

Article

Toward Eco-Intelligent Concrete for Resilient Urban Infrastructure: Explainable Surrogate Optimization and LLM-Assisted Low-Clinker Mix Design

Junyi Zhang ¹, Haidong Yang ¹ , Guo Hu ^{1,*}, Jun Wu ² and Xiaotian Wu ³

¹ School of Urban Rail Transportation, Shanghai University of Engineering Science, Shanghai 201620, China; m405124223@sues.edu.cn (J.Z.); m405123416@sues.edu.cn (H.Y.)

² School of Civil Engineering, Shanghai Normal University, Shanghai 201418, China; cvewujun@shnu.edu.cn

³ School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; 2251221107@stu.xaut.edu.cn

* Correspondence: huguo@sues.edu.cn

Abstract

Concrete mix design increasingly requires rapid screening of mixture proportions against mechanical and resource-efficiency targets. This study develops an explainable, cement-reduction-oriented computational screening framework based on the UCI Concrete Compressive Strength dataset. The dataset records Portland cement, fly ash, blast furnace slag, aggregates, water, superplasticizer, curing age, and compressive strength, but does not report clinker factor, material-specific emission factors, or durability performance. Cement dosage is therefore minimized only as a surrogate objective; the study does not claim quantified embodied-carbon optimization. A LightGBM surrogate was trained using raw mixture variables and domain-informed ratios. On the held-out test set, the model achieved an R^2 of 0.946 and a mean absolute error of 2.639 MPa. Shapley Additive Explanations were used to examine the statistical influence of mixture variables, with the water-to-binder ratio emerging as the dominant predictor. Constraint-filtered Monte Carlo sampling and non-dominated sorting were then used to screen Pareto-efficient binder allocations. For a 28-day target of 45 MPa, repeated searches identified candidates with a mean Portland cement dosage of 164.4 kg/m³, 44.2% below the mean of empirical mixtures in the same strength band. The evaluated numerical modules were embedded in a Streamlit prototype in which a DeepSeek large language model performs only intent parsing and report generation. The main contribution is this tool-augmented separation of language interaction from deterministic engineering computation. The resulting mixtures remain computational candidates and should next be validated experimentally and assessed using material-specific life-cycle carbon and durability data.



Academic Editor: Antonio Caggiano

Received: 15 July 2026

Revised: 31 July 2026

Accepted: 5 August 2026

Published: 6 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: concrete mix screening; cement reduction; explainable machine learning; Pareto filtering; tool-augmented large language model; supplementary cementitious materials

1. Introduction

Concrete production must respond simultaneously to mechanical performance requirements and pressure to reduce resource use and greenhouse-gas emissions [1,2]. Replacing part of ordinary Portland cement (OPC) with supplementary cementitious materials (SCMs), including fly ash, blast furnace slag, and recycled powders, is one established route toward lower-clinker binders [3–9]. The use of industrial by-products also connects concrete

technology with the broader concept of construction-sector industrial symbiosis, in which material exchanges are organized across firms and supply chains [10]. The environmental benefit of an individual mixture, however, cannot be inferred from cement mass alone: it depends on clinker factor, SCM processing, transport, regional emission factors, service life, and durability.

SCM-containing mixtures exhibit coupled effects among cement, mineral admixtures, water, aggregate, chemical admixtures, and curing age. Traditional proportioning relies on rules such as Abrams' Law [11], trial batching, and performance testing. Conservative designs can produce overstrength and overdesign [12], while broad trial searches can be costly. Data-driven screening can help prioritize candidates for testing by representing non-linear effects such as the latent hydraulic reaction of slag and delayed pozzolanic reaction of fly ash [13–15]. It should be treated as a complement to, rather than a replacement for, empirical proportioning and laboratory validation.

Machine learning prediction, transfer learning, metaheuristic optimization, and multi-objective concrete proportioning are each well-established research directions [16–34]. Recent studies in Buildings, for example, have examined strength–cost–CO₂ trade-offs, AI-assisted geopolymer proportioning, ensemble-based low-carbon design platforms, and interpretable prediction of strength and workability [35–38]. Likewise, LightGBM combined with Shapley Additive Explanations (SHAP) should not by itself be regarded as a novel contribution; SHAP describes the behavior of a fitted model and does not establish material causality [39–43].

The closest recent comparator is GreenMix-Pareto [44], which uses a similar benchmark-scale concrete dataset and combines physics-guided gradient boosting, monotonicity constraints, conformal uncertainty calibration, feasibility learning, and hybrid Bayesian optimization/NSGA-II to optimize strength and quantified environmental indicators. The present study does not reproduce those uncertainty or life-cycle objectives and does not claim algorithmic superiority. Its numerical pipeline instead provides supporting infrastructure for a different contribution: a tool-augmented large language model (LLM) interface in which language interaction is separated from all deterministic prediction, constraint filtering, and Pareto calculations.

Three practical gaps motivate this architecture. First, benchmark accuracy alone does not reveal whether a model has learned trends consistent with basic concrete proportioning; ratio-based features and SHAP can provide a statistical consistency check, subject to the limits of correlated features [39,42]. Second, engineers often need an inverse screening task—given a target age and strength, identify admissible candidates with lower cement dosage—whereas many published models address forward prediction only. Third, research scripts often require manual parameter entry and programming expertise. A natural language interface can reduce this operational barrier only if the LLM is prevented from inventing numerical results and every recommendation remains traceable to the evaluated back-end tools [45–47].

Accordingly, this study develops a cement-reduction-oriented screening workflow rather than a “low-carbon” or “low-clinker” optimizer. The term “eco-intelligent concrete” denotes the explainable, data-driven design workflow rather than a smart, self-sensing, or self-actuating material, while “resilient urban infrastructure” denotes the intended application context rather than a validated infrastructure-level outcome. Likewise, “low-clinker mix design” denotes the longer-term design direction; because clinker fraction is not reported, the implemented objective in this study is limited to OPC-dose reduction. The UCI dataset records OPC dosage and two SCM dosages, but it contains no cement chemistry, clinker factor, material-specific emission factor, or life-cycle inventory. Minimizing OPC dosage is therefore used as a surrogate screening objective and is not equated with quanti-

fied carbon reduction. The generated mixtures are candidates for laboratory evaluation, not construction-ready prescriptions.

The study makes three scoped contributions. First, it implements and demonstrates a tool-augmented LLM architecture in which the language model is limited to intent parsing and report generation, while evaluated deterministic modules retain numerical authority. Second, it builds a transparent supporting pipeline that combines domain-informed ratios, benchmark surrogate comparison, SHAP-based statistical interpretation, constraint-filtered Monte Carlo sampling, and non-dominated sorting. These individual methods are not claimed as new; their role is to provide an auditable numerical toolchain for the interface. Third, it reports a clearly bounded empirical comparison of cement dosage within the UCI data domain and identifies the experimental, uncertainty, durability, and life-cycle information required before any engineering application. The technical workflow and the step-by-step procedure are presented at the beginning of Section 2.

2. Methodology

This section translates the study objective into five auditable steps. First, the UCI records are curated and ratio-based features are constructed. Second, four surrogate models are evaluated and the selected model is checked by cross-validation. Third, SHAP is used to inspect model behavior. Fourth, Monte Carlo candidates are filtered by declared proportioning constraints and ranked by non-dominated sorting. Fifth, the numerical tools are exposed to a tool-augmented LLM interface through structured inputs and outputs. Figure 1 summarizes this sequence; the figure is descriptive, while the equations and implementation details that follow define each computational step.

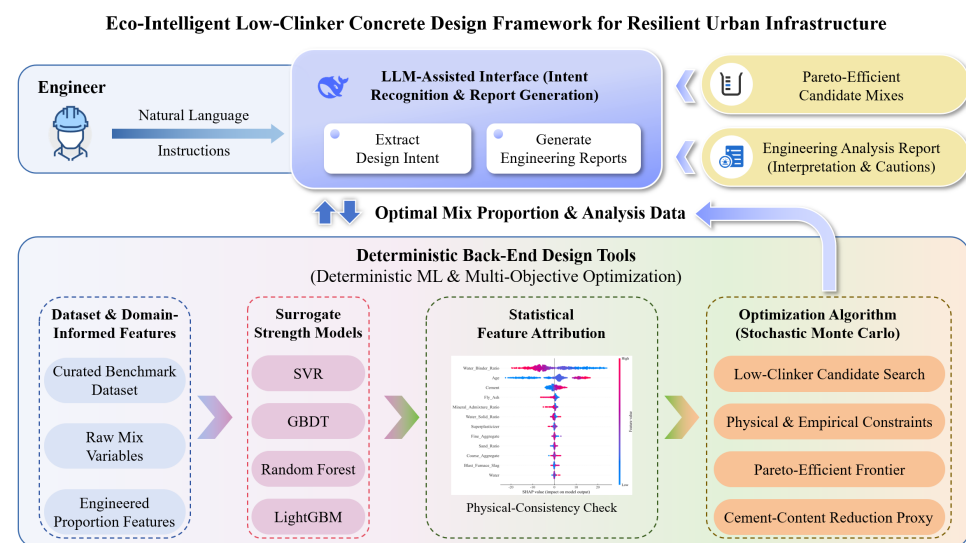


Figure 1. Workflow of the proposed computational screening framework.

2.1. Dataset Description and Domain-Specific Feature Engineering

2.1.1. Dataset Description and Preprocessing

The Concrete Compressive Strength dataset was obtained from the UCI Machine Learning Repository [48]. It contains 1030 laboratory-tested records with eight quantitative inputs: Portland cement dosage (C), blast furnace slag (S), fly ash (F), water (W), superplasticizer (SP), coarse aggregate (CA), fine aggregate (FA), and curing age (A). Age ranges from 1 to 365 days, and the target is measured compressive strength, f_c , in MPa. The dataset documentation does not identify the cement strength class, clinker fraction, chemical composition, SCM source or reactivity, aggregate grading, curing environment, slump, air content, or durability. Accordingly, C is treated as the reported Portland cement

mass, while S and F are the only reported supplementary binder constituents; no inference is made about the actual clinker factor or embodied carbon of a record.

The database was curated before surrogate development. First, 25 exact duplicate records were removed so that repeated rows would not artificially inflate prediction accuracy. Second, an interquartile range (IQR) filter was used to flag statistical extremes [49]. For a variable x , the first and third quartiles are Q_1 and Q_3 , and $IQR = Q_3 - Q_1$. A retained observation satisfies Equation (1).

$$Q_1 - 1.5 \times IQR \leq x_i \leq Q_3 + 1.5 \times IQR \quad (1)$$

Applying this rule across the variables excluded 94 records, mainly because of extreme age, water, or superplasticizer values, and left 911 retained records. The term “statistical extreme” does not imply experimental error: uncommon but legitimate high-performance mixtures may fall outside an IQR envelope. The exclusion was a modeling choice that narrows the interpolation domain and may reduce applicability to extreme formulations. The reported model and optimization results are therefore valid only for the retained benchmark domain; external validation and an all-record sensitivity study are still required before the workflow is applied to unusual mixtures.

Figure 2 presents the distributions of the retained records. Measured strengths extend from approximately 2 to 77 MPa across ages of 1–365 days, with many observations at the conventional 28-day test age [50]. In this manuscript, “lower-strength” and “higher-strength” refer descriptively to values below about 30 MPa and above about 60 MPa within this retained dataset; they are not code-based strength classes and should not be interpreted as fixed 28-day grades. All-age data are used for surrogate training, whereas the Pareto plots and demonstration case in Section 3 fix age at 28 days unless stated otherwise.

2.1.2. Domain-Specific Feature Engineering

Raw material masses do not fully express the relative proportions that govern hydration, aggregate packing, and strength development. Accordingly, four dimensionless features were defined before model training. In Equations (2)–(5), C , S , F , W , FA , and CA denote Portland cement, slag, fly ash, water, fine aggregate, and coarse aggregate, respectively. Total binder is $B = C + S + F$, and total aggregate is $G = FA + CA$.

The water-to-binder ratio (W/B) expresses water relative to total binder:

$$W/B = \frac{W}{C + S + F} \quad (2)$$

The sand ratio (SR) expresses fine aggregate as a fraction of total aggregate:

$$SR = \frac{FA}{FA + CA} \quad (3)$$

The mineral admixture ratio (MAR) expresses the reported SCM mass as a fraction of total binder:

$$MAR = \frac{S + F}{C + S + F} \quad (4)$$

The water-to-solid ratio (W/S) expresses water relative to the reported binder and aggregate solids:

$$W/S = \frac{W}{C + S + F + FA + CA} \quad (5)$$

The W/B ratio is associated with capillary porosity and compressive strength. SR represents the fine-aggregate fraction of the aggregate skeleton and is related to workability, packing, and segregation resistance [51]. MAR describes the fraction of the reported binder

mass supplied by slag and fly ash; because cement chemistry is unavailable, MAR is not treated as a measured clinker-replacement factor. W/S supplies an additional indicator of the mixture's overall water condition.

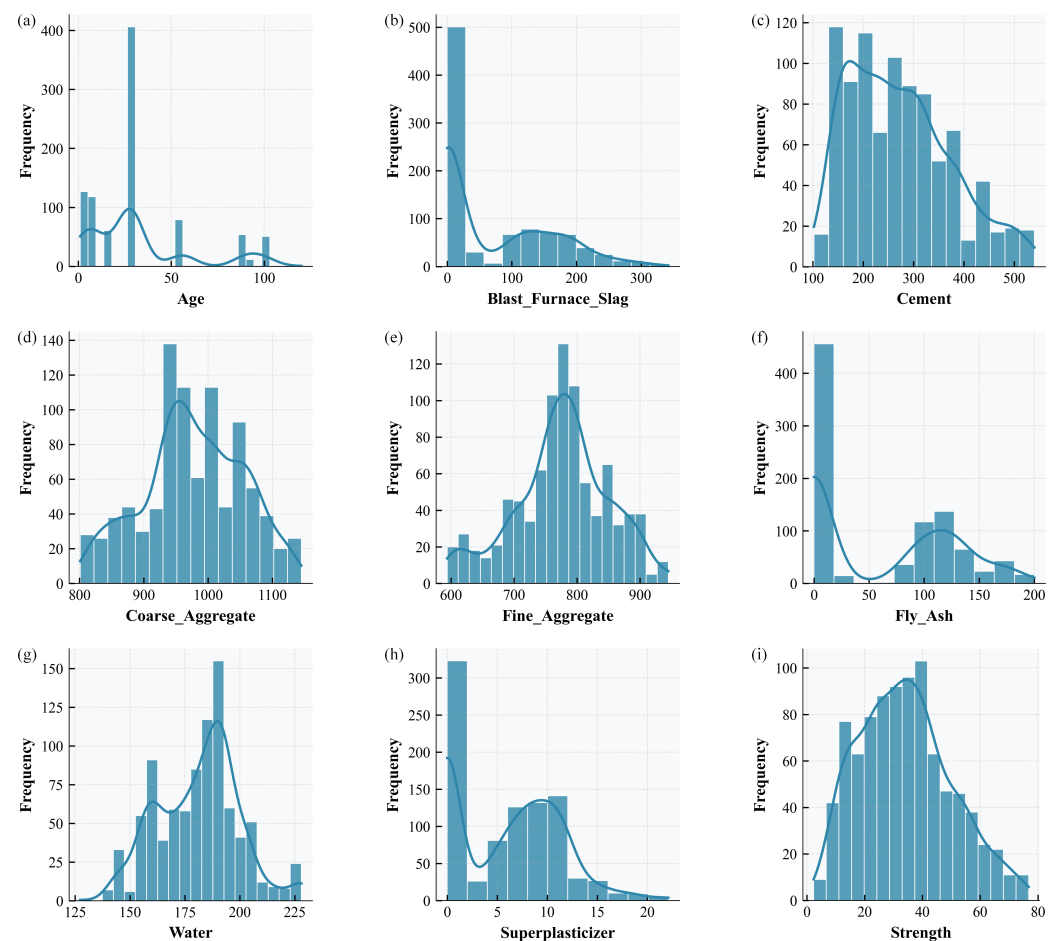


Figure 2. Distributions after preprocessing: (a) age, (b) slag, (c) cement, (d) coarse aggregate, (e) fine aggregate, (f) fly ash, (g) water, (h) superplasticizer, and (i) compressive strength.

The four features were used together with the eight raw variables. Figure 3 introduces their pairwise Pearson correlations before the nonlinear models are fitted. Both axes list the same feature set. Each circle corresponds to one row–column feature pair; circle area is proportional to the absolute correlation magnitude, while the blue–red scale gives the signed coefficient from -1 to $+1$. The diagonal therefore equals $+1$. W/B has a stronger negative linear correlation with strength than cement dosage alone because the ratio jointly represents water availability and binder mass. This is an initial descriptive check only: Pearson coefficients do not represent nonlinear interactions or causation, which are considered later using tree models and SHAP [52].

2.2. Development of Machine Learning Surrogate Models

The objective of surrogate modeling was to develop a fast and accurate predictor of concrete compressive strength that could be embedded into the subsequent multi-objective optimization framework. Four representative algorithms were selected for comparison: Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Decision Trees (GBDTs), and Light Gradient Boosting Machine (LightGBM). These models were chosen because they represent different learning mechanisms, including kernel-based regression, parallel ensemble learning, sequential boosting, and histogram-based gradient boosting.

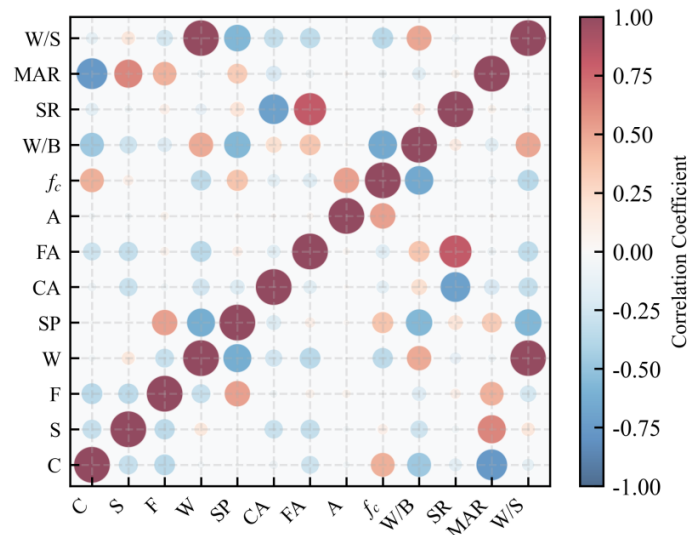


Figure 3. Bubble-coded Pearson correlations among raw and engineered variables.

The retained dataset was randomly divided into training and testing subsets using an 80/20 ratio. The training set was used for model fitting and hyperparameter tuning, while the testing set was reserved for evaluating generalization performance. To reduce the risk of instability associated with a single arbitrary split, model selection was based not only on final testing metrics but also on consistency between training and testing errors. Particular attention was given to potential overfitting because the surrogate model directly affects the reliability of the optimized mixture proportions.

2.2.1. Support Vector Regression (SVR)

SVR was included as a kernel-based baseline [53]. It fits an ϵ -insensitive regression function, and a radial basis function kernel represents nonlinear mixture–strength relationships. Figure 4 summarizes the two concepts; model suitability was assessed empirically on the same train–test split used for all benchmarks rather than being inferred from a general algorithmic claim.

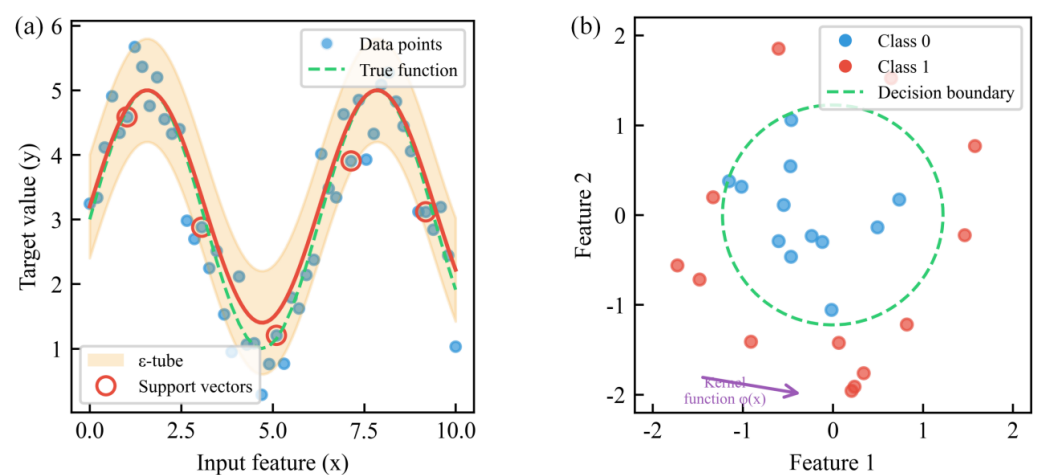


Figure 4. SVR principles: (a) the ϵ -insensitive tube and (b) kernel mapping.

2.2.2. Random Forest (RF)

RF was included as a bagging ensemble [54]. Bootstrap trees with random feature subsets were averaged to represent nonlinear thresholds and provide a comparison with the sequential boosting models.

2.2.3. Gradient Boosting Decision Trees (GBDTs)

GBDTs were included as a sequential boosting baseline [55]. Each tree fits the residual gradient of the current ensemble, allowing coupled effects among mixture variables to be represented without an explicit constitutive equation.

2.2.4. Light Gradient Boosting Machine (LightGBM)

LightGBM was the fourth candidate because its histogram binning and leaf-wise growth can provide efficient prediction on structured tabular data [56]. To define how a tree split is selected, let $I = I_L \cup I_R$ be the observations at a parent node and its left and right children, and let g_i and h_i be the first- and second-order derivatives of the training loss for observation i . For a candidate split at bin d , the regularized reduction in the second-order loss approximation is

$$Gain(d) = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (6)$$

Here, λ regularizes leaf weights and γ is the minimum gain required to create a split. The candidate with the largest positive gain is selected, so Equation (6) links histogram search to the training objective rather than appearing as an isolated formula.

As shown in Figure 5, LightGBM expands the leaf with the largest gain. The number of leaves, maximum depth, learning rate, feature fraction, bagging fraction, and minimum observations per leaf were tuned while monitoring held-out error to control overfitting.

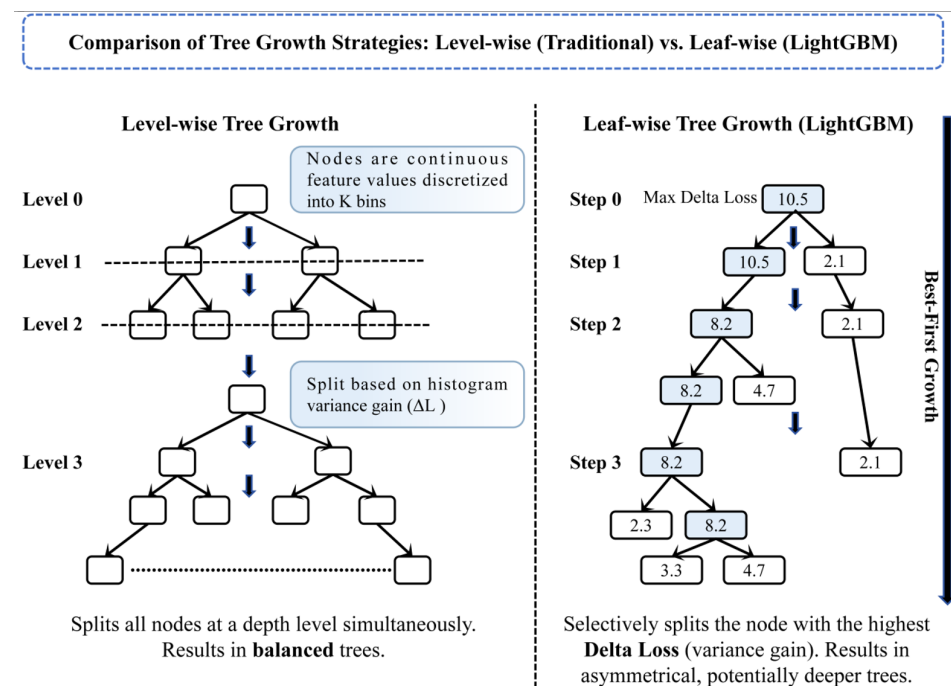


Figure 5. Level-wise and LightGBM leaf-wise tree growth.

2.2.5. K-Fold Cross-Validation

To further examine the robustness of the selected surrogate model and reduce the dependence of model evaluation on a single random train–test split, a k-fold cross-validation

procedure was additionally conducted on the retained dataset. In this study, five-fold cross-validation was adopted because it provides a practical balance between training efficiency and variance estimation for a medium-sized concrete database. The 911 retained samples were randomly partitioned into five approximately equal folds. In each iteration, four folds were used for model training, while the remaining fold was used for validation. This process was repeated five times so that each fold served once as the validation subset.

The same input feature set, including both raw mixture variables and domain-informed features, was used in all folds. The preprocessing steps, feature construction procedure, and evaluation metrics were kept consistent with the main train–test experiment. For the LightGBM model, the main hyperparameters were fixed consistently during cross-validation to avoid fold-specific overfitting. The final model was evaluated using the mean and standard deviation of R^2 , RMSE, and MAE across the five folds.

This cross-validation analysis was used as an additional stability check rather than as a replacement for the independent testing result. If the cross-validation metrics remain close to the 80/20 testing metrics, the surrogate model can be considered more stable within the empirical domain of the dataset. Conversely, large fold-to-fold variation would indicate that the model is sensitive to data partitioning and that the subsequent optimization results should be interpreted with greater caution.

2.3. Explainable AI Using SHAP

2.3.1. Role of Interpretability in Eco-Intelligent Concrete Design

High prediction accuracy alone is insufficient for practical material design. In concrete engineering, a surrogate model should not only predict compressive strength, but also exhibit feature-response patterns that are consistent with established material knowledge. For example, excessive W/B ratio should generally reduce strength because of increased capillary porosity, while the effect of fly ash may depend strongly on curing age owing to delayed pozzolanic reactions. If a model contradicts these basic trends, its optimized outputs would be difficult to justify in engineering practice.

Therefore, Shapley Additive Explanations (SHAP) were used to interpret the trained surrogate model [42]. In this study, SHAP was not used as direct proof of hydration mechanisms. Instead, it was employed as a statistical interpretation tool to examine whether the learned relationships between mixture variables and predicted compressive strength are broadly consistent with known concrete behavior. This distinction is important because data-driven feature attribution can reveal model logic, but it cannot replace independent experimental, microstructural, or durability validation.

2.3.2. Mathematical Formulation of SHAP

SHAP is based on Shapley values from cooperative game theory. It treats the prediction of a machine learning model as a cooperative game, in which the input features act as players and the model output is the payoff. For a trained model f and feature i , the Shapley value ϕ_i is calculated as

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_x(S \cup \{i\}) - f_x(S)] \quad (7)$$

Here, F denotes the complete set of input features, specifically the concrete mix components and engineered ratios, and $|F|$ is the total number of features. S represents a subset of features excluding feature i . The term $f_x(S \cup \{i\}) - f_x(S)$ captures the exact marginal effect of introducing feature i into subset S . The leading combinatorial coefficient weights this calculation, ensuring that the marginal contribution is averaged systematically across all possible feature permutations.

2.3.3. Global and Local Interpretation Using TreeExplainer

Because the final surrogate model is tree-based, the TreeExplainer algorithm was used to calculate SHAP values efficiently [43]. TreeExplainer reduces the computational cost of exact Shapley-value estimation for decision-tree ensembles and enables both global and local interpretation.

Two levels of interpretation were conducted. First, global SHAP summary plots were used to rank the relative importance of input variables and identify whether high or low feature values contributed positively or negatively to compressive strength prediction. This analysis provides an overall view of which mixture parameters dominate the model output. Second, SHAP dependence plots were used to examine nonlinear effects and feature interactions. Particular attention was given to the W/B ratio and fly ash content because these variables are directly related to concrete strength development and cement-reduction screening. The W/B dependence plot was used to check whether the model reproduced the expected inverse strength trend associated with Abrams' Law. The fly ash dependence plot was used to examine whether the model captured age-dependent statistical effects associated with delayed pozzolanic contribution.

This interpretability step serves as a physical-consistency check before the surrogate model is embedded into multi-objective optimization. It does not eliminate the need for experimental verification, but it helps reduce the risk of using a high-accuracy model whose internal decision logic is inconsistent with fundamental concrete proportioning principles.

2.4. Cement-Reduction-Oriented Multi-Objective Screening

2.4.1. Formulation of the Inverse Design Problem

The trained surrogate performs forward strength prediction. The inverse screening task is to identify mixtures near a specified strength and age while reducing the reported Portland cement dosage. For all optimization results in Section 3, age is fixed at 28 days. This screening objective is narrower than low-clinker or low-carbon design because the dataset does not provide clinker fraction or emission factors.

In this study, the decision variable vector $\mathbf{x}$ was defined as

$$\mathbf{x} = [C, S, F, W, SP, CA, FA]^T \quad (8)$$

where C, S, F, W, SP, CA, and FA represent cement, blast furnace slag, fly ash, water, super-plasticizer, coarse aggregate, and fine aggregate, respectively. Curing age A was treated as a scenario parameter rather than an optimization variable and was fixed according to the user-defined design condition.

The multi-objective optimization problem was formulated as

$$\begin{cases} \min f_1(\mathbf{x}) = |\hat{f}_c(\mathbf{x}) - f_{\text{target}}| \\ \min f_2(\mathbf{x}) = C \end{cases} \quad (9)$$

where $\hat{f}_c(\mathbf{x})$ is the compressive strength predicted by the LightGBM surrogate model, and f_{target} is the user-defined target strength. The first objective minimizes the deviation between the predicted and target compressive strengths. The second objective minimizes reported cement dosage. It is used only as a surrogate resource-efficiency objective, not as an embodied-carbon metric. A life-cycle objective would require material-specific emission factors, clinker fraction, SCM allocation and processing, transport, admixture production, curing, service life, and durability data.

2.4.2. Physical and Empirical Constraints

Unconstrained sampling can produce implausible proportions. The search was therefore limited by the study-defined bounds in Equation (10). These bounds are scenario settings for this benchmark, not universal code limits.

$$\left\{ \begin{array}{l} 2350 \leq \sum_{i=1}^7 x_i \leq 2450 \\ 0.25 \leq \frac{W}{C+S+F} \leq 0.80 \\ 0.35 \leq \frac{FA}{FA+CA} \leq 0.45 \\ x_j^{(\min)} \leq x_j \leq x_j^{(\max)}, j = 1, 2, \dots, p \end{array} \right. \quad (10)$$

The first expression constrains the nominal sum of the seven reported constituent masses per cubic meter. Because air content and constituent specific gravities are unavailable, this sum is not a measured density; the interval 2350–2450 kg/m³ is a narrow screening window centered on 2400 kg/m³, consistent with normal-density proportioning and the absolute-volume framework of ACI PRC-211.1-22 [57]. Fresh density would need to be verified in accordance with ASTM C138/C138M [58]. The W/B and SR intervals were selected to retain a broad normal-concrete proportioning domain while filtering visibly extreme ratios [51,57]. Finally, $x_j^{(\min)}$ and $x_j^{(\max)}$ are the retained-data bounds for variable j , and p is the number of optimized variables. The last constraint prevents numerical extrapolation beyond that empirical domain.

In addition, the mineral admixture ratio was constrained as

$$MAR = \frac{S + F}{C + S + F} \leq 0.60 \quad (11)$$

The MAR cap of 0.60 is likewise a study-defined scenario boundary. It was selected to permit high SCM use while avoiding an unconstrained search at the edge of a dataset that lacks reactivity, exposure, and curing descriptors. It is not attributed to ACI 318 and is not a blanket replacement limit. ACI requirements depend on material, exposure, and performance conditions, while proportioning guidance requires trial batches [57]. No full multi-bound sensitivity analysis was performed in the present study; only repeated-seed stability was evaluated. Consequently, every reported Pareto set is conditional on the declared bounds, and MAR should be recalibrated using site-specific SCM performance and durability tests.

2.4.3. Constraint-Filtered Sampling and Pareto Extraction

The search procedure is a transparent Monte Carlo screening method, followed by constraint filtering and non-dominated sorting; it is not an evolutionary optimizer. The hybrid candidate pool contains the 911 retained records as empirical anchors and randomly sampled virtual mixtures within the retained variable bounds. Random search was chosen as an auditable baseline with low implementation overhead for the seven-variable screening space [59]. It is not claimed to be equivalent or superior to NSGA-II, differential evolution, or Bayesian optimization. In particular, GreenMix-Pareto uses uncertainty-aware Bayesian search and NSGA-II refinement [44]; those methods are a more advanced option when calibrated uncertainty and environmental objectives are available.

Candidate mixtures were generated within the observed material bounds, filtered by nominal constituent mass, W/B, SR, MAR, and individual variable limits, and then augmented with the four engineered features. The LightGBM surrogate predicted 28-day strength for the admissible candidates. For a user-specified target, a candidate was non-dominated when no other admissible point had both lower cement dosage and smaller absolute target deviation. For the global strength–cement plots in Section 3, non-dominance

is presented equivalently as the trade-off between maximizing predicted strength and minimizing cement dosage.

The extracted front is the sampled efficiency boundary conditional on the surrogate, random candidate pool, and constraints. It is not proof of a globally optimal formulation. For a user-defined target, the interface returns the non-dominated candidate with the smallest strength deviation inside the stated tolerance and labels it for experimental verification.

2.4.4. Repeated-Seed Stability and Constraint Dependence

Because Monte Carlo results depend on the sampled points, the same procedure was repeated under independent random seeds to test whether similar candidates recur. This evaluates sampling stability only. The MAR cap, W/B interval, SR interval, and nominal mass window were held fixed; therefore, the repeated runs do not constitute a constraint sensitivity analysis.

The front is shaped jointly by the surrogate, empirical distribution, random candidate pool, and declared constraints. Returned mixtures are therefore computational candidates, not construction-ready prescriptions. Slump, air content, setting, segregation, measured strength, shrinkage, and durability must be tested before any practical use, and boundary sensitivity should be reported when local materials are introduced.

2.5. LLM-Assisted Interactive Framework

2.5.1. Motivation for a Tool-Augmented Architecture

The surrogate and screening procedure provide a deterministic numerical workflow, but their direct use requires programming, parameter entry, and interpretation of model outputs. The interface is intended to lower this operational barrier for preliminary screening.

Large language models (LLMs) are effective at processing natural language and generating human-readable explanations, but they are not suitable as standalone numerical engines for structural or material design. Unconstrained generative reasoning may produce arithmetic mistakes, unsupported claims, or fabricated mixture recommendations. Therefore, this study adopts a tool-augmented architecture in which the LLM does not calculate strength, optimize mix proportions, or modify the physical constraints. Instead, deterministic back-end tools perform all numerical operations, while the LLM is used only for semantic intent recognition and engineering report synthesis [46,47,60].

2.5.2. Architecture and Workflow

The interactive prototype was developed using Streamlit and organized into three layers [61]. The user-interface layer accepts a query such as “Screen a cement-reduced mixture for 45 MPa at 28 days.” The deterministic layer performs feature construction, LightGBM prediction, constraint filtering, and Pareto ranking. The report layer converts the structured result into a readable explanation. Figure 6 shows the separation of these roles.

The LLM first extracts the user intent and target strength from the input query. The extracted information is converted into standardized execution keys, such as target strength, design age, and optimization mode. These keys are then passed to the deterministic optimization module. The LightGBM surrogate model and Pareto search algorithm evaluate admissible candidate mixtures and return structured numerical outputs, including material dosages, predicted strength, cement content, and SCM replacement ratio.

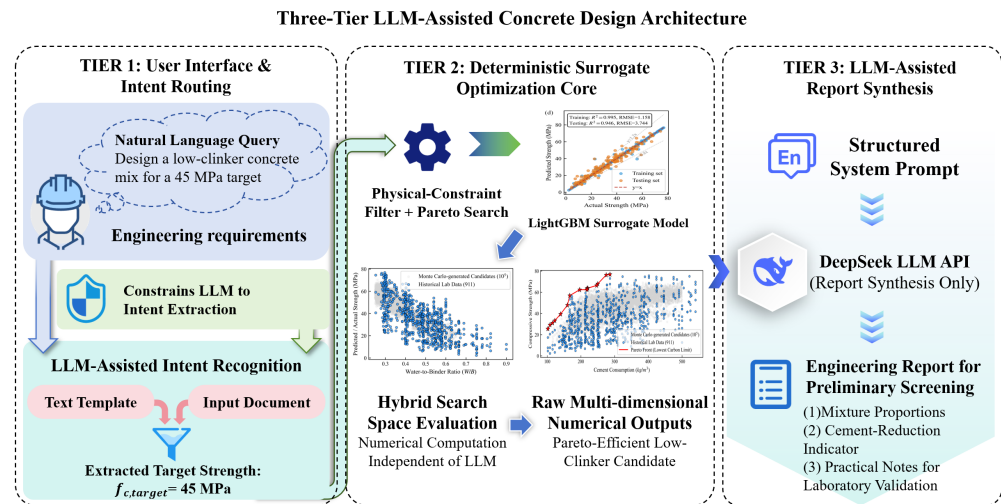


Figure 6. Tool-augmented LLM workflow with deterministic numerical computation.

After the numerical results are produced, the LLM generates an engineering-style report based on the structured output. The report includes the optimized mixture proportions, the predicted strength, the cement-reduction implication, and practical cautions regarding experimental validation and durability assessment. Because the LLM receives only structured data from the deterministic module, it functions as an explanatory interface rather than a numerical decision-maker. This design reduces the risk of ungrounded generative calculation, but it does not eliminate language-model failure.

2.5.3. Safeguards, Failure Modes, and User Verification

The prototype uses four architectural safeguards. First, the system prompt restricts the LLM to intent extraction and report wording. Second, numerical tools accept standardized execution keys rather than free-form mixture values. Third, the report is generated only after the deterministic module returns a structured result; model-written numbers are not used as optimization inputs. Fourth, every report states that the candidate requires engineering review and experimental validation.

These controls do not constitute a formal safety certification. Failure can still occur when a request is ambiguous, uses inconsistent units, specifies a target outside the retained data domain, or produces a malformed or unsupported intent. The present study did not quantify hallucination rate, adversarial-prompt robustness, schema-conformance rate, or user error. A user must verify the parsed target and age, confirm that all dosages come from the deterministic result, check constraint compliance, and reject any candidate outside the documented scope. Future deployment should add strict schema validation, out-of-domain rejection, logged tool traces, regression tests for known failure cases, and human approval before report export.

2.6. Performance Evaluation Metrics

The predictive performance of the surrogate models was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). These metrics were calculated as follows [62]:

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

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

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

where n is the number of samples, y_i is the experimentally measured compressive strength of the i -th sample, $\hat{y}_i$ is the model-predicted compressive strength, and $\bar{y}$ is the mean measured compressive strength. R^2 measures the proportion of variance explained by the model. $RMSE$ gives greater weight to large prediction errors and is therefore important for identifying potentially unsafe deviations in structural material prediction. MAE represents the average absolute prediction error and provides an intuitive measure of practical prediction accuracy.

In addition to these conventional metrics, the consistency between training and testing performance was considered when selecting the final surrogate model. A model with high training accuracy but substantially lower testing accuracy would indicate possible overfitting and would be unsuitable for downstream optimization. Therefore, the final model was selected based on both high testing performance and acceptable generalization behavior. The selected surrogate model was then used for SHAP interpretation and multi-objective optimization.

For the present study, the screening results should be interpreted within the prediction uncertainty and data-domain limitations of the surrogate model. The current framework provides a deterministic screening tool for identifying promising cement-reduced candidates. Future extensions may incorporate conformal prediction, quantile regression, Bayesian ensemble methods, or experimentally calibrated safety factors to transform the current deterministic optimization into a fully uncertainty-aware robust design framework [63,64].

3. Results and Discussion

3.1. Performance Evaluation of Machine Learning Surrogate Models

The predictive performance of the four candidate surrogate models was first evaluated to determine which model was most suitable for subsequent explainable analysis and multi-objective mix optimization. Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Decision Trees (GBDTs), and LightGBM were trained using the same retained dataset and domain-informed input features. The dataset was divided into a training subset and an unseen testing subset using an 80/20 split. Because the optimization module relies directly on surrogate predictions, both fitting ability and generalization performance were considered.

Table 1 summarizes the training-phase results. All models exhibited acceptable fitting capacity, indicating that the selected raw variables and engineered features contain sufficient information for compressive strength prediction. Among the four models, LightGBM achieved the highest training accuracy, with an R^2 of 0.995, RMSE of 1.158 MPa, and MAE of 0.235 MPa. RF and GBDT also showed strong fitting performance, whereas SVR produced slightly higher training errors. The near-perfect training performance of LightGBM indicates strong nonlinear fitting capacity, but it also requires careful interpretation because excessive divergence between training and testing performance may suggest overfitting. Therefore, the testing results were used as the primary basis for model selection.

Table 1. Performance evaluation of the four benchmark models on the training dataset.

Model	R^2	RMSE (MPa)	MAE (MPa)
LightGBM	0.995	1.158	0.235
Random Forest	0.984	2.020	1.329
GBDT	0.979	2.275	1.618
SVR	0.945	3.691	2.148

Testing results are compared in Figure 7. Separate two-dimensional panels are used for R^2 and error metrics so that unlike scales are not distorted by a three-dimensional perspective.

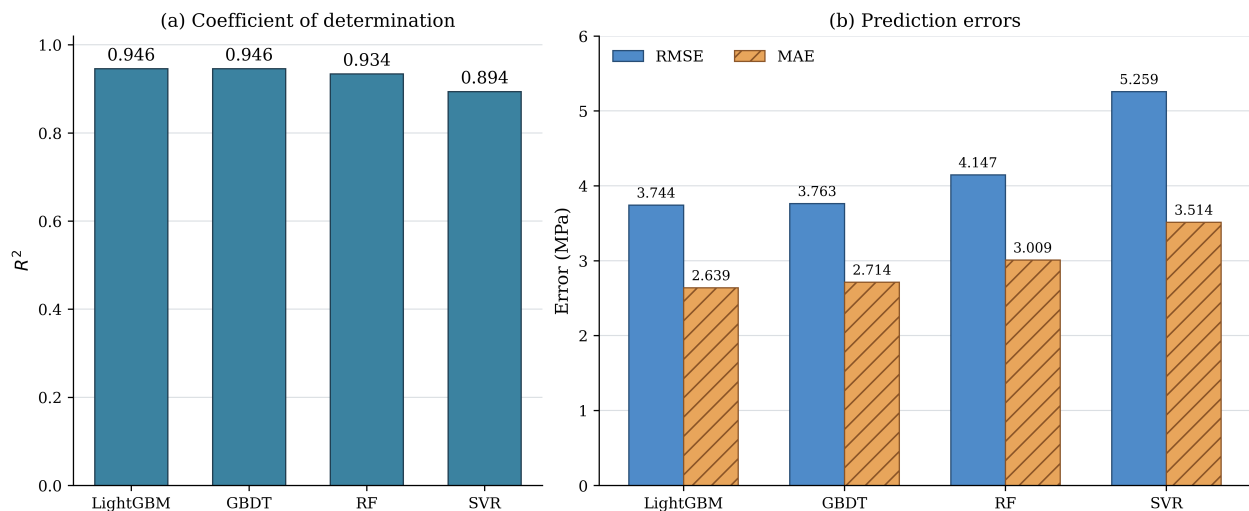


Figure 7. Held-out testing performance of the four surrogate models.

The tree-based ensembles outperformed the kernel-based SVR model on this split. SVR achieved an R^2 of 0.894 and RMSE of 5.259 MPa, whereas RF and GBDT reduced the held-out errors.

LightGBM achieved the best testing performance, with an R^2 of 0.946, RMSE of 3.744 MPa, and MAE of 2.639 MPa. Although GBDT produced the same rounded testing R^2 of 0.946, LightGBM provided slightly lower error and better computational efficiency because of its histogram-based split strategy and leaf-wise growth mechanism [56]. These characteristics make LightGBM suitable for rapid screening of large candidate mixture spaces in the subsequent Pareto optimization stage. However, the difference between the LightGBM training R^2 and testing R^2 also indicates that the model should be used as a high-fidelity screening surrogate rather than as a replacement for experimental validation. This distinction is particularly important when the optimization algorithm explores mixture combinations that are sparse in the original dataset.

To further examine whether the reported model performance was overly dependent on the random 80/20 split, five-fold cross-validation was conducted for the selected LightGBM surrogate model. Table 2 summarizes the validation results for each fold. The cross-validation results were highly comparable to the independent testing result, with a mean R^2 of 0.946, mean RMSE of 3.741 MPa, and mean MAE of 2.637 MPa. The corresponding standard deviations were 0.005 for R^2 , 0.155 MPa for RMSE, and 0.098 MPa for MAE. These small fold-to-fold variations indicate that the model performance is reasonably stable within the current benchmark dataset.

Nevertheless, the cross-validation results should not be interpreted as external validation. The folds were generated from the same curated UCI database, and therefore the analysis only evaluates internal data-partition stability. Independent datasets from different laboratories, material sources, curing regimes, and regional batching practices would still be required to confirm the broader generalizability of the surrogate model.

Figure 8 presents scatter plots of measured versus predicted compressive strength for the four models. The red dashed diagonal line represents perfect prediction. Compared with the other models, the LightGBM predictions are more tightly distributed around the diagonal line, especially within the main strength range of approximately 20–70 MPa. This indicates that the model effectively captures the dominant statistical relationships in the dataset. A slightly larger dispersion can be observed in the very-high-strength region. This

may be related to two factors. First, high-strength concrete is generally more sensitive to small variations in water content, superplasticizer dosage, specimen preparation, and curing conditions [65]. Second, high-strength mixtures are less densely represented in the benchmark dataset, which may reduce the reliability of surrogate extrapolation in this region.

Table 2. Five-fold cross-validation performance of the LightGBM surrogate model.

Fold	R^2	RMSE (MPa)	MAE (MPa)
1	0.938	3.983	2.789
2	0.944	3.758	2.673
3	0.947	3.627	2.558
4	0.950	3.689	2.606
5	0.951	3.648	2.559
Mean	0.946	3.741	2.637
Std.	0.005	0.155	0.098

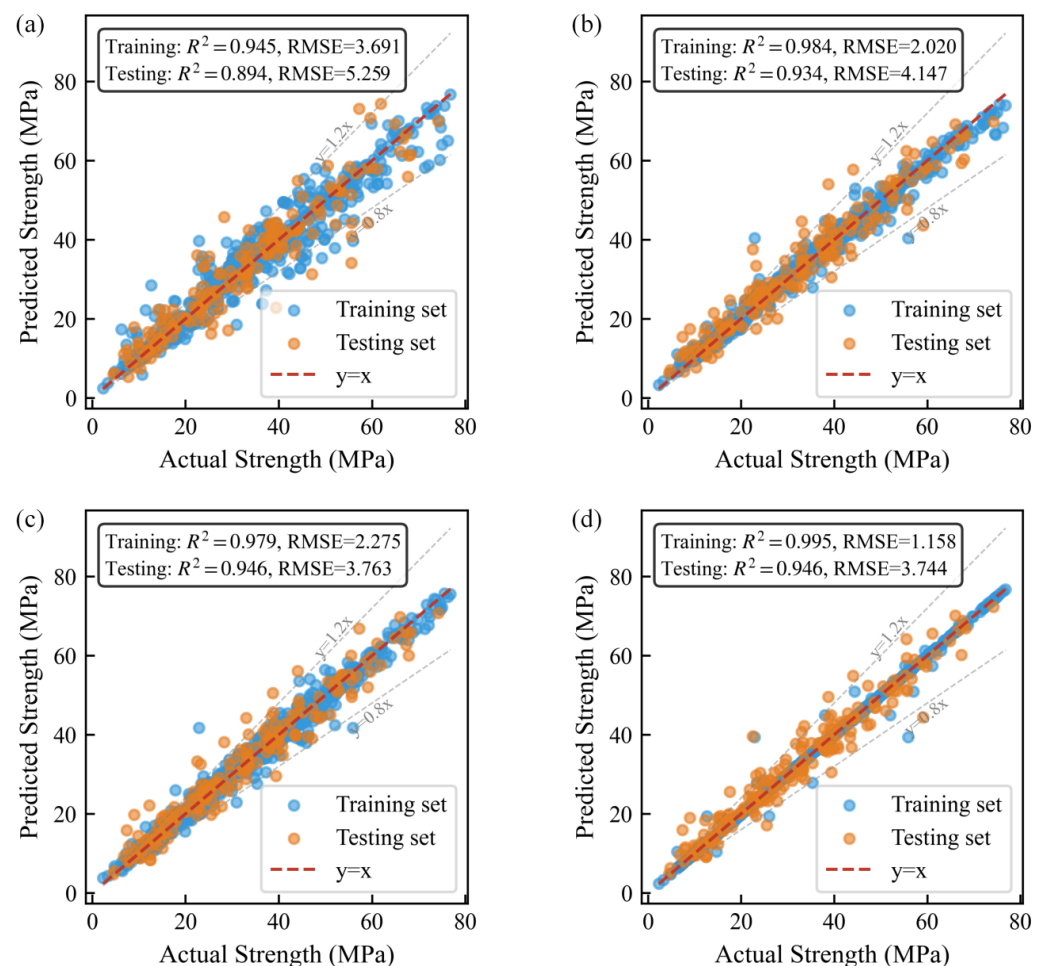


Figure 8. Measured versus predicted strength: (a) SVR, (b) RF, (c) GBDT, and (d) LightGBM.

Therefore, LightGBM was selected as the final surrogate model for SHAP interpretation and multi-objective optimization. The model provides a useful balance between prediction accuracy, computational efficiency, and compatibility with tree-based explanation tools. Nevertheless, the optimized mixtures reported in the following sections should be interpreted as data-driven candidates within the model's empirical domain rather than construction-ready mix designs.

3.2. SHAP-Based Interpretation of Mixture Effects

3.2.1. Global Feature Importance and Directional Effects

After selecting LightGBM as the final surrogate model, SHAP analysis was used to examine the statistical contribution of each input variable to the predicted compressive strength. This step was introduced to improve model transparency and to assess whether the learned feature-response patterns are broadly consistent with established concrete proportioning knowledge. It should be emphasized that SHAP analysis explains the behavior of the trained model. It provides statistical attribution, but it does not directly prove hydration mechanisms or replace microstructural and experimental validation [42].

Figure 9 reports the global SHAP distribution. Its axes and point colors are interpreted below rather than embedded in a long caption.

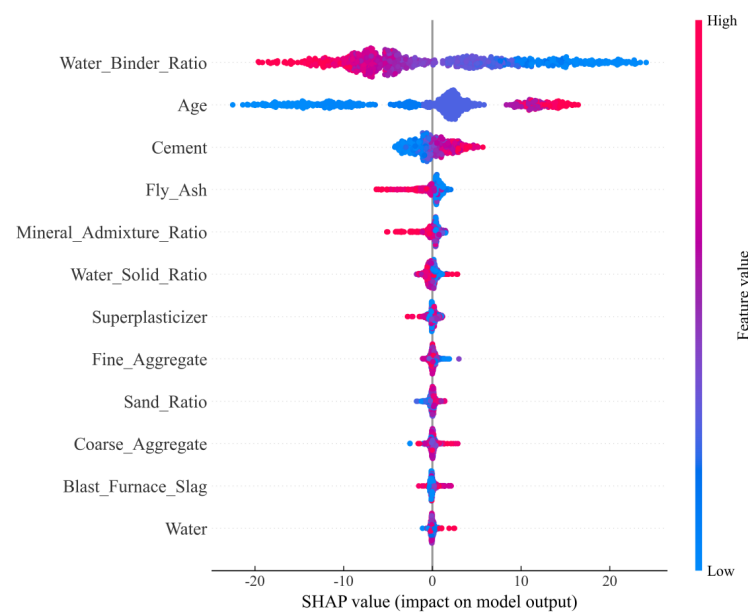


Figure 9. Global SHAP importance and directional effects.

The y-axis ranks variables according to their mean absolute SHAP values, while the x-axis indicates the marginal contribution of each feature to the predicted compressive strength. The color scale represents the relative magnitude of feature values.

The engineered water-to-binder ratio (W/B) exhibits the strongest influence on the prediction results. High W/B values are primarily located on the negative SHAP side, whereas low W/B values tend to contribute positively to predicted strength. This trend is consistent with the well-established understanding that excessive mixing water increases capillary porosity and weakens the hardened cementitious matrix [51]. Curing age also shows a clear positive contribution, reflecting the continued development of hydration products and microstructural densification over time.

The dominance of W/B over cement dosage is plausible for two reasons. First, W/B combines water and total binder in one variable and therefore captures a proportioning relationship that cement mass alone cannot express. A high cement dosage can coexist with either a high or a low W/B and thus with different strength outcomes. Second, SHAP allocates credit within a correlated feature set: W/B is constructed from cement, slag, fly ash, and water, so its importance partly absorbs information shared with those raw variables. The ranking is therefore model- and dataset-specific, not a causal hierarchy of material mechanisms.

The effects of supplementary cementitious materials, including fly ash and blast furnace slag, are more dispersed. Unlike W/B ratio, their SHAP distributions do not show

a simple monotonic pattern. High or low values of fly ash and slag can appear on both positive and negative sides of the SHAP axis, suggesting that their influence depends on the surrounding mixture context, such as water availability, cement content, curing age, and superplasticizer dosage. This observation is consistent with the known complexity of SCM-based binders, where slag and fly ash may improve later-age strength and packing density but may also reduce early-age reactivity when replacement levels are excessive or curing conditions are insufficient [13].

The global SHAP plot therefore indicates that the surrogate model relies on physically meaningful proportional features, especially W/B ratio, while also capturing more conditional effects associated with SCMs. However, because global ranking cannot fully describe nonlinear thresholds or interaction patterns, local dependence analysis was further conducted for W/B ratio and fly ash dosage.

3.2.2. Water-to-Binder Ratio Dependence

Figure 10 presents the SHAP dependence plot for the W/B ratio. The horizontal axis represents the raw W/B value, and the vertical axis represents its SHAP contribution to predicted compressive strength. The plot shows a clear nonlinear negative trend. When the W/B ratio is relatively low, the corresponding SHAP values are generally positive, indicating that lower water demand relative to binder content improves predicted strength. As W/B increases, the SHAP contribution decreases and eventually becomes negative.

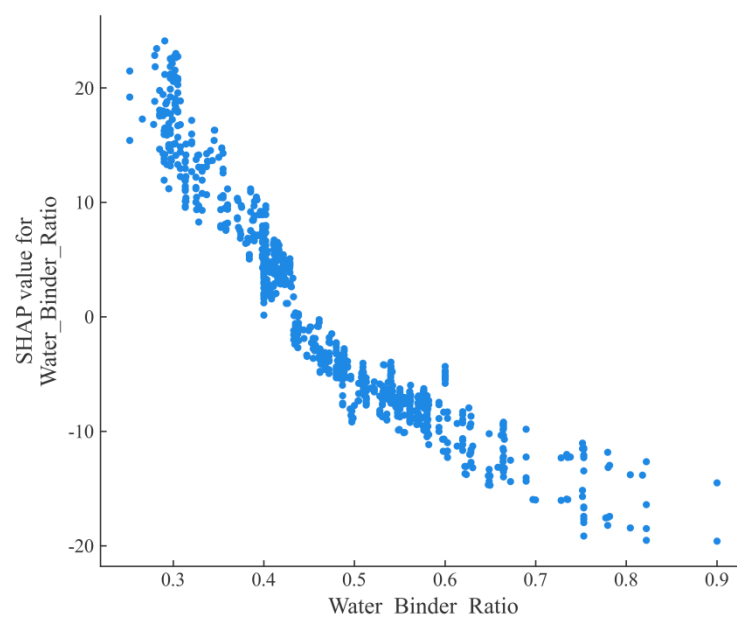


Figure 10. SHAP dependence of predicted strength on W/B.

A transition region appears around $W/B \approx 0.40$. Below this range, the model tends to associate the mixture with higher strength potential. Above this range, the negative SHAP contribution becomes more pronounced, indicating that increased water content begins to dominate the strength prediction. This trend is consistent with Abrams' Law and with the conventional understanding that higher W/B ratio increases pore connectivity and reduces matrix strength [11].

The vertical dispersion of the points at a given W/B value is relatively limited compared with the dispersion observed for SCM-related variables. This suggests that W/B ratio acts as a primary and relatively stable predictor within the current dataset. From a modeling perspective, this result supports the use of W/B ratio as a domain-informed feature. From a material-design perspective, it indicates that any cement-reduction strat-

egy must still respect the fundamental water-demand constraint. Replacing cement with SCMs cannot compensate for an excessively high W/B ratio if the target strength is to be maintained.

3.2.3. Fly Ash–Age Interaction

Figure 11 shows the SHAP dependence plot for fly ash dosage, with the color scale representing curing age. Compared with the W/B dependence plot, the fly ash plot exhibits greater vertical dispersion, especially at higher fly ash contents. This indicates that fly ash does not exert a uniform effect on strength prediction. Its contribution depends strongly on the curing-age context and surrounding mixture proportions.

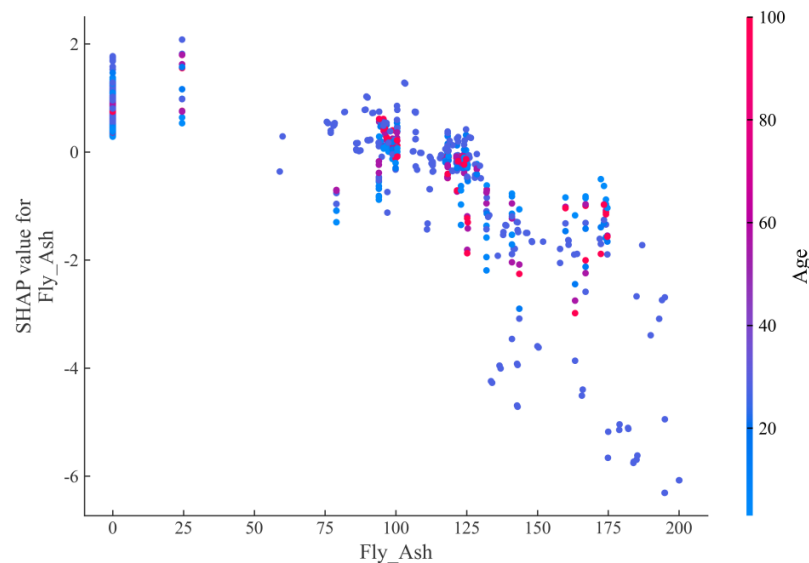


Figure 11. Fly ash SHAP dependence colored by curing age.

At low fly ash dosages, SHAP values remain close to zero, implying that small fly ash additions have limited independent influence on the model prediction. As fly ash content increases, a more evident age-dependent pattern appears. Early-age samples tend to show negative SHAP values, suggesting that high fly ash replacement may reduce early strength development. In contrast, later-age samples are more likely to move toward neutral or positive SHAP regions, indicating that the model recognizes the delayed contribution of fly ash under sufficient curing time.

This pattern is consistent with the delayed pozzolanic reaction of fly ash, in which early dilution of cement may reduce initial strength, while later reaction with calcium hydroxide may contribute to matrix densification and strength recovery [13]. Nevertheless, this interpretation should be treated as a model-based statistical explanation. The current dataset does not include direct microstructural observations, hydration heat measurements, pore-structure data, or durability testing. Therefore, the SHAP result should be understood as evidence that the surrogate model learned a plausible age-dependent trend, rather than as independent proof of reaction kinetics.

For subsequent screening, this finding indicates that the target age must be explicit. A high-fly-ash candidate associated with later-age strength may be unsuitable for a rapid-repair scenario requiring early strength. This is a hypothesis for testing, not a direct mixture-design rule.

3.3. Pareto Analysis for Cement-Reduction Screening

3.3.1. Sampled Search Space and Efficiency Boundary

The LightGBM surrogate was coupled to constraint-filtered sampling to identify candidates that trade predicted strength against reported cement dosage. Cement mass is a surrogate screening objective, not a calculated clinker or carbon objective. Life-cycle comparison would require the missing cement, SCM, transport, service-life, and durability inventories [5–7]. Figure 12 shows the resulting 28-day strength–cement landscape.

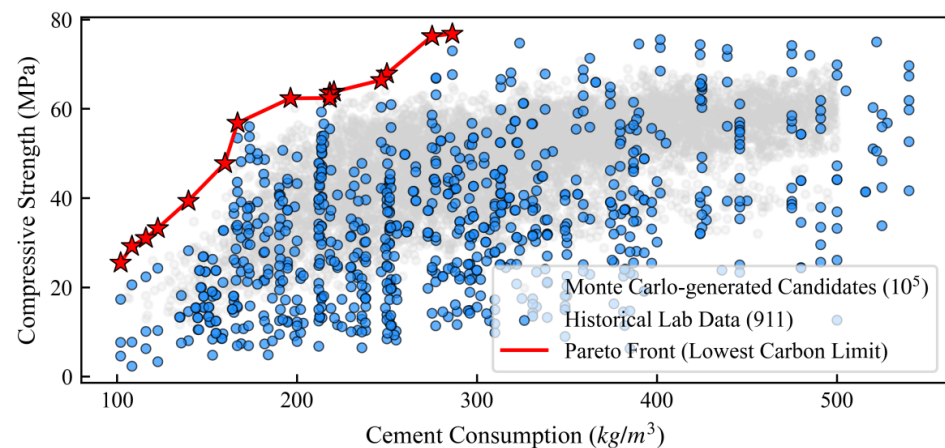


Figure 12. Sampled 28-day strength–cement space and extracted Pareto boundary.

The horizontal axis represents cement content, and the vertical axis represents predicted 28-day compressive strength. The hybrid search space combines the retained historical mixtures with Monte Carlo-generated candidates filtered by the declared constraints. The historical mixtures provide empirical anchors within the original data domain, while the virtual candidates allow the model to explore additional feasible combinations.

A clear Pareto boundary emerges along the upper-left edge of the data cloud. The algorithm identified 15 non-dominated solutions along this boundary, consisting of both historical anchor points and virtual formulations. These Pareto solutions represent mixtures for which cement content cannot be further reduced without either reducing predicted strength or moving farther away from the target performance requirement. In the medium-to-high-strength range, especially around 40–60 MPa, several Pareto solutions are located to the left of the historical mixture cluster. This suggests that the historical dataset contains mixtures with substantial cement redundancy relative to their measured or predicted strength.

The front identifies cement-efficient regions within the sampled benchmark domain, but it is not a definitive construction recommendation. It is shaped by the surrogate, data distribution, sampling strategy, and imposed constraints. Every point remains a candidate for fresh- and hardened-property testing and durability assessment.

3.3.2. Physical-Consistency Check Using the W/B–Strength Relationship

Figure 13 projects both historical and generated candidate mixtures onto the W/B ratio–compressive strength plane. This visualization was used as a physical-consistency check for the optimization results. A general inverse relationship can be observed: as W/B ratio increases, predicted compressive strength decreases. The Monte Carlo-generated candidates follow the same broad trend as the historical mixtures, suggesting that the constrained search process did not create obviously unrealistic candidate clusters.

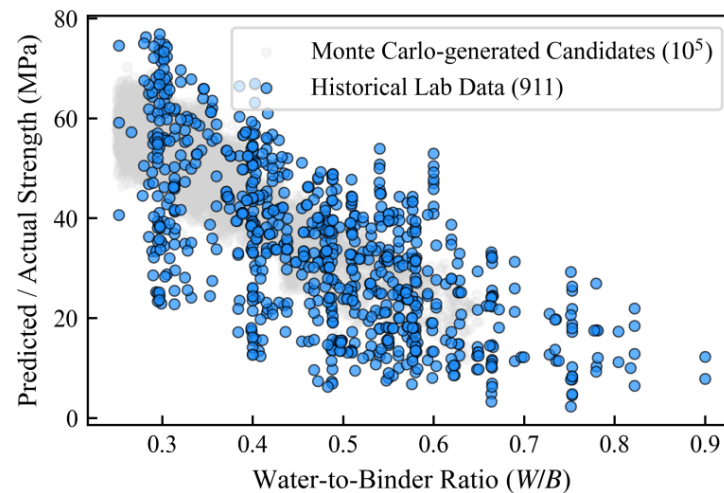


Figure 13. W/B–strength consistency check for sampled candidates.

This result is important because optimization algorithms can sometimes exploit weaknesses in a surrogate model by identifying numerically favorable but physically questionable mixtures. The agreement between generated candidates and the expected W/B–strength trend indicates that the imposed physical constraints and engineered features help maintain the generated mixtures within a plausible concrete-design domain. Nevertheless, this check only evaluates one major proportioning relationship. Other properties, including slump, setting time, shrinkage, chloride resistance, freeze–thaw resistance, and thermal cracking sensitivity, were not included in the current optimization objective and must be added before broader engineering use.

3.3.3. Evolution of Binder Composition Along the Pareto Front

To further understand the material allocation strategy identified by the optimization process, five representative Pareto-optimal mixtures were selected across the strength spectrum. These mixtures correspond to predicted strengths of 25.5, 33.2, 62.3, 63.8, and 76.8 MPa. Table 3 lists their mixture proportions, while Figure 14 visualizes the evolution of binder composition under the $MAR \leq 0.60$ constraint.

In the low-to-medium-strength region, represented by the 25.5 and 33.2 MPa mixtures, the optimization reaches the MAR upper limit of 0.600. Both mixtures adopt a cement–slag binary binder system, with cement accounting for only 40% of the total binder. This indicates that, within the current dataset and constraint system, slag replacement is an efficient strategy for reducing cement content when the target strength is moderate. The W/B ratios of these two mixtures remain relatively high, which is compatible with their lower target strengths.

Table 3. Five representative 28-day Pareto points from the sampled search.

Predicted Strength (MPa)	Cement (kg/m ³)	Slag (kg/m ³)	Fly Ash (kg/m ³)	Total Binder (kg/m ³)	MAR	W/B	Cement (%)	Binder System
25.5	102.0	153.0	0.0	255.0	0.600	0.753	40.00	Cement–Slag Binary
33.2	122.6	183.9	0.0	306.5	0.600	0.664	40.00	Cement–Slag Binary
62.3	196.3	204.3	70.4	471.0	0.583	0.302	41.70	Ternary (C–S–F)
63.8	220.5	242.3	51.4	514.2	0.571	0.296	42.90	Ternary (C–S–F)
76.8	286.3	200.9	0.0	487.2	0.412	0.297	58.80	Cement–Slag Binary

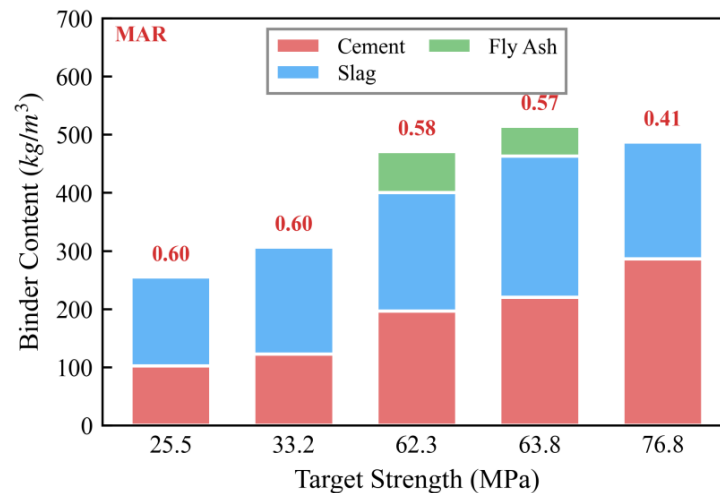


Figure 14. Binder composition of five sampled Pareto points.

When the strength target increases to the 62–64 MPa range, the binder system transitions toward a ternary cement–slag–fly ash composition. The 62.3 MPa mixture contains 196.3 kg/m³ cement, 204.3 kg/m³ slag, and 70.4 kg/m³ fly ash, while the 63.8 MPa mixture contains 220.5 kg/m³ cement, 242.3 kg/m³ slag, and 51.4 kg/m³ fly ash. Both mixtures maintain high SCM replacement levels while sharply reducing W/B ratio to approximately 0.30. This suggests that, in the current optimization space, high strength is achieved not by simply increasing cement content, but by combining low W/B ratio with optimized SCM allocation.

The adjacent 62.3 and 63.8 MPa points were retained deliberately to show the local slope of the sampled front, not because both are equally attractive recommendations. Under the two plotted objectives, the 63.8 MPa point remains non-dominated because no sampled candidate achieved at least 63.8 MPa with cement dosage at or below 220.5 kg/m³. Nevertheless, gaining 1.5 MPa required 24.2 kg/m³ more cement and 43.2 kg/m³ more total binder than the 62.3 MPa point. Its low incremental strength efficiency makes it a poor knee-point choice, and it would normally be rejected unless the additional predicted strength were specifically required. Pareto membership alone therefore does not determine the preferred engineering solution.

At the highest selected strength level of 76.8 MPa, fly ash is eliminated from the Pareto-optimal binder system, and the mixture returns to a cement–slag binary composition. Cement content increases to 286.3 kg/m³, while slag remains substantial at 200.9 kg/m³. The MAR decreases to 0.412, and W/B remains close to 0.30. This result may indicate that the model favors slag over fly ash in the high-strength region because slag generally contributes more actively to strength development under suitable alkalinity and curing conditions [66]. However, this high-strength solution should be interpreted cautiously. The high-strength region of the original dataset is relatively sparse, and the elimination of fly ash may partly reflect the data distribution learned by the surrogate model rather than a universal material rule. Additional high-strength ternary-binder experiments would be required to confirm whether this trend holds beyond the current dataset.

Overall, the sampled front contains different cement-allocation patterns across the strength range. Moderate-strength points tend to reach the imposed MAR cap, whereas higher-strength points combine lower W/B with more selective SCM allocation. These patterns are hypotheses for targeted mixture testing, not general binder-design rules.

3.3.4. Stability of Pareto Candidates Under Repeated Sampling

Because the Pareto front was extracted from a stochastic Monte Carlo search space, the stability of the selected cement-reduced solution may be affected by random sampling. To examine whether the 45 MPa candidate mixture was an isolated outcome of a single random trial, the optimization procedure was repeated ten times. In each run, the same retained historical dataset, feature-engineering procedure, LightGBM surrogate model, physical constraints, and target strength of 45 MPa were retained. Only the random seed used for generating virtual candidate mixtures was changed.

Table 4 summarizes the selected Pareto-efficient candidate from each repeated run. The cement contents ranged from 161.8 to 167.0 kg/m³, while the predicted compressive strengths remained within 44.7–45.2 MPa. The mean cement content was 164.4 kg/m³, with a standard deviation of 1.6 kg/m³. The mean predicted strength was 44.9 MPa, with a standard deviation of 0.2 MPa. The MAR and W/B values also showed limited variation, suggesting that the repeated searches converged toward a similar binder allocation region.

These results indicate that the 45 MPa cement-reduced candidate is not a singular numerical artifact generated by one Monte Carlo trial. Instead, the repeated searches consistently identified mixtures with similar cement contents, SCM replacement ratios, and W/B ratios within the imposed feasible domain. Nevertheless, this analysis only evaluates stochastic search stability. It does not quantify experimental variability, material-source variability, or prediction uncertainty of the surrogate model. Therefore, the selected mixture should still be regarded as a computationally screened candidate requiring laboratory verification.

Table 4. Repeated-sampling stability for the 28-day, 45 MPa candidate.

Run	Cement (kg/m ³)	Predicted Strength (MPa)	MAR	W/B
1	164.4	44.8	0.586	0.392
2	162.9	45.1	0.590	0.389
3	166.2	44.9	0.582	0.395
4	163.7	45.0	0.588	0.391
5	165.1	44.7	0.585	0.394
6	161.8	45.2	0.594	0.387
7	167.0	44.8	0.580	0.396
8	164.0	44.9	0.587	0.392
9	163.2	45.1	0.591	0.390
10	165.8	44.8	0.584	0.393
Mean	164.4	44.9	0.587	0.392
Std.	1.6	0.2	0.004	0.003

3.4. Tool-Augmented LLM Demonstration

3.4.1. Deployment of the Interactive Interface

The final step was to embed the deterministic surrogate and screening modules into an LLM-assisted interface. The purpose of this interface is to reduce the operational barrier for using data-driven mix screening. In conventional workflows, engineers must manually operate scripts, set constraints, run optimization routines, and interpret numerical outputs. The proposed interface allows the user to submit a natural-language design request, while the back-end system performs deterministic prediction and candidate ranking. Figure 15 shows the Streamlit interface.

The interface provides a code-free environment in which users can initiate a design request, specify target strength, and view the resulting mix proportions and engineering interpretation. Compared with conventional graphical interfaces that depend mainly on fixed forms and manual parameter inputs, the language-assisted interface allows more

flexible engineering queries. This design may improve accessibility for preliminary material screening in batching plants, design offices, and research laboratories.

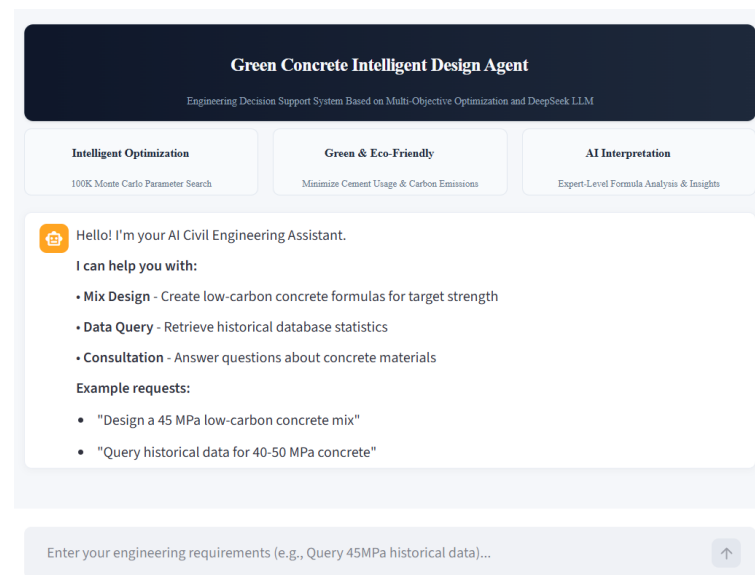


Figure 15. Streamlit interface for the tool-augmented screening workflow.

Figure 16 illustrates the operational workflow. When a user submits a query such as “Screen a cement-reduced mixture for 45 MPa at 28 days,” the LLM first extracts the target strength and identifies the user intent. The extracted information is then converted into standardized execution keys and passed to the deterministic optimization module. All strength prediction, constraint filtering, and Pareto search operations are performed by the back-end LightGBM and optimization algorithms rather than by the generative model.

This separation is essential for engineering reliability. The LLM is not allowed to invent mixture proportions or conduct numerical optimization independently. Instead, it functions as a semantic parser and report generator. This architecture reduces the risk of ungrounded numerical generation while preserving the usability advantage of natural-language interaction. Nevertheless, the current interface has not yet been evaluated through a systematic user study, hallucination-rate test, or large-scale field deployment. Therefore, the present result should be regarded as a prototype demonstration of interface feasibility rather than a completed industrial software validation.

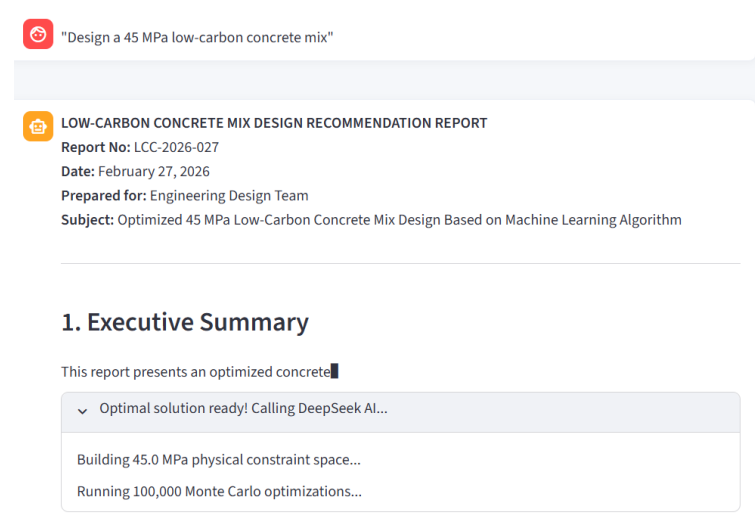


Figure 16. Natural-language request and deterministic screening execution.

3.4.2. Example Case: 45 MPa Cement-Reduced Candidate

To demonstrate the workflow, a 45 MPa target strength was used as an example design request. The optimization module searched the physically admissible candidate space and identified a Pareto-efficient mixture with a predicted strength of 44.8 MPa. The selected mixture required 164.4 kg/m³ of Portland cement. This value was compared with 142 historical mixtures in the original dataset whose strengths were within the 40–50 MPa range. These historical mixtures had an average cement content of 294.6 kg/m³, with a standard deviation of 92.5 kg/m³. Therefore, the selected candidate corresponds to a 44.2% reduction in cement content relative to this empirical benchmark.

A direct ACI PRC-211.1 benchmark was not calculated because the UCI records do not provide the required project inputs, including specified slump, nominal maximum aggregate size and grading, specific gravities, air content, exposure class, or trial-batch adjustments [57]. Claiming an ACI-designed comparator without these inputs would create a false precision. The 142-record strength-band comparison is therefore reported explicitly as an internal empirical benchmark, not as evidence that the candidate outperforms ACI proportioning. A future experimental study should design a code-informed control mixture from complete local material data and compare strength, workability, durability, and cement efficiency under the same test conditions.

This reduction should be interpreted as a cement-content reduction within the benchmark dataset, not as a complete life-cycle carbon reduction. Cement dosage is a dominant contributor to concrete's embodied carbon, but full carbon accounting would also require the emission factors of slag and fly ash, local SCM availability, transport distance, grinding or processing energy, admixture dosage, curing regime, durability-related service life, and maintenance requirements [5–7]. Therefore, the result demonstrates the potential of the framework for identifying cement-efficient candidates, while further life-cycle and experimental validation remains necessary before practical adoption. Figure 17 shows how these qualifications are communicated with the structured numerical output.

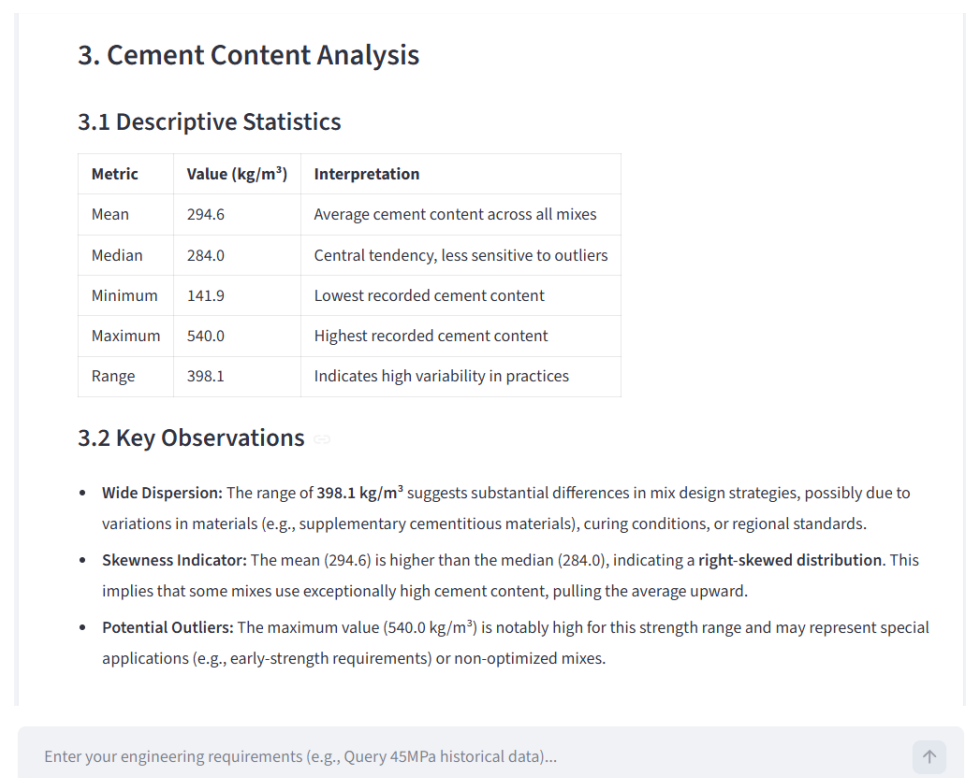


Figure 17. Structured candidate output and generated engineering report.

The report presents the screened component dosages, predicted strength, SCM replacement information, and explanatory notes in an engineering-readable format. Instead of exposing raw numerical arrays, the interface translates the optimization result into a practical material-design summary. This is useful for early-stage decision-making because engineers can quickly identify whether a candidate mixture is worth further laboratory testing.

The main advantage of the interface is not that it replaces engineering judgment, but that it connects deterministic optimization with accessible communication. The LLM provides explanation and formatting, while the numerical core remains controlled by evaluated deterministic algorithms and physical constraints. This design is aligned with the requirements of safety-sensitive material design, where interpretability, traceability, and constraint control are more important than unconstrained generative flexibility.

3.5. Scope and Limitations

The results establish an internally evaluated computational workflow, not an externally validated concrete design method. The UCI database contains only Portland cement, fly ash, slag, conventional aggregates, water, superplasticizer, age, and strength. It omits cement chemistry, regional by-products, aggregate grading, many chemical admixtures, curing environment, fresh properties, durability, and life-cycle inventories. Cross-validation partitions this same database and cannot demonstrate transfer to another laboratory, material source, or region. The outlier filter further narrows the applicable domain.

The optimization objective is limited to predicted compressive strength and reported cement dosage. It does not calculate clinker content or carbon emissions and does not test workability, setting, air content, shrinkage, cracking, transport properties, freeze–thaw resistance, or service life. SHAP attributes the fitted model rather than material causality, and repeated Monte Carlo runs measure sampling stability rather than predictive or experimental uncertainty. Finally, the LLM interface has not undergone systematic user, hallucination, adversarial, or field evaluation. These boundaries preclude construction use without experimental calibration and professional review.

3.6. Targeted Research Outlook

3.6.1. Multi-Source Regional Dataset Expansion

The first priority is a harmonized database from multiple laboratories and regions. It should record cement type and chemistry, SCM source and reactivity, aggregate grading and specific gravity, admixture chemistry, curing temperature and humidity, and comparable fresh and hardened properties. Laboratory- or region-held-out validation should replace random internal folds as the principal generalization test.

3.6.2. Durability-Coupled Robust Optimization

Candidate selection should jointly consider strength, slump retention, setting, shrinkage, chloride penetration, carbonation, sulfate and freeze–thaw resistance, and service-life indicators. Material-specific life-cycle inventories should then replace cement dosage with quantified carbon, energy, and resource objectives. Code-informed control mixtures and laboratory trial batches are needed to benchmark practical value.

3.6.3. Uncertainty-Aware Surrogate Prediction

Prediction intervals and chance constraints should be added through conformal calibration, quantile models, or Bayesian ensembles. Robust optimization should propagate both surrogate uncertainty and plausible material-source variability. GreenMix-Pareto [44] provides a relevant comparator for this next stage, while systematic sensitivity analysis should quantify the effect of every feasibility bound.

3.6.4. Industrial Deployment and Field Validation of the LLM Tool

The interface requires strict input schemas, unit normalization, out-of-domain rejection, immutable tool logs, prompt-injection tests, failure-case regression suites, and explicit human approval. Controlled user studies should measure intent-extraction accuracy, unsupported statement rate, correction burden, and reproducibility. Only after laboratory validation of returned mixtures and monitored pilot use should field deployment be considered.

4. Conclusions

The study shows that ratio-informed features, a tree-based strength surrogate, statistical explanation, and constraint-filtered sampling can be assembled into a traceable preliminary screening chain. Within the retained UCI domain, LightGBM gave the strongest held-out performance, W/B carried the largest model attribution, and repeated sampling located a stable 45 MPa region with substantially lower reported cement dosage than the internal strength-band benchmark. These findings describe model behavior and a sampled trade-off; they do not establish a new material mechanism or a validated low-carbon mixture.

The contribution to knowledge is the explicit separation of numerical authority from language interaction in concrete proportioning. The surrogate and Pareto components provide deterministic, inspectable tools, while the LLM is restricted to converting user intent into execution keys and structured results into prose. This architecture offers a reproducible way to use an LLM in a safety-sensitive materials workflow without treating generative text as an engineering calculator.

The contribution to practice is correspondingly limited but useful: the prototype can rank cement-reduced candidates for laboratory testing and communicate why a candidate was returned, which constraints governed it, and what must still be verified. It cannot replace ACI proportioning, trial batching, durability design, life-cycle assessment, or professional approval.

The evidence remains bounded by one curated benchmark, no external or experimental validation, a cement-mass surrogate instead of quantified carbon, fixed feasibility bounds, and an unevaluated LLM user interface. The next research step is therefore not immediate deployment but the four-stage program defined in Section 3.6: multi-source validation, durability- and life-cycle-coupled objectives, uncertainty-aware robust screening, and controlled laboratory and field evaluation of both the mixtures and the interface.

Author Contributions: Methodology, J.Z. and H.Y.; software, J.Z. and H.Y.; validation, G.H.; investigation, X.W.; data curation, X.W.; writing—original draft preparation, J.Z. and H.Y.; writing—review and editing, J.Z. and J.W.; visualization, G.H. and J.W.; supervision, G.H. and J.W.; project administration, G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 42377201.

Data Availability Statement: Raw data comprising eight mix components and approximately 1000 samples are available from the UCI Machine Learning Repository (DOI: <https://doi.org/10.24432/C5PK67>). The prototype source code is openly available at <https://github.com/yhd630/streamlit-ai-green-concrete> (accessed on 4 August 2026).

Acknowledgments: The authors declare that no specific funding was received for this research. During the preparation and revision of this manuscript, the authors used ChatGPT (OpenAI) to assist with English-language editing, readability improvement, and editorial refinement. The DeepSeek LLM used in the proposed prototype is a component of the research methodology and is described

separately in Section 2.5. All AI-assisted content was critically reviewed, verified, and revised by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

ACI	American Concrete Institute
GBDT	Gradient Boosting Decision Tree
IQR	Interquartile range
LightGBM	Light Gradient Boosting Machine
LLM	Large language model
MAE	Mean absolute error
MAR	Mineral admixture ratio
ML	Machine learning
OPC	Ordinary Portland cement
RF	Random Forest
RMSE	Root mean square error
SCM	Supplementary cementitious material
SHAP	Shapley Additive Explanations
SR	Sand ratio
SVR	Support Vector Regression
W/B	Water-to-binder ratio
W/S	Water-to-solid ratio
XAI	Explainable Artificial Intelligence

References

1. IPCC . Climate Change 2022: Impacts, Adaptation and Vulnerability. In *Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2022. <https://doi.org/10.1017/9781009325844>.
2. Andrew, R.M. Global CO₂ emissions from cement production. *Earth Syst. Sci. Data* **2018**, *10*, 195–217. <https://doi.org/10.5194/essd-10-195-2018>.
3. Amitani, A.; Miller, S.A. Material performance, economic feasibility, and greenhouse gas emissions of natural pozzolans in concrete: A review and meta-analysis. *Resour. Conserv. Recycl.* **2026**, *228*, 108769. <https://doi.org/10.1016/j.resconrec.2025.108769>.
4. Kashani, A.; Hameed, R.; Khondoker, M.T.H.; Li, W.; Kim, T.; Tam, V.W.Y. Carbon utilization for low-carbon concrete: Prospects for carbonated steel slag and recycled concrete. *Resour. Conserv. Recycl.* **2026**, *226*, 108698. <https://doi.org/10.1016/j.resconrec.2025.108698>.
5. Nair, S.R.; Nagarajan, P.; Thampi, S.G.; Das, S. Life cycle-based sustainability assessment of optimized 3D printable concrete mixes incorporating SCMs and recycled aggregates. *Results Eng.* **2025**, *28*, 108002. <https://doi.org/10.1016/j.rineng.2025.108002>.
6. Nehdi, M.L.; Marani, A.; Zhang, L. Is net-zero feasible: Systematic review of cement and concrete decarbonization technologies. *Renew. Sustain. Energy Rev.* **2024**, *191*, 114169. <https://doi.org/10.1016/j.rser.2023.114169>.
7. Ozcelikci, E.; Hu, M.; van der Meide, M.; Sahmaran, M. Life cycle assessment of SCM substitution in various CDW-based geopolymer concretes and sensitivity analyses on allocation methods. *Waste Manag.* **2025**, *205*, 115034. <https://doi.org/10.1016/j.wasman.2025.115034>.
8. Xu, L.; Huang, R.; Li, M.; Li, K.; Leng, F.; Ran, B.; Wang, J. From demolition to decarbonization: Synergistic hydration and sustainability of thermally activated recycled concrete powders. *J. Environ. Manag.* **2026**, *401*, 128902. <https://doi.org/10.1016/j.jenvman.2026.128902>.
9. Zhou, Z.; Li, S.; Hu, G.; Wu, J.; Yao, C.; Niu, F. On the use of recycled polyethylene terephthalate fiber in one-part geopolymer stabilized soft soil: Tensile performance and sustainability analysis. *Dev. Built Environ.* **2025**, *21*, 100641. <https://doi.org/10.1016/j.dibe.2025.100641>.
10. Genc, O.; Kurt, A. Biologically inspired optimization of construction sector eco-industrial park networks using food web metrics. *Sci. Rep.* **2026**, *16*, 23693. <https://doi.org/10.1038/s41598-026-54667-x>.
11. Sear, L.K.A.; Dews, J.; Kite, B.; Harris, F.C.; Troy, J.F. Abrams law, air and high water-to-cement ratios. *Constr. Build. Mater.* **1996**, *10*, 221–226. [https://doi.org/10.1016/0950-0618\(95\)00079-8](https://doi.org/10.1016/0950-0618(95)00079-8).

12. Antunes, R. Concrete overstrength: Assessment of field strength seeking insights for overdesign optimization. *CivilEng* **2022**, *3*, 51–65. <https://doi.org/10.3390/civileng3010004>.
13. Han, T.; Huang, J.; Sant, G.; Neithalath, N.; Goel, A.; Kumar, A. Modeling hydration kinetics of sustainable cementitious binders using an advanced nucleation and growth approach. *Constr. Build. Mater.* **2023**, *404*, 133327. <https://doi.org/10.1016/j.conbuildmat.2023.133327>.
14. Lallas, Z.N.; Gombedra, M.J.; Mendonca, F. Review of supplementary cementitious materials with implications for age-dependent concrete properties affecting precast concrete. *PCI J.* **2023**, *68*, 46–64. <https://doi.org/10.15554/pci68.6-01>.
15. Luca, M.; Michelle, H.; Amir, M.; Robert, S.; Maria, J. Application of maturity to estimate the strength development of high-early-strength concrete mixtures using isothermal calorimetry. *J. Mater. Civ. Eng.* **2024**, *36*, 04024087. <https://doi.org/10.1061/JMCEE7.MTENG-17168>.
16. Al-Naghi, A.A.A.; Ahmad, A.; Amin, M.N.; Algasse, O.; Alnawmasi, N. Sustainable optimisation of GGBS-based concrete: De-risking mix design through predictive machine learning models. *Case Stud. Constr. Mater.* **2025**, *23*, e04900. <https://doi.org/10.1016/j.cscm.2025.e04900>.
17. Yadav, A.A.; Hirekhan, S.G.; Bhandari, P.S.; Bhagat, R.M.; Ranit, A.B.; Shelare, S.; Soudagar, M.E.M.; Sharma, S.; Raja, V.K.B.; Mahapatro, A.; et al. AI-driven sustainable concrete mix design: Hybrid deep Q-learning and genetic algorithms-based multi-objective machine learning optimizations for high structural strength, low cost, and low carbon footprints. *Structures* **2025**, *82*, 110443. <https://doi.org/10.1016/j.istruc.2025.110443>.
18. Yuan, Z.; Zheng, W.; Qiao, H. Machine learning based optimization for mix design of manufactured sand concrete. *Constr. Build. Mater.* **2025**, *467*, 140256. <https://doi.org/10.1016/j.conbuildmat.2025.140256>.
19. Chen, Z.; Li, J.; Xu, X. Prediction of concrete failure behavior under high impact using machine learning. *Eng. Appl. Artif. Intell.* **2025**, *156*, 111036. <https://doi.org/10.1016/j.engappai.2025.111036>.
20. Kim, J. Challenges with hard-to-learn data in developing machine learning models for predicting the strength of multi-recycled aggregate concrete. *Appl. Soft Comput.* **2025**, *175*, 113110. <https://doi.org/10.1016/j.asoc.2025.113110>.
21. Liu, Z.; Zhou, H.; Zhao, X.; Zhao, W.; Guo, A. Physics-enhanced machine learning surrogate modelling for real-time seismic fragility assessment of reinforced concrete bridges. *Eng. Struct.* **2026**, *348*, 121772. <https://doi.org/10.1016/j.engstruct.2025.121772>.
22. Liu, J.; Wang, S.; Chen, K.; Wang, K.; Fan, Y.; Wang, Q. Synthetic data augmentation and integrated prediction framework for low-carbon recycled concrete: Leveraging CTGAN and multiple ML models stacking. *J. Clean. Prod.* **2026**, *543*, 147599. <https://doi.org/10.1016/j.jclepro.2026.147599>.
23. Zhao, W.; Ma, C.; Xiao, X.; Zhao, J.; Feng, X. Design of recycled concrete mix proportions based on ensemble algorithms and multi-objective optimization. *Constr. Build. Mater.* **2025**, *499*, 144122. <https://doi.org/10.1016/j.conbuildmat.2025.144122>.
24. Liu, J.; Kong, X.; Wang, S.; Peng, L.; Wang, Q.; Zhang, Y.; Bao, X. Multi-objective material-structure integrated optimization of recycled aggregate CFST stub columns for mechanical performance, reliability, and sustainability. *Expert Syst. Appl.* **2026**, *324*, 132597. <https://doi.org/10.1016/j.eswa.2026.132597>.
25. La, H.; Uddin, M.; Nguyen, N.H.T.; Nguyen, T. Hybrid NSGA-III and surrogate model framework for sustainable concrete mix design: Balancing strength, energy, and carbon. *Appl. Soft Comput.* **2025**, *185*, 114018. <https://doi.org/10.1016/j.asoc.2025.114018>.
26. Wang, C.; Chu, W.; Liu, J.; Zhao, Y.; Liu, F.; Zhang, Y.; Rashidi, M.; Yu, Y. Unified axial capacity prediction of FRP-confined concrete across natural and recycled aggregate systems using a Kolmogorov-Arnold Networks based cross-domain transfer-learning framework. *Structures* **2026**, *90*, 112255. <https://doi.org/10.1016/j.istruc.2026.112255>.
27. Kong, X.; Fan, Y.; Liu, J.; Xi, M.; Zhang, Y.; Yu, Y. Lightweight Kolmogorov-Arnold Network with dual-objective optimization for axial capacity prediction of square coal gangue concrete-filled steel tube stub columns based on finite element simulation. *Eng. Appl. Artif. Intell.* **2026**, *176*, 114843. <https://doi.org/10.1016/j.engappai.2026.114843>.
28. Yu, X.; Li, J.; Liu, J.; Li, Y.; Chen, A.; Zhang, Y.; Chang, H. Cross-scenario impact damage prediction in concrete using peridynamics-trained Kolmogorov-Arnold networks with transfer learning. *Int. J. Impact Eng.* **2026**, *215*, 105764. <https://doi.org/10.1016/j.ijimpeng.2026.105764>.
29. Liang, D.; Xue, X.; Liu, J.; Cui, Y.; Sun, Z.; Zhang, Y.; Zhang, T.; Xu, L. Matching metaheuristics to machine learning for concrete drying shrinkage prediction toward cost-effective mix design. *Dev. Built Environ.* **2026**, *27*, 100964. <https://doi.org/10.1016/j.dibe.2026.100964>.
30. Kaplanvural, İ. Volumetric water content estimation of concrete by particle swarm optimization of GPR data. *Constr. Build. Mater.* **2023**, *375*, 130995. <https://doi.org/10.1016/j.conbuildmat.2023.130995>.
31. Dash, P.K.; Parhi, S.K.; Patro, S.K.; Panigrahi, R. Efficient machine learning algorithm with enhanced cat swarm optimization for prediction of compressive strength of GGBS-based geopolymer concrete at elevated temperature. *Constr. Build. Mater.* **2023**, *400*, 132814. <https://doi.org/10.1016/j.conbuildmat.2023.132814>.
32. Tian, H.; Zhang, Y.; Feng, Q.; Qiao, H. Performance analysis and strength prediction of polymer-modified manufactured sand concrete using artificial bee colony optimization algorithm. *Mater. Today Commun.* **2025**, *49*, 113854. <https://doi.org/10.1016/j.mtcomm.2025.113854>.

33. Truong, D.N.; Chou, J.S. Multi-objective scientific approach to problem solving-inspired optimization integrated with the finite element method for automated structural design. *Autom. Constr.* **2026**, *183*, 106770. <https://doi.org/10.1016/j.autcon.2026.106770>.
34. Fan, Y.; Liu, J.; Tao, Y.; Sun, Y.; Li, Z.; Li, J.; Wang, K.; Ma, J.; Fan, D.; Xu, L. A transfer-learning framework to alleviate data scarcity in cross-slope wind pressure modeling. *Results Eng.* **2026**, *29*, 109376. <https://doi.org/10.1016/j.rineng.2026.109376>.
35. Zhang, Y.; Peng, J.; Wang, Z.; Xi, M.; Liu, J.; Xu, L. Machine learning-assisted sustainable mix design of waste glass powder concrete with strength–cost–CO₂ emissions trade-offs. *Buildings* **2025**, *15*, 2640. <https://doi.org/10.3390/buildings15152640>.
36. Siddiq, M.U.; Anwar, M.K.; Almansour, F.H.; Qurashi, M.A.; Adeel, M. AI-driven optimization of fly ash-based geopolymer concrete for sustainable high strength and CO₂ reduction: An application of hybrid Taguchi–Grey–ANN approach. *Buildings* **2025**, *15*, 2081. <https://doi.org/10.3390/buildings15122081>.
37. Nwetlawung, Z.E.; Lin, Y.-H. Development of an optimization algorithm for designing low-carbon concrete materials standardization with blockchain technology and ensemble machine learning methods. *Buildings* **2025**, *15*, 2809. <https://doi.org/10.3390/buildings15162809>.
38. Alqurashi, M. Data-driven insights into concrete flow and strength: Advancing smart material design using machine learning strategies. *Buildings* **2025**, *15*, 2244. <https://doi.org/10.3390/buildings15132244>.
39. Kamarthi, A.; Kaliyamoorthy, B. An interpretable automated optimized machine learning for predicting concrete compressive strength. *Expert Syst. Appl.* **2026**, *298*, 129656. <https://doi.org/10.1016/j.eswa.2025.129656>.
40. Cravero, F.; Ponzoni, I.; Diaz, M.F.; Vazquez, G.E. From prediction to understanding: A review of XAI applications and innovations in materials science. *Comput. Mater. Sci.* **2026**, *267*, 114493. <https://doi.org/10.1016/j.commatsci.2026.114493>.
41. Yang, H.; Hu, G.; Liu, L.; Li, Y.; Deng, Y.; Wu, J. A physics-informed and SHAP-enhanced modeling framework for predicting strength of cement-stabilized soil. *Case Stud. Constr. Mater.* **2025**, *23*, e05280. <https://doi.org/10.1016/j.cscm.2025.e05280>.
42. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2017.
43. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
44. Abdennaji, T.S.; Ksibi, A.; Flah, A.; Batra, V.; Tipu, R.K. GreenMix-Pareto: Uncertainty-aware, physics-guided multi-objective optimization of low-carbon concrete mix designs. *Ain Shams Eng. J.* **2026**, *17*, 104009. <https://doi.org/10.1016/j.asej.2026.104009>.
45. Wang, H.; Zhang, Y.; Liu, J.; Zhao, Y.; He, R.; Xu, S.; Wang, W.; Lyu, X. Data-driven reliability-based optimization framework for concrete-encased steel composite columns toward cost-effective design: Methodology and interactive tool. *Adv. Struct. Eng.* **2026**, 13694332261447524. <https://doi.org/10.1177/13694332261447524>.
46. Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K.; Cao, Y. ReAct: Synergizing reasoning and acting in language models. In *Proceedings of the 11th International Conference on Learning Representations*, Kigali, Rwanda, 1–5 May 2023.
47. Schick, T.; Dwivedi-Yu, J.; Dessì, R.; Raileanu, R.; Lomeli, M.; Hambro, E.; Zettlemoyer, L.; Cancedda, N.; Scialom, T. Toolformer: Language models can teach themselves to use tools. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 68539–68551.
48. Yeh, I.C. *Concrete Compressive Strength*; UCI Machine Learning Repository: Irvine, CA, USA, 1998. <https://doi.org/10.24432/C5PK67>.
49. Tukey, J.W. *Exploratory Data Analysis*; Addison-Wesley: Reading, MA, USA, 1977.
50. ASTM C39/C39M-21; Standard Test Method for Compressive Strength of Cylindrical Concrete Specimens. ASTM International: West Conshohocken, PA, USA, 2021.
51. Mehta, P.K.; Monteiro, P.J.M. *Concrete: Microstructure, Properties, and Materials*; McGraw-Hill Education: New York, NY, USA, 2014.
52. Zhang, C.; Zhu, Z.; Liu, F.; Yang, Y.; Wan, Y.; Huo, W.; Yang, L. Efficient machine learning method for evaluating compressive strength of cement stabilized soft soil. *Constr. Build. Mater.* **2023**, *392*, 131887. <https://doi.org/10.1016/j.conbuildmat.2023.131887>.
53. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. <https://doi.org/10.1007/BF00994018>.
54. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>.
55. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
56. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2017.
57. ACI Committee 211. *Selecting Proportions for Normal-Density and High-Density Concrete—Guide (ACI PRC-211.1-22)*; American Concrete Institute: Farmington Hills, MI, USA, 2022.
58. ASTM C138/C138M-23; Standard Test Method for Density (Unit Weight), Yield, and Air Content (Gravimetric) of Concrete. ASTM International: West Conshohocken, PA, USA, 2023.
59. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
60. DeepSeek-AI; Bi, X.; Chen, D.; Chen, G.; Chen, S.; Dai, D.; Deng, C.; Ding, H.; Dong, K.; Du, Q.; et al. DeepSeek LLM: Scaling open-source language models with longtermism. *arXiv* **2024**, arXiv:2401.02954.
61. Streamlit Inc. *Streamlit: The Fastest Way to Build and Share Data Apps*; Streamlit Inc.: San Francisco, CA, USA, 2020.

62. Chen, Q.; Hu, G.; Wu, J. Comparative study on the prediction of the unconfined compressive strength of the one-part geopolymer stabilized soil by using different hybrid machine learning models. *Case Stud. Constr. Mater.* **2024**, *21*, e03439. <https://doi.org/10.1016/j.cscm.2024.e03439>.
63. Vovk, V.; Gammerman, A.; Shafer, G. *Algorithmic Learning in a Random World*; Springer: New York, NY, USA, 2005.
64. Meinshausen, N. Quantile regression forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999.
65. Aitcin, P.C. *High Performance Concrete*; E & FN Spon: London, UK, 1998.
66. Liu, L.; Zhou, A.; Deng, Y.; Cui, Y.; Yu, Z.; Yu, C. Strength performance of cement/slag-based stabilized soft clays. *Constr. Build. Mater.* **2019**, *211*, 909–918. <https://doi.org/10.1016/j.conbuildmat.2019.03.256>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.