



Knowledge graph-guided data-driven design of ultra-high-performance concrete (UHPC) with interpretability and physicochemical reaction discovery capability

Pengwei Guo, Weina Meng, Yi Bao^{*}

Department of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, United States

ARTICLE INFO

Keywords:

Interpretable artificial intelligence
Machine learning
Knowledge graph
Solid wastes
Physicochemical reactions
Ultra-high-performance concrete

ABSTRACT

Traditional methods for designing concrete materials typically rely on labor-intensive laboratory experiments, resulting in time and cost inefficiencies. Recently, designing concrete using artificial intelligence (AI) methods has shown high efficiency, but existing AI methods often rely solely on data, which can lead to violation with scientific principles and result in models lacking reasoning abilities. To overcome these challenges, this paper presents an interpretable knowledge graph-guided data-driven design approach. By integrating advanced computing techniques with domain knowledge via knowledge graphs, this approach enables the interpretation of data-driven models and uncovers the underlying mechanisms behind predictions. This approach is applied to ultra-high-performance concrete (UHPC) involving complex physicochemical reactions. The domain knowledge about UHPC is imparted using a knowledge graph, and UHPC properties are predicted using a machine learning model considering mixing proportions, processing methods, and physiochemical properties of materials via natural language processing. The results show that the knowledge graph displays crucial design variables and their effects on UHPC properties, aiding in selecting variables for machine learning models and interpreting their results. The prediction accuracy of the machine learning model reached 0.95. The research paves the way for more transparent and scientific AI models for material design and AI-enabled discovery of scientific knowledge.

1. Introduction

Concrete is the most used structural material worldwide. The annual consumption is about 30 billion tons [1]. Although its unit cost and carbon footprint are not as high as steel and aluminum [2], the used volume is substantial, making concrete an important contributor to carbon emissions. Concrete production and utilization cause two environmental problems: (1) The production of concrete depletes natural resources and releases greenhouse gases. In 2017, the consumption of cement for producing concrete was more than 4 billion tons worldwide [3], which contributed to 2.5 gigatons of CO₂ emission, accounting for 7% of the total CO₂ emission [3]. (2) Construction and demolition (C&D) activities generate more than 3 billion tons of waste concrete annually [4,5]. The two problems are mapped to climate change and waste management, which are two grand challenges identified by the National Academy of Engineering (NAE) of the United States [6]. The challenges are exacerbated by the poor infrastructure condition. Restoration of civil infrastructure is another NAE grand challenge. These

challenges converge to the idea of valorizing solid wastes to produce high-performance green concrete for resilient and sustainable infrastructure.

Ultra-high-performance concrete (UHPC) is a family of advanced concrete featuring superior mechanical properties, workability, and durability [7–9]. The compressive strength is 120 MPa or higher at 28 days [8]; UHPC has adapted fresh properties that facilitate the placement and quality control [9,10]; and chopped fibers are used to achieve high crack resistance [11]. These properties make UHPC an unique candidate for structural applications that require high mechanical strengths and long durability, such as the construction and rehabilitation of bridges [12,13], tunnels [14], and high-rise buildings [15]. However, the high cost and high carbon footprint of UHPC have hindered its applications. Recent research has shown that the cost and carbon footprint of UHPC can be reduced by using various solid wastes to partially replace the raw ingredients (e.g., cement or sand) of UHPC [16–23]. Valorizing solid wastes in producing UHPC has trifold benefits: (1) The material cost and carbon footprint are reduced. (2) The volume

^{*} Corresponding author.

E-mail address: yi.bao@stevens.edu (Y. Bao).

of solid wastes for landfill is reduced. (3) UHPC has superior properties for improving the resilience of infrastructure.

However, the development of low-carbon cost-effective UHPC blended with solid wastes is challenging. The traditional methods are performance-based and demand a large number of trial-and-error laboratory tests. The laboratory tests of UHPC are typically conducted via several steps to optimize the mixture design variables step by step [8]. Those tests are costly, labor-intensive, and time-consuming. The evaluation of the mechanical properties of concrete takes 28 days or longer typically [24], because of the hardening kinetics of cement-based materials. When the type or dosage of a raw ingredient is changed, the daunting tests must be replicated, hindering the utilization of solid wastes because the physicochemical properties of wastes often change. For example, the particle size gradation and chemical composition of two batches of fly ash are usually different even if they are produced from the same plant [25,26]. Although many concrete mixtures containing solid wastes have been developed, it is difficult to extend their applications because the concrete properties will be changed when the physicochemical properties of wastes are changed, making it unrealistic to utilize solid wastes to produce concrete in practices.

Under such a circumstance, artificial intelligence (AI) approaches emerged as an alternative method to design concrete [27,28]. The AI approaches are used to design concrete via a prediction-optimization framework: First, machine learning models are trained using prior test data to relate the design variables (e.g., water-to-binder ratio, sand-to-binder ratio, and binder combination) to the material properties (e.g., compressive strength, tensile strength, and porosity) [29]. Second, the machine learning predictive models are integrated with optimization techniques to maximize the mechanical properties while minimizing the material cost and the carbon footprint [27]. Machine learning models have exhibited the ability to predict the fresh and hardened properties of UHPC with high accuracy and high efficiency [30–33]. Recently, auto-tuned machine learning techniques have been developed to enable auto-discovery of low-carbon cost-effective UHPC [28,29].

Although it is promising to utilize AI-assisted approaches to design concrete, limitations have been identified from current approaches: First, most AI-based design utilizes data-driven machine learning models to predict concrete properties, without considering the domain knowledge about concrete. Although data-driven models can achieve high accuracy in training and testing processes, those models cannot reveal the underlying scientific principles. Second, machine learning models are unlikely to be applicable when new ingredients (e.g., solid wastes) are used, because the models are trained using datasets with specific ingredients. New ingredients are unseen to the trained models, creating a major barrier for using AI methods to design UHPC with solid wastes because wastes involve large variations in particle size gradation and chemical composition.

Recently, important efforts have been made to address the above-mentioned limitations. First, attempts have been made to utilize knowledge about concrete to generate or augment the dataset used to train machine learning models. For example, the particle packing theory was referenced to guide the generation of datasets [34], and a micromechanics-based formula was used to generate data that satisfy the micromechanics theory [35]. Second, the physicochemical information of raw materials has been utilized to develop machine learning models that consider the physicochemical properties (e.g., particle size gradation and chemical composition) of raw materials rather than the engineering names (e.g., Class C and Class F fly ash) [29], making it possible to consider various solid wastes using the same machine learning model. However, existing machine learning methods generate black-box models lacking interpretability. Shapley additive explanations (SHAP) were employed to achieve “explainable machine learning”, but the term “explainable” refers to the evaluation of the significance of design variables at a data level [36,37], without explaining any mechanism such as the physical and chemical reactions of concrete. To attain

truly explainable machine learning, it is necessary to incorporate domain knowledge of physicochemical reactions, such as cement hydration, pozzolanic reaction, and generation of calcium silicate hydrate.

This paper presents an approach to achieve interpretable machine learning that explains the underlying physicochemical mechanisms of concrete for the first time. The idea is to incorporate scientific knowledge about concrete by a knowledge graph that explicitly describes the underlying mechanisms. The graph is created by domain experts using the knowledge in current literature and is updatable when new knowledge becomes available in future publications. This paper proposes to utilize the knowledge graph in two ways: (1) Guide the generation of machine learning models. (2) Interpret the prediction results from the machine learning models through an interpretable AI framework that integrates knowledge graph and machine learning techniques. There are two main research objectives: (1) To develop the interpretable machine learning approach. (2) To implement the approach into designing low-carbon cost-effective UHPC. This paper has three main novelties: (1) An interpretable knowledge-guided data-driven approach is presented. By integrating advanced machine learning techniques with domain knowledge via knowledge graphs, this approach enables the interpretation of data-driven models and uncovers the underlying mechanisms behind machine learning-based predictions. (2) A guide for creating knowledge graphs is presented and utilized to generate a knowledge graph for UHPC materials. The utilization method and benefits of the knowledge graph are elaborated. (3) The innovative approach is applied to predict UHPC properties, interpret machine learning results, and discover scientific knowledge.

The remainder of the paper is structured as follows: Section 2 discusses the basic concepts of machine learning-based design methods in comparison with the current design methods. Section 3 outlines the methods developed or applied in this research. Section 4 presents and discusses the results obtained from machine learning and the interpretation of results based on a knowledge graph. Section 5 concludes the new findings and future opportunities.

2. Proposed concept

The concept of machine learning-based design methods has been compared with the concept of current design methods for UHPC and other types of materials, as depicted in Fig. 1. The current methods utilize domain knowledge about UHPC and are mainly based on laboratory experiments [8,38]. The domain knowledge is utilized to design the laboratory experiments and interpret the experimental data, qualitatively [39]. Laboratory experiments are conducted to evaluate the effects of design variables on concerned properties of UHPC, quantitatively. The experimental results are utilized to determine the optimal design and generate new knowledge to enrich domain knowledge.

For example, domain knowledge about UHPC was applied to design laboratory experiments such as the selection of tested mixtures, mixing protocols, curing schemes, and property evaluation methods [8]. With the domain knowledge, the number of candidate mixtures was limited to save experimental efforts while achieving sufficient information to obtain the optimal UHPC mixture.

Laboratory experiments typically generated two types of knowledge, which are (i) the effects of the design variables (e.g., water-to-binder ratio, sand-to-binder ratio, and binder combination) on the concerned properties (e.g., compressive strengths, elastic modulus, and durability), and (ii) the underlying mechanisms of the effects of design variables. For example, increasing the water-to-binder ratio increases the porosity of matrix and thus weakens the compressive strength; and increasing the cement replacement percentage of fly ash reduces the volume of hydration products (e.g., calcium hydroxide) and thus reduces the compressive strengths [7–9].

In current practices of machine learning-based design, machine learning models are used to replace laboratory experiments, without considering the domain knowledge about concrete. In this research, a

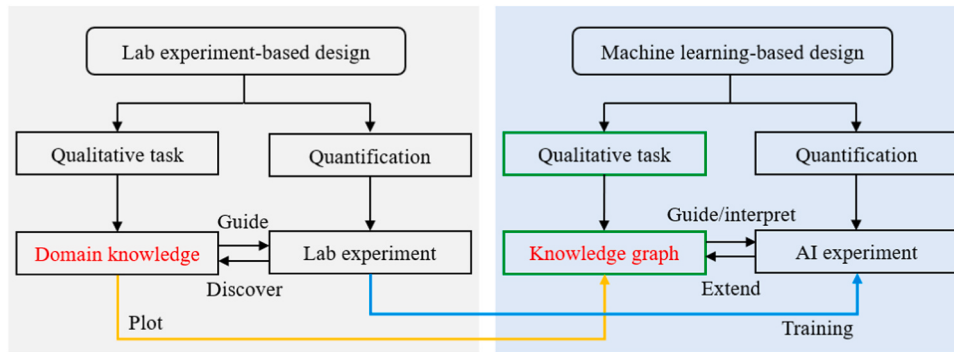


Fig. 1. Concept of knowledge-guided interpretable AI designer versus the conventional methods.

knowledge graph is incorporated into an AI designer, aiming at imparting the domain knowledge into the AI designer. The knowledge graph plays the role of domain knowledge (Fig. 1) and interacts with AI experiments which replace laboratory experiments in two aspects: (1) The knowledge graph is utilized to guide the design of AI experiments and interpret the AI experiment results; and (2) the AI experiment results are utilized to evaluate the knowledge graph. If the results from AI experiments agree with the qualitative trends embedded in the knowledge graph, the knowledge graph will be used to interpret the experiment results. Otherwise, if the AI results are inconsistent with the knowledge graph or beyond its scope, either the knowledge graph will be extended as informed by the machine learning model, or the machine learning model will be corrected. More details are available in Section 4 of this paper.

The machine learning-based design philosophy is interconnected with the experiment-based design philosophy in two aspects: (1) Laboratory experiments generate data that are used to train machine learning models that predict material properties and are used to conduct AI experiments. (2) Domain knowledge is utilized to create knowledge graphs. The methods of creating knowledge graphs using domain knowledge are elaborated in Section 3. In brief, a knowledge graph embodies domain knowledge in a graphic form. The generation of knowledge graphs can be performed using natural language processing technologies that extract knowledge from publications such as books and papers automatically [40,41]. The knowledge graph updated based on AI experiments informs new domain knowledge, as discussed in Section 4. In other words, the integration of knowledge graphs and machine learning creates a new pathway to generating new knowledge.

3. Methods

The explainable machine learning framework includes seven steps, as shown in Fig. 2: (1) Establish a knowledge graph according to existing knowledge. (2) Identify relevant variables such as the mixture design

variables, physiochemical information of raw ingredients, and experimental conditions based on the knowledge graph. The variables can be categorized into numerical data, categorical data, and textual data. (3) Collect data from publications. (4) Establish the dataset. Categorical data are converted into numerical data through one-hot encoding, and textual data are converted into numerical data through word vectorization. (5) Optimize hyperparameters. (6) Train machine learning models using the dataset. (7) Interpret results using the knowledge graph.

3.1. Knowledge graph

A knowledge graph is a semantic network that utilizes a directed labeled graph to describe domain knowledge. A knowledge graph has three components, which are nodes, edges, and labels, as shown in Fig. 3. Any entity can be a node, and nodes are connected using edges (i. e., arrowlines) that define the relationships between nodes. Both the nodes and edges are well defined by labels. A knowledge graph provides an organized way to represent domain knowledge, facilitates data analysis and data retrieval, promotes the discovery of new insights by enabling users to explore the relationships between entities and extract new information based on sophisticated data analysis. Knowledge graphs have been utilized in multiple fields such as natural language processing [42], recommendation systems [43], and knowledge management [44], but knowledge graphs have not been applied to material design.

To create a knowledge graph for UHPC, the first step is to define the design problem. In this research, the design problem is defined as the design of low-carbon cost-effective UHPC blended with solid wastes such as fly ash and slag as supplementary cementitious materials (SCMs). When the laboratory experiment-based methods are used to design UHPC, domain knowledge is required to properly design the laboratory experiments by considering four aspects, which are the available raw ingredients, processing methods, concerned properties,

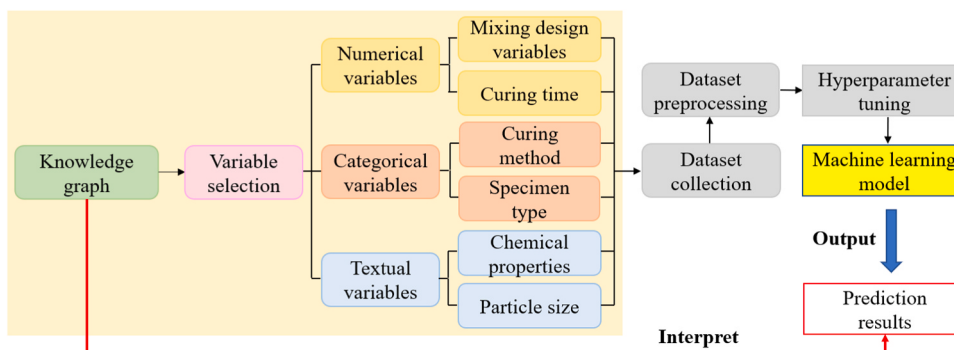


Fig. 2. Explainable machine learning framework for predicting material properties of UHPC.

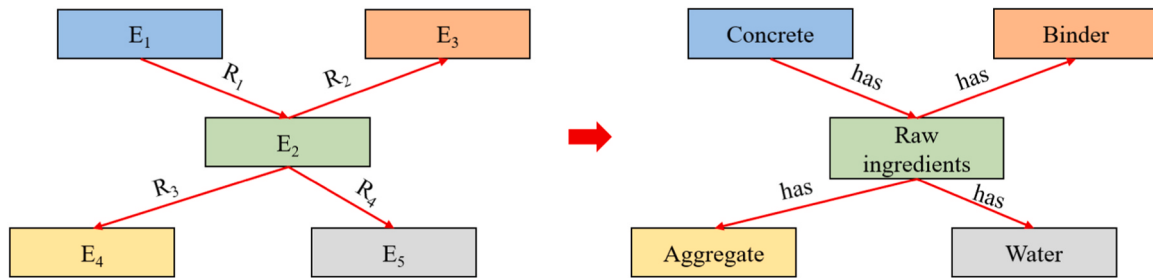


Fig. 3. Example of a knowledge graph. “R” denotes relationship, and “E” denotes entities.

and test mixtures. These four aspects are briefly discussed as follows:

- (1) Raw ingredients include the binders (e.g., cement and SCMs), aggregates, fillers, fibers, water, and admixtures. It is essential to understand the mechanisms of different ingredients that affect the concerned properties of UHPC. For example, partially replacing Portland cement with fly ash can affect the compressive strength of UHPC through influencing the hydraulic reactions and microstructures [45]. Such mechanisms are closely associated with the physicochemical properties of fly ash. Therefore, it is essential to include the knowledge about how the physicochemical properties (e.g., particle size gradation, chemical compositions) of fly ash impact the concerned properties of UHPC.
- (2) Processing methods for UHPC encompass raw ingredient pre-treatment, mixing protocols, curing schemes, and testing methods. For this study, no additional pre-treatment of raw ingredients is necessary, and a common mixing protocol for UHPC is employed. Therefore, the considered factors are mainly the curing scheme and testing methods [46]. It is essential to list the knowledge about how the curing schemes (e.g., steam curing) and testing methods (e.g., specimen type) impact the concerned properties of UHPC.
- (3) The concerned properties often include the fresh and hardened properties. In this research, the focus was placed on the compressive strength of UHPC at 28 days since UHPC has to achieve a compressive strength of 120 MPa or higher at 28 days [8]. In practices, it is also important to consider fresh properties such as the mini-slump spread and other important rheological properties because they play important roles in the mechanical properties and durability of UHPC [47]. More properties can be considered using the same methods.
- (4) Determining the test mixtures involves not only domain knowledge but also experiences. In current practices, the determination of test mixtures often consider three aspects. First, it is beneficial to find existing mixtures of low-carbon cost-effective UHPC blended with solid wastes as the reference mixtures, whose mixture proportionings can be modified to develop new mixtures. Second, existing theories such as particle packing density models can be utilized to provide a general guide for the mixture design [7]. Third, design of experiments tools based on factorial design or statistical design methods can be used [8]. Currently, there is lack of consensus on a unified design method that can be generalized for all cases. Trial-and-error methods are often incorporated into designing UHPC when new design variables are involved because there is lack of similar existing mixtures. The machine learning-based design methods have advantages in selecting test mixtures for two reasons. First, machine learning models are trained using large datasets of existing mixtures that have similarities with the target mixtures. The range of each design variable can be determined based on existing mixtures. Second, machine learning models are time-efficient in conducting AI experiments because the models can predict material properties based on the material design rapidly [48].

The above discussion reveals the scope of knowledge that should be included in the knowledge graph. It must be noted that the relevant knowledge is dependent on and should be tailored to the specific design problem. With the defined scope in this research, a knowledge graph was created by extracting relevant information from publications such as books, papers, and social media [7,26,49–52]. The knowledge graph is shown in Fig. 4. The labeled nodes (i.e., text boxes) show the entities such as “UHPC”, “raw ingredients”, “processing methods”, and “key properties”. The edges (i.e., arrowlines) show the relationships between the entities. For example, the edge between “UHPC” and “Raw ingredients” describes that UHPC has raw ingredients.

The same method can be applied to create new knowledge graphs or extend the knowledge graph by considering other types of ingredients, physicochemical properties (e.g., density and water absorption), and key properties (e.g., flowability and durability), as well as other factors such as the pre-treatment methods for raw ingredients and different mixing protocols of concrete.

3.2. Variable selection and data analysis

The knowledge graph was utilized to guide variable selection by qualitatively identifying the relevant factors that affect UHPC properties. The factors were then utilized as the input variables of the machine learning models. This research classified the input variables into numerical data (e.g., mixture design variables, curing times), textual data (e.g., chemical and physical properties of raw materials), and categorical data (e.g., curing methods, specimen type). The output variable is the compressive strength of UHPC at 28 days. Based on the selected variables, a total of 488 UHPC mixtures were extracted from existing publications. The dataset was divided into a training set (80%) and a testing set (20%). Since there are limited available data, the dataset was not divided to have a validation set, consistent with previous research in references [28,35].

3.2.1. Numerical data

The statistics of the numerical data of the dataset are shown in Table 1 [45,46,49,50,52,55–63]. The numerical data include various mixture design variables such as the cement-to-binder ratio, fly ash-to-binder ratio, silica fume-to-binder ratio, slag-to-binder ratio, superplasticizer-to-binder ratio, water-to-binder ratio, sand-to-binder ratio, fiber volume ratio, and curing ages.

To avoid multicollinearity, a correlation analysis was performed to determine the Pearson correlation coefficient for the input variables of the numerical dataset [53]. The results of the correlation analysis are shown in Fig. 5. The maximum Pearson correlation coefficient of the input variables is 0.64, lower than 0.7, indicating that multicollinearity does not occur. Therefore, it is appropriate to use this dataset to train machine learning models. When the maximum Pearson correlation coefficient is larger than 0.7, multicollinearity will occur, and it will be necessary to modify the dataset to eliminate multicollinearity, as elaborated in reference [28].

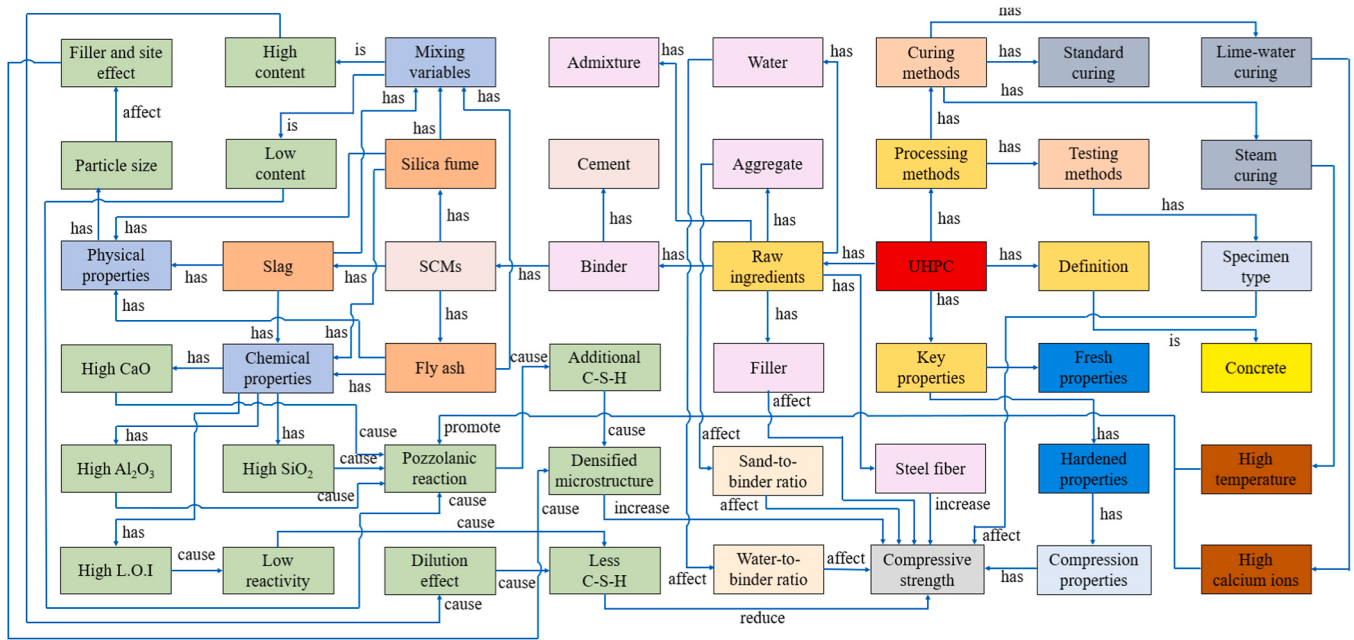


Fig. 4. Knowledge graph created to represent the knowledge of UHPC blended with solid wastes as supplementary cementitious composites. Different colors of text boxes are utilized to show the hierarchical structure of the graph.

Table 1
Description of selected mixture design variables of UHPC.

Number	Variables	Range	Mean	S.D.	Skewness	Kurtosis
1	Cement-to-binder ratio	0.05–1.00	0.65	0.21	-0.50	-0.58
2	Silica fume-to-binder ratio	0–0.25	0.12	0.08	-0.05	-0.74
3	Fly ash-to-binder ratio	0–0.70	0.12	0.14	1.61	2.37
4	Slag-to-binder ratio	0–0.84	0.12	0.19	1.60	1.74
5	Superplasticizer-to-binder ratio	0–0.03	0.01	0.004	0.94	1.22
6	Water-to-binder ratio	0.13–0.23	0.18	0.02	0.74	0.10
7	Sand-to-binder ratio	0.53–1.45	1.11	0.20	0.07	-0.07
8	Fiber volume ratio (%)	0–3	1.23	1.10	0.03	-1.46
9	Curing ages (days)	1–91	29.59	31.48	1.12	-0.03

Note: “S.D.” is the standard deviation. “Skewness” and “kurtosis” describe the shape of a probability distribution. A symmetrical distribution has a skewness of 0. A normal distribution has a kurtosis of 3.

3.2.2. Textual data

Different batches of solid waste usually have different physico-chemical properties although they have the same engineering name (e. g., Class C or Class F fly ash). For example, the particle size gradations and chemical compositions of two batches of Class C fly ash can be quite different [54]. When different materials are used to produce concrete, they have different effects on concrete properties [20]. Thus, it is inappropriate to use the engineering names of wastes to represent the materials in the development of machine learning models without considering the physicochemical properties of ingredients. Recent research has shown that it is essential and useful to consider the physicochemical properties as input variables of the machine learning models for predicting the properties of UHPC blended with various solid wastes [29].

The statistics of important physicochemical properties that affect the compressive strength of UHPC are listed in Table 2, including the types, ranges, mean values, kurtosis, and skewness of the physicochemical properties. The types include the percentages of CaO, SiO₂, and Al₂O₃, as well as the loss of ignition (L.O.I.) and median particle size (D₅₀) [45, 46, 49, 50, 52, 55–63].

The percentages of CaO, SiO₂, and Al₂O₃ in the binder system have significant influences on the hydraulic reactivity and pozzolanic reactivity of the binder system, and affect the compressive strength of UHPC. It is important to note that different SCMs have different chemical

compositions. In addition to chemical composition, some materials such as off-specification fly ash (OSFA) [51] and bottom ash [64] have a high amount of carbon, which results in a high L.O.I. and low reactivity of the materials, therefore compromising the compressive strength of UHPC. A ternary diagram of the chemical compositions of cement, fly ash, slag, and silica fume is shown in Fig. 6.

The particle size gradation is another important factor that can affect the compressive strength of UHPC. Fine particles can fill the pores in cementitious matrix and densify the microstructure, known as the filler effect [65]. The partial replacement of cement by fine mineral cementitious materials can accelerate the rate of hydration reactions, known as the seed effect [66]. The particle size gradation of raw materials can significantly impact UHPC properties.

3.2.3. Categorical data

The compressive strength of UHPC is dependent on the curing methods and specimen type. For example, steam curing has been used to enhance the compressive strength of UHPC through accelerating the hydration reactions of cementitious materials and producing more hydration products [67]. The test results of the compressive strength of UHPC are associated with the size of the specimens. Typically, the compressive strength decreases with the increase of the specimen size [68], known as the size effect. The curing methods and specimen types are shown in Table 3.

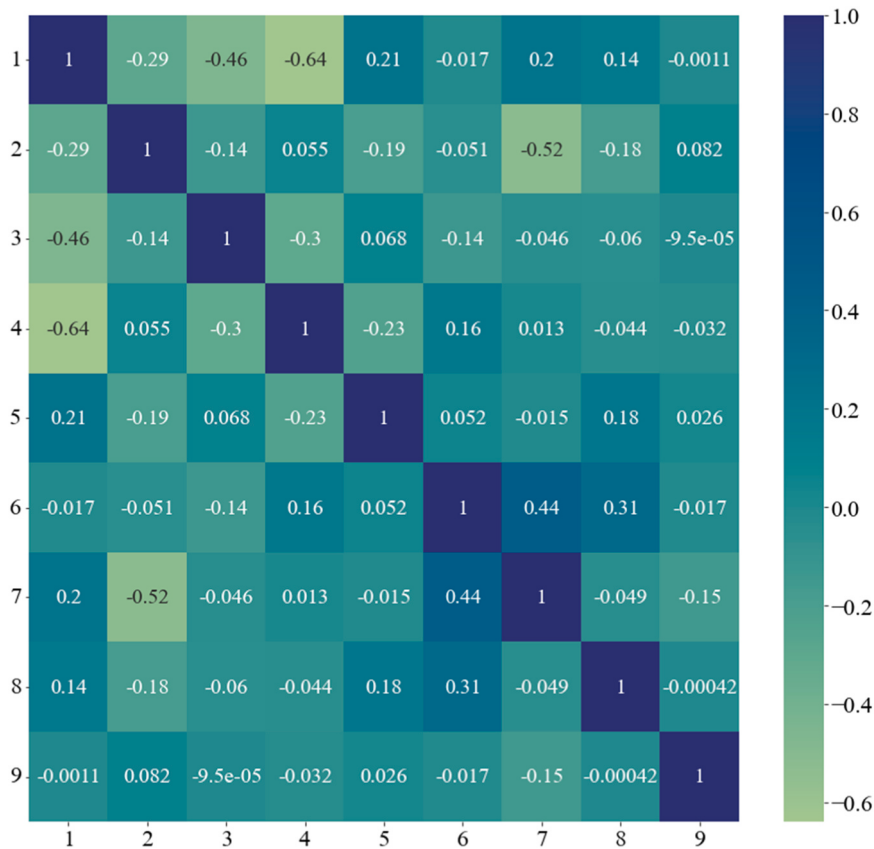


Fig. 5. Correlation matrix of the input numerical data. Numbers 1–9 in the heatmap correspond to the variable numbers in Table 1.

Table 2
Statistical data of physicochemical information of cementitious materials.

Number	Materials	Properties	Range	Mean	S.D.	Skewness	Kurtosis
1	Cement	CaO (%)	56.6–68.7	63.50	2.45	1.58	2.00
		SiO ₂ (%)	17.4–22.4	21.0	1.25	-1.47	2.65
		Al ₂ O ₃ (%)	2.5–8.7	4.72	1.41	1.53	3.30
		L.O.I. (%)	0–2.9	1.59	0.78	-0.18	-0.66
		D ₅₀ (μm)	7.8–46.2	18.8	8.40	2.07	5.72
2	Silica fume	CaO (%)	0.3–1.9	0.73	0.61	1.41	0.19
		SiO ₂ (%)	90.4–97.8	94.2	1.76	-0.07	0.40
		Al ₂ O ₃ (%)	0.1–1.0	0.51	0.29	0.29	-1.13
		L.O.I. (%)	0.3–3.9	1.91	1.29	0.40	-1.09
		D ₅₀ (μm)	0.08–1.8	0.33	0.54	2.69	5.94
3	Fly ash	CaO (%)	1.2–28.1	7.85	8.47	1.75	2.06
		SiO ₂ (%)	16.7–58.0	45.22	10.63	-1.71	3.74
		Al ₂ O ₃ (%)	10.2–38.0	26.17	6.79	-0.69	1.89
		L.O.I. (%)	0.2–49.8	5.95	13.91	3.37	11.55
		D ₅₀ (μm)	2.9–75.0	12.57	18.90	3.51	12.51
4	Slag	CaO (%)	32.4–45.9	38.40	5.18	0.29	-1.57
		SiO ₂ (%)	23.2–45.2	32.07	6.87	1.00	0.42
		Al ₂ O ₃ (%)	9.2–13.9	12.26	1.78	-0.89	-1.00
		L.O.I. (%)	0.4–5.1	1.89	1.58	0.82	-0.68
		D ₅₀ (μm)	0.72–104.0	19.9	26.67	3.00	9.82

Note: “S.D.” is the standard deviation. “Skewness” and “kurtosis” describe the shape of a probability distribution. A symmetrical distribution has a skewness of 0. A normal distribution has a kurtosis of 3.

The considered curing methods include: (1) standard curing, at room temperature and relative humidity higher than 95% [69]; (2) lime-saturated water curing [39]; and (3) steam curing (high temperature, high humidity). Three types of specimens for compressive testing were considered: (1) cubic specimens measuring 51 mm in side length [24]; (2) cubic specimens measuring 100 mm in side length [70]; and (3) prism specimens measuring 160 mm × 40 mm × 40 mm [71].

3.2.4. Output variable

A statistical plot of the compressive strengths of concrete in the dataset is shown in Fig. 7. The plot shows the distribution of the compressive strengths of the considered concrete [72]. With different curing ages, the compressive strengths of different concrete mixtures range from 4.1 MPa to 178.4 MPa, with an average value of 110.3 MPa and a standard deviation of 32.9 MPa. The data points of compressive strength follow the normal distribution approximately.

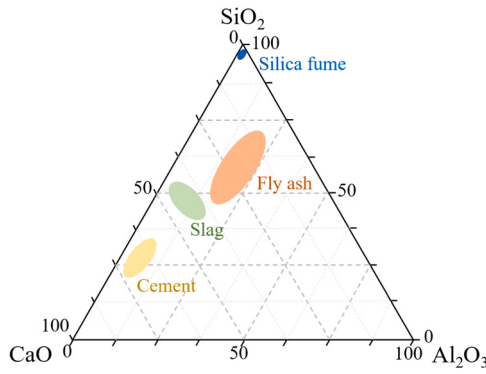


Fig. 6. Ternary diagram (CaO, SiO₂, and Al₂O₃) of Portland cement and typical types of SCMs.

Table 3
Description of selected mixture design variables of UHPC.

Number	1	2	3
Curing method	Standard curing	Lime-water curing	Steam curing
Specimen type	51 × 51 × 51	100 × 100 × 100	160 × 40 × 40

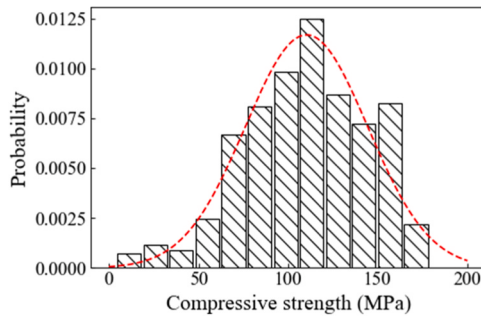


Fig. 7. Statistical analysis of the compressive strengths of concrete mixtures included in the dataset.

3.3. Data preprocessing

3.3.1. One-hot encoding

Categorical data were transformed into numerical data using the one-hot encoding method [73]. The numerical data were then utilized to train the machine learning model. One-hot encoding can improve the accuracy and the generalizability of machine learning models by providing more categorical information. An example of one-hot encoding is shown in Fig. 8.

3.3.2. Word vectorization

Word vectorization is a natural language processing method that maps words to corresponding vectors with real numbers. In this research, the text used to describe the physicochemical properties of

cementitious materials was stored in a dictionary, as shown in Eq. (1). DictVectorizer was used to extract the categorical and numerical features from the dictionary [74]. The categorical features were automatically one-hot encoded, meaning that each unique category was represented by a binary feature, and the numerical features were retained. The categorical and numerical features were turned into sparse matrices.

{‘Type’: ‘Cement’, ‘CaO’: 64.2, ‘SiO₂’: 20.1, ‘Al₂O₃’: 5.1, ‘LOI’: 2.4, ‘D50’: 14.9}(1)

3.4. Machine learning methods

3.4.1. Machine learning algorithms

Machine learning has exhibited outstanding performance in modeling nonlinear relationships between the input variables and the compressive strength of UHPC [28]. Among various machine learning algorithms, ensemble learning algorithms have shown promising performance since they combine the results from different predictive models to achieve an integrated predictive model with high accuracy and high generalizability [27].

CatBoost is an ensemble machine learning method developed based on gradient boosting decision trees, which has a limited number of parameters, supports categorical variables, and provides high accuracy [75]. The employment of CatBoost reduces the need for optimizing many hyperparameters and minimizes the possibility of overfitting, making the machine learning model more generalizable. In addition to CatBoost, other ensemble learning models, such as random forest, LightGBM, and XGBoost, were also considered in this research.

3.4.2. Hyperparameter optimization

Hyperparameter optimization is critically important for the performance of a machine learning model. This research adopted Bayesian optimization to automatically and iteratively optimize the hyperparameters following the Gaussian process [76]. Each iteration was based on a Gaussian function fitted in the previous iteration, aiming to find better hyperparameters compared with the previous iteration. In this research, cross-validation was performed in the optimization process to improve the accuracy and generalizability of machine learning models. Bayesian optimization builds a probabilistic model of the objective function that maps the hyperparameters to the performance of the model [76], as shown in Eq. (2):

$$x_p = \underset{x \in X}{\operatorname{argmax}} f(x) \quad (2)$$

where $f(x)$ is an objective function that needs to be minimized; x_p is the set of hyperparameters that result in the lowest objective function; and X is the search spacing of the hyperparameters. The value of the objective function is evaluated on the testing set. The optimized hyperparameters are listed in Table 4.

3.4.3. Performance metrics

To evaluate the performance of machine learning models, three typical performance metrics were adopted, which are the mean absolute error (MAE), root mean squared error (RMSE), and coefficient of

Curing method	Standard curing	Water curing	Steam curing
Standard curing	1	0	0
Standard curing	1	0	0
Steam curing	0	0	1
Water curing	0	1	0
Steam curing	0	0	1

Fig. 8. Example of converting curing methods into numerical data using one-hot encoding.

Table 4
Hyperparameters of the CatBoost model.

Name	Searching space	Optimal hyperparameters
Iterations	100–2000	1000
Learning rate	0.05–1.0	0.087
Bagging temperature	0.05–7.4	0.636
Border count	10–300	193
Colsample bylevel	0.1–1.0	0.30
Depth	4–20	6
L2_leaf_reg	0.001–7.4	2.33
Random strength	0.001–2.7	0.024

determination (R^2), respectively. The mathematical definitions of the three metrics are shown in Eqs. (3) to (5).

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |P_i - A_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (P_i - A_i)^2} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - \bar{A})^2}{\sum_{i=1}^n (A_i - \bar{A})^2} \quad (5)$$

where n is the total number of data; i is the i^{th} data; P is the predicted compressive strength; A is the actual compressive strength; and $\bar{A}$ is the average value of actual compressive strength.

4. Results and discussion

4.1. Property prediction

With the optimal hyperparameters, machine learning models were trained using the dataset to predict the compressive strength of UHPC. The comparison of experimental results and predicted results from the CatBoost model is shown in Fig. 9, showing that the prediction results from the CatBoost model are consistent with the experimental results. The errors of the prediction results are lower than 15 MPa, which is acceptable to most applications.

The CatBoost model is compared with the other ensemble machine learning models in Table 5. The XGBoost model showed the highest accuracy on the training set. The minimum values of MAE, RMSE, and R^2 were 2.41 MPa, 3.58 MPa, and 0.99, respectively. The CatBoost model achieved the highest accuracy on the testing set. The minimum values of MAE, RMSE, and R^2 were 5.37 MPa, 7.68 MPa, and 0.95, respectively. Fig. 10 shows the correlation between the predicted and actual experimental results of each model on both the training and testing sets.

Fig. 11 shows the Taylor diagram of different machine learning

Table 5
Comparison of CatBoost model with other advanced ensemble learning models.

Model	Dataset	MAE (MPa)	RMSE (MPa)	R^2
Random forest	Training	3.25	4.72	0.98
	Testing	7.20	9.88	0.92
LightGBM	Training	3.80	6.08	0.98
	Testing	7.83	10.58	0.90
XGboost	Training	2.41	3.58	0.99
	Testing	5.97	8.66	0.94
CatBoost	Training	2.91	4.01	0.99
	Testing	5.37	7.68	0.95

models, including multi-layer perception (MLP), support vector regressor (SVR), random forest (RF), LightGBM, XGBoost, and CatBoost. The Taylor diagram uses Pearson correlation (R), root mean square error (RMSE), and standard deviation (SD) to compare the prediction performance and actual observations. The CatBoost model shows the highest accuracy since its prediction result is closest to the reference point, suggesting that the CatBoost model has the best prediction accuracy compared with the other machine learning models on the adopted dataset.

4.2. Knowledge-informed development of sustainable UHPC

The function of the knowledge graph, which is used to interpret the prediction results of the machine learning model, is demonstrated via a case study. The case study presents the feasibility of using the proposed methods to develop novel sustainable UHPC blended with solid wastes.

4.2.1. Raw materials and mixture design

In this case study, various types of SCMs were considered, and, for each type of SCM, the physicochemical properties were varied. For example, different types of fly ash were considered, and they have different particle size gradations and chemical compositions. The different chemical compositions of OSFA and specification-grade fly ash (i.e., Class C and Class F) are shown in Fig. 12. Compared with specification-grade fly ash, OSFA contains a higher amount of carbon, medium levels of silica and aluminum, and lower amount of calcium. The high amount of carbon and low amount of calcium compromise the reactivity of OSFA.

The knowledge graph (Fig. 4) was used to find solutions to mitigate the strength reduction effect of OSFA. Seven factors influencing the compressive strength were identified and plotted in Fig. 13. Slag was used as a SCM to increase the pozzolanic reactivity of the binder system, aiming at increasing the 28-day compressive strength of UHPC [8].

Table 6 lists eight UHPC mixtures investigated in this study. Type I ordinary Portland cement was adopted. OSFA and slag produced from a local plant in New Jersey were considered. Masonry sand with a median size of 525 μm was used as fine aggregate. Straight steel fibers were used

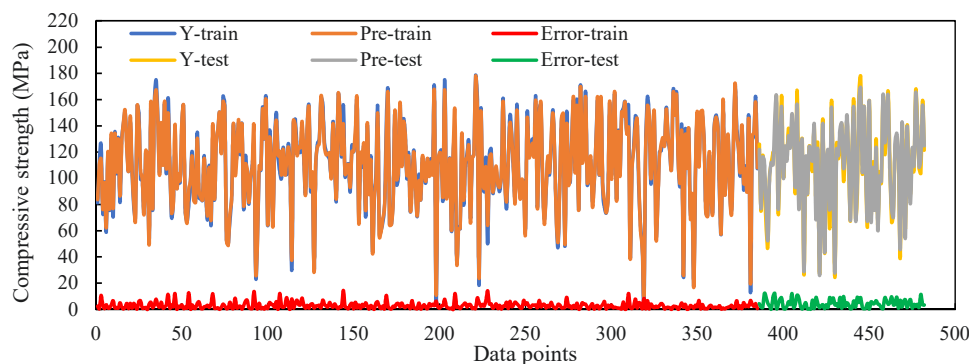


Fig. 9. Prediction performance of the CatBoost model. “Y” represents the experimental results; “Pre” represents the predicted results; and “Error” represents the errors.

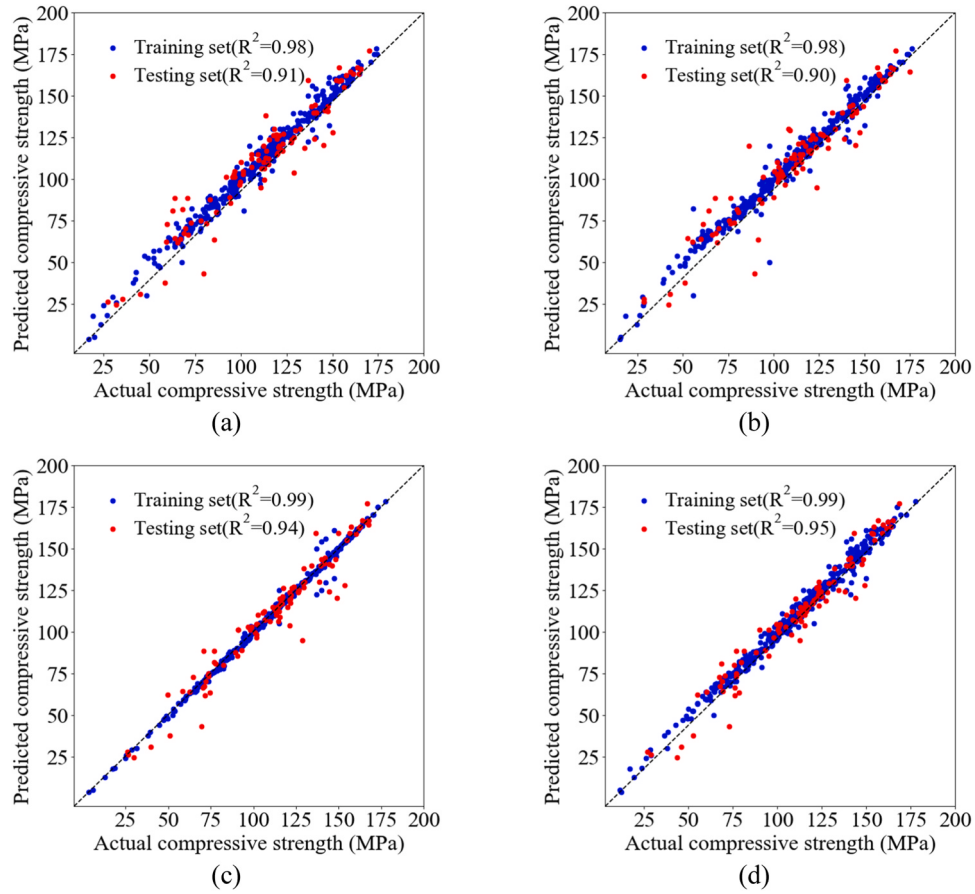


Fig. 10. Comparison of prediction and actual experimental results: (a) random forest, (b) LightGBM, (c) XGBoost, and (d) CatBoost.

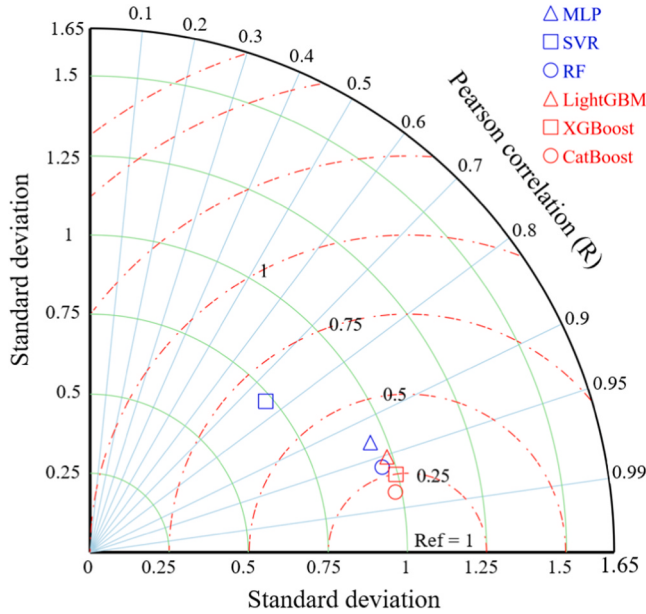


Fig. 11. Taylor diagram used to evaluate the prediction performance of machine learning models.

enhance the crack resistance. The steel fibers were 13 mm in length and 200 μ m in diameter. A polycarboxylate-based superplasticizer with 34.4% solid content and specific gravity of 1.05 was used to improve the flowability. To reduce the economic and environmental impacts, cement

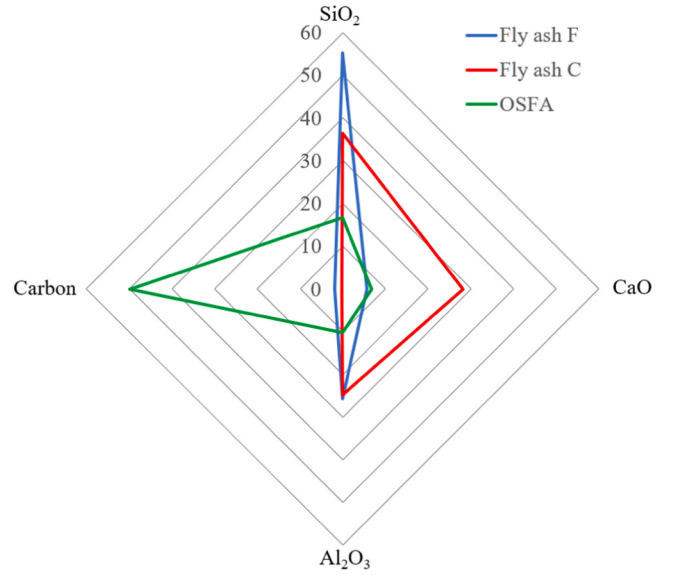


Fig. 12. Comparison of the chemical compositions of different types of fly ash in a radar chart.

was partially replaced by OSFA and slag. For all mixtures, the water-to-binder ratio was 0.23; the sand-to-binder ratio was 1.0; and the steel fiber content was 2% by the volume of the mixture.

The physical and chemical properties of raw ingredients are shown in Table 7. The OSFA had a high carbon content that does not conform to ASTM C618 [77]. To evaluate the pozzolanic reactivity of OSFA, the

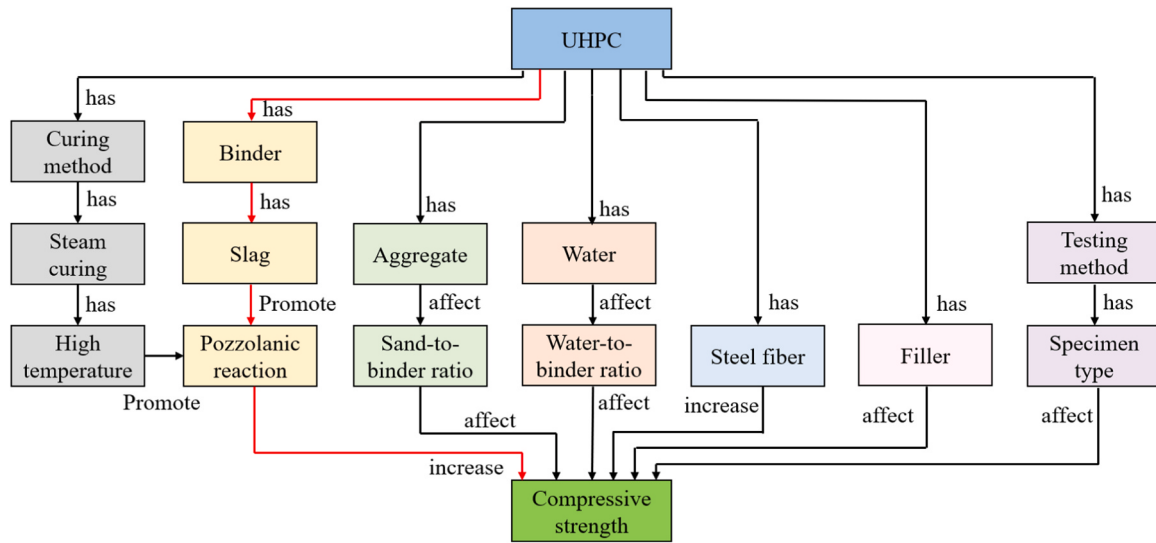


Fig. 13. Potential method proposed to improve the compressive strengths of UHPC mixtures.

Table 6
Mixture proportions of UHPC.

Mixture	Cement	OSFA	Slag	Sand	Superplasticizer	Water	Steel fiber
Control	1.00	0	0	1.00	0.0052	0.23	2
OSFA10	0.90	0.10	0	1.00	0.0051	0.23	2
OSFA20	0.80	0.20	0	1.00	0.0054	0.23	2
OSFA30	0.70	0.30	0	1.00	0.0055	0.23	2
OSFA20SL20	0.60	0.20	0.20	1.00	0.0042	0.23	2
OSFA20SL30	0.50	0.20	0.30	1.00	0.0042	0.23	2
OSFA20SL40	0.40	0.20	0.40	1.00	0.0042	0.23	2
OSFA20SL60	0.20	0.20	0.60	1.00	0.0042	0.23	2

Table 7
Chemical and physical properties of raw materials.

Materials	Cement	OSFA	Slag
CaO	68.1	2.4	43.8
SiO ₂	22.4	16.7	36.2
Al ₂ O ₃	2.7	11.1	10.2
Fe ₂ O ₃	2.2	6.7	0.8
MgO	0.9	0.9	5.1
SO ₃	2.3	3.9	2.2
Na ₂ O	0.2	0.3	0.2
K ₂ O	0.1	1.2	0.4
L.O.I	1.3	49.8	0.7
D ₅₀	13.7	5.8	12.4

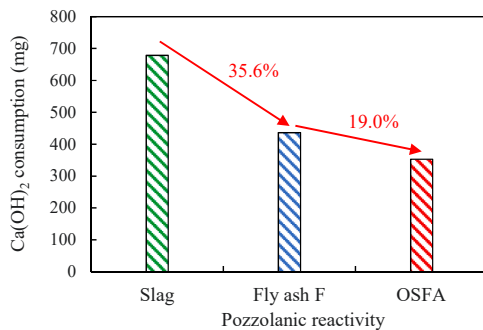


Fig. 14. Comparison of the pozzolanic reactivity results of slag, Class F fly ash, and OSFA.

Chapelle test was performed following NF P18–513 [78].

The reactivity results of OSFA and other SCMs are compared in Fig. 14. The reactivity results of OSFA, Class F fly ash, and slag were 353.1, 436.0, and 678.0 mg Ca(OH)₂/g, respectively, indicating that the reactivity of the OSFA is the lowest among the SCMs.

4.2.2. Machine learning-based prediction

With the methods in Section 2, the CatBoost model was employed to predict the compressive strength of UHPC. A comparison between prediction results and the actual experimental results is shown in Fig. 15, indicating that the machine learning model provides reasonable predictions.

The replacement of cement with up to 30% OSFA decreased the compressive strength. When the binder system contained 20% OSFA, the compressive strength was enhanced with the addition of up to 40% slag as a replacement for cement. However, if the percentage of slag replacement exceeded 40%, the compressive strength decreased.

4.2.3. Interpretation of prediction results

To demonstrate the use of the knowledge graph for the interpretation of the machine learning prediction results, this subsection discusses the underlying mechanisms of the effects of OSFA and slag on the compressive strength. The mechanisms were identified from the knowledge graph (Fig. 4) by the arrows between the nodes automatically. To facilitate the discussion, two graphs (Fig. A1 and Fig. A2) were produced to highlight the red directed arrows that reflect the pathways of mechanisms. In this study, the knowledge graph was manually constructed, and the underlying mechanisms were elucidated by domain experts. In practice, the AI designer can leverage natural language processing to automate the generation of the knowledge graph and identify the pathways graph within it once a raw ingredient and an

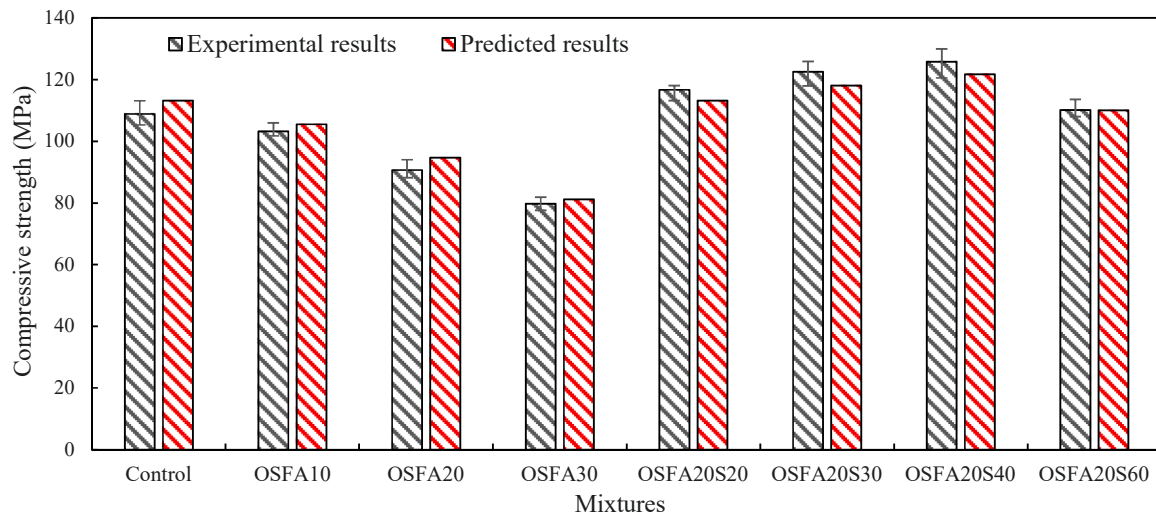


Fig. 15. Comparison of the prediction and the experimental results of UHPC blended with OSFA.

output property are selected [40,41]. Additionally, the AI designer can generate corresponding explanations based on the identified pathways.

(1) Effect of OSFA

According to the machine learning model, when the cement replacement percentage of OSFA increases from 0% to 30%, the 28-day compressive strength of UHPC will decrease from 113.2 MPa to 81.2 MPa. In Fig. A1, the red solid lines indicate the primary factors that affect the compressive strength, and the red dashed lines indicate the factors that potentially influence the compressive strength. The directed arrows identified from the knowledge graph (Fig. A1) explicitly reveal the underlying mechanisms for the reduction of the compressive strength of UHPC when the OSFA content was increased:

- OSFA has a high value of L.O.I., which reduces the reactivity of the binder system when OSFA is blended. The low reactivity reduces the produced amount of hydration products, ultimately decreasing the compressive strength, due to the dilution effect.
- OSFA has fine particle sizes and can act as a filler in the binder system by filling fine pores. In addition, OSFA can serve as the seeds for the precipitation of C-S-H. Both the filler effect and the seed effect of OSFA contribute to increasing the compressive strength.
- OSFA has pozzolanic reactions, although the pozzolanic reactivity of OSFA is relatively low compared with specification-grade fly ash. The pozzolanic reactions tend to enhance the compressive strength of UHPC.

The function of the proposed approach is reflected in three aspects: (i) All the above three effects were identified explicitly from the knowledge graph. (ii) Negative and positive effects exist simultaneously. (iii) The negative dilution effect dominates the results since the carbon content of OSFA is high.

(2) Effect of slag

The reduction of the compressive strength of UHPC due to the use of OSFA was mitigated by using slag in the binder system, as evidenced by mixtures OSFA20S20 and OSFA20S60. When the cement replacement percentage of slag increases from 0% to 40%, the compressive strength will increase from 94.7 MPa to 121.7 MPa. When the replacement percentage increases from 40% to 60%, the compressive strength of UHPC will decrease from 121.7 MPa to 110.1 MPa.

In Fig. A2, the red solid lines show the primary factors that affect the

compressive strength, and the red dashed lines show the factors that potentially influence the compressive strength. The directed arrows identified from the knowledge graph explicitly reveal the underlying mechanisms for the change of the 28-day compressive strength of UHPC when the slag content increased:

- The high pozzolanic reactivity of slag tends to refine the microstructure and increase the compressive strength of UHPC. Slag has amorphous silica which reacts with calcium hydroxide, producing calcium silicate hydrates [79]. Slag has high contents of amorphous silica and calcium hydroxide, thus achieving high pozzolanic reactivity.
- The fine particle size of the slag contributes to densifying the microstructure of UHPC via the filler effect and promoting the hydraulic reactions of cement via the site effect, generating more C-S-H to refine the microstructure and increase the compressive strength.
- The reactivity of slag is lower than the reactivity of cement. When slag is used to replace cement at an excessive percentage, the mechanical strengths will be reduced due to the dilution effect.

Again, the function of the proposed approach is reflected in three aspects: (1) The above three effects were identified explicitly from the knowledge graph. (2) Negative and positive effects exist simultaneously. (3) Whether the compressive strength will increase or not is dependent on the cement replacement percentage of slag.

The case study demonstrates the use of the knowledge graph in interpreting the results from the machine learning models through qualitatively explaining the physicochemical mechanisms. The prediction and interpretation capabilities gained from this research can be utilized to evaluate the effects and underlying mechanisms of other variables, such as the fiber content, water-to-binder ratio, and curing methods, on the compressive strength of UHPC.

In general, the knowledge graph does not include all the knowledge required to interpret the results obtained from the machine learning models. When the results are beyond the scope of the knowledge graph, the knowledge graph provides a reference to support the further development of the knowledge graph. This is discussed in Section 3.3.

4.3. Generation of scientific knowledge

The interactions or coupling effects of the OSFA content and slag content on the compressive strength of UHPC were investigated using the machine learning model, as shown in Fig. 16. When the OSFA content is zero, the compressive strength increases with the increase of the

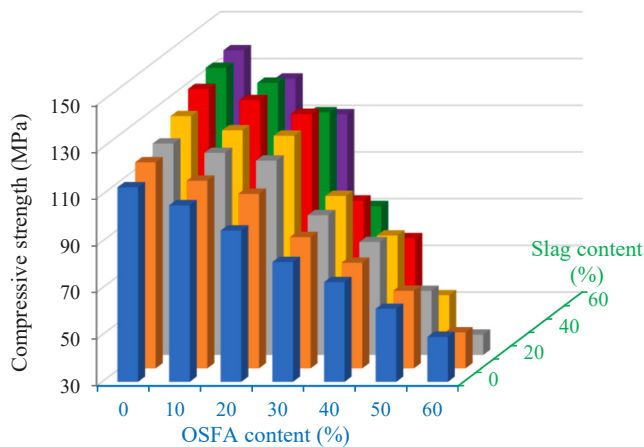


Fig. 16. Interaction of OSFA and slag content on 28-day compressive strength of UHPC.

slag content, which is in the range of 0–60%. When the OSFA content is between 10% and 50%, for each OSFA content, the compressive strength first increases and then decreases with the increase of the slag content. When the OSFA content reaches 60%, the compressive strength decreases with the increase of the slag content monotonically.

Interpretation of such interactive effects are not directly available in the knowledge graph in Fig. 4, therefore informing the need for extending the knowledge graph, aiming to interpret the interactive effects. The process of identifying and addressing knowledge gaps using the machine learning results helps generate new knowledge, as shown in Fig. 17. In this process, the knowledge graph serves as a reference which is used to identify knowledge gaps. The identified knowledge gaps are utilized to extend the knowledge graph with knowledge that is either available in literature unseen to the knowledge graph or generated through new experiments.

In this research, the knowledge graph can be extended to embody the interactive effects based on available literature [80], which is assumed to be unseen during the generation of the knowledge graph in Fig. 4. Although the knowledge graph in Fig. 4 does not provide the interpretation of the interactive effects of OSFA and slag directly, it indeed provides useful hints for extending the knowledge graph: (1) It explicitly shows that the OSFA has a dilution effect on the compressive strength, and the dilution effect dominates in the concerned system. (2) It explicitly shows that the slag has lower reactivity than the cement, also having a dilution effect. The two mechanisms imply that when OSFA is used, the dilution effects of OSFA and slag will likely be combined and thus interact with each other. The implied content is supported by reference [80]. When the cement content was low, the calcium hydroxide produced by the hydration of cement would be insufficient for the dissolution of siloxane (Si-O-Si) bonds in SCMs and the production of C-S-H [80], thus compromising the compressive strength.

The above investigations reflect the interconnection between the machine learning model and the knowledge graph. First, a knowledge graph of UHPC was created based on available literature and the

concerned problem. The knowledge graph was then utilized to guide the establishment of the machine learning model for predicting the compressive strength of UHPC. Next, the machine learning model was utilized to predict the UHPC property, considering various variables. On one hand, the prediction capability can be integrated with an optimization method to discover the optimal UHPC mixtures [81]. On the other hand, the prediction results from the machine learning model can be interpreted using the knowledge graph and utilized to assess the knowledge graph. In this research, the individual effects of OSFA and slag were directly interpreted by the knowledge graph, and the interactive effects of OSFA and slag as indicated by the machine learning model informed the need for extending the knowledge graph. The incorporation of knowledge graph also enables the identification of machine learning prediction results that are inconsistent with existing theories and knowledge, which will then guide further research to examine the machine learning model and the existing theory. The knowledge graph-based interpretable AI designer proposed in this paper offers an alternative way to efficiently design materials and generate knowledge.

5. Conclusions

This paper presents an idea of incorporating domain knowledge into machine learning-based design of concrete for knowledge-guided interpretable AI designer for sustainable concrete. The knowledge graph-based interpretable AI designer offers a new path to generating knowledge about concrete. The AI designer has been implemented into designing low-carbon cost-effective UHPC utilizing solid wastes through a case study. This paper demonstrates the use of a knowledge graph-based interpretable AI designer in interpreting the machine learning prediction results and generating knowledge about the effects of individual variables and their interactive effects. Based on the above investigations, the following conclusions can be drawn:

- A knowledge graph is an effective way to explicitly describe the domain knowledge about concrete in a computer-understandable manner. Knowledge graphs can be integrated with machine learning models to enable quantitative predictions and qualitative interpretations, simultaneously and seamlessly. On one hand, knowledge graphs guide the establishment of machine learning models and interpret machine learning prediction results. On the other hand, machine learning models provide results to evaluate and develop knowledge graphs. Inconsistency between machine learning prediction results and knowledge graphs informs the need for generating knowledge and improving the machine learning models. The knowledge graph-based interpretable AI designer offers an alternative way to developing new materials and generating new knowledge efficiently.
- Knowledge graphs explicitly show the key design variables of UHPC, which facilitates the variable selection task in the establishment of machine learning models for predicting UHPC properties. The identified variables were utilized as input variables to develop a high-fidelity machine learning model that is not only driven by the

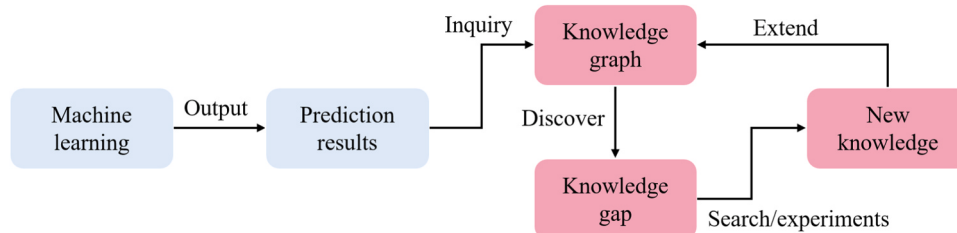


Fig. 17. Extending knowledge graph based on knowledge gap identified from prediction results.

collected dataset but also complaint with the domain knowledge about concrete.

- Knowledge graphs are easier to understand and can play a crucial role in satisfying engineers' needs to explain machine learning results by providing a structured representation of domain knowledge. Unlike traditional method (e.g., SHAP analysis), knowledge graphs facilitate interpretability by establishing relationships between different entities and concepts. Engineers can trace back the reasoning behind a prediction by following the connections within the knowledge graph.
- The proposed machine learning model considers the physicochemical information of raw ingredients and different experimental conditions. The physicochemical information of raw ingredients enables the machine learning model to consider various types of wastes with different physicochemical information, different from the machine learning models that use engineering names to represent the types of ingredients. The consideration of the physicochemical information of raw ingredients largely enhances the generalizability of the machine learning model for designing UHPC with various types of solid wastes. The predictive accuracy (R^2) of compressive strength exceeds 0.95 on the testing dataset.
- Feature engineering techniques such as one-hot encoding and word vectorization serve to transform categorical and textual data, respectively, making them suitable for machine learning applications. This is also essential for leveraging the physiochemical properties of raw materials for predicting the properties of UHPC.

As the first attempt to incorporate a knowledge graph into an AI designer, this paper shows the feasibility of integrating domain knowledge with data-driven machine learning methods and creates new opportunities for future research. Some future research opportunities are listed:

- The knowledge graph developed in this research is limited to a narrow scope for the design of UHPC mixtures with limited types of raw ingredients. It is interesting to develop other knowledge graphs to consider more types of raw ingredients including various wastes and comprehensively evaluate the performance of the proposed approach in a large scope of material research.
- The development of the idea of integrating knowledge graph with machine learning was based on concrete in this research. It is

envisioned that the idea can be extended to other disciplines such as the other areas of civil engineering (e.g., structural and geotechnical engineering), biomedical science and engineering, and material science and engineering.

- Automatic generation and update of knowledge graph plays an important role in practices. It is important to develop effective methods to automate the extraction of knowledge from available literature. It is promising to develop natural language processing techniques for this purpose.

CRediT authorship contribution statement

Pengwei Guo: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Yi Bao:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Weina Meng:** Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Weina Meng received funding from the National Science Foundation of the United States. Yi Bao received funding from United States Army Corps of Engineers.

Data Availability

Data will be made available on request.

Acknowledgement

This research was funded by the National Science Foundation of the United States [award number: CMMI-2046407] and the Army Corps of Engineers through Industrial Base Resilience Initiative [award number: W91278-16-D-0007].

Appendix

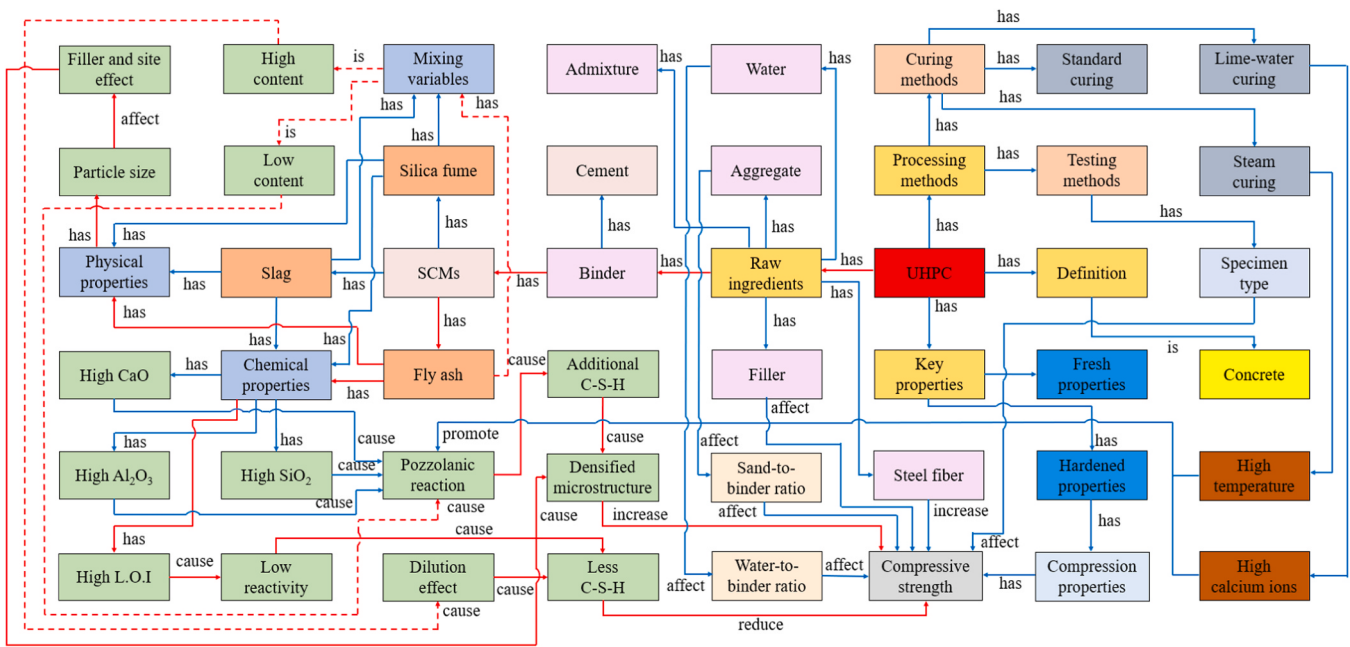


Fig. A1. Effects of OSFA on the compressive strength of UHPC. The red lines in the graph indicate the primary factors that affect compressive strength. The red dashed lines indicate potential factors that may influence the compressive strength. Different colors of text boxes are utilized to show the hierarchical structure of the graph.

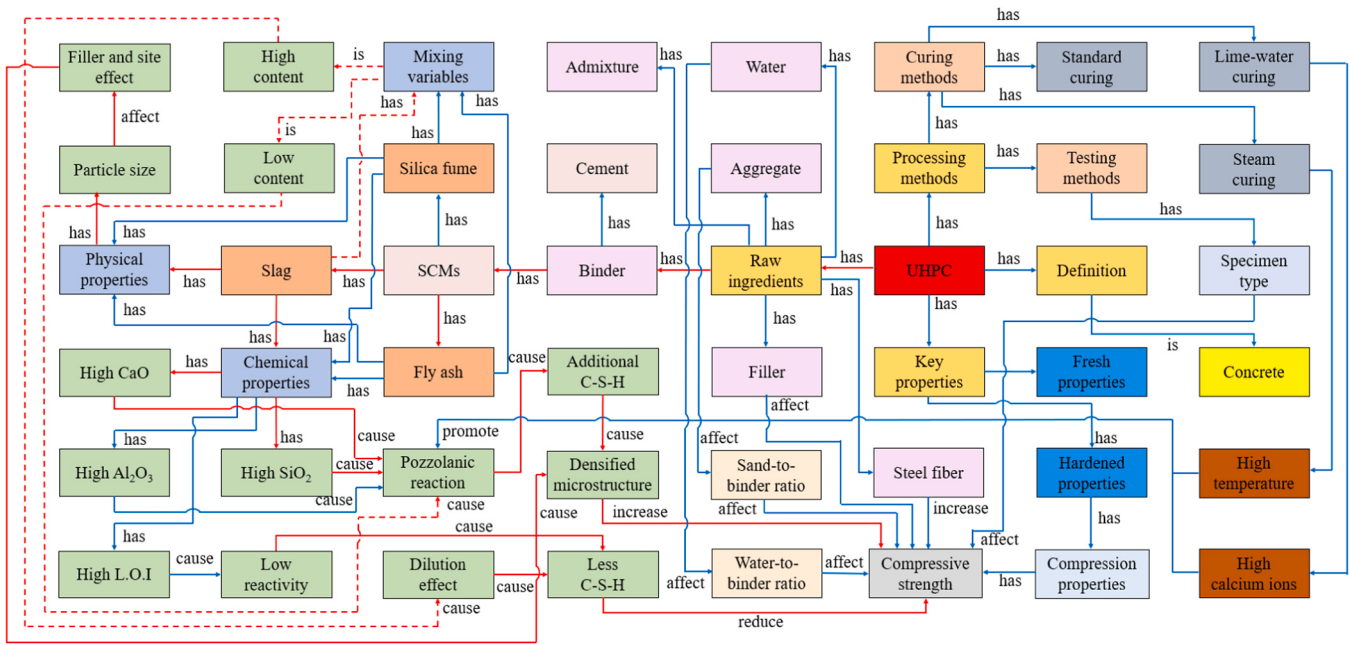


Fig. A2. Effects of slag on the compressive strength of UHPC. The red lines in the graph indicate the primary factors that affect compressive strength. The red dashed lines indicate potential factors that may influence the compressive strength. Different colors of text boxes are utilized to show the hierarchical structure of the graph.

References

- ## References
- [1] P.J. Monteiro, S.A. Miller, A. Horvath, Towards sustainable concrete, *Nat. Mater.* 16 (7) (2017) 698–699, <https://doi.org/10.1038/nmat4930>.
 - [2] G. Kilgore, Carbon Footprint of Building Materials. 2023. (<https://8billiontrees.com/carbon-offsets-credits/carbon-footprint-of-building-materials>).
 - [3] Unsustainable: concrete and cement. 2022. (<https://2150-vc.medium.com>).
 - [4] X. Li, X. Lv, X. Zhou, W. Meng, Y. Bao, Upcycling of waste concrete in eco-friendly strain-hardening cementitious composites: mixture design, structural performance, and life-cycle assessment, *J. Clean. Prod.* 330 (2022) 129911, <https://doi.org/10.1016/j.jclepro.2021.129911>.
 - [5] A. Akhtar, A.K. Sarmah, Construction and demolition waste generation and properties of recycled aggregate concrete: a global perspective, *J. Clean. Prod.* 186 (2018) 262–281, <https://doi.org/10.1016/j.jclepro.2018.03.085>.
 - [6] N.A.E. Grand Challenges for Engineering. (<http://www.engineeringchallenges.org/challenges.aspx>).
 - [7] J. Du, W. Meng, K.H. Khayat, Y. Bao, P. Guo, Z. Lyu, A. Abu-obeidah, H. Nassif, H. Wang, New development of ultra-high-performance concrete (UHPC), *Compos. Part B: Eng.* 224 (2021) 109220, <https://doi.org/10.1016/j.compositesb.2021.109220>.
 - [8] W. Meng, M. Valipour, K.H. Khayat, Optimization and performance of cost-effective ultra-high performance concrete, *Mater. Struct.* 50 (1) (2017) 1–16, <https://doi.org/10.1617/s11527-016-0896-3>.
 - [9] L. Teng, W. Meng, K.H. Khayat, Rheology control of ultra-high-performance concrete made with different fiber contents, *Cem. Concr. Res.* 138 (2020) 106222, <https://doi.org/10.1016/j.cemconres.2020.106222>.

- [10] W. Meng, K.H. Khayat, Improving flexural performance of ultra-high-performance concrete by rheology control of suspending mortar, *Compos. Part B Eng.* 117 (2017) 26–34, <https://doi.org/10.1016/j.compositesb.2017.02.019>.
- [11] P. Guo, W. Meng, J. Du, B. Han, Y. Bao, Lightweight ultra-high-performance concrete (UHPC) with expanded glass aggregate: development, characterization, and life-cycle assessment, *Constr. Build. Mater.* 371 (2023) 130441, <https://doi.org/10.1016/j.conbuildmat.2023.130441>.
- [12] J. Du, P. Guo, Z. Liu, W. Meng, Highly thixotropic ultra-high-performance concrete (UHPC) as an overlay, *Constr. Build. Mater.* 366 (2023) 130130, <https://doi.org/10.1016/j.conbuildmat.2022.130130>.
- [13] Y. Bu, M. Li, C. Wei, Z. Cheng, C. Cui, Y. Bao, Experimental and analytical studies on flexural behavior of composite bridge decks with orthotropic steel deck and ultra-high-performance concrete (UHPC) slab under negative moment, *Eng. Struct.* 274 (2023) 115190, <https://doi.org/10.1016/j.engstruct.2022.115190>.
- [14] Z. Wang, J. Fei, W. Ma, X. Chen, Full-scale loading experiments on performance of UHPC joints for prefabricated mountain tunnel, *Tunn. Undergr. Space Technol.* 131 (2023) 104784, <https://doi.org/10.1016/j.tust.2022.104784>.
- [15] N.V. Tuan, P.H. Hanh, L.T. Thanh, M.N. Soutos, and C.I. Goodier, Ultra high performance concrete using waste materials for high-rise buildings. 2010. (<https://hdl.handle.net/2134/9773>).
- [16] Z.-h. He, M.-l. Shen, J.-y. Shi, Ç. Yalçinkaya, S.-g. Du, Q. Yuan, Recycling coral waste into eco-friendly UHPC: Mechanical strength, microstructure, and environmental benefits, *Sci. Total Environ.* 836 (2022) 155424, <https://doi.org/10.1016/j.scitotenv.2022.155424>.
- [17] X. Wang, R. Yu, Z. Shui, Q. Song, Z. Liu, Z. Liu, S. Wu, Optimized treatment of recycled construction and demolition waste in developing sustainable ultra-high performance concrete, *J. Clean. Prod.* 221 (2019) 805–816, <https://doi.org/10.1016/j.jclepro.2019.02.201>.
- [18] L. Yu, L. Huang, H. Ding, Rheological and mechanical properties of ultra-high-performance concrete containing fine recycled concrete aggregates, *Materials* 12 (22) (2019) 3717, <https://doi.org/10.3390/ma12223717>.
- [19] L. Yang, C. Shi, Z. Wu, Mitigation techniques for autogenous shrinkage of ultra-high-performance concrete—a review, *Compos. Part B Eng.* 178 (2019) 107456, <https://doi.org/10.1016/j.compositesb.2019.107456>.
- [20] N. Soliman, A. Tagnit-Hamou, Partial substitution of silica fume with fine glass powder in UHPC: Filling the micro gap, *Constr. Build. Mater.* 139 (2017) 374–383, <https://doi.org/10.1016/j.conbuildmat.2017.02.084>.
- [21] S.-H. Kang, S.-G. Hong, J. Moon, The use of rice husk ash as reactive filler in ultra-high performance concrete, *Cem. Concr. Res.* 115 (2019) 389–400, <https://doi.org/10.1016/j.cemconres.2018.09.004>.
- [22] S. Qaidi, Y. Al-Kamaki, I. Hakeem, A.F. Dulaimi, Y. Özkılıç, M. Sabri, V. Sergeev, Investigation of the physical-mechanical properties and durability of high-strength concrete with recycled PET as a partial replacement for fine aggregates, *Front. Mater.* 10 (2023) 1101146, <https://doi.org/10.3389/fmats.2023.1101146>.
- [23] Q. Chang, L. Liu, M.U. Farooqi, B. Thomas, Y.O. Özkılıç, Data-driven based estimation of waste-derived ceramic concrete from experimental results with its environmental assessment, *J. Mater. Res. Technol.* 24 (2023) 6348–6368, <https://doi.org/10.1016/j.jmrt.2023.04.223>.
- [24] ASTM C109/C109M-20, Standard Test Method for Compressive Strength of Hydraulic Cement Mortars. 2020. DOI: (https://doi.org/10.1520/C0109_C0109M-20).
- [25] M.A. Bahedh, M.S. Jaafar, Ultra high-performance concrete utilizing fly ash as cement replacement under autoclaving technique, *Case Stud. Constr. Mater.* 9 (2018) e00202, <https://doi.org/10.1016/j.cscm.2018.e00202>.
- [26] P.S. Dong, N.V. Tuan, L.T. Thanh, N.C. Thang, V.H. Cu, J.-H. Mun, Compressive strength development of high-volume fly ash ultra-high-performance concrete under heat curing condition with time, *Appl. Sci.* 10 (20) (2020) 7107, <https://doi.org/10.3390/app10207107>.
- [27] S. Mahjoubi, R. Barhemat, W. Meng, Y. Bao, AI-guided auto-discovery of low-carbon cost-effective ultra-high-performance concrete (UHPC), *Resour. Conserv. Recycl.* 189 (2023) 106741, <https://doi.org/10.1016/j.resconrec.2022.106741>.
- [28] S. Mahjoubi, W. Meng, Y. Bao, Auto-tune learning framework for prediction of flowability, mechanical properties, and porosity of ultra-high-performance concrete (UHPC), *Appl. Soft Comput.* 115 (2022) 108182, <https://doi.org/10.1016/j.asoc.2021.108182>.
- [29] S. Mahjoubi, R. Barhemat, W. Meng, Y. Bao, Deep learning from physicochemical information of concrete with an artificial language for property prediction and reaction discovery, *Resour., Conserv., Recycl.* 190 (2023) 106870, <https://doi.org/10.1016/j.resconrec.2023.106870>.
- [30] C. Tavares, Z. Grasley, Machine learning-based mix design tools to minimize carbon footprint and cost of UHPC. Part 2: cost and eco-efficiency density diagrams, *Clean. Mater.* (2022) 100094.
- [31] E. Sadrossadat, H. Basarir, A. Karrech, M. Elchalakani, Multi-objective mixture design and optimisation of steel fiber reinforced UHPC using machine learning algorithms and metaheuristics, *Eng. Comput.* (2021) 1–14, <https://doi.org/10.1007/s00366-021-01403-w>.
- [32] T.G. Wakjira, A.A. Kutty, M.S. Alam, 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 (2024) 135114, <https://doi.org/10.1016/j.conbuildmat.2024.135114>.
- [33] P. Guo, S. Mahjoubi, K. Liu, W. Meng, Y. Bao, Self-updatable AI-assisted design of low-carbon cost-effective ultra-high-performance concrete (UHPC), *Case Stud. Constr. Mater.* 19 (2023) e02625, (<https://doi.org/10.1016/j.cscm.2023.e02625>).
- [34] D. Fan, R. Yu, S. Fu, L. Yue, C. Wu, Z. Shui, K. Liu, Q. Song, M. Sun, C. Jiang, Precise design and characteristics prediction of Ultra-High Performance Concrete (UHPC) based on artificial intelligence techniques, *Cem. Concr. Compos.* 122 (2021) 104171, <https://doi.org/10.1016/j.cemconcomp.2021.104171>.
- [35] P. Guo, W. Meng, M. Xu, V.C. Li, Y. Bao, Predicting mechanical properties of high-performance fiber-reinforced cementitious composites by integrating micromechanics and machine learning, *Materials* 14 (12) (2021) 3143, <https://doi.org/10.3390/ma14123143>.
- [36] Y. Huang, Z. Huo, G. Ma, L. Zhang, F. Wang, J. Zhang, Multi-objective optimization of fly ash-slag based geopolymer considering strength, cost and CO₂ emission: a new framework based on tree-based ensemble models and NSGA-II, *J. Build. Eng.* 68 (2023) 106070, <https://doi.org/10.1016/j.jobe.2023.106070>.
- [37] A.A. Alabdullah, M. Iqbal, M. Zahid, K. Khan, M.N. Amin, F.E. Jalal, Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis, *Constr. Build. Mater.* 345 (2022) 128296, <https://doi.org/10.1016/j.conbuildmat.2022.128296>.
- [38] W. Meng, K.H. Khayat, Effect of hybrid fibers on fresh properties, mechanical properties, and autogenous shrinkage of cost-effective UHPC, *Acids Mater. J.* 30 (4) (2018) 04018030, [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002212](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002212).
- [39] W. Meng, K. Khayat, Effects of saturated lightweight sand content on key characteristics of ultra-high-performance concrete, *Cem. Concr. Res.* 101 (2017) 46–54, <https://doi.org/10.1016/j.cemconres.2017.08.018>.
- [40] D. Dessi, F. Osborne, D.R. Recupero, D. Buscaldi, E. Motta, Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain, *Future Gener. Comput. Syst.* 116 (2021) 253–264, <https://doi.org/10.1016/j.future.2020.10.026>.
- [41] N. Kertkeidkachorn, R. Ichise, An automatic knowledge graph creation framework from natural language text, *IEICE Trans. Inf. Syst.* 101 (1) (2018) 90–98, <https://doi.org/10.1587/transinf.2017SWP0006>.
- [42] X. Han, Y. Ning, Text-enhanced multi-granularity temporal graph learning for event prediction, 2022 IEEE Int. Conf. Data Min. (ICDM) (2022) 171–180, <https://doi.org/10.1109/ICDM54844.2022.00027>.
- [43] K. Wu, J. Erickson, W.H. Wang, Y. Ning, Equipping recommender systems with individual fairness via second-order proximity embedding, 2022 IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min. (ASONAM) (2022) 171–175, <https://doi.org/10.1109/ASONAM55673.2022.10068703>.
- [44] J.M. Gomez-Perez, J.Z. Pan, G. Vetere, H. Wu, Enterprise Knowledge Graph: An introduction, Springer, 2017, pp. 1–14. (https://link.springer.com/chapter/10.1007/978-3-319-45654-6_1).
- [45] R. Jing, Y. Liu, P. Yan, Uncovering the effect of fly ash cenospheres on the macroscopic properties and microstructure of ultra high-performance concrete (UHPC), *Constr. Build. Mater.* 286 (2021) 122977, <https://doi.org/10.1016/j.conbuildmat.2021.122977>.
- [46] T. Chen, X. Gao, M. Ren, Effects of autoclave curing and fly ash on mechanical properties of ultra-high performance concrete, *Constr. Build. Mater.* 158 (2018) 864–872, <https://doi.org/10.1016/j.conbuildmat.2017.10.074>.
- [47] J. Du, P. Guo, W. Meng, Effect of water-based nanoclay and ambient temperature on rheological properties of UHPC pastes, *Constr. Build. Mater.* 370 (2023) 130733, <https://doi.org/10.1016/j.conbuildmat.2023.130733>.
- [48] S.C. Nayak, S.K. Nayak, S.K. Panda, Assessing compressive strength of concrete with extreme learning machine, *J. Soft Comput. Civ. Eng.* 5 (2) (2021) 68–85. (<https://doi.org/10.22115/SCCE.2021.286525.1320>).
- [49] T. Ahmed, M. Elchalakani, A. Karrech, M. Dong, M. Mohamed Ali, H. Yang, ECO-UHPC with high-volume class-F fly ash: new insight into mechanical and durability properties, *J. Mater. Civ. Eng.* 33 (7) (2021) 04021174, [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0003726](https://doi.org/10.1061/(ASCE)MT.1943-5533.0003726).
- [50] A. Shanmugasundaram, K. Jayakumar, Effect of curing regimes on microstructural and strength characteristics of UHPC with ultra-fine fly ash and ultra-fine slag as a replacement for silica fume, *Arab. J. Geosci.* 15 (4) (2022) 345, <https://doi.org/10.1007/s12517-022-09617-y>.
- [51] J. Du, Z. Liu, C. Christodoulatos, M. Conway, Y. Bao, W. Meng, Utilization of off-specification fly ash in preparing ultra-high-performance concrete (UHPC): mixture design, characterization, and life-cycle assessment, *Resour., Conserv., Recycl.* 180 (2022) 106136, <https://doi.org/10.1016/j.resconrec.2021.106136>.
- [52] Z. Wu, C. Shi, K. Khayat, Influence of silica fume content on microstructure development and bond to steel fiber in ultra-high strength cement-based materials (UHSC), *Cem. Concr. Compos.* 71 (2016) 97–109. (<https://doi.org/10.1016/j.resconrec.2022.106741>).
- [53] M.S. Sulaiman, M.M. Abood, S.K. Sinnakaudan, M.R. Shukor, G.Q. You, X. Z. Chung, Assessing and solving multicollinearity in sediment transport prediction models using principal component analysis, *ISH J. Hydraul. Eng.* 27 (sup1) (2021) 343–353, <https://doi.org/10.1080/09715010.2019.1653799>.
- [54] X. Guo, H. Shi, W.A. Dick, Compressive strength and microstructural characteristics of class C fly ash geopolymer, *Cem. Concr. Compos.* 32 (2) (2010) 142–147, <https://doi.org/10.1016/j.cemconcomp.2009.11.003>.
- [55] P.S. Dong, N.V. Tuan, L.T. Thanh, N.C. Thang, V.H. Cu, J.-H. Mun, Compressive strength development of high-volume fly ash ultra-high-performance concrete under heat curing condition with time, *Appl. Sci.* 10 (20) (2020) 7107, <https://doi.org/10.3390/app10207107>.
- [56] Z. Wu, C. Shi, W. He, L. Wu, Effects of steel fiber content and shape on mechanical properties of ultra high performance concrete, *Constr. Build. Mater.* 103 (2016) 8–14, <https://doi.org/10.1016/j.conbuildmat.2015.11.028>.
- [57] Ç. Yalçinkaya, O. Çopuroğlu, Hydration heat, strength and microstructure characteristics of UHPC containing blast furnace slag, *J. Build. Eng.* 34 (2021) 101915, <https://doi.org/10.1016/j.jobe.2020.101915>.

- [58] X. Zhang, S. Zhao, Z. Liu, F. Wang, Utilization of steel slag in ultra-high performance concrete with enhanced eco-friendliness, *Constr. Build. Mater.* 214 (2019) 28–36, <https://doi.org/10.1016/j.conbuildmat.2019.04.106>.
- [59] Z. Wu, C. Shi, W. He, Comparative study on flexural properties of ultra-high performance concrete with supplementary cementitious materials under different curing regimes, *Constr. Build. Mater.* 136 (2017) 307–313, <https://doi.org/10.1016/j.conbuildmat.2017.10.074>.
- [60] P. Ganesh, A.R. Murthy, Tensile behaviour and durability aspects of sustainable ultra-high performance concrete incorporated with GGBS as cementitious material, *Constr. Build. Mater.* 197 (2019) 667–680, <https://doi.org/10.1016/j.conbuildmat.2018.11.240>.
- [61] Y. Shi, G. Long, C. Ma, Y. Xie, J. He, Design and preparation of ultra-high performance concrete with low environmental impact, *J. Clean. Prod.* 214 (2019) 633–643, <https://doi.org/10.1016/j.jclepro.2018.12.318>.
- [62] M. Abdellatif, S.M. AL-Tam, W.E. Elemam, H. Alanazi, G.M. Elgendy, A. M. Tahwia, Development of ultra-high-performance concrete with low environmental impact integrated with metakaolin and industrial wastes, *Case Stud. Constr. Mater.* 18 (2023) e01724, <https://doi.org/10.1016/j.cscm.2022.e01724>.
- [63] Y. Shi, G. Long, X. Zen, Y. Xie, T. Shang, Design of binder system of eco-efficient UHPC based on physical packing and chemical effect optimization, *Constr. Build. Mater.* 274 (2021) 121382, <https://doi.org/10.1016/j.conbuildmat.2020.121382>.
- [64] Ł. Uzarowicz, Z. Zagórski, Mineralogy and chemical composition of technogenic soils (Technosols) developed from fly ash and bottom ash from selected thermal power stations in Poland, *Soil Sci. Annu.* 66 (2) (2015) 82–91, <https://doi.org/10.1515/ssa-2015-0023>.
- [65] P. Guo, Y. Bao, W. Meng, Review of using glass in high-performance fiber-reinforced cementitious composites, *Cem. Concr. Compos.* 120 (2021) 104032, <https://doi.org/10.1016/j.cemconcomp.2021.104032>.
- [66] W. Meng, K.H. Khayat, Effect of graphite nanoplatelets and carbon nanofibers on rheology, hydration, shrinkage, mechanical properties, and microstructure of UHPC, *Cem. Concr. Res.* 105 (2018) 64–71, <https://doi.org/10.1016/j.cemconres.2018.01.001>.
- [67] J. Yang, H. Hu, X. He, Y. Su, Y. Wang, H. Tan, H. Pan, Effect of steam curing on compressive strength and microstructure of high volume ultrafine fly ash cement mortar, *Constr. Build. Mater.* 266 (2021) 120894, <https://doi.org/10.1016/j.conbuildmat.2020.120894>.
- [68] J. Lee, C. Soutis, A study on the compressive strength of thick carbon fibre–epoxy laminates, *Compos. Sci. Technol.* 67 (10) (2007) 2015–2026, <https://doi.org/10.1016/j.compscitech.2006.12.001>.
- [69] X. Xu, Z. Jin, Y. Yu, N. Li, Impact properties of Ultra High Performance Concrete (UHPC) cured by steam curing and standard curing, *Case Stud. Constr. Mater.* 17 (2022) e01321, <https://doi.org/10.1016/j.cscm.2022.e01321>.
- [70] BS EN 12390-3:2002 Testing hardened concrete. Compressive strength of test specimens DOI: <https://doi.org/10.3403/BSEN12390>.
- [71] GBT 17671-1999, Method of Testing Cements - Determination of Strength. 1999, <https://www.chinesestandard.net/PDF.aspx/GBT17671-1999>.
- [72] H. Wickham, Data analysis. 2016. DOI: https://doi.org/10.1007/978-3-319-24277-4_9.
- [73] P. Guo, J. Du, Y. Bao, W. Meng, Real-time video recognition for assessing plastic viscosity of ultra-high-performance concrete (UHPC), *Measurement* (2022) 110809, <https://doi.org/10.1016/j.measurement.2022.110809>.
- [74] A.H. Patoary, M.J.B. Kibria, A. Kaium, Implementation of automated Bengali parts of speech tagger: an approach using deep learning algorithm, 2020 IEEE Reg. 10 Symp. (TENSYP) (2020) 308–311, <https://doi.org/10.1109/TENSYP50017.2020.9230907>.
- [75] A.V. Dorogush, V. Ershov, and A. Gulin, CatBoost: gradient boosting with categorical features support. *ArXiv preprint* 2018, <https://arxiv.org/abs/1810.11363>.
- [76] A.H. Victoria, G. Maragatham, Automatic tuning of hyperparameters using Bayesian optimization, *Evol. Syst.* 12 (2021) 217–223, <https://doi.org/10.1007/s12530-020-09345-2>.
- [77] ASTM C618-22, Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete. DOI: <https://doi.org/10.1520/C0618-22>.
- [78] E. Ferraz, S. Andrejkovicova, W. Hajjaji, A.L. Velosa, A.S. Silva, and F. Rocha, t methodology. *Acta Geodynamica et Geomaterialia*, 2015. 12(3): p. 289–298. DOI: <https://doi.org/10.13168/AGG.2015.0026>.
- [79] P. Guo, W. Meng, H. Nassif, H. Gou, Y. Bao, New perspectives on recycling waste glass in manufacturing concrete for sustainable civil infrastructure, *Constr. Build. Mater.* 257 (2020) 119579, <https://doi.org/10.1016/j.conbuildmat.2020.119579>.
- [80] S. Urhan, Alkali silica and pozzolanic reactions in concrete. Part 1: interpretation of published results and an hypothesis concerning the mechanism, *Cem. Concr. Res.* 17 (1) (1987) 141–152, [https://doi.org/10.1016/0008-8846\(87\)90068-8](https://doi.org/10.1016/0008-8846(87)90068-8).
- [81] S. Mahjoubi, R. Barhemat, P. Guo, W. Meng, Y. Bao, Prediction and multi-objective optimization of mechanical, economical, and environmental properties for strain-hardening cementitious composites (SHCC) based on automated machine learning and metaheuristic algorithms, *J. Clean. Prod.* 329 (2021) 129665, <https://doi.org/10.1016/j.jclepro.2021.129665>.