DOI: 10.58895/ksp/1000162754-15 erschienen in:
Proceedings. 33. Workshop Computational Intelligence - Berlin, 23. – 24. November 2023
DOI: 10.5445/KSP/1000162754 | https://www.ksp.kit.edu/site/books/e/10.5445/KSP/1000162754/

Holistic Modeling of Ultra-High Performance Concrete Production Process: Synergizing Mix Design, Fresh Concrete Properties, and Curing Conditions

Farzad Rezazadeh P.[1], Axel Dürrbaum[1], Gregor Zimmermann[2], Andreas Kroll[1]

[1]Department of Measurement and Control, University of Kassel
E-Mail: {farzad.rezazadeh, axel.duerrbaum, andreas.kroll}@mrt.uni-kassel.de

[2]German Technologies and Engineering Concepts, G.tecz Engineering GmbH
E-Mail: zimmermann@gtecz.com

Abstract

Concrete, the second most consumed resource worldwide after water [1], plays a fundamental role in construction. However, modeling the production process of concrete is challenging due to the incompletely understood physical and chemical relationships among its ingredients and various influencing factors, and the scarcity of data. Current models predominantly only rely on mix design (recipe) data, often overlooking the properties of fresh concrete, the interactions that result from curing conditions, and disturbances. This paper introduces a holistic view that integrates mix design, fresh concrete properties, and curing conditions to enhance predictive models for ultra-high performance concrete (UHPC) quality. This analysis highlights the significant effect of average power consumption, fresh concrete temperature, and curing storage conditions on the quality of concrete.

1 Introduction

Concrete is formulated from cement, aggregates (both fine and coarse), water, and occasionally, admixtures. Its production process begins with the combination of these raw materials, followed by a curing process to ensure the end product quality (Figure 1). The curing process typically requires maintaining specific moisture and temperature conditions for 28 days. This allows the cement to undergo hydration, the pivotal chemical reaction that imparts strength to concrete. The process's intricacy lies in the delicate balance of these components and stages, as well as its susceptibility to external environmental influences, resulting in potential variances in concrete quality [2]. The basic composition of conventional concrete is predominantly characterized by the amalgamation of primary constituents: Cement, fine and coarse aggregates, and water. However, advancements in concrete technology have underscored the integration of supplementary cementitious materials to optimize specific mechanical and rheological properties. Materials such as fly ash, silica fume, blast furnace slag, and superplasticizers, when judiciously incorporated into the mix, can enhance both the compressive strength (CS) and the workability of concrete. This yields, e.g., high-performance and ultra-high performance concrete (Table 1 [3]).

Table 1: Differences between conventional (CC), high-performance (HPC), and ultra-high performance concrete (UHPC) recipes and properties [3]. CS: Compressive strength.

Concrete type	Cement in kg/m^3	Water/binder in %	Workability in mm	CS in MPa
CC	260 – 380	0.45 – 0.65	-	20 – 50
HPC	400 – 700	< 0.4	455 – 810	50 – 100
UHPC	800 – 1000	0.2 – 0.3	260	> 100

Concrete production presents a multitude of challenges that influence the quality and consistency of the end product. The complexity of the process is influenced by the intrinsic properties of the raw materials, the mixing conditions and tools, the environmental factors, and the storage conditions (Figure 2).

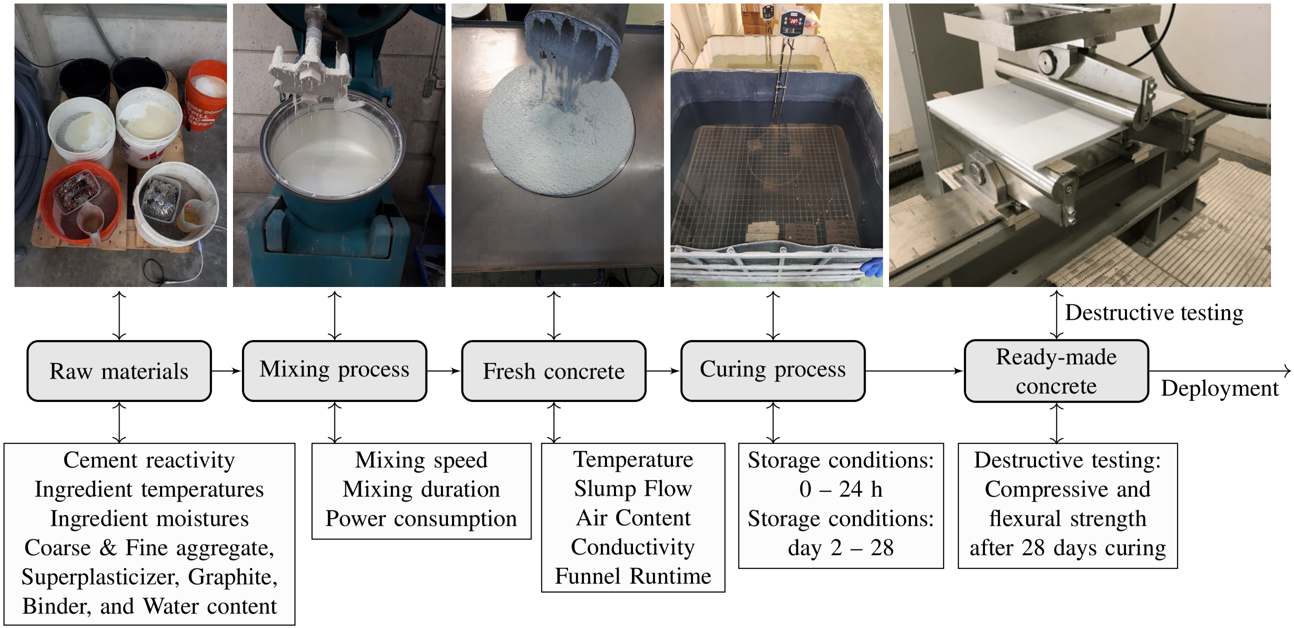


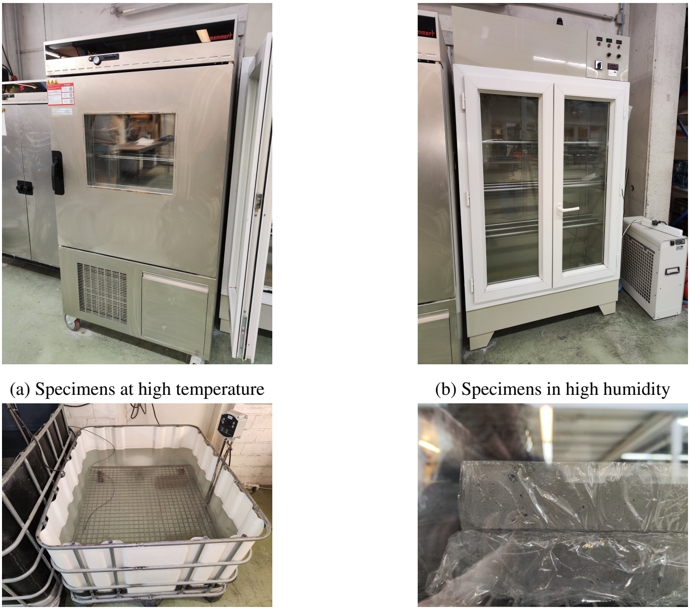
Figure 1: Concrete production and testing process: Mix design, fresh concrete properties, and curing conditions [4]

Traditional paradigms in concrete production modeling have predominantly gravitated towards mix designs, emphasizing input proportions and types [5, 6]. The mere act of mixing predetermined quantities does not invariably guarantee uniformity or the sought properties in the resultant concrete. Characteristics of fresh concrete, such as temperature and workability, are quintessential in predicting the final product quality [7]. Adding to this complexity, the curing process is inherently dynamic. Adjustments made herein, be it due to external environmental conditions or the targeted properties of the concrete, can substantially reshape its micro-structure, and by extension, its macro-behavior. Overlooking these complex nuances could culminate in a limited understanding of the production process, potentially manifesting as inconsistencies, inefficiencies, or even structural vulnerabilities [4].

The primary objective of this contribution, therefore, is to grasp the extent to which these multifaceted factors might shape the process and discern strategies to modulate or adapt them, ensuring reproducible outcomes. In light of these complexities and challenges, a comprehensive framework is proposed in this contribution to model the concrete production process. Designed to eclipse the constraints of traditional recipe-centric models, this framework assimilates insights from fresh concrete characteristics and delves deep into the intricacies of the curing process. Our contribution in this work can be summarized as:

- Determining the important influencing factors on the concrete process.
- Generating data based on the Taguchi orthogonal array L-50 [8] and the characteristics of fresh concrete.
- Adjusting different curing conditions to analyze their impact.
- Concrete process modeling based on four different approaches: Mix design, fresh concrete, curing conditions, and the entire production process, along with analysis of the results.

In our previous study [5], it was observed that two benchmark datasets, which neglected to consider environmental, mix process, and curing conditions in their content, exhibited distinctive behaviors when modeled using data-driven algorithms. In this paper, our primary focus is to analyze the exact impact of these omissions on modeling the concrete production process. Unlike in our



(a) Specimens at high temperature

(b) Specimens in high humidity

(c) Specimens under water

(d) Specimens in plastic foil

Figure 2: Illustration of different curing conditions in the concrete production process

previous work where multiple data-driven algorithms were compared, only the Gradient Boosting method will be used in this study to discern the effects of different modeling approaches.

2 Traditional Data-driven Concrete Modeling

Concrete quality estimation models largely fall into traditional models and machine learning approaches. The well-known Abram's law [9] relates the water-cement ratio (W/C) to the compressive strength (CS) after 28 days:

$$CS = \frac{b_1}{b_2^{W/C}}, \tag{1}$$

where b_1 and b_2 are empirical constants. Enhancing this, Zain et al. [10] introduced multiple linear regression, yielding

$$CS = b_0 + b_1 \frac{W}{C} + b_2 CA + b_3 FA + C. \tag{2}$$

Here, W denotes water volume, C represents cement, CA stands for coarse aggregate, and FA signifies fine aggregate. However, both methodologies neglect the ambient influences, mixing conditions, and the influence of the fresh concrete characteristics and curing conditions.

Modeling the concrete production process using traditional algorithms is challenging due to the many partially known effects on CS. Ling et al. [11] found that among Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Tree, SVM was the superior method for studying the impact of environmental factors on CS. In contrast, Hoang et al. [12] determined that Gaussian Process Regression outperformed both ANN and SVM in estimating CS. Ensemble learning regression, however, provided the most accurate results, as indicated by [13]. Nevertheless, these studies overlook the properties of fresh concrete and curing conditions in their models.

Ozbay et al. [14] explored the mix proportions of high-strength self-compacting concrete using Taguchi's L-18 experimental design, focusing on six pivotal factors to achieve an optimal design. Notably, their work did not consider the potential influence of environmental factors and curing conditions on concrete production. Safranek [15] delved into the role of the mixing protocol, particularly examining the effects of mixing speed and time, in concrete production. Their findings suggest that UHPC necessitates an extended mixing period compared to its conventional counterpart to ensure uniformity. However, mixing at too high a speed could initiate thermal consequences, which might interfere with the chemical processes during blending. Cazacliu et al. [16] embarked on an investigation focusing on the importance of power usage patterns during the mixing process.

Assessing the workability of fresh concrete is vital, with the slump flow test being a key method [7, 17]. Kemer et al. [18] refined the correlation between yield stress and slump results. Hoang and Pham [19] employed LS-SVR for slump prediction. Farzampour [20] explored the relationship between environmental conditions during curing, and the impact of various cement types on concrete's compressive strength. Their findings highlighted that both severe weather conditions in the curing process and the water-to-cement ratio can significantly affect concrete quality.

While various subprocesses of concrete production have been investigated, modeling the entire process considering all major influences remains unexplored.

3 Holistic Concrete Production Modeling

In this contribution, a holistic modeling of the concrete production process is presented, integrating aspects of environmental factors, mix design, fresh concrete properties, and curing conditions. The Gradient Boosting (GB) algorithm [21, 22], in conjunction with recursive feature elimination (RFE) technique [23], is employed for this purpose. The selection of RFE was based on a comparative analysis with other standard methods, namely forward feature selection and backward feature elimination [24]. Among these techniques, RFE demonstrated superior performance, and as such, the outcomes of the other methods will not be discussed further. As for the choice of GB, the primary focus of this study is not to identify the optimal algorithm for modeling the concrete production process but rather to discern the influence of various factors on the final product's quality. Both Random Forest [25] and GB were considered in preliminary tests, with GB yielding better results. It's noteworthy that for techniques like RFE, only algorithms capable of inherently determining feature importance are viable, further justifying our choice.

The developed framework operates on a computer powered by an Intel(R) Core(TM) i9-10900X CPU with 64 GB RAM. Leave-One-Out Cross Validation (LOOCV) [26] learning processes with random initialization are conducted to validate result consistency. Subsequently, the average performance of Gradient Boosting is reported and analyzed, based on the test data garnered through the LOOCV process."

3.1 Modeling Approaches

In the context of mixing design, we refer to the specific recipes or raw material combination, along with their desired proportions (Figure 3). This modeling approach (MA 1) also encompasses the optimal mixing approach, including the appropriate speed and duration for mixing. The second modeling approach (MA 2) evaluates only the fresh concrete properties. Additionally, this modeling approach takes into account the average power consumption during the mixing process. The distinction between mixer adjustments (speed and duration) and average power consumption stems from two main factors. Firstly, mixer adjustments are controllable variables influenced by the type of raw materials, concrete type, and the desired attributes of the end product. Secondly, once water is introduced to the mixture, the subsequent mixing and the corresponding average power consumption after that offer insights into the rheological characteristics of fresh concrete. Because of that, mixer adjustments are analyzed in the first modeling approach (mix design), and average power consumption is examined in the second one. Unlike the second modeling approach that considers only the fresh concrete properties, in the third approach (MA 3), the effects of fresh concrete properties together with curing conditions are investigated (Figure 3).

During evaluation, the accuracy of the GB algorithm is assessed in the RFE process by selecting different numbers of features (3, 4, 5, 6, and 7) for three distinct modeling approaches. As fourth modeling approach (MA 4), a holistic approach integrates all modeling approaches to model the concrete production process (Figure 3). In this comprehensive attitude, the optimal number of features (3, 4, 5, 6, and 7) is re-evaluated using RFE to gauge the performance of the GB algorithm. The results for each phase are then analyzed to determine the most effective combination of factors from both modeling approaches for predicting concrete quality after a curing period of 28 days.

3.2 Recursive Feature Elimination

Recursive feature elimination is a method designed to address the issue of feature selection for machine learning algorithms. By training a model iteratively

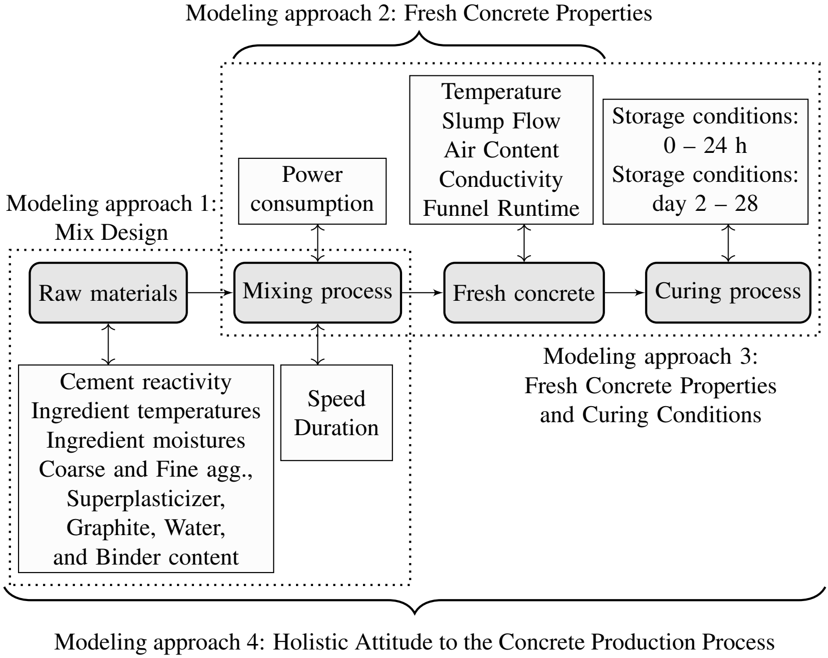


Figure 3: Illustrating the holistic approach from raw material selection to the curing process, emphasizing mix design, fresh concrete properties, and curing conditions.

and eliminating systematically the least important features in each iteration, RFE ensures that only the most impactful features are retained. Gradient Boosting, by its nature, assigns feature importances based on how often a feature is employed to split the data across all trees. This ability makes GB an appropriate choice for RFE, as it can objectively rank features and provide a clear criterion for elimination. This recursive process continues until the desired number of features is retained (Algorithm 1).

3.3 Gradient Boosting Algorithm

Gradient Boosting is a machine learning algorithm that aims to construct a robust predictive model by iteratively building a series of weak learners. Typically, these learners are decision trees. The algorithm iterates by adjusting the

Algorithm 1.: Recursive Feature Elimination with Gradient Boosting

Input: Training data $X \in \mathbb{R}^{N \times D}$ with N samples and D features, targets $y \in \mathbb{R}^N$, Gradient Boosting model, desired number of features to select k
Output: Selected feature set S
Train the Gradient Boosting model on all features in X to obtain feature importances I.
$S \leftarrow \{1, \dots, D\}$ ▷ Initialize feature set with all features
$n \leftarrow D$ ▷ Initialize with total number of features
while $n > k$ **do**
    Remove the feature with the lowest importance from S and corresponding entry from I.
    Retrain the Gradient Boosting model on features in S to obtain updated importances I.
    $n \leftarrow n - 1$
return Feature set S with the k top important features

weights of incorrectly predicted instances, ensuring that the following weak learner focuses more on these challenging instances. The entire process is governed by a predefined loss function, which the algorithm seeks to minimize (Algorithm 2).

4 Experiment Design, Data Collection, and Data Preprocessing

4.1 Controllable Influencing Factors

In order to achieve uniform quality and reproducibility in concrete production, identifying variables that affect consistency is crucial. This includes factors like mixing procedures, storage conditions, the presence of admixtures, and environmental influences, such as temperature and humidity. Tables 2 and 3 list the key factors that were chosen from an initial pool of 25 factors. In this study, cement is categorized as Cement-reactivity-class = 1 if it had been stored for long periods (more than one year), and as Cement-reactivity-class = 2 if it had been shortly stored (less than 3 months). Table 3 also detailed two curing scenarios: Storage-conditions-1T/C (first day storage conditions after mixing)

Algorithm 2.: Gradient Boosting Algorithm

1: **Input:** Training data $X \in \mathbb{R}^{N \times D}$ with N samples and D features, targets $y \in \mathbb{R}^N$, Number of boosting rounds M. $L(y_i, \gamma)$: Loss function measuring the discrepancy between the true target y_i and prediction γ.
2: **Output:** Final boosted model $F_M(x)$
3: Initialize the model with a constant (mean value):

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, \gamma)$$

4: **for** $m = 1$ **to** M **do**
5: **for** each data point i **do**
6: Compute the negative gradient (pseudo-residuals):

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

7: Fit a weak learner $h_m(x)$ to pseudo-residual using $\{x_i, r_{im}\}$
8: Compute multiplier:

$$\gamma_m = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

9: Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

10: **return** $F_M(x) = F_0(x) + \sum_{m=1}^{M} \gamma_m h_m(x)$

and Storage-conditions-28T/C (storage from day 2 to 28). During the first day, concrete was stored at 95 % humidity (Storage-conditions-1C = 1) or at 40 % humidity (Storage-conditions-1C = 2). From days 2 to 28, it was kept at 40 % humidity (Storage-conditions-28C = 1) or submerged in water (Storage-conditions-28C = 2). Given the costly and time-consuming nature of data collection in concrete production, 50 experiments were planned. Considering the factors detailed in Tables 2 and 3, and the constraints of the maximum number of experiments, the Taguchi Orthogonal Array L-50 was employed for data generation. The Taguchi Orthogonal Array ensures data robustness and an equal distribution of data points [8]. After curing for 28 days, the CS of the

Table 2: Factors of mix design: In this investigation, factor values span a range, represented by their designated levels (L: Level). For each category, two distinct aggregates are utilized: coarse and fine. These aggregates are labeled as type I and II within their categories.

Factor	**Abb**	**Unit**	**L 1**	**L 2**	**L 3**	**L 4**	**L 5**
Cement reactivity class	CRC	-	1	2	-	-	-
Ingredient moisture	IM	kg (%)	3.042 (4 %)	2.925 (0 %)	3.159 (8 %)	3.276 (12 %)	3.364 (15 %)
Ingredient temperature	IT	°C	10	20	25	30	40
Coarse aggregate I	CA-I	kg	6.900	6.000	5.400	6.300	5.100
Coarse aggregate II	CA-II	kg	8.925	10.500	11.550	9.975	12.075
Fine aggregate I	FA-I	kg	5.100	6.000	6.600	5.700	6.900
Fine aggregate II	FA-II	kg	0.863	0.750	0.675	0.788	0.638
Superplasticizer	SP	kg	0.290	0.323	0.306	0.355	0.339
Graphite	GP	kg	0.045	0.000	0.090	0.135	0.225
Mixing speed	MS	rad/s	200	350	500	350	350
Mixing duration	MD	s	300	300	300	210	480

specimens was determined using a destructive method. For each experiment, six specimens were tested, i.e. a total of 300 specimens were produced.

4.2 Fresh Concrete Properties

After each mixing process, the properties of fresh concrete are measured. A comprehensive overview of the general characteristics of each property can be found in Table 4. Fresh concrete temperature depends on the concrete mix condition [15], environmental factors, and raw material temperatures. Chemical reactions, notably cement hydration, can affect temperature too. High temperatures reduce workability, and low temperatures can extend setting times.

The air content test gauges the volume of air in fresh concrete as a percentage of its total volume, affecting durability and strength. While higher air content enhances workability and freeze-thaw resistance, it diminishes compressive strength.The average power consumption in concrete production indicates the

Table 3: Factors of the Curing Process: In this investigation, curing condition factors vary across a range, represented by their designated levels. The numbers before T and C denote the curing period in days. (T: Temperature; C: Class; L: Level)

Factor	**Abb**	**Unit**	**L 1**	**L 2**	**L 3**	**L 4**	**L 5**
Storage-conditions-1T	SC-1T	°C	20	20	10	30	40
Storage-conditions-1C	SC-1C	-	1	2	2	2	2
Storage-conditions-28T	SC-28T	°C	20	20	10	30	40
Storage-conditions-28C	SC-28C	-	1	2	2	2	2

mean power utilized for mixing raw materials and overcoming mixture resistance throughout the entire duration of the process. Environmental factors and mixer properties can influence the average power needs. Similarly, chemical reactions, notably between water and cement, can modify the average power demands. High average power consumption might hint at issues like insufficient water, while low average power may suggest a weak mix. Electrical conductivity in fresh concrete reflects largely the ionic content in the liquid phase. This property can indicate the water-to-cement ratio, vital for workability and durability.

The slump flow test evaluates the flowability of fresh concrete, particularly for fluid mixes like self-compacting concrete. Concrete is placed in a slump cone with an outlet diameter of 120 mm. When the cone is lifted, the concrete spreads, and after t = 30, 60, and 120 seconds, the diameter of the spread gives the slump flow test value [7]. High values suggest increased flowability, which can lead to issues like segregation, while low values might pose placement challenges. The funnel runtime assesses the flowability of fresh self-compacting concrete by timing its flow through a V-shaped funnel. Extended funnel times indicate workability concerns, while short times indicate risks like segregation or bleeding.

4.3 Data Preprocessing

In data preprocessing, steps were taken to ensure data integrity. Manual checks are conducted to verify the absence of outliers. The L-50 Taguchi Orthogonal

Table 4: Observed Quantities Related to Fresh Concrete

Factor	Abb	Unit	Min	Mean	Max	STD
Fresh Concrete Temp.	FCT	°C	17.60	25	31.90	3.94
Air Content	AC	%	0.40	1.93	7	1.53
Average Power Consumption	PC	kW	0.37	0.90	1.40	0.24
Slump Flow	SF	mm	120	327.33	395	53.36
Conductivity	CD	V	4.54	4.62	4.74	0.05
Funnel Runtime	FR	s	4	8.05	15	2.69

Array minimizes collinearity risks, and no issues were found. Six missing values in the fresh concrete characteristics led to the exclusion of related experiments. As a result of excluding the related experiments due to the six missing values in the fresh concrete characteristics, the analysis is based on the remaining 44 datapoints. In the project, min-max normalization was chosen due to the presence of varied scales and feature types, the absence of negative values, and the lack of outliers. Additionally, the use of the Taguchi Orthogonal Array inherently facilitated the application of min-max normalization to ensure consistent interpretation across all factors. If $X \in \mathbb{R}^{N \times D}$, each entry can be denoted as X_{ij}, where i ranges from 1 to N and j ranges from 1 to D:

$$X'_{ij} = \frac{X_{ij} - \min_j(X)}{\max_j(X) - \min_j(X)} \tag{3}$$

4.4 General setting for experiments

In the 50 experiments, the same mixing tool is employed (Figure 1). Environmental conditions for material storage and production are controlled, mitigating seasonal influences. All experiments utilized a single material batch for consistent properties. Both the old and new cement were of the same type and originated from the same factory, and production conditions. The mixer chamber temperature was measured before each experiment. Given that the laboratory's ambient temperature was consistently maintained at 20 °C, the

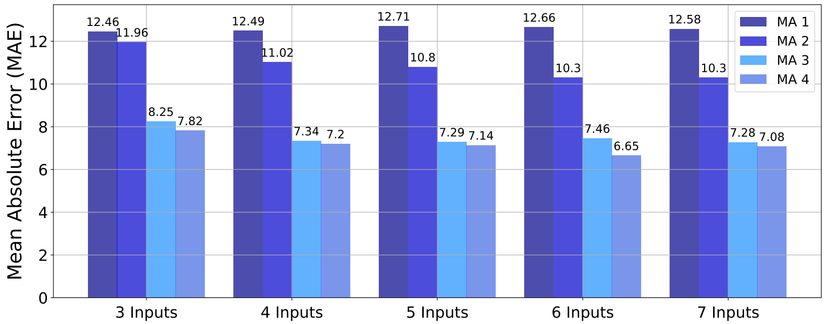


Figure 4: Comparing the prediction accuracy of Gradient Boosting across different modeling approaches (MAs) and also the number of features to be selected by REF. The barplot succinctly illustrates the average performance on the test data, derived from 44 LOOCV iterations. (MA 1: Mix Design, MA 2: Fresh Concrete Properties, MA 3: Fresh Concrete Properties & Curing Conditions, MA 4: Entire Concrete Production Process)

mixer chamber temperature was also close to 20 °C. As a result, this factor did not introduce any variability into the process.

5 Results and Discussion

In this study, the prediction accuracy of Gradient Boosting under four modeling approaches and the number of features selected (either three, four, five, six, or seven) by RFE are analyzed (Figure 4). The objective is to find the combinations of influencing factors from the entire concrete production process that would result in optimal model accuracy. When comparing the prediction accuracy of the models using the same number of selected features across the four modeling approaches (Figure 4), model training on the entire concrete production process consistently yielded the lowest MAE for all modeling approaches. Specifically, utilizing the complete concrete production process with six features yielded the most accurate results, achieving an MAE of 6.65. The mix design consistently exhibited the largest error, indicating that this subset of data might not be as informative for predictions compared to either the fresh concrete and curing conditions data or the comprehensive data from the entire process. In summary, adding more features doesn't always guarantee enhanced performance across all modeling approaches. The data underscores the need for careful feature selection. In the evaluation of feature contribution frequency across the considered modeling approaches, distinct patterns emerged (Figure 5). Within the mix design modeling approach, Ingredient-temperature (44 times) and Mixing-duration (44 times) distinctly stood out, highlighting their central role in modeling the recipe. Additionally, Superplasticizer (40 times), Mixing-speed (41 times), and Graphite (36 times) are of notable significance, reinforcing their essential roles in the mix design modeling approach. Conversely, Cement-reactivity-class (1 time) and Coarse-aggregate-II (2 times) showed minimal importance.

Although the modeling approach was based on fresh concrete data, it isn't elaborated upon in the discussion. This is due to the fact that only 6 fresh concrete features exist, which matches the number of inputs selected in the considered modeling method. In the fresh concrete properties and curing conditions modeling approach, average Power-consumption (44 times), Storage-conditions-28-T (44 times), and Storage-conditions-1-T (44 times) are consistently selected, emphasizing their significant roles in the modeling. Furthermore, Fresh-concrete-temperature (42 times) and Air-content (41 times) made significant appearances, underscoring their relevance. In contrast, electrical conductivity and Slump-flow are less influential.

In the entire concrete production process, the terms average Power consumption, Fresh-concrete-temperature, and Storage-conditions-28T each appeared 44 times, underscoring their critical roles. Storage-conditions-1T (43 times) and Superplasticizer (38 times) also held significant positions. However, features like electrical conductivity (1 time), Funnel-runtime (8 times), Air-content (12 times), and Slump-flow (2 times) are less prominent. To culminate, when examining combinations for a comprehensive representation of the concrete production process, the data from the entire process suggest that average Power-consumption, Fresh-concrete-production, Storage-conditions-28T, Storage-conditions-1T, Superplasticizer, and Graphite are the most vital. This combination promises a comprehensive and accurate modeling of the concrete production process.

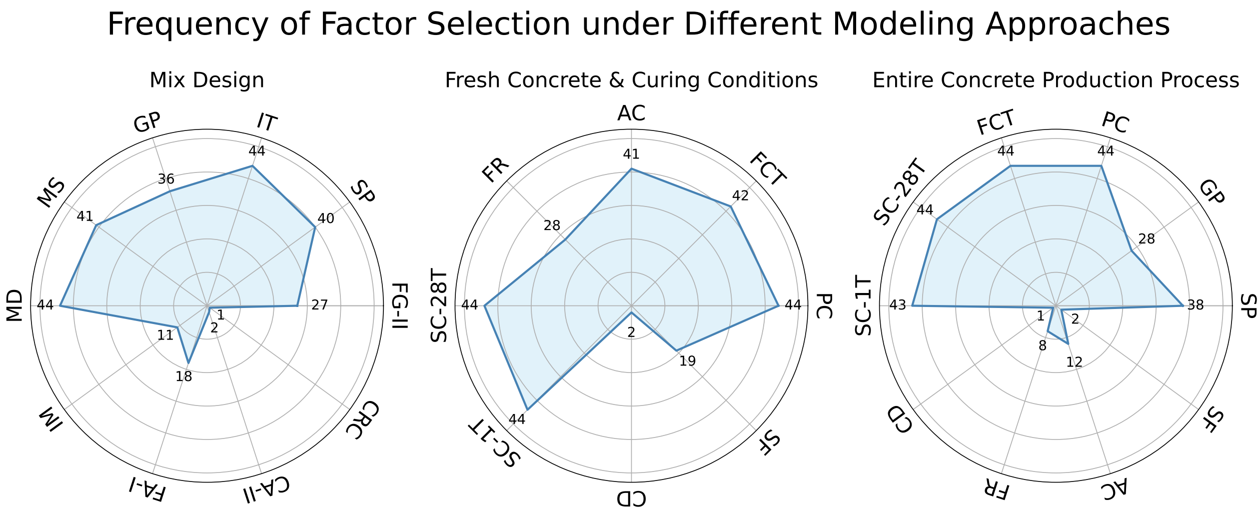


Figure 5: Illustration of how often each of the six influencing factors is selected during the Gradient Boosting training process based on LOOCV by RFE in different modeling approaches. There is a total of 44 iterations. Results from the modeling approach based solely on fresh concrete data are not shown. This is because there are only six features for fresh concrete, and the number of inputs considered for selection in the modeling approaches is also six. This implies that all features were selected in every iteration. **SC-28T:** Storage-conditions-28T, **PC:** Average Power Consumption, **SC-1T:** Storage-conditions-1T, **FCT:** Fresh Concrete Temperature, **SP:** Superplasticizer, **GP:** Graphite, **AC:** Air Content, **MD:** Mixing duration, **IT:** Ingredient temperature, **MS:** Mixing speed, **FR:** Funnel Runtime, **FG-II:** Fine-aggregate-II, **SF:** Slump Flow, **FA-I:** Fine-aggregate-I, **IM:** Ingredient moisture, **CD:** Conductivity, **CA-II:** Coarse-aggregate-II, **CRC:** Cement-reactivity-class, **SC-28C:** Storage-conditions-28C, **CA-I:** Coarse-aggregate-I, **SC-1C:** Storage-conditions-1C

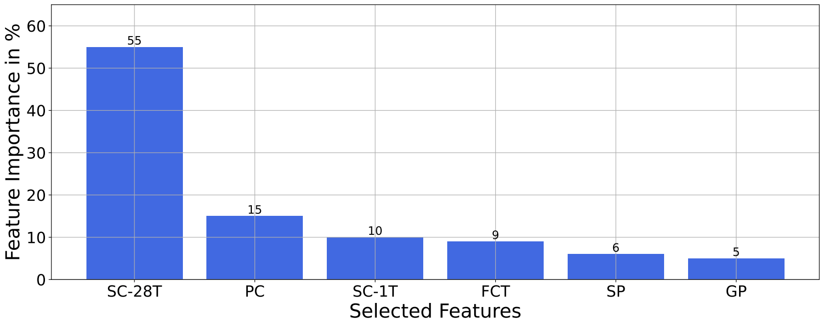


Figure 6: Illustrating the percentage importance of features selected for predicting the compressive strength of concrete, as determined by the chosen GB algorithm, using six inputs and the entire concrete production process as the modeling approach. The SC-28T feature exhibits the highest importance. **SC-28T:** Storage-conditions-28T, **PC:** Average Power Consumption, **SC-1T:** Storage-conditions-1T, **FCT:** Fresh Concrete Temperature, **SP:** Superplasticizer, **GP:** Graphite

In Figure 6, a detailed breakdown of the feature importance is presented. This breakdown was determined from a model that was identified from a series of models trained using various approaches based on LOOCV. Among all these modeling approaches, the one that delivered the best accuracy performance was selected. Within this chosen approach, several models were generated due to the nature of LOOCV. From these models, the one exhibiting a performance closest to the average performance over LOOCV was selected. The feature importances displayed in Figure 6 are derived from this specific model. The chart illustrates that the feature Storage-conditions-28T is of the highest importance, contributing 55 % to the decision-making process of the model. This is followed by average Power-consumption at 15 %, with the remaining features each contributing less than 11 %. In general, that means during the monitoring of the concrete production process, from mix design to the final fresh concrete state, one can predict the eventual quality of the end product. If this predicted quality falls short or is not up to the desired standard, modifications can be made to the curing conditions. By implementing these suitable adjustments, it becomes feasible to achieve the desired quality for the final product, ensuring that the concrete aligns with or surpasses the established benchmarks.

6 Conclusion and Future work

In our previous study [5], it was observed that two benchmark datasets, which neglected to consider environmental, mix process, and curing conditions in their content, exhibited distinctive behaviors when modeled using data-driven algorithms. The presented research underscores the intricacies inherent in the concrete production process and the significance of incorporating mix design, fresh concrete properties, and curing conditions to enhance predictive models for UHPC quality. With this perspective in mind, modifications can be made to the curing conditions. By implementing these suitable adjustments, it becomes feasible to achieve the desired quality for the final product, ensuring that the concrete aligns with or surpasses the established benchmarks.

This contribution also emphasizes that it is not necessary for modeling to measure all factors in the concrete production process. This insight is particularly valuable for concrete plants, considering the costs associated with sensors and the monitoring process. This investigation identified the crucial factors pivotal in enhancing the predictive model's precision, namely: average Power-consumption, Fresh-concrete-temperature, Storage-conditions-28T, Storage-conditions-1T, Superplasticizer, and Graphite. However, it's worth noting that this study was conducted under laboratory conditions. In a real concrete plant, the situation might differ. For instance, controlling the curing process is tough. Wear of the mixing tools and outdoor storage of raw materials, especially before mixing in harsh weather, can impact product quality.

For our subsequent steps, we aim to generate more data and delve deeper into modeling the concrete production process.

Acknowledgment

This project (HA project no. 1011/21-13) is financed with funds of LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz, Förderlinie 3: KMU-Verbundvorhaben (State Offensive for the Development of Scientific and Economic Excellence).

LOEWE

Exzellente Forschung für
Hessens Zukunft

HESSEN

Hessisches Ministerium
für Wissenschaft und Kunst

We thank *Tino Sablotny*, *Michelle Apholz*, and *Amin Abrishambaf* from *G.tecz Engineering GmbH* for carrying out the experiments, and joint work on the analysis of the results.

References

[1] T. Czigler, S. Reiter, P. Schulze and K. Somers. "Laying the foundation for zero-carbon cement". In: *McKinsey and Company Chemicals Insights* 9. 2020.

[2] Z. Li, J. Yoon, R. Zhang, F. Rajabipour, W. V. Srubar III, I. Dabo and A. Radlińska. "Machine learning in concrete science: applications, challenges, and best practices". In: *Npj Computational Materials* 8.1. p. 127. 2022.

[3] M. T. Marvila, A. R. G. de Azevedo, P. R. de Matos, S. N. Monteiro and C. M. F. Vieira. "Materials for production of high and ultra-high performance concrete: review and perspective of possible novel materials". In: *Materials* 14.15. p. 4304. 2021.

[4] F. Rezazadeh P., A. Dürrbaum, G. Zimmermann and A. Kroll. "Leveraging ensemble structures to elucidate the impact of factors that influence the quality of ultra–high performance concrete". In: *IEEE Symposium Series on Computational Intelligence (SSCI)* Accepted. 2023.

[5] F. Rezazadeh and A. Kroll. "Predicting the compressive strength of concrete up to 28 days-ahead: comparison of 16 machine learning algorithms on benchmark datasets". In: *Workshop Computational Intelligence* 1, p. 52. 2022.

[6] I. C. Yeh. “Modeling of strength of high-performance concrete using artificial neural networks”. In: *Cement and Concrete Research* 28, pp. 1797–1808. 1998.

[7] A. Dürrbaum, F. Rezazadeh P. and A. Kroll. “Automatic camera-based advanced slump flow testing for improved reliability” In: *IEEE Sensors* Accepted. 2023.

[8] G. Taguchi. “System of experimental design: engineering methods to optimize quality and minimize costs”. In: *UNIPUB/Kraus International Publications, American Supplier Institute* 1987.

[9] S. Propovics. “Analysis of concrete strength versus water-cement ratio relationship”. In: *Materials Journal* 87.5. pp. 517–529. 1990.

[10] M. F. M. Zain, S. M. Abd, K. Sopian, M. Jamil and A. I. Che-Ani. “Mathematical regression model for the prediction of concrete strength”. In: *WSEAS International Conference. Mathematics and Computers in Science and Engineering* 10. 2008.

[11] H. Ling, C. Qian, W. Kang, C. Liang and H. Chen. “Combination of support vector machine and k-fold cross validation to predict compressive strength of concrete in marine environment”. In: *Construction and Building Materials* 206. pp. 355–363. 2019.

[12] N. D. Hoang, A. D. Pham, Q. L. Nguyen and Q. N. Pham. “Estimating compressive strength of high performance concrete with Gaussian process regression model”. In: *Advances in Civil Engineering*. 2016.

[13] J. Yu, R. Pan, and Y. Zhao. “High-dimensional, small-sample product quality prediction method based on MIC-stacking ensemble learning”. In: *Applied Sciences* 12.1. p. 23. 2021.

[14] E. Ozbay, A. Oztas, A. Baykasoglu, and H. Ozbebek. “Investigating mix proportions of high strength self compacting concrete by using Taguchi method”. In: *Construction and Building Materials* 23. p. 694. 2009.

[15] K. Safranek. “Influence of different mixing processes on the strength of ultra-high strength concretes”. Available at:

https://repositum.tuwien.at/handle/20.500.12708/10881 Diploma thesis, Vienna University of Technology, reposiTUm. 2007.

[16] B. Cazacliu, and N. Roquet. “Concrete mixing kinetics by means of power measurement”. In: *Cement and Concrete Research* 39. p. 182. 2009.

[17] H. Y. Aydogmus et al. “A comparative assessment of bagging ensemble models for modeling concrete slump flow”. In: *Computers and Concrete* 16. pp. 741–757. 2015.

[18] H. Kemer, R. Bouras, M. Sonebi, N. Mesboua, and A. Benmounah. “Slump test: a new empirical model for high yield stress materials”. In: *Journal of Applied Engineering Sciences* 11. pp. 107–112. 2021.

[19] N. D. Hoang, and A. D. Pham. “Estimating concrete workability based on slump test with Least squares support vector regression”. In: *Journal of Construction Engineering* 2016.

[20] A. Farzampour. “Compressive behavior of concrete under environmental effects”. In: *Compressive Strength of Concrete* pp. 92–104. 2019.

[21] J. H. Friedman. “Greedy function approximation: a gradient boosting machine”. In: *Annals of Statistics* pp. 1189–1232. 2001.

[22] F. Pedregosa, et al. “Scikit-learn: Machine learning in Python”. In: *Journal of Machine Learning Research* 12. p. 2825. 2011.

[23] X. Chen and J. C. Jeong. “Enhanced recursive feature elimination”. In: *International Conference on Machine Learning and Applications (ICMLA)* pp. 429–435. 2007.

[24] F. J. Ferri, P. Pudil, M. Hatef, and J. Kittler. “Comparative study of techniques for large-scale feature selection”. In: *Machine Intelligence and Pattern Recognition* 16. pp. 403–413. 1994.

[25] L. Breiman. “Random forests”. In: *Machine Learning* 45. pp. 5–32. 2001.

[26] P. Refaeilzadeh, L. Tang and H. Liu. “Cross-validation”. In: *Encyclopedia of Database Systems* 5. pp. 532–538. 2009.