

XGBoost-Based Predictive Framework for High-Performance Concrete Compressive Strength with SHAP-Guided Explainability

Amarsinh B. Landage¹, Aniket P. Bhusari², Naaz B. Ni³, Mahvish N. Mirkar⁴
¹Assistant Professor, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India, ^{2,3,4}Research Scholar, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India

Abstract— Accurate estimation of concrete compressive strength is central to structural safety, construction quality control, and sustainable material use. Conventional destructive testing mandates curing durations of 7 to 28 days, creating costly delays and obstacles to rapid mix design iteration.

This work presents a machine learning framework to estimate the compressive strength of high-performance concrete (HPC) from eight mix-design parameters and curing age. Six algorithms were evaluated — Linear Regression, Support Vector Regression (SVR), Random Forest, Gradient Boosting, eXtreme Gradient Boosting (XGBoost), and Deep Neural Networks (DNN) — on 1,030 experimental specimens from Yeh's (1998) benchmark dataset. XGBoost consistently outperformed all alternatives on R^2 , RMSE, and MAE. SHAP interpretability analysis was applied to identify dominant features and validate established concrete chemistry principles, including Abrams' Law.

The final model was deployed as a cloud-hosted web application returning predictions in under 100 ms, providing a practical non-destructive alternative for real-time construction site decision-making.

Keywords— Compressive Strength Prediction, XGBoost, Gradient Boosting, SHAP Analysis, High-Performance Concrete, Mix Design Optimization, Feature Importance.

I. INTRODUCTION

Concrete is the most widely used structural material worldwide, with applications spanning foundations, pavements, bridges, and high-rise buildings. Its compressive strength is the principal design parameter governing load-bearing performance and must be reliably quantified for structural safety and economic efficiency. Conventional strength determination relies on destructive compression testing of specimens cured over defined intervals, typically 7, 14, or 28 days. While scientifically sound, this process delays formwork removal, increases testing expenditure, and cannot assess concrete already embedded in a completed structure.

Machine learning (ML) offers an effective alternative by learning complex, high-dimensional nonlinear relationships between mix parameters and resulting strength from experimental data, enabling real-time prediction without physical specimen preparation. However, the opacity of ensemble methods has restricted adoption in safety-critical civil engineering contexts. This study addresses both concerns by integrating a systematic multi-model comparison with SHAP-based interpretability, producing a transparent,

accurate, and practically deployable strength prediction framework for HPC.

II. DATASET AND EXPLORATORY ANALYSIS

The study used Yeh's (1998) benchmark HPC dataset [6], available from the UCI Machine Learning Repository, comprising 1,030 records spanning diverse concrete mix formulations. Eight continuous input features and one continuous output variable (compressive strength) describe each record. Table I summarizes the dataset statistics.

TABLE I. DESCRIPTIVE STATISTICS OF DATASET FEATURES [6]

Feature	Min	Max	Mean	Std	Unit
Cement	102.0	540.0	281.2	104.5	kg/m ³
BF Slag	0.0	359.4	73.9	86.2	kg/m ³
Fly Ash	0.0	200.1	54.2	63.9	kg/m ³
Water	121.8	247.0	181.6	21.4	kg/m ³
Superplasticizer	0.0	32.2	6.2	8.3	kg/m ³
Coarse Aggregate	801.0	1145.0	972.9	77.8	kg/m ³
Fine Aggregate	594.0	992.6	773.6	80.9	kg/m ³
Curing Age	1.0	365.0	45.7	63.4	days
Comp. Strength	2.33	82.60	35.3	16.8	MPa

The feature set captures all principal physicochemical parameters affecting concrete performance. Cement (102–540 kg/m³) is the primary binder; blast furnace slag and fly ash act as supplementary cementitious materials. Water content (121.8–247.0 kg/m³) governs hydration and capillary porosity. Superplasticizer enables workability at reduced water-to-cement (w/c) ratios. Coarse and fine aggregates form the inert skeletal volume. Curing age (1–365 days) reflects temporal strength gain. No missing values were identified. Correlation analysis confirmed expected trends: cement content and curing age correlated positively with strength; water content showed a clear inverse relationship aligned with Abrams' Law. No physically unrealistic outliers were detected.

III. METHODOLOGY

A. Data Preprocessing and Feature Engineering

The dataset was partitioned into training/validation (85%, $n = 875$) and held-out test (15%, $n = 155$) subsets. Z-score standardization ($z = (x - \mu)/\sigma$) via StandardScaler was

applied exclusively to training data to prevent leakage. An engineered water-to-cement (w/c) ratio feature was added to explicitly encode Abrams' Law. Pairwise multicollinearity checks confirmed no feature pair exceeded $|r| > 0.95$, so no dimensionality reduction was required. Five-fold stratified cross-validation with GridSearchCV was used for hyperparameter tuning, maximizing cross-validation R^2 as the selection criterion.

B. Machine Learning Model Implementation

Six algorithms spanning the complexity spectrum were assessed. Ordinary Least Squares (OLS) regression provided a linear baseline. SVR with RBF kernel was optimized over $C \in \{0.1, 1, 10, 100\}$ and $\gamma \in \{0.001, 0.01, 0.1, \text{'scale'}\}$. Random Forest used bootstrap aggregation over $n_{\text{estimators}} \in \{100, 200, 500\}$ and $\text{max_depth} \in \{10, 15, 20\}$. Gradient Boosting used $n_{\text{estimators}} \in \{100, 200\}$, $\text{learning_rate} \in \{0.05, 0.1\}$, $\text{max_depth} \in \{3, 5\}$. XGBoost extended gradient boosting with second-order Taylor loss approximation, L1/L2 regularization, and column subsampling ($\text{max_depth} \in \{3, 5, 7\}$; $\text{learning_rate} \in \{0.01, 0.05, 0.1\}$). A DNN with three fully connected layers (128–64–32 neurons), ReLU activation, dropout, and early stopping completed the comparison. Model ranking used a composite score: R^2 (40%), RMSE (40%), MAE (10%), CV standard deviation (10%).

C. SHAP Interpretability Framework

To lift the black-box limitation of the optimal ensemble model, SHAP (SHapley Additive exPlanations) values were computed using cooperative game theory (Shapley, 1953). Each input feature receives a Shapley value representing its exact marginal contribution to a given prediction, averaged over all possible feature permutation subsets. SHAP satisfies local accuracy, consistency, and missingness, qualifying it as the gold standard for ensemble interpretability. Summary plots visualize contribution magnitude and direction across all test samples, with color encoding representing feature value magnitude (red = high, blue = low).

D. Prediction System Architecture

The operational system consists of three layers: a User Input Interface with out-of-range validation; a Processing Core performing full preprocessing and XGBoost inference; and an Output Layer returning predicted strength (MPa) with bootstrap-estimated 95% confidence intervals. Implemented in Python using Streamlit, the trained model is serialized via Pickle and exposed through /predict and /model_info REST API endpoints.

IV. RESULTS AND DISCUSSION

A. Comparative Model Performance

Table II presents performance metrics for all six models on the held-out test set, ranked by descending R^2 . Results confirm that predictive accuracy scales with algorithmic capacity to capture nonlinear physicochemical relationships.

TABLE II. COMPARATIVE PERFORMANCE METRICS ON HELD-OUT TEST SET

Model	R^2	RMSE (MPa)	MAE (MPa)	Time (s)
XGBoost	0.9099	4.7499	3.3963	0.81
Grad. Boosting	0.9079	4.8005	3.6269	0.61

Random Forest	0.8807	5.4648	3.9008	1.53
DNN	0.8550	6.0238	4.0533	4.02
SVR	0.8354	6.4185	4.4472	0.09
Linear Regression	0.6041	9.9547	7.8809	0.03

XGBoost achieved the highest R^2 of 0.9099, accounting for approximately 91% of total strength variance, with the lowest RMSE (4.7499 MPa) and MAE (3.3963 MPa). Its marginal but consistent advantage over Gradient Boosting ($\Delta R^2 = 0.0020$, $\Delta \text{RMSE} = 0.051$ MPa) reflects the added benefit of second-order gradient optimization, built-in regularization, and feature subsampling. The linear baseline yielded an RMSE more than double that of XGBoost (9.9547 MPa), confirming that HPC strength development is governed by nonlinear interactions structurally incompatible with linear statistical models.

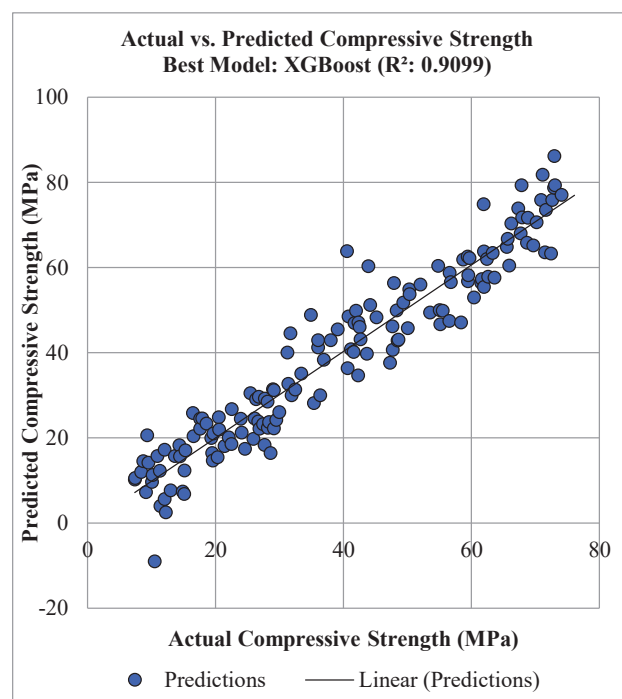


Fig. 1. Actual vs. Predicted Compressive Strength (XGBoost, $R^2 = 0.9099$)

Scatter analysis of XGBoost predictions revealed a tight cluster around the perfect prediction line ($y = x$) across the full operational range of 2.33–82.60 MPa, with no systematic directional bias and normally distributed residuals centered at zero, indicating well-calibrated model performance. A slight increase in prediction uncertainty was observed above 70 MPa, attributable to the relatively sparse representation of ultra-high-strength specimens in the training data.

B. SHAP Feature Importance and Physical Validation

Table III presents the SHAP-derived feature importance rankings alongside material science interpretations, demonstrating that XGBoost independently recovered well-established concrete chemistry principles from experimental data without any domain knowledge encoded in the model architecture.

TABLE III. SHAP FEATURE IMPORTANCE RANKINGS

Rank	Feature	Material Science Significance
1	Curing Age	Primary +ve driver; C-S-H gel accumulation, capillary porosity reduction
2	Cement Content	Strong +ve; governs total hydration potential and C-S-H volume
3	Water Content	Dominant -ve; validates Abrams' Law — excess water forms load-reducing capillary pores
4	BF Slag	Moderate +ve; latent hydraulic reactivity beyond 28 days
5	Superplasticizer	Indirect enabler; reduces w/c at target workability
6	Fly Ash	Low-moderate; pozzolanic contribution at extended ages
7-8	Aggregates	Lowest; inert volumetric fillers, secondary role

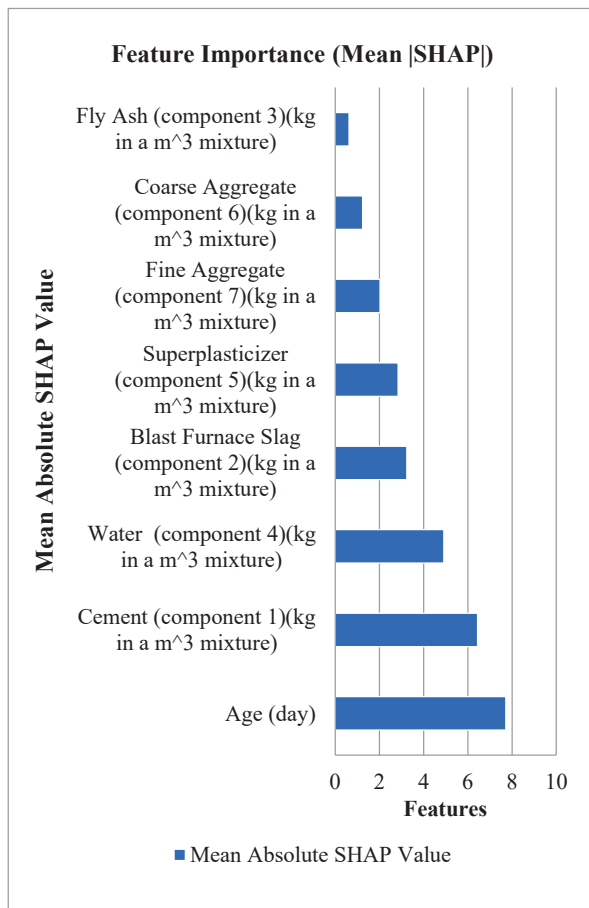


Fig. 2. SHAP Feature Importance (Mean |SHAP| values)

Curing age ranked first, reflecting self-limiting diffusion-controlled C-S-H gel kinetics that progressively densify paste microstructure over time. Cement content ranked second, with the model correctly identifying diminishing strength returns at high dosages from heat of hydration and shrinkage effects. Water content produced the most physically significant negative SHAP values, independently validating Abrams' Law (1918). Excess water beyond hydration requirements ($w/c \approx 0.25$) occupies paste volume, evaporating to form capillary pores that reduce the

effective load-bearing cross-section and create crack initiation pathways. This autonomous rediscovery of a century-old empirical law from unlabeled patterns provides compelling evidence that model predictions are physically grounded rather than artifacts of overfitting. Blast furnace slag, superplasticizer, and fly ash ranked as secondary contributors consistent with their delayed or indirect physicochemical roles. Aggregate components showed the lowest influence, as expected for inert volumetric fillers.

C. Engineering Implications and System Demonstration

The alignment between SHAP-attributed importances and established concrete science underpins practical deployability. As a forward predictor for sustainable mix designs incorporating industrial by-products, the system can substantially reduce destructive quality control testing, accelerating project delivery and lowering costs. The model processes queries in under 100 ms on standard cloud infrastructure, enabling seamless integration into real-time construction workflows.

The framework was operationalized as ConcretIQ v2.0, a cloud-hosted Streamlit web application. The interface accepts all eight mix parameters across two input tabs (Binders & Water; Aggregates & Admixtures), computes the w/c ratio in real time with bounds flagging, and presents results as predicted strength in MPa with grade classification per IS 456:2000, a mix composition chart, and a strength development projection from 1 to 365 days on a logarithmic time scale.

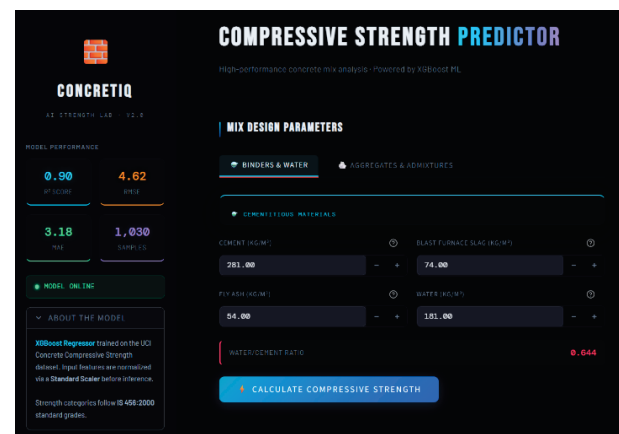


Fig. 3 (a) Binders & Water input tab

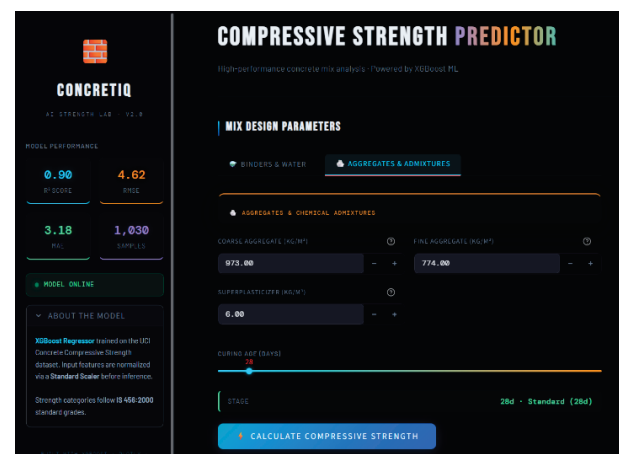


Fig. 3 (b) Aggregates & Admixtures tab

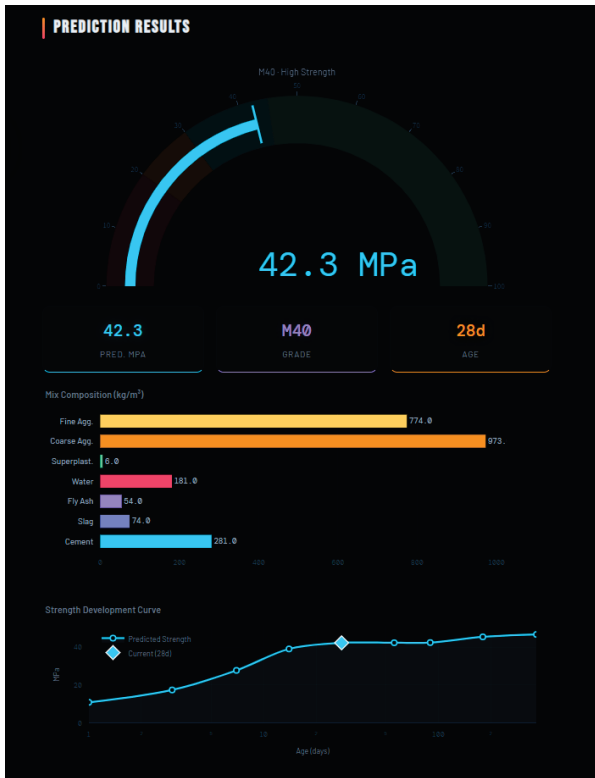


Fig. 3 (c) Prediction results panel

Fig. 3. ConcretIQ v2.0 Deployed Web Application Interface

A disclaimer note at the base of the results panel explicitly states that predictions are indicative and that laboratory testing remains mandatory for structural design, ensuring responsible AI deployment in safety-critical contexts. All six validation criteria were fulfilled, with sub-100 ms inference latency confirmed on cloud infrastructure.

V. CONCLUSION

This study developed, validated, and deployed a machine learning framework for HPC compressive strength prediction, jointly addressing predictive accuracy and engineering interpretability. Table IV summarizes the principal quantitative outcomes.

TABLE IV. SUMMARY OF KEY RESEARCH OUTCOMES

Parameter	Quantified Outcome
Optimal Algorithm	XGBoost — superior across all three metrics
R ² (XGBoost)	0.9099 (91% of total variance explained)
RMSE (XGBoost)	4.7499 MPa (5–8% of mid-range strength)
MAE (XGBoost)	3.3963 MPa (±5 MPa engineering tolerance)
Linear Baseline RMSE	9.9547 MPa (confirms nonlinear nature)
Primary SHAP Finding	Age & cement dominate; water validates Abrams' Law
Inference Latency	Sub-100 ms — compatible with real-time site use

Four principal conclusions were drawn. First, HPC strength development is empirically confirmed as

fundamentally nonlinear: Linear Regression's R² of 0.6041 and RMSE of 9.9547 MPa demonstrate the structural inadequacy of linear statistical approaches. Second, XGBoost delivers the optimal solution for tabular concrete data, with superior generalization from second-order gradient approximation, regularization, and column subsampling at 0.81 s training time. Third, SHAP analysis conclusively establishes scientific grounding: curing age and cement content dominate as primary strength drivers; water content independently validates Abrams' Law; supplementary cementitious materials and admixtures are correctly ranked as secondary contributors. Fourth, the deployed ConcretIQ v2.0 framework achieves sub-100 ms inference compatible with real-time construction decision support.

Future directions include integration of IoT sensor inputs (curing temperature, humidity), dataset expansion to geopolymers and recycled aggregate concrete, containerized enterprise deployment (Docker / AWS / Google Cloud), and BIM platform integration for automated structural design workflows.

REFERENCES

- [1] W. B. Chaabene, M. Flah, and M. L. Nehdi, "Machine learning prediction of mechanical properties of concrete: Critical review," *Constr. Build. Mater.*, vol. 260, p. 119889, 2020.
- [2] O. O. Omotayo, C. Arum, and C. M. Ikumapayi, "Assessment of machine learning methods for concrete compressive strength prediction," *J. Soft Comput. Civ. Eng.*, vol. 8, no. 4, pp. 116–140, 2024.
- [3] Y. Gamil, "Machine learning in concrete technology: A review," *Front. Built Environ.*, vol. 9, p. 1145591, 2023.
- [4] P. G. Asteris et al., "Predicting concrete compressive strength using hybrid ensembling of surrogate ML models," *Cem. Concr. Res.*, vol. 145, p. 106449, 2021.
- [5] Q. Han, C. Gui, J. Xu, and G. Lacidogna, "A generalized method to predict HPC compressive strength by improved random forest," *Constr. Build. Mater.*, vol. 226, pp. 734–742, 2019.
- [6] I. C. Yeh, "Modeling of strength of high-performance concrete using artificial neural networks," *Cem. Concr. Res.*, vol. 28, no. 12, pp. 1797–1808, 1998.
- [7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD*, 2016, pp. 785–794.
- [8] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [9] A. Ahmad et al., "Prediction of compressive strength of fly ash concrete using individual and ensemble algorithms," *Materials*, vol. 14, no. 4, p. 794, 2021.
- [10] P. L. Ng and Y. Ding, "Machine learning prediction and SHAP analysis of concrete strength," *J. Civ. Eng. Urban Planning*, vol. 7, no. 2, pp. 122–128, 2020.
- [11] M. A. DeRousseau, J. R. Kasprzyk, and W. V. Srubar, "Computational design optimization of concrete mixtures: A review," *Cem. Concr. Res.*, vol. 109, pp. 42–53, 2019.
- [12] I. B. Mustapha et al., "Comparative analysis of gradient-boosting ensembles for quaternary blend concrete strength," *Int. J. Concr. Struct. Mater.*, vol. 18, no. 1, p. 20, 2024.
- [13] L. Tang, "Machine learning-based prediction of concrete compressive strength and interpretability analysis," *J. Civ. Eng. Urban Planning*, vol. 7, no. 2, pp. 122–138, 2025.
- [14] R. Cook et al., "Prediction of concrete compressive strength: Critical comparison of hybrid vs. standalone ML models," *J. Mater. Civ. Eng.*, vol. 31, no. 12, p. 04019255, 2019.
- [15] G. A. Lyngdoh et al., "Prediction of concrete strengths enabled by missing data imputation and interpretable ML," *Cem. Concr. Compos.*, vol. 128, p. 104414, 2022.