Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.086222

ARTICLE

Interpretable Machine Learning for Compressive and Flexural Strength Prediction of Fly Ash Blended 3D Printed Concrete with Uncertainty Quantification

Jia Chen[1], **Zhicheng Liao**[1], **Mengdi Hou**[2] **and Jianbo Huang**[1,3,*]

[1]School of Computer Applications, Guilin University of Technology, Guilin, China
[2]Guangxi Key Laboratory of Machine Vision and Intelligent Control, Wuzhou University, Wuzhou, China
[3]Centrale Méditerranée, Technopôle de Château-Gombert, 38 rue Frédéric Joliot-Curie, Marseille, France

*Corresponding Author: Jianbo Huang. Email: jianbo.huang@centrale-med.fr

Received: 29 May 2026; Accepted: 24 June 2026; Published: 13 August 2026

ABSTRACT: Fly ash (FA) blended 3D printed concrete (3DPC) offers improved sustainability but requires strength prediction models validated at the mix-composition level rather than within familiar formulations. This study applies leave-one-mix-out (LOMO) cross-validation to benchmark eight machine learning algorithms on 126 experimental records spanning seven FA-blended 3DPC compositions (FA 0–15 wt.%, W/B 0.30–0.35, age 1–28 days). ExtraTrees and ElasticNet achieve the highest composition-level generalisation for compressive strength (CS, R^2 = 0.786 ± 0.253) and flexural strength (FS, R^2 = 0.915 ± 0.061, RMSE = 0.301 MPa), respectively. SHAP analysis identifies FA replacement percentage as the dominant CS predictor (mean |SHAP| = 3.92 MPa, negative effect) and curing age as the dominant FS predictor (1.06 MPa), a reversal consistent with distinct failure mechanisms under compression and flexure. FA substitution at 5%–7.5% with W/B = 0.30 is identified as the optimal design range, with content limited to 5% when structural loading within three days is required. LOMO residual prediction intervals are well calibrated from 50% to 99% nominal coverage, with empirical deviations not exceeding 1.4 percentage points; at the 90% level, interval half-widths are ±4.60 MPa for CS and ±0.494 MPa for FS. Design-space mapping over the FA% × age prediction grid at W/B = 0.30 identifies minimum curing ages of 3 days at FA ≤ 7.5 wt.% and 13.5 days at FA = 10 wt.% for simultaneous CS and FS target compliance; Sobol variance decomposition confirms FA% (S_1 = 0.553) and curing age (S_1 = 0.526) as the dominant sensitivity drivers for CS and FS, respectively, in close agreement with the SHAP attribution. A graphical user interface integrating the two best models returns simultaneous CS and FS predictions with 90% calibrated prediction intervals from three mix design inputs, supporting uncertainty-aware strength estimation without programming expertise. Both models are local interpolators conditioned on the experimental design space (FA 0–15 wt.%, W/B 0.30–0.35, age 1–28 days); predictions outside these boundaries have not been validated and should be interpreted with caution. The dataset and source code are publicly available at https://github.com/lucassivan/ML-FA-3DPC.

KEYWORDS: 3D printed concrete; fly ash; leave-one-mix-out cross-validation; machine learning; SHAP; prediction intervals

1 Introduction

Three-dimensional printed concrete (3DPC) has gained traction as a formwork-free, low-waste construction method capable of fabricating complex geometries with high dimensional accuracy [1]. Extrusion-based layer deposition eliminates manual formwork, reduces material waste, and shortens project timelines, attributes that make 3DPC competitive with conventional casting for both residential and infrastructure applications [2,3]. A persistent challenge, however, is the formulation of printable mortars that balance extrudability, shape retention, and structural performance simultaneously [1,4,5].

Ordinary Portland cement (OPC) constitutes 15%–45% of 3DPC mortar by weight and contributes a disproportionate share of its embodied carbon [1,5]. Partial replacement of OPC with fly ash (FA), a coal combustion by-product rich in reactive silica and alumina, lowers CO_2 intensity while refining paste microstructure through pozzolanic gel formation and spherical particle packing [6–8]. FA also reduces water demand and improves workability when combined with appropriate superplasticiser dosages [9–11], though slower pozzolanic reactivity suppresses early-age strength relative to OPC-only controls, with recovery largely complete by 28 days [7,8,12–14].

FA's influence on 3DPC extends across fresh and hardened states. In the fresh state, FA incorporation retards stiffening and extends open time, a property beneficial to layer adhesion in continuous print runs [15–18]; Kaya et al. [19] quantified this tradeoff, showing that FA at 60% replacement reduces green strength by 35.4% whereas an equivalent slag content raises it by 163.41%. In the hardened state, Mishra et al. [20] reported that a 70% OPC and 30% GGBS blend achieved 18.3% higher shape retention and 18.5% lower permeable porosity than the OPC control; broader work confirms that FA and GGBFS at 20%–30% substitution improve 28-day compressive and flexural strength, with the magnitude governed by replacement level, W/B ratio, and curing duration [21–24].

Machine learning (ML) offers a route to systematic strength prediction without the resource cost of exhaustive mix trials, and ensemble regressors have demonstrated strong performance across diverse concrete formulations [25–27]. For 3DPC, several recent studies have predicted compressive and flexural strength from mix composition [28–34]; Li et al. [34] achieved R^2 of 0.929 and 0.967 for the two targets using CatBoost with SHAP attribution, while Zhang et al. [32] and Schossler et al. [30] reported comparable accuracy using random forest and Bayesian-optimised regressors. These studies share a cross-validation design in which specimens from the same mix appear in both training and test partitions, so reported accuracy reflects age-axis interpolation for familiar compositions rather than generalisation to unseen formulations. Consequently, the $R^2 > 0.997$ reported by Iqbal et al. [35] may overstate generalisation capability due to same-mix data leakage. Prediction uncertainty is equally unaddressed: without calibrated prediction intervals, point estimates alone provide no basis for risk-aware mix design decisions. The mechanistic contrast between compressive and flexural strength responses to FA content and curing age has also not been characterised at the feature-attribution level for FA-blended 3DPC.

This study addresses these gaps using 126 experimental records from seven FA-blended 3DPC compositions [35] (FA 0–15 wt.%, W/B 0.30–0.35). Leave-one-mix-out (LOMO) cross-validation withholds all records from one complete mix composition per fold, testing whether a model trained on known compositions can predict an unseen formulation; this directly resolves the data-leakage limitation identified above. Eight regression algorithms are benchmarked under this stricter protocol, with ExtraTrees and ElasticNet emerging as optimal for CS and FS, respectively, and the divergent rankings are explained through Sobol sensitivity indices. Distribution-free prediction intervals built from LOMO out-of-mix residuals provide empirically calibrated uncertainty bounds, filling the gap in risk-aware mix design support. Combined SHAP attribution and Sobol variance decomposition characterise the mechanistic contrast between CS and FS responses to FA content and curing age at the feature level; results are translated into design-space compliance maps and a graphical user interface for practical strength estimation.

2 Dataset and Preliminary Analysis

2.1 Dataset Description

The dataset originates from a controlled single-laboratory study on extrusion-based 3D-printed concrete in which fly ash (FA) partially replaces ordinary Portland cement at five substitution levels (0, 5, 7.5, 10, and 15 wt.%) and the water-to-binder ratio (W/B) is fixed at three levels (0.30, 0.34, and 0.35) [35]. These seven mix proportions were each cast and evaluated at six standard curing ages (1, 3, 7, 14, 21, and 28 days) with triplicate specimens per age point, giving 126 records in total. Both compressive strength (CS) and flexural strength (FS) were measured on every specimen, enabling simultaneous dual-target regression rather than independent single-output models.

Sand content (1000 g per batch) and accelerator dosage (1.4 wt.%) are invariant across all records and are excluded from the predictor set. Among the remaining variables, FA replacement percentage and total cement content are perfectly collinear by construction ($r = -1.00$); only FA% is retained as the more interpretable descriptor. Superplasticizer (SP) was applied solely to the W/B = 0.30 group (SP = 0.3% of binder mass) and set to zero for both W/B = 0.34 and 0.35 formulations, producing near-perfect anti-correlation between W/B and SP ($r = -0.991$). The effective predictor space therefore contains three variables (FA%, W/B, and curing age), with SP providing a partially redundant fourth signal.

Table 1 reports the range, central tendency, and zero-entry frequency for all retained variables.

Table 1: Statistical summary of input and output variables.

Variable	Description	Unit	Range	Mean	SD	Zero (%)
FA (%)	FA replacement level	wt.%	0–15.0	5.36	5.44	42.9
W/B	Water-to-binder ratio	—	0.30–0.35	0.313	0.021	—
SP (%)	Superplasticizer dosage	% binder	0–0.30	0.214	0.136	28.6
Age	Curing age	days	1–28	12.3	9.76	—
CS	Compressive strength (target)	MPa	25.15–61.50	40.23	8.19	—
FS	Flexural strength (target)	MPa	3.60–9.10	5.56	1.29	—

The CS range (25.15–61.50 MPa) and FS range (3.60–9.10 MPa) are consistent with the moderate-to-high strength window accessible to printable paste formulations operating within the W/B range studied. The dataset's fully balanced structure (18 records per mix, 3 replicates per age per mix) is well-suited to Leave-One-Mix-Out (LOMO) cross-validation, in which each fold withholds one complete mix composition and tests generalisation to an unseen binder–W/B combination (Section 3.2).

2.2 Exploratory Data Analysis

Fig. 1 presents the marginal distributions of all input variables and both strength targets as histograms with kernel density estimate overlays. CS is approximately symmetric (Q1–Q3: 32.9–46.2 MPa), with modest right-skew above 55 MPa corresponding to the lowest-W/B, zero-FA mixes. FS is comparably unimodal, concentrated between 4.5 and 6.5 MPa, with a pronounced right tail driven by the same high-performing mix group. Age is uniformly distributed across six discrete points. FA% is right-skewed: the zero-FA control formulations account for 42.9% of all records, with the remaining four substitution levels equally represented.

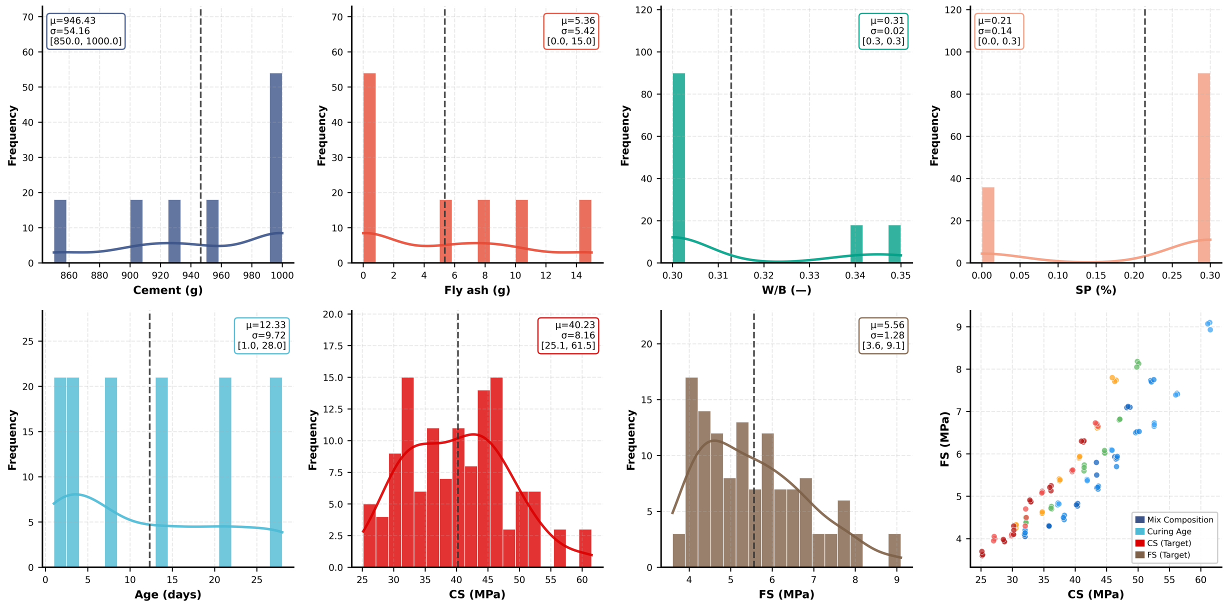

Figure 1: Marginal distributions of input variables and strength targets. Histograms with KDE overlays; dashed vertical lines indicate the mean; μ, σ, and [min, max] are shown in each panel. Bottom-right panel: CS vs. FS scatter plot coloured by mix design.

Fig. 2 presents the Pearson correlation matrix for all four input variables and both target outputs. Curing age is the dominant predictor of both CS ($r = +0.731$, $p < 0.001$) and FS ($r = +0.870$, $p < 0.001$). The stronger age–FS coupling reflects the particular sensitivity of bending failure to continued paste densification and pore-structure refinement: as hydration products fill capillary pores, tensile-governed crack initiation under flexural loading is progressively retarded to a greater degree than axial crushing resistance. FA% ranks second: $r = -0.573$ ($p < 0.001$) against CS and $r = -0.311$ ($p < 0.001$) against FS. W/B and SP do not individually reach significance against either target ($|r| < 0.17$, $p > 0.058$); the narrow W/B window (0.30–0.35) imposed by printability constraints limits the capillary porosity variation attributable to water content alone. CS and FS are tightly coupled ($r = +0.922$, $p < 0.001$), motivating their joint rather than independent prediction.

Fig. 3 traces mean strength development by mix design. Across all seven mixes, mean CS rises from 30.9 MPa at day 1 to 49.0 MPa at day 28 (+58.5%) and mean FS from 4.20 to 7.53 MPa (+79.3%); FS is proportionally more age-sensitive in all mixes. Within the FA-substituted series at W/B = 0.30, 28-day CS declines from 61.3 MPa (FA0) through 49.9 MPa (FA5) to 41.2 MPa (FA15%), with FS following the same rank order (9.0 to 6.3 MPa). Reducing W/B from 0.35 to 0.30 in the FA0 series raises 28-day CS by 12.9 MPa (+26.6%) and FS by 1.9 MPa (+27.2%). Early-age gain (days 1–7) is steeper than the 14–28-day window for both targets, with the rate differential between mixes greatest at day 7.

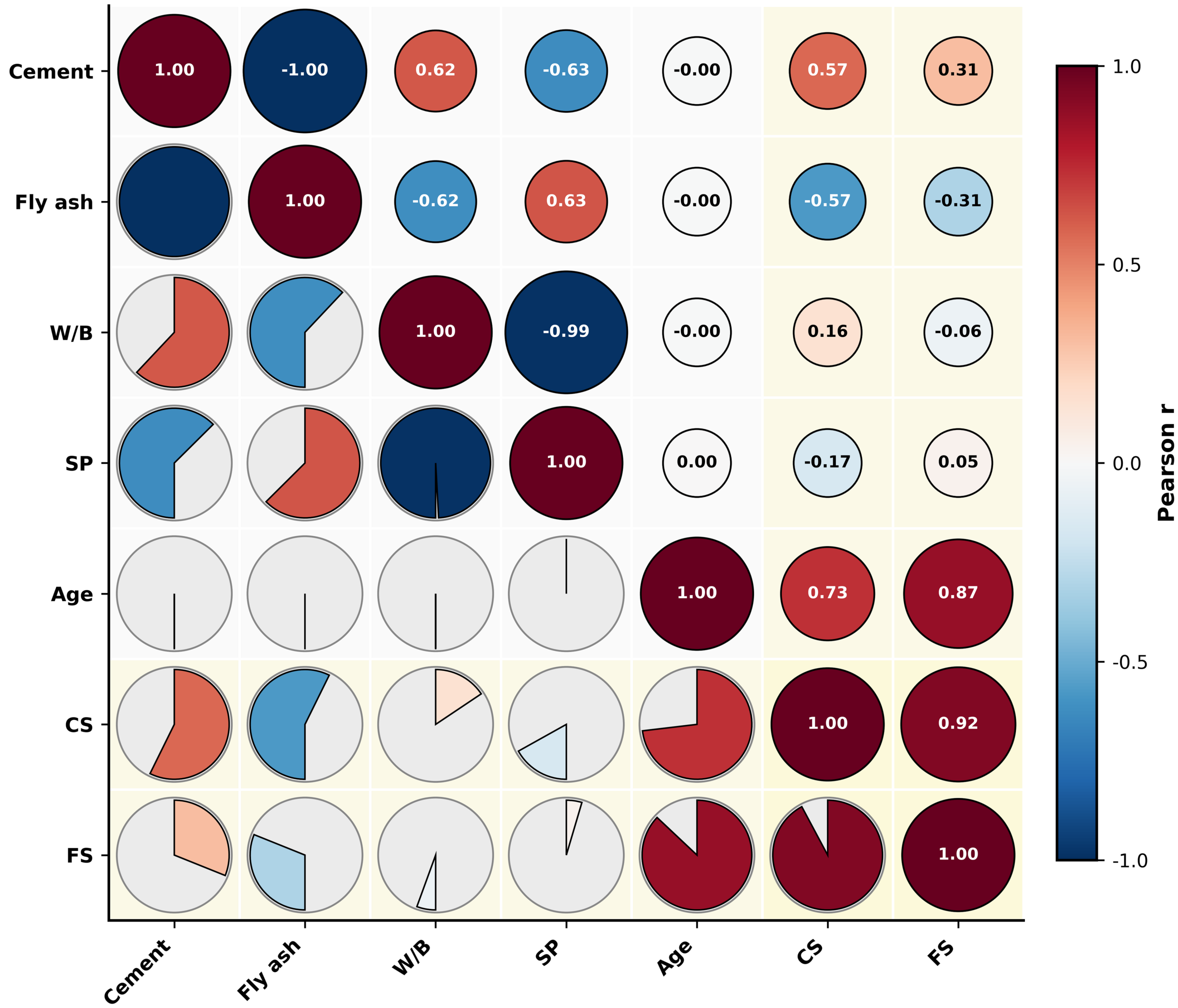

Figure 2: Pearson correlation heatmap for input variables and strength targets. Upper triangle: bubble chart (circle area $\propto |r|$, colour = r). Lower triangle: fan/pie chart (sector angle = $360|r|$, direction encodes sign). Diagonal: r = 1.00.

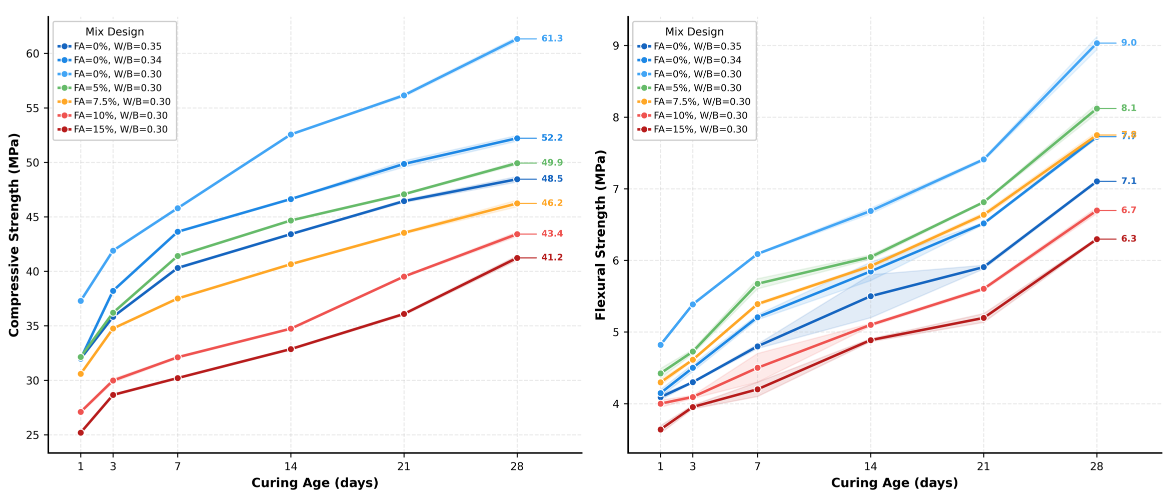

Figure 3: Mean strength development curves by mix design. Shaded bands: ±1 SD across three replicates per age. Data labels show 28-day values. (**left**) Compressive strength; (**right**) flexural strength.

3 Methodology

3.1 Feature Construction and Predictor Selection

Two constant-valued columns (sand: 1000 g; accelerator: 1.4 wt.%) are excluded. Among the remaining variables, total cement mass and FA% are perfectly collinear ($r = -1.00$); only FA% is retained. Superplasticizer dosage and W/B are near-perfectly anti-correlated ($r = -0.991$) because SP was applied exclusively at W/B = 0.30; SP is dropped and W/B is retained. Because SP was exclusively co-applied with the lowest W/B level, the W/B predictor implicitly encodes the combined influence of reduced capillary porosity and SP-assisted rheology modification; the model cannot separate these two contributions. Predictions at W/B = 0.30 therefore reflect the joint effect of a lower water content and a 0.3% SP dosage as used in the original experiments. Users who apply the model to W/B = 0.30 mixes without SP, or to higher-W/B mixes with SP, should exercise caution, as such combinations fall outside the original experimental design and the implicit SP–W/B coupling may not hold. Two features are constructed from the three base variables, yielding a final predictor set of five (Table 2).

Table 2: Predictor set (3 base + 2 engineered).

Feature	Expression	Physical Rationale
FA (%)	Dataset value	Fly ash replacement fraction; governs pozzolanic reaction potential
W/B	Dataset value	Water-to-binder ratio; primary control on gel-space porosity
Age (days)	Dataset value	Curing duration
ln(Age)	ln(Age)	Captures log-linear strength gain; allows linear models to represent curvature without implicit inference
FA% × Age	FA(%) × Age	Encodes the time-varying FA strength penalty: high FA suppresses early-age CS and FS, with the gap narrowing at 28 days as pozzolanic reactions proceed

The two engineered features were selected on physical grounds rather than by automated search. Concrete strength gain follows an approximately logarithmic time curve governed by the evolving gel-space ratio; ln(Age) linearises this relationship so that ElasticNet can capture age-related curvature without requiring polynomial terms. The FA% × Age term encodes the time-dependent pozzolanic modifier effect: at early ages FA dilutes the cement clinker and suppresses strength; the gap narrows as pozzolanic CSH formation proceeds between 1 and 28 days, so the effective sign of FA's contribution to strength evolves with curing duration. Alternative candidates evaluated included a ln(Age) × FA% interaction (captures the same coupling on a log-time scale but provided no improvement over the linear-time version within the 1–28 day window at normal curing temperatures) and quadratic terms W/B^2 and $(FA\%)^2$ (the narrow W/B range of 0.05 units and five-level FA grid produced near-collinear quadratic columns that contributed no independent information to held-out LOMO performance). No other interaction terms reduced seven-fold mean RMSE, so only the two physically motivated features were retained.

CS and FS are modelled as separate regression targets within the same cross-validation framework.

3.2 Leave-One-Mix-Out (LOMO) Cross-Validation

The dataset's factorial structure (seven mixes × six ages × three replicates) enables a composition-level hold-out strategy. In each LOMO fold, all 18 records from one mix are withheld as the test set; the remaining 108 records from the six other mixes form the training set. Seven folds are executed in sequence so every record serves as a test point exactly once. Performance is reported as mean ± standard deviation of R^2, RMSE, and MAE across the seven folds.

This design tests whether a model trained on known compositions can predict the strength of a mix it has never encountered, which is the operationally relevant question for mix-design guidance. Random hold-out splits instead place specimens from the same mix in both partitions; the model then only needs to interpolate along the age axis for a familiar composition, which substantially inflates apparent accuracy without testing generalisation capacity. This strategy is a specific instance of group k-fold cross-validation in which each complete mix composition constitutes one group.

Hyperparameter optimisation (Section 3.4) is performed exclusively on each fold's training partition; the test fold has no influence on model selection.

3.3 Machine Learning Models

Eight regression algorithms spanning four model families are evaluated (Table 3). Models sensitive to input scale (ElasticNet, SVR, GPR) receive z-score normalised features fitted on each fold's training partition; tree-based models use the raw feature matrix.

Table 3: Machine learning models evaluated.

Family	Model	Key Design Characteristics
Linear	ElasticNet	Combined L1 + L2 regularisation; handles residual near-collinearity through coordinated shrinkage
Kernel	SVR	RBF kernel; structural risk minimisation; competitive on small training sets
Tree Ensemble	Random Forest	Bootstrap aggregation of decorrelated trees; variance reduction via bagging
Tree Ensemble	Extra Trees	Randomised split thresholds; further variance reduction at modest bias cost
Gradient Boosting	XGBoost	Regularised sequential tree fitting; L1 + L2 penalties on tree weights
Gradient Boosting	LightGBM	Leaf-wise growth; same regularisation as XGBoost plus `num_leaves` control
Gradient Boosting	CatBoost	Ordered boosting; reduces prediction shift on small datasets
Probabilistic	GPR	Matérn($\nu = 1.5$) + WhiteKernel prior; kernel hyperparameters via log marginal likelihood

3.4 Hyperparameter Optimisation

Model hyperparameters are tuned independently for each LOMO fold using the Tree-structured Parzen Estimator (TPE) implemented in Optuna. Within each fold, TPE evaluates 30 trials, each scored by 5-fold cross-validated RMSE on the training partition. The test fold has no influence on model selection. TPE was selected over grid search and random search for its sequential sample efficiency within a fixed trial budget, and over Gaussian-process-based Bayesian optimisation for its direct handling of mixed integer and continuous hyperparameter spaces without requiring a parametric surrogate. Comparing HPO strategies is outside the scope of this study; the 30-trial budget is held uniform across all eight models to ensure a fair benchmark.

For full-dataset retraining (used in SHAP analysis), the hyperparameter set from the fold achieving the highest test R^2 is adopted for each model. Table 4 lists the resulting configurations.

Table 4: Optimised hyperparameters (from best LOMO fold per model and target).

Model	CS Parameters	FS Parameters
ElasticNet	α = 0.050, l_1-ratio = 0.013	α = 0.0075, l_1-ratio = 0.951
SVR	RBF, C = 12.3, ε = 0.046	RBF, C = 153.9, ε = 0.057
Random Forest	n_{est} = 267, depth = 6, min_leaf = 1	n_{est} = 227, depth = 10, min_leaf = 1
Extra Trees	n_{est} = 205, depth = 7, min_leaf = 1	n_{est} = 287, depth = 11, min_leaf = 2
XGBoost	n_{est} = 393, depth = 6, lr = 0.271, subsample = 0.51	n_{est} = 175, depth = 3, lr = 0.063, subsample = 0.57
LightGBM	n_{est} = 338, depth = 3, leaves = 40, lr = 0.099	n_{est} = 297, depth = 3, leaves = 60, lr = 0.068
CatBoost	iter = 228, depth = 4, lr = 0.060, l_2 = 0.053	iter = 528, depth = 3, lr = 0.027, l_2 = 0.029
GPR	Matérn(ν = 1.5) + WhiteKernel; auto kernel fit	same

3.5 *Evaluation Metrics*

Three metrics quantify predictive accuracy on the test partition of each LOMO fold:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum|y_i - \hat{y}_i| \tag{3}$$

R^2 measures the proportion of variance explained; RMSE penalises large errors quadratically and shares units with the target variable; MAE provides a scale-interpretable average error. Final performance is summarised as mean ± standard deviation across the seven folds.

3.6 *Uncertainty Quantification*

Two complementary methods quantify prediction uncertainty. The primary method constructs empirical prediction intervals (PIs) from the distribution of held-out absolute residuals. For nominal coverage α, the PI half-width equals the α-quantile of all 126 $|y - \hat{y}|$ values collected across the seven LOMO folds. Because each residual derives from an out-of-mix prediction, the intervals capture generalisation error under composition-level hold-out without distributional assumptions. The method is applied to both CS (ExtraTrees) and FS (ElasticNet). The validity of pooling residuals across folds rests on the approximate exchangeability of out-of-mix prediction errors across compositions; empirical coverage curves in Section 4.5 confirm that this assumption holds within the experimental design space.

As a supplementary measure, non-parametric bootstrap resampling (B = 500) estimates in-sample prediction variance for CS. In each iteration, ExtraTrees is retrained on a bootstrap resample of the 126 observations and evaluated on the full dataset; the 2.5th–97.5th percentile band forms the 95% CI at each sample point. Bootstrap is not applied to FS: ElasticNet's strong L1 regularisation produces near-identical coefficient sets across resamples, yielding CI widths below 0.02 MPa and empirical coverage below 30%. This behaviour reflects genuine stability of the FS coefficient estimates rather than model mis-specification:

with 126 observations and only three to four active predictors, the dominant linear drivers of FS (curing age, FA%, W/B) are precisely identified, leaving little parametric uncertainty for resampling to reveal. The LOMO residual PI, whose FS calibration deviates by at most 1.4 percentage points from the nominal level across the 50%–99% range (Section 4.5), already provides a distribution-free, coverage-validated uncertainty bound capturing composition-level generalisation error directly, serving the same practical role as split-conformal prediction applied at the mix level. Among alternative uncertainty quantification frameworks, split-conformal prediction partitions data into a fixed training and calibration set to obtain marginal coverage guarantees; the LOMO residual PI achieves an analogous guarantee at the composition level by pooling out-of-mix residuals across all seven folds, which yields more stable empirical quantiles from 126 residuals than a single hold-out split would allow. Quantile regression forests (QRF) directly estimate conditional quantiles by retaining the full leaf-node residual distribution; however, with only 18 records per composition fold, stable within-fold quantile estimation is difficult, and pointwise QRF intervals do not in general guarantee coverage at the composition level. Bayesian machine learning approaches, including Gaussian processes and Bayesian neural networks, produce full posterior predictive distributions but require prior specification and entail higher computational cost; moreover, their predictive variance primarily reflects within-distribution parameter uncertainty rather than composition-level generalisation error, which is the uncertainty most relevant to practitioners selecting new mix designs.

3.7 Design-Space Analysis and Input Sensitivity

The two best-performing models (ExtraTrees for CS, ElasticNet for FS) are applied to a uniform prediction grid of 1736 points covering FA content from 0 to 15 wt.% at 0.5 wt.% increments and curing age from 1 to 28 days at 0.5-day increments, with W/B fixed at 0.30. At each grid point, predicted CS and FS are offset by the 90% LOMO PI half-widths (Section 3.6) to form conservative lower bounds: $CS_{lower} = \hat{CS} - 4.60$ MPa and $FS_{lower} = \hat{FS} - 0.494$ MPa. A grid point is classified as structurally safe when both lower bounds simultaneously satisfy the design targets CS ≥ 30 MPa and FS ≥ 4 MPa. The minimum curing age for each FA level is the lowest age at which the joint safety condition first holds.

Global input sensitivity is assessed by variance-based Sobol analysis. The three base inputs (FA%, W/B, Age) are sampled uniformly over [0, 15 wt.%], [0.28, 0.36], and [1, 28 days] using Saltelli quasi-random sampling with $N = 2048$ base samples, yielding 16,384 model evaluations per target [36]. First-order indices S_1 quantify each variable's independent contribution to output variance; total-order indices S_T include all interaction effects. One-at-a-time ±1σ tornado diagrams complement the variance decomposition, perturbing each input individually by ±one standard deviation around the representative baseline FA = 7.5 wt.%, W/B = 0.32, Age = 14 days, with $\sigma_{FA} = 2.5$ wt.%, $\sigma_{W/B} = 0.02$, and $\sigma_{Age} = 7$ days.

4 Results and Discussion

4.1 LOMO Cross-Validation Performance

Table 5 and Fig. 4 report the mean and standard deviation of R^2, RMSE, and MAE across seven LOMO folds for each model and target. Rankings differ substantially between CS and FS, indicating that the two properties respond differently to the predictor set and to the training-set size available in each fold.

Compressive strength. ExtraTrees achieves the highest mean R^2 (0.786) with RMSE 2.57 MPa. The two closely ranked alternatives, ElasticNet ($R^2 = 0.730$) and GPR ($R^2 = 0.730$), differ in stability: ElasticNet has a lower standard deviation (0.327 vs. 0.425 for GPR), whereas GPR produces a negative R^2 in the FA15-WB-0.30 fold ($R^2 = -0.216$), flagging numerical instability under certain hold-out compositions. The three gradient boosting frameworks rank below the ensemble trees for CS; the poorer boosting performance likely reflects the limited training partition size (108 records per fold), which restricts residual-fitting convergence within the 30-trial Optuna budget.

Table 5: LOMO cross-validation performance (mean ± SD, seven folds).

Model	CS R^2	CS RMSE (MPa)	CS MAE (MPa)	FS R^2	FS RMSE (MPa)	FS MAE (MPa)
ExtraTrees	**0.786 ± 0.253**	**2.569 ± 2.177**	**2.360 ± 2.079**	0.835 ± 0.158	0.403 ± 0.274	0.370 ± 0.269
ElasticNet	0.730 ± 0.327	2.693 ± 1.783	2.412 ± 1.660	**0.915 ± 0.061**	**0.301 ± 0.077**	**0.249 ± 0.066**
GPR	0.730 ± 0.425	2.604 ± 1.757	2.369 ± 1.569	0.600 ± 0.700	0.525 ± 0.360	0.483 ± 0.349
SVR	0.709 ± 0.444	3.003 ± 2.859	2.788 ± 2.828	0.725 ± 0.168	0.557 ± 0.211	0.487 ± 0.159
CatBoost	0.661 ± 0.166	3.542 ± 1.260	3.299 ± 1.216	0.780 ± 0.228	0.462 ± 0.262	0.427 ± 0.260
RandomForest	0.547 ± 0.203	4.077 ± 1.447	3.860 ± 1.372	0.750 ± 0.254	0.486 ± 0.278	0.452 ± 0.273
XGBoost	0.512 ± 0.230	4.282 ± 1.602	4.033 ± 1.528	0.702 ± 0.272	0.538 ± 0.334	0.504 ± 0.341
LightGBM	0.476 ± 0.379	4.170 ± 2.081	3.625 ± 2.060	0.591 ± 0.332	0.658 ± 0.310	0.549 ± 0.304

Note: Bold: best model per target.

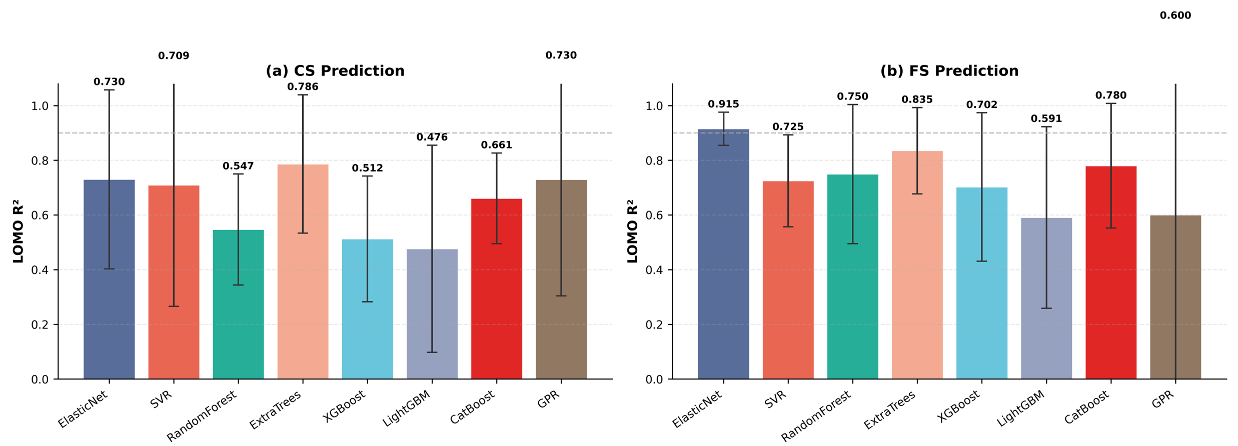

Figure 4: LOMO cross-validation R^2 for all eight models. Error bars: ±1 SD across seven folds; dashed line: R^2 = 0.90. (**a**) CS prediction; (**b**) FS prediction.

Flexural strength. ElasticNet ranks first for FS by a clear margin ($R^2 = 0.915 \pm 0.061$, RMSE = 0.301 MPa), with a fold standard deviation less than half that of any other model. GPR shows severe instability for FS: $R^2 = 0.600 \pm 0.700$, with one fold producing $R^2 = -0.973$ (FA15-WB-0.30), reflecting numerical conditioning failures in the Matérn kernel under small, high-leverage training sets. LightGBM ranks last for both targets, explained by its sensitivity to shallow trees and the small per-fold training size.

The model rankings between CS and FS diverge in a physically interpretable way: CS exhibits stronger nonlinear interactions between W/B, FA%, and age, which ExtraTrees captures through high-dimensional splits, while FS follows a smoother, more monotonic response that a regularised linear model can represent compactly. The SHAP attribution quantifies this contrast directly: the FA% × Age interaction term contributes 0.77 MPa for CS but only 0.15 MPa for FS, and ln(Age) drops from 2.62 MPa to 0.02 MPa across the two targets; in FS the three base predictors (Age, FA%, W/B) collectively account for over 97% of total SHAP attribution, leaving no role for the interaction features that give tree-based models their advantage in CS.

4.2 Best-Model Performance by Fold

ExtraTrees for CS. Six of seven folds return $R^2 \geq 0.73$. The two weakest folds are FA0-WB-0.30 (R^2 = 0.264, RMSE = 7.12 MPa) and FA0-WB-0.35 (R^2 = 0.735, RMSE = 2.97 MPa), both OPC-only mixes. When all training mixes contain FA replacement, the model must extrapolate to a domain without pozzolanic reaction; strength gain is governed purely by OPC hydration. The four FA-blend folds yield R^2 of 0.978, 0.979, 0.845, and 0.744, confirming that ExtraTrees generalises well within the FA-replacement domain. Averaged across these four FA-blend folds, the mean R^2 is 0.887, compared with 0.500 for the two lowest-performing pure OPC folds (FA0-WB-0.30 and FA0-WB-0.35). The third OPC composition (FA0-WB-0.34) achieves R^2 = 0.957, comparable to the FA-blend range, because the remaining two OPC folds are retained in its training set and the model does not need to extrapolate outside the observed W/B range.

ElasticNet for FS. Fold-level R^2 ranges from 0.800 (FA10-WB-0.30) to 0.978 (FA0-WB-0.34), with all seven folds above 0.80. RMSE ranges from 0.18–0.42 MPa against an overall FS SD of 1.29 MPa. The sparse optimal solution (α = 0.0075, l_1-ratio = 0.951) retains only three to four non-negligible coefficients, confirming that most FS variance is captured by a low-dimensional combination of Age, FA%, and W/B. Predicted vs. measured scatter for both best models is shown in Fig. 5.

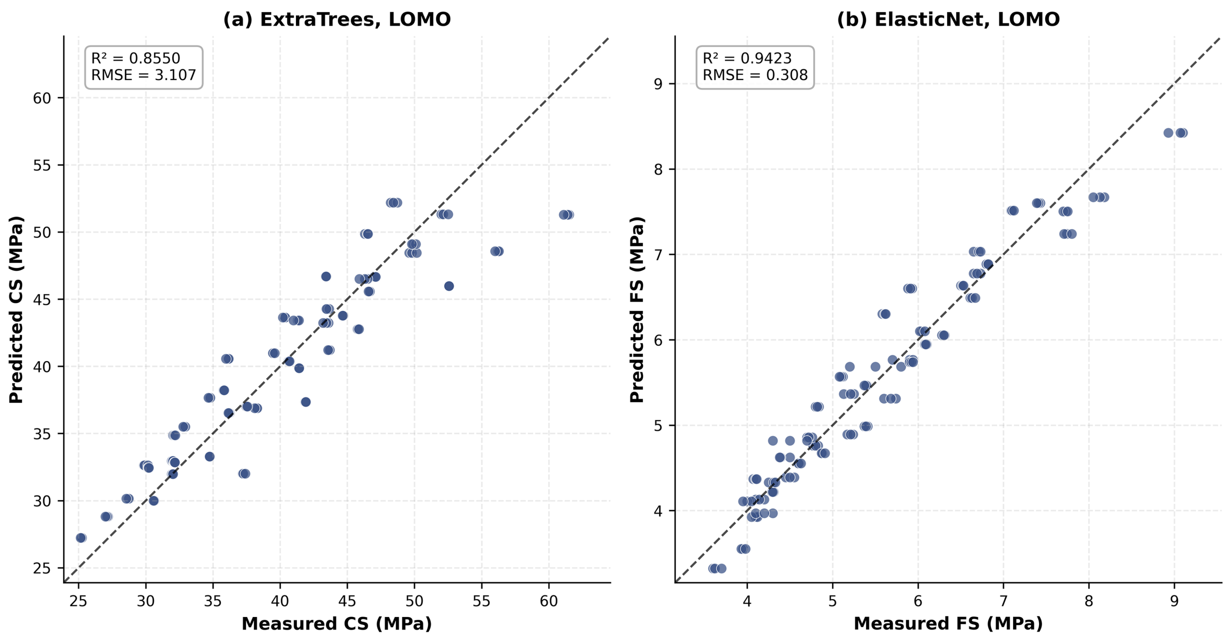

Figure 5: LOMO predicted vs. measured values for the best-performing model per target. Dashed line: 1:1 reference. (**a**) ExtraTrees, CS; (**b**) ElasticNet, FS.

4.3 SHAP Feature Importance and Dependence

Global importance. Mean absolute SHAP values are reported in Table 6, with bar charts in Fig. 6 and beeswarm plots in Fig. 7.

Table 6: Mean absolute SHAP values for full-dataset models (MPa).

Rank	CS Feature (ExtraTrees)	CS $\overline{\lvert SHAP \rvert}$	FS Feature (ElasticNet)	FS $\overline{\lvert SHAP \rvert}$
1	FA (%)	3.920	Age (days)	1.057
2	Age (days)	2.859	FA (%)	0.488
3	ln(Age)	2.615	W/B	0.433
4	W/B	1.375	FA% × Age	0.153
5	FA% × Age	0.768	ln(Age)	0.022

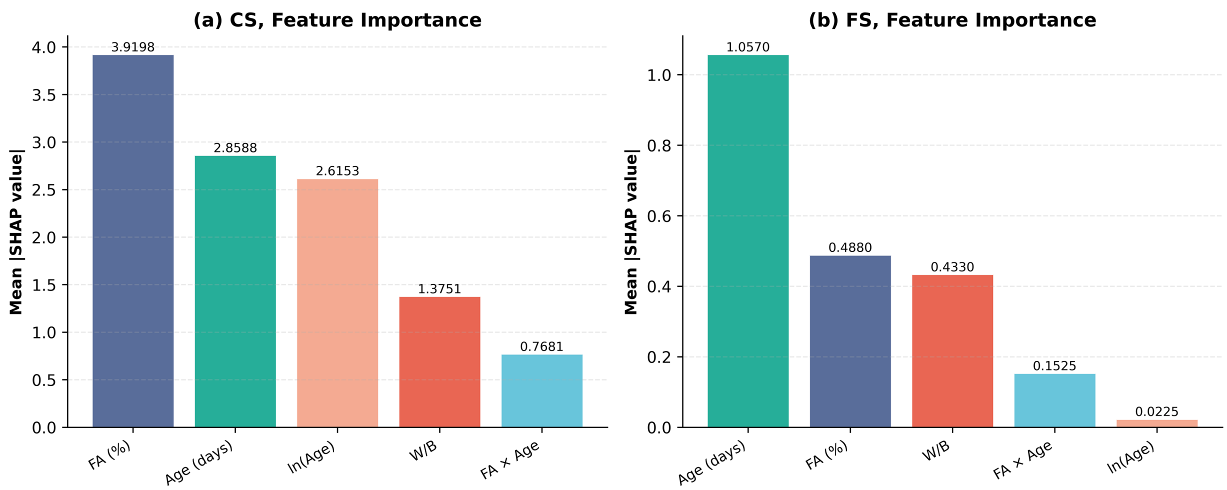

Figure 6: Mean absolute SHAP feature importance for the full-dataset models. (**a**) ExtraTrees for CS; (**b**) ElasticNet for FS.

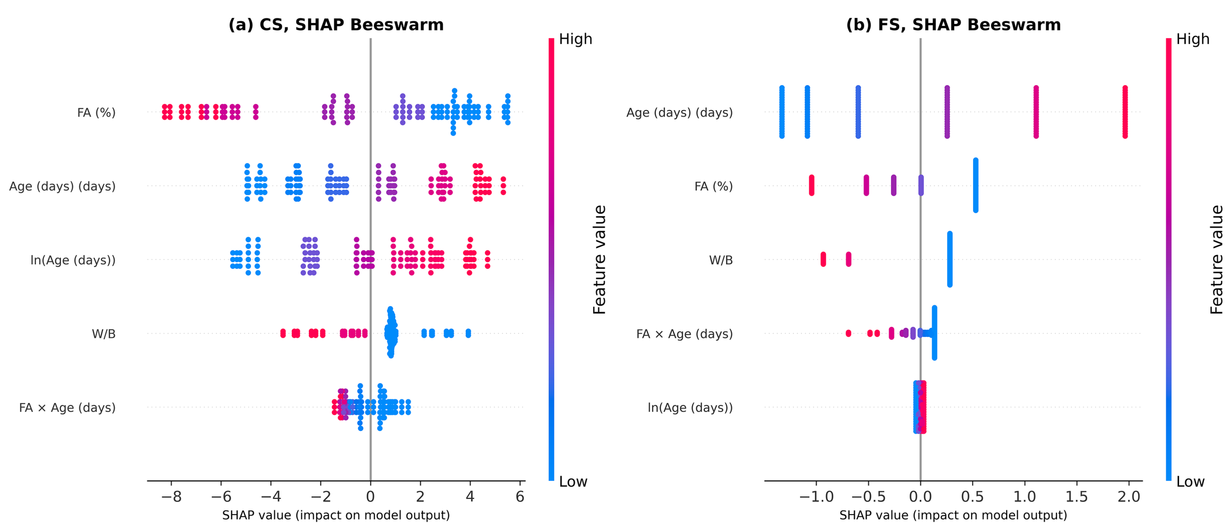

Figure 7: SHAP beeswarm plots for the full-dataset models. Each point represents one sample; colour encodes feature value (red = high, blue = low). (**a**) ExtraTrees, CS; (**b**) ElasticNet, FS.

Compressive strength (ExtraTrees). FA replacement percentage is the single most influential predictor (mean |SHAP| = 3.92 MPa), with a consistently negative marginal effect: as FA% increases from 0% to 15%, the SHAP contribution to CS decreases monotonically. Within the 1–28 day curing window covered by the dataset, FA substitution acts primarily as an OPC diluent: replacing clinker with a less reactive material reduces early-age hydration product formation and lowers CS. The secondary pozzolanic mechanism, in which reactive silica and alumina from FA consume portlandite to generate additional CSH, reverses this dilution penalty only beyond 28–90 days of curing, a range not represented in the training data; the SHAP attribution therefore reflects the dilution-dominated net effect within the observed window. The combined contribution of Age (2.86 MPa) and ln(Age) (2.62 MPa), together 47% of total importance, confirms that curing-time evolution is the second dominant axis. W/B ranks fourth (1.38 MPa); its negative marginal effect (lower W/B yields higher CS) emerges clearly at W/B = 0.30 relative to 0.34 and 0.35. The FA% × Age term (0.768 MPa, 7% of total) reflects the fold-level evidence that the FA penalty on CS partially diminishes as curing age increases.

Flexural strength (ElasticNet). The importance hierarchy is reversed: Age dominates (1.057 MPa) while FA% ranks second (0.488 MPa) and W/B third (0.433 MPa). The near-zero importance of ln(Age) (0.022 MPa) is consistent with FS development being well described by a linear-in-age term; the log transformation provides no incremental information once raw age enters the model. The W/B ranking for FS (third vs. fourth for CS) and the reversal of the Age-vs.-FA% order suggest that the FS response surface is more strongly driven by curing kinetics than by mix composition, while the opposite holds for CS. SHAP dependence plots for the top two features per target are shown in Fig. 8.

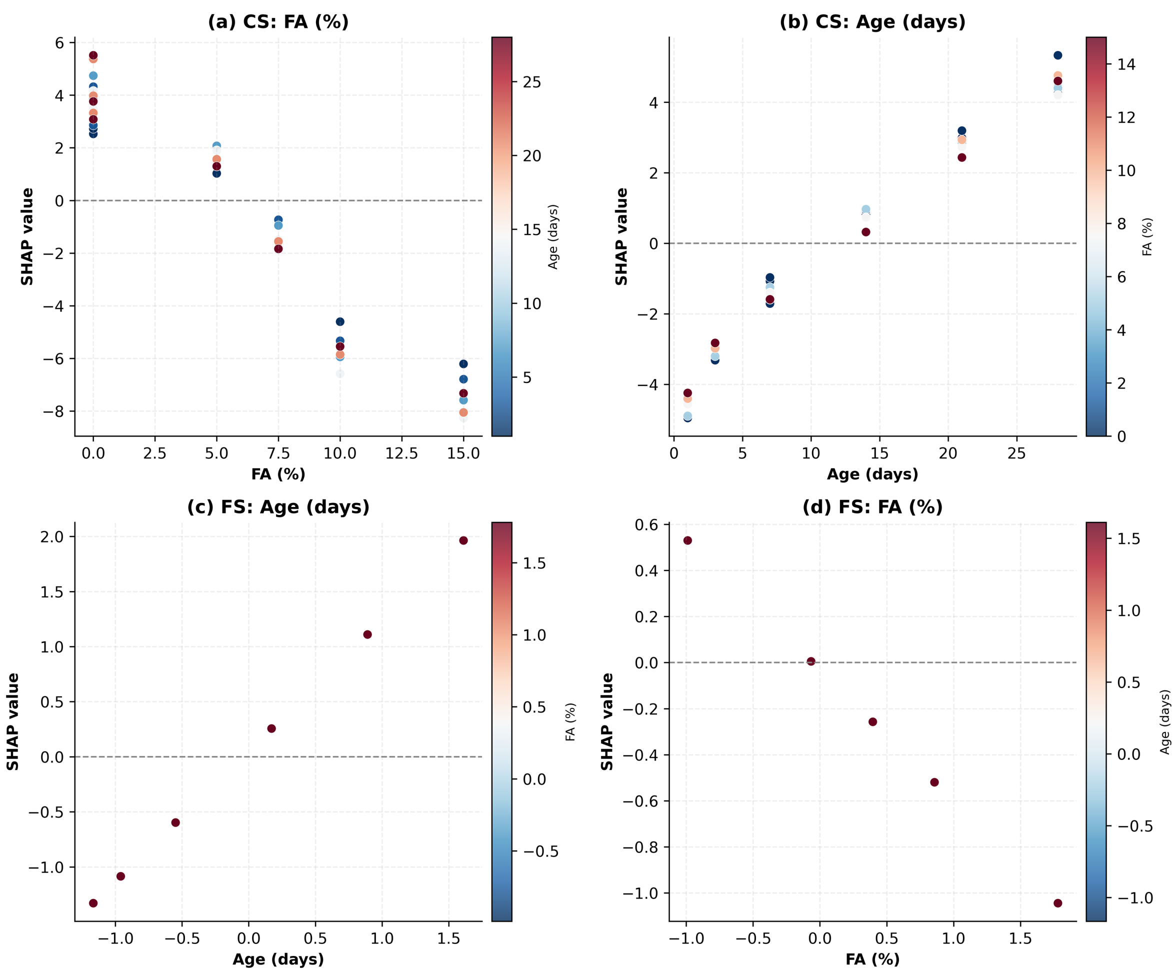

Figure 8: SHAP dependence plots for the top-two features per target. Colour encodes the complementary top feature. (**a**) CS: FA (%); (**b**) CS: age (days); (**c**) FS: age (days); (**d**) FS: FA (%).

4.4 Practical Implications for Mix Design

FA% as the primary CS lever. FA replacement percentage contributes 3.92 MPa (34% of total feature attribution) to CS. Mixes at FA = 0%–5% show the smallest SHAP penalties; mixes at FA = 10%–15% incur increasingly large CS reductions at 3–7-day ages. The LOMO folds for FA5-WB-0.30 and FA7.5-WB-0.30 achieve R^2 = 0.979 and RMSE below 1 MPa, confirming that 5%–7.5% FA at W/B = 0.30 represents the optimal and most reliably predicted design range. For structural applications where 28-day CS governs, FA above 10% should be accompanied by a W/B reduction; in the OPC-only series, reducing W/B from 0.35 to 0.30 raised 28-day CS by 12.9 MPa (+26.6%), though W/B variation for FA-blended mixes was not investigated in this dataset.

FA% × age interaction and the 3DPC printability window. Elevated FA replacement slows early setting and retains workability, which is beneficial for printability and layer adhesion, but suppresses early-age strength. The positive SHAP contribution of FA% × Age at high ages supports a strategy of using FA = 7.5%–10% for print runs where extended open time is needed, noting that 28-day CS remains approximately 15 MPa (FA = 7.5 wt.%) and 18 MPa (FA = 10 wt.%) below the OPC-only reference at W/B = 0.30. For applications requiring post-print structural loading within 3 days, FA% should remain below 5%.

Age as the dominant FS lever. Age carries 49% of total FS SHAP attribution (1.057 MPa), while FA% and W/B contribute comparably (0.488 and 0.433 MPa). The most effective way to increase FS within the tested composition space is to extend curing duration rather than to adjust mix proportions. The ElasticNet RMSE of 0.30 MPa against a dataset mean of 5.56 MPa (5.4% relative error) provides a reliable tool for specifying minimum curing ages to reach FS thresholds.

Model applicability bounds. ExtraTrees generalises well within the FA-blend domain but degrades for pure OPC mixes (FA = 0%, RMSE up to 7.12 MPa). ElasticNet maintains R^2 > 0.80 across all seven LOMO folds, suitable for mix-level interpolation across the full FA% range. Both models should be treated as local interpolators; extrapolation beyond W/B 0.30–0.35 or FA% 0%–15% is not validated.

4.5 Prediction Intervals and Uncertainty Calibration

The LOMO residual PIs are well calibrated across the full 50%–99% nominal coverage range for both targets (Fig. 9). Empirical coverage deviates from the nominal level by at most 0.6 percentage points for CS and 1.4 percentage points for FS.

At the 90% nominal level, the PI half-width is ±4.60 MPa for CS and ±0.494 MPa for FS (Fig. 10). The CS half-width represents 56% of the dataset SD (8.19 MPa) and 12.6% of the data range (36.4 MPa). It is conservative for the four FA-blend folds (fold RMSE below 1 MPa) and slightly non-conservative for pure OPC folds. The FS half-width of ±0.494 MPa equals 38% of one FS SD (1.29 MPa), consistent with ElasticNet's low fold-level variability.

The Bootstrap 95% CI for CS has a mean width of 2.15 MPa and in-sample coverage of 92.1% (Fig. 11). Its width is narrower than the LOMO PI because it quantifies ExtraTrees parameter variance under training-set perturbation rather than generalisation error to unseen mixes. The two intervals measure distinct uncertainty sources: the Bootstrap CI bounds sensitivity of the predicted mean to training data composition; the LOMO PI bounds the expected prediction–observation deviation for a held-out mix.

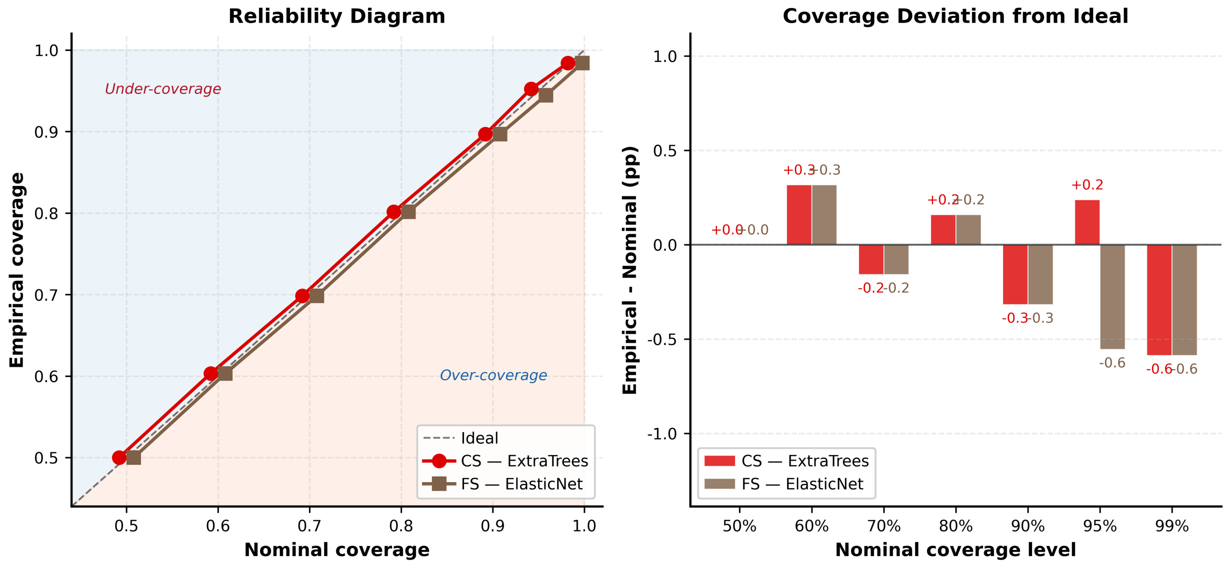

Figure 9: Reliability diagram and empirical coverage deviation for LOMO residual prediction intervals. Dashed line: perfect calibration. (**left**) CS (ExtraTrees); (**right**) FS (ElasticNet).

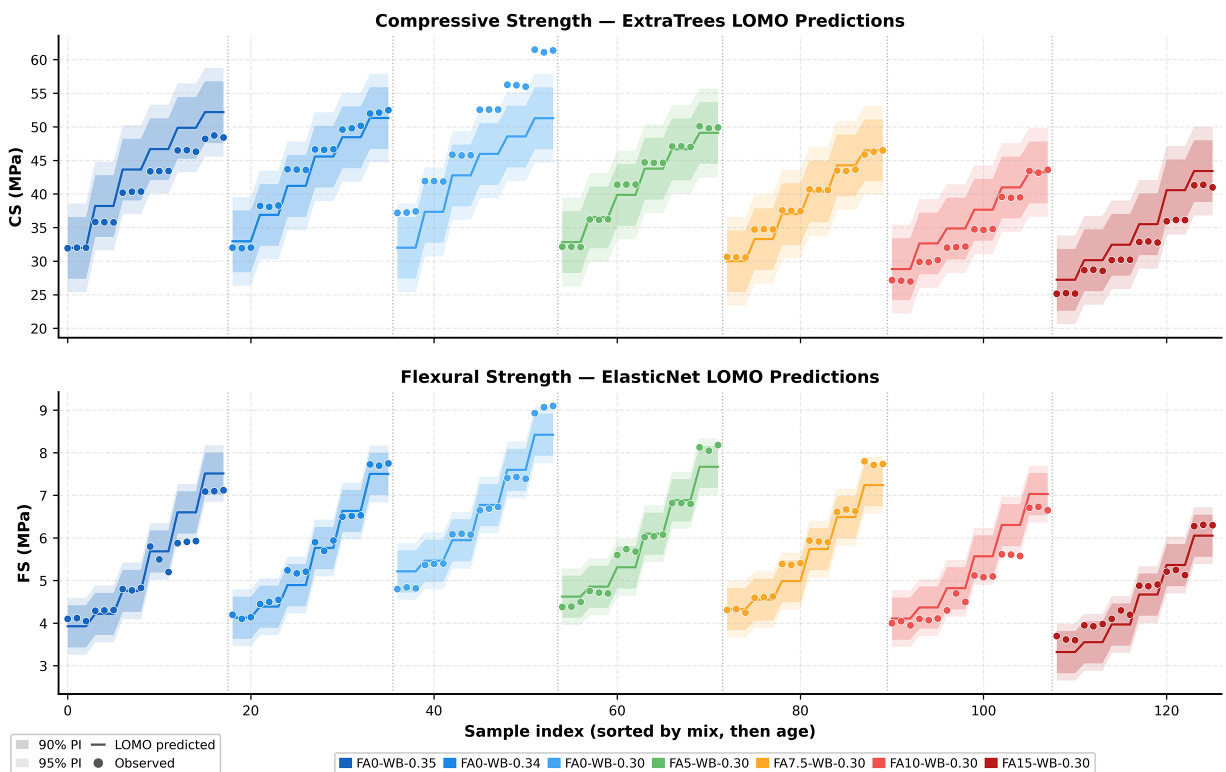

Figure 10: LOMO prediction intervals at 90% nominal coverage for all 126 samples, sorted by measured value. Shaded bands: 90% PI; filled circles: measured values. (**top**) CS; (**bottom**) FS.

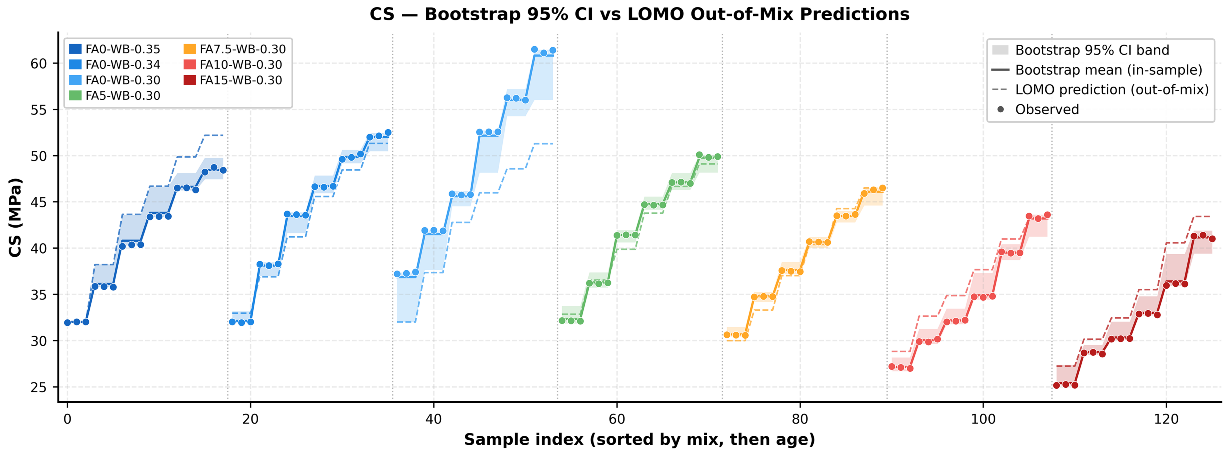

Figure 11: Bootstrap 95% confidence intervals vs. LOMO out-of-mix predictions for CS (ExtraTrees). Grey band: bootstrap 95% CI; coloured markers: LOMO out-of-mix predictions.

4.6 Design-Space Mapping and Global Input Sensitivity

The 90% PI half-widths established in Section 4.5 (±4.60 MPa for CS, ±0.494 MPa for FS) serve as conservative offsets from mean predictions to define safety boundaries across the FA% × Age parameter space at W/B = 0.30.

CS and FS response surfaces. Predicted CS spans 25.2 to 61.2 MPa over the FA% × Age grid (Fig. 12). The FA% penalty on CS is present at all ages: raising FA from 0 to 15 wt.% reduces predicted CS by approximately 13 MPa at day 3, widening to approximately 20 MPa at day 28 as OPC-dominant mixes continue to gain strength more rapidly than FA-blended counterparts within the 28-day curing window. Predicted FS ranges from 3.54 to 8.51 MPa over the same grid (Fig. 12) and follows a near-monotonic age trajectory; the FA% gradient is gentler and more uniform than for CS, consistent with ElasticNet's near-linear FS surface.

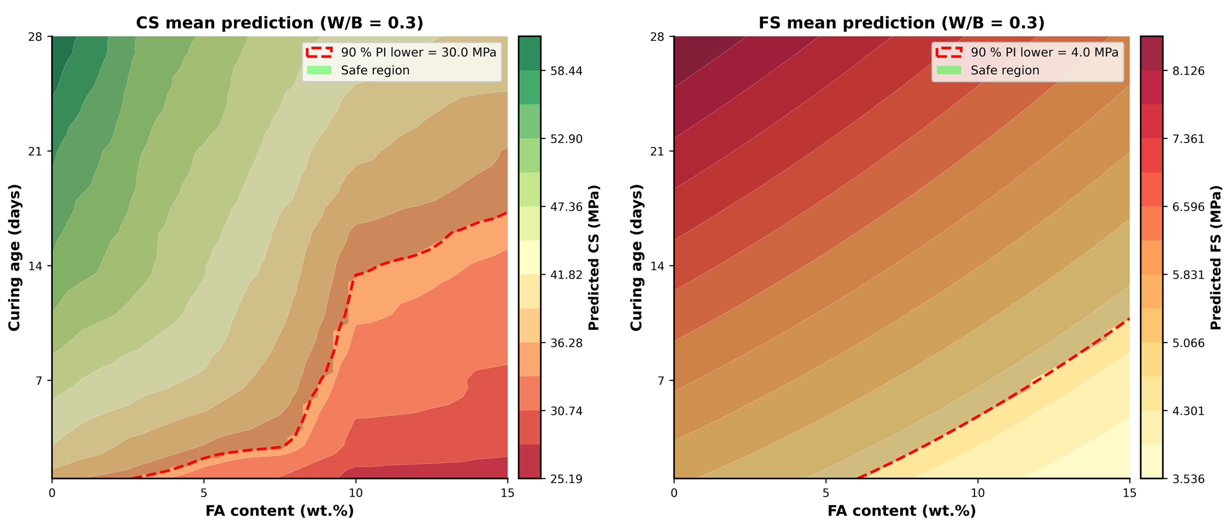

Figure 12: Predicted CS and FS response surfaces over the FA% × curing age design space at W/B = 0.30. Dashed red contour: 90% PI lower bound equal to the design target. Shaded green: safe region where the lower bound exceeds the target. (**left**) CS (ExtraTrees); (**right**) FS (ElasticNet).

Safe design regions. Applying the 90% PI lower bounds to each grid point identifies the FA%–Age combinations satisfying both targets simultaneously (Fig. 13). The CS constraint is consistently the binding limit: at every FA level tested, the minimum curing age for CS compliance equals or exceeds that for FS compliance (Table 7). At FA = 10 wt.%, CS requires curing to at least 13.5 days while FS compliance is achieved by day 5; at FA ≤ 7.5 wt.%, both targets are met within three days.

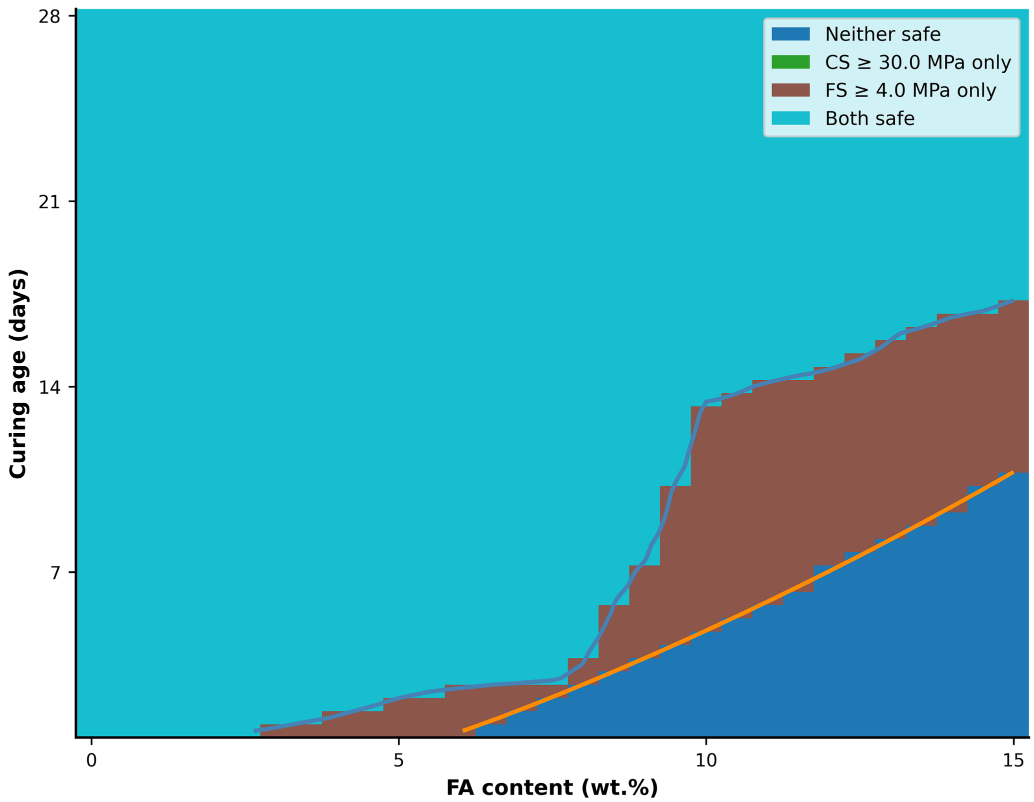

Figure 13: Combined safe-region map at W/B = 0.30. Colours indicate: cyan, both CS and FS targets met; brown, FS target met only; dark blue, neither target met. Boundary lines indicate the 90% PI lower bounds for CS (blue) and FS (orange).

Table 7: Minimum curing age for simultaneous CS ≥ 30 MPa and FS ≥ 4 MPa compliance at W/B = 0.30, based on 90% PI lower bounds. CS constraint is binding at all FA levels.

FA (wt.%)	Min. Age, CS ≥ 30 MPa (days)	Min. Age, FS ≥ 4 MPa (days)	Min. Age, Both Targets (days)
0	1.0	1.0	1.0
5	2.5	1.0	2.5
7.5	3.0	2.5	3.0
10	13.5	5.0	13.5
15	17.5	11.0	17.5

Minimum curing age requirements. Table 7 and Fig. 14 report the minimum curing age for simultaneous CS and FS compliance at each FA level. For FA ≤ 7.5 wt.%, the threshold is reached by day 3; it extends to 13.5 days at FA = 10 wt.% and 17.5 days at FA = 15 wt.%. The large step between 7.5% and 10% marks the transition from a regime where OPC hydration provides sufficient early-age CS to one where pozzolanic reaction delay becomes the governing constraint. For print runs requiring load-bearing capacity within one week, FA should not exceed 7.5 wt.% at W/B = 0.30; for extended open-time operations (Section 4.4), FA = 10 wt.% is viable provided a minimum 14-day curing period is maintained.

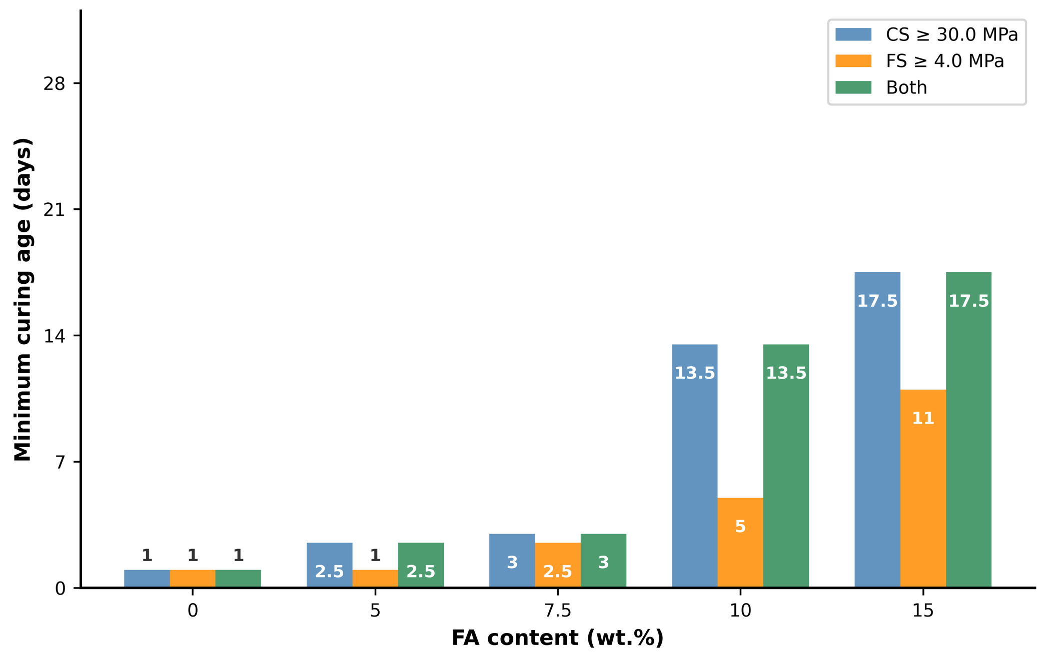

Figure 14: Minimum curing age to reach simultaneous CS and FS design targets at W/B = 0.30 for each FA level.

All design-space results are conditioned on W/B = 0.30, the lowest and most mechanically favourable value in the dataset. At W/B = 0.34–0.35, predicted FS declines: the tornado sensitivity analysis indicates a change of ±0.49 MPa per ±0.02 increase in W/B at the reference baseline (FA = 7.5 wt.%, Age = 14 days), so the FS compliance boundary shifts toward longer minimum curing ages at higher W/B. The CS constraint, which is binding at all FA levels at W/B = 0.30, remains the governing limit at higher W/B but with a reduced compliance margin, particularly at FA ≥ 10 wt.% where the predicted CS already approaches the 30 MPa threshold. Raising the CS compliance target to 40 MPa substantially alters the design guidance. At FA = 7.5 wt.% and W/B = 0.30, the representative 28-day prediction yields a 90% PI lower bound of 41.6 MPa (Section 4.7), indicating that the three-day compliance window identified at CS ≥ 30 MPa no longer holds and curing close to the full 28-day window would be required at this FA level. At FA ≥ 10 wt.%, achieving a 40 MPa lower PI bound within the 28-day training range would be marginal. The FS constraint would remain non-binding under this adjustment, given the wide margin above the 4 MPa target throughout the FA ≤ 10 wt.% range.

Global sensitivity analysis. Sobol first-order indices for the three base inputs are shown in Fig. 15. For CS, FA% accounts for the largest share of output variance ($S_1 = 0.553$), with Age second ($S_1 = 0.412$) and W/B negligible ($S_1 = 0.005$). For FS the ordering reverses: Age dominates ($S_1 = 0.526$), with FA% and W/B contributing comparably ($S_1 = 0.242$ and 0.228). Total-order indices track first-order indices closely across all cases ($S_T - S_1 \leq 0.03$), indicating that pairwise interactions contribute little additional variance. The pattern of S_1 values nonetheless reveals a structural difference between the two targets. For CS, W/B carries a near-zero first-order index ($S_1 = 0.005$), meaning it produces no measurable independent main effect within the 0.30–0.35 printability window; its contribution to CS variance is channelled through interactions with FA% and Age, consistent with the gel-space ratio being jointly controlled by water content and clinker dilution. For FS, the three inputs share variance more evenly (S_1 = 0.526, 0.242, 0.228), and all track closely in S_T, confirming the near-additive response structure that allows ElasticNet to perform well without capturing higher-order couplings. The tornado diagrams (Fig. 16) quantify these sensitivities directionally. For CS at the baseline (FA = 7.5 wt.%, W/B = 0.32, Age = 14 days, $\hat{\text{CS}}$ = 40.7 MPa), increasing FA% by one standard deviation (2.5 wt.%) reduces predicted CS by 5.92 MPa; the symmetric decrease raises it by 4.36 MPa. Age perturbation (±7 days) produces a ±2.86–3.13 MPa response, while W/B variation at ±0.02 yields no measurable change, consistent with the near-zero Sobol index and the narrow W/B range available in the training data. For FS at the same baseline ($\hat{\text{FS}}$ = 5.3 MPa), age perturbation yields the largest change (±0.76 MPa), followed by W/B (±0.49 MPa) and FA% (±0.33 MPa). The Sobol decomposition and SHAP attribution (Section 4.3) are in close agreement for both targets, confirming that the driver shift from FA% for CS to Age for FS is a structurally robust feature of the dataset and not an artifact of any single interpretability method.

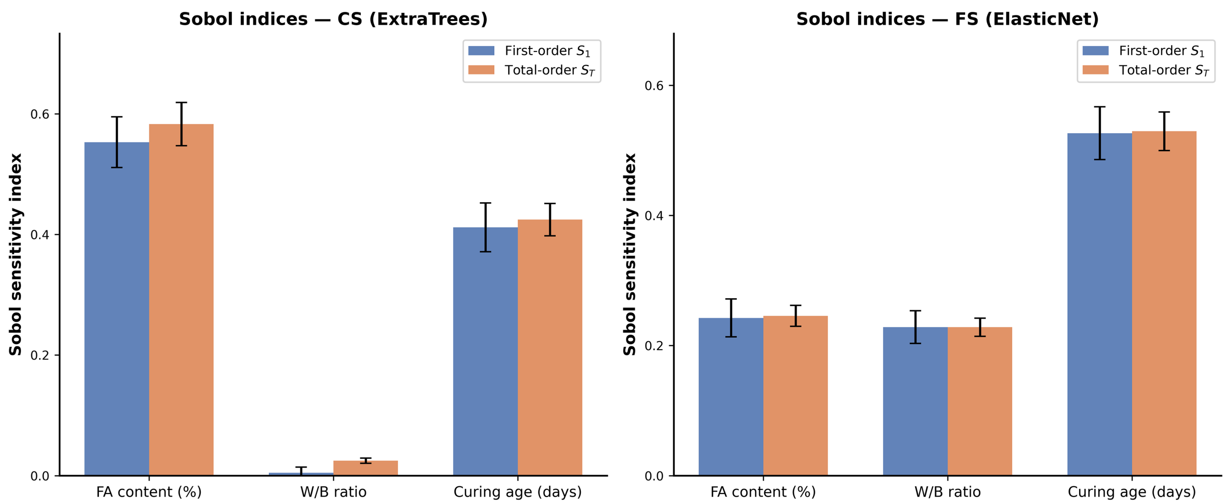

Figure 15: Variance-based sobol sensitivity indices for CS and FS. Bars: first-order S_1 (independent contribution) and total-order S_T (including interactions). Error bars: 95% bootstrap confidence intervals. (**left**) CS (ExtraTrees); (**right**) FS (ElasticNet).

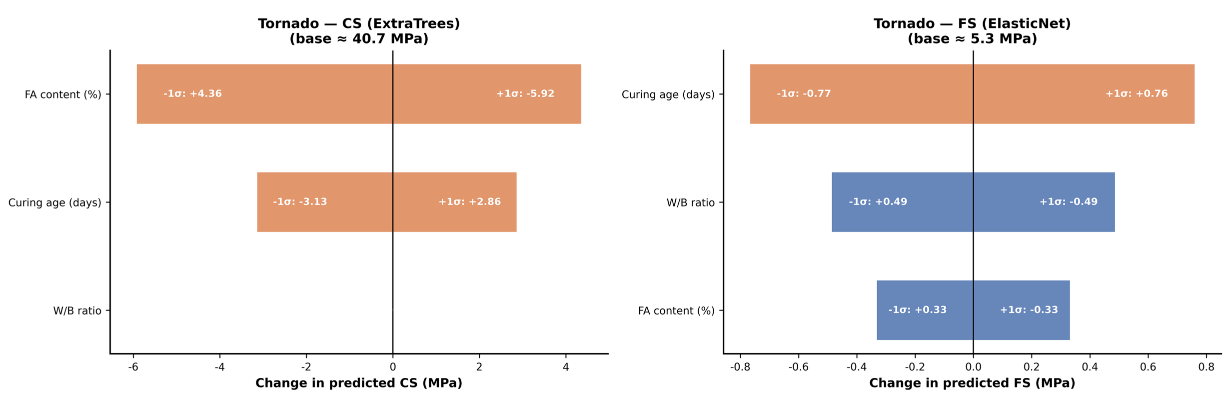

Figure 16: One-at-a-time $\pm 1\sigma$ tornado diagrams at baseline FA = 7.5 wt.%, W/B = 0.32, Age = 14 days. Bar length indicates predicted strength change; labels inside bars show the directional magnitude. (**left**) CS (ExtraTrees, base = 40.7 MPa); (**right**) FS (ElasticNet, base = 5.3 MPa).

4.7 *Graphical User Interface*

A standalone graphical user interface (GUI) was developed in Python to enable practical strength prediction without programming expertise (Fig. 17). Three active inputs are required: FA replacement content (wt.%), W/B ratio, and curing age. On submission, the interface returns simultaneous CS and FS predictions with 90% LOMO-calibrated prediction intervals. For the representative composition FA = 7.5 wt.%, W/B = 0.30, Age = 28 days, the GUI yields CS = 46.2 MPa (90% PI: [41.6, 50.8] MPa) and FS = 7.31 MPa (90% PI: [6.81, 7.80] MPa). The GUI is freely available at https://github.com/lucassivan/ML-FA-3DPC. Users are advised that CS predictions at FA = 0% carry substantially higher uncertainty than those within the FA-blend range; for reliable design guidance the interface is best applied at FA ≥ 5 wt.%, where composition-level generalisation has been validated. Both models are lightweight and require no dedicated hardware. The GUI trains both models from the dataset at startup in a background thread (typically within a few seconds on a standard laptop) and returns predictions instantaneously thereafter. The application runs locally without an internet connection or external server, making it suitable for on-site use during mix-design review.

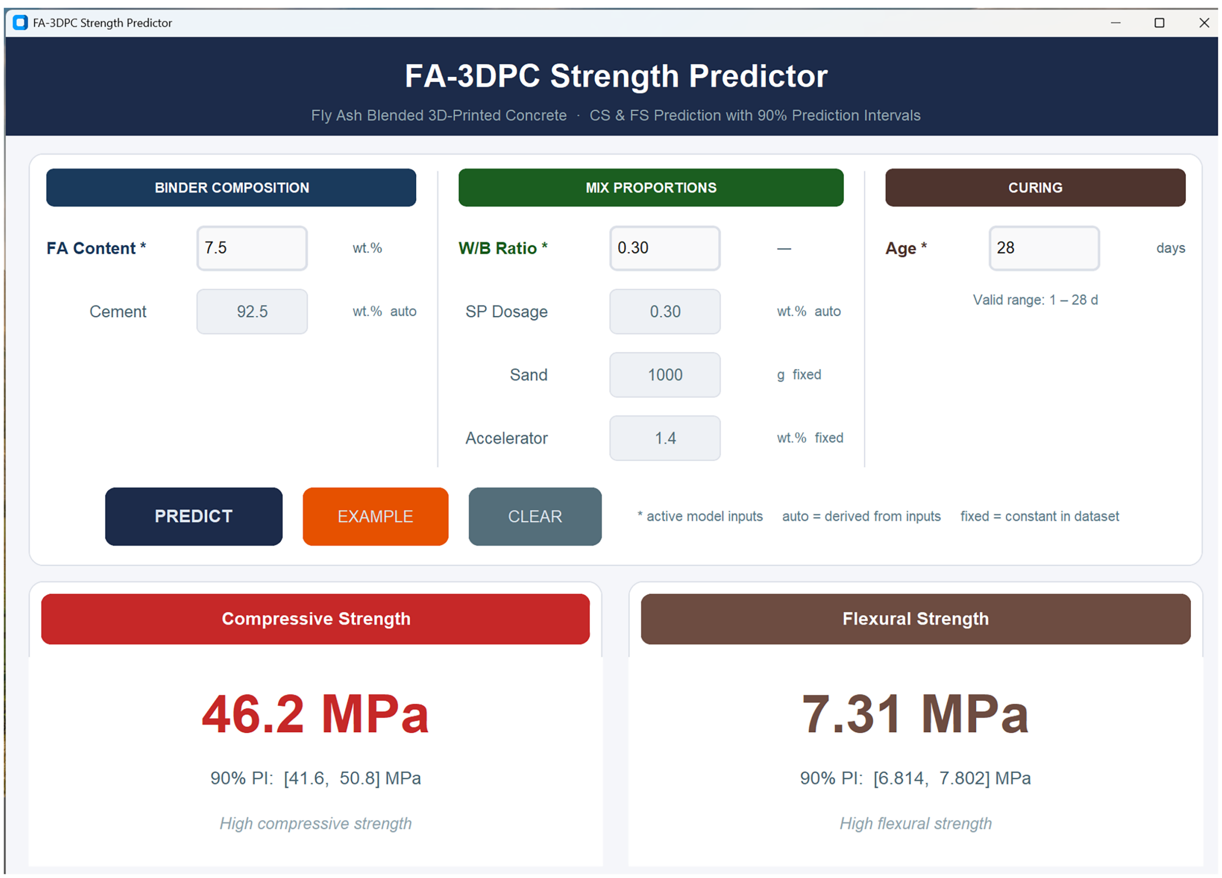

Figure 17: Graphical user interface for simultaneous CS and FS prediction with 90% calibrated prediction intervals. Inputs: FA replacement content (%), W/B ratio, and curing age (days).

5 Conclusions

This study benchmarked eight machine learning algorithms on 126 experimental records from seven FA-blended 3DPC compositions using leave-one-mix-out (LOMO) cross-validation, and augmented predictive models with SHAP attribution, calibrated prediction intervals, and a graphical user interface. The following conclusions are drawn.

1. ExtraTrees achieves the highest composition-level generalisation for CS (R^2 = 0.786 ± 0.253, RMSE = 2.57 MPa), while ElasticNet achieves the best and most stable performance for FS (R^2 = 0.915 ± 0.061, RMSE = 0.301 MPa). The divergent rankings reflect the distinct physical response mechanisms of the two targets: CS responds to complex nonlinear interactions between FA%, W/B, and age, whereas FS follows a smoother, monotonic pattern that a regularised linear model captures compactly.
2. SHAP analysis reveals an inverted importance hierarchy: FA% is the dominant CS predictor (mean |SHAP| = 3.92 MPa, consistently negative), while curing age dominates FS (1.057 MPa). This reversal is physically consistent with the distinct sensitivity of compressive and flexural failure mechanisms to clinker dilution vs. continued paste densification.
3. LOMO residual prediction intervals are well calibrated across the 50%–99% nominal coverage range, with empirical deviations not exceeding 1.4 percentage points. At the 90% level, half-widths are ±4.60 MPa for CS and ±0.494 MPa for FS, providing actionable bounds for uncertainty-aware mix design decisions.
4. FA substitution at 5%–7.5% with W/B = 0.30 is identified as the optimal design range. For applications requiring structural loading within three days, FA% should be limited to 5% to preserve early-age performance.
5. Design-space mapping at W/B = 0.30 identifies minimum curing ages of 3 days at FA ≤ 7.5 wt.% and 13.5 days at FA = 10 wt.% for simultaneous CS ≥ 30 MPa and FS ≥ 4 MPa compliance, based on 90% PI lower bounds. Sobol analysis attributes 55.3% of CS output variance to FA% (S_1 = 0.553) and 52.6% of FS output variance to curing age (S_1 = 0.526).
6. The current models condition predictions solely on mix composition; incorporating printing-process variables, including layer height, print speed, and ambient temperature, as additional inputs would close the gap between mix-design optimisation and on-site strength realisation, where process deviation often dominates composition-level uncertainty. Among these variables, interlayer time interval and layer height are expected to exert the greatest influence on FS, as both govern the contact quality and degree of partial hydration at the printed interface where tensile-governed crack initiation under bending preferentially occurs; print speed and extrusion pressure primarily affect bead geometry and layer density, with consequences mainly for CS. Environmental temperature modulates hydration kinetics for both OPC and FA components, with the effect amplified in FA-blend systems where the pozzolanic reaction is more temperature-sensitive than pure OPC hydration. Systematic collection of process metadata alongside mechanical testing in future multi-source studies would enable these variables to enter the modelling framework as additional predictors without altering the LOMO cross-validation structure adopted here.

Several limitations of this study should be acknowledged when applying the models in practice. The dataset originates from a single laboratory and spans seven mix compositions; the trained models are strictly local interpolators within the experimental design space (FA 0–15 wt.%, W/B 0.30–0.35, age 1–28 days). Predictions at FA > 15 wt.%, W/B below 0.30 or above 0.35, or curing ages beyond 28 days represent extrapolations that have not been validated and may yield unreliable outputs. The ExtraTrees model for CS shows markedly higher error on pure OPC compositions (FA = 0%, mean RMSE ≈ 5.1 MPa) compared with FA-blend folds (1.2 MPa), reflecting the absence of pozzolanic chemistry in the training data when the OPC-only fold is withheld; CS predictions at FA = 0% should therefore be treated as indicative only. Additionally, all specimens were fabricated on a single extrusion-based printing system; the influence of process variables such as print speed, nozzle geometry, and interlayer time interval on prediction accuracy under differing hardware conditions remains untested. The small number of distinct compositions also manifests directly in the fold-level variability of LOMO performance: the standard deviation of ExtraTrees R^2 across seven folds reaches 0.253 for CS, indicating that the identity of the withheld composition substantially affects the test result. With the factorial grid restricted to five FA levels and three W/B levels, potentially important combinations, such as intermediate W/B values paired with elevated FA content, remain untested; a more diverse compositional design covering a wider FA–W/B parameter space would reduce fold sensitivity and provide more reliable estimates of composition-level generalisation. External validation against an independent dataset remains a priority for future work.

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: conceptualisation, Jia Chen and Jianbo Huang; methodology, Jia Chen and Jianbo Huang; software, Jia Chen and Zhicheng Liao; formal analysis, Jia Chen; investigation, Mengdi Hou; data curation, Mengdi Hou; writing—original draft preparation, Jia Chen; writing—review and editing, Jianbo Huang; visualisation, Zhicheng Liao; supervision, Jianbo Huang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The experimental dataset and source code are openly available at https://github.com/lucassivan/ML-FA-3DPC.

Ethics Approval: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

3DPC	Three-dimensional printed concrete
CS	Compressive strength
FA	Fly ash
FS	Flexural strength
GPR	Gaussian process regression
GUI	Graphical user interface
LOMO	Leave-one-mix-out
MAE	Mean absolute error
ML	Machine learning
OPC	Ordinary Portland cement
PI	Prediction interval
RMSE	Root mean square error
SHAP	SHapley Additive exPlanations
SP	Superplasticizer
SVR	Support vector regression
TPE	Tree-structured Parzen Estimator
W/B	Water-to-binder ratio
S_1	Sobol first-order sensitivity index
S_T	Sobol total-order sensitivity index

References

1. Dilawar Riaz R, Usman M, Ali A, Majid U, Faizan M, Jalil Malik U. Inclusive characterization of 3D printed concrete (3DPC) in additive manufacturing: a detailed review. Constr Build Mater. 2023;394:132229. doi:10.1016/j.conbuildmat.2023.132229.
2. Besklubova S, Tan BQ, Zhong RY, Spicek N. Logistic cost analysis for 3D printing construction projects using a multi-stage network-based approach. Autom Constr. 2023;151(4):104863. doi:10.1016/j.autcon.2023.104863.
3. Wang X, Jia L, Jia Z, Zhang C, Chen Y, Ma L, et al. Optimization of 3D printing concrete with coarse aggregate via proper mix design and printing process. J Build Eng. 2022;56:104745. doi:10.1016/j.jobe.2022.104745.
4. Khan SA, Ghazi SMU, Amjad H, Imran M, Khushnood RA. Emerging horizons in 3D printed cement-based materials with nanomaterial integration: a review. Constr Build Mater. 2024;411:134815. doi:10.1016/j.conbuildmat.2023.134815.
5. Zhong H, Zhang M. 3D printing geopolymers: a review. Cem Concr Compos. 2022;128(3):104455. doi:10.1016/j.cemconcomp.2022.104455.
6. Rivera F, Martínez P, Castro J, López M. Massive volume fly-ash concrete: a more sustainable material with fly ash replacing cement and aggregates. Cem Concr Compos. 2015;63(4):104–12. doi:10.1016/j.cemconcomp.2015.08.001.
7. Oner A, Akyuz S, Yildiz R. An experimental study on strength development of concrete containing fly ash and optimum usage of fly ash in concrete. Cem Concr Res. 2005;35(6):1165–71. doi:10.1016/j.cemconres.2004.09.031.
8. Siddique R. Performance characteristics of high-volume class F fly ash concrete. Cem Concr Res. 2004;34(3):487–93. doi:10.1016/j.cemconres.2003.09.002.
9. Tkaczewska E. Effect of the superplasticizer type on the properties of the fly ash blended cement. Constr Build Mater. 2014;70:388–93. doi:10.1016/j.conbuildmat.2014.07.096.
10. Aparna S, Sathyan D, Anand KB. Microstructural and rate of water absorption study on fly-ash incorporated cement mortar. Mater Today Proc. 2018;5(11):23692–701. doi:10.1016/j.matpr.2018.10.159.
11. Hsu S, Chi M, Huang R. Effect of fineness and replacement ratio of ground fly ash on properties of blended cement mortar. Constr Build Mater. 2018;176(6):250–8. doi:10.1016/j.conbuildmat.2018.05.060.
12. Chalee W, Soeurt R, Pachana P, Songpiriyakij S. Improvement of high-volume fly ash cementitious material using single alkali activation. Int J Concr Struct Mater. 2021;15(1):44. doi:10.1186/s40069-021-00482-9.
13. Babalu R, Anil A, Sudarshan K, Amol P. Compressive strength, flexural strength, and durability of high-volume fly ash concrete. Innov Infrastruct Solut. 2023;8(5):154. doi:10.1007/s41062-023-01120-x.
14. Mohamed OA, Najm O, Ahmed E. Alkali-activated slag and fly ash as sustainable alternatives to OPC: sorptivity and strength development characteristics of mortar. Clean Mater. 2023;8(10):100188. doi:10.1016/j.clema.2023.100188.
15. Panda B, Tan MJ. Experimental study on mix proportion and fresh properties of fly ash based geopolymer for 3D concrete printing. Ceram Int. 2018;44(9):10258–65. doi:10.1016/j.ceramint.2018.03.031.
16. Marczyk J, Ziejewska C, kadek GS, Korniejenko K, Łach M, Góra M, et al. Hybrid materials based on fly ash, metakaolin, and cement for 3D printing. Materials. 2021;14(22):6874. doi:10.3390/ma14226874.
17. Guo X, Yang J, Xiong G. Influence of supplementary cementitious materials on rheological properties of 3D printed fly ash based geopolymer. Cem Concr Compos. 2020;114(6):103820. doi:10.1016/j.cemconcomp.2020.103820.
18. Kondepudi K, Subramaniam KVL. Formulation of alkali-activated fly ash-slag binders for 3D concrete printing. Cem Concr Compos. 2021;119(8):103983. doi:10.1016/j.cemconcomp.2021.103983.
19. Kaya E, Ciza B, Yalçınkaya Ç, Felekoğlu B, Yazıcı H, Çopuroğlu O. A comparative study on the effectiveness of fly ash and blast furnace slag as partial cement substitution in 3D printable concrete. J Build Eng. 2025;108:112841. doi:10.1016/j.jobe.2025.112841.
20. Mishra SK, Upadhyay B, Das BB. 3D printing aspects of fly ash and GGBS admixed binary and ternary blended cementitious mortar. Eur J Environ Civ Eng. 2025;29(15):3393–423. doi:10.1080/19648189.2025.2521381.
21. Liu X, Wang N, Zhang Y, Ma G. Optimization of printing precision and mechanical property for powder-based 3D printed magnesium phosphate cement using fly ash. Cem Concr Compos. 2024;148(4):105482. doi:10.1016/j.cemconcomp.2024.105482.

22. Tseng KC, Chi M, Yeih W, Huang R. Influence of slag/fly ash as partial cement replacement on printability and mechanical properties of 3D-printed concrete. Appl Sci. 2025;15(7):3933. doi:10.3390/app15073933.
23. Wang B, Zhai M, Yao X, Wu Q, Yang M, Wang X, et al. Printable and mechanical performance of 3D printed concrete employing multiple industrial wastes. Buildings. 2022;12(3):374. doi:10.3390/buildings12030374.
24. Xu Z, Zhang D, Li H, Sun X, Zhao K, Wang Y. Effect of FA and GGBFS on compressive strength, rheology, and printing properties of cement-based 3D printing material. Constr Build Mater. 2022;339(3):127685. doi:10.1016/j.conbuildmat.2022.127685.
25. Nguyen-Sy T, Wakim J, To QD, Vu MN, Nguyen TD, Nguyen TT. Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. Constr Build Mater. 2020;260(5):119757. doi:10.1016/j.conbuildmat.2020.119757.
26. Peng C, Xiang S, Liu W, Wang X, Li B, Wu S. Prediction of compressive strength and calibration of mesoscopic parameters of particle discrete element model based on explainable machine learning. Powder Technol. 2026;476(3):122427. doi:10.1016/j.powtec.2026.122427.
27. Du X, Liu W, Peng C, Huang B, Xu C, Sun Y, et al. Study on micro-parameters of parallel bond model based on machine learning algorithm. Comput Part Mech. 2025;12(5):2637–54. doi:10.1007/s40571-025-00938-9.
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. Akter U, Rezvi SE, MdJP R, Kishor SK. Optimization of data-driven ensemble models using firefly algorithm for enhanced 3D printed concrete strength prediction. In: Proceedings of the 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN); 2025 Jul 31–Aug 2; Rangpur, Bangladesh. Piscataway, NJ, USA: IEEE; 2025. p. 1–6. doi:10.1109/QPAIN66474.2025.11171958.
30. Schossler RT, Ullah S, Alajlan Z, Yu X. Data-driven analysis in 3D concrete printing: predicting and optimizing construction mixtures. AI Civ Eng. 2025;4(1):1. doi:10.1007/s43503-024-00044-4.
31. Hematibahar M, Kharun M, Fediuk R, Vatin NI, Porvadov MG, Sabitov LS. Predicting the flexural strength of 3D-printed geopolymer reinforced concrete using machine learning techniques. Mater Phys Mech. 2025;53(4):22–34. doi:10.18149/MPM.5342025_2.
32. 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.
33. 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.
34. 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.
35. 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.
36. Herman J, Usher W. SALib: an open-source Python library for sensitivity analysis. J Open Source Softw. 2017;2(9):97. doi:10.21105/joss.00097.