random Forest Regressor were formulated and tested using the standard performance metrics. The findings indicated that the performance of the models can be quite different based on the capability to reproduce nonlinear relationships in the data. The best-performing model was the Random Forest Regressor, as it had the highest R 2 and lowest error values. The other models were not as accurate and robust as its ensemble learning mechanism allowed. The results support the idea that machine learning, especially ensemble methods can be used as a useful alternative to conventional experimental methods to predict concrete compressive strength. The paper details the increasing role of machine learning in civil engineering, where data-driven methods have the potential to promote efficiency, minimise testing time and aid in decision-making in construction practices.

In future research, a more sophisticated model, including Extreme Gradient Boosting (XGBoost) and Artificial Neural Networks (ANN), can be investigated to enhance accuracy in prediction even further. Moreover, more and larger datasets should be used to promote generalization of the model. Such models could also be practically and reliably applied in real-world construction projects, which could further confirm their practical usefulness.

References

1. Al-Hashem, M. N., Amin, M. N., Raheel, M., Khan, K., Alkadhim, H. A., Imran, M., Ullah, S., & Iqbal, M. (2022). Predicting the Compressive Strength of Concrete Containing Fly Ash and Rice Husk Ash Using ANN and GEP Models. *Materials*, *15*(21), 7713. https://doi.org/10.3390/ma15217713
2. Ansari, S. S., Ansari, H., Khateeb, A., & Ibrahim, S. M. (2024). Comparative study of machine learning models for predicting the compressive strength of concrete using Non-Destructive Testing methods. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2024.04.009
3. 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. https://doi.org/10.1016/j.conbuildmat.2020.119889
4. DeRousseau, M. A., Laftchiev, E., Kasprzyk, J. R., Rajagopalan, B., & Srubar, W. V. (2019). A comparison of machine learning methods for predicting the compressive strength of field-placed concrete. *Construction and Building Materials*, *228*, 116661. https://doi.org/10.1016/j.conbuildmat.2019.08.042
5. Elshaarawy, M. K., Alsaadawi, M. M., & Hamed, A. K. (2024). Machine learning and interactive GUI for concrete compressive strength prediction. *Scientific Reports*, *14*(1), 16694. https://doi.org/10.1038/s41598-024-66957-3
6. 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. https://doi.org/10.1016/j.conbuildmat.2019.117000
7. Gupta, S., & Sihag, P. (2022). Prediction of the compressive strength of concrete using various predictive modeling techniques. *Neural Computing and Applications*, *34*(8), 6535–6545. https://doi.org/10.1007/s00521-021-06820-y
8. 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. https://doi.org/10.1016/j.conbuildmat.2019.07.315
9. Li, Q.-F., & Song, Z.-M. (2022). High-performance concrete strength prediction based on ensemble learning. *Construction and Building Materials*, *324*, 126694. https://doi.org/10.1016/j.conbuildmat.2022.126694
10. Ling, H., Qian, C., Kang, W., Liang, C., & Chen, H. (2019). Combination of Support Vector Machine and K-Fold cross validation to predict compressive strength of concrete in marine

environment. *Construction and Building Materials*, *206*, 355–363. https://doi.org/10.1016/j.conbuildmat.2019.02.071

11. Liu, Y. (2022). High-Performance Concrete Strength Prediction Based on Machine Learning. *Computational Intelligence and Neuroscience*, *2022*(1), 5802217. https://doi.org/10.1155/2022/5802217
12. 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. https://doi.org/10.1016/j.jobe.2018.01.007
13. Nunez, 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. https://doi.org/10.1016/j.conbuildmat.2021.125279
14. Sah, A. K., & Hong, Y.-M. (2024). Performance Comparison of Machine Learning Models for Concrete Compressive Strength Prediction. *Materials*, *17*(9), 2075. https://doi.org/10.3390/ma17092075
15. Silva, P. F. S., Moita, G. F., & Arruda, V. F. (2020). *Machine learning techniques to predict the compressive strength of concrete*. https://www.scipedia.com/public/Silva_et_al_2020a
16. Song, H., Ahmad, A., Farooq, F., Ostrowski, K. A., Maślak, M., Czarnecki, S., & Aslam, F. (2021). Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. *Construction and Building Materials*, *308*, 125021. https://doi.org/10.1016/j.conbuildmat.2021.125021
17. Yeh, I.-C. (2007). *Concrete compressive strength* [Dataset]. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength
18. 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. https://doi.org/10.1016/j.cemconres.2018.09.006
19. Zhang, W., Guo, J., Ning, C., Cheng, R., & Liu, Z. (2024). Prediction of concrete compressive strength using a Deepforest-based model. *Scientific Reports*, *14*(1), 18918. https://doi.org/10.1038/s41598-024-69616-9