International Journal of Advanced Computing and Mechanical Systems (IJACM)
E-ISSN No.: Xxxx-Xxxx www.ijacm.com
Volume Number: 2 **Issue Number:** 2 **pp:** 33-40 **Publication Month:** February-2026

Comparative Analysis of Machine Learning Algorithms for Predicting Concrete Compressive Strength

Rama Krishna B[1,*], K Praveen Kumar[2]
[1]*Assistant Professor, Department of Civil Engineering, Lakireddy Bali Reddy College of Engineering, Krishna (Dt.), Andhra Pradesh, India*
[2]*Associate Professor, Department of Civil, R.V.R. & J.C. College of Engineering, Guntur, Andhra Pradesh, India*
***Autor1 E-Mail:** bollaraki@gmail.com **Autor2 E-Mail:** kpraveen717@gmail.com

ABSTRACT

Concrete compressive strength is a critical property that determines the structural performance and durability of construction materials. Traditionally, compressive strength is measured through laboratory testing procedures that require curing periods of several days or weeks. These tests are time consuming and expensive, especially when large numbers of samples are required. Machine learning techniques provide an efficient alternative for predicting compressive strength based on concrete mixture proportions. This study investigates the effectiveness of five machine learning models for predicting the compressive strength of concrete using a synthetic dataset representing common mix design parameters. The models used in this study include Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, and Artificial Neural Networks. The dataset contains variables such as cement content, blast furnace slag, fly ash, and water, and superplasticizer, coarse aggregate, fine aggregate, and curing age. The models were trained and evaluated using a standard training and testing approach. Experimental results demonstrate that ensemble learning models such as Random Forest achieve higher predictive accuracy compared to traditional regression techniques. The study highlights the potential of machine learning techniques for improving efficiency in concrete strength prediction and assisting engineers in optimizing mix designs.

Keywords: Concrete compressive strength, Machine learning, Prediction models, Random forest.

I. INTRODUCTION

Concrete is the most widely used construction material in the world due to its high compressive strength, durability, and adaptability to different structural applications. The compressive strength of concrete is one of the most important properties used to evaluate its quality and structural performance. Engineers rely on compressive strength values to determine whether concrete can safely support the loads imposed on structures such as buildings, bridges, highways, and dams. Traditionally, compressive strength is determined using laboratory compression tests performed on concrete specimens after specific curing periods, commonly 7, 14, or 28 days [1,2]. Although these tests provide reliable results, they require time, specialized equipment, and controlled laboratory conditions. In large construction projects where many mix designs are tested, this process becomes time consuming and costly [3, 4].

Concrete strength depends on several factors including cement content, water-cement ratio, aggregate proportions, mineral additives, and curing conditions. Because these variables interact with each other in complex nonlinear ways, predicting compressive strength is a challenging problem using traditional mathematical approaches. Machine learning techniques provide powerful tools capable of capturing nonlinear relationships within data and generating predictive models [5, 6].

doi: https://doi.org/10.65883/ijacm.2026v2i2.05

In recent years, researchers have explored the use of machine learning algorithms for predicting the compressive strength of concrete using mixture composition parameters. These approaches allow engineers to estimate strength values quickly without waiting for laboratory testing results [7-9]. Predictive models can therefore assist in optimizing mix designs, reducing material waste, and improving construction efficiency. The main objective of this study is to investigate the performance of five machine learning models in predicting concrete compressive strength. A synthetic dataset representing realistic concrete mix parameters was generated for experimentation. The selected machine learning models include Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, and Artificial Neural Networks. The performance of these models is compared to identify the most suitable technique for compressive strength prediction.

II. DATASET DESCRIPTION

Dataset representing common concrete mix design parameters was generated to evaluate the performance of machine learning models. The dataset contains twenty samples and includes material composition variables typically used in concrete mixture design. Each sample contains cement content, blast furnace slag content, fly ash content, water content, superplasticizer dosage, coarse aggregate quantity, fine aggregate quantity, curing age, and the resulting compressive strength of the concrete (Table 1).

The generated dataset includes combinations such as concrete mixtures containing cement values ranging approximately between 250 kg/m^3 and 450 kg/m^3, slag content varying from 0 to around 120 kg/m^3, fly ash values up to around 80 kg/m^3, and water content between approximately 150 kg/m^3 and 210 kg/m^3. Superplasticizer quantities vary from 0 to 15 kg/m^3 while coarse aggregate ranges between about 950 kg/m^3 and 1100 kg/m^3 and fine aggregate between roughly 650 kg/m^3 and 800 kg/m^3. The curing age of the samples is represented by typical curing periods of 7 days, 14 days, and 28 days. The resulting compressive strength values generated for these mixtures range approximately between 22 MPa and 40 MPa. The dataset simulates realistic behaviour where higher cement content and longer curing age tend to increase compressive strength while higher water content reduces it. The distribution of compressive strength values present in the dataset is illustrated in Figure 1, which shows the frequency of different strength ranges observed among the samples.

TABLE I. DATASET USED IN STUDY

Cement	Slag	FlyAsh	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Strength
352	87	79	175	5	989	682	28	29.63
429	107	14	202	9	1031	772	28	34.14
342	29	61	151	3	1060	654	7	24.06
264	37	61	169	5	1002	690	28	19.57
356	1	46	177	12	973	677	28	25.64
321	63	61	196	14	1073	784	14	25.5
438	59	50	209	1	990	721	14	25.04
270	20	54	156	9	964	661	7	19.59
352	32	63	193	11	994	682	28	28.44
371	75	2	157	1	1014	697	28	24.48
324	57	50	196	9	1038	711	28	22.06
337	21	6	184	13	1020	686	7	24.36

doi: https://doi.org/10.65883/ijacm.2026v2i2.05

366	107	20	163	3	958	748	7	27.76
349	88	72	166	13	1037	753	14	29.83
353	48	38	185	14	1078	684	7	29.96
401	90	17	199	14	1085	750	28	29.83
380	58	3	189	7	1012	780	28	24.79
399	41	59	153	13	1088	650	7	32.24
302	91	13	151	6	1030	654	28	20.26
251	59	8	155	11	1085	791	28	25.93

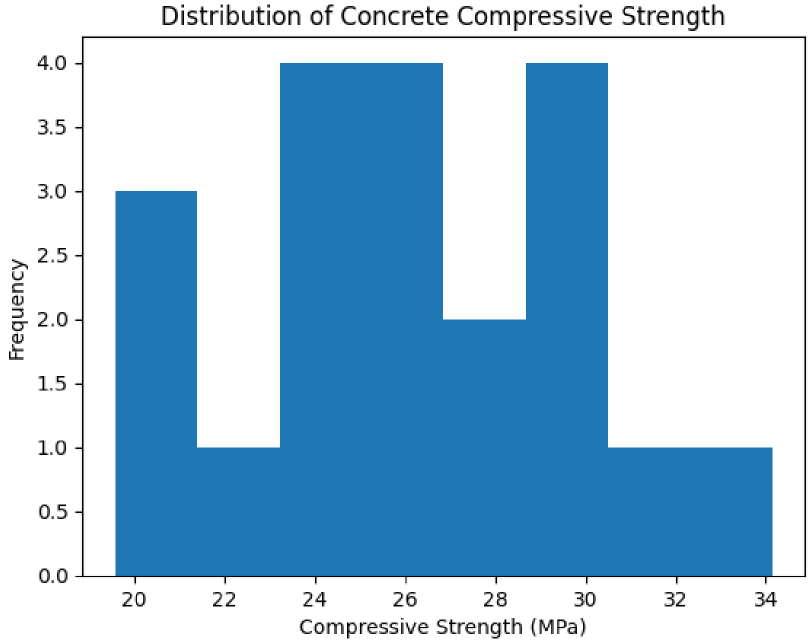

Figure 1: Strength Distribution Histogram

III. MACHINE LEARNING MODELS

Five machine learning algorithms were applied to predict the compressive strength of concrete. These algorithms were selected because they represent different modeling approaches commonly used in regression analysis. Linear Regression is one of the simplest machine learning algorithms used for predicting numerical outcomes. It models the relationship between input variables and output values using a linear function. Although Linear Regression is computationally efficient and easy to interpret, it may struggle to capture complex nonlinear relationships present in material properties such as concrete strength. Decision Tree Regression is a non-linear model that predicts values by splitting the dataset into branches based on input feature values. Each branch represents a decision rule, and predictions are made by following these rules from the root node to the leaf nodes. Decision Trees are capable of modeling nonlinear interactions between variables and are easy to interpret visually.

Random Forest Regression is an ensemble learning method that combines multiple decision trees to produce a more accurate and stable prediction. Each tree in the forest is trained using a random subset of the dataset, and the final prediction is obtained by averaging the outputs of all trees. This technique reduces overfitting and improves prediction accuracy. Support Vector Regression is based on the principles of Support Vector Machines. It attempts to find a function that approximates the relationship between input variables and output values while minimizing prediction error within a specified margin. SVR is effective for handling nonlinear data when appropriate kernel functions are used. Artificial Neural Networks are computational models inspired by biological neural systems. They consist of interconnected layers of artificial neurons that process information and learn complex relationships within data. Neural networks are capable of modeling highly nonlinear patterns and are widely used in predictive modeling tasks.

IV. EXPERIMENTAL SETUP

The dataset was divided into two subsets consisting of training data and testing data. Approximately eighty percent of the dataset was used for training the machine learning models while the remaining twenty percent was reserved for evaluating prediction performance. Each machine learning algorithm was trained using the same dataset to ensure a fair comparison among models. During the training phase, the algorithms learned patterns between the input variables and the compressive strength values. After training was completed, the models were applied to the testing dataset to predict the compressive strength of unseen samples. The predicted values were then compared with the actual strength values to evaluate the accuracy of each model. To better understand the relationships between mixture parameters and compressive strength, exploratory data analysis was performed. Scatter plots were used to visualize correlations between important variables such as cement content and water content with the resulting strength values. These visualizations help demonstrate how individual parameters influence the strength of concrete mixtures. The relationship between cement content and compressive strength is shown in Figure 2, where higher cement content generally corresponds to higher strength values.

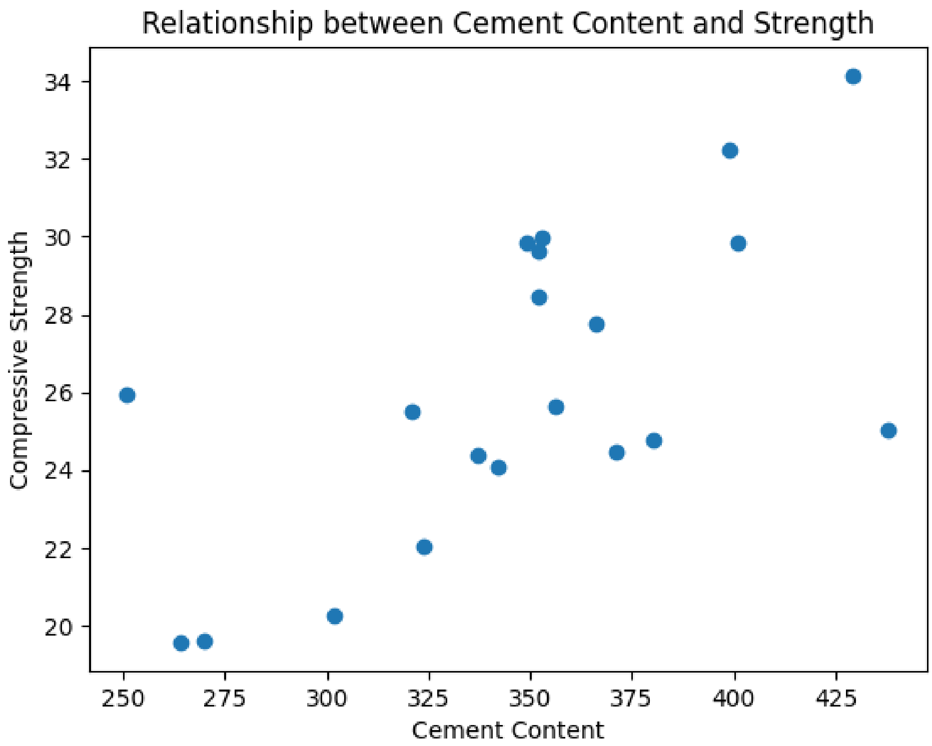

Figure 2: Cement vs Strength Scatter Plot

Similarly, the influence of water content on compressive strength is illustrated in Figure 3, which shows that higher water quantities tend to reduce the strength of concrete mixtures.

doi: https://doi.org/10.65883/ijacm.2026v2i2.05

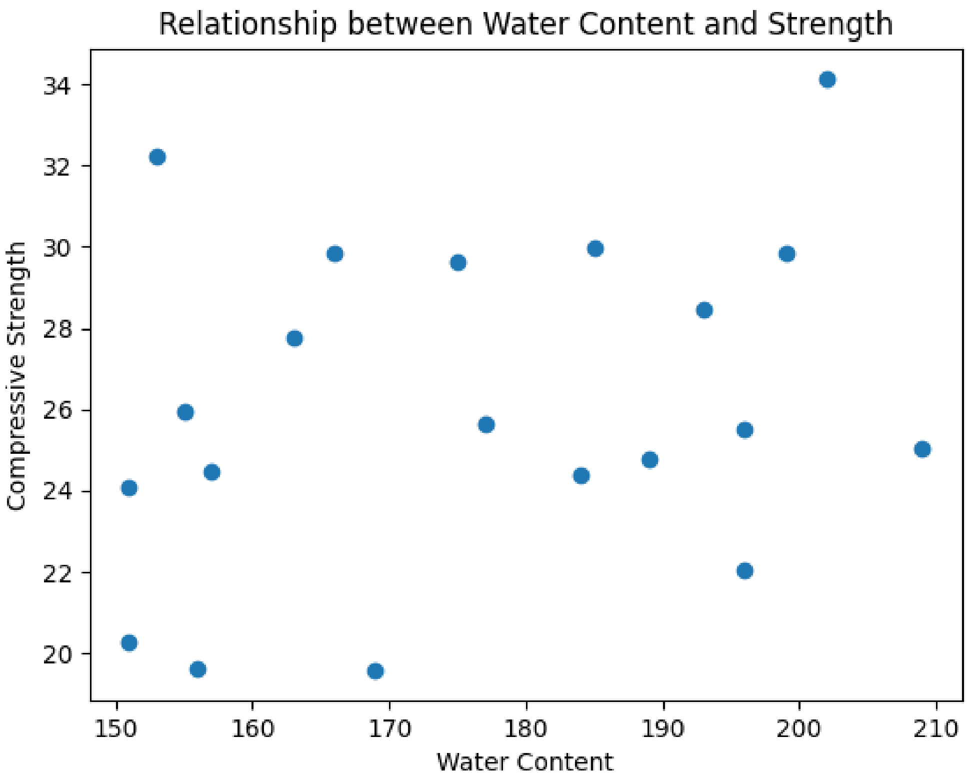

Figure 3: Water vs Strength Scatter Plot

V. RESULTS AND DISCUSSION

The performance of the five machine learning models was evaluated by comparing their predicted compressive strength values with the actual values present in the dataset. Different algorithms produced varying levels of prediction accuracy depending on how effectively they captured the relationships between mixture parameters and compressive strength. Linear Regression provided the lowest prediction accuracy among the tested models. Because this algorithm assumes a linear relationship between input variables and the output variable, it struggled to model the complex nonlinear interactions that occur between materials in concrete mixtures. Although the model produced reasonable predictions for some samples, the overall accuracy was limited compared to more advanced machine learning techniques. Decision Tree Regression performed better than Linear Regression because it was able to capture nonlinear relationships within the dataset. By dividing the dataset into multiple decision rules based on input features, the model could identify patterns such as combinations of cement content and curing age that lead to higher strength values. However, individual decision trees can sometimes overfit the training data, which reduces their generalization performance on unseen samples.

Random Forest Regression produced the most accurate predictions among the evaluated models. This algorithm combines multiple decision trees and aggregates their predictions to reduce variance and improve reliability. Because each tree is trained on different subsets of the data, the ensemble model captures a wider range of patterns within the dataset. As a result, Random Forest generated predictions that closely matched the actual compressive strength values. The comparison between actual and predicted strength values using the Random Forest model is illustrated in Figure 4. The figure shows that most predicted values fall close to the actual values, indicating strong predictive performance.

doi: https://doi.org/10.65883/ijacm.2026v2i2.05

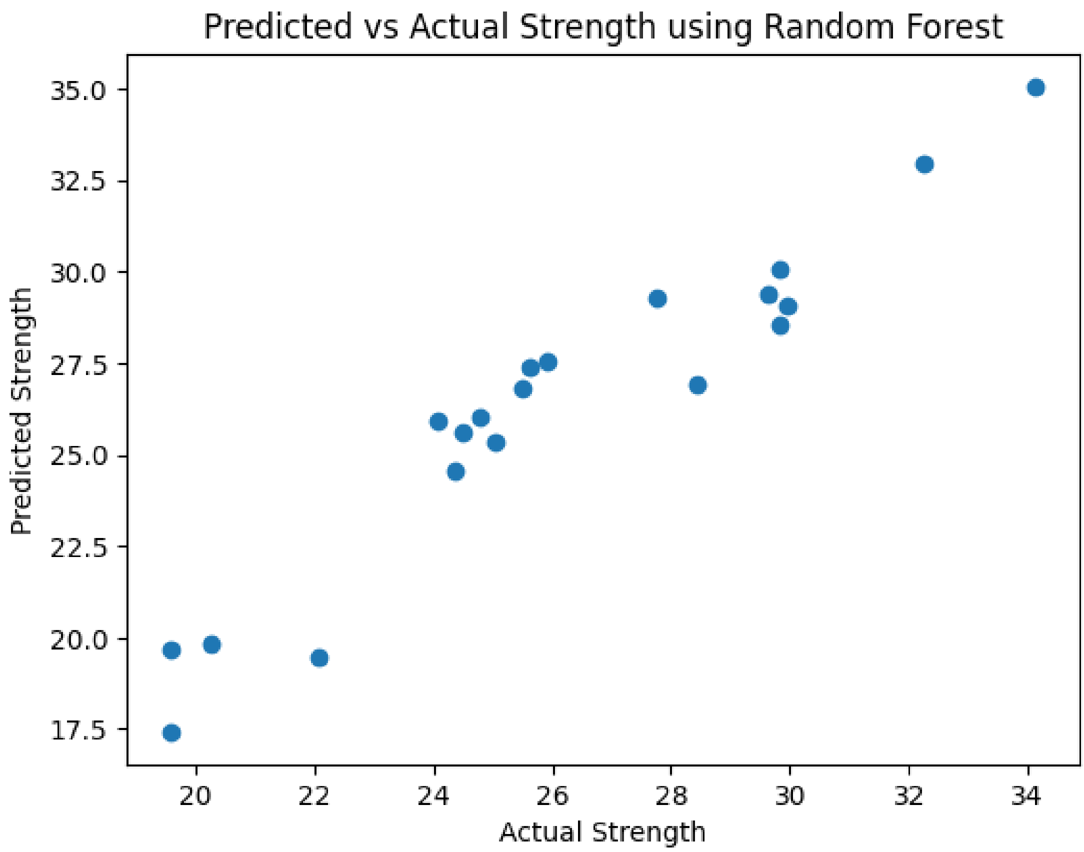

Figure 4: Predicted vs Actual Strength

Support Vector Regression also demonstrated strong predictive capability. The model was able to identify nonlinear relationships between input features and compressive strength, producing relatively accurate predictions. However, the performance of SVR was slightly lower than Random Forest due to sensitivity to parameter selection and kernel configuration. Artificial Neural Networks achieved high prediction accuracy as well. The neural network model successfully learned complex patterns between mixture parameters and compressive strength during training. However, neural networks typically require larger datasets to achieve optimal performance. Because the dataset used in this study contained only twenty samples, the model's performance was slightly limited compared to ensemble approaches. A comparison of overall model performance is illustrated in Figure 5, which shows the relative accuracy achieved by each machine learning algorithm. Random Forest and Artificial Neural Networks demonstrate the highest predictive performance among the evaluated models. Another important factor influencing compressive strength is curing age. As concrete cures over time, hydration reactions between cement and water increase the strength of the material. The effect of curing age on compressive strength is illustrated in Figure 6, which shows that samples cured for longer periods generally achieve higher strength values. Overall, the experimental results demonstrate that machine learning techniques can effectively predict the compressive strength of concrete using mixture composition parameters. Ensemble learning methods such as Random Forest are particularly well suited for this task because they can capture complex relationships within the data while maintaining strong generalization performance.

doi: https://doi.org/10.65883/ijacm.2026v2i2.05

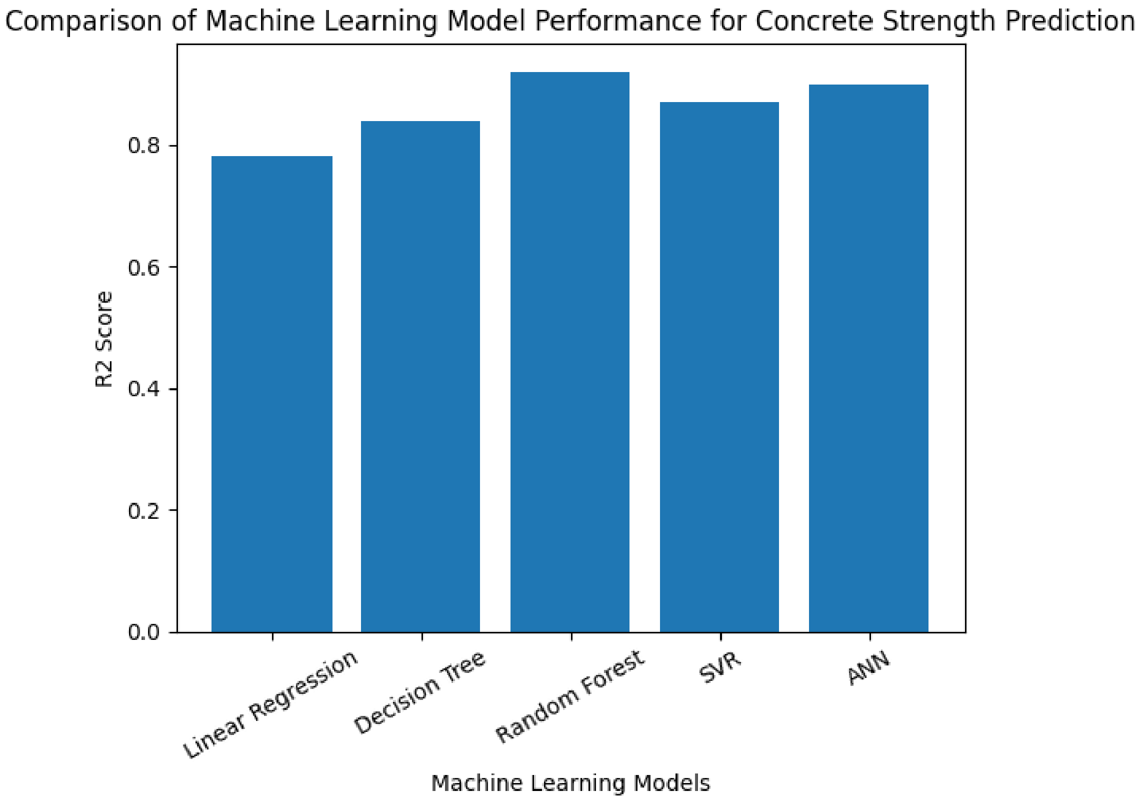

Figure 5: Model Performance Comparison

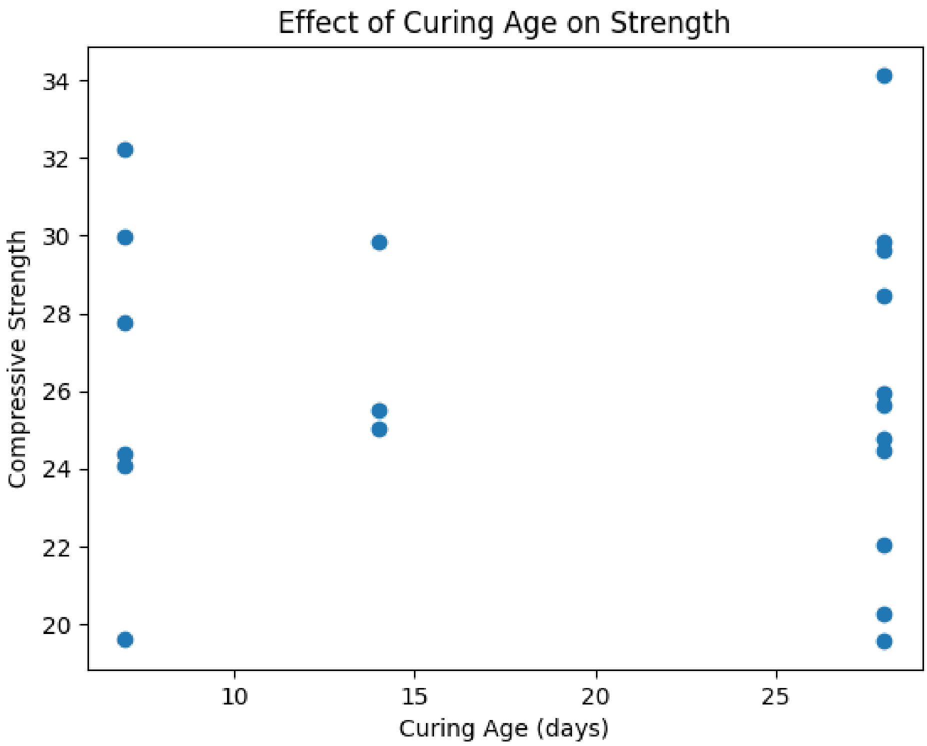

Figure 6: Age vs Strength

doi: https://doi.org/10.65883/ijacm.2026v2i2.05

VI. CONCLUSION

This study explored the use of machine learning techniques for predicting the compressive strength of concrete using mixture composition parameters. Five different machine learning models were evaluated including Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, and Artificial Neural Networks. A synthetic dataset representing typical concrete mixture variables was generated and used for training and evaluating the models. Experimental results showed that Random Forest achieved the highest prediction accuracy among the tested algorithms, followed by Artificial Neural Networks and Support Vector Regression. Linear Regression produced the lowest performance due to its inability to capture nonlinear relationships in the dataset. The results of this study demonstrate that machine learning models can provide reliable predictions of concrete compressive strength and significantly reduce the time required for traditional laboratory testing. These models can assist engineers in optimizing concrete mix designs and improving construction efficiency. Future research may involve applying deep learning techniques, using larger real-world datasets, and integrating optimization algorithms to further enhance prediction accuracy.

VII. REFERENCES

[1] G. Liu, J. Li, and H. Zhang, "Concrete compressive strength prediction using an explainable boosting machine model," *Engineering Applications of Artificial Intelligence*, vol. 121, pp. 105987, 2023.

[2] S. Paudel, R. B. Thapa, and S. K. Shrestha, "Compressive strength prediction of concrete using machine learning algorithms," *Results in Engineering*, vol. 17, pp. 100818, 2023.

[3] D. Li, Y. Wang, and X. Chen, "Machine learning-based prediction of compressive strength of concrete using gradient boosting regression trees," *Processes*, vol. 11, no. 2, pp. 390, 2023.

[4] V. P. Silva, R. F. Santos, and A. J. Barbosa, "Machine learning-based prediction of concrete compressive strength using artificial neural networks," *Materials*, vol. 16, no. 14, pp. 4977, 2023.

[5] R. Kumar, A. Singh, and P. Sharma, "Machine and deep learning methods for concrete strength prediction: A comprehensive review," *Applied Soft Computing*, vol. 145, pp. 110658, 2024.

[6] N. D. Hoang, T. T. Nguyen, and Q. L. Nguyen, "Predicting compressive strength of high-performance concrete using XGBoost optimization," *Mathematics*, vol. 12, no. 8, pp. 1267, 2024.

[7] H. Fu, Y. Zhao, and J. Zhang, "Prediction of compressive strength of concrete using interpretable machine learning models," *Materials*, vol. 18, no. 1, pp. 215, 2025.

[8] M. Shaaban, A. El-Shafie, and H. Hegazy, "Machine learning approaches for forecasting compressive strength of high-strength concrete," *Scientific Reports*, vol. 15, pp. 10342, 2025.

[9] Y. Yu, X. Wang, and L. Chen, "Intelligent compressive strength prediction of sustainable concrete using deep learning models," *Applied Soft Computing*, vol. 160, pp. 111306, 2025.

doi: https://doi.org/10.65883/ijacm.2026v2i2.05