2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Leveraging Ensemble Structures to Elucidate the Impact of Factors that Influence the Quality of Ultra–High Performance Concrete

1st Farzad Rezazadeh P.
Department of Measurement & Control
University of Kassel
Kassel, Germany
farzad.rezazadeh@mrt.uni-kassel.de

2nd Axel Dürrbaum
Department of Measurement & Control
University of Kassel
Kassel, Germany
axel.duerrbaum@mrt.uni-kassel.de

3rd Gregor Zimmermann
German Technologies & Engineering Concepts
G.tecz Engineering GmbH
Kassel, Germany
zimmermann@gtecz.com

4th Andreas Kroll
Department of Measurement & Control
University of Kassel
Kassel, Germany
andreas.kroll@mrt.uni-kassel.de

***Abstract*—Concrete is an essential material ubiquitously employed in construction. Yet, deciphering the factors that influence its quality is a formidable challenge due to partially understood physical relationships, the high dimensionality of the data, and its limited availability. This study introduces an ensemble framework designed to address these challenges. It uses a combination of individual methods within an ensemble configuration to identify the critical features that determine concrete quality. Within this framework, diverse base methods are harmonized using an average-based technique, leading to a robust final verdict. After selecting the potential influencing factors, 50 experiments are conducted using the Taguchi Orthogonal Array (L-50) to generate the data points. The proposed ensemble learning framework underscores the substantial impact of storage conditions during the curing time on the final quality of concrete.**

***Index Terms*—Concrete, feature importance, ensemble learning**

I. Introduction

Concrete – comprising cement, fine and coarse aggregates, water, and, occasionally admixtures – is essential for many structures, including buildings, bridges, roads, and dams. The production process of concrete, as shown in Fig. 1, can contain highly variable influencing factors. These depend not only on the ingredients and methods used during production, but also on diverse environmental factors [1]. The concrete production process commences with the procurement of raw materials. These materials undergo a mixing process, determined by parameters such as speed and duration, resulting in fresh concrete. The freshly mixed concrete then undergoes a curing process under specific storage conditions, typically over a period of 28 days, before evaluation. Upon finalizing the concrete recipe, an initial prediction of the 28-day concrete quality can be generated. If the predicted quality is insufficient, subsequent modifications to the mixing and curing process parameters can enhance the outcome. Moreover, after mixing, a refined prediction can be obtained based on fresh concrete properties, data from the mixing process, and the recipe. Such predictive insights facilitate strategic adjustments in curing parameters, contributing to the final product quality control.

Concrete is graded into types such as conventional, high-performance, and ultra-high-performance concrete (UHPC), with each displaying unique properties suited to specific construction requirements (Table I). High-performance concrete surpasses conventional concrete in product quality, thanks to the inclusion of supplementary materials and customized mix proportions, whereas UHPC features exceptional resilience due to concentrated binder materials and fiber inclusion [2]. Determining feature importance in concrete production refines experimental designs and helps researchers create accurate predictive models. It also optimizes the process, guiding decisions on materials and procedures, and supports effective monitoring strategies. This determination is crucial for efficient, cost-effective, and quality production [1]. The concrete production process presents numerous challenges. These include issues associated with high dimensionality and small datasets, where generating a single data point takes at least 28 days. The process also involves partially understood physics and chemistry. Furthermore, various factors, such as storage and curing conditions, impact the quality of the end product. Addressing these challenges requires advanced statistical and computational tools, such as machine learning algorithms, which are adept at handling high-dimensional problems. However, these tools often need large amounts of data to function effectively. Hence, a key objective of research and development in the field of concrete technology is to overcome these hurdles [3].

Therefore, a more holistic approach – one that takes into account as many important factors as possible – is required to

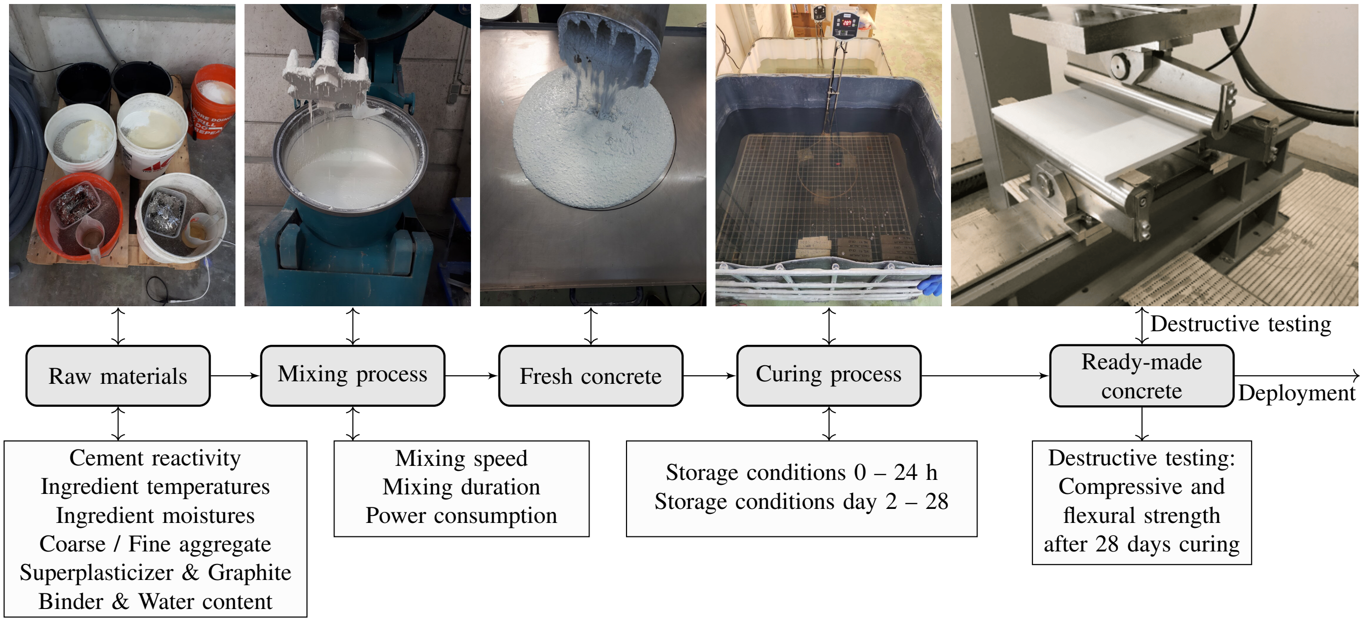

Fig. 1: Concrete production and testing process: process and influencing factors

ensure a comprehensive understanding and optimization of the process. The main contributions of this work are as follows:

- Identifying the most relevant factors that influence the concrete production process
- Conducting experiments based on the L-50 Taguchi Orthogonal Array [4] to generate data under industrial conditions
- Proposing a ensemble structure to determine the importance of selected factors affecting the concrete production process.

II. Related Work

The concrete production process has been studied, yet the depth of exploration into influential factors remains limited. Previous research typically emphasized isolated factors, assuming their broader significance. For instance, in traditional concrete production methods, according to Abram's law [5], an empirical relationship based on the water-to-cement ratio is used to estimate the compressive strength (CS). This relationship is given by

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

where CS represents the CS after 28 days of curing time, and b_1 and b_2 are empirical constants. Reference [6] enhanced Abram's law using multiple linear regression for better estimation accuracy:

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

where W is the water volume, C the amount of cement, CA the amount of coarse aggregate, FA the amount of fine aggregate, and b_0, b_1, b_2, and b_3 are empirical constants. The prevalent dataset named "compressive-strength" [7] aggregates laboratory data from multiple literature sources. A major problem is that while the gathered experimental data have common input factors, they may lack consistency in concrete production conditions, affecting the reliability of the dataset in real-world applications. Reference [8] focused solely on the effects of cement type and curing conditions, concluding that extreme environmental conditions and variations in the water-cement ratio significantly influence concrete quality. Similarly, [9] explored the role of mixing regimes on concrete production, concluding that ultra-high performance concrete (UHPC) demands extended mixing durations for optimal homogeneity, but high speeds may negatively affect the chemical processes involved. Meanwhile, [10] emphasized the importance of power consumption signals during mixing, indicating its significance in the process. Each of these works, while providing valuable insights, narrowly focuses on specific aspects of concrete production, overlooking the complex interplay of multiple influencing factors in the overall process.

The challenge of analyzing high-dimensional data with limited samples is recognized as a complex task. Some researchers, such as [11] and [12], have proposed feature importance methods with two sequential layers. While innovative, these approaches run the risk of eliminating important features in their initial stages. Recent advancements in ensemble feature selection techniques, as observed in studies like [13] and [14], have predominantly been tailored for bioinformatics

TABLE I: Differences between conventional (CC), high-performance (HPC), and ultra-high performance concrete (UHPC) recipes and properties [2]. CS: Compressive strength.

Concrete type	Cement in $\mathrm{kg/m^3}$	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

classification tasks. A recurring theme in these works is the emphasis on subsampling and feature subsetting methods, strategies that may inadvertently overlook crucial inter-feature relationships, especially in smaller datasets such as those encountered in concrete production. Furthermore, the limited diversity of base algorithms in their ensemble structures might compromise the accuracy of feature importance determination.

Diverging from the fragmented approach of previous studies, in this contribution, the influencing factors on concrete production are considered using an ensemble structure with a pool of relevant base algorithms selected based on their diversity performance. This is done with the aim of determining their collective importance in a regression problem.

III. Determination of Feature Importance

The determination of feature importance is facilitated through a variety of techniques typically classified under three principal categories: filter, wrapper, and embedded methods. Filter methods, characterized by their simplicity and speed, evaluate each feature independently, ranking them in accordance with specific statistical metrics. Wrapper methods, on the other hand, perform evaluations at a subset level, identifying the most optimal combination in terms of predictive performance with a designated machine learning algorithm. This approach encompasses the learning algorithm and uses its performance as a metric to assess feature utility. Embedded methods seamlessly integrate feature selection into the training process. This characteristic makes them exclusive to certain learning algorithms. Table II provides a summary of the techniques employed in this study, distributed in accordance with the aforementioned classifications. Each technique is concisely elaborated upon, with a focus on its unique attributes related to the processing of intricate high-dimensional data where sample sizes are restricted. In this contribution, these algorithms are utilized as the base algorithms of the ensemble structure proposed, specifically tailored to determine feature importance in the concrete production process.

IV. Ensemble-based feature importance Determination

Discerning feature importance in complex, small high-dimensional datasets, especially prevalent in concrete production processes, is a daunting task. Ensemble learning is an efficacious strategy to address these challenges by providing a more comprehensive view compared to relying solely on a single model or algorithm [30], [31]. Ensemble learning, akin to the human behavior principle of the “wisdom of the crow”, leverages collective insights garnered from a diverse group of algorithms. Each of these algorithms scrutinizes the data from its unique perspective. The fundamental expectation is that the errors from individual algorithms would be compensated by others, thereby elevating the overall predictive prowess of the ensemble over singular algorithms [32]. The principle of ensemble learning resonates with the mathematical concept encapsulated within the Law of Large Numbers, which states

$$\lim_{n\to\infty} \bar{X}_n = \mu. \tag{3}$$

Here, $\bar{X}_n$ symbolizes the sample mean of n observations and μ signifies the expected value or population mean. This equation proclaims that as the sample size n burgeons, the sample mean $\bar{X}_n$ increasingly converges towards the expected mean of the entire population μ. Given that this contribution is singularly focused on the determination of feature importance, the Law of Large Numbers accentuates the reliability and robustness of the proposed framework.

The performance error of an ensemble framework is inherently bound to be less than or equivalent to the mean performance error of its individual constituents [33]. This error is reduced by the ensemble ambiguity as represented by

$$Error_{ensemble} = \frac{1}{N}\sum Error_{individuals} - Ambiguity, \tag{4}$$

where $Error_{individuals}$ denotes the errors of the ensemble components, N is the total count of components, and $Ambiguity$ encapsulates the diversity in performance across individual components. This mathematical representation elucidates the interplay between ensemble error, individual errors, and diversity. A typical way to quantify the diversity or ambiguity is through the computation of the Pearson correlation coefficient of pairwise base components in the ensemble [34],

$$\rho(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}. \tag{5}$$

The Pearson correlation coefficient ($\rho(X,Y)$) gauges only the linear association between two variables, X and Y. Within this formulation, x_i and y_i signify specific values of X and Y, whereas $\bar{x}$ and $\bar{y}$ indicate their respective means. The element n represents the number of paired values. Another way to quantify the diversity is through Spearman correlation. Spearman correlation, given its foundation in ranks, discerns monotonic nonlinear relationships with greater efficacy as

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \tag{6}$$

where d_i is the difference between the two ranks of each observation and n is the number of observations. Consequently, in this work, the Spearman correlation has been proposed to quantify the diversity among the base algorithms. The associated base algorithms are selected by calculating the correlation for pairwise features and removing one of the features that has a correlation higher than 95 % with another. The average of normalized ranked outcomes of the selected base components is calculated to determine the importance of each feature. The primary feature selection groups and their associated algorithms employed in this framework are detailed in Table II. Based on the Law of Large Numbers and their unique data exploration aspects, 16 algorithms are chosen, making them well-suited to address challenges associated with complex, high-dimensional datasets. Fig. 2 illustrates the proposed ensemble structure incorporating the diverse feature importance identification algorithms. While filter methods look at individual characteristics of each feature, wrapper methods evaluate subsets of features based on their contribution to the

TABLE II: Overview of the methods used in the study to determine the importance of features.

	Method	Description (diverse aspect)
Filter	Partial Correlation Coefficient [15]	The Partial Correlation Coefficient is a measure used in feature selection that assesses the degree of association between two variables, while holding other variables constant (removing the effect of other variables).
	F-value [16]	It is calculated by the F-statistic between each feature and the target. A higher f-value suggests a stronger correlation.
	P-value [16]	The p-value is a statistical approach used to determine feature importance, where the p-value of each feature from a statistical test is evaluated. A lower p-value suggests a stronger correlation.
	RreliefF [17]	It estimates feature relevance by continually updating feature weights based on their ability to distinguish between nearest instances of different labels, thereby identifying features that contribute most effectively to the task.
Wrapper	Recursive Feature Elimination [18]	It is a feature ranking algorithm. It iteratively constructs a model, ranks features based on their importance, and eliminates the least significant ones. This refines the feature set to emphasize those that contribute the most to model prediction. A Support Vector Machine with Leave-One-Out cross-validation and Root Mean Square Error is utilized.
	Sequential Feature Selection [19]	It incrementally selects features based on predetermined criteria. Starting with zero features, it progressively adds or removes them to maximize the predictive performance of the model. This aids in identifying the most relevant feature subset. A Support Vector Machine with Leave-One-Out cross-validation and Root Mean Square Error is utilized.
	Partial Least Squares [20]	It is a dimensionality reduction technique. It projects both the predictors and the response onto a new space. PLS identifies directions in the predictor space that account for the variance in both the predictors and the response. This makes it an effective method for determining feature importance, especially in cases of high dimensionality and limited data.
	Boruta [21]	It is based on the random forest algorithm; it iteratively evaluates and removes features deemed less important, effectively retaining those that are statistically significant, and thereby provides a robust solution for identifying relevant features.
Embedded	Lasso [22]	Utilizing regularization ($\|x\|_1$), it assigns insignificant features a zero weight by imposing a constraint on the sum of absolute values of model parameters (addressing high dimensionality and a small amount of data).
	Ridge Regression [23]	Leveraging regularization ($\|x\|_2$), ridge regression balances the magnitude of coefficients by incorporating a penalty equal to the sum of the squares of the model parameters (addressing high dimensionality and a small amount of data).
	Elastic Net [24]	It combines Lasso's feature selection with Ridge's overfitting prevention by integrating both L1 and L2 penalties.
	Decision Tree [25]	It is a structure where each node represents a feature, each branch a decision rule, and each leaf node an outcome; by following paths from root to leaf, it prioritizes significant features based on the information gain (addressing nonlinearity).
	Random Forest [26]	It is a bagging ensemble learning method that constructs multiple decision trees. For regression, it outputs the label based on the mean prediction. This process inherently ranks features in terms of their importance based on their impact across the various trees, addressing nonlinearity and high dimensionality.
	Gradient Boosting [27]	It is a derivative-based method in the form of a boosting ensemble of weak prediction models, typically decision trees. It identifies important features by fitting new models to the residuals of the prior one, enhancing the impact of significant features in the final prediction and addressing nonlinearity and high dimensionality.
Other	Gaussian Process Regression [28] (Permutation [26])	It leverages Permutation testing in conjunction with Gaussian process models, assessing feature importance by measuring the impact on model performance when the values of a particular feature are randomly shuffled. Larger impacts signify higher feature relevance, addressing nonlinearity, high dimensionality, and a small amount of data.
	Total-Sobol Index [29]	It is a powerful tool in sensitivity analysis that measures the effect of an input variable on the variance of the output; it identifies influential features by attributing the change in the output variance to the variance of individual input variables, thereby, highlighting the importance of different features (addressing nonlinearity and high dimensionality).

performance of a model, and embedded methods integrate feature selection into the model training process. The "Other" category includes unique algorithms such as Gaussian process regression, permutation testing, and the Total Sobol index. The former gauges feature significance based on shuffled feature value impacts, while the latter utilizes output variance changes to underscore individual input variable importance. This configuration aims to alleviate any single method bias, potentially leading to critical features being overlooked. To achieve robust performance, the ensemble structure adheres to the following criteria [31]:

- Performance independence among individual methods.
- Decentralized decision-making, avoiding reliance on a single method.
- Diverse data manipulation afforded by the unique attributes / aspects of each method.
- Aggregatable decisions from individual methods to formulate the final output.

The proposed framework starts with an empty weight list for each of the 16 features in the dataset (Algorithm 1). For each algorithm in the ensemble, it assigns weights to every feature based on its importance. Once all weights are gathered, the framework calculates pairwise Spearman correlation coefficients. If any pair of features has a correlation greater than 0.95, one of the features from the pair is removed. The

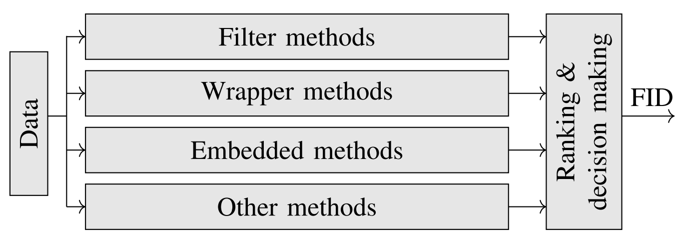

Fig. 2: Ensemble-based Feature Importance Determination (FID): 16 different base algorithms from filter, wrapper, embedded categories, and others (including Gaussian process regression, permutation testing, and the Total Sobol index) selected for their diversity and in accordance with the Law of Large Numbers.

mean weight for each remaining feature is then computed, and features are finally ranked based on their mean weights to determine their overall importance.

V. Data Collection

A. Influencing factors

In the pursuit of uniform product quality in the concrete production process, even with consistent recipes and material properties, it is essential to discern variables contributing to product variation. This may encompass procedural elements such as mixing, storage, and curing conditions, admixture presence, and environmental factors, such as ambient temperature and humidity. The selected factors (from a pool of 25 identified factors) in Table III represent a balanced cross-section of the process and its context, fostering a comprehensive understanding for achieving consistent, high-quality concrete production that is reproducible. The storage conditions for concrete ingredients, specifically temperature and moisture levels, are critical factors that influence their quality from the initial stage of the mixing process. These conditions can trigger undesired chemical reactions and impact the water absorption capacity of the ingredients. Notably, the duration of cement storage can diminish its reactivity. Hence, in this study, cement stored for extended periods was classified as Cement-reactivity-class = 1, while freshly stored cement was classified as Cement-reactivity-class = 2. As illustrated in Table III, two other specific storage conditions are considered: Storage-conditions-1T/C and Storage-conditions-28T/C. Storage-conditions-1T/C refers to the storage environment for fresh concrete, typically at the end of the mixing process, for the first 24 hours (or the first day). On the first day, fresh concrete is stored either in a climate cabinet with a relative humidity of 95 % (Storage-conditions-1C = 1) or in air (Storage-conditions-1C = 2). Conversely, Storage-conditions-28T/C signifies the storage conditions of the concrete from the second to the 28th day. In this period, the concrete specimens are stored either in air with a relative humidity of 40 % (Storage-conditions-28C = 1) or underwater (Storage-conditions-28C = 2).

B. Experiments

Due to the cost and time-intensive nature of data generation in the concrete production process, 50 experiments were planned for this study. Therefore, given the constraints of data generation, the number of factors, their respective lower and upper bounds, and the considered levels (Table III), the Taguchi Orthogonal Array L-50 was utilized for data generation. This statistical method ensures robustness against noise and equal data distribution in the input space [4]. In addition to the factors included in the L-50 table, the "power consumption" of each experiment during the mixing process was recorded, and the average was added to the data as a new influencing factor for further analysis Subsequent to a 28-day curing period, CS assessments were undertaken with six specimens, and flexural strength (FS) assessments were done using three specimens for each experiment, both through standardized destructive testing methods.

Algorithm 1 Feature Importance Ranking with Ensemble Structure: the dataset comprises of 16 features.

Initialize: Start with an empty list for each feature in the dataset.
for each algorithm in the ensemble **do**
  for each feature in the dataset **do**
    Weight: Assign a weight to the feature from 0 to 1 based on its importance.
    Store Weight: Append the weight to the list of weights for the corresponding feature.
  end for
end for
Calculate Correlations: Compute pairwise Spearman correlation coefficients for all features.
for each pair of features **do**
  if correlation coefficient > 0.95 **then**
    Remove: Remove one feature from the high correlated pair.
  end if
end for
for each remaining feature **do**
  Mean Weight: Calculate the mean weight from the list of weights.
end for
Final Rank: Order the features based on their mean weights to obtain the final ranking of feature importance.

C. Data Preprocessing

The absence of outliers was confirmed throughout the dataset by manual inspection. In general, using the Taguchi orthogonal array for the design of experiments reduces the risk of collinearity within the data, and there was no collinearity issue. Data normalization is performed as follows:

$$K'_{ij} = \frac{K_{ij} - \min_j(K)}{\max_j(K) - \min_j(K)} \tag{7}$$

If $K \in \mathbb{R}^{N \times D}$, any particular element of the dataset can be represented as K_{ij}. In this representation, i is taken to be the observation number, which can vary from 1 to N, while j is considered the feature number, ranging from 1 to D.

Four different averaging methods were implemented to calculate the trend in the output specimens of each experiment: conventional, median, trimmed, and averaging after excluding the specimen with the greatest impact on variance among all specimens per experiment. The comparison revealed negligible differences between the methods, leading to the adoption of conventional averaging for this work.

D. General setting for experiments

In all 50 experiments the same mixing tool was used (Fig. 1). The environmental conditions for the storage of materials and for production remained constant, ensuring none was affected by seasonal changes. The experiments draw from a single batch of materials to keep the properties uniform. The old and new cements are of the same type and are delivered by the same company. Modifications to the selected factors were performed within the laboratory environment, contributing to the controlled variation in the dataset. In all experiments, the mixing process was conducted in a closed room at 20°C.

VI. Results

Fig. 3 presents five distinct bar plots. Each plot underscores a facet of our dataset. The initial bar plot demonstrates the average results from all filter methods, showcasing the relative importance of influencing factors. Similarly, the second and third plots articulate the average outcomes derived from wrapper methods and embedded methods, accentuating the significance of factors within these contexts. The fourth bar plot accentuates the average results specific to the Gaussian process regression and the Total Sobol index method. Finally, the last bar plot illustrates the outcomes from the ensemble-averaging method employed. This needs to be considered, filter and wrapper techniques typically prioritize features that either exhibit strong individual relevance or contribute significantly to the performance of a specific model. Conversely, embedded methods adopt a more holistic approach, factoring in intricate feature relationships, which often leads to a broader distribution of importance across features. This can sometimes result in lower perceived importance for individual features, particularly in cases of feature interdependencies or redundancies (Fig. 3). On the other hand, CS, a measure of a material's ability to resist axial load, is closely related to FS. Flexural strength (FS) is a measure of a material's resistance to bending, and both often share similar influencing factors due to their reliance on the material's inherent properties. However, they are not identical because different stress states are involved: uniaxial compression in the former and bending-induced tension and compression in the latter. These different stress states can cause variations in how a feature contributes to strength measurement, resulting in the observed differences in feature importance between the two groups. In Fig. 4, features are ranked based on their importance determined by all base algorithms and also the ensemble-averaging procedure, with Rank 16 representing the most important feature.

Based on the Spearman correlation coefficients, there is a high correlation between the results from the p-value, f-Value, and Partial-Least-Squares (PLS) methods and those from the Forward and Backward Sequential Feature Selection (F/B-SFS) when considering CS. As a result, the f-value and B-SFS results were removed. For FS, only the p-value and f-value are correlated, yet the f-value was removed. Consequently, the ensemble-averaging results (Fig. 3) and ensemble-ranking results (Fig. 4) are based on the remaining base algorithms' outcomes.

VII. Discussion

The comparison of feature importance for predicting CS and FS reveals distinct patterns (Fig. 3). The Storage-conditions-28T feature is paramount for both strengths, highlighting the crucial role of temperature from the 2nd day to the 28th day during curing. The feature Storage-conditions-1T, indicating the temperature on the first day post-mixing, also ranks highly in both lists, emphasizing the importance of early temperature management. Storage-conditions-28C and Storage-conditions-1C, which reflect storage class during curing, significantly impact both CS and FS. This influence suggests that optimizing storage conditions during curing can enhance the final concrete quality even if the initial prediction post-mixing is sub-optimal.

Subsequently, the temperature of raw materials and power consumption (average) during mixing emerge as key factors. Notably, the importance of power consumption (average) exceeds that of mixer speed and mixing time. Therefore, to streamline the process, only power consumption may be considered. The results also indicate that the amount of superplasticizer and graphite are significant. For CS, our analysis shows that the type of fine aggregate (Fine-aggregate-I or Fine-aggregate-II) influences more than the type of its coarse counterpart. This might suggest the vital role of fine aggregate in UHPC's dense microstructure, contributing to its high CS. For FS, the importance of Coarse-aggregate-I outweighs that of Fine-aggregate-I. This indicates the significant role of coarse aggregate in stress distribution within concrete. However, the influence of Coarse-aggregate-II and Fine-aggregate-II is nearly equal and more subdued. Features such as Mixing speed, Mixing duration, and Cement-reactivity-class rank lower on both lists.

Highlighting feature importance in concrete production not only refines experimental designs, aiding the creation of accurate predictive models, but also streamlines the entire process. It directs decisions on materials and operational methods,

TABLE III: The levels indicate the range of values factors can take. Fresh concrete is stored in the climate cabinet at 95 % relative humidity (Storage-conditions-1C = 1) or in air (Storage-conditions-1C = 2) on day one. From Day 2 to Day 28, it's stored at 40 % relative humidity (Storage-conditions-28C = 1) or underwater (Storage-conditions-28C = 2). Two types of aggregates are used - coarse and fine - referred to as I and II within its category.

Factor	Unit	Level 1	Level 2	Level 3	Level 4	Level 5
Cement-reactivity-class	-	1	2	-	-	-
Storage-conditions-28T	°C	20	20	10	30	40
Storage-conditions-28C	-	1	2	2	2	2
Ingredient moisture	kg (%)	3.042 (4 %)	2.925 (0 %)	3.159 (8 %)	3.276 (12 %)	3.364 (15 %)
Coarse-aggregate-I	kg	6.900	6.000	5.400	6.300	5.100
Coarse-aggregate-II	kg	8.925	10.500	11.550	9.975	12.075
Fine-aggregate-I	kg	5.100	6.000	6.600	5.700	6.900
Fine-aggregate-II	kg	0.863	0.750	0.675	0.788	0.638
Superplasticizer	kg	0.290	0.323	0.306	0.355	0.339
Storage-conditions-1T	°C	20	20	10	30	40
Storage-conditions-1C	-	1	2	2	2	2
Mixing speed	rad/s	200	350	500	350	350
Mixing duration	s	300	300	300	210	480
Ingredient temperature	°C	10	20	25	30	40
Graphite	kg	0.045	0.000	0.090	0.135	0.225

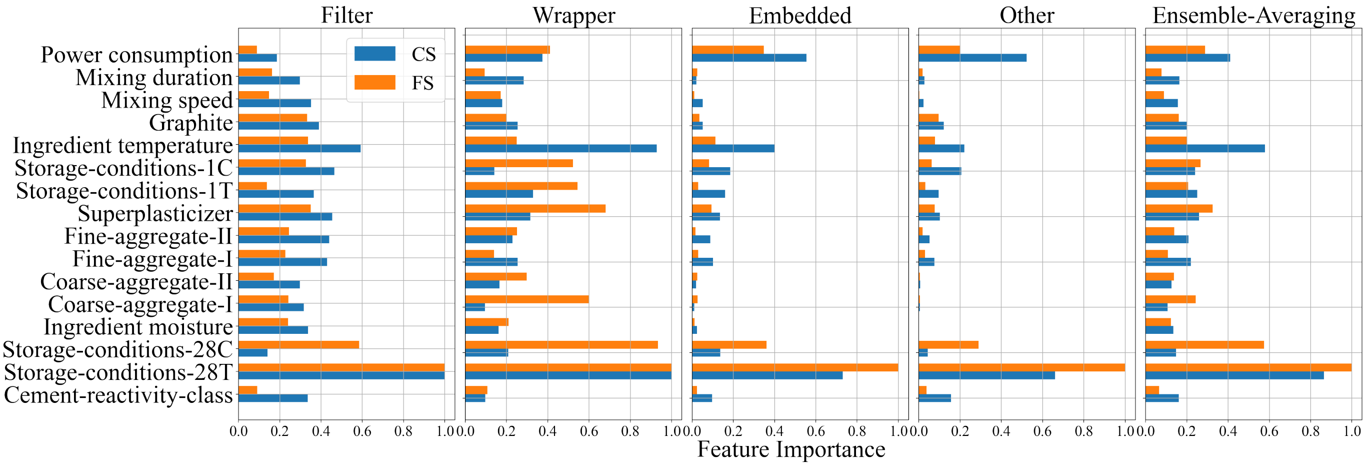

Fig. 3: Feature importance assessment using different methods (CS: compressive strength, FS: flexural strength).

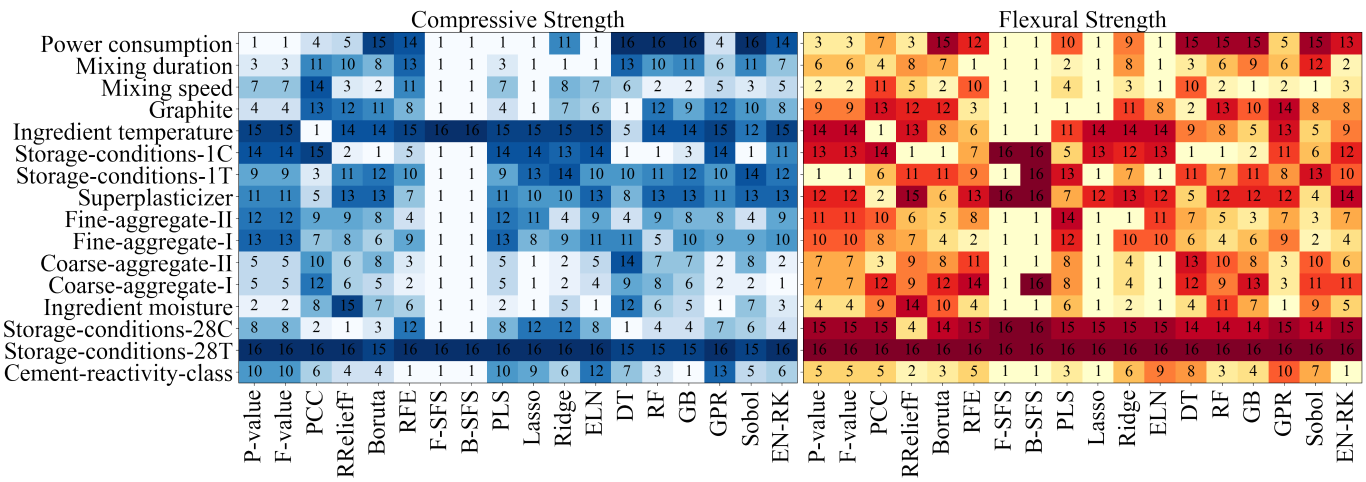

Fig. 4: Feature importance assessment using different methods. Features rank from 1 to 16 (most important), with equal ranks indicating similar importance. (PCC: Partial Correlation Coefficient, RFE: Recursive Feature Elimination, F/B-SFS: Forward/Backward Sequential Feature Selection, PLS: Partial Least Squares, ELN: Elastic Net, DT: Decision Tree, RF: Random Forest, GB: Gradient Boosting, GPR: Gaussian Process Regression, EN-RK: Ensemble-Ranking)

establishes efficient monitoring strategies, and underscores the significance of particular factors. Notably, the findings indicate the potential reduction of certain examined factors, like cement reactivity, mixing speed, duration, and ingredient moisture. This streamlining carries cost benefits and, considering the limited data size available in concrete production, becomes pivotal for effective modeling of the process.

Fig. 4 showcases the diverse aspects of dataset analysis via multiple algorithms. For both CS and FS, RF, GB, and Boruta provided congruent results, especially between RF and Boruta. RreliefF, which is a model-free approach, mirrored RF in determining feature importance. The comparisons of p-value, f-value, and PLS in CS were similar to those in FS; however, PLS varied in FS, suggesting that the p-value alone might be sufficient. Examination of RFE, F-SFS, and B-SFS for CS revealed matching trends. Features identified by F-SFS and B-SFS aligned with those by RFE. However, RFE allocated importance to additional features, a benefit not seen in FS. For CS and FS, regularization-based algorithms (Lasso, Ridge, Elastic Net) showed similar trends. Lasso and Elastic Net nullified irrelevant features, favoring feature selection. In contrast, Ridge weighed all features, refining feature importance.

VIII. Conclusion

In the context of a complex process like concrete production, where high dimensionality, small datasets, and uncertain chemical reactions are the norm, an ensemble structure is particularly beneficial for determining feature importance. This multivariate approach is particularly adept at managing the intricate nature of the concrete production process, capturing more complexity than a single method. In this study, the most relevant influencing factors are considered collectively. Fifty experiments were conducted. The findings underscore the fact that the storage conditions of fresh concrete up to Day 28 are of paramount importance. While the reactivity of the cement, mixing speed, and duration are integral to the mixing process, their impact on the final product's strength seems to be less pronounced than the storage conditions during the curing period. Therefore, they can be excluded from further investigation. This knowledge also enables a refinement of the prediction after the mixing process. If a modification is necessary, it can be obtained post-mixing by adjusting the appropriate settings for curing conditions.

Future work could explore the robustness of these findings across different UHPC formulations. Furthermore, it is planned to conduct additional experiments to construct a model based on the most significant identified influencing factors.

IX. 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 *Ahmed Hooshang Noori*, for cooperating on the implementation of ensemble structure, and *Tino Sablotny* and *Michelle Apholz* from *G.tecz Engineering GmbH* for carrying out the experiments, and joint work on the analysis of results.

References

[1] Z. Li, et al., "Machine learning in concrete science: applications, challenges, and best practices," Npj computational materials, vol. 8, p. 127, 2022.

[2] 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," Materials, vol. 14, p. 4304, 2021.

[3] 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," Workshop Computational intelligence, vol. 1, p. 52, 2022.

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

[5] S. Propovics, "Analysis of concrete strength versus water-cement ratio relationship," Materials Journal vol. 87.5, pp. 517–529, 1990.

[6] 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," mathematics and computers in science and engineering vol. 10, 2008.

[7] I. C. Yeh, "Modeling of strength of high-performance concrete using artificial neural networks," Cement and Concrete research, vol. 28, pp. 1797–1808, 1998.

[8] A. Farzampour, "Compressive behavior of concrete under environmental effects," Compressive Strength of Concrete, pp. 92–104, 2019.

[9] K. Safranek, "Influence of different mixing processes on the strength of ultra-high strength concretes," Diploma Thesis, Vienna University of Technology, repositTUm, 2007.

[10] B. Cazacliu, and N. Roquet, "Concrete mixing kinetics by means of power measurement," Cement and Concrete Research, vol. 39, pp. 182–194, 2009.

[11] J. Yu, Y. Zhao, R. Pan, X. Zhou, and Z. Wei, "Prediction of the Critical Temperature of Superconductors Based on Two-Layer Feature Selection and the Optuna-Stacking Ensemble Learning Model," ACS Omega, vol. 8, pp. 3078–3090, 2023.

[12] K. K. Yun, S. W. Yoon, and D. Won, "Interpretable stock price forecasting model using genetic algorithm-machine learning regressions and best feature subset selection," Expert Systems with Applications, vol. 213, p. 118803, 2023.

[13] J. Ruyssinck, V. A. Huynh-Thu, P. Geurts, T. Dhaene, P. Demeester, and Y. Saeys, "NIMEFI: gene regulatory network inference using multiple ensemble feature importance algorithms," PLoS One, vol. 9.3, p. e92709, 2014.

[14] Y. Saeys, T. Abeel, and Y. Van de Peer, "Robust feature selection using ensemble feature selection techniques," In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part II 19, pp. 313–325, 2008.

[15] B. L. Brown, and S. B. Hendrix, "Partial correlation coefficients," Wiley StatsRef: Statistics Reference Online, 2014.

[16] J. T. Zhang, and X. Liang, "One-way ANOVA for functional data via globalizing the pointwise F-test," Scandinavian Journal of Statistics, vol. 41, pp. 51–71, 2014.

[17] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," Machine learning, vol. 53, pp. 23–69, 2003.

[18] X. W. Chen, and J. C. Jeong, "Enhanced recursive feature elimination," International conference on machine learning and applications (ICMLA 2007), pp. 429–435, 2007.

[19] D. W. Aha, and R. L. Bankert, "A comparative evaluation of sequential feature selection algorithms," Learning from data, pp. 1199–206, 1996.

[20] M. Ehre, I. Papaioannou, and D. Straub, "Global sensitivity analysis in high dimensions with PLS-PCE," Reliability Engineering and System Safety, vol. 198, p. 106861, 2020.

[21] M. B. Kursa, A. Jankowski, and W. R. Rudnicki, "Boruta–a system for feature selection," Fundamenta Informaticae, vol. 101, pp. 271–285, 2010.

[22] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, pp. 267–288, 2018.

[23] R. M. Rifkin and R. A. Lippert, "Notes on regularized least squares," MIT-CSAIL-TR-2007-025, CBCL-268, 2007.

[24] H. Zou, and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 67, pp. 301–320, 2005.

[25] M. Dumont, R. Marée, L. Wehenkel and P. Geurts, "Fast multi-class image annotation with random subwindows and multiple output randomized trees," International conference on computer vision theory and applications (VISAPP), vol. 2, pp.196–203, 2009.

[26] L. Breiman, "Random forests," Machine learning, vol. 45.1, pp. 5–32, 2009.

[27] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Annals of statistics, vol. 29, pp. 1189–1232, 2001.

[28] C. E. Rasmussen, "Gaussian processes in machine learning," Summer school on machine learning, Springer: Berlin, Germany, pp. 63–71, 2003.

[29] M. Lamboni, B. Iooss, A. L. Popelin, and F. Gamboa, "Derivative-based global sensitivity measures: general links with Sobol'indices and numerical tests," Mathematics and computers in simulation, vol. 87, pp. 45–54, 2013.

[30] J. Yu, R. Pan, and Y. Zhao, "High-dimensional, small-sample product quality prediction method based on mic-stacking ensemble learning," Applied Sciences, vol. 12, p. 23, 2022.

[31] O. Sagi, and L. Rokach, "Ensemble learning: a survey," Wiley interdisciplinary reviews: data mining and knowledge discovery, vol. 8, p. e1249, 2018.

[32] G. Brown, J. L. Wyatt, P. Tino, and Y. Bengio, "Managing diversity in regression ensembles," Journal of machine learning research, vol. 6, pp. 1621–1650, 2005.

[33] A. Krogh, and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," Advances in neural information processing systems, vol. 7, 1994.

[34] H. Dutta, "Measuring Diversity in Regression Ensembles," IICAI , vol. 9, p. 17, 2009.