

"AI-Powered Predictive Analytics for Concrete Compressive Strength with Material Impact Interpretation"

Kalpesh Wani¹, Prashant Shimpi²

PG Student, Department of Computer Engineering, Godavari College of Engineering, Jalgaon, India¹

Assistant Professor, Department of Computer Engineering, Godavari College of Engineering, Jalgaon, India²

Abstract: Because it directly affects the quality, longevity, and safety of structures, concrete compressive strength is a crucial factor in the construction sector. Conventional laboratory-based strength testing is expensive, time-consuming, and labor-intensive. This study suggests an AI-powered predictive analytics framework for evaluating concrete compressive strength utilizing machine learning approaches in order to get over these restrictions. A concrete compressive strength dataset that includes mix design factors such as cement, water, fly ash, blast furnace slag, superplasticizer, coarse aggregate, fine aggregate, and curing age is used in this work. To find the most effective model for precise strength prediction, several machine learning regression models are put into practice and assessed. Evaluation metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R2 are used to evaluate the models' performance. Additionally, explainable AI algorithms are used for material impact interpretation in order to assess how each input information contributes to the prediction of compressive strength. The experimental findings show that sophisticated ensemble-based regression models outperform conventional regression techniques in terms of prediction accuracy. This study provides engineers with an efficient and comprehensible AI solution for strength assessment and concrete mix design optimization.

Keywords: Concrete Compressive Strength, Artificial Intelligence, Machine Learning, Predictive Analytics, Regression Models, Explainable AI, SHAP.

I. INTRODUCTION

The longevity, safety, and load-bearing capacity of structures are directly impacted by concrete's compressive strength, making it one of the most crucial quality characteristics in the construction sector. Compressive strength is often measured using laboratory testing techniques including cube and cylinder tests, which are precise but expensive, time-consuming, and reliant on curing time. With the quick development of data-driven technologies and artificial intelligence (AI), machine learning-based predictive models have emerged as an effective alternative for estimating concrete strength using mix design parameters such as cement content, water, fly ash, slag, aggregates, superplasticizer, and curing age. AI-powered predictive analytics can efficiently learn complex non-linear relationships between these parameters and strength values, enabling faster decision-making, reducing experimental workload, and improving construction planning. Moreover, AI models can provide deeper insights into the impact of each material component on compressive strength, supporting engineers in optimizing mix proportions and enhancing quality control. In order to make the prediction process more dependable and practically relevant, this research focuses on creating an AI-based framework for precise concrete compressive strength prediction coupled with material impact interpretation.

II. LITERATURE SURVEY

Numerous researchers have investigated machine learning (ML) and artificial intelligence (AI) methods for forecasting concrete compressive strength using mix design data. Using ensemble models like AdaBoost, Random Forest, and XGBoost in addition to non-ensemble methods like MLP and Linear Regression, Paudel et al. (2023) found that ensemble approaches improve prediction accuracy, but there are still significant issues with limited dataset generalization and lack of real-time validation [1]. Similar to this, Zviazhynski et al. (2025) suggested a Random Forest-based model in conjunction with uncertainty optimization and just-in-time learning, which enhanced prediction performance using fresh-state measurements. However, the study was constrained by a small dataset size and inadequate generalization across various concrete types [2]. Using experimental concrete datasets, Beskopylny et al. (2022) assessed CatBoost, KNN, and SVR models and found that CatBoost performed better with less prediction errors. Despite this, the authors pointed out that generalization under different construction conditions is limited [3]. Additionally, Liu (2022) employed XGBoost,

Random Forest, and SVR models to predict the strength of high-performance concrete (HPC), with XGBoost yielding the highest accuracy. Nevertheless, the study found that the absence of real-world validation and dataset size restrictions were major limitations [4]. Together, these findings show that boosting and ensemble-based AI models perform better than conventional regression methods; yet, there are still significant research gaps due to problems including low dataset diversity, inadequate external validation, and real-time applicability.

III. RESEARCH METHODOLOGY

This study offers a predictive analytics framework driven by AI that uses machine learning regression approaches to estimate the compressive strength of concrete. Eight input parameters (cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, and curing age) and one output parameter (concrete compressive strength in MPa) make up the dataset used in this study, which comprises 1030 samples. To guarantee quality, the dataset is first examined for missing values and cleaned. The dataset is separated into training and testing sets for model building and validation following preprocessing. To create predictive models and evaluate their effectiveness, several AI and machine learning methods are used.

A. Performance Measures for Prediction

A. Predictive Performance Measures The following measures are used to assess regression-based AI models' performance:

- (i) Mean Absolute Error (MAE): Calculates the average absolute difference between actual and forecasted values.
- (ii) Mean Squared Error (MSE): Calculates the average squared deviation between the actual and forecasted values.
- (iii) Root Mean Squared Error (RMSE): Often used to quantify forecast error in MPa, RMSE is the square root of MSE.
- (iv) The coefficient of determination, or R² score, indicates how well the model accounts for the variation in compressive strength values. Better prediction accuracy is indicated by a higher R².

B. Methods Used in Machine Learning:

In this work, the following machine learning regression approaches are used:

- (i) Linear Regression: This baseline regression method models linear connections between compressive strength and input factors.
- (ii) Support Vector Regression (SVR): This potent regression method handles non-linear patterns in concrete mix datasets by utilizing kernel functions.
- (iii) Random Forest Regressor: An ensemble learning technique that reduces overfitting and increases prediction accuracy by combining several decision trees.
- (iv) XGBoost Regressor: An ensemble technique based on gradient boosting that offers quick training times and great accuracy for structured datasets.

C. Artificial Intelligence Techniques

AI-based methods are also used to improve prediction performance and interpretability:

- (i) Feature Importance Analysis: This method identifies the mix factors that have the greatest impact on compressive strength.
- (ii) Explainable AI (SHAP Analysis): This technique improves the transparency of AI models by interpreting each input feature's contribution to strength prediction.
- (iii) Hyperparameter tuning: To increase prediction accuracy and lower error rates, model parameters are optimized.
- (iv) Model Comparison Framework: The optimal AI model for predicting concrete strength is found by methodically comparing several models.

Ultimately, the model that performs the best is chosen based on assessment metrics and verified by comparing the actual and anticipated compressive strength values.

IV. RESULT AND DISCUSSION

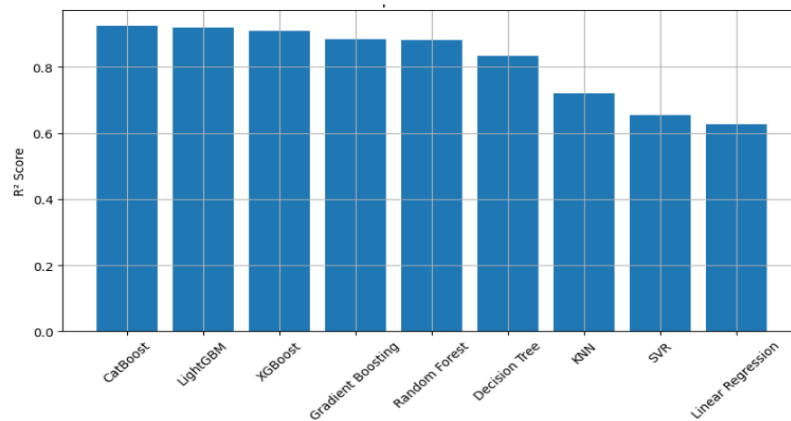
In this work, various machine learning regression models were used to estimate concrete compressive strength utilizing an AI-based predictive analytics framework. MAE, MSE, RMSE, and R² score were used to assess the models. The findings demonstrate that ensemble AI models outperformed conventional techniques in terms of prediction accuracy. With R² = 0.924821 and RMSE = 4.401342, CatBoost outperformed all other algorithms, followed by LightGBM and XGBoost. The most important variables influencing compressive strength are curing age, cement content, and water quantity, according to Explainable AI (SHAP) analysis, which was also used to explain the material impact. Overall, the findings show that AI-based boosting models offer precise and comprehensible predictions for estimating concrete strength.

	Model	MAE	MSE	RMSE	R2
8	CatBoost	3.058018	19.371808	4.401342	0.924821
7	LightGBM	2.961945	20.675881	4.547074	0.919760
6	XGBoost	3.123326	23.194762	4.816094	0.909985
5	Gradient Boosting	4.135014	30.176898	5.493350	0.882889
4	Random Forest	3.788964	30.700321	5.540787	0.880857
3	Decision Tree	4.293786	42.581023	6.525414	0.834750
2	KNN	6.768476	72.448762	8.511684	0.718839
1	SVR	7.514947	88.959415	9.431830	0.654764
0	Linear Regression	7.745559	95.970940	9.796476	0.627553

Fig.1 compares the R² scores of different AI/ML models used for concrete strength prediction.

A higher R² score indicates better prediction accuracy.

In this study, **CatBoost achieved the highest R² score (0.924821)**, making it the best-performing model.



Best AI Model: CatBoost

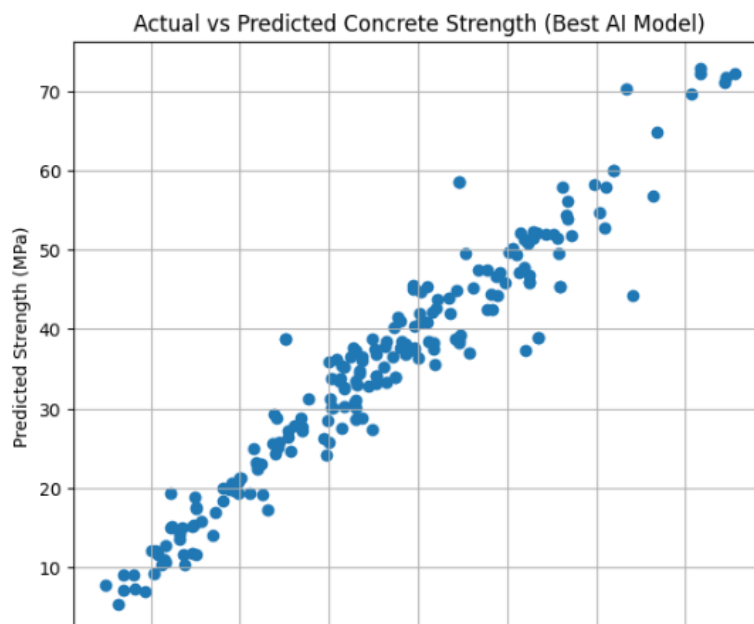


Fig.2 shows the Actual vs Predicted compressive strength values using the best AI model.

If the predicted points are close to the actual values, it indicates high accuracy. In this study, the CatBoost model predictions closely match the actual strength values, proving good performance.

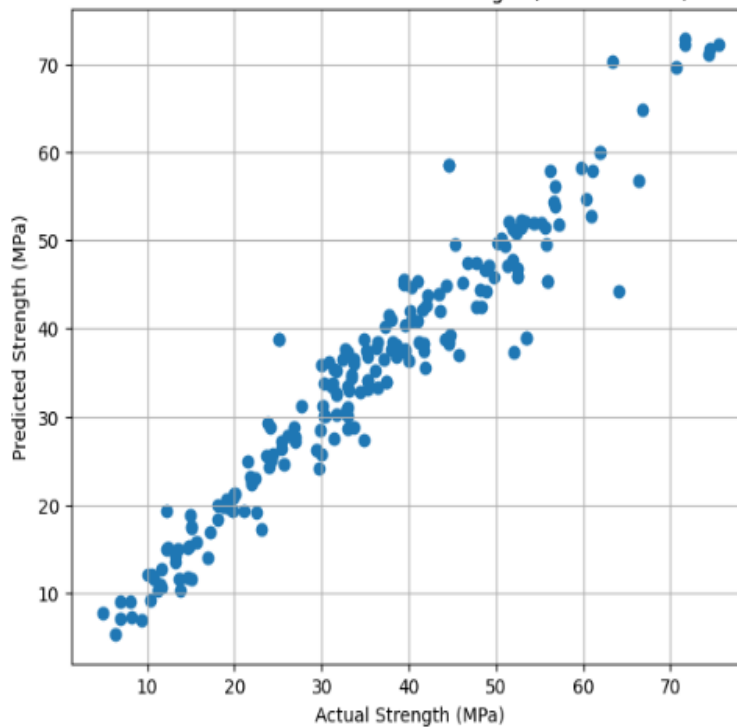
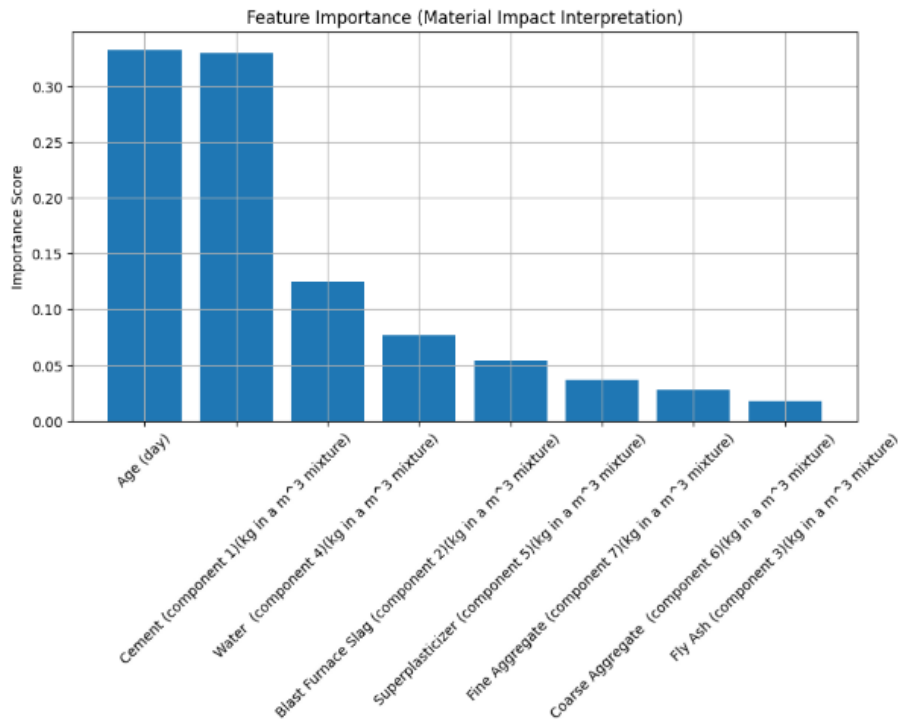


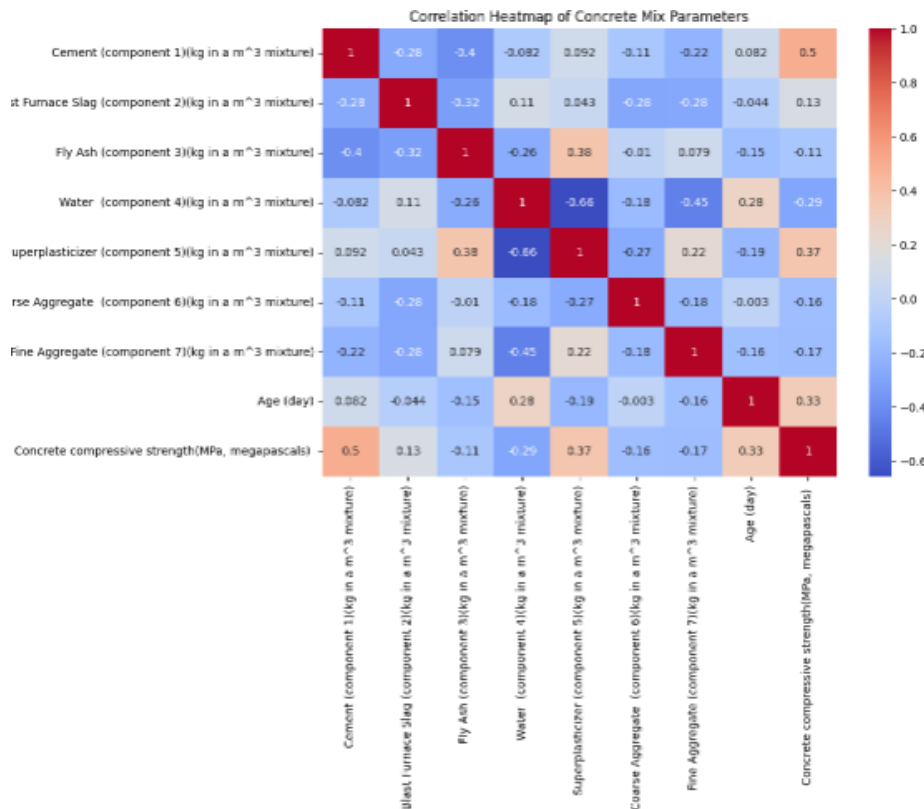
Fig 3 This plot is a **Residual vs Fitted (Predicted) plot with a LOWESS trend line**. It shows the relationship between predicted values and residual errors to evaluate model performance. The LOWESS line helps identify any systematic patterns in the residuals, indicating possible bias or non-linearity in the model. If the trend line is approximately flat around zero, it suggests that the model is well-fitted and generalizes effectively.

		Feature Importance
7	Age (day)	0.332129
0	Cement (component 1)(kg in a m ³ mixture)	0.329845
3	Water (component 4)(kg in a m ³ mixture)	0.124917
1	Blast Furnace Slag (component 2)(kg in a m ³ m...	0.076409
4	Superplasticizer (component 5)(kg in a m ³ mix...	0.053926
6	Fine Aggregate (component 7)(kg in a m ³ mixture)	0.036556
5	Coarse Aggregate (component 6)(kg in a m ³ mi...	0.027922
2	Fly Ash (component 3)(kg in a m ³ mixture)	0.018296

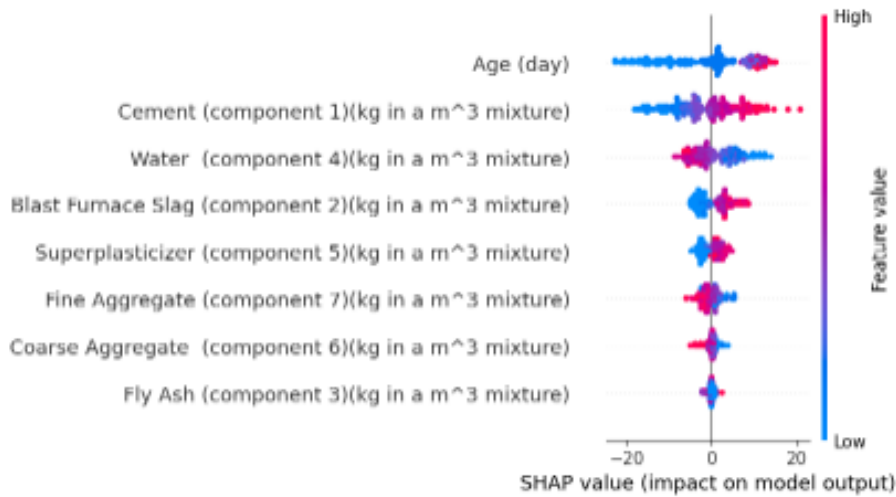
Fig 4 “The feature importance analysis indicates the relative contribution of each input variable to the model’s prediction, helping identify the most influential parameters affecting concrete compressive strength.”



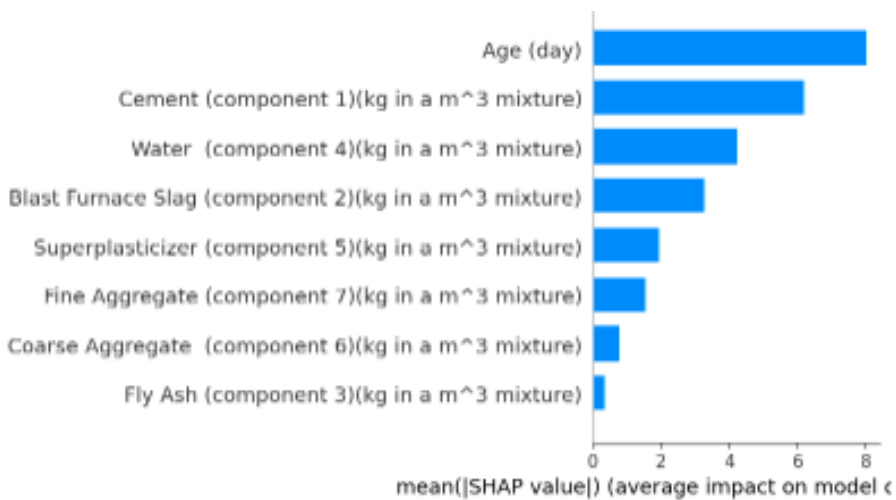
The bar chart represents feature importance derived from the Random Forest model. It illustrates the relative contribution of each input variable to the prediction of concrete compressive strength. Higher values indicate stronger influence of the corresponding material properties on the model output.



The correlation heatmap visualizes the relationships among concrete mix parameters. It helps identify positive and negative dependencies between input variables and compressive strength. This analysis assists in understanding the influence of material composition on concrete performance.



“SHAP explains not only which features are important, but also how they affect concrete strength prediction.”



The SHAP bar plot represents the global feature importance by averaging the absolute SHAP values for each input variable. It identifies the most influential parameters affecting concrete compressive strength prediction, providing a more reliable interpretation compared to traditional feature importance methods.

```

===== FINAL AI MODEL RESULT =====
Best Model: CatBoost
MAE: 3.058017718587141
RMSE: 4.401341646888823
R2 Score: 0.9248213219097271
    
```

The final model evaluation was performed using MAE, RMSE, and R² score to assess prediction accuracy and model reliability. Lower error values and higher R² indicate better model performance.

Model summary:

The best-performing model for forecasting concrete compressive strength was shown to be the CatBoost model. With R² = 0.9248, the model demonstrated strong predictive performance, indicating high accuracy and good generalization on the test dataset. The model's dependability for strength estimation is demonstrated by the error metrics MAE = 3.0580 and RMSE = 4.4013, which verify that the prediction error is minimal.



V. CONCLUSION

This study introduced a predictive analytics framework driven by AI that uses various machine learning regression models to estimate the compressive strength of concrete. The experimental findings demonstrated that ensemble-based AI models outperformed conventional regression methods in terms of accuracy. With the highest R^2 score of 0.924821 and the lowest RMSE of 4.401342, CatBoost outperformed all other models tested, making it the most dependable model for strength prediction. Furthermore, Explainable AI (SHAP) analysis aided in the interpretation of material impact and demonstrated that the most significant elements influencing compressive strength are water quantity, cement content, and curing age. All things considered, the suggested method offers an accurate and comprehensible AI solution that may help engineers with construction quality control and concrete mix design optimization.

VI. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Department of Computer Engineering, Godavari College of Engineering, Jalgaon, for providing the necessary facilities and support to carry out this research work. We are thankful to our guide and mentor **Prof. Prashant Shimpi** for his valuable guidance, continuous encouragement, and technical support throughout the project. We also acknowledge the support of all faculty members and friends who directly or indirectly contributed to the successful completion of this research.

REFERENCES

- [1]. "Prediction of Compressive Strength of Concrete using Machine Learning Techniques," S. Paudel et al., 2023.
- [2]. "Predicting Concrete Strength using Machine Learning and Fresh-State Measurements," B. Zviazhynski et al., 2025.
- [3]. "Concrete Strength Prediction using Machine Learning Methods," A. Beskopylny et al., 2022.
- [4]. Liu, "Machine Learning-Based High-Performance Concrete Strength Prediction," 2022.
- [5]. I. Yeh, "Modeling of Strength of High Performance Concrete Using Artificial Neural Networks," *Cement and Concrete Research*, vol. 28, no. 12, pp. 1797–1808, 1998.
- [6]. The "Concrete Compressive Strength Dataset" is available in the UCI Machine Learning Repository.
- [7]. XGBoost: A Scalable Tree Boosting System, T. Chen and C. Guestrin, *Proceedings of the 22nd ACM SIGKDD*, pp. 785–794, 2016.