

# Monitoring of ultra-high performance concrete manufacturing for reproducible quality and waste reduction

Received: 7 September 2025

Accepted: 15 December 2025

Published online: 26 December 2025

Cite this article as: Rezazadeh F., Abrishambaf A., Zimmermann G. *et al.* Monitoring of ultra-high performance concrete manufacturing for reproducible quality and waste reduction. *Sci Rep* (2025). <https://doi.org/10.1038/s41598-025-32975-y>

Farzad Rezazadeh, Amin Abrishambaf, Gregor Zimmermann & Andreas Kroll

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

# Monitoring of ultra-high performance concrete manufacturing for reproducible quality and waste reduction

Farzad Rezazadeh<sup>1,\*</sup>, Amin Abrishambaf<sup>2</sup>, Gregor Zimmermann<sup>3</sup>, and Andreas Kroll<sup>1</sup>

<sup>1</sup>Department of Measurement and Control Engineering, University of Kassel, Kassel, 34125, Germany

<sup>2</sup>QuantumFusion GmbH, Waldeck, 34513, Germany

<sup>3</sup>MAITERIA UG, Kassel, 34131, Germany

\*farzad.rezazadeh@mrt.uni-kassel.de

## ABSTRACT

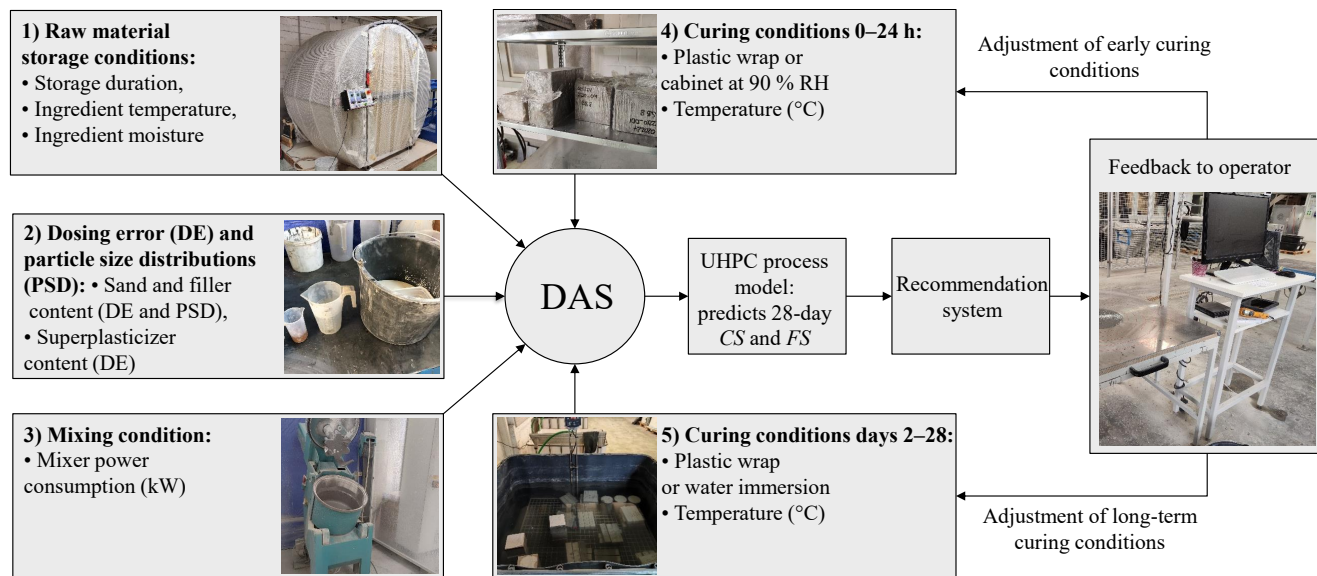
Ultra-high performance concrete (UHPC) combines exceptional strength and durability, yet its industrial production is hampered by batch-to-batch variability that generates costly off-specification waste. Leveraging a 150-batch design-of-experiments dataset based on systematic variations of a single reference UHPC mix, this study takes a holistic view of the UHPC manufacturing chain and quantifies how fluctuations in raw material quality, storage conditions, dosing errors, mixer energy demand, and curing regimes affect the 28-day compressive strength. Ten diverse machine learning algorithms are benchmarked; the best-performing model explains  $\geq 75\%$  of the strength variance with a prediction error  $\leq 10\%$  under leave-one-out cross-validation. SHapley Additive exPlanations reveal that long-term curing temperature and humidity dominate strength development, followed by ingredient moisture and silica fume impurity. These insights are operationalized in an at-line, operator-in-the-loop recommendation system that explores the curing envelope and proposes end-of-mix, batch-specific adjustments before curing starts. In five validation cases, curing adjustments rescued 5/5 underperforming batches, eliminating  $75\text{ m}^3$  of off-specification UHPC and – considering cement only with  $600\text{ kg/m}^3$  and  $15\text{ m}^3$  per batch of white Portland cement – avoided  $\approx 41\text{ t CO}_2\text{e}$  (cement-only;  $0.913\text{ kg CO}_2\text{e/kg}$ , A1–A3). The framework therefore not only elucidates the main sources of UHPC quality inconsistency but also provides a practical, data-driven tool to rescue off-specification products, minimize waste, and cut associated  $\text{CO}_2$  emissions.

## Introduction

Ultra-high performance concrete (UHPC) is an advanced cementitious composite with outstanding mechanical performance and durability. Owing to a densely packed microstructure obtained by optimizing ultrafine particles and steel fibers<sup>1,2</sup>, UHPC routinely achieves compressive strengths above 120 MPa and tensile strengths exceeding 5 MPa – several times those of conventional concrete<sup>3–5</sup>. The resulting low porosity affords exceptional resistance to water ingress, chemical attack, and freeze–thaw cycling<sup>6–9</sup>; autogenous self-healing further prolongs service life<sup>10</sup>.

These properties have propelled UHPC into a wide spectrum of demanding applications. In bridge construction, it enables slender, lighter decks with extended design lives<sup>11</sup>. Its flowability and early strength permit intricate architectural façades, and its durability in aggressive environments suits marine and offshore structures<sup>12</sup>. In repair and retrofit projects, UHPC lowers maintenance and life-cycle costs<sup>13</sup>. For instance, Russell and Graybeal<sup>4</sup> report that, for decks, full-depth UHPC waffle panels reduced slab dead load by  $\approx 56\%$  compared with conventional concrete. In a short-span case study, a composite timber–ultra-high performance fiber-reinforced concrete (UHPFRC) superstructure weighed 20,890 kg versus 63,422 kg for a conventional reinforced concrete (RC) alternative (an  $\approx 67\%$  reduction); the authors also note that UHPFRC decks typically reduce deck weight by a factor of  $\sim 3$  relative to RC<sup>14</sup>. At the component scale, designs of UHPC bridge piers achieved  $\approx 3.5\text{--}36.6\%$  reductions in concrete volume<sup>15,16</sup>. Regarding costs, a Federal Highway Administration (FHWA) life-cycle cost (LCC) analysis for a signature suspension bridge found that a UHPC overlay rehabilitation option (Installation Strategy 2;  $\approx 3.75\text{ in.}$  partial-depth replacement) reaches break-even with new precast deck replacement at an actual overlay service life of  $\approx 24\text{ years}$ <sup>17</sup>. In the same short-span study, life-cycle costs for the timber–UHPFRC solution were  $\approx 30\%$  lower (the RC alternative's LCC was  $\approx 43\%$  higher)<sup>14</sup>. At the network scale, systematic UHPFRC interventions are projected to save up to 18.5 billion Swiss francs (CHF) over an 80-year horizon relative to demolition–reconstruction strategies<sup>18</sup>. Accordingly, the material is reshaping modern construction practice and supporting the transition to more sustainable infrastructure.

Although UHPC offers superior performance, its manufacture is both resource- and carbon-intensive. The material contains a high cement content and demands tight process control; consequently, it is costly to produce and associated with elevated



**Figure 1.** At-line end-of-mix recommendation system for UHPC production. The data acquisition system ingests process data into a predictive process model and a recommendation engine, which jointly monitor the current batch and propose pre-curing adjustments to keep the predicted 28-day strength within specification. Feedback is given to the operator (operator-in-the-loop). (*CS* = compressive strength; *FS* = flexural strength)

CO<sub>2</sub> emissions. Moreover, steel fibers and the high dosage of superplasticizer commonly required in UHPC also contribute to higher cost and embodied carbon<sup>19</sup>. A persistent industrial challenge is achieving batch-to-batch reproducibility: even when the same reference mix is followed, slight deviations in process conditions can yield off-specification (off-spec) batches that must be scrapped, thereby compounding costs and carbon footprint<sup>20</sup>. These realities highlight the need for a decision-support system that allows operators to adjust processing parameters, predict performance, and uphold quality, ultimately reducing waste, lowering costs, and ensuring consistent high-performance output.

Accurately forecasting 28-day mechanical properties further complicates UHPC production. Conventional empirical relationships have advanced concrete science, yet they cannot capture the multifactorial chemical and physical interactions that govern high-performance concretes<sup>21</sup>. Accordingly, more sophisticated, data-driven modeling approaches are required to navigate the complex variable space and deliver reliable property predictions<sup>22,23</sup>.

Recent studies highlight the promise of machine learning (ML) techniques for predicting the properties of cementitious materials<sup>23–27</sup>. Ling *et al.*<sup>28</sup> reported that support vector regression (SVR) surpassed artificial neural networks (ANN) and decision trees (DT) in capturing environmental effects on concrete strength, whereas Hoang and Pham<sup>29</sup> identified Gaussian process regression (GPR) as superior to both ANN and SVR. Nguyen *et al.*<sup>30</sup> further showed that extreme gradient boosting (XGB)<sup>31</sup> accurately predicts UHPC compressive strength, expediting mixture development at lower cost. However, most prior ML-assisted UHPC studies stop at prediction or mix-design exploration; they do not deliver at-line recommendations that can adjust processing parameters for the current batch before curing begins.

Despite these advances, ML deployment in UHPC modeling faces notable obstacles<sup>23,32</sup>. Chief among them is data sparsity: UHPC experiments are expensive and time-consuming, limiting dataset size and risking model overfitting. Existing datasets – such as those compiled by Yeh<sup>33,34</sup> – aggregate high-performance concrete (HPC) data from diverse sources and have been widely used in studies. However, data collected from disparate sources risk leading to redundant experiments and introduce inconsistencies in material quality and production conditions, which may undermine the reliability of predictive models, even when large datasets are employed. A practical remedy is to generate high-quality data systematically. Design of Experiments (DoE) frameworks maximize information gain per trial and thus support reliable model training while minimizing laboratory effort<sup>35</sup>.

This study tackles the persistent challenge of delivering consistent UHPC quality at industrial scale. A comprehensive dataset comprising 150 mixtures, created under controlled conditions to establish a high-quality data foundation<sup>36</sup>, is employed. To emulate real-world variability, the experimental design systematically perturbed six uncertainty sources – raw material storage, particle size distribution, dosing errors, mixer power demand, and early- and late-age curing regimes – while holding the reference mix proportions constant<sup>36</sup>. Thus, the machine learning models developed here are specific to this single UHPC formulation and are not universal across recipes.

To maximize the available data for modeling while ensuring reliable evaluation on unseen data, a leave-one-out cross-

validation (LOOCV)<sup>37</sup> approach, with varied model initializations and hyperparameter-optimization runs in each fold, is applied across ten ML algorithms. The best-performing model (BPM) is embedded in a recommendation engine (Fig. 1) that monitors incoming batches and proposes curing adjustments to keep strength on target.

The results advance process understanding on two fronts. First, outdoor storage exposes aggregates and binders to uncontrolled temperature and moisture swings, which translate directly into strength variability. Second, curing temperature and humidity dominate mechanical performance, especially under the seasonal and geographic variations typical of field production. Collectively, these insights inform data-driven control strategies that minimize waste, stabilize batch quality, and reduce the embodied CO<sub>2</sub> in UHPC production.

This study (i) assembles a targeted 150-batch dataset that perturbs some source of uncertainty while holding the base UHPC recipe fixed; (ii) benchmarks ten ML models with nested cross-validation; (iii) uses SHapley Additive exPlanations (SHAP) to reveal the most influential factors; (iv) demonstrates practical impact by rescuing five underperforming batches to 118–122 MPa via model-recommended curing adjustments; and, crucially, (v) implements at-line, operator-in-the-loop recommendation engine that exhaustively searches the curing envelope (3,844 set-points) and returns a plan in ~30 s.

## Methods

### Materials and Experiments

The dataset used in this study is created using a DoE framework that maximized input-space coverage and efficiently revealed the factors influencing UHPC performance<sup>36</sup>. First, 50 experiments were conducted using a Taguchi L50 orthogonal array for factor screening<sup>35</sup>. Next, the design space from phase one is augmented with Latin hypercube sampling (LHS) to place another 100 experiments between existing data points and improve uniform coverage of the input space<sup>38,39</sup>. This methodical approach produced 150 well-balanced UHPC batches that capture realistic industrial variability for subsequent analysis<sup>20,36,40</sup>.

All experiments employed a fixed reference UHPC mixture formulated for façade panel applications, composed of white Portland cement (600 kg/m<sup>3</sup>), silica fume (100 kg/m<sup>3</sup>), quartz fillers (450 kg/m<sup>3</sup> combined), two silica-based sands (1,100 kg/m<sup>3</sup> combined), water (195 kg/m<sup>3</sup>), and a polycarboxylate-based superplasticizer (21 kg/m<sup>3</sup>) to yield 15 m<sup>3</sup> of UHPC product. The fixed reference recipe is designed to yield a UHPC product with a final compressive strength of 120 MPa. The cement used is 42.5R white Portland type I. Silica fume with over 95 % SiO<sub>2</sub> content enhances the mix's density and reduces porosity, contributing significantly to its mechanical properties and durability. Quartz powder, with a silicon oxide content of 99.6 % and a  $d_{50}$  of 27 µm, serves as filler type I. Filler type II is an add-on developed to further improve workability and enhance strength. Two grades of silica-based sands are included: one with a  $d_{50}$  of 0.32 mm and another with a  $d_{50}$  of 0.94 mm, both with a purity of 99.2 % SiO<sub>2</sub>. This combination of sand grades optimizes particle packing and minimizes voids, ensuring a strong and dense UHPC structure. Controlled variations were systematically introduced to mimic realistic industrial conditions, explore their impacts on UHPC properties, and understand causal effects in the reproducibility challenges of UHPC even when a reference recipe is followed.

The factors investigated are listed in Table 1. Moisture content (*IM*) and temperature (*IT*) of the raw materials were varied to replicate differences caused by storage environments. Typically, such materials in the concrete industry are stored outdoors, which causes variations in temperature and humidity at the moment of use. Graphite (*GRP*) content in silica fume represents impurities in this fine material. The study also considers the impact of the particle size distribution of sands (*SAI*, *SAII*) and fillers (*FLI*, *FLII*), along with measurement errors in dosing sand, filler, and superplasticizer (*SPP*). Average mixer power (*APW*) during the mixing process is recorded as an informative variable for predicting final UHPC quality, which may be affected by these variations in raw materials.

Finally, curing regimes – early age (day 1) and late age (days 2–28) – with temperature profiles of 10–40 °C and conditions such as plastic wrapping, water immersion, or exposure to air at high relative humidity were examined to understand the effect of seasonal and daily environmental conditions on the final UHPC product.

### Data-driven Modeling of the UHPC Production Process

To model the UHPC manufacturing process, a diverse set of ten ML algorithms is employed. The algorithms used are multiple linear regression (MLR)<sup>41</sup>, partial least squares (PLS)<sup>41,42</sup>, kernel ridge regression (KRR)<sup>41,43</sup>, k-nearest neighbors (KNN)<sup>41,44</sup>, support vector regression (SVR)<sup>41,45</sup>, decision trees (DT)<sup>41,45</sup>, random forest (RF)<sup>41,46</sup>, gradient boosting (GB)<sup>41,47</sup>, extreme gradient boosting (XGB)<sup>31</sup>, and Gaussian process regression (GPR)<sup>41,48</sup>. Hyperparameters for these models are optimized using an evolutionary algorithm<sup>49</sup>. These algorithms, which are typically appropriate for modeling small datasets, are widely applied in industrial prediction tasks<sup>50</sup>, particularly for forecasting the mechanical properties of concrete<sup>51</sup>.

The selected modeling approaches are grouped into three broad categories: linear models, nonlinear single models, and nonlinear ensemble models. A linear relationship between the inputs and the output is assumed by linear models. MLR is included as a baseline; PLS is used as a linear latent-factor method effective for handling correlated features or sparse data in small datasets; and KRR is employed, extending ordinary ridge regression with a nonlinear kernel to better tackle

**Table 1.** The dataset ( $X \in \mathbb{R}^{13 \times 150}$ ) is based on a primary ultra-high performance concrete recipe and includes variations in material quality, potential measurement errors, mixer power consumption, and curing conditions. *CCI* comprises two curing classes: specimens stored at 90 % relative humidity (Class = 1) or encased in plastic wrap (Class = 2); *CC28* likewise comprises two classes: specimens encased in plastic wrap (Class = 1) or submerged in water (Class = 2). The cement and silica fume contents remain constant in all experiments. The reference UHPC recipe is designed to yield a 28-day compressive strength of 120 MPa; the observed ranges (85–136 MPa for *CS28* and 8–25 MPa for *FS28*) demonstrate that reproducibility cannot be guaranteed even when identical mix proportions are followed.

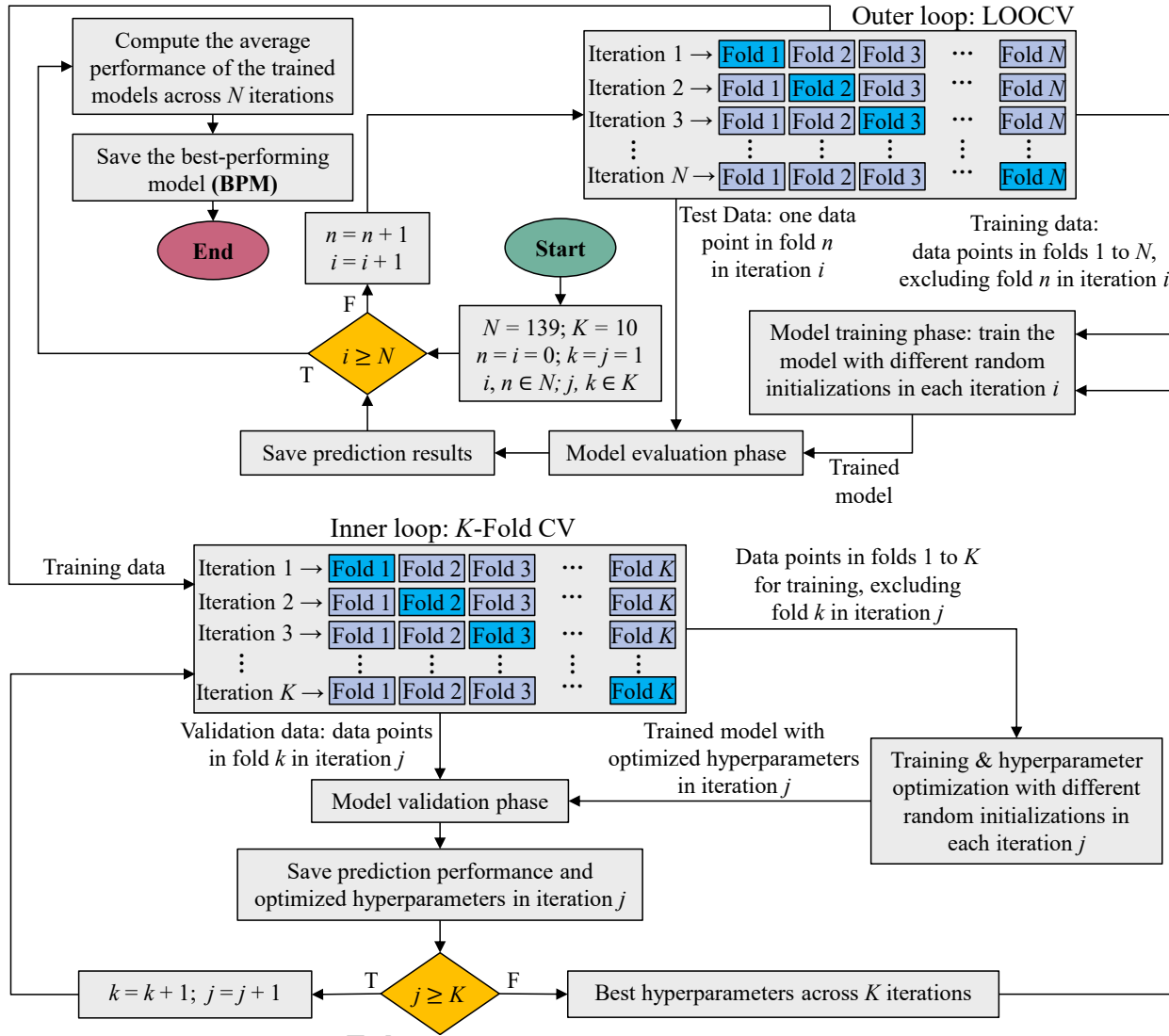
Group	Factor name	Var.	Unit	Mean	Median	SD	Min.	Max.
Raw materials	Ingredient moisture	<i>IM</i>	kg	3.13	3.15	0.16	2.92	3.36
	Ingredient/water temperature	<i>IT</i>	°C	24.20	25	9.05	10	40
	Graphite	<i>GRP</i>	kg	0.08	0.09	0.07	0	0.22
Particle size distributions and measurement errors	Sand I	<i>SAI</i>	kg	5.98	6.00	0.59	5.10	6.90
	Sand II	<i>SAII</i>	kg	10.53	10.50	1.04	8.92	12.07
	Filler I	<i>FLI</i>	kg	6.00	6.00	0.59	5.10	6.90
	Filler II	<i>FLII</i>	kg	0.75	0.75	0.07	0.63	0.86
	Superplasticizer	<i>SPP</i>	kg	0.32	0.32	0.02	0.29	0.35
Mixing condition	Average power consumption	<i>APW</i>	kW	1.04	1.06	0.19	0.36	1.40
Curing conditions	Curing temperature, day 1	<i>CTI</i>	°C	24.60	20	9.72	10	40
	Curing class, day 1	<i>CCI</i>	Class	–	–	–	1	2
	Curing temperature, days 2–28	<i>CT28</i>	°C	22.15	20	9.38	10	40
	Curing class, days 2–28	<i>CC28</i>	Class	–	–	–	1	2
Outputs	Compressive strength at day 28	<i>CS28</i>	MPa	109.83	110.57	11.83	85.06	135.71
	Flexural strength at day 28	<i>FS28</i>	MPa	16.98	17.29	3.60	8.15	24.57

high-dimensional data with limited observations. These linear models are high-bias (low-variance) estimators and thus serve as baseline benchmarks for predictive performance.

Nonlinear single models are stand-alone algorithms capable of capturing nonlinear patterns. In this category, SVR and GPR are included; both are kernel-based methods adept at modeling complex functions – notably, Bayesian priors are leveraged by GPR, which tends to perform well even on small datasets. Additionally, KNN – a simple non-parametric approach that makes predictions based on the local neighborhood of a query point in the feature space – and DT modeling, which provides an interpretable rule-based model that captures nonlinear feature interactions (and also serves as a baseline for tree-based ensemble methods), are included. These single-model approaches were selected to represent a diverse range of learning paradigms – from instance-based learning (KNN) to kernel-based learning (SVR and GPR) and rule-based learning (DT) – covering both parametric and non-parametric techniques.

In nonlinear ensemble models, multiple base learners are combined to improve generalization, especially when dealing with small datasets<sup>52</sup>. In this study, RF is included from the bagging family, which constructs an ensemble of decision trees on bootstrapped subsets of the training data to reduce variance and better handle high-dimensional feature spaces. From the boosting family, GB and its advanced variant, XGB, are employed; an ensemble of decision trees is trained sequentially by these algorithms, with each successive tree correcting errors made by the model thus far<sup>52</sup>. These algorithms were selected due to their demonstrated success in structured regression tasks – particularly in the context of concrete production applications<sup>53,54</sup>. By evaluating multiple ensemble approaches, the performance gains of more complex, high-capacity models relative to simpler models can be observed.

To ensure reliable model development and evaluation on unseen data, a nested training-evaluation loop is implemented, as illustrated in Fig. 2. In the outer loop, a LOOCV strategy is adopted whereby, in each iteration, all data points except one are used for training, while the excluded data point is reserved as unseen data for the final evaluation of the model's performance. In the inner loop, prior to the final model training, the optimization algorithm employs a 10-fold cross-validation strategy on the training data to determine the optimal hyperparameters. Subsequently, the model is built using the complete training dataset with these optimized hyperparameters. The procedure is repeated for every data point, each time using independent model initializations and separate hyperparameter-optimization runs for all ten ML algorithms. Ultimately, BPM for each mechanical property is identified through comparative analysis. Model quality is assessed using three metrics calculated on unseen data (collected from each iteration): coefficient of determination ( $R^2$ ) (1), mean absolute error (*MAE*) (2), and mean



**Figure 2.** Model training and evaluation workflow. The outer loop uses leave-one-out cross-validation (LOOCV), holding out one sample per iteration to estimate performance on unseen data. The inner loop performs 10-fold cross-validation on the training set to optimize hyperparameters. Held-out predictions from all LOOCV folds are concatenated to compute the coefficient of determination ( $R^2$ ), mean absolute error (MAE), and mean absolute percentage error (MAPE). (BPM: best-performing model)

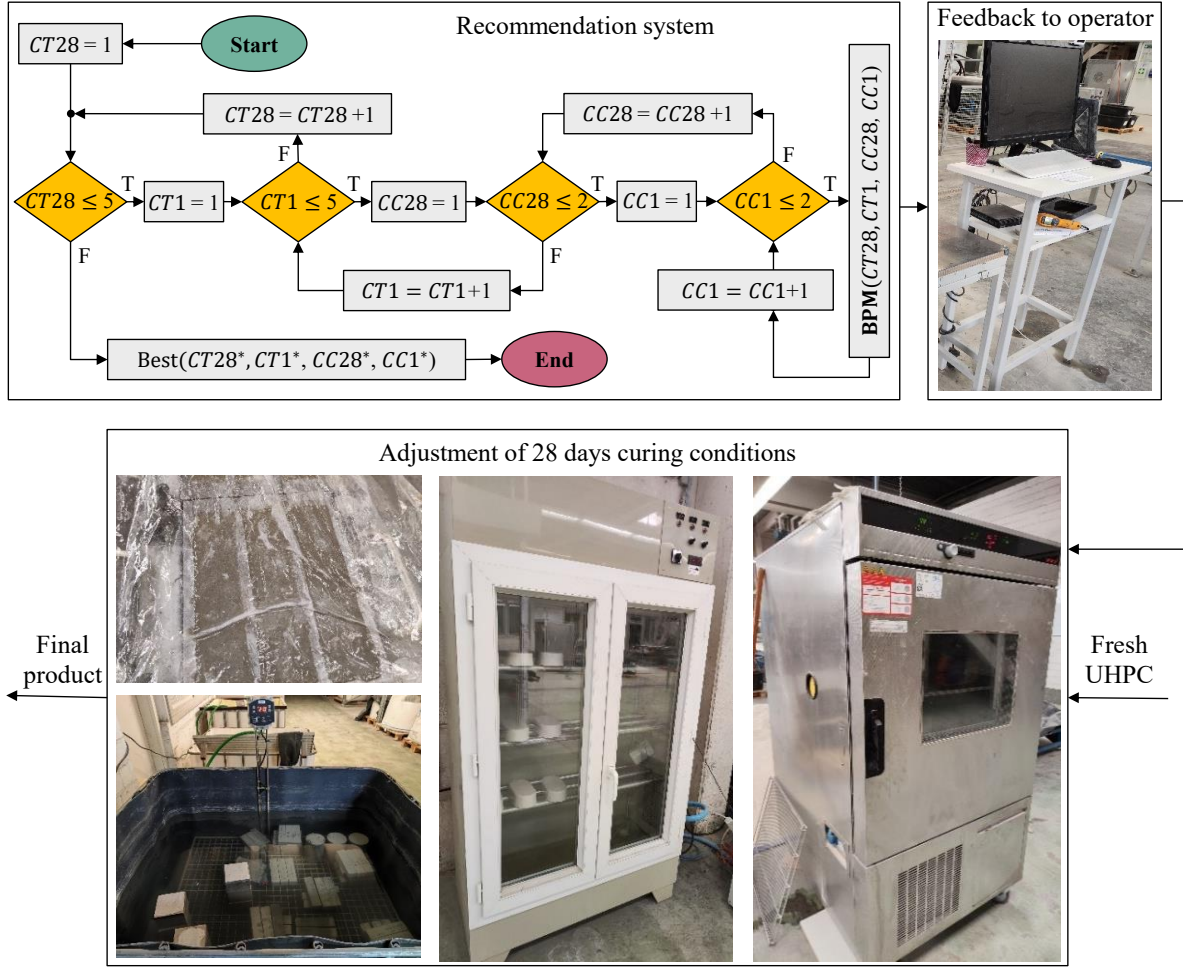
absolute percentage error (MAPE) (3):

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

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

where  $y_i$  is the  $i$ -th true value,  $\hat{y}_i$  is the corresponding predicted value, and  $N$  is the total number of data points.



**Figure 3.** Recommendation engine workflow for curing-setpoint selection. Based on the best-performing model (BPM), the system forecasts the 28-day strength at the end of mixing and proposes batch-specific curing adjustments before curing starts to keep the product within specification. Abbreviations for variables are defined in Table 1.

### Model Interpretability

Model predictions are interpreted by means of SHAP<sup>55</sup>, a game-theoretic framework that decomposes a prediction into additive feature attributions. For a sample  $\mathbf{x}$  with feature set  $F$  and  $M = |F|$ , the model output is represented as

$$f(\mathbf{x}) = \phi_0 + \sum_{j=1}^M \phi_j, \quad (4)$$

where  $\phi_0 = \mathbb{E}_{\mathbf{X}}[f(\mathbf{X})]$  denotes the expected model output under a background distribution, and  $\phi_j$  is the Shapley value for feature  $j$ , defined as the weighted average marginal contribution of  $j$  across all coalitions  $S \subseteq F \setminus \{j\}$ <sup>55,56</sup>,

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(M - |S| - 1)!}{M!} (f_{S \cup \{j\}}(\mathbf{x}_{S \cup \{j\}}) - f_S(\mathbf{x}_S)). \quad (5)$$

This attribution satisfies local accuracy, consistency, and missingness axioms<sup>55</sup>. During the outer LOOCV, SHAP values are computed only for held-out instances, using the fold-specific trained model and a background set drawn from the corresponding training partition to avoid information leakage. For linear estimators, the linear SHAP formulation is used; for non-linear estimators, the model-agnostic KernelSHAP estimator is employed.<sup>55</sup> Global influence is summarized as the mean absolute SHAP magnitude,  $\text{mean}(|\phi_j|)$ , across all held-out predictions, whereas signed SHAP values are used to analyze the direction and magnitude of batch-level effects in subsequent figures.

**Table 2.** Batches excluded from modeling and their exclusion rationale (11 of 150)<sup>36</sup>.

Experiment number	Rationale for exclusion
5, 17, 75, 99, 101	Human error attributable to inadvertent use of material(s) in different composition/volume than the fixed reference recipe; labeled as an outlier.
30	Abnormal early-age curing with suppressed hydration led to unrealistically low 24 h strengths relative to process settings; the batch is judged non-representative of the population and excluded.
36	Abnormally high entrained air together with poor flowability and prolonged efflux behavior indicated non-representative rheology; flagged by univariate checks and by the multivariate consistency screening.
41	Inconsistent fresh-state and hardened measurements (aberrant <i>CS28</i> and <i>FS28</i> ) indicated a procedural/testing anomaly; excluded as non-representative.
47	Physically inconsistent rheology (very large slump-flow concurrent with long V-funnel time) indicated measurement instability; flagged by univariate checks and the multivariate consistency screening.
128	Missing critical targets ( <i>CS28</i> and <i>FS28</i> ); unusable for supervised learning without imputation of outcomes; excluded by integrity check.
148	Unusually high 24 h strengths inconsistent with recorded process conditions, suggestive of logging or testing irregularity; excluded as non-representative.

### Model-Based Recommendation System for Adjusting Curing Conditions

Once BPM for each mechanical property is established, it is embedded in an at-line (end-of-mix), operator-in-the-loop recommendation engine to safeguard the UHPC production line against off-spec output (illustrated in Fig. 3). In practice, operators input the actual mix recipe and process parameters into the BPM, which immediately forecasts the 28-day strength at the end of mixing – before any curing decisions become irreversible. If this initial prediction falls outside the 120 MPa target specification, the engine automatically triggers an exhaustive nested grid search over the four adjustable curing parameters. This model-based recommendation system can thus proactively suggest batch-specific curing adjustments to avoid potential off-spec products or to improve suboptimal quality, thereby reducing UHPC waste.

The search itself is exhaustive in the strict sense; every feasible combination of the four curing variables – initial curing temperature (*CT1*), initial curing condition class (*CCI*), final curing temperature (*CT28*), and final curing condition class (*CC28*) – is systematically evaluated. All other input features (mix composition, etc.) remain fixed. The continuous temperature parameters (*CT1* and *CT28*) are scanned across 10–40 °C in 1 °C increments, while the categorical parameters *CCI* and *CC28* are each tested at their two experimentally validated levels (see Table 1). In total, this yields a discrete search space of  $31 \times 2 \times 31 \times 2 = 3,844$  possible curing condition combinations. For each candidate combination, the BPM predicts the 28-day compressive strength; the algorithm then identifies the combination that minimizes the absolute deviation from the 120 MPa target (i.e. the predicted strength closest to 120 MPa, meeting the design specification and preventing off-spec products).

Because every point in the search space is evaluated exactly once, this brute-force approach is immune to local minima and requires no complex hyperparameter tuning – advantages that make it straightforward to implement in practice. Even with thousands of evaluations, the computation remains fast: the full grid search completes in roughly 30 s on a standard plant workstation, well within the normal production cycle time.

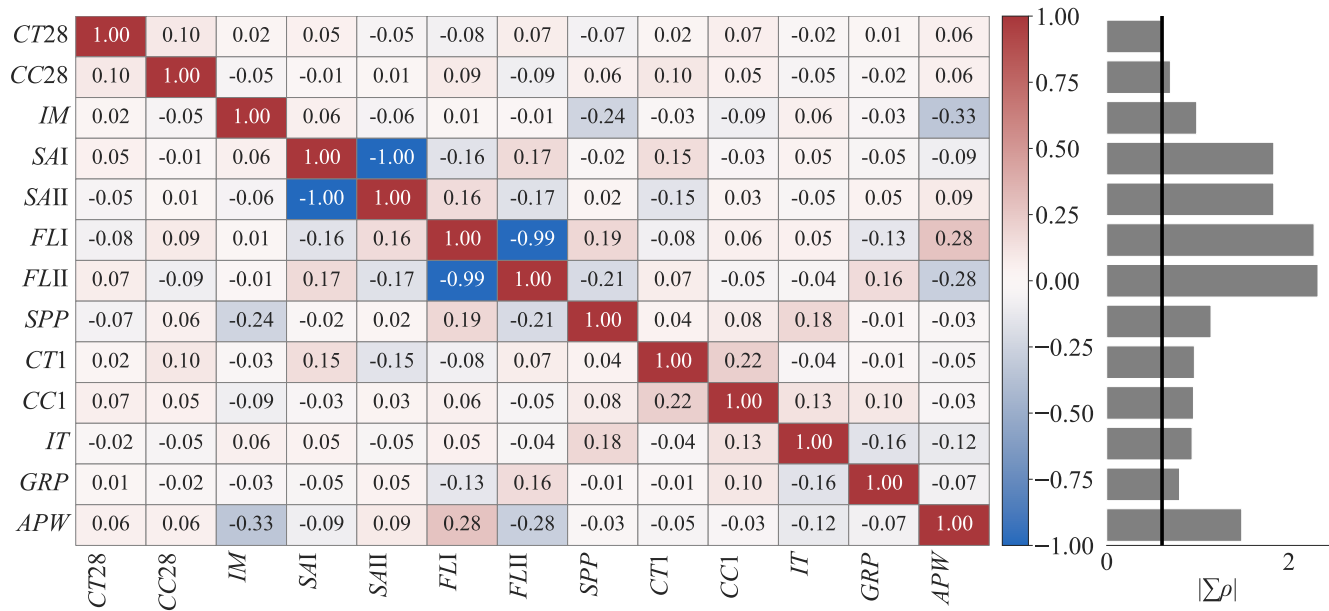
## Results and Discussion

### Data Preprocessing

The modeling inputs comprised 12 controllable factors along with the recorded average mixing power during mixing (APW), while the 28-day mechanical strengths were designated as outputs (Table 2).

Before model development, eleven batches (Exp. 5, 17, 30, 36, 41, 47, 57, 99, 101, 128, 148) were excluded as outliers, see Table 2. Exclusion followed a two-step procedure: (i) an integrity check for missing critical targets (*CS28*, *FS28*); (ii) univariate plausibility screening of fresh-state and strength measurements against the dataset's empirical ranges. The identified outliers and the specific anomalies prompting their exclusion are summarized in Table 2. Eliminating these 11 irregular cases yielded a refined dataset of 139 experiments for subsequent modeling.

Prior to model development, inter-feature dependencies are assessed via Pearson correlation analysis (Fig. 4) to mitigate potential multicollinearity<sup>57</sup>. The nearly perfect negative correlations observed between sand type I (*SAI*) and sand type II (*SAII*), and between filler type I (*FLI*) and filler type II (*FLII*), highlight the delicate balance between these inputs. Such relationships emphasize that the relative proportions fundamentally influence matrix packing density and, therefore, the microstructure of



**Figure 4.** Correlation structure of the UHPC dataset batches). Left: heatmap of pairwise Pearson correlation coefficients ( $\rho$ ); right: bar plot of the row-wise sum of absolute correlations ( $|\sum \rho|$ ), a scalar summary of how strongly each variable co-varies with the remaining feature space. Lower  $|\sum \rho|$  indicates lower collinearity and greater uniqueness; features with the lowest values – *CT28*, *CC28*, *IM*, *GRP* – retain comparatively unique information for subsequent modeling. Abbreviations for variables are defined in Table 1.

UHPC, which ultimately affects mechanical behavior. Certain pairs of highly correlated features – (*SAI*, *SAII*) and (*FLI*, *FLII*) – are subsequently identified for removal from the feature pool to prevent redundancy prior to further analytical steps.

Distributions of all input and output variables after outlier exclusion and the correlation analysis are shown as violin plots of z-scored values (each variable normalized to zero mean and unit variance) (Fig. 5). Each violin depicts the probability density of the 139 recorded values for that feature, with horizontal lines marking the median and interquartile range (embedded within the violin).

LOOCV is employed to robustly evaluate the predictive models on the final dataset ( $N = 139$ ). In each LOOCV iteration, a single batch is held out as the test case, and the remaining 138 batches are used for training the model. Prior to training, feature values in the training fold are normalized to a common scale (0–1) using min–max scaling; the same scaling parameters are then applied to normalize the held-out sample to ensure consistency. No information from the test batch is used in fitting the scaler or the model. This process is repeated 139 times so that each batch is left out once. Thus, every observation is predicted exactly once (using a model fitted to all other observations), and performance statistics are computed from the aggregate of these held-out predictions.

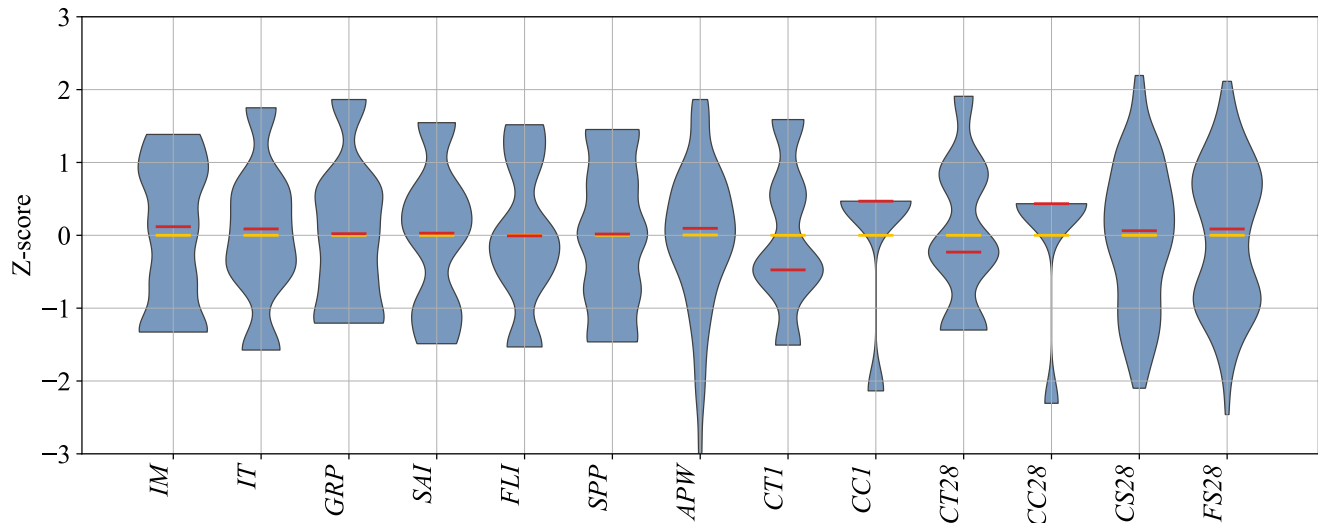
### Evaluation of UHPC Models for Predicting Final Mechanical Properties

Table 3 summarizes the performance of ten ML models for predicting the mechanical properties of UHPC. For compressive strength at day 28 (*CS28*), the PLS model exhibited the best overall performance, achieving the highest  $R^2$  of 74.89 %, the lowest *MAE* of 4.88 MPa, and the lowest *MAPE* of 4.43 %. Notably, MLR, despite its simplicity, performed competitively with an  $R^2$  of 74.54 %. For flexural strength at day 28 (*FS28*), SVR demonstrated superior predictive performance by attaining the highest  $R^2$  (77.91 %). Furthermore, RF and XGB achieved comparable  $R^2$  values of 77.58 % and 77.10 %, respectively; RF yielded the lowest *MAPE* (8.10 %), whereas XGB attained the lowest *MAE* (1.29 MPa).

Overall, the results indicate that PLS and MLR are well-suited for predicting *CS28*, whereas SVR, RF, GB, and XGB are recommended for *FS28* owing to their predictive capabilities and minimal error rates. In particular, the accuracy of PLS makes it especially appealing for *CS28* prediction, while SVR – because of its predictive performance and its simplicity of implementation – is regarded as the most effective for *FS28*. Accordingly, PLS and SVR are selected as the BPMs for *CS28* and *FS28*, respectively.

### Confidence Interval Analysis of Best-Performing UHPC Models

Table 4 reports calibration and predictive performance with 95 % confidence intervals (CIs). For *CS28* (PLS), calibration is statistically indistinguishable from the identity line: slope  $b = 0.98$  (95 % CI: [0.88, 1.07]), intercept  $a = 2.6$  MPa (95 %



**Figure 5.** Empirical distributions of inputs and outputs used in modeling. For each variable, the white dot marks the median, the thick bar indicates the interquartile range, and the envelope shows a kernel-density estimate of the sample distribution. Notably, 28-day compressive strengths range from 85 MPa to 136 MPa (mean 110 MPa), underscoring substantial batch-to-batch variation even under a fixed recipe. Abbreviations for variables are defined in Table 1.

**Table 3.** Performance of ten machine-learning models for predicting 28-day UHPC strength under leave-one-out cross-validation (LOOCV). Metrics are computed from held-out predictions across all folds:  $R^2$  (percent), mean absolute error ( $MAE$ , MPa), and mean absolute percentage error ( $MAPE$ , percent) for compressive strength ( $CS28$ ) and flexural strength ( $FS28$ ). Higher  $R^2$  and lower  $MAE$  /  $MAPE$  indicate better performance. Because all training data are drawn from a single UHPC recipe with controlled perturbations, models are local to that mix. Best values per column are shown in bold.

Model	Abbr.	$R^2$ (%)		$MAE$ (MPa)		$MAPE$ (%)	
		$CS28$	$FS28$	$CS28$	$FS28$	$CS28$	$FS28$
Multiple Linear Regression	MLR	74.54	73.27	4.91	1.51	4.45	9.41
<b>Partial Least Squares</b>	<b>PLS</b>	<b>74.89</b>	73.22	<b>4.88</b>	1.52	<b>4.43</b>	9.43
Kernel Ridge Regression	KRR	72.99	76.57	5.08	1.36	4.62	8.53
K-Nearest Neighbors	KNN	56.09	60.69	6.47	1.76	6.10	10.33
<b>Support Vector Regression</b>	<b>SVR</b>	72.29	<b>77.91</b>	5.16	1.33	4.66	8.31
Decision Tree	DT	57.26	72.90	6.43	1.46	5.87	9.03
Random Forest	RF	68.35	77.58	5.36	1.30	4.87	<b>8.10</b>
Gradient Boosting	GB	67.88	77.10	5.41	1.31	4.91	8.24
Extreme Gradient Boosting	XGB	68.58	77.28	5.35	<b>1.29</b>	4.84	8.18
Gaussian Process Regression	GPR	71.19	76.88	5.26	1.34	4.78	8.46

CI: [-7.9, -13.1] MPa), and bias  $\approx 0$  MPa (95 % CI: [-1.0, 1.0] MPa). Accuracy remained high and stable under resampling ( $R^2_{LOOCV} = 0.749$  [0.669, 0.807],  $MAE = 4.88$  MPa [4.35, 5.44],  $MAPE = 4.43\%$  [3.96, 4.94]), and the parametric 95 % prediction interval for an individual batch at the mean level is  $\pm 11.7$  MPa (about 10 % of 110 MPa). For  $FS28$  (SVR), strong LOOCV performance is observed ( $R^2 = 0.779$ ,  $MAE = 1.33$  MPa,  $MAPE = 8.31\%$ ); the regression slope 0.85 (95 % CI: [0.78, 0.92]) and intercept 2.5 MPa (95 % CI: [1.3, 3.7] MPa) indicate mild compression at the extremes, while the mean bias remains  $\approx 0$  MPa (95 % CI: [-0.25, 0.31] MPa). Metric CIs were obtained by nonparametric bootstrap (10,000 resamples) applied to the LOOCV held-out pairs.

### Decoding UHPC Strength Variability through SHAP: Implications for Low-Waste, High-Quality Manufacturing

The global SHAP analysis of the studied dataset reveals that curing parameters dominate 28-day strength variability, followed by ingredient moisture ( $IM$ ) and impurity in silica fume ( $GRP$ ); all other mix constituents are secondary (Fig. 6).

For both mechanical properties in the secondary curing phase, for example, for compressive strength, batches cured

**Table 4.** Comparison of PLS (*CS28*) and SVR (*FS28*) with 95% confidence intervals (CIs).

Metric	Units	PLS ( <i>CS28</i> )		SVR ( <i>FS28</i> )	
		Value	95% CI	Value	95% CI
Regression slope (pred. vs. obs.)	—	0.98	[0.88, 1.07]	0.85	[0.78, 0.92]
Regression intercept	MPa	2.6	[−7.9, 13.1]	2.5	[1.3, 3.7]
$R^2$	%	74.9	[66.9, 80.7]	77.9	[69.9, 83.7]
<i>MAE</i>	MPa	4.88	[4.35, 5.44]	1.33	[1.17, 1.49]
<i>MAPE</i>	%	4.43	[3.96, 4.94]	8.31	[7.13, 9.30]
Prediction bias (pred−obs)	MPa	≈ 0.0	[−1.0, 1.0]	≈ 0.0	[−0.25, 0.31]
95% prediction interval <sup>a</sup>	MPa	±11.7	—	±3.3	—

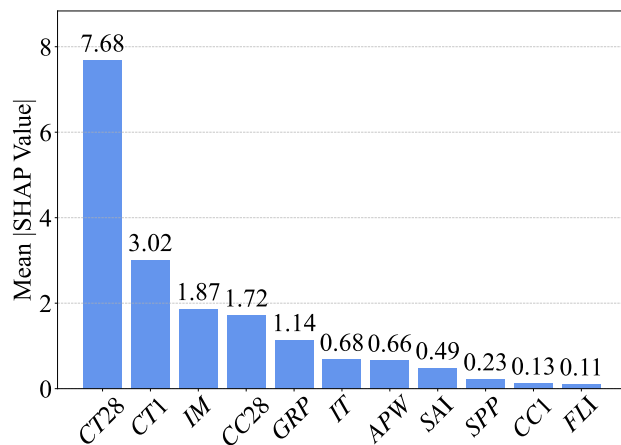
<sup>a</sup> Half-width about the point prediction for a single new observation at the mean level (parametric PI); relative half-widths are ~±10% for *CS28* (mean ~110 MPa) and ~±19% for *FS28* (mean ~17 MPa).

at high temperature (e.g. 40 °C instead of the reference 20 °C) attained higher strengths – for *CS28* the mean SHAP at 40 °C is +17.90 MPa vs. −1.99 MPa at 20 °C ( $\Delta CT28 = +19.89$  MPa); for *FS28* it is +3.44 MPa vs. −0.21 MPa ( $\Delta CT28 = +3.65$  MPa), whereas curing at a cold 10 °C resulted in lower strengths – for *CS28* the mean SHAP at 10 °C is −11.93 MPa vs. −1.99 MPa at 20 °C ( $\Delta CT28 = -9.94$  MPa); for *FS28* it is −3.03 MPa vs. −0.21 MPa ( $\Delta CT28 = -2.82$  MPa) (Fig. 7a). Mechanistically, insufficient curing heat leaves the microstructure under-developed (lower degree of hydration, higher porosity), whereas elevated-temperature curing fosters more complete cement hydration and a denser matrix – greatly enhancing strength. Across both 28-day strengths, *CC28* is among the most influential variables, yet its practical effect points in opposite directions (see *CC28* in Fig. 7). Globally, its mean |SHAP| contribution is about 1.72 MPa for compressive strength (*CS28*, see Fig. 6a) and 1.29 MPa for flexural strength (*FS28*, see Fig. 6b). At the level of the two curing classes, switching from the reference plastic-wrap condition (class 2) to continuous water immersion (class 1) shows absolute mean SHAP values of +5.61 MPa (class 1) vs. −0.99 MPa (class 2) for *CS28* ( $\Delta CC28 = +6.60$  MPa), and −4.24 MPa (class 1) vs. +0.74 MPa (class 2) for *FS28* ( $\Delta CC28 = -4.98$  MPa). This inversion arises because immersion (class 1) maintains full saturation that densifies the matrix and boosts compressive capacity, while simultaneously softening the interfacial transition zone and fiber–matrix bond, which depresses tensile-type (flexural) response. These findings confirm that the reference UHPC mix is highly sensitive to curing regime – without strict control of curing temperature and moisture, the intended performance (~120 MPa) cannot be reliably attained. Indeed, the largest strength shortfalls are traced to suboptimal curing conditions, and adjusting those conditions is shown to rescue the product’s strength in underperforming batches.

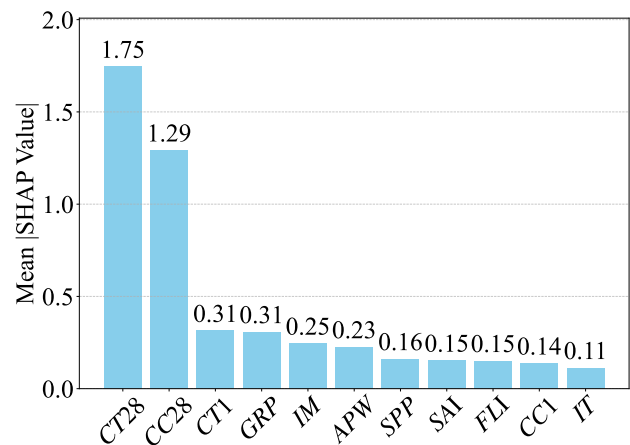
Notably, early-age curing conditions show a somewhat different effect (Fig. 7). Raising the curing temperature on day 1 (*CTI*) actually decreased the eventual 28-day strength when quantified against the reference level: for *CS28*, the mean SHAP at 40 °C is −5.30 MPa vs. +1.82 MPa at 20 °C ( $\Delta CTI = -7.12$  MPa) (Fig. 7a); for *FS28*, it is −0.71 MPa vs. +0.28 MPa ( $\Delta CTI = -0.99$  MPa), whereas a cooler initial cure (e.g. 10°C) yielded a slight strength improvement – for *CS28*, +5.38 MPa vs. +1.82 MPa ( $\Delta CTI = +3.56$  MPa); for *FS28*, +0.31 MPa vs. +0.28 MPa ( $\Delta CTI = +0.03$  MPa) (Fig. 7b). This inverse relationship suggests that overly rapid hydration in the first hours – due to high initial heat – may induce thermal stresses or an unfavorable early microstructure that ultimately reduce strength, whereas a milder initial cure leads to a more refined matrix by 28 days. By contrast, the curing environment on day 1 (*CCI*, i.e. 90 % RH versus sealed curing) had only a negligible influence on final strength (for *CS28*, −0.42 MPa at class 1 vs. +0.06 MPa at class 2;  $\Delta CCI = -0.48$  MPa; for *FS28*, −0.43 MPa vs. +0.07 MPa;  $\Delta CCI = -0.51$  MPa), well inside normal scatter. As long as the fresh UHPC is prevented from drying out in the first 24 h, additional humidity during that period offers little further benefit.

Notably, the two *CC28* curing classes contribute SHAP values of virtually identical magnitude to both mechanical responses. Given that the mean compressive strength at 28 days (*CS28*) is roughly (~110 MPa), whereas the mean flexural strength at 28 days (*FS28*) is only about (~17 MPa) (Table 1), an equal absolute contribution translates into a far larger relative change for *FS28*.

The next major contributor to strength variability is the state of the raw materials – foremost, the initial moisture content (*IM*) of the ingredients (Fig. 6). SHAP shows that if the cement, sand, and powders carry even slightly higher moisture (e.g. 3.364 kg vs. the 2.925 kg reference), the 28-day strength drops – for *CS28*, the absolute level values are −2.83 MPa (3.364 kg) vs. +2.94 MPa (2.925 kg) with  $\Delta IM = -5.77$  MPa (Fig. 7a); for *FS28*, −0.41 MPa vs. +0.34 MPa with  $\Delta IM = -0.75$  MPa (Fig. 7b). The extra free water raises the effective water-to-cement ratio, producing a more porous matrix and thus lower strength. Conversely, very dry ingredients yielded a positive deviation (e.g. for *CS28*, +2.94 MPa at 2.925 kg), indicating the reference mix is calibrated for relatively low moisture. This sensitivity confirms that the mix design is not forgiving to uncontrolled raw material storage conditions – strict control of raw material dryness is needed to hit the strength target



(a) Global SHAP importance (mean |SHAP|, MPa) for predicting CS28 with the PLS model.



(b) Global SHAP importance (mean |SHAP|, MPa) for predicting FS28 with the SVR model.

**Figure 6.** Global feature influence measured by mean absolute SHAP ( $|\text{SHAP}|$ , MPa) across all batches; larger bars indicate stronger influence on the predicted property. Global influence is reported as  $\text{mean}(|\phi_j|)$ <sup>58</sup>. Abbreviations are defined in Table 1.

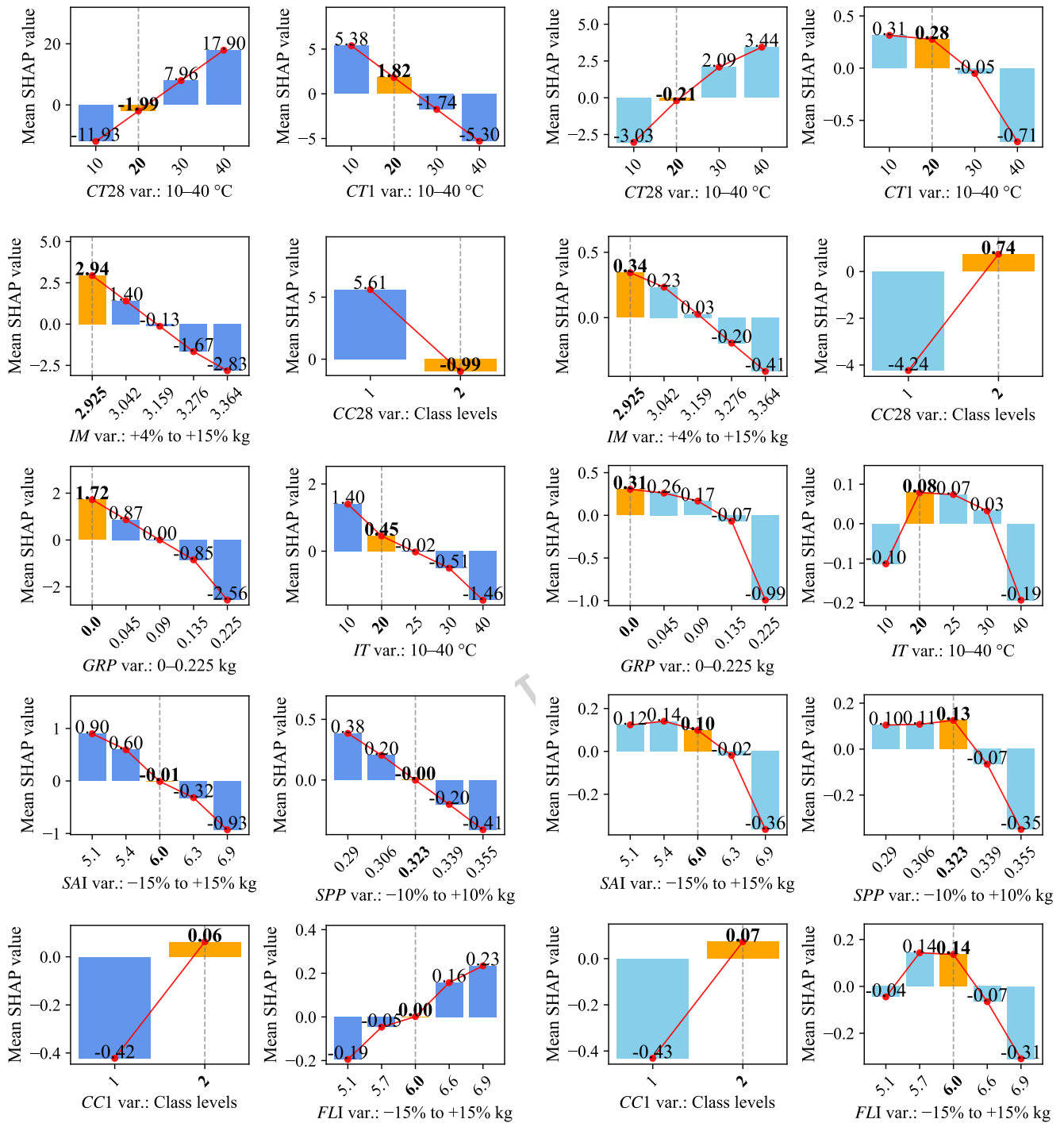
consistently.

Another critical material factor is ingredient purity, represented by the graphite content in silica fume (*GRP*). The reference formulation assumes high-purity silica fume (0% *GRP*); at this level a strength advantage is observed. As impurity levels rise (simulating lower-grade silica fume), both mechanical properties systematically decline – for *CS28*, the absolute level value at the highest *GRP* tested (0.225) is  $-2.56$  MPa vs.  $+1.72$  MPa at 0.000 ( $\Delta\text{GRP} = -4.29$  MPa) (Fig. 7a); for *FS28*,  $-0.99$  MPa vs.  $+0.31$  MPa ( $\Delta\text{GRP} = -1.30$  MPa) (Fig. 7b). These impurities likely act as inert inclusions or interfere with pozzolanic reactions, undermining the microstructure. Hence, adhering to nominal mix proportions is insufficient if raw material *quality* varies; maintaining consistent ingredient quality is vital for consistent outcomes.

Variations in the particle size distribution and dosing of aggregates and admixtures from the reference also influence the 28-day properties. Given that *SAII* and *FLII* are removed from the feature pool because of their high correlation with *SAI* and *FLI*, respectively; the sand and filler composition indices (*SAI* and *FLI*) represent the proportion of type I vs. type II sand and filler in the mix, which affects packing density and thus strength. Deviation from the reference blend can be detrimental. When *SAI* is pushed higher (favoring more of sand type I), the model output drops (SHAP turning negative at high *SAI*, see Fig. 7), while a lower-than-reference *SAI* (more of sand type II) improves strength. This behavior is attributed to the fact that sand type II, with its coarser grading and potentially more angular shape, improves the interlocking and particle packing in the granular skeleton, thereby enhancing the compressive strength. Conversely, a higher fraction of finer sand (type I) can increase the paste demand and reduce the packing efficiency, leading to higher porosity and lower strength. This suggests that the type II sand (used more when *SAI* is low) provided better grading or particle shape, so reducing type I sand beyond the baseline benefited the mix. Conversely for fillers, a greater fraction of filler type I (higher *FLI* than the reference) yielded higher strength, whereas substituting more of filler type II (lower *FLI*) hurt performance. This trend reflects the superior space-filling capacity of filler type I, which is finer and better able to occupy voids between cement grains and sand particles. Filler type II, while enhancing workability, appears less effective in optimizing the particle packing density, resulting in a more porous matrix and diminished mechanical performance. These trends highlight that maintaining a consistent particle size distribution and blend of sources as in the reference mix is critical – otherwise the matrix packing efficiency changes, altering the microstructure and strength. Such deviations in sand and filler grading commonly arise from batch-to-batch variation in raw material supply, measurement inaccuracies during dosing, or insufficient precision in scale calibration.

In addition, accurate dosing of the chemical admixture is vital. The SHAP analysis indicates that overshooting the reference *SPP* (e.g. due to a dosing error adding a few hundredths of a percent more) has a negative effect on both compressive and flexural strength (Fig. 7), likely because excessive superplasticizer can cause segregation or retardation. Slightly under-dosing below the reference might marginally improve strength (since the reference dosage may have been a bit high), but generally, variability in *SPP* leads to inconsistent workability and air entrainment.

The only critical finding when comparing *CS28* and *FS28* is the importance of *IT* (Fig. 6). Relative to the 20°C reference, cooling the constituents to 10°C modestly improves *CS28* by about  $+0.95$  MPa ( $+1.40$  MPa vs.  $+0.45$  MPa), whereas heating them to 40°C reduces it by approximately  $-1.91$  MPa ( $-1.46$  MPa vs.  $+0.45$  MPa); the corresponding shift in *FS28* is

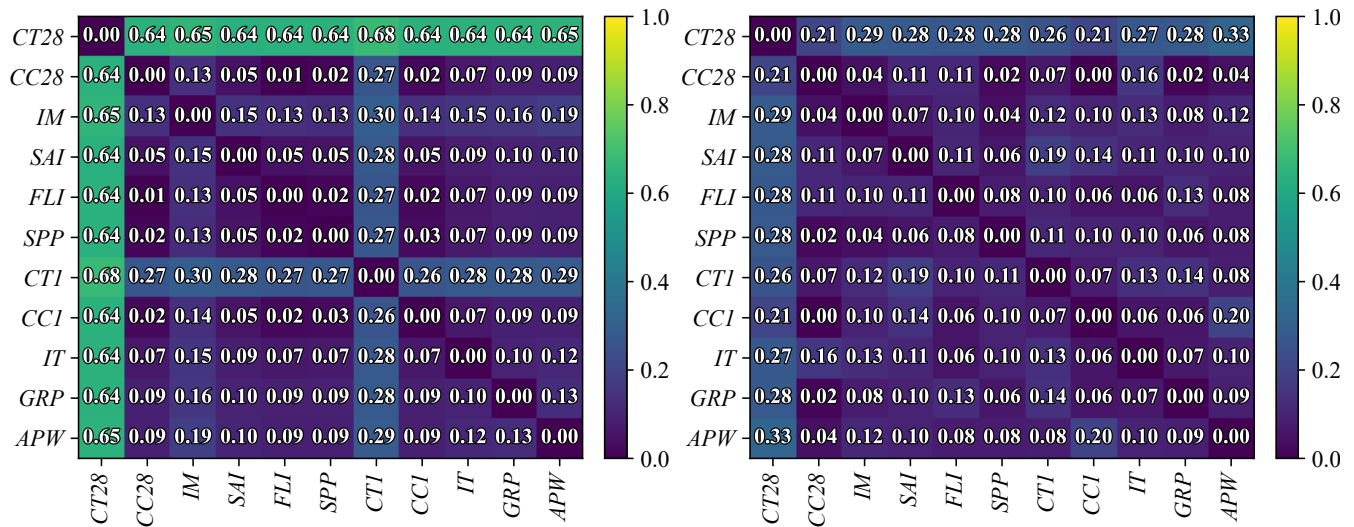


**(a)** Batch-wise signed SHAP contributions (MPa) to the predicted CS28; positive values increase the prediction, negative values decrease it.

**(b)** Batch-wise signed SHAP contributions (MPa) to the predicted FS28; positive values increase the prediction, negative values decrease it.

**Figure 7.** SHAP-based interpretation of the best-performing models at the batch level. Panels show signed SHAP values ( $\phi_j$ , MPa) per parameter and batch<sup>58</sup>. Orange bars (dashed lines) indicate reference levels from the baseline UHPC recipe. Abbreviations are defined in Table 1.

small (at 10 °C: -0.10 vs. +0.08 MPa,  $\Delta IT = -0.18$  MPa; at 40 °C: -0.19 vs. +0.08 MPa,  $\Delta IT = -0.27$  MPa), well within normal scatter. Because similar effects on flexural strength are anticipated, further investigation is warranted before definitive



**(a)** Pairwise interaction heatmap for CS28 (PLS). The ALE-based Friedman  $H$ -statistic ( $0 \leq H \leq 1$ ) is shown; strong non-additivity is centered on  $CT28$ , with the largest pair  $CT28 \times CTI$  ( $H = 0.677$ ). Numeric annotations denote  $H$ .

**(b)** Pairwise interaction heatmap for FS28 (SVR). No strong interactions are present; the largest effects are moderate and  $CT28$ -centered (e.g.,  $CT28 \times GRP$ ,  $H = 0.289$ ;  $CT28 \times IM$ ,  $H = 0.288$ ). Numeric annotations denote  $H$ .

**Figure 8.** Pairwise interaction structure of the models. Heatmaps report the ALE-based Friedman  $H$ -statistic for each feature pair ( $H \approx 0$  additive;  $H > 0.30$  strong two-way effect). Panel (a) shows pronounced curing-centered non-additivity for CS28, while panel (b) shows predominantly additive behavior for FS28 with only mild two-factor synergies. Abbreviations are defined in Table 1.

conclusions are drawn – perhaps the temperature span examined is not wide enough to reveal the full influence of  $IT$ .

In summary, the interpretative analysis confirms that strictly adhering to a fixed UHPC recipe is not sufficient. Curing temperature and humidity in particular govern the extent to which the UHPC mix reaches its potential strength, eclipsing other factors. Fluctuations in raw material moisture and purity further explain why ostensibly identical batches can diverge in quality. Meanwhile, the formulation shows encouraging robustness to small errors in dosing and gradation, which is advantageous for practical production. These insights highlight the need for a smart manufacturing approach: by monitoring the critical parameters and providing feedback (e.g. adjusting curing protocols for a given batch), one can at-line, i.e., before curing for the current batch. This holistic understanding of sensitivity and robustness will be vital for minimizing waste and reliably achieving the ultra-high performance that UHPC formulations are meant to deliver.

### Pairwise Interaction Effects on CS28 and FS28

Pairwise interaction effects are quantified using the ALE-based<sup>59</sup> Friedman  $H$ -statistic<sup>60</sup> ( $H \in [0, 1]$ ), where  $H \approx 0$  indicates additivity<sup>61</sup> and  $H > 0.30$  indicates a strong two-way effect<sup>62</sup> (Fig. 8). The  $H$ -statistic analysis reveals that 28-day curing temperature ( $CT28$ ) has strong two-way interaction effects with virtually every other factor ( $H \approx 0.64$ – $0.68$ ). In particular, the  $CT28 \times CTI$  (early curing temperature) pair achieves  $H = 0.677$ , indicating a highly non-additive influence ( $H > 0.3$  is considered strong). This single interaction contributes nearly half of the model's explained variance ( $H^2 \approx 46\%$ ). All other  $CT28$  combinations (e.g., with internal moisture  $IM$ , mixing energy  $APW$ , ingredient temperature  $IT$ , etc.) yield  $H \approx 0.64$ – $0.65$  (roughly 40–43% variance each), underscoring that the effect of  $CT28$  on CS28 is strongly contingent on those conditions rather than additive. The early curing temperature ( $CTI$ ) itself shows consistent moderate interactions with several parameters ( $H \approx 0.26$ – $0.30$ ) – for instance,  $CTI$  paired with  $IM$ ,  $APW$ , silica/filler indices ( $SAI$ ,  $FLI$ ), or superplasticizer dosage ( $SPP$ ) all have  $H$  between 0.27 and 0.30. Apart from these curing-related synergies, most other factor pairs exhibit much lower  $H$  values (typically  $< 0.2$ ), suggesting that their influences on CS28 are approximately additive with no significant two-factor coupling. For example, the largest interaction outside the curing variables is between  $IM$  and  $APW$  at  $H \approx 0.185$  (moderate), whereas many other combinations (e.g.,  $CC28$  with  $FLI$  or  $SPP$ ) are near zero and essentially additive.

In contrast to CS28, the 28-day flexural strength model is governed more by additive effects, with no factor pair exceeding the strong interaction threshold. The highest two-factor interactions in FS28 are only moderate in magnitude ( $H < 0.30$ ). Notably,  $CT28$  again emerges as an interactive factor, but to a lesser extent: its joint effects with mixture composition (graphite content,  $GRP$ ) and internal moisture ( $IM$ ) are the largest ( $H = 0.289$  and  $0.288$ , respectively).  $CT28$ 's interactions with packing density ( $SAI$ ,  $H \approx 0.281$ ) and ingredient temperature ( $IT$ ,  $H \approx 0.275$ ) are similar in scale, as are those with mixing power,

**Table 5.** Original vs. optimized curing set points for five underperforming batches. The table lists the original curing conditions alongside the model-recommended optimized values for early-age curing (*CT1*, *CCI*) and long-term curing (*CT28*, *CC28*), as well as the corresponding 28-day compressive strengths achieved. Abbreviations are defined in Table 1. (Exp. Nr.: Experiment number)

Exp. Nr.	<i>IM</i>	<i>SAI</i>	<i>FLI</i>	<i>SPP</i>	<i>IT</i>	<i>GRP</i>	<i>APW</i>	Original					Optimized				
								<i>CT1</i>	<i>CCI</i>	<i>CT28</i>	<i>CC28</i>	<i>CS28</i>	<i>CT1</i>	<i>CCI</i>	<i>CT28</i>	<i>CC28</i>	<i>CS28</i>
52	3.364	6	6.9	0.339	40	0.045	1.07	40	2	10	2	85.70	33	1	31	1	117.56
59	3.276	6.9	5.7	0.323	20	0.225	1.04	40	2	40	2	105.11	34	1	36	2	121.74
107	3.042	6	5.1	0.323	25	0.225	1.21	10	2	20	1	107.87	26	2	33	2	119.58
132	2.925	6	6	0.323	25	0	1.18	20	1	10	2	95.94	15	2	25	1	118.26
148	3.364	5.4	5.7	0.29	25	0	1.18	20	1	20	1	107.86	10	1	27	2	120.68

superplasticizer, and filler (all  $H \approx 0.26$ – $0.27$ ). These values imply only moderate departures from additivity, meaning the effect of *CT28* on *FS28*, while context-dependent, is far less pronounced than it was for *CS28*. Aside from *CT28*-related pairs, very few interactions in the *FS28* system are notable – for example, the combination of initial curing condition (*CCI*) and mixing energy (*APW*) shows a moderate  $H \approx 0.21$ . Most other factor pairs (including all involving *CC28*, the 28-day curing condition type) register  $H \lesssim 0.15$  and can be considered nearly additive. In summary, the *FS28* outcome appears predominantly additive with only mild two-factor synergies, whereas *CS28* is distinctly non-additive with strong coupled effects centered on the curing temperature variables.

### Batch-Specific Curing Optimization Restores UHPC Strength

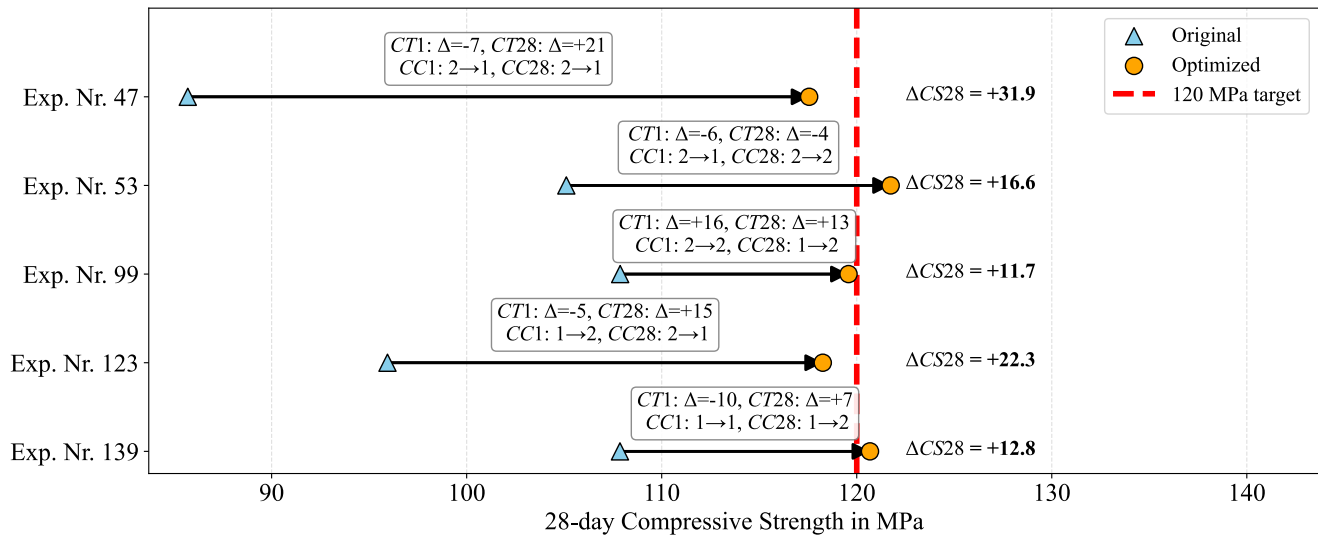
The previous section established that the 28-day compressive strength of the studied UHPC mix is highly sensitive to curing, moisture content, and the purity and gradation of constituent materials. A large variance in final strength (85–136 MPa, Table 1) across nominally identical batches underscores that relying on a fixed recipe is insufficient.

To translate these insights into actionable guidance, the BPM (PLS model; Table 3) for the 28-day compressive strength is embedded in the designed recommendation system (Fig. 3). Five underperforming batches are selected from the investigated UHPC dataset<sup>36</sup> to test the system. For each batch, the original process record (including *IM*, *SAI*, *FLI*, *SPP*, *IT*, *GRP*, and the original curing schedule) is passed to the recommendation engine. The model's recommended curing schedule is used to manufacture a new batch. Table 5 summarizes the input and output variables, and Fig. 9 records the differences between original and recommended curing conditions. The remanufactured batches achieved significant strength gains of 11.7–31.9 MPa, with an average increase of roughly 19 MPa. In other words, adjusting the curing conditions allowed previously rejected batches to reach or exceed the 120 MPa design target, thereby salvaging material that would otherwise have been discarded.

The corrective power of the recommendation system is illustrated by experiment 47. The original batch, cured at 40 °C under plastic wrap during the first 24 h and at 10 °C thereafter, reached only 85.7 MPa. The model prescribed a moderate reduction of the initial temperature (40 → 33 °C), a switch from plastic wrap to 90 % RH air (*CC1*: 2 → 1), and a substantial increase of the long-term temperature (10 → 31 °C) while replacing plastic wrap with water immersion (*CC28*: 2 → 1). These adjustments raised the measured strength to 117.6 MPa. Comparable gains are obtained for the four other underperforming batches (experiments 53, 99, 123, 139; Table 5), all of which converge to 118–122 MPa after optimization.

The improvements are realized through moderate adjustments rather than extreme values. Across the five cases, the long-term curing temperature *CT28* is increased by 7–21 °C relative to the original conditions, but none of the optimized schedules used the maximum 40 °C identified in the global sensitivity analysis (Fig. 7). This moderation reflects trade-offs with the other variables: for example, if the mix is already extremely dry or the superplasticizer dosage is slightly high, then an excessive increase in curing temperature can accelerate hydration beyond the optimum, leading to internal cracking or increased porosity (Table 5). A moderate elevation ensures sufficient hydration while limiting thermal stress.

Instead, values in the 25–36 °C range are recommended. The early-age curing temperature *CT1* is lowered by 5–10 °C in three experiments but raised by 16 °C in one case (Fig. 9). This variability highlights that the optimum depends on the entire vector of material properties: while the overall trend from the previous section suggested that reducing *CT1* improves 28-day strength and raising *CT28* increases strength, the recommendation engine identified that moderate early heat (e.g., 25–30 °C) can be beneficial when raw materials are extremely dry or have lower impurity levels (Table 5). For example, in experiment 99 the original batch is cured at 10 °C and still underperformed; the model recommended increasing *CT1* to 26 °C. In this case, the initial moisture of the dry ingredients is low and the sand grading/filler grading deviated from the reference. A moderate increase in early heat accelerates hydration just enough to offset the slower reaction kinetics associated with dry powders and sub-optimal packing. Excessive heat would induce thermal stresses, but a gentle rise to 25–30 °C can improve early microstructure without causing microcracking. Thus, the optimal curing temperature is context-dependent rather than a



**Figure 9.** Batch-specific curing optimization restores UHPC strength. Arrows indicate the shift from original (triangles) to optimized (circles) 28-day compressive strength; labels show  $\Delta CS_{28}$  and recommended changes in  $CT1$ ,  $CT28$ ,  $CC1$ ,  $CC28$ . The red dashed line marks the 120 MPa target. Modest adjustments recover 11.7–31.9 MPa, bringing all batches to 118–122 MPa. Abbreviations are defined in Table 1. (Exp. Nr.: Experiment number)

fixed 10 °C.

The global analysis indicated that switching from plastic wrapping to continuous water immersion (for  $CC28$ ) increases compressive strength by about 5.6 MPa (Fig. 7a). In practice, the optimal choice is not uniform (Table 5): some remanufactured batches benefited from immersion (class 1), while others performed better under plastic wrap (class 2). This discrepancy arises because the PLS model balances compressive strength against the other measured properties; if the raw material moisture is already high, additional water immersion can push the effective water-to-cement ratio above the optimum, whereas using plastic wrap helps maintain a lower pore volume.

In summary, the variability observed in UHPC production arises from the cumulative effect of multiple uncertainties. Even small deviations in raw material moisture ( $IM$ ), sand grading ( $SAI$  vs.  $SAII$ ), filler grading ( $FLI$ ,  $FLII$ ), superplasticizer dosage ( $SPP$ ), ingredient temperature ( $IT$ ), and impurity in silica fume ( $GRP$ ) can interact in complex ways. For example, an extra 0.4 kg of moisture in the powders increases the effective water-to-cement ratio and lowers the density of the hardened matrix; simultaneously, a higher proportion of the less well-graded sand type (high  $SAI$ ) reduces packing density. When these shifts coincide with an overdosage of superplasticizer, the result is higher porosity and more air entrainment, which drastically lower compressive strength. Such combined deviations explain why some batches originally attained only ~85 MPa (Table 5).

The case studies demonstrate that at-line, operator-in-the-loop, data-driven adjustment of curing conditions is a powerful lever to ensure that UHPC mixes consistently achieve their design strength. Although the global SHAP analysis provides valuable insights into the relative importance of each variable, the optimal curing profile for a specific batch depends on the entire constellation of process parameters. Employing a predictive model to navigate this multidimensional space yields tailored recommendations that restore performance, reduce waste, and contribute to more sustainable concrete production.

### Industrial Deployment Considerations

The framework has been implemented as an at-line, operator-in-the-loop decision-support system. End-of-mix predictions are generated so that any required adjustments can precede curing, and the exhaustive grid evaluation of candidate curing set points (~3,844 combinations) is completed in ~30 s on a standard workstation (Intel® Core™ i9-10900X CPU, 64 GB RAM), which is compatible with typical batch cycle times. Sensor integration has been established as follows. Ambient temperature and relative humidity in the production area are continuously measured and controlled, and curing temperature (for early-age and long-term phases) is monitored by calibrated probes; relative humidity in environmental chambers is likewise monitored to enforce prescribed profiles. Material dosing is executed with high accuracy using calibrated scales to eliminate weighing and admixture-dosage errors. The granular fractions ( $SAI$  /  $SAII$  and  $FLI$  /  $FLII$ ) are prepared to the specified proportions and their particle size distributions are periodically characterized by scanning electron microscopy to ensure reproducible gradation. During mixing, mixer speed and duration are precisely adjustable, the mixer controller is physically connected to the acquisition computer, and the instantaneous power signal  $P(t)$  is streamed at 1 Hz by a Python interface. The stream is

persisted per second, and the average energy consumption are computed and forwarded to the batch record. A graphical user interface (GUI) has been developed for operator interaction: batch-level inputs (ingredient temperatures and moisture, verified dosing amounts, intended curing schedule, and ambient conditions) are entered, the trained model executes in the background, and the predicted 28-day compressive strength together with any recommended curing adjustments are displayed for operator approval and implementation at-line.

However, the present system has been developed and validated in a controlled precast production context, where formwork, cycle times, and curing environments are repeatable. Direct transfer to cast-in-place construction is not anticipated, as field operations typically lack tight environmental control and exhibit non-standard cycle times with constrained sensing and actuation. Furthermore, industrial-scale deployment further presupposes a sufficiently dense sensing and data-integration infrastructure (e.g., continuous curing/maturity monitoring, at-line dosing verification, and ambient-condition logging). Retrofitting legacy plants may entail prohibitive costs and downtime due to capital expenditure, integration and calibration effort, and limited digital connectivity; consequently, near-term benefits are most readily realized in digitally mature precast facilities, whereas broader adoption will depend on lower-cost sensing solutions and lightweight integration.

## Conclusions and Future Work

Ultra-high performance concrete (UHPC) production is prone to batch-to-batch variability in the mechanical properties, even when identical mix proportions are used. This inconsistency often yields off-spec batches that must be discarded, leading to material waste and added CO<sub>2</sub> emissions. Addressing this challenge, the present study developed an at-line end-of-mix ML recommendation system to identify process-induced strength deviations and enable timely corrections to uphold quality.

A comprehensive 150-batch design-of-experiments captured key sources of process uncertainty – raw material quality, storage conditions, dosing errors, and curing regimes – while the base mix recipe remained fixed. Ten diverse ML algorithms were trained on this dataset; the best-performing model (BPM) explained  $\geq 75\%$  of strength variance with a prediction error  $\leq 10\%$  under leave-one-out cross-validation. SHAP analysis showed that long-term curing temperature and humidity dominate strength gain, followed by ingredient moisture and impurity in the silica fume. Guided by these insights, the BPM was embedded in an at-line decision-support tool that proposes batch-specific curing adjustments whenever a strength shortfall is predicted. Curing temperature and humidity thus become control knobs: if upstream variables stray from their ideal ranges, the system selects modified curing settings that bring the predicted 28-day strength (CS28) within the target specification of CS28 = 120 MPa.

Validation with five underperforming batches confirmed the system's efficacy. Model-recommended adjustments – typically  $\pm 20^\circ\text{C}$  temperature shifts or humidity changes – raised compressive strength from an initial 86–108 MPa to 118–122 MPa. The maximum recovery reached 31.9 MPa, and every batch met or exceeded the 120 MPa target. Hence, timely, data-driven intervention can *rescue* batches that would otherwise be rejected, eliminating waste, reducing reprocessing, and lowering the embodied CO<sub>2</sub> of production. Because the reference UHPC uses white Portland cement (600 kg/m<sup>3</sup>) and the plant batch volume is 15 m<sup>3</sup>, the five salvaged batches (75 m<sup>3</sup>) contain 45 t of cement; on a cement-only, cradle-to-gate basis (A1–A3) using 0.913 kg CO<sub>2</sub>e kg<sup>-1</sup> for white Portland cement, this corresponds to  $\approx 41$  t CO<sub>2</sub>e avoided.

The next step is to scale the framework to full industrial trials, integrating live data streams and larger batch volumes. Additionally, applying this experimental–ML methodology to other cementitious systems, additional UHPC types, and recycled concrete could demonstrate its broader relevance to sustainable, reproducible materials manufacturing.

## Data availability

The datasets generated during and/or analysed during the current study are available in the DaKS – University of Kassel's research data repository, <https://doi.org/10.48662/daks-56>.

## Code availability

The underlying code generated during the current study is available from the corresponding author on reasonable request.

## Funding

This project is funded by the University of Kassel through the *BiTWerk* project.

## Author contributions

F.R. conceived and designed the study; developed the methodology and software; curated the data; performed formal analysis and validation; prepared the visualizations; and wrote the original draft. A.A. contributed to conceptualization, data curation,

investigation, and validation, and contributed to writing the original draft. G.Z. contributed to conceptualization, acquired funding, provided resources, and administered the project. A.K. contributed to conceptualization, acquired funding, provided resources, supervised the research, and revised the manuscript. All authors approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## References

1. Abbas, S., Soliman, A. M. & Nehdi, M. L. Exploring mechanical and durability properties of ultra-high performance concrete incorporating various steel fiber lengths and dosages. *Constr. Build. Mater.* **75**, 429–441, DOI: [10.1016/j.conbuildmat.2014.11.017](https://doi.org/10.1016/j.conbuildmat.2014.11.017) (2015).
2. Li, J., Wu, Z., Shi, C., Yuan, Q. & Zhang, Z. Durability of ultra-high performance concrete – a review. *Constr. Build. Mater.* **255**, 119296, DOI: [10.1016/j.conbuildmat.2020.119296](https://doi.org/10.1016/j.conbuildmat.2020.119296) (2020).
3. Richard, P. & Cheyrezy, M. Composition of reactive powder concretes. *Cem. Concr. Res.* **25**, 1501–1511, DOI: [10.1016/0008-8846\(95\)00144-1](https://doi.org/10.1016/0008-8846(95)00144-1) (1995).
4. Russell, H. G. & Graybeal, B. A. Ultra-high performance concrete: A state-of-the-art report for the bridge community. Tech. Rep. FHWA-HRT-13-060, Federal Highway Administration, McLean, VA (2013).
5. Graybeal, B., Crane, C. K., Perry, V., Corvez, D. & Ahlborn, T. M. Advancing ultra-high-performance concrete. *Concr. Int.* **41**, 41–45 (2019).
6. Abrishambaf, A., Pimentel, M. & Nunes, S. Influence of fibre orientation on the tensile behaviour of ultra-high performance fibre reinforced cementitious composites. *Cem. Concr. Res.* **97**, 28–40, DOI: [10.1016/j.cemconres.2017.03.007](https://doi.org/10.1016/j.cemconres.2017.03.007) (2017).
7. Charron, J.-P., Denarié, E. & Brühwiler, E. Permeability of ultra high performance fiber reinforced concretes (UHPC) under high stresses. *Mater. Struct.* **40**, 269–277, DOI: [10.1617/s11527-006-9105-0](https://doi.org/10.1617/s11527-006-9105-0) (2007).
8. Tran, N. T., Tran, T. K., Jeon, J. K., Park, J. K. & Kim, D. J. Fracture energy of ultra-high-performance fiber-reinforced concrete at high strain rates. *Cem. Concr. Res.* **79**, 169–184, DOI: [10.1016/j.cemconres.2015.09.011](https://doi.org/10.1016/j.cemconres.2015.09.011) (2016).
9. Yoo, D.-Y., Lee, J.-H. & Yoon, Y.-S. Effect of fiber content on mechanical and fracture properties of ultra high performance fiber reinforced cementitious composites. *Compos. Struct.* **106**, 742–753, DOI: [10.1016/j.compstruct.2013.07.033](https://doi.org/10.1016/j.compstruct.2013.07.033) (2013).
10. Xi, B., Huang, Z., Al-Obaidi, S. & Ferrara, L. Healing capacity of ultra high performance concrete under sustained through crack tensile stresses and aggressive environments. *Cem. Concr. Compos.* **145**, 105355, DOI: [10.1016/j.cemconcomp.2023.105355](https://doi.org/10.1016/j.cemconcomp.2023.105355) (2024).
11. Brühwiler, E. & Denarié, E. Rehabilitation and strengthening of concrete structures using ultra-high performance fibre reinforced concrete. *Struct. Eng. Int.* **23**, 450–457, DOI: [10.2749/101686613X13627347100437](https://doi.org/10.2749/101686613X13627347100437) (2013).
12. Charron, J.-P., Denarié, E. & Brühwiler, E. Transport properties of water and glycol in an ultra high performance fiber reinforced concrete (UHPC) under high tensile deformation. *Cem. Concr. Res.* **38**, 689–698, DOI: [10.1016/j.cemconres.2007.12.006](https://doi.org/10.1016/j.cemconres.2007.12.006) (2008).
13. Meda, A. & Rosati, G. Design and construction of a bridge in very high performance fiber-reinforced concrete. *J. Bridg. Eng.* **8**, 281–287, DOI: [10.1061/\(ASCE\)1084-0702\(2003\)8:5\(281\)](https://doi.org/10.1061/(ASCE)1084-0702(2003)8:5(281)) (2003).
14. Bertola, N., Kälin, E. & Brühwiler, E. Environmental and life-cycle cost evaluation of a composite timber-UHPC bridge. In *Proceedings of the Third International Interactive Symposium on Ultra-High Performance Concrete (IIS-UHPC 2023)*, 1–8, DOI: [10.21838/uhpc.16641](https://doi.org/10.21838/uhpc.16641) (Iowa State University Digital Press, Des Moines, IA, USA, 2023).
15. Joe, C. D. & Moustafa, M. A. Cost and ecological feasibility of using UHPC in bridge piers. In *Proceedings of the First International Interactive Symposium on Ultra-High Performance Concrete (IIS-UHPC 2016)*, 1–8, DOI: [10.21838/uhpc.2016.103](https://doi.org/10.21838/uhpc.2016.103) (Iowa State University Digital Press, Des Moines, IA, USA, 2016).
16. Fan, J., Shao, Y., Bandelt, M. J., Adams, M. P. & Ostertag, C. P. Sustainable reinforced concrete design: The role of ultra-high performance concrete (UHPC) in life-cycle structural performance and environmental impacts. *Eng. Struct.* **316**, 118585, DOI: [10.1016/j.engstruct.2024.118585](https://doi.org/10.1016/j.engstruct.2024.118585) (2024).
17. Haber, Z. B., McDonagh, M., Foden, A. & Sadasivam, S. Ultra-high performance concrete (UHPC) overlays: An example of lifecycle cost analysis. Tech Note FHWA-HRT-23-012, United States. Federal Highway Administration. Office of Research, Development, and Technology, McLean, VA (2022).

18. Bertola, N., Küpfer, C. & Brühwiler, E. UHPFRC intervention reduces environmental and economic costs in bridge management. *Res. Sq.* DOI: [10.21203/rs.3.rs-6549136/v1](https://doi.org/10.21203/rs.3.rs-6549136/v1) (2025). Preprint.
19. Alsaman, A. *et al.* Comparative sustainability assessment of proprietary and non-proprietary ultra-high performance concrete mixtures. *Infrastructures* **10**, 245, DOI: [10.3390/infrastructures10090245](https://doi.org/10.3390/infrastructures10090245) (2025).
20. Rezazadeh P, F., Dürrbaum, A., Zimmermann, G. & Kroll, A. Leveraging ensemble structures to elucidate the impact of factors that influence the quality of ultra-high performance concrete. In *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, 180–187, DOI: [10.1109/SSCI52147.2023.10371800](https://doi.org/10.1109/SSCI52147.2023.10371800) (IEEE, Mexico City, Mexico, 2023).
21. Wakjira, T. G., Kutty, A. A. & Alam, M. S. A novel framework for developing environmentally sustainable and cost-effective ultra-high-performance concrete (uhpc) using advanced machine learning and multi-objective optimization techniques. *Constr. Build. Mater.* **416**, 135114, DOI: [10.1016/j.conbuildmat.2024.135114](https://doi.org/10.1016/j.conbuildmat.2024.135114) (2024).
22. Rezazadeh P, F., Dürrbaum, A., Zimmermann, G. & Kroll, A. Holistic modeling of ultra-high performance concrete production process: Synergizing mix design, fresh concrete properties, and curing conditions. In *Proceedings - 33. Workshop Computational Intelligence: Berlin, 23.-24. November 2023*, 215–237, DOI: [10.58895/ksp/1000162754-15](https://doi.org/10.58895/ksp/1000162754-15) (KIT Scientific Publishing, Berlin, Germany, 2023).
23. Aylas-Paredes, B. K. *et al.* Data driven design of ultra high performance concrete prospects and application. *Sci. Reports* **15**, 9248, DOI: [10.1038/s41598-025-94484-2](https://doi.org/10.1038/s41598-025-94484-2) (2025).
24. Golafshani, E. M., Behnood, A., Kim, T., Ngo, T. & Kashani, A. Metaheuristic optimization based-ensemble learners for the carbonation assessment of recycled aggregate concrete. *Appl. Soft Comput.* **159**, 111661, DOI: [10.1016/j.asoc.2024.111661](https://doi.org/10.1016/j.asoc.2024.111661) (2024).
25. Kumar, R., Rai, B. & Samui, P. A comparative study of prediction of compressive strength of ultra-high performance concrete using soft computing technique. *Struct. Concr.* **24**, 5538–5555, DOI: [10.1002/suco.202200850](https://doi.org/10.1002/suco.202200850) (2023).
26. Kandiri, A., Sartipi, F. & Kioumars, M. Predicting compressive strength of concrete containing recycled aggregate using modified ANN with different optimization algorithms. *Appl. Sci.* **11**, 485, DOI: [10.3390/app11020485](https://doi.org/10.3390/app11020485) (2021).
27. Marani, A., Jamali, A. & Nehdi, M. L. Predicting ultra-high-performance concrete compressive strength using tabular generative adversarial networks. *Materials* **13**, 4757, DOI: [10.3390/ma13214757](https://doi.org/10.3390/ma13214757) (2020).
28. Ling, H., Qian, C., Kang, W., Liang, C. & Chen, H. Combination of support vector machine and k-fold cross validation to predict compressive strength of concrete in marine environment. *Constr. Build. Mater.* **206**, 355–363, DOI: [10.1016/j.conbuildmat.2019.02.071](https://doi.org/10.1016/j.conbuildmat.2019.02.071) (2019).
29. Hoang, N.-D., Pham, A.-D., Nguyen, Q.-L. & Pham, Q.-N. Estimating compressive strength of high performance concrete with gaussian process regression model. *Adv. Civ. Eng.* **2016**, 1–8, DOI: [10.1155/2016/2861380](https://doi.org/10.1155/2016/2861380) (2016).
30. Nguyen, N.-H., Abellán-García, J., Lee, S., García-Castaño, E. & Vo, T. P. Efficient estimating compressive strength of ultra-high performance concrete using XGBoost model. *J. Build. Eng.* **52**, 104302, DOI: [10.1016/j.job.2022.104302](https://doi.org/10.1016/j.job.2022.104302) (2022).
31. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, 785–794, DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785) (Association for Computing Machinery, San Francisco, CA, USA, 2016).
32. Li, Z. *et al.* Machine learning in concrete science: applications, challenges, and best practices. *npj Comput. Mater.* **8**, 127, DOI: [10.1038/s41524-022-00810-x](https://doi.org/10.1038/s41524-022-00810-x) (2022).
33. Yeh, I.-C. Modeling of strength of high-performance concrete using artificial neural networks. *Cem. Concr. Res.* **28**, 1797–1808, DOI: [10.1016/S0008-8846\(98\)00165-3](https://doi.org/10.1016/S0008-8846(98)00165-3) (1998).
34. Yeh, I.-C. Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cem. Concr. Compos.* **29**, 474–480, DOI: [10.1016/j.cemconcomp.2007.02.001](https://doi.org/10.1016/j.cemconcomp.2007.02.001) (2007).
35. Taguchi, G. *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs* (UNIPUB/Kraus International Publications; American Supplier Institute, White Plains, NY; Dearborn, MI, 1987).
36. Rezazadeh, F., Dürrbaum, A., Abrishambaf, A., Zimmermann, G. & Kroll, A. Mechanical properties of ultra-high performance concrete (UHPC). Dataset, DOI: [10.48662/daks-56](https://doi.org/10.48662/daks-56) (2025). Available since 2025-02-24; License: CC BY 4.0.
37. Refaeilzadeh, P., Tang, L. & Liu, H. Cross-validation. In *Encyclopedia of Database Systems*, 532–538, DOI: [10.1007/978-0-387-39940-9\\_565](https://doi.org/10.1007/978-0-387-39940-9_565) (Springer US, Boston, MA, 2009).

38. Stein, M. Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29**, 143–151, DOI: [10.1080/00401706.1987.10488205](https://doi.org/10.1080/00401706.1987.10488205) (1987).
39. Carnell, R. optaugmentlhs.r [R script]. GitHub repository: [bertcarnell/lhs](https://github.com/bertcarnell/lhs) (2019). Part of the lhs package (GPL-3). Accessed 12 February 2025.
40. Rezazadeh, F., Abrishambaf, A., Dürrbaum, A., Zimmermann, G. & Kroll, A. Investigating reproducibility of ultra-high performance concrete with consistent mechanical properties: A modeling pipeline for complex manufacturing. In Schulte, H., Hoffmann, F. & Mikut, R. (eds.) *Proceedings - 34. Workshop Computational Intelligence: Berlin, 21.–22. November 2024*, 143–148, DOI: [10.58895/ksp/1000174544-9](https://doi.org/10.58895/ksp/1000174544-9) (KIT Scientific Publishing, Berlin, Germany, 2024).
41. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
42. Wegelin, J. A. A survey of partial least squares (pls) methods, with emphasis on the two-block case. Technical Report 371, University of Washington, Department of Statistics, Seattle, WA, USA (2000).
43. Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning (MIT Press, Cambridge, MA, 2012).
44. Fix, E. & Hodges Jr., J. L. Discriminatory analysis. nonparametric discrimination: Consistency properties. *Int. Stat. Rev.* **57**, 238–247, DOI: [10.2307/1403797](https://doi.org/10.2307/1403797) (1989).
45. Hamel, L. H. *Knowledge Discovery with Support Vector Machines*. Wiley Series on Methods and Applications in Data Mining (John Wiley & Sons, Hoboken, NJ, 2009).
46. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32, DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324) (2001).
47. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *The Annals Stat.* **29**, 1189–1232, DOI: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451) (2001).
48. Rasmussen, C. E. Gaussian processes in machine learning. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2–14, 2003; Tübingen, Germany, August 4–16, 2003, Revised Lectures*, vol. 3176 of *Lecture Notes in Computer Science*, 63–71, DOI: [10.1007/978-3-540-28650-9\\_4](https://doi.org/10.1007/978-3-540-28650-9_4) (Springer, Berlin, Heidelberg, 2004).
49. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evol. Comput.* **6**, 182–197, DOI: [10.1109/4235.996017](https://doi.org/10.1109/4235.996017) (2002).
50. Ge, Z., Song, Z., Ding, S. X. & Huang, B. Data mining and analytics in the process industry: The role of machine learning. *IEEE Access* **5**, 20590–20616, DOI: [10.1109/ACCESS.2017.2756872](https://doi.org/10.1109/ACCESS.2017.2756872) (2017).
51. Rezazadeh, F. & Kroll, A. Predicting the compressive strength of concrete up to 28 days-ahead: Comparison of 16 machine learning algorithms on benchmark datasets. In *Proceedings - 32. Workshop Computational Intelligence: Berlin, 1. – 2. Dezember 2022*, 53–75, DOI: [10.58895/ksp/1000151141-4](https://doi.org/10.58895/ksp/1000151141-4) (KIT Scientific Publishing, Berlin, Germany, 2022).
52. Sagi, O. & Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **8**, DOI: [10.1002/widm.1249](https://doi.org/10.1002/widm.1249) (2018).
53. Taffese, W. Z., Zhu, Y. & Chen, G. Ensemble-learning model based ultimate moment prediction of reinforced concrete members strengthened by UHPC. *Eng. Struct.* **305**, 117705, DOI: [10.1016/j.engstruct.2024.117705](https://doi.org/10.1016/j.engstruct.2024.117705) (2024).
54. Nguyen, M. H., Nguyen, T.-A. & Ly, H.-B. Ensemble xgboost schemes for improved compressive strength prediction of uhpc. *Structures* **57**, 105062, DOI: [10.1016/j.istruc.2023.105062](https://doi.org/10.1016/j.istruc.2023.105062) (2023).
55. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 4765–4774 (Curran Associates, Inc., Long Beach, CA, USA, 2017). [1705.07874](https://arxiv.org/abs/1705.07874).
56. Shapley, L. S. A value for  $n$ -person games. In Kuhn, H. W. & Tucker, A. W. (eds.) *Contributions to the Theory of Games, Volume II*, 307–317, DOI: [10.1515/9781400881970-018](https://doi.org/10.1515/9781400881970-018) (Princeton University Press, Princeton, NJ, 1953).
57. Benesty, J., Chen, J., Huang, Y. & Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*, vol. 2 of *Springer Topics in Signal Processing*, 1–4, DOI: [10.1007/978-3-642-00296-0\\_5](https://doi.org/10.1007/978-3-642-00296-0_5) (Springer, Berlin, Heidelberg, 2009).
58. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67, DOI: [10.1038/s42256-019-0138-9](https://doi.org/10.1038/s42256-019-0138-9) (2020).
59. Apley, D. W. & Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. Royal Stat. Soc. Ser. B* **82**, 1059–1086, DOI: [10.1111/rssb.12377](https://doi.org/10.1111/rssb.12377) (2020).

60. Friedman, J. H. & Popescu, B. E. Predictive learning via rule ensembles. *The Annals Appl. Stat.* **2**, 916–954, DOI: [10.1214/07-AOAS148](https://doi.org/10.1214/07-AOAS148) (2008).
61. Molnar, C. *Interpretable Machine Learning* (2024). Feature Interaction & ALE chapters.
62. Herbinger, J. *et al.* Decomposing global feature effects based on generalized functional anova. *J. Mach. Learn. Res.* **25**, 1–49 (2024).

ARTICLE IN PRESS