International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

A Study of Machine Learning Algorithms for Concrete Compressive Strength Prediction

R. Harshitha Merlin[1], Dr. D. Preethi[2]

[1]Department of Computer Science, Mount Carmel College Autonomous, Bengaluru, India

[2]Assistant Professor, Department of Computer Science, Mount Carmel College Autonomous, Bengaluru, India

Abstract: ***The toughness of concrete is judged by its compressive power, and the compressive strength is typically determined using a traditional crushing test on a concrete cylinder. To achieve proper outcomes, it is recommended that one waits 28 days before testing the cylinder. Using machine learning techniques, this process can be accelerated. This paper includes a compressive strength examination of concrete as well as the development of machine learning models to predict compressive strength using machine learning methods such as Random Forest Regression, CatBoost, Light GBM, and ANN. The efficiency of different algorithms is assessed, and the model with the least RMSE (Root Mean Squared Error) is ultimately selected to forecast concrete compressive strength.***

Keywords: artificial intelligence; machine learning; concrete; compressive strength; prediction; regression

1. Introduction

Concrete is the most often used building material on earth. It can be used to construct buildings, bridges, roads, and other infrastructures. By blending aggregates, cement, sand, water, and several additional admixtures, concrete is created homogeneous. Even with correct mixing, microcracks may form because of variations in the thermal and mechanical properties of coarse particles and cement matrix, resulting in concrete failure.[1]Concrete's compressive strength ranges from 2500 psi (17 MPa) to 11450 psi (80 MPa). Concrete must meet the requirements of NBC (National Building Code). For testing the quality of concrete, 2 kinds of tests are performed: i) Compressive strength test ii) Slump test.The ability of a material or structure to carry loads on its surface without cracking or deflection is referred to as compressive strength. In this test, a push force is applied to both faces of the concrete specimen, and the highest compression that the concrete can withstand without failing is recorded. Engineers must construct miniature concrete cylinders/cubes using various combinations of raw components and strength tests will be done for differences when each raw material is changed.[2] The preparation and testing of several prototypes takes a long time and a lot of effort. Furthermore, this system is vulnerable to human mishap.The objective of this research is to i) evaluate the best model that can be used for predicting concrete's compressive strength, thereby improving the speed of the process and reducing human-prone errors, ii) find the correlation between the variables and iii) find the most important feature among all the other features.

Concrete compressive strength is an important property that is typically evaluated after 4 weeks of normal curing. Until recently, when support vector regression (SVR), artificial neural networks (ANNs), random forest (RF), as well as other forms of computational intelligence have been effectively implemented in concrete study, concrete characteristics where typically, linear and non-linear regression methods are used to forecast.[3]

2. Literature Review

De-Cheng Feng et al presented a new intelligent approach for forecasting concrete compressive strength based on the AdaBoost algorithm in this work. [4].To facilitate a full evaluation of the concrete manufacturing process, influential aspects from five viewpoints were gathered by J. Xu *et al*. [5].In A. K. Jha *et al* study, the first eight features are elements impacting concrete compressive strength, while the final characteristic is the value of concrete compressive strength. Some of the models used are the Linear Regressor, Ridge Regressor, Lasso Regressor, Decision Tree Regressor, Random Forest Regressor, AdaBoost Regressor, and Gradient Boosting Regressor. The Gradient Boosting Regressor and Random Forest Regressor performed well in terms of RMSE and r2 in the presented dataset. [6]

S. Khursheed *et al* used machine learning algorithms such as Minimax Probability Machine Regression (MPMR), Relevance Vector Machine (RVM), Genetic Programming (GP), Emotional Neural Network (ENN), and Extreme Learning Machine (ELM) to predict the 28-day compressive strength of fly ash concrete. [7]. In A. Ahmad *et al*'s study, a traditional technique is used where bias and variance decrease for the test set are examined using K-fold cross validation. H. Song *et al* assessed the outcome of cross-validation using Mean absolute error (MAE), Mean square error (MSE), and Root mean square error (RMSE). [8]The database of rubberized concrete samples was built using a systematic collection of data from rubberized concrete testing in M. Hadzima-Nyarko *et al* research. [9]Individual (ANN) and ensemble (AdaBoost and boosting) models were used in the A. Ahmad *et al* investigation to give output, with nine input components including Na_2SiO_3, NaOH, SiO_2, Na_2O, the molarity of NaOH, and the age of the cure (CS).[10]

3. Methodology

Several machine learning techniques were used in this study to predict the compressive strength of concrete. The regression techniques used were Random Forest, Artificial

Neural Network (ANN), Light Gradient Boosting Machine (LGBM), and CatBoost.[13] [14]These many regressors are tested to see which one fits the best for calculating concrete compressive strength.

Table 1: Attributes with units

Sl.	Attributes	Units	Max	Min
1.	Cement	Kg/m^3	540	102
2.	Blast Furnace	Kg/m^3	359.4	0
3.	Fly Ash	Kg/m^3	200.1	0
4.	Water	Kg/m^3	247	121.8
5.	Superplasticizer	Kg/m^3	32.2	0
6.	Coarse Aggregate	Kg/m^3	1145	801
7.	Fine Aggregate	Kg/m^3	992.6	594
8.	Age	Days	365	1

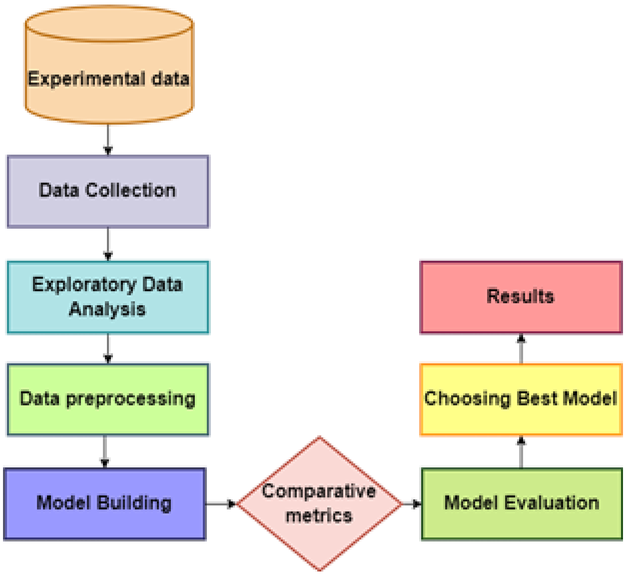

Figure 1: System Framework

3.1. Data Collection

The materials for the study were gathered from the UCI repository. There are 1031 observations in the dataset, each having 9 attributes. [15] [16]The first eight features indicate elements influencing concrete compressive strength, while the final attribute is the value of concrete compressive strength. The compressive strength of concrete is measured in mega Pascals (MPa). The eight characteristics for each observation are cement, blast furnace slag, fly ash, water, super plasticizer, coarse aggregate, fine aggregate, and concrete age. [17] The variables are all quantitative in nature. Table 1 shows the attributes along with maximum and minimum values that are used in this study.

3.2. Exploratory Data Analysis

Using the Seaborn library, a correlation matrix was plotted. Compressive Strength and Cement have a strong positive association, Super Plasticizer and Water have a significant negative association, while Super Plasticizer and Fly Ash and Fine Aggregate have positive correlations. A scatterplot from the same library was utilized. It was discovered that compressive strength rises with the amount of cement used and that concrete strength rises with decreased water usage.

3.3. Data Preprocessing

There are duplicate values in the data utilized in this study. Duplicates in the training set might cause the model to gain biases toward the repeated samples, making it difficult for the model to generalize to new data.1030 was the original number of observations. It is observed that it has 1005 values after removing duplicates. After that, the data is divided, with the features separated from the target variable. The independent variable subset is then divided into an 80 percent training set and a 20 percent test set.

3.4. Model Building

Different regressors were used to create the models. Decision tree regressor, Random Forrest Regressor, Artificial Neural Networks (ANN), Light GBM, and CatBoost Regressor are the regressors employed.

3.4.1. Random Forest Regressor

Random Forest Regression is a supervised learning technique that performs regression by using ensemble learning method. Ensemble learning takes predictions from numerous machine learning algorithms and combines them to give a more precise forecast than a single system. A Random Forest creates numerous decision trees during training and uses the mean of the classes as the tree's forecast. It excels at a wide variety of problems, particularly those with non-linear relationships.

Artificial Neural Network (ANN)

Regression ANNs predict the outcome of a variable based on its parameters. The advantage of Artificial Neural Networks over Linear Regression is that linear regression can only learn linear relationships between features and targets, and so cannot learn complex non-linear correlations.[18]

3.4.2. Light GBM (Gradient Boosting Machine)

Light GBM is a fast and scalable gradient boosting framework based on the decision tree technique, which is used in ranking, classification, and a variety of other machine learning applications. It separates the tree leaf by leaf with the best fit because it is based on decision tree algorithms.

3.4.3. CatBoost Regressor

CATBOOST is an open-source machine learning package. Catboost's capacity to handle missing data and categorical data without encoding is one of its most notable features. It's a boosting method, thus it builds trees sequentially and reduces error with each iteration.

3.5. Evaluation of Models

The RMSE, R-squared, and MAPE metrics were used to evaluate the models.

3.5.1. Root Mean Square Error (RMSE)

The average standard deviation of the differences between forecasted and measured values is expressed by the RMSE. The RMSE is a measure of prediction accuracy that combines the magnitudes of errors in forecasting different times. [19]

RMSE can be mathematically formulated as:

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

Where,
RMSE = root-mean-square error,
i = variable i
N = Number of non-missing data points
y_i = actual values
$\hat{y}_i$ = predicted values

3.5.2. R-squared

The strength of fit of a regression model is represented by R-squared, a statistical metric. [20] 1 is the optimal r-square value. When the r-square value is close to 1, the model fits better. The residual sum of squares (SSres) is compared to the total sum of squares using the R-square formula (SStot).
R-squared can be mathematically formulated as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where,
SS_{res}= residual sum of squares
SS_{tot}= total sum of squares

3.5.3. Mean Absolute Percentage Error (MAPE)

A forecasting system's accuracy is measured by its Mean Absolute Percentage Error (MAPE). It is expressed as a percentage and is calculated by subtracting actual values from real values for each time period.

MAPE can be mathematically formulated as:

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Where,
n = number of fitted points
A_t = actual value
F_t = forecast value

3.6. Feature Importance

The feature relevance according to different regressors was investigated after the errors were calculated and the best model was found.

4. Results

For the given dataset, the various models mentioned before were employed to accomplish the prediction. The models were assessed using the RMSE, R-squared, and MAPE metrics.

Table 2: Performance of different regressors

Sl.	Model	RMSE	R^2	MAPE
1.	Random Forest	5.29	0.89	0.12
2.	ANN	6.46	0.84	0.17
3.	Light GBM	4.71	0.91	0.09
4.	CatBoost	4.33	0.93	0.10

From the table above, Light GBM and CatBoost regressors have performed well.

But LightGBM seems to outdo CatBoost in terms of feature importance.

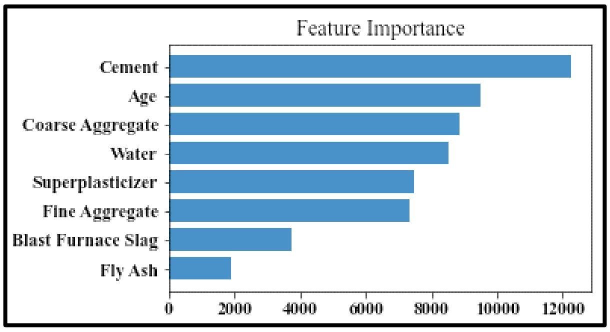

Figure 2: Feature importance by LGBM

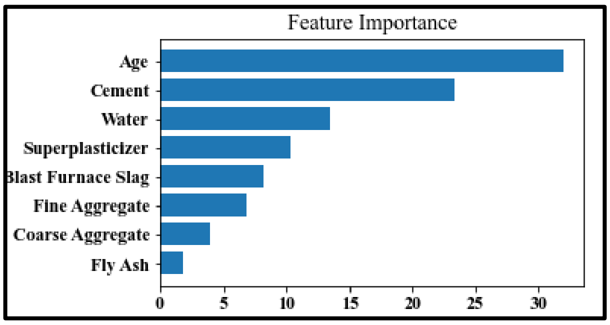

Figure 3: Feature importance by CatBoost

Ensemble learners frequently outperform traditional machine learning algorithms, and ensemble learners performed well in this experiment in terms of RMSE, R2, and MAPE. The most essential features for such algorithms are cement and age. Experiments in the laboratory have also revealed that the compressive strength of concrete improves as the number of days increases. As a consequence, machine learning models provide outcomes that are consistent with laboratory findings.

5. Conclusion

The development of algorithms for forecasting concrete compressive strength is critical. Because the compressive strength of concrete is critical to infrastructure longevity, the factors impacting concrete should be thoroughly investigated. Some elements may have been overlooked, and these traits may have an impact on concrete's compressive strength. To address these constraints, more robust algorithms can be devised. Furthermore, the value of a characteristic varies from algorithm to algorithm. Feature selective approaches, such as wrapper methods, may be utilised in the future to pick features while developing predictive models. Future research into new prediction models and analyses of the impact of elements affecting concrete compressive strength will make infrastructure construction more economical.

References

[1] M. Timur Cihan, "Prediction of Concrete Compressive Strength and Slump by Machine Learning Methods," *Advances in Civil Engineering*, vol. 2019, pp. 1–11, Nov. 2019, doi: 10.1155/2019/3069046.
[2] B. A. Salami*et al*, "Data-driven model for ternary-blend concretecompressive strength prediction using machine learning approach," *Construction and*

Building Materials, vol. 301, p. 124152, Sep. 2021, doi: 10.1016/j.conbuildmat.2021.124152.

[3] F. Deng*et al*, "Compressive strength prediction of recycled concrete based on deep learning," *Construction and Building Materials*, vol. 175, pp. 562–569, Jun. 2018, doi: 10.1016/j.conbuildmat.2018.04.169.

[4] D.-C. Feng *et al.*, "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach," *Construction and Building Materials*, vol. 230, p. 117000, Jan.2020,doi:10.1016/j.conbuildmat.2019.117000.

[5] J. Xu *et al*, "Comprehensive Machine Learning-Based Model for Predicting Compressive Strength of Ready-Mix Concrete," *Materials*, vol. 14, no. 5, p. 1068, Jan. 2021, doi: 10.3390/ma14051068.

[6] A.K.Jha *et al*, "Evaluation of Factors Affecting Compressive Strength of Concrete using Machine Learning," *IEEE Explore*,2020,p.70,doi: 10.1109/ 9213199.

[7] S. Khursheed*et al*, "Compressive strength prediction of fly ash concrete by using machine learning techniques," *Innovative Infrastructure Solutions*, vol. 6, no. 3, May 2021, doi: 10.1007/s41062-021-00506-z.

[8] A. Ahmad *et al*, "Comparative Study of Supervised Machine Learning Algorithms for Predicting the Compressive Strength of Concrete at High Temperature," *Materials*, vol. 14, no. 15, p. 4222, Jul. 2021, doi: 10.3390/ma14154222.

[9] H. Song *et al.*, "Predicting the compressive strength of concrete with fly ash admixture usingmachinelearningalgorithms," *Construction and Building Materials*, vol. 308, p. 125021, Nov. 2021, doi: 10.1016/j.conbuildmat.2021.125021.

[10] M. Hadzima-Nyarko *et al*, "Machine learning approaches for estimation of compressive strength of concrete," *The European Physical Journal Plus*, vol. 135, no. 8, Aug. 2020, doi: 10.1140/epjp/s13360-020-00703-2.

[11] A. Ahmad *et al.*, "Prediction of Geopolymer Concrete Compressive Strength Using Novel Machine Learning Algorithms," *Polymers*, vol. 13, no. 19, p. 3389, Oct. 2021, doi: 10.3390/polym13193389.

[12] B. A. Young *et al*, "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*, vol. 115, pp. 379–388, Jan. 2019, doi: 10.1016/j.cemconres.2018.09.006.

[13] N. Kumar and V. Naranje, "Prediction of Cement Strength using Machine Learning Approach," *Amity International Conference on Artificial Intelligence (AICAI)*, 2019, pp. 239-243, doi: 10.1109/AICAI.2019.8701370.

[14] M. Azimi-Pour*et al*, "Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete," *Construction and Building Materials*, vol. 230, p. 117021, Jan. 2020,doi:10.1016/j.conbuildmat.2019.11702.

[15] O. R. Abuodeh*et al*, "Assessment of compressive strength of Ultra-high Performance Concrete using deep machine learning techniques," *Applied Soft Computing*, vol. 95, p. 106552, Oct. 2020, doi: 10.1016/j.asoc.2020.106552.

[16] P. G. Asteris*et al*, "Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models," *Cement and Concrete Research*, vol. 145, p. 106449, Jul. 2021, doi: 10.1016/j.cemconres.2021.106449.

[17] M.-C. Kang*et al*, "Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete," *Construction and Building Materials*, vol. 266, p. 121117, Jan. 2021,doi:10.1016/j.conbuildmat.2020.121117.

[18] M. R. Kaloop*et al*, "Compressive strength prediction of high-performance concrete using gradient tree boosting machine," *Construction and Building Materials*,vol.264,p.120198,Dec.2020,doi:10.1016/j.conbuildmat.2020.120198.

[19] A. Iqtidar *et al.*, "Prediction of Compressive Strength of Rice Husk Ash Concrete through Different Machine Learning Processes," *Crystals*, vol. 11, no. 4, p. 352, Apr. 2021, doi: 10.3390/cryst11040352.

[20] A. K. Al-Shamiri*et al*, "Non-Tuned Machine Learning Approach for Predicting the Compressive Strength of High-Performance Concrete," *Materials*, vol. 13, no. 5, p. 1023, Feb. 2020, doi: 10.3390/ma13051023.