

ARTICLE

# Machine Learning for Compressive Strength Prediction of 3D-Printed Concrete: Feature Engineering, Statistically Validated Model Selection, and Applicability Boundaries

Jia Chen<sup>1</sup>, Zhicheng Liao<sup>1</sup>, Mengdi Hou<sup>2</sup> and Jianbo Huang<sup>1,3,\*</sup>

<sup>1</sup>School of Computer Applications, Guilin University of Technology, Guilin, China

<sup>2</sup>Guangxi Key Laboratory of Machine Vision and Intelligent Control, Wuzhou University, Wuzhou, China

<sup>3</sup>Centrale Méditerranée, Technopôle de Château-Gombert, 38 rue Frédéric Joliot-Curie, Marseille, France

\*Corresponding Author: Jianbo Huang. Email: [huang.jianbo@glut.edu.cn](mailto:huang.jianbo@glut.edu.cn)

Received: 16 May 2026; Accepted: 21 July 2026; Published: 28 August 2026

**ABSTRACT:** Extrusion-based 3D-printed concrete (3DPC) imposes a dual constraint on mix design: fresh-state printability and hardened compressive strength must both be maintained within a narrow water-to-binder window, making data-driven prediction tools essential for reducing experimental iteration. This study evaluates 20 regression algorithms on 254 experimental records spanning plain printable mortars to high-fibre reinforced composites (CS: 11.1–189.0 MPa). Four physically motivated composite variables encoding cement blend potency, cumulative supplementary cementitious material (SCM) substitution, fibre volumetric stiffness, and water-to-sand ratio are constructed; Boruta-based selection retains 11 of 17 candidate features. CatBoost achieves the highest 30-run mean performance ( $R^2 = 0.8968 \pm 0.0505$ ; single-run metrics on the seed-42 partition: RMSE = 8.82 MPa, MAPE = 9.68%) under frequentist Bonferroni correction ( $\alpha^* = 0.0026$ ) and Bayesian Savage-Dickey comparison; XGBoost is the only alternative not statistically separated from CatBoost after correction ( $p = 0.085$ ;  $BF_{10} = 9.23$  indicates only moderate evidence for a small advantage). Multi-method interpretability analysis identifies the water-to-binder ratio as the dominant predictor with an ALE effect range of 57.57 MPa; local interpretable model-agnostic explanations reveal a sign reversal of this effect across the strength spectrum, capturing the printability–strength coupling specific to layer-by-layer deposition. For practical mix design,  $W/B \leq 0.28$  consistently projects CS above 80 MPa within the compiled dataset; the 95% prediction interval half-width of 12 MPa is recommended as a design safety margin. Trained model is well-calibrated for OPC/SAC-based 3DPC within the compiled W/B window (0.15–0.65), achieving MAPE  $\leq 9.68\%$ . Cross-system transfer to 262 ECC/SHCC specimens yields a 12.0 MPa systematic overestimation attributable to material-system differences rather than model overfitting, delineating the applicability boundary for deployment. A graphical user interface implementing the optimized model is released to support mix design without programming expertise. Source codes are publicly available at <https://github.com/lucassivan/ML-3DPC>.

**KEYWORDS:** 3D-printed concrete; machine learning; compressive strength prediction; feature engineering; uncertainty quantification; external validation; interpretability

## 1 Introduction

Extrusion-based additive manufacturing of concrete, commonly referred to as three-dimensional printed concrete (3DPC), has attracted sustained research interest over the past decade owing to its capacity to fabricate structurally complex elements without formwork, reduce construction waste by up to 60%, and substantially cut labor costs relative to conventional cast-in-place methods [1–4]. Large-scale deployments

now encompass residential buildings exceeding 1000 m<sup>2</sup>, long-span pedestrian and vehicular bridges, bespoke architectural installations, and load-bearing wall structures with systematically characterized compressive and shear capacity [4–6], collectively demonstrating that 3DPC has progressed from laboratory curiosity to field-deployable technology [6,7]. Compressive strength (CS) governs the structural load-bearing capacity of printed elements and is accordingly the primary design target in mix development. However, achieving reliable CS in 3DPC is complicated by a printability constraint absent in conventional concrete: the fresh mixture must simultaneously sustain continuous extrusion under pump pressure and retain sufficient green strength to support each successive deposited layer without deformation [2,8]. This dual requirement compresses the workable water-to-binder (W/B) range to a narrow band, within which small compositional perturbations produce disproportionate consequences for both extrudability and hardened strength [9,10].

The mix design space for 3DPC is substantially more complex than that of ordinary structural concrete. Beyond the standard W/B ratio and aggregate proportions, printable cementitious composites routinely incorporate sulfoaluminate cement or blended ordinary Portland cement (OPC) for early-strength control, supplementary cementitious materials (SCMs) including fly ash, ground granulated blast-furnace slag (GGBS), and silica fume to manage rheology and pozzolanic reactivity, water-reducing admixtures (WRA) to extend the printability window, and short-cut fibres spanning wide modulus and aspect-ratio ranges to mitigate the anisotropic brittleness of layer-by-layer deposition [2,7]. These constituents interact through mechanisms that govern printability and hardened performance simultaneously: pozzolanic substitution refines long-term pore structure and alters thixotropic recovery; fibre incorporation modifies fresh-state yield stress and hardened crack-bridging capacity; W/B is bounded below by extrudability limits and above by buildability requirements [9,10]. Navigating this multi-variable space through trial-and-error experimentation is time-intensive and resource-costly, motivating data-driven prediction tools [11,12].

Gradient boosting algorithms have emerged as the leading family for 3DPC CS prediction across independent benchmarks, consistently outperforming linear and instance-based alternatives [13–15]. Li et al. [14] assembled 254 CS records spanning plain mortars to fibre-reinforced composites and benchmarked four algorithms, finding CatBoost superior at  $R^2 = 0.929$ ; SHapley Additive exPlanations (SHAP) and partial dependence plots identified W/B and OPC content as primary predictors. Uddin et al. [15] evaluated six algorithms on 299 records, with XGBoost, LightGBM, and CatBoost reaching  $R^2$  up to 0.98; SHAP confirmed W/B dominance. Alyami et al. [13] extended the comparison to seven algorithms including gene expression programming, achieving correlation coefficients above 0.986. Zhang et al. [16] independently confirmed random forest superiority on a separate 3DPC dataset ( $R^2 = 0.96$ – $0.98$ ). Neural architectures and hybrid optimization approaches have also been explored: Iqbal et al. [17] reported CNN at  $R^2 = 0.95$  on 200 records; Khodadadi et al. [18] combined metaheuristic weight optimization with data augmentation; Katlav and Turk [19] achieved  $R^2 = 0.951$  with gradient boosting paired with metaheuristic solvers; Liu et al. [20] applied CNN-LSTM with particle swarm optimization; and Izadgoshasb et al. [21] and Mütteveli Özkan and Aldemir [22] applied ANN-based and finite-element-augmented approaches, respectively. Machine learning (ML) has been extended to rheological properties [10], interlayer and reinforcement bond strength [23–26], eco-friendly aggregate substitutions [11], and geopolymer 3DPC [27], confirming that the same mix-design feature space governs multiple performance targets across binder systems [12,28–30]. Interpretability has been dominated by SHAP across independent datasets [13,15,31–33]; Arif et al. [32] supplemented SHAP with partial dependence plots and individual conditional expectation curves, Uddin et al. [31] derived GEP-based symbolic equations, and Albostami et al. [33] paired GEP sensitivity analysis with SHAP; more recently, Abbas and Alsaif [34] applied explainable data-driven modeling to interpret nonlinear synergies among binder components and proportions for 3DPC mix design. Accumulated local effects (ALE) and Local Interpretable Model-agnostic Explanations (LIME) have not been applied in prior independent 3DPC

CS prediction studies. On uncertainty, Ma et al. [24] provided distributional bond-strength predictions and Wang et al. [26] augmented training data with deep generative adversarial networks, yet no study has compared multiple uncertainty quantification methods on the same CS model. Published models are evaluated exclusively on internal hold-out sets from the same experimental population; cross-system external validation has not been attempted [13–15].

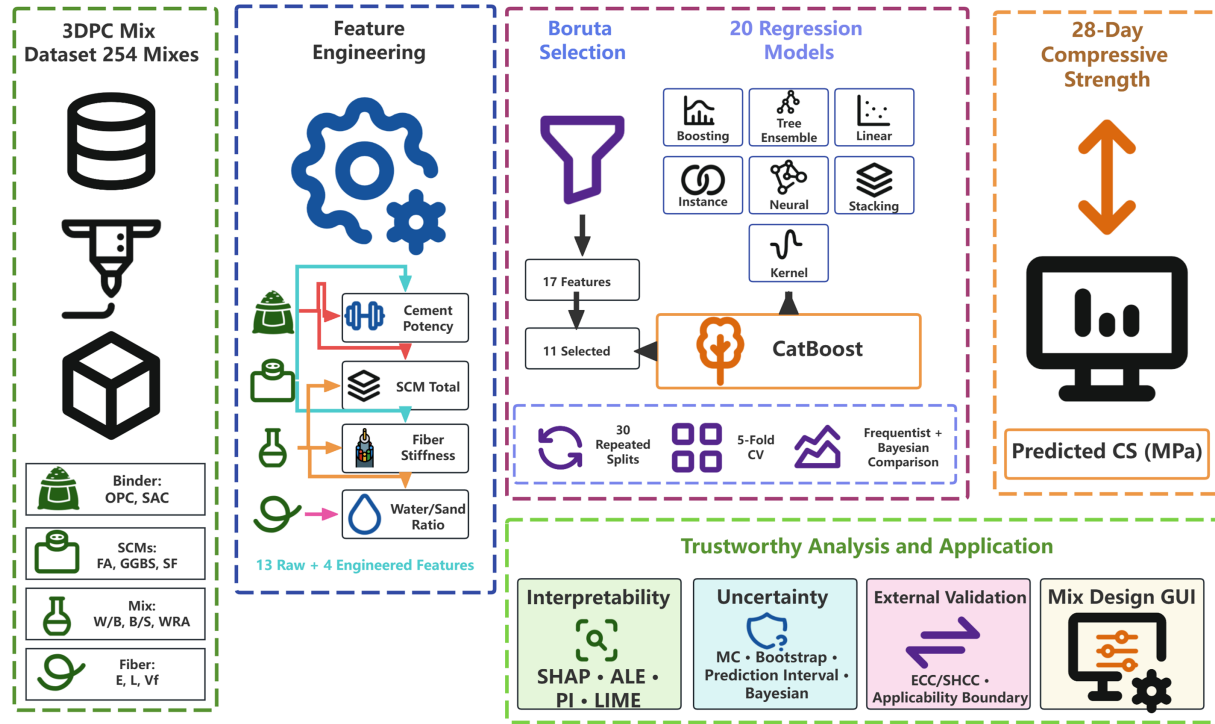
Five methodological gaps collectively limit the reliability and transferability of published 3DPC CS models. First, algorithm comparisons have evaluated no more than eight models under identical conditions, leaving the optimal predictor family unresolved [13–15]. Second, domain-informed feature engineering has not been attempted; composite variables capturing cement blend potency, cumulative SCM substitution, fibre volumetric stiffness, and water-to-sand ratio remain untested [14]. Third, all published metrics derive from a single random partition; with dataset sizes typically below 300, this yields split-sensitive estimates, and neither repeated frequentist testing nor Bayesian model comparison has been applied [13–16,32]. Fourth, prediction uncertainty has not been characterized with multiple complementary methods on the same CS model [24]. Fifth, cross-system external validation has not been performed, leaving generalization boundaries unknown [13–15]. Table 1 positions the present study against representative prior work across these five dimensions.

**Table 1:** Methodological comparison of representative 3DPC CS prediction studies.

| Study                 | Algorithms   | Samples | Feature Eng. | Repeated Val. | External Val. | Uncertainty | Interpretability       |
|-----------------------|--------------|---------|--------------|---------------|---------------|-------------|------------------------|
| Li et al. [14]        | 4            | 254     | No           | No            | No            | No          | SHAP, PDP              |
| Uddin et al. [15]     | 6            | 299     | No           | No            | No            | No          | SHAP                   |
| Alyami et al. [13]    | 7            | 299     | No           | No            | No            | No          | SHAP                   |
| Arif et al. [32]      | 3            | 299     | No           | No            | No            | No          | SHAP, PDP, ICE         |
| Uddin et al. [31]     | 8 + GEP      | 299     | No           | No            | Partial       | No          | SHAP                   |
| Zhang et al. [16]     | 5            | —       | No           | No            | No            | No          | None                   |
| Katlav & Turk [19]    | GBM + 4 opt. | 202     | No           | 5-fold CV     | No            | No          | SHAP, ICE              |
| Albostami et al. [33] | 4            | —       | No           | No            | No            | No          | SHAP, GEP              |
| Ma et al. [24]        | 1 (NGBoost)  | —       | No           | No            | No            | No          | SHAP                   |
| Khodadadi et al. [18] | 5 + meta     | —       | No           | No            | No            | No          | SHAP                   |
| This study            | 20           | 254     | Yes          | Bayesian      | Yes           | Yes         | SHAP + ALE + PI + LIME |

This study addresses these gaps through an ML framework for 3DPC CS prediction trained on 254 experimental records compiled from 24 peer-reviewed sources [14], spanning plain printable mortars to high-fibre reinforced composites (CS: 11.1–189.0 MPa). Six contributions are made. (1) Twenty regression algorithms spanning seven algorithmic families are benchmarked under identical preprocessing and validation conditions to determine the optimal predictor for 3DPC CS. (2) Four physically motivated composite features are constructed, encoding cement blend potency, cumulative SCM substitution, fibre volumetric stiffness, and water-to-sand ratio. (3) Thirty independent repeated experiments with frequentist paired testing under Bonferroni correction and Bayesian inference via the Savage-Dickey Bayes factor provide statistically defensible model rankings with effect sizes. (4) Four interpretability methods, SHAP, ALE, permutation importance, and LIME, jointly characterize global feature hierarchy and local prediction behavior. (5) Four uncertainty quantification methods, Monte Carlo simulation, bootstrap resampling, residual-based prediction intervals, and Bayesian credible intervals, decompose epistemic and aleatoric prediction error. (6) Cross-system external validation on an independent fibre-reinforced cementitious dataset establishes the applicability boundary of the trained model. A graphical user interface implementing the

calibrated CatBoost model supports practical mix design application without programming expertise. Fig. 1 summarizes the overall architecture of the proposed framework, from dataset compilation and composite feature engineering through Boruta selection, benchmarking of 20 regression models with statistically validated selection of CatBoost, to the trustworthy analysis and application layer comprising interpretability, uncertainty quantification, external validation, and the mix design interface.



**Figure 1:** Architecture of the proposed ML framework for 3DPC CS prediction.

## 2 Preliminary

### 2.1 Data Overview

The dataset contains 254 compressive strength records for 3D-printed concrete (3DPC), compiled from [14]. All specimens were tested perpendicular to the printing-layer direction at a standard 28-day curing age; because the curing period is uniform across the entire dataset, curing age is not retained as a predictor variable. The target variable CS spans 11.1–189.0 MPa (mean 64.1 MPa, SD 31.6 MPa), with the majority of observations concentrated below 80 MPa, reflecting the relatively moderate strength levels of printable paste formulations compared with cast ultra-high-performance concrete (UHPC). The compiled dataset and its full source attribution are publicly available at <https://github.com/lucassivan/ML-3DPC>. The two verifiable inclusion criteria applied during compilation are: (1) all retained records report CS measured with the loading axis perpendicular to the printed-layer interface, ensuring loading-direction homogeneity; (2) a uniform 28-day curing age applies across all source studies. The 24 contributing studies span multiple international testing standards and specimen geometries; cross-study variability in these procedural factors represents an uncontrolled source of measurement uncertainty absorbed into the model residuals and partially reflected in the 95% prediction interval half-width of 12 MPa.

Table 2 summarizes the variables employed in this study.



**Table 2:** Input and output variables for 3DPC CS prediction.

| Variable | Description                   | Unit        | Range                | Mean  | SD    | Zero (%) |
|----------|-------------------------------|-------------|----------------------|-------|-------|----------|
| f(OPC)   | OPC strength grade            | MPa         | 0–52.5               | 40.9  | 15.7  | 11.8     |
| n(OPC)   | OPC binder fraction           | —           | 0–1.0                | 0.691 | 0.314 | 11.8     |
| f(SAC)   | SAC strength grade            | MPa         | 0–42.5               | 8.2   | 16.8  | 80.7     |
| n(SAC)   | SAC binder fraction           | —           | 0–0.6                | 0.075 | 0.193 | 80.7     |
| n(FA)    | Fly ash binder fraction       | —           | 0–0.75               | 0.110 | 0.163 | 53.5     |
| n(GGBS)  | GGBS binder fraction          | —           | 0–0.3                | 0.067 | 0.103 | 65.4     |
| n(SF)    | Silica fume binder fraction   | —           | 0–0.27               | 0.056 | 0.066 | 44.1     |
| n(W/B)   | Water-to-binder ratio         | —           | 0 <sup>a</sup> –0.65 | 0.297 | 0.084 | —        |
| n(B/S)   | Binder-to-sand ratio          | —           | 0–4.5                | 1.091 | 0.635 | —        |
| n(WRA)   | WRA dosage                    | % of binder | 0–18.0               | 0.613 | 1.223 | 20.9     |
| n(Fb-E)  | Fiber elastic modulus         | GPa         | 0–200.0              | 18.0  | 38.3  | 59.8     |
| n(Fb-L)  | Fiber length                  | mm          | 0–16.0               | 3.37  | 4.61  | 59.8     |
| n(Fb)    | Fiber volume fraction         | %           | 0–4.0                | 0.312 | 0.566 | 60.2     |
| CS       | Compressive strength (target) | MPa         | 11.1–189.0           | 64.1  | 31.6  | —        |

Note: <sup>a</sup>The minimum of 0 for n(W/B) corresponds to a single record in which water content was not reported in the original source and was assigned zero in the compiled dataset; the minimum physically measured W/B across all other records is 0.15.

Thirteen input variables are organized into four groups. Binder composition is characterized by strength-grade and binder-fraction pairs for ordinary Portland cement (OPC) and sulfoaluminate cement (SAC): f(OPC) and n(OPC) encode the grade (MPa) and the fraction of total binder mass contributed by OPC, with the corresponding quantities f(SAC) and n(SAC) for SAC. Supplementary cementitious materials are represented by the binder fractions of fly ash n(FA), ground granulated blast-furnace slag n(GGBS), and silica fume n(SF). Mix proportions are captured by the water-to-binder ratio n(W/B), the binder-to-sand ratio n(B/S), and the water-reducing admixture dosage n(WRA) expressed as a percentage of binder mass. Fiber reinforcement is described by three continuous parameters: fiber elastic modulus n(Fb-E), fiber length n(Fb-L), and fiber volume fraction n(Fb), which encode fiber type implicitly through physical properties rather than through categorical dummies.

Several variables exhibit pronounced sparsity. SAC-related fields are non-zero in only 19.3% of records, consistent with the limited adoption of sulfoaluminate cement in 3DPC practice. GGBS (34.6% non-zero) and FA (46.5% non-zero) appear in fewer than half of the mixes, while all three fiber descriptors are non-zero in approximately 40% of records, reflecting the substantial proportion of fiber-free formulations. Particle size distribution (PSD) data for the constituent materials and detailed fine aggregate characterisation were not systematically reported across the 24 source studies and are therefore absent from the compiled dataset. Fine aggregate in extrusion-based 3DPC is intrinsically constrained to a narrow grading range (maximum aggregate size typically below 2 mm) by nozzle geometry requirements, limiting between-study variability attributable to aggregate characteristics; the binder-to-sand ratio n(B/S) implicitly encodes the aggregate dilution effect, and PSD-related variability in binder water demand is partially absorbed by n(W/B). Residual variability in aggregate properties across source studies represents an uncontrolled source of prediction error that contributes to the model's prediction interval width.

The full CS range (11.1–189.0 MPa) is modelled as a single population because the governing mechanisms (capillary porosity controlled by W/B, pozzolanic refinement from SCMs, and crack-arresting from fibres) apply continuously across it, and the multi-method interpretability analysis (Section 4.5) reveals no contradictory attribution patterns across strength regimes. Stratification into narrow strength bands is not feasible at the current dataset scale: the high-performance tier ( $CS > 120$  MPa) contains only 18 records (7.1%), insufficient for reliable standalone sub-model training. This sparsity limits calibration accuracy at the high-strength extreme, as reflected in the reduced prediction interval coverage reported in Section 4.4.

## 2.2 Exploratory Data Analysis

Fig. 2 presents the Pearson correlation matrix for all 13 input features and CS. The water-to-binder ratio  $n(W/B)$  is the sole dominant linear predictor, registering  $r = -0.628$  ( $p < 0.001$ ) and exceeding the next largest absolute correlation by a factor of 2.6. This magnitude reflects the central role of W/B in governing paste rheology and effective capillary porosity in 3DPC formulations: printability constraints impose a lower bound on W/B, while strength requirements impose an upper bound, compressing the workable range and amplifying its influence on mechanical output. A secondary cluster of moderately significant predictors includes  $n(GGBS)$  ( $r = +0.238$ ),  $n(Fb-E)$  ( $r = +0.229$ ),  $f(OPC)$  ( $r = +0.214$ ), and  $f(SAC)$  ( $r = -0.195$ ), all statistically significant at  $p < 0.01$ . The negative sign of  $f(SAC)$  is physically consistent with the lower long-term strength potential of sulfoaluminate cement under 28-day standard curing. In contrast,  $n(B/S)$  ( $r = -0.016$ ) and  $n(OPC)$  ( $r = +0.001$ ) show no linear association with CS, implying that their effects are either non-linear or interaction-dependent. The strongest inter-feature correlation is between  $f(OPC)$  and  $n(SAC)$  ( $r = -0.96$ ), a structural consequence of the near-complete compositional separation between OPC-based and SAC-based mix families in the compiled dataset. The three fiber descriptors are strongly intercorrelated ( $r = 0.57$ – $0.68$  for Fb-E, Fb-L, and Fb), reflecting that fiber specifications are typically fixed together within individual experimental programs.

Fig. 3 characterizes CS distribution and mix composition across binder systems and fiber types. OPC-GGBS blends achieve the highest mean CS (85.8 MPa,  $n = 21$ ); OPC-FA mixes show the widest spread ( $SD = 40.8$  MPa,  $n = 82$ ). W/B vs. CS yields  $r = -0.665$  ( $n = 253$ , excluding the single record with unreported W/B; see Table 2, note<sup>a</sup>); steel and carbon fiber specimens cluster at high CS and low W/B. OPC constitutes the dominant binder fraction across all CS ranges (0.68–0.71); 60.2% of specimens contain no discrete fibers.

Fig. 4 examines feature-CS relationships through three complementary visualizations. (a)  $n(W/B)$  leads all predictors at  $r = -0.628$ ; eight of thirteen features pass the  $p < 0.05$  significance threshold;  $n(B/S)$  and  $n(OPC)$  show no linear association with CS. (b) W/B quartile density plots reveal a 54.6 MPa mean gap between Q1 ( $W/B \leq 0.25$ , mean CS = 98.6 MPa) and Q4 ( $W/B > 0.35$ , mean CS = 44.0 MPa), confirming that W/B alone partitions the dataset into largely non-overlapping strength regimes. (c) Steel and carbon fiber mixes achieve the highest mean CS (94.7 MPa,  $SD = 58.5$  MPa); PVA/PP-reinforced mixes occupy the lowest strength tier (mean 43.8 MPa). (d) High-CS mixes are distinguished by lower  $n(W/B)$ , higher  $f(OPC)$ , and higher  $n(Fb-E)$ .

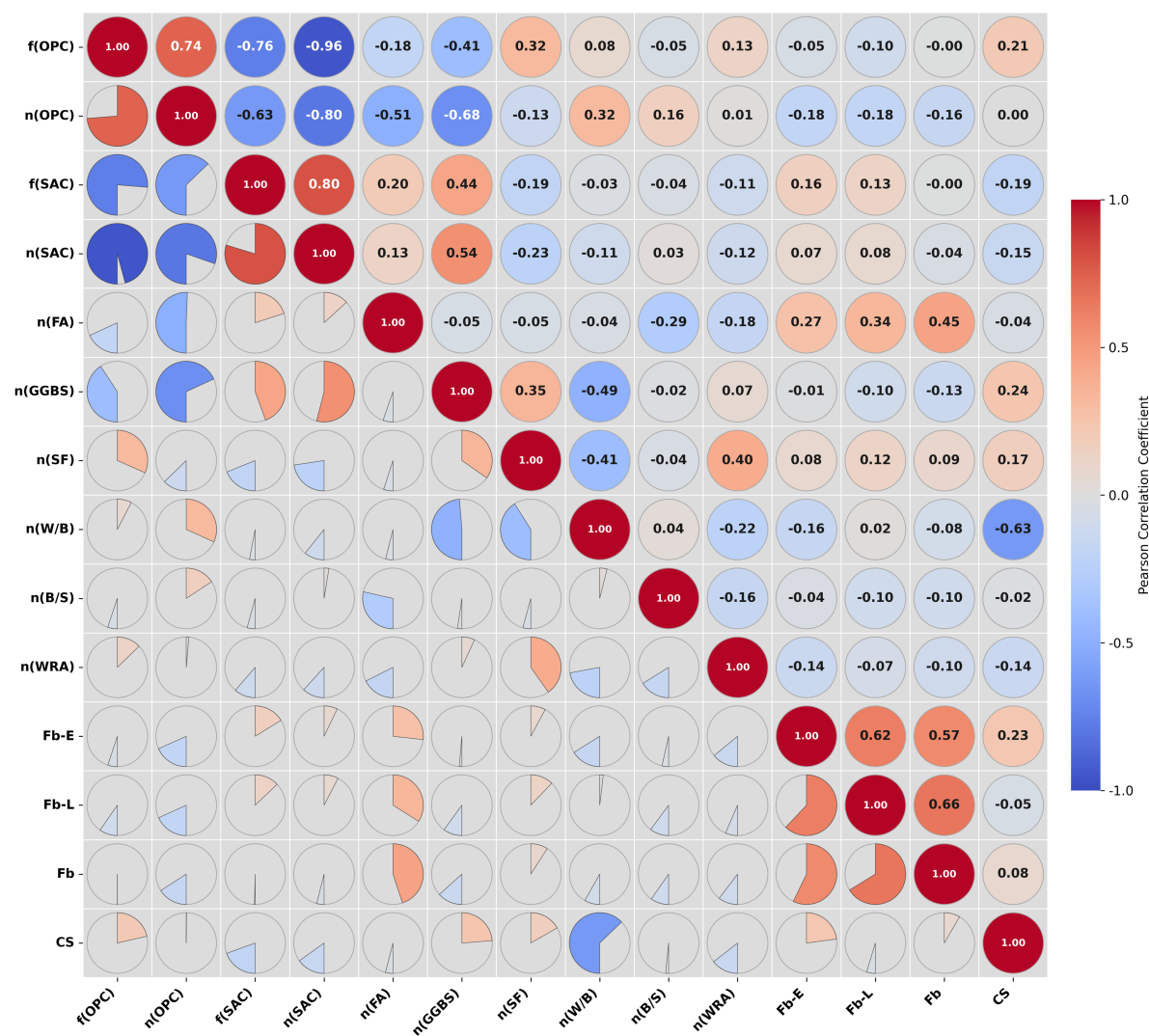


Figure 2: Pearson correlation matrix for all 13 input features and CS.

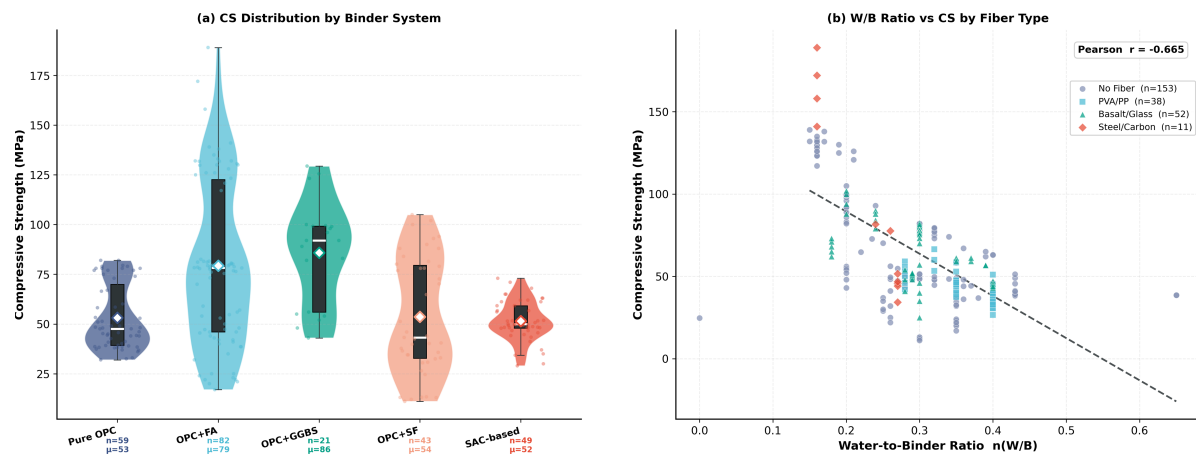


Figure 3: (Continued)

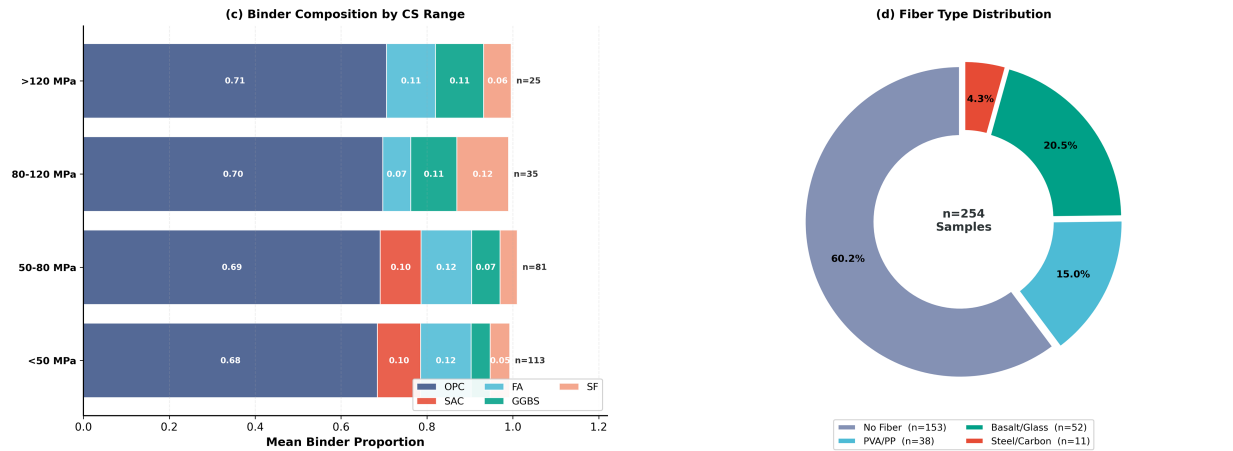


Figure 3: Dataset composition and CS distribution.

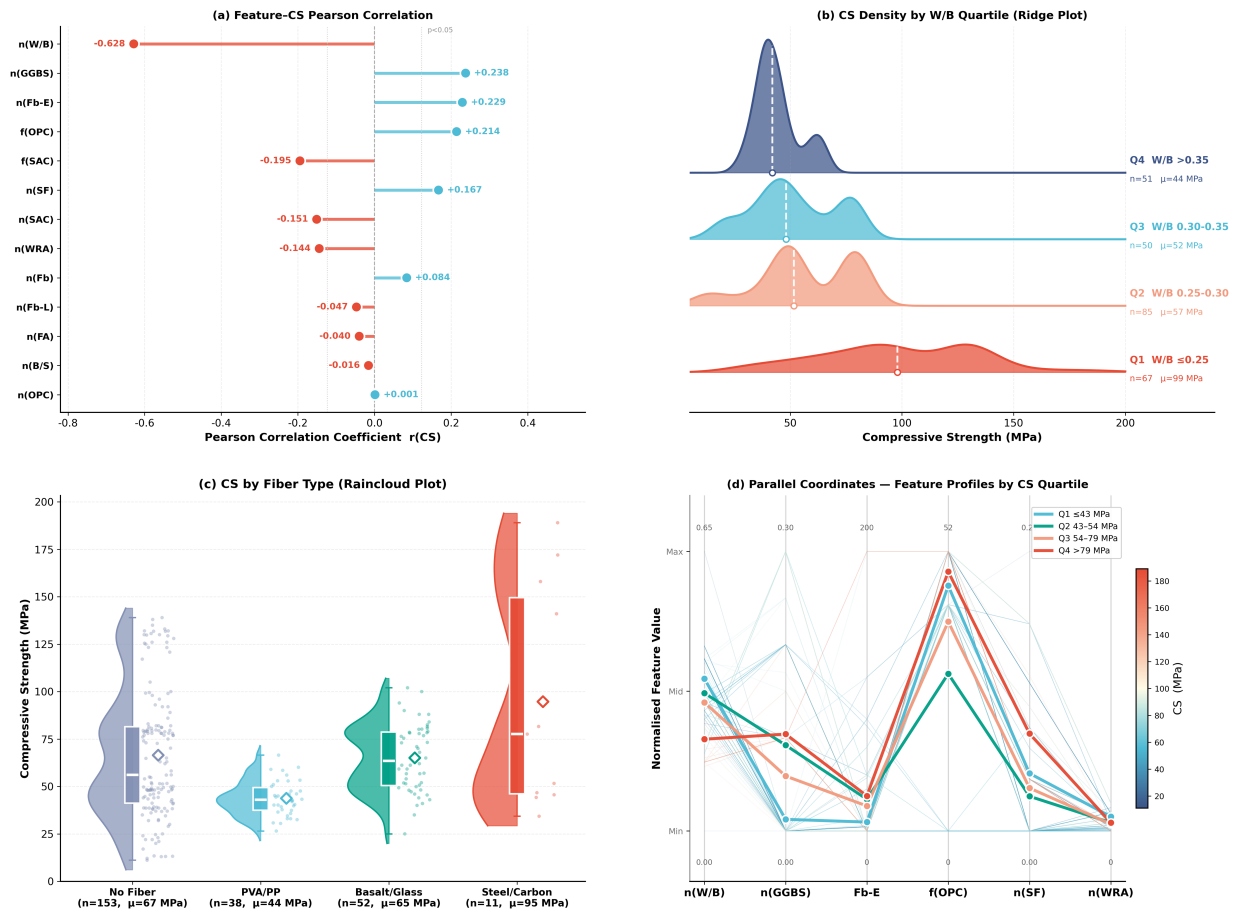


Figure 4: Feature-CS relationship analysis.

These patterns establish that CS prediction across the compiled dataset requires capturing non-linear interactions that linear correlation cannot fully represent. The isolation of n(W/B) as the dominant predictor, the near-perfect anti-correlation between f(OPC) and n(SAC), and the non-informative linear associations

of  $n(B/S)$  and  $n(OPC)$  motivate the composite feature engineering and Boruta-based selection applied in [Section 3](#).

### 3 Methods

This section describes the six-stage modelling framework: composite feature engineering, Boruta-based feature selection, regression model benchmarking, training and evaluation protocol, repeated-experiment statistical comparison, and prediction uncertainty quantification.

#### 3.1 Feature Engineering

Four composite variables were defined to encode interaction effects among cement type, pozzolanic substitution, fiber phase, and mix proportions that raw features represent only in isolation ([Table 3](#)).

**Table 3:** Engineered features and their physical rationale.

| Feature                | Expression                                  | Physical Rationale                                       |
|------------------------|---------------------------------------------|----------------------------------------------------------|
| Effective_Cement_Grade | $f(OPC) \cdot n(OPC) + f(SAC) \cdot n(SAC)$ | Weighted strength potential of the cement blend          |
| SCM_Total              | $n(FA) + n(GGBS) + n(SF)$                   | Aggregate supplementary cementitious substitution        |
| Fiber_Stiffness        | $n(Fb) \cdot n(Fb-E)$                       | Volumetric stiffness index of the fiber phase            |
| Water_Sand_Ratio       | $n(W/B) \cdot n(B/S)$                       | Effective water-to-sand ratio via $W/B \times B/S = W/S$ |

Effective\_Cement\_Grade collapses cement proportion and grade into a single index of net strength potential, recovering information lost when the two variables are modelled independently. SCM\_Total aggregates all pozzolanic substitutions to capture their cumulative influence on hydration kinetics and pore structure. Fiber\_Stiffness approximates the volumetric stiffness contribution of the fiber phase following classical composite mechanics. Water\_Sand\_Ratio derives the water-to-sand ratio algebraically, simultaneously encoding fluidity and aggregate dilution. The feature space expanded from 13 raw inputs to 17 variables after engineering.

#### 3.2 Boruta-Based Feature Selection

Feature relevance was assessed with the Boruta algorithm, a Random Forest-based wrapper that tests all features simultaneously against a null distribution constructed from permuted shadow copies. At each iteration, the algorithm trains a Random Forest on the combined real-and-shadow dataset and compares each feature's importance against the maximum shadow importance through a Z-score:

$$Z_j = \frac{VIM_j - \mu_{shadow}}{\sigma_{shadow}} \quad (1)$$

where  $VIM_j$  is the variable importance of feature  $j$ , and  $\mu_{shadow}$ ,  $\sigma_{shadow}$  are the mean and standard deviation of shadow feature importances across iterations. Features with importance consistently exceeding the shadow threshold are assigned Confirmed status; those consistently below it are rejected. Given the limited training size ( $n = 203$ ), Tentative features were retained alongside Confirmed ones to avoid discarding borderline predictors. The post-hoc interpretability analysis ([Section 4.5](#)) validates this decision: Fiber\_Stiffness ranks



third among all 11 selected features in ALE importance (24.81 MPa), confirming non-trivial predictive contribution beyond what its rejected constituent  $n(\text{Fb-E})$  provides individually.

Boruta was applied once to the training partition; the resulting binary feature mask was fixed for all subsequent analyses. Table 4 summarizes the outcome.

**Table 4:** Boruta feature selection results (17 features evaluated).

| Status    | Features                                                                                                                                                               | Count |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| Confirmed | $n(\text{OPC})$ , $n(\text{FA})$ , $n(\text{SF})$ , $n(\text{W/B})$ , $n(\text{B/S})$ , $n(\text{WRA})$ , $n(\text{Fb})$ ,<br>Effective_Cement_Grade, Water_Sand_Ratio | 9     |
| Tentative | SCM_Total, Fiber_Stiffness                                                                                                                                             | 2     |
| Rejected  | $f(\text{OPC})$ , $f(\text{SAC})$ , $n(\text{SAC})$ , $n(\text{GGBS})$ , $n(\text{Fb-E})$ , $n(\text{Fb-L})$                                                           | 6     |

Eleven features entered the final model input. The six rejected variables include both cement grade scalars and fiber/slag descriptors whose information is largely absorbed by the composite engineered features or correlated with retained variables. Notably,  $n(\text{GGBS})$  registers a moderate linear correlation with CS ( $r = +0.238$ ,  $p < 0.01$ ) yet is rejected: as a constituent of  $\text{SCM\_Total} = n(\text{FA}) + n(\text{GGBS}) + n(\text{SF})$ , its predictive information is subsumed by the composite term, leaving no residual importance above Boruta's shadow threshold.

### 3.3 Regression Models

Twenty regression algorithms spanning seven algorithmic families were evaluated under identical pre-processing and validation conditions; full hyperparameter configurations are given in Table 5. The gradient boosting group comprises CatBoost, XGBoost, LightGBM, GBR, and HistGB. Tree ensemble methods include Random Forest, Extra Trees, Bagging, AdaBoost, and a single Decision Tree as an interpretable baseline. Six linear models (LR, Ridge, Lasso, ElasticNet, BayesianRidge, and SGD) establish the performance floor under linearity assumptions. K-Nearest Neighbors (KNN) and Support Vector Regression (SVR) represent instance-based and kernel-based learning, respectively. A two-hidden-layer Multilayer Perceptron (MLP) covers neural network-based regression, and a StackingRegressor combining RF, XGBoost, and CatBoost with a Ridge meta-learner completes the benchmark. The twenty algorithms span a complexity spectrum from inherently interpretable linear estimators to high-capacity ensemble methods, enabling simultaneous assessment of the CS dataset's capacity to support non-linear learning.

**Table 5:** Machine learning model configurations.

| Category          | Model    | Key Parameters                                                                   |
|-------------------|----------|----------------------------------------------------------------------------------|
| Gradient Boosting | CatBoost | iterations = 400, depth = 4, lr = 0.1                                            |
|                   | XGBoost  | n_est = 300, max_depth = 4, lr = 0.1, subsample = 0.8,<br>min_child_weight = 3   |
|                   | LightGBM | n_est = 300, max_depth = 4, lr = 0.1, subsample = 0.8,<br>min_child_samples = 10 |
|                   | GBR      | n_est = 200, max_depth = 4, lr = 0.1, subsample = 0.8,<br>min_samples_leaf = 2   |
|                   | HistGB   | max_iter = 300, max_depth = 5, lr = 0.1                                          |

(Continued)

**Table 5 (continued)**

| Category       | Model             | Key Parameters                                                                    |
|----------------|-------------------|-----------------------------------------------------------------------------------|
| Tree Ensemble  | RF                | n_est = 200, max_depth = 8, min_samples_leaf = 4                                  |
|                | ET                | n_est = 200, max_depth = 8, min_samples_leaf = 4                                  |
|                | AdaBoost          | n_est = 100, lr = 0.8, base DT(depth = 4, leaf = 3)                               |
|                | Bagging           | n_est = 50, base DT(max_depth = 6)                                                |
|                | DecisionTree      | max_depth = 4, min_samples_leaf = 5                                               |
| Linear         | LR/Ridge          | default/ $\alpha = 1.0/\alpha = 0.1/\alpha = 0.1$ , l1_ratio = 0.5/default priors |
|                | /Lasso/ElasticNet |                                                                                   |
|                | /BayesianRidge    |                                                                                   |
|                | SGD               | max_iter = 2000, L2 penalty                                                       |
| Instance-based | KNN               | $k = 8$                                                                           |
| Kernel-based   | SVR               | RBF kernel, $C = 50$ , $\gamma = \text{scale}$                                    |
| Neural Network | MLP               | layers = (50,25), max_iter = 2000, early stopping                                 |
| Meta-Ensemble  | Stacking          | base = [RF, XGB, CatBoost], meta = Ridge( $\alpha = 1.0$ ), cv = 5                |

### 3.4 Training and Evaluation Protocol

All models were trained on the 203-sample training partition and evaluated on the fixed 51-sample test set; the 80/20 ratio was selected to maximize training coverage at  $n < 300$  while preserving a test set of sufficient size ( $n = 51$ ) for stable metric estimation. Cross-validation  $R^2$  (CV\_Mean  $\pm$  CV\_Std) was obtained by 5-fold on the training partition; the dataset  $\mathcal{D}$  was partitioned into five disjoint folds:

$$\mathcal{D} = \bigcup_{k=1}^5 \mathcal{F}_k, \quad \mathcal{F}_i \cap \mathcal{F}_j = \emptyset \quad (i \neq j) \quad (2)$$

Each fold  $\mathcal{F}_k$  served once as the validation set while the remaining four folds trained the model. The generalization gap was defined as:

$$\text{Gen\_gap} = R_{\text{train}}^2 - R_{\text{test}}^2 \quad (3)$$

### 3.5 Repeated Experiments and Statistical Comparison

A single 80/20 split yields performance estimates sensitive to the particular random partition, a concern amplified for datasets of fewer than 300 samples. To obtain distributional performance estimates and support rigorous model comparison, 30 independent experiments were conducted. Each run drew a new random 80/20 split while keeping the Boruta-selected feature mask fixed.

For each pairwise comparison, the 30-run  $R^2$  differences  $\Delta_i = R_{\text{CatBoost},i}^2 - R_{\text{other},i}^2$  were tested for normality with the Shapiro-Wilk statistic. When normality held, a two-sided paired  $t$ -test was applied; otherwise the Wilcoxon signed-rank test was used. With 19 pairwise comparisons against CatBoost, the Bonferroni-corrected significance threshold is  $\alpha^* = 0.05/19 \approx 0.003$ . Practical significance was assessed with Cohen's  $d$ :

$$d = \frac{\bar{\Delta}}{s_p}, \quad s_p = \sqrt{\frac{s_{\text{ref}}^2 + s_{\text{other}}^2}{2}} \quad (4)$$

where  $\bar{\Delta}$  is the mean of the 30 pairwise differences and  $s_p$  is the pooled standard deviation of the 30-run  $R^2$  values for the two models. The 95% confidence interval for  $\bar{\Delta}$  was estimated analytically via the  $t$ -distribution ( $df = 29$ ) and by 2000-resample non-parametric bootstrap.

Bayesian model comparison used the Savage-Dickey density ratio to compute the Bayes factor  $BF_{10}$ , quantifying relative evidence for  $H_1 : \Delta > 0$  against  $H_0 : \Delta = 0$ . A conjugate normal-normal model was adopted with prior  $\mathcal{N}(0, 0.1^2)$ . Given the 30-run difference sample with mean  $\bar{\Delta}$  and standard deviation  $s_{\Delta}$ , the posterior mean and variance are:

$$\mu_{\text{post}} = \frac{\mu_0/\sigma_0^2 + n\bar{\Delta}/s_{\Delta}^2}{1/\sigma_0^2 + n/s_{\Delta}^2}, \quad \sigma_{\text{post}}^2 = \frac{1}{1/\sigma_0^2 + n/s_{\Delta}^2} \quad (5)$$

The Bayes factor follows from the Savage-Dickey ratio:

$$BF_{10} = \frac{\mathcal{N}(0; \mu_0, \sigma_0^2)}{\mathcal{N}(0; \mu_{\text{post}}, \sigma_{\text{post}}^2)} \quad (6)$$

Evidence is interpreted on the conventional scale:  $BF_{10} < 1$  favors  $H_0$ ; 1–3 weak, 3–10 moderate, 10–30 strong, 30–100 very strong, and  $>100$  extreme evidence for  $H_1$ . Posterior probability of  $H_1$  and 95% credible intervals were derived from 50,000 posterior samples drawn from the analytical normal posterior.

### 3.6 Uncertainty Quantification

Four methods are applied to the held-out test set to quantify prediction uncertainty of the best-performing model. Let  $\hat{y}_i$  denote the predicted value for test sample  $i$  and  $n$  the test sample count. The mean prediction, predictive standard deviation, relative uncertainty index, and empirical coverage rate are defined as:

$$\bar{\hat{y}} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i \quad (7)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2} \quad (8)$$

$$U(\%) = \frac{\sigma}{\bar{\hat{y}}} \times 100 \quad (9)$$

$$\text{Coverage} = \frac{\#\{y_i \in \text{PI}_i\}}{n} \quad (10)$$

Monte Carlo simulation retrains the model on 1000 bootstrap-resampled training sets and evaluates on the fixed test set, yielding a distribution of predictions from which the mean, standard deviation, and a per-sample 2.5th–97.5th percentile bootstrap interval of the predicted response are computed. Because this interval quantifies the sampling variability of the model prediction (the estimated conditional mean at each test point) induced by finite training data rather than the scatter of individual observations, it is an epistemic confidence-type interval, distinct from the prediction intervals defined below. Bootstrap resampling applies the same resample-refit-predict procedure and yields a mean prediction and 95% percentile interval that closely reproduce the Monte Carlo estimate, providing an independent check on this epistemic uncertainty. Prediction intervals are constructed from training residuals using a  $t$ -distribution correction for finite sample size, and their empirical coverage rate is compared against the nominal 95% level. Bayesian credible intervals

are derived by placing an inverse-Gamma prior on residual variance, drawing 5000 posterior samples, and computing 2.5th and 97.5th percentiles of the resulting predictive distribution.

## 4 Experiments and Results

### 4.1 Experimental Configuration and Setup

Random seeds were fixed throughout to ensure reproducibility. The complete pipeline covered data loading and cleaning, StandardScaler normalization (fit on training partition only), Boruta feature selection, model training, and performance evaluation. For the repeated-experiment framework, 30 independent 80/20 splits were generated and all 20 models were retrained and evaluated on each partition, with the Boruta feature mask held constant. The 5-fold cross-validation was conducted independently within each of the 30 runs.

### 4.2 Evaluation Metrics

Model performance was quantified using twelve metrics computed on the held-out test set ( $n = 51$ ), including root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE). Let  $y_i$  and  $\hat{y}_i$  denote observed and predicted values for sample  $i$ ,  $\bar{y}$ , the mean of observations, and  $n$ , the sample count.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (15)$$

$$\text{Max Error} = \max_i |y_i - \hat{y}_i| \quad (16)$$

$$\text{MdAE} = \text{median}(|y_i - \hat{y}_i|) \quad (17)$$

$$\text{MAD} = \text{median}(|e_i - \tilde{e}|), \quad e_i = y_i - \hat{y}_i, \quad \tilde{e} = \text{median}(e_i) \quad (18)$$

$$\text{EVS} = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)} \quad (19)$$

$R^2$  and EVS coincide when residual mean is zero; their divergence indicates systematic prediction bias. CV\_Mean and CV\_Std (Eq. (2)) and Gen\_gap (Eq. (3)) complete the twelve-metric suite.

### 4.3 Model Performance Comparison and Analysis

Table 6 reports the evaluation for all 20 models on the fixed seed-42 partition, sorted by test  $R^2$ , together with cross-validation stability and generalization gap. The top tier is occupied by gradient boosting models and the stacking meta-ensemble, all achieving  $R^2 > 0.83$ . CatBoost ( $R^2 = 0.8675$ , RMSE = 8.82 MPa, MAPE = 9.68%) and XGBoost ( $R^2 = 0.8638$ ) rank first among single-model estimators; Stacking ( $R^2 = 0.8676$ ) marginally exceeds CatBoost by less than 0.0001 on this partition. Tree ensemble methods (RF, ET, Bagging, AdaBoost) form a second tier at  $R^2 = 0.78$ –0.84. Linear models yield  $R^2 \approx 0.35$ –0.39, confirming that the CS–composition relationship is substantially nonlinear. SVR achieves a competitive test  $R^2 = 0.81$  but a cross-validation  $R^2$  of only 0.587 (CV\_Std = 0.114), attributable to feature standardization applied before fold splitting; the held-out test result is unaffected and is the reliable estimate. The MLP result ( $\bar{R}^2 = -0.65 \pm 2.93$  across 30 runs) reflects training instability rather than a meaningful performance estimate: with  $n = 203$  training samples, the two-hidden-layer configuration (50–25 neurons,  $\approx 1300$  parameters) is overparameterised relative to the data volume, causing gradient descent to fail reliably within 2000 iterations across many random partitions. The large variance ( $\pm 2.93$ ) confirms that outcomes are dominated by random initialisation rather than the composition–strength relationship.

**Table 6:** Single-run (seed = 42) performance of all 20 models on the test set ( $n = 51$ ).

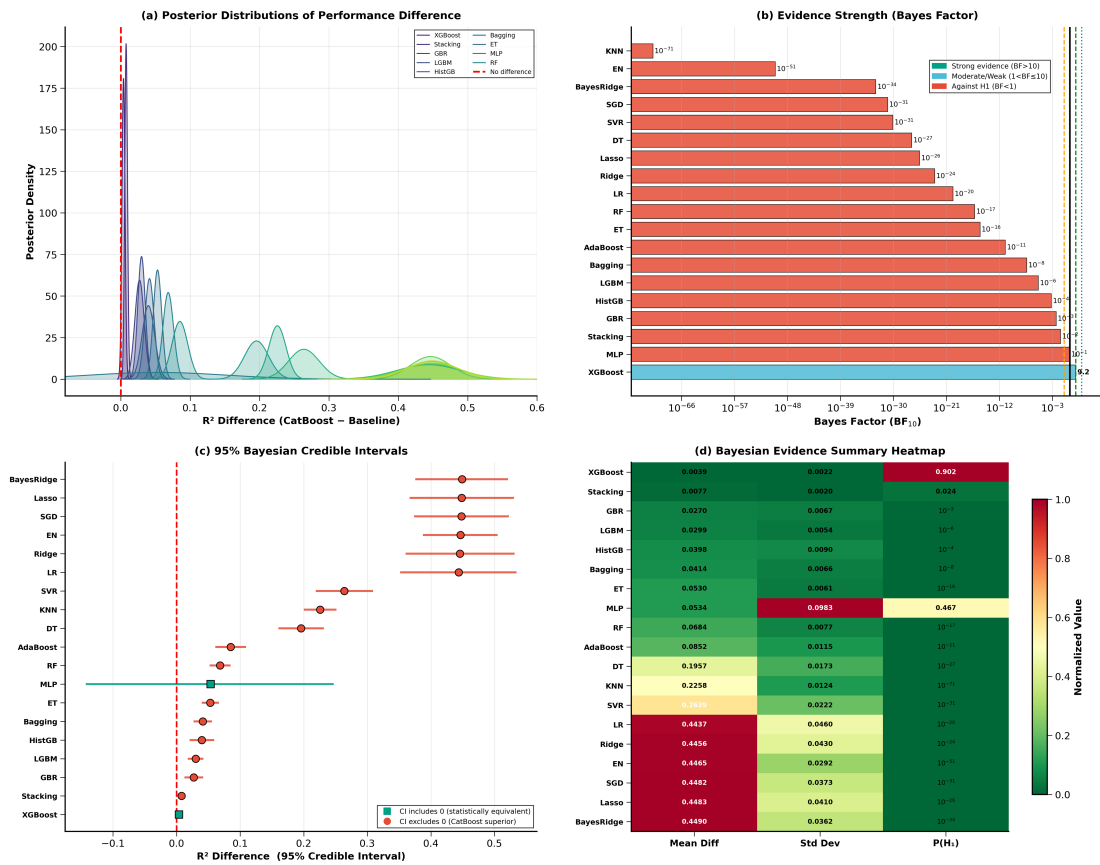
| Model         | $R^2$  | RMSE    | MAE     | MAPE (%) | MSE    | Max Error | MdAE    | MAD     | EVS    | CV_Mean | CV_Std | Gen_Gap |
|---------------|--------|---------|---------|----------|--------|-----------|---------|---------|--------|---------|--------|---------|
| Stacking      | 0.8676 | 8.8205  | 6.2568  | 9.90     | 77.80  | 29.50     | 4.3783  | 4.1923  | 0.8721 | 0.9011  | 0.0288 | 0.0959  |
| CatBoost      | 0.8675 | 8.8237  | 6.2011  | 9.68     | 77.86  | 28.62     | 4.4069  | 4.0841  | 0.8774 | 0.9038  | 0.0310 | 0.0988  |
| XGBoost       | 0.8638 | 8.9460  | 6.4593  | 10.10    | 80.03  | 28.60     | 4.5476  | 4.3546  | 0.8760 | 0.9017  | 0.0273 | 0.1041  |
| HistGB        | 0.8439 | 9.5798  | 6.9352  | 10.51    | 91.77  | 28.43     | 5.4079  | 5.0995  | 0.8618 | 0.8751  | 0.0378 | 0.1078  |
| ET            | 0.8434 | 9.5941  | 6.9243  | 10.25    | 92.05  | 27.89     | 4.2156  | 5.3125  | 0.8621 | 0.8567  | 0.0268 | 0.0793  |
| GBR           | 0.8403 | 9.6876  | 6.7305  | 10.49    | 93.85  | 30.73     | 4.7660  | 3.6775  | 0.8544 | 0.8864  | 0.0367 | 0.1282  |
| LightGBM      | 0.8392 | 9.7218  | 6.8870  | 10.56    | 94.51  | 28.52     | 4.6233  | 4.0706  | 0.8533 | 0.8881  | 0.0316 | 0.1220  |
| Bagging       | 0.8122 | 10.5056 | 7.5738  | 10.94    | 110.37 | 29.54     | 4.4610  | 4.9362  | 0.8377 | 0.8520  | 0.0541 | 0.1219  |
| SVR           | 0.8060 | 10.6771 | 7.1597  | 11.79    | 114.00 | 35.10     | 4.2236  | 4.1231  | 0.8067 | 0.5872  | 0.1139 | −0.0648 |
| AdaBoost      | 0.7987 | 10.8764 | 8.5063  | 12.46    | 118.30 | 27.22     | 6.9900  | 5.9115  | 0.8414 | 0.8135  | 0.0631 | 0.1195  |
| RF            | 0.7782 | 11.4168 | 8.1432  | 11.34    | 130.34 | 38.21     | 5.2257  | 5.1066  | 0.8071 | 0.8352  | 0.0478 | 0.1196  |
| KNN           | 0.7081 | 13.0977 | 9.3936  | 14.13    | 171.55 | 49.73     | 7.8250  | 6.9300  | 0.7303 | 0.6183  | 0.0658 | 0.0660  |
| DT            | 0.6107 | 15.1266 | 10.5606 | 14.59    | 228.81 | 53.69     | 6.0273  | 5.3214  | 0.6415 | 0.6614  | 0.1099 | 0.1392  |
| MLP           | 0.5025 | 17.0992 | 13.3736 | 20.74    | 292.38 | 42.39     | 12.1196 | 9.8981  | 0.6061 | 0.4785  | 0.3331 | 0.1784  |
| ElasticNet    | 0.3901 | 18.9329 | 13.9546 | 22.67    | 358.45 | 82.80     | 10.9353 | 10.8077 | 0.4696 | 0.5117  | 0.0837 | 0.2199  |
| BayesianRidge | 0.3755 | 19.1579 | 13.9882 | 23.03    | 367.03 | 85.58     | 12.7163 | 9.4381  | 0.4586 | 0.4730  | 0.0571 | 0.2399  |
| Lasso         | 0.3604 | 19.3877 | 14.0101 | 23.25    | 375.88 | 88.07     | 12.9907 | 8.8245  | 0.4452 | 0.4123  | 0.1341 | 0.2568  |
| SGD           | 0.3604 | 19.3882 | 14.0168 | 23.24    | 375.90 | 88.27     | 11.7765 | 9.7734  | 0.4469 | 0.4267  | 0.1041 | 0.2548  |
| Ridge         | 0.3595 | 19.4016 | 13.9785 | 23.23    | 376.42 | 87.56     | 13.2834 | 8.7777  | 0.4460 | 0.4167  | 0.1278 | 0.2588  |
| LR            | 0.3492 | 19.5570 | 13.9690 | 23.28    | 382.48 | 88.60     | 13.1665 | 8.7625  | 0.4377 | 0.3792  | 0.1894 | 0.2696  |

CatBoost is the recommended algorithm for 3DPC CS prediction: it achieves the highest 30-run mean  $R^2 = 0.8968 \pm 0.0505$  and is not statistically separated from XGBoost after Bonferroni correction ( $p = 0.085$ ,  $d = 0.08$ , 95% CI spanning zero) while maintaining lower cross-run variance. The seed-42 result ( $R^2 = 0.8675$ ) falls 0.58 standard deviations below the 30-run mean, indicating a moderately harder-than-average test partition rather than a representative performance estimate. Results are presented in Table 7 and Fig. 5.

**Table 7:** 30-run statistical comparison (CatBoost as reference model).

| Model         | Mean $R^2 \pm \text{Std}$ | CV $R^2$            | $p$ -Value | 95% CI of $\bar{\Delta}$ | Cohen's $d$ | BF <sub>10</sub> | Evidence           |
|---------------|---------------------------|---------------------|------------|--------------------------|-------------|------------------|--------------------|
| CatBoost      | 0.8968 $\pm$ 0.0505       | 0.8797 $\pm$ 0.0269 | —          | —                        | —           | —                | Reference          |
| XGBoost       | 0.8929 $\pm$ 0.0526       | 0.8756 $\pm$ 0.0275 | 0.085      | [−0.0006, 0.0085]        | 0.08        | 9.23             | Moderate ( $H_1$ ) |
| Stacking      | 0.8891 $\pm$ 0.0537       | 0.8732 $\pm$ 0.0285 | <0.001**   | [0.0037, 0.0118]         | 0.15        | 0.02             | Against $H_1$      |
| GBR           | 0.8697 $\pm$ 0.0765       | 0.8593 $\pm$ 0.0319 | <0.001**   | [0.0134, 0.0409]         | 0.42        | 0.00             | Against $H_1$      |
| LightGBM      | 0.8668 $\pm$ 0.0715       | 0.8544 $\pm$ 0.0309 | <0.001**   | [0.0190, 0.0411]         | 0.49        | 0.00             | Against $H_1$      |
| HistGB        | 0.8567 $\pm$ 0.0842       | 0.8439 $\pm$ 0.0328 | <0.001**   | [0.0216, 0.0586]         | 0.58        | 0.00             | Against $H_1$      |
| Bagging       | 0.8553 $\pm$ 0.0753       | 0.8408 $\pm$ 0.0305 | <0.001**   | [0.0280, 0.0551]         | 0.65        | 0.00             | Against $H_1$      |
| ET            | 0.8436 $\pm$ 0.0551       | 0.8173 $\pm$ 0.0323 | <0.001**   | [0.0407, 0.0656]         | 1.01        | 0.00             | Against $H_1$      |
| RF            | 0.8280 $\pm$ 0.0763       | 0.8021 $\pm$ 0.0325 | <0.001**   | [0.0531, 0.0845]         | 1.06        | 0.00             | Against $H_1$      |
| AdaBoost      | 0.8105 $\pm$ 0.0956       | 0.8090 $\pm$ 0.0386 | <0.001**   | [0.0627, 0.1100]         | 1.13        | 0.00             | Against $H_1$      |
| DT            | 0.6950 $\pm$ 0.1182       | 0.6918 $\pm$ 0.0520 | <0.001**   | [0.1657, 0.2378]         | 2.22        | 0.00             | Against $H_1$      |
| KNN           | 0.6675 $\pm$ 0.0633       | 0.6223 $\pm$ 0.0314 | <0.001**   | [0.2037, 0.2549]         | 4.00        | 0.00             | Against $H_1$      |
| SVR           | 0.6192 $\pm$ 0.1182       | 0.6031 $\pm$ 0.0492 | <0.001**   | [0.2311, 0.3241]         | 3.05        | 0.00             | Against $H_1$      |
| ElasticNet    | 0.4085 $\pm$ 0.2015       | 0.5147 $\pm$ 0.0540 | <0.001**   | [0.4257, 0.5508]         | 3.32        | 0.00             | Against $H_1$      |
| BayesianRidge | 0.3802 $\pm$ 0.2481       | 0.4776 $\pm$ 0.0746 | <0.001**   | [0.4373, 0.5960]         | 2.89        | 0.00             | Against $H_1$      |
| SGD           | 0.3764 $\pm$ 0.2543       | 0.4614 $\pm$ 0.0818 | <0.001**   | [0.4383, 0.6025]         | 2.84        | 0.00             | Against $H_1$      |
| Lasso         | 0.3578 $\pm$ 0.2825       | 0.4340 $\pm$ 0.1049 | <0.001**   | [0.4470, 0.6310]         | 2.66        | 0.00             | Against $H_1$      |
| Ridge         | 0.3503 $\pm$ 0.2971       | 0.4367 $\pm$ 0.1039 | <0.001**   | [0.4492, 0.6437]         | 2.56        | 0.00             | Against $H_1$      |
| LR            | 0.3341 $\pm$ 0.3207       | 0.3998 $\pm$ 0.1268 | <0.001**   | [0.4568, 0.6686]         | 2.45        | 0.00             | Against $H_1$      |
| MLP           | −0.650 $\pm$ 2.925        | 0.6518 $\pm$ 0.1249 | <0.001**   | [0.4655, 2.6275]         | 0.75        | 0.88             | Against $H_1$      |

Note: \*\* Bonferroni-corrected significance ( $\alpha^* = 0.0026$ ).

**Figure 5:** Bayesian statistical comparison of CatBoost against 19 baseline models.



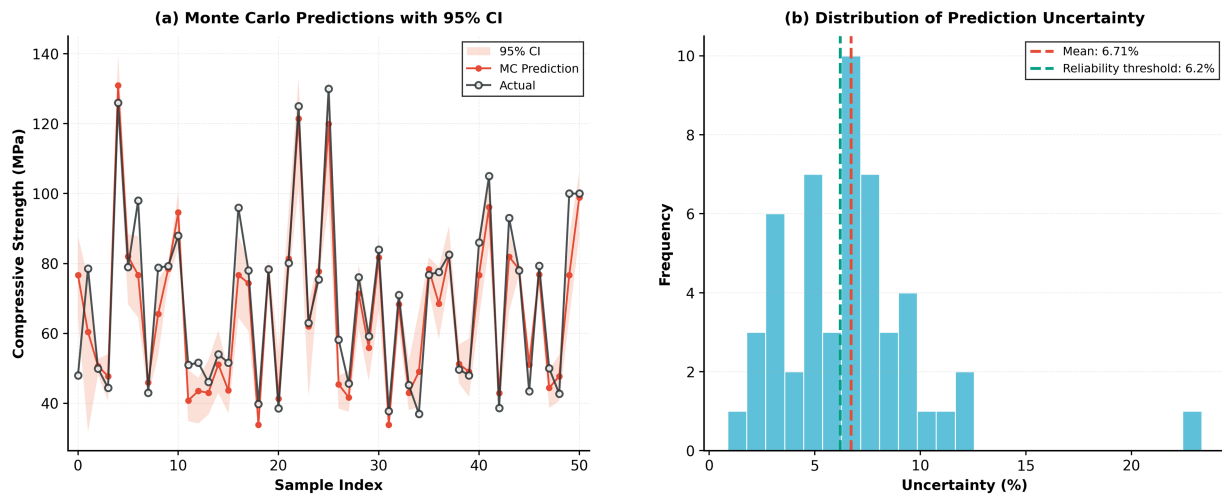
Frequentist testing finds 18 of 19 comparisons Bonferroni-significant; XGBoost is the sole exception and is not statistically separated from CatBoost. Bayesian analysis reveals a  $p$ -value/effect-size tension for Stacking: despite frequentist significance ( $p < 0.001$ ),  $BF_{10} = 0.02$  indicates the mean difference  $\bar{\Delta} \approx 0.008$  is too small to shift the posterior, and Bayesian inference supports  $H_0$ . Tree ensemble methods show large effect sizes ( $d = 1.01$ – $1.13$ ); linear models and KNN are separated by  $d > 2.4$ , confirming that gradient boosting captures nonlinear composition–strength interactions that simpler architectures cannot. The MLP collapses to a mean  $R^2$  of  $-0.65 \pm 2.93$ , reflecting instability on 203 training samples.

#### 4.4 Uncertainty Analysis

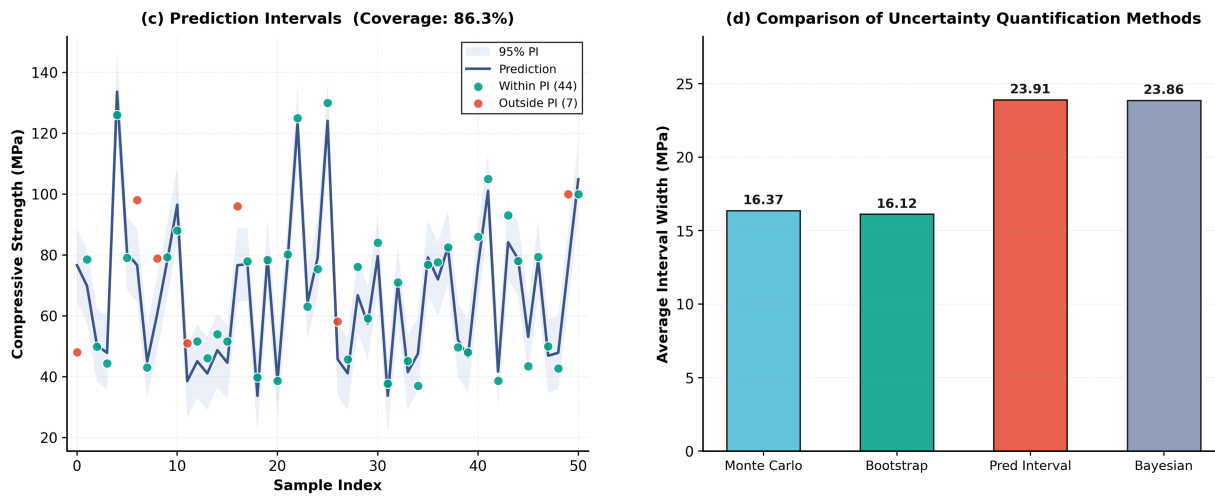
Four complementary methods quantify prediction uncertainty of the optimized CatBoost model on the held-out test set ( $n = 51$ ): Monte Carlo simulation ( $n = 1000$ ), Bootstrap resampling ( $n = 1000$ ), residual-based prediction intervals, and Bayesian credible intervals. Results are summarized in Table 8 and visualized in Fig. 6.

**Table 8:** Uncertainty analysis of the CatBoost model (test set,  $n = 51$ ).

| Method                     | Mean (MPa) | Std (MPa) | $U\%$ | 95% CI Width (MPa) | Coverage | $R^2$               | RMSE (MPa)       |
|----------------------------|------------|-----------|-------|--------------------|----------|---------------------|------------------|
| Monte Carlo ( $n = 1000$ ) | 66.47      | 4.27      | 6.71  | 16.37              | —        | $0.8223 \pm 0.0462$ | $10.14 \pm 1.23$ |
| Bootstrap ( $n = 1000$ )   | 66.49      | —         | —     | 16.12              | —        | —                   | —                |
| Prediction Interval (95%)  | —          | —         | —     | 23.91              | 86.3%    | —                   | —                |
| Bayesian Credible          | 66.92      | 6.08      | —     | 23.86              | —        | —                   | —                |



**Figure 6:** (Continued)



**Figure 6:** Uncertainty quantification of the CatBoost model.

The Monte Carlo analysis yields an overall uncertainty of 6.71%, indicating low parametric sensitivity to training-set composition given the small sample size and wide CS range (11.1–189.0 MPa). Bootstrap resampling independently confirms this, with a mean interval width of 16.12 MPa nearly identical to the Monte Carlo estimate (16.37 MPa). The residual-based prediction interval achieves 86.3% empirical coverage against the 95% nominal level; this shortfall is consistent with training residuals underestimating test-set error under mild overfitting ( $\text{Gen\_gap} = 0.099$ ). Stratifying the test set by strength range shows that coverage is lowest in the 80–120 MPa band (72.7%, 8 of 11 specimens), where prediction residuals are largest; the above-120 MPa stratum contains only 3 test specimens, too few for a reliable coverage estimate at the highest-strength extreme. The shortfall is attributable to the uniform-margin construction of the residual-based PI: a single global residual SD is applied regardless of predicted strength, so strata with systematically larger residuals are penalised disproportionately. The Bayesian credible interval yields an almost identical width (23.86 vs. 23.91 MPa), confirming convergence of both residual-based methods. The approximately 1.5-fold width difference between the resampling methods ( $\sim 16$  MPa) and the residual-based methods ( $\sim 24$  MPa) reflects their complementary scopes: Monte Carlo and Bootstrap capture epistemic uncertainty from finite training data, while prediction and Bayesian intervals additionally incorporate aleatoric noise. For structural design calculations, a minimum CS safety margin of 12 MPa (the half-width of the 95% prediction interval) is recommended when using model predictions as the nominal strength target.

#### 4.5 Model Interpretability Analysis

Four complementary methods—SHAP, ALE, permutation importance, and LIME—were applied to the optimized CatBoost model to attribute predictions to input variables and verify physical consistency.

**SHAP analysis (Fig. 7).**  $n(\text{W/B})$  is the most influential predictor globally, with a mean  $|\text{SHAP}|$  of 11.73 MPa and a dispersion spanning  $-15$  to  $+42$  MPa across the test set, reflecting the central role of water-to-binder ratio in governing paste rheology and strength development.  $n(\text{WRA})$  (4.54 MPa),  $n(\text{B/S})$  (3.12 MPa), and  $\text{Effective\_Cement\_Grade}$  (3.06 MPa) occupy the secondary tier. The  $n(\text{W/B})$  dependence plot exhibits a non-monotonic SHAP trajectory: negative contributions at sub-average W/B ratios give way to large positive values above the mean, capturing the printability-window constraint characteristic of 3DPC systems.

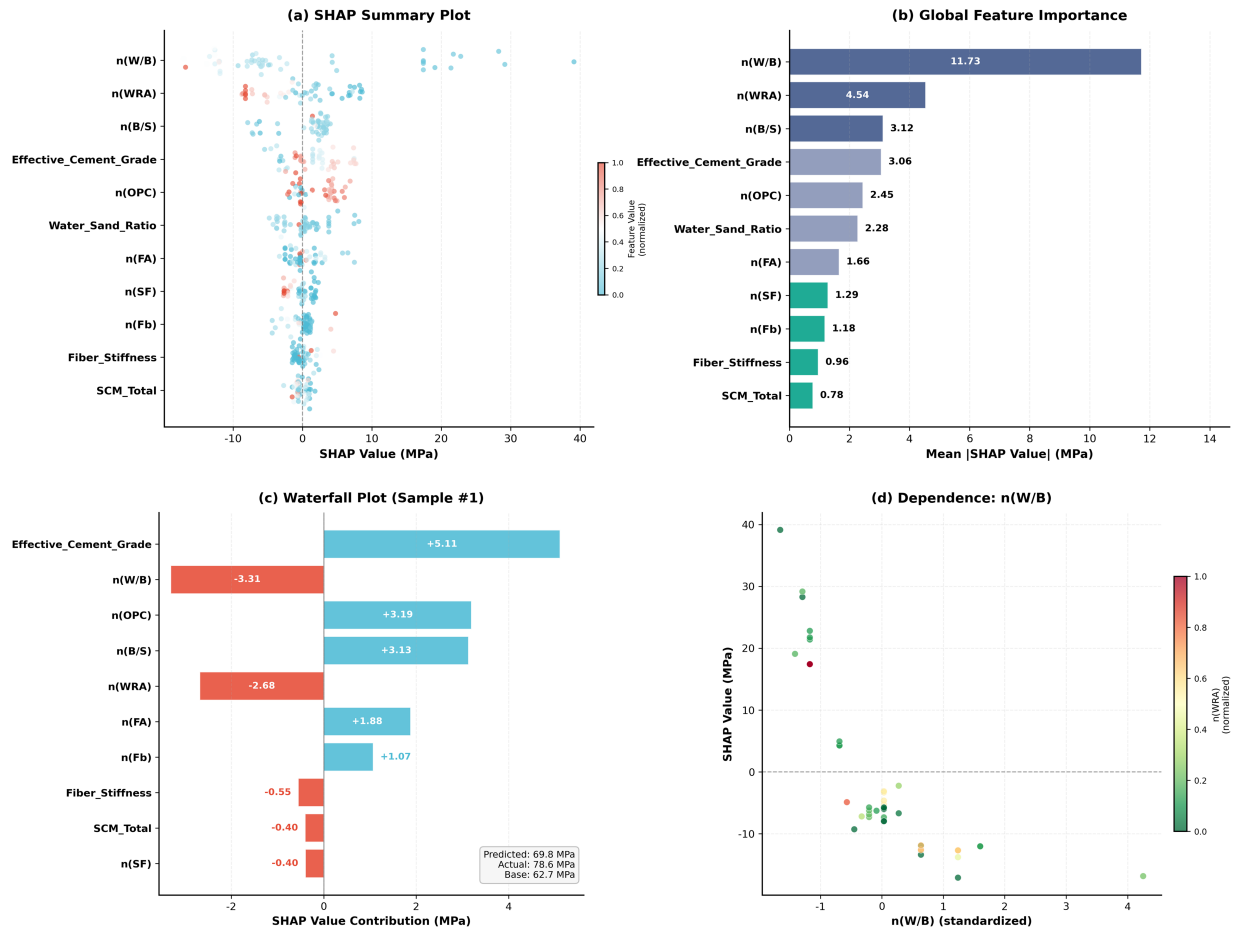


Figure 7: SHAP explainability analysis of the CatBoost model.

**ALE analysis (Fig. 8).** The ALE importance index confirms n(W/B) as the dominant variable (57.57 MPa), with an effect range of  $-14.85$  to  $+42.72$  MPa illustrating pronounced nonlinearity. n(OPC) ranks second (26.36 MPa), followed by Fiber\_Stiffness (24.81 MPa), n(B/S) (14.49 MPa), and Effective\_Cement\_Grade (13.87 MPa). The Fiber\_Stiffness ALE range (up to  $+22.26$  MPa) demonstrates that the composite fiber index captures volumetric stiffness contributions beyond what fiber volume fraction alone encodes.

**Permutation importance (Fig. 9).** Across all four scoring metrics, n(W/B) retains unambiguous primacy: its permutation reduces test  $R^2$  by  $0.812 \pm 0.145$  (approximately 81% of explained variance) and increases RMSE by  $14.66 \pm 1.81$  MPa. n(OPC) ranks second under  $R^2$  and RMSE metrics. Spearman correlations between the four metric-derived importance vectors range from  $\rho = 0.982$  to  $1.000$ , confirming the hierarchy is robust to the choice of loss function.

**LIME analysis (Fig. 10).** Local explanations across four prediction quartiles reveal how the dominant driver shifts with predicted strength. At low predictions ( $\leq 46.7$  MPa), n(W/B) imposes a large negative contribution (weight  $\approx -11.5$ ), indicating that above-average W/B is the primary inhibitor of low-regime CS. In the high-prediction range ( $> 78.9$  MPa), n(W/B) sustains a large positive role alongside n(B/S) and Effective\_Cement\_Grade, indicating that high-performance mixes benefit from simultaneous optimization of multiple variables. The sign reversal of n(W/B) captures the non-monotonic workability–strength coupling characteristic of 3DPC, where sub-optimal W/B impairs layer bonding while mixes near the printability optimum exploit full binder reactivity.

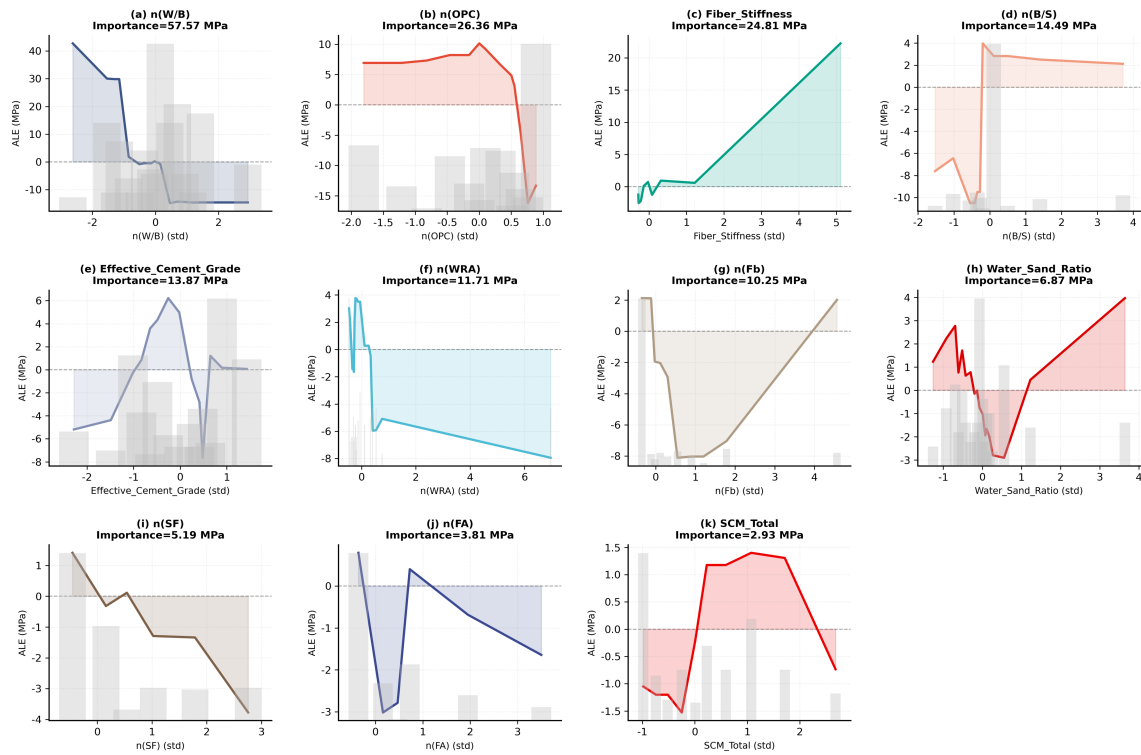


Figure 8: ALE analysis for all 11 Boruta-selected features, ranked by ALE importance.

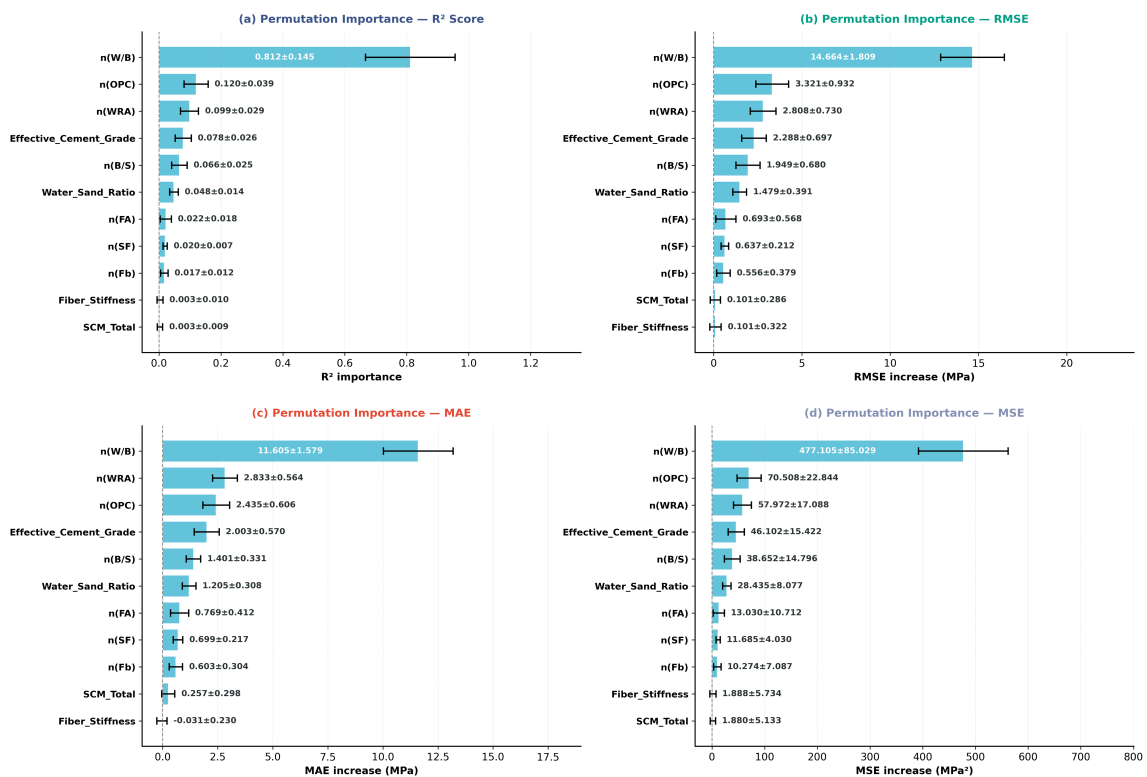


Figure 9: Permutation importance under four scoring metrics.

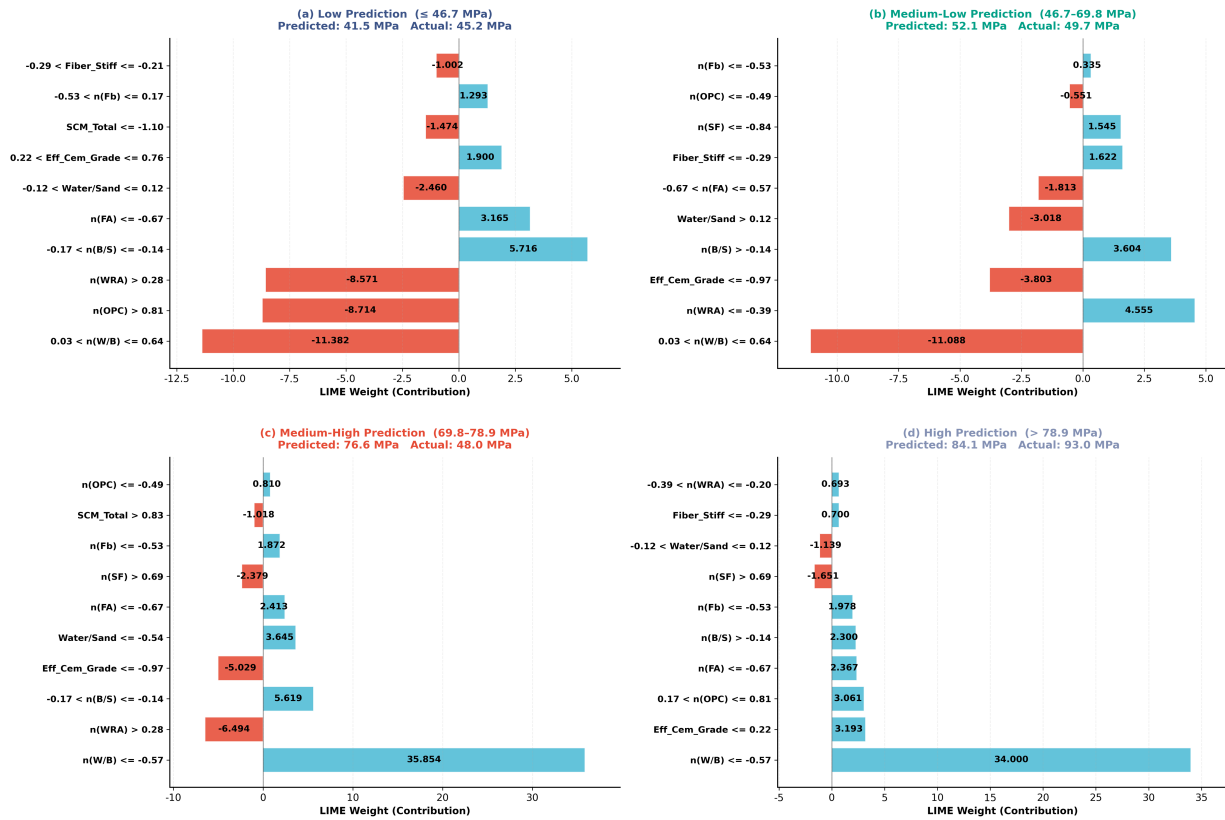


Figure 10: LIME local explanations.

The Spearman correlation between SHAP- and permutation-based rankings is  $\rho = 0.936$ ;  $n(\text{W/B})$  is identified as the dominant predictor by all four methods, and three of the five top-ranked features ( $n(\text{W/B})$ ,  $n(\text{B/S})$ , and Effective\_Cement\_Grade) are shared between the SHAP and ALE hierarchies. The partial divergence between SHAP and ALE rankings ( $\rho = 0.573$ ) for mid-tier features reflects each method's sensitivity to feature correlation structure rather than a physical contradiction.

The four interpretability analyses converge on three actionable design levers for 3DPC CS. First,  $n(\text{W/B})$  is the primary control variable: the ALE profile spans 57.57 MPa across the printable window, with  $\text{W/B} \leq 0.28$  consistently projecting CS above 80 MPa while  $\text{W/B} \geq 0.40$  constrains CS below 55 MPa irrespective of other variables. Second, the OPC binder fraction  $n(\text{OPC})$  governs a secondary CS increment of up to 26 MPa (ALE importance 26.36 MPa); increasing the OPC fraction toward 0.80 delivers meaningful CS gains once W/B is already optimized. Cement strength grade  $f(\text{OPC})$  was not retained by Boruta selection, indicating that the OPC proportion rather than its nominal grade is the actionable lever within the compiled dataset. Third, high-stiffness fiber reinforcement (steel or carbon,  $\text{Fb-E} \geq 100$  GPa) contributes up to +22 MPa via the Fiber\_Stiffness pathway; PVA/PP fibers at comparable volume fractions yield negligible CS benefit for compression-critical elements.

#### 4.6 Cross-System Transfer Evaluation

The trained model is calibrated for OPC- and SAC-based 3DPC within the compiled compositional bounds ( $\text{W/B}$ : 0.15–0.65,  $\text{B/S}$ : 0.1–4.5); application outside this material system requires recalibration on

representative target-system data. To assess this boundary quantitatively, an independent engineered cementitious composite (ECC)/strain-hardening cementitious composite (SHCC) dataset [35] comprising 262 specimens was used for blind external testing. ECC/SHCC was selected because it shares three features with 3DPC, no coarse aggregate, the same OPC-FA-GGBS-SF binder platform, and short-fiber reinforcement, making it the closest available cross-system validation target. The external CS range is 4.0–120.0 MPa (mean 43.6 MPa, SD 20.3 MPa).

Table 9 summarizes the feature availability and mapping between the two datasets.

**Table 9:** Feature availability and mapping between the 3DPC training dataset and the ECC/SHCC external validation dataset.

| Feature Category                     | Training Dataset (3DPC, $n = 254$ )                              | External Dataset (ECC/SHCC, $n = 262$ )               | Mapping Strategy                                                                                        |
|--------------------------------------|------------------------------------------------------------------|-------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| Common—direct                        | $n(W/B)$ , $n(B/S)$ , $n(WRA)$ , $n(Fb-E)$ , $n(Fb-L)$ , $n(Fb)$ | Same quantities, same units                           | Direct use after range check                                                                            |
| Common—unit conversion               | $n(OPC)$ , $n(FA)$ , $n(GGBS)$ , $n(SF)$ as fractions (0–1)      | Same quantities as percentages (0%–100%)              | Divide by 100                                                                                           |
| Training-only                        | $f(OPC)$ (OPC strength grade, 0–52.5 MPa); $f(SAC)$ , $n(SAC)$   | Absent                                                | $f(OPC)$ fixed at 42.5 MPa; $f(SAC) = n(SAC) = 0$                                                       |
| External-only                        | Absent                                                           | Nine individual fiber-type columns; curing conditions | Fiber volumes summed to $n(Fb)$ ; volume-weighted modulus mapped to $n(Fb-E)$ ; curing fields discarded |
| Out-of-range (clipped in Strategy B) | $n(Fb-E)$ : 0–200 GPa; $n(Fb)$ : 0%–4.0%                         | $n(Fb-E)$ : 4–684 GPa; $n(Fb)$ : 0.5%–7.5%            | Cap at training maxima                                                                                  |

Three evaluation strategies are examined (Table 10). Strategy A applies only unit conversion to all 262 samples. Strategy B additionally caps out-of-distribution fiber inputs at training maxima ( $n(Fb-E)$  at 200 GPa,  $n(Fb)$  at 4.0%) to prevent extrapolation from ECC-specific carbon fibers (up to 684 GPa) and higher fiber loadings (up to 7.5%). Strategy C restricts evaluation to the 202 samples within the training feature bounds. Together, Strategies A through C constitute a progressive domain-adaptation analysis: each step applies an additional layer of input-space recalibration, enabling the contribution of feature-space mismatch to be separated from that of target-variable distribution shift.

**Table 10:** External validation performance by mapping strategy.

| Strategy                               | $n$ | $R^2$  | RMSE (MPa) | MAE (MPa) | MAPE (%) | Pearson $R$ |
|----------------------------------------|-----|--------|------------|-----------|----------|-------------|
| A: Unit conversion only                | 262 | −0.336 | 23.44      | 18.88     | 75.70    | 0.314       |
| B: Unit conversion + fiber clipping    | 262 | −0.246 | 22.63      | 18.24     | 74.08    | 0.347       |
| C: Within-training-distribution subset | 202 | −0.095 | 21.13      | 16.36     | 60.46    | 0.301       |

The progressive improvement from Strategy A ( $R^2 = -0.336$ ) to Strategy B ( $R^2 = -0.246$ ) to Strategy C ( $R^2 = -0.095$ ) isolates the contribution of each recalibration layer: input-space adaptation (fiber clipping) and domain filtering each reduce the transfer error, yet a systematic overestimation of 12.0 MPa persists across all three strategies. This stability confirms that the residual gap is attributable to target-variable distribution shift rather than feature-space extrapolation, and is explained by three material-system boundaries: (1) a 20 MPa mean CS offset between 3DPC (mean 64.1 MPa) and ECC (mean 43.6 MPa); (2) fiber-type incompatibility, as ECC predominantly employs PVA and PE fibers (25–170 GPa) outside the steel-fiber-dominated 3DPC training distribution; and (3) low ECC binder-to-sand ratios (typically below 1.0), outside the high-paste 3DPC regime (B/S up to 4.5). The preserved Pearson correlation ( $R = 0.347$ , Strategy B; Fig. 11) confirms that the model maintains the relative strength ordering within ECC, consistent with the dominant role of  $n(W/B)$  identified in Section 4.5; for a practitioner adapting this framework to ECC/SHCC, this stable 12.0 MPa offset constitutes a first-order recalibration constant, with the preserved rank-ordering capacity enabling mix-design screening prior to full recalibration on target-system data. Bland-Altman analysis (Fig. 12) shows 93.5% of samples within the 95% limits of agreement  $[-49.6, +25.7]$  MPa. Full recalibration with target-system training data is required to achieve within-domain accuracy; the present cross-system analysis delineates the applicability boundary and quantifies the systematic offset that any such recalibration must address.

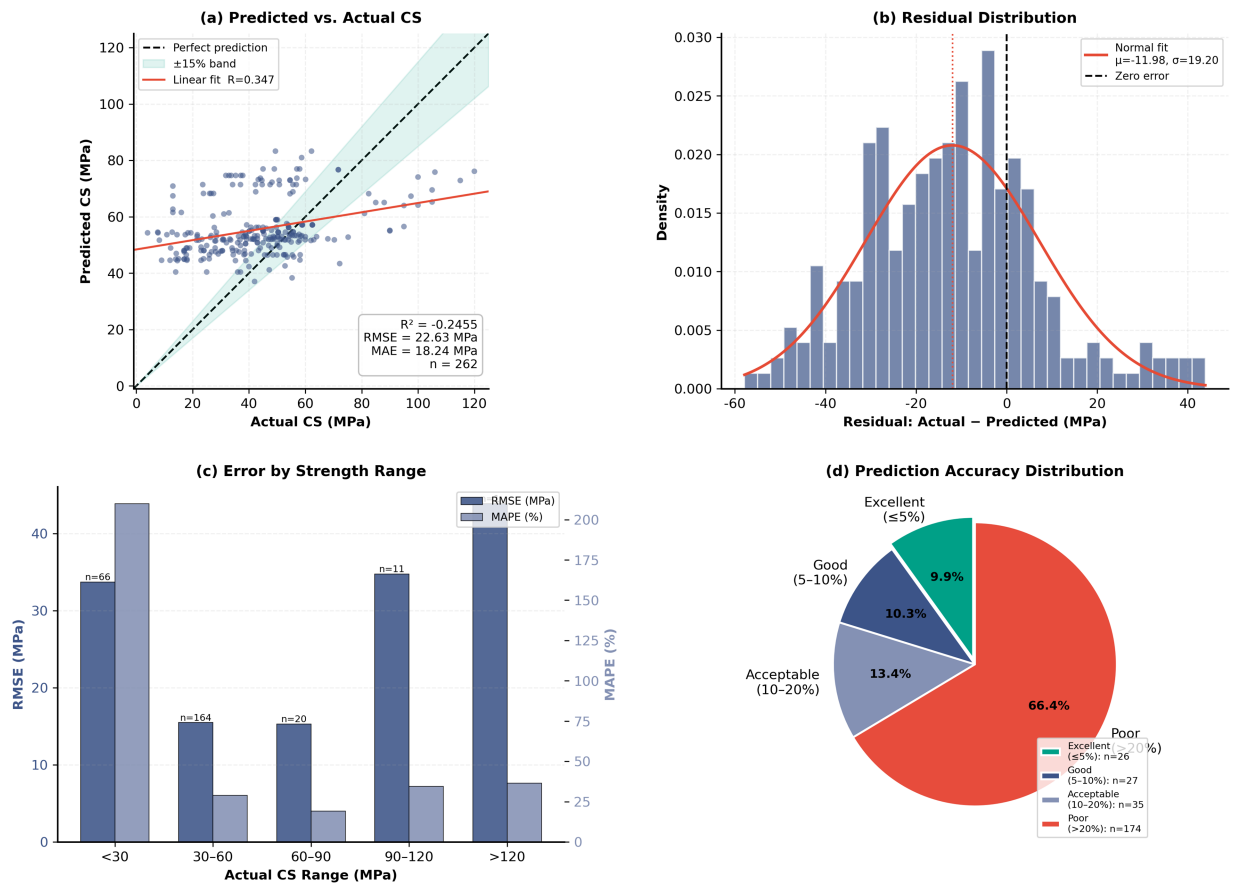
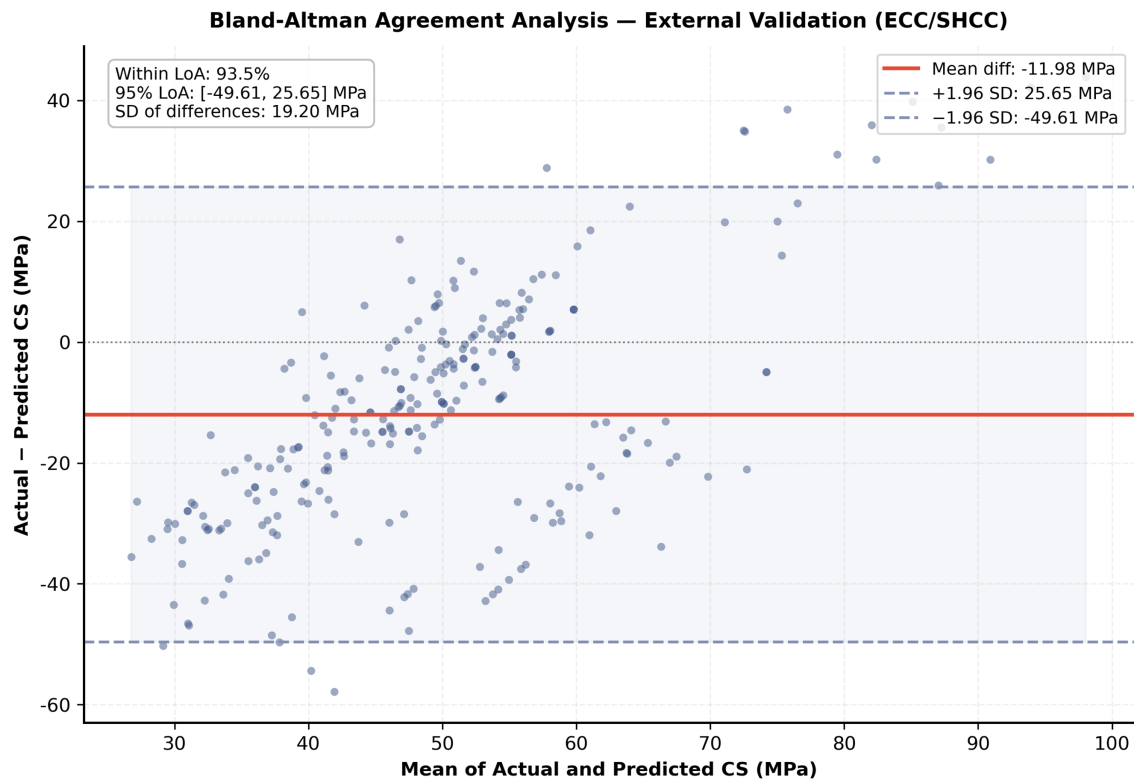


Figure 11: External validation analysis on the ECC/SHCC dataset.





**Figure 12:** Bland-Altman agreement plot for the ECC/SHCC external validation.

To demonstrate a practical recalibration workflow, the 262 ECC/SHCC specimens were partitioned into a calibration set and an independent held-out test set at four size ratios (10%, 20%, 30%, 40%); 20 independent random splits were evaluated at each ratio to avoid dependence on a single partition. Two correction methods were applied to the uncalibrated Strategy B predictions: bias correction, which subtracts the mean residual estimated from the calibration set; and linear recalibration, which fits an ordinary least-squares intercept and slope to the calibration set and applies the resulting correction to the test set. Both methods use only calibration set observations and make no use of the test set during parameter estimation.

**Table 11** reports the results. Bias correction reduces the systematic overestimation from +11.98 MPa to below +1.4 MPa with as few as 26 calibration specimens (10%), confirming that the offset identified above constitutes an estimable recalibration constant. Linear recalibration achieves comparable bias reduction but does not improve RMSE or MAE over the simpler offset correction, consistent with the predominance of translational rather than multiplicative error between the two material systems.

The explained variance nonetheless remains low across all conditions ( $R^2 = 0.057\text{--}0.090$  for bias correction;  $R^2 = 0.025\text{--}0.083$  for linear recalibration), indicating that removing the systematic offset does not recover the composition–strength rank ordering within the ECC/SHCC domain. The large standard deviations of  $R^2$  ( $\pm 0.056\text{--}\pm 0.129$ ) reflect sensitivity to which specimens enter the calibration set, a consequence of the broad ECC/SHCC compositional space relative to the 3DPC training distribution. Increasing the calibration set from 10% to 40% yields only marginal RMSE improvement (19.61 to 19.33 MPa for bias correction), indicating that the residual scatter is not reducible by adding more calibration data within a two-parameter correction framework. These results quantify the limit of post-hoc offset calibration and confirm that achieving within-domain accuracy comparable to the 3DPC training performance requires retraining the model on representative ECC/SHCC data.

**Table 11:** Practical recalibration performance on the ECC/SHCC external dataset (262 specimens). At each calibration ratio, 20 independent random splits were performed; metrics are reported as mean  $\pm$  SD on the held-out test set. The uncalibrated baseline corresponds to strategy B in Table 10.

| Cal. Set          | Method          | $n_{cal}$ | $n_{test}$ | $R^2$             | RMSE (MPa)       | MAE (MPa)        | MAPE (%)       | Bias (MPa)      |
|-------------------|-----------------|-----------|------------|-------------------|------------------|------------------|----------------|-----------------|
| Uncalibrated      | —               | 0         | 262        | -0.246            | 22.63            | 18.24            | 74.1           | +11.98          |
| 10% ( $n = 26$ )  | Bias correction | 26        | 236        | $0.057 \pm 0.064$ | $19.61 \pm 0.57$ | $15.41 \pm 0.49$ | $53.7 \pm 6.0$ | $0.03 \pm 4.22$ |
|                   | Linear          | 26        | 236        | $0.025 \pm 0.129$ | $19.91 \pm 1.04$ | $15.42 \pm 1.01$ | $54.2 \pm 6.1$ | $0.24 \pm 3.97$ |
| 20% ( $n = 52$ )  | Bias correction | 52        | 210        | $0.070 \pm 0.056$ | $19.53 \pm 0.60$ | $15.34 \pm 0.51$ | $55.8 \pm 4.5$ | $1.36 \pm 3.56$ |
|                   | Linear          | 52        | 210        | $0.060 \pm 0.060$ | $19.63 \pm 0.65$ | $15.15 \pm 0.61$ | $56.2 \pm 4.6$ | $1.45 \pm 3.65$ |
| 30% ( $n = 79$ )  | Bias correction | 79        | 183        | $0.076 \pm 0.102$ | $19.37 \pm 0.75$ | $15.16 \pm 0.61$ | $53.6 \pm 6.5$ | $0.70 \pm 3.37$ |
|                   | Linear          | 79        | 183        | $0.063 \pm 0.104$ | $19.51 \pm 0.92$ | $14.96 \pm 0.72$ | $53.8 \pm 6.5$ | $0.54 \pm 3.45$ |
| 40% ( $n = 105$ ) | Bias correction | 105       | 157        | $0.090 \pm 0.077$ | $19.33 \pm 0.67$ | $15.15 \pm 0.56$ | $53.1 \pm 6.0$ | $0.41 \pm 2.43$ |
|                   | Linear          | 105       | 157        | $0.083 \pm 0.075$ | $19.41 \pm 0.83$ | $14.88 \pm 0.59$ | $53.2 \pm 6.0$ | $0.23 \pm 2.47$ |

#### 4.7 GUI Development

A standalone graphical user interface (GUI) was implemented to make the calibrated CatBoost model accessible to mix designers without programming expertise. The interface accepts 13 mix design parameters organized into four color-coded panels covering binder composition, supplementary cementitious materials, mix proportions, and fiber reinforcement (Fig. 13). Boruta-selected predictors are marked with a star symbol, communicating directly which inputs carry the greatest predictive weight. The 30-run mean  $R^2 = 0.897 \pm 0.051$  and single-run test metrics (RMSE = 8.82 MPa, MAPE = 9.68%) are displayed persistently in the title banner as a quantitative reliability reference.

**3D Printed Concrete | Compressive Strength Predictor**  $R^2: 0.897 \pm 0.051$  RMSE 8.82 MPa MAPE 9.68%

**Binder Composition**

OPC Grade (MPa): 52.5    ★ OPC Fraction (0-1): 0.75    SAC Grade (MPa): 0    SAC Fraction (0-1): 0

**Supplementary CM**

★ Fly Ash (0-1): 0.10    GGBS (0-1): 0.10    ★ Silica Fume (0-1): 0.05

**Mix Proportions**

★ W/B Ratio (—): 0.16    ★ B/S Ratio (—): 1.0    ★ WRA Dosage (% binder): 0.4

**Fiber Reinforcement**

Fiber Modulus (GPa): 0    Fiber Length (mm): 0    ★ Fiber Vol. Frac. (%): 0

★ Boruta-selected feature (11/17 used by model)

**PREDICT** **EXAMPLE** **CLEAR**

**Predicted Compressive Strength: 137.9 MPa**

**Figure 13:** Graphical user interface for 3DPC compressive strength prediction.

The typical workflow is: (1) enter the target mix proportions within the labelled input ranges; (2) read the predicted CS from the color-coded output panel (green: CS < 80 MPa; orange: 80–120 MPa; red: >120 MPa); (3) iterate W/B and cement grade, the two primary levers identified in Section 4.5, to converge on the desired strength class while monitoring the uncertainty bounds. For a reference mix drawn from the compiled dataset, the GUI returns a predicted CS of 137.9 MPa against a recorded value of 130.6 MPa, a point error of 5.6%, within the model's overall MAPE. The GUI is publicly available at <https://github.com/lucassivan/ML-3DPC>.

## 5 Conclusions

A machine learning framework for 3D-printed concrete compressive strength prediction was developed and benchmarked across 20 algorithms on 254 experimental records. Five principal conclusions emerge.

- (1) CatBoost is the recommended algorithm for 3DPC CS prediction: 30-run mean  $R^2 = 0.8968 \pm 0.0505$ , with single-run (seed-42) RMSE = 8.82 MPa and MAPE = 9.68%. XGBoost is the only model not statistically separated from CatBoost after Bonferroni correction ( $p = 0.085$ ,  $d = 0.08$ ), though the Bayesian Bayes factor indicates moderate evidence for a small CatBoost advantage ( $BF_{10} = 9.23$ ); together these identify a family-level rather than model-level winner. The Stacking discordance ( $p < 0.001$  yet  $BF_{10} = 0.02$ ) illustrates that  $p$ -values alone are unreliable for practical model selection at this dataset scale.
- (2) Composite features contribute signal beyond raw inputs. Fiber\_Stiffness ranks third in ALE importance (24.81 MPa): although its constituent  $n(\text{Fb})$  was individually confirmed by Boruta, the composite additionally encodes the fiber elastic modulus  $n(\text{Fb-E})$ —rejected by Boruta on its own—capturing a volumetric-stiffness interaction that neither raw variable represents in isolation. Fiber type and volume fraction should therefore be specified jointly rather than independently in 3DPC mix design.
- (3)  $n(\text{W/B})$  governs CS with a non-monotonic ALE profile spanning 57.57 MPa across the printable window. For mix design,  $\text{W/B} \leq 0.28$  consistently targets CS above 80 MPa; each 0.05 increase in W/B within the printable window is associated with a mean CS reduction of approximately 15 MPa.
- (4) Prediction uncertainty is well-characterized across four quantification methods; the 95% prediction interval half-width of 12 MPa is recommended as a minimum safety margin when using model predictions as the nominal CS target in structural design.
- (5) The model is well-calibrated for OPC- and SAC-based 3DPC within W/B 0.15–0.65, where MAPE  $\leq 9.68\%$  is expected. Cross-system transfer to ECC/SHCC yields a 12.0 MPa systematic overestimation attributable to material-system differences rather than model overfitting. A bias correction using as few as 26 target-domain specimens reduces this offset to near zero, but  $R^2$  remains below 0.09 (Table 11), confirming that full retraining on representative ECC/SHCC data is required for within-domain accuracy.

Process parameters (nozzle geometry, printing speed, layer height, interlayer time, and extrusion rate) are absent from the predictor set because they are inconsistently reported across the 24 source studies constituting the compiled dataset; this is a data-availability constraint rather than a modelling choice. Variability in CS attributable to cross-study printing procedure differences is consequently pooled into the model residuals rather than captured by dedicated features, making the model composition-conditional rather than process-aware. Prospective work should integrate printing process parameters (layer height, extrusion rate, interlayer delay), expand training coverage above 150 MPa where current data are sparse, and apply multi-objective optimization to jointly maximize CS, printability, and carbon efficiency.

**Acknowledgement:** Not applicable.

**Funding Statement:** The authors received no specific funding for this study.

**Author Contributions:** The authors confirm contribution to the paper as follows: conceptualization, Jia Chen and Jianbo Huang; methodology, Jia Chen; software, Zhicheng Liao; validation, Jia Chen and Mengdi Hou; formal analysis, Mengdi Hou; investigation, Jia Chen and Zhicheng Liao; resources, Jianbo Huang; data curation, Mengdi Hou; writing—original draft preparation, Jia Chen; writing—review and editing, Jia Chen, Zhicheng Liao, Mengdi Hou and Jianbo Huang; visualization, Zhicheng Liao; funding acquisition, Jianbo Huang; supervision, Jianbo Huang; project administration, Jianbo Huang. All authors reviewed and approved the final version of the manuscript.

**Availability of Data and Materials:** The data and source codes that support the findings of this study are openly available on GitHub at <https://github.com/lucassivan/ML-3DPC>.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Hassan A, Alomayri T, Noaman MF, Zhang C. 3D printed concrete for sustainable construction: a review of mechanical properties and environmental impact. *Arch Computat Methods Eng*. 2025;32(5):2713–43. doi:10.1007/s11831-024-10220-5.
2. Girskas G, Kligys M. 3D concrete printing review: equipment, materials, mix design, and properties. *Buildings*. 2025;15(12):2049. doi:10.3390/buildings15122049.
3. Garcés G, García-Alvarado R, Bunster V, Muñoz-Sanguinetti C. Additive construction 4.0: a systematic review of 3D concrete printing for construction 4.0. *Eng Constr Archit Manag*. 2025;33(8):6539–68. doi:10.1108/ECAM-07-2024-0875.
4. Zhang Y, Zhu X, Li M, Zhang C, Zhang Y, Du X, et al. 3D printing technology in concrete construction. *Nat Rev Clean Technol*. 2025;1(4):288–303. doi:10.1038/s44359-025-00047-z.
5. Liu H, Wang Y, Zhu C, Wu Y, Liu C, He C, et al. Design of 3D printed concrete masonry for wall structures: mechanical behavior and strength calculation methods under various loads. *Eng Struct*. 2025;325(S1):119374. doi:10.1016/j.engstruct.2024.119374.
6. Iqbal I, Kasim T, Besklubova S, Inqiad WB, Nowakowski DJ, Rahman M. Exploring knowledge domains and future research directions in 3D printed concrete: a bibliometric and systematic review. *Innov Infrastruct Solut*. 2026;11(1):35. doi:10.1007/s41062-025-02396-x.
7. Nan B, Qiao Y, Leng J, Bai Y. Advancing structural reinforcement in 3D-printed concrete: current methods, challenges, and innovations. *Materials*. 2025;18(2):252. doi:10.3390/ma18020252.
8. Sapata A, Šinka M, Šahmenko G, Korat Bensa L, Hanžič L, Šter K, et al. Establishing benchmark properties for 3D-printed concrete: a study of printability, strength, and durability. *J Compos Sci*. 2025;9(2):74. doi:10.3390/jcs9020074.
9. Gu Y, Zheng S, Ma H, Long WJ, Khayat KH. Effect of absorption kinetics of superabsorbent polymers on printability and interlayer bond of 3D printing concrete. *Cem Concr Compos*. 2024;151(3):105609. doi:10.1016/j.cemconcomp.2024.105609.
10. Gao H, Jin L, Chen Y, Chen Q, Liu X, Yu Q. Rheological behavior of 3D printed concrete: influential factors and printability prediction scheme. *J Build Eng*. 2024;91:109626. doi:10.1016/j.jobe.2024.109626.
11. Zhu R, Egbe KJI, Salehi H, Shi Z, Jiao P. Eco-friendly 3D printed concrete with fine aggregate replacements: fabrication, characterization and machine learning prediction. *Constr Build Mater*. 2024;413(22):134905. doi:10.1016/j.conbuildmat.2024.134905.

12. Malik UJ, Riaz RD, Rehman SU, Usman M, Riaz RE, Hamza R. Advancing mix design prediction in 3D printed concrete: predicting anisotropic compressive strength and slump flow. *Case Stud Constr Mater.* 2024;21:e03510. doi:10.1016/j.cscm.2024.e03510.
13. Alyami M, Khan M, Fawad M, Nawaz R, Hammad AWA, Najeh T, et al. Predictive modeling for compressive strength of 3D printed fiber-reinforced concrete using machine learning algorithms. *Case Stud Constr Mater.* 2024;20(9):e02728. doi:10.1016/j.cscm.2023.e02728.
14. Li J, Wu L, Huang Z, Xu Y, Liu K. Machine learning-driven prediction of mechanical properties for 3D printed concrete. *Comput Concr.* 2025;35(3):263–79. doi:10.12989/CAC.2025.35.3.263.
15. Uddin MN, Ye J, Deng B, Li L, Yu K. Interpretable machine learning for predicting the strength of 3D printed fiber-reinforced concrete (3DP-FRC). *J Build Eng.* 2023;72(24):106648. doi:10.1016/j.jobbe.2023.106648.
16. Zhang Y, Cui S, Yang B, Wang X, Liu T. Research on 3D printing concrete mechanical properties prediction model based on machine learning. *Case Stud Constr Mater.* 2025;22(21):e04254. doi:10.1016/j.cscm.2025.e04254.
17. Iqbal I, Kasim T, Besklubova S, Mustafa A, Rahman M, Alabduljabbar H, et al. Passive determination of anisotropic compressive strength of 3D printed concrete using multiple neural networks enhanced with explainable machine learning (XML). *Sci Rep.* 2025;15(1):43885. doi:10.1038/s41598-025-11068-w.
18. Khodadadi N, Roghani H, De Caso F, El-kenawy EM, Yesha Y, Nanni A. Machine learning approach for the flexural strength of 3D-printed fiber-reinforced concrete based on the meta-heuristic algorithm. *Struct Concr.* 2025;26(4):4103–42. doi:10.1002/suco.70195.
19. Katlav M, Turk K. Explainable hybrid machine learning approach for mechanical performance of 3D-printed strain-hardening cementitious composites (3DP-SHCC). *J Sustain Cem Based Mater.* 2026;2026(7):1–31. doi:10.1080/21650373.2026.2646341.
20. Liu S, Liu T, Alqurashi M, Abdou Elabbasy AA, Alanazi N, Shakor P. Advancing 3D-printed fiber-reinforced concrete for sustainable construction: a comparative optimization based study of hybrid machine intelligence models for predicting mechanical strength and CO<sub>2</sub> emissions. *Case Stud Constr Mater.* 2025;23(12):e05259. doi:10.1016/j.cscm.2025.e05259.
21. Izadgoshasb H, Kandiri A, Shakor P, Laghi V, Gasparini G. Predicting compressive strength of 3D printed mortar in structural members using machine learning. *Appl Sci.* 2021;11(22):10826. doi:10.3390/app112210826.
22. Müteveli Özkan İG, Aldemir A. Machine-learning networks to predict the ultimate axial load and displacement capacity of 3D printed concrete walls with different section geometries. *Structures.* 2024;66:106879. doi:10.1016/j.istruc.2024.106879.
23. Abid KA, Syed SA, Khan M. Explainable machine learning-based model for predicting interlayer bond strength in 3D printed concrete. *Multiscale Multidiscip Model Exp Des.* 2025;8(9):415. doi:10.1007/s41939-025-00997-8.
24. Ma XR, Wang XL, Chen SZ. Trustworthy machine learning-enhanced 3D concrete printing: predicting bond strength and designing reinforcement embedment length. *Autom Constr.* 2024;168(5):105754. doi:10.1016/j.autcon.2024.105754.
25. Syed SA, Abid KA, Khan M. An interpretable machine learning approach for predicting reinforcement bond performance in 3D concrete printing. *Multiscale Multidiscip Model Exp Des.* 2025;8(10):435. doi:10.1007/s41939-025-01028-2.
26. Wang X, Banthia N, Yoo DY. Reinforcement bond performance in 3D concrete printing: explainable ensemble learning augmented by deep generative adversarial networks. *Autom Constr.* 2024;158:105164. doi:10.1016/j.autcon.2023.105164.
27. Hematibahar M, Kharun M, Fediuk R, Vatin NI, Porvadov MG, Sabitov LS. Predicting the flexural strength of 3D-printed geopolymers reinforced concrete using machine learning techniques. *Mag Civ Eng.* 2025;53(4):22–34. doi:10.18149/MPM.5342025\_2.
28. Ali A, Riaz RD, Malik UJ, Abbas SB, Usman M, Shah MU, et al. Machine learning-based predictive model for tensile and flexural strength of 3D-printed concrete. *Materials.* 2023;16(11):4149. doi:10.3390/ma16114149.
29. Iqbal I, Bin Inqiad W, Kasim T, Besklubova S, Adil MM, Rahman M. Strength characterisation of fly ash blended 3D printed concrete enhanced with explainable machine learning. *Case Stud Constr Mater.* 2026;24(2):e05682. doi:10.1016/j.cscm.2025.e05682.

30. Asif U. Comparative analysis of evolutionary computational methods for predicting mechanical properties of fiber-reinforced 3D printed concrete. *Innov Infrastruct Solut.* 2025;10(6):259. doi:10.1007/s41062-025-02057-z.
31. Uddin MN, Ye J, Haque MA, Yu K, Li L. A novel compressive strength estimation approach for 3D printed fiber-reinforced concrete: integrating machine learning and gene expression programming. *Multiscale Multidiscip Model Exp Des.* 2024;7(5):4889–910. doi:10.1007/s41939-024-00439-x.
32. Arif M, Jan F, Rezzoug A, Afridi MA, Luqman M, Khan WA, et al. Data-driven models for predicting compressive strength of 3D-printed fiber-reinforced concrete using interpretable machine learning algorithms. *Case Stud Constr Mater.* 2024;21(4):e03935. doi:10.1016/j.cscm.2024.e03935.
33. Albostami AS, Mohammad M, Ismael BH, Al-Hamd RKS. Optimized strength predictions for 3D printed fiber-reinforced concrete: machine learning-driven insights. *Structures.* 2025;82:110400. doi:10.1016/j.istruc.2025.110400.
34. Abbas YM, Alsaif A. Explainable data-driven modeling for optimized mix design of 3D-printed concrete: interpreting nonlinear synergies among binder components and proportions. *Comput Model Eng Sci.* 2025;145(2):1789–819. doi:10.32604/cmes.2025.073088.
35. Altayeb M. Engineered cementitious composites dataset/strain hardening cementitious composites data for FDNN ensemble predictive model [Internet]. Mendeley; 2021 [cited 2026 Apr 28]. Available from: <https://data.mendeley.com/datasets/584ydx3xm/1>.