

Knowledge-guided data-driven design of ultra-high-performance geopolymer (UHPG)

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:

Explainable machine learning
Interpretable artificial intelligence
Knowledge graph
Multi-objective optimization
Physicochemical information
Ultra-high performance geopolymer

ABSTRACT

Geopolymer has been identified as a promising family of sustainable construction materials alternative to cement-based materials. However, designing geopolymer utilizing solid wastes is a challenging task given the large variations of solid wastes in their physical and chemical properties. To overcome this challenge, this paper proposes a knowledge graph-guided data-driven approach to design geopolymer utilizing solid wastes, aimed at achieving high mechanical properties, low material cost, and low carbon emission, while largely improving material discovery efficiency. The proposed approach seamlessly integrates knowledge graph, machine learning, and multi-objective optimization, and has been utilized to design ultra-high performance geopolymer (UHPG). This approach has two main novelties: (1) The incorporation of knowledge graph imparts geopolymer domain knowledge, making the machine learning model interpretable and compliant with domain knowledge. (2) The consideration of physical and chemical properties of raw materials enables the utilization of various solid wastes. The results show that the proposed approach can reasonably predict geopolymer properties, interpret prediction results, and optimize UHPG design.

1. Introduction

Concrete is the most used structural material worldwide. In 2022, the annual consumption was more than 30 billion tons [1]. Such a high consumption volume makes concrete a main contributor to carbon emissions although its unit cost and unit carbon emission are not as high as other popular construction materials such as steel and aluminum [2]. The mechanical strengths of concrete are primarily dependent on the binder, and Portland cement is the most popular binder used in concrete. The manufacturing process of cement is energy intensive and involves high carbon emissions. In 2017, the consumption of cement to produce concrete exceeded 4 billion tons, resulting in 250 million tons of carbon emissions, which accounted for 7 % of the total emission [1]. Aiming to achieve carbon neutrality by 2050 [3], the development and utilization of innovative binders are important missions in the concrete and construction industries.

Geopolymer has been identified as a promising alternative solution that does not use cement while still achieving desired binding performance. Geopolymer often uses solid wastes such as fly ash and slag which are industrial by-products [4]. The utilization of wastes further improves the sustainability of geopolymer by mitigating pollution,

making geopolymer a competitive solution. By just considering the elimination of cement from the mixture while neglecting the other benefits such as the reduction of landfill-induced emissions, the carbon emission of geopolymer can be reduced by 60 % compared with conventional concrete using Portland cement [5].

In addition to the benefits of eliminating cement, geopolymer has exhibited high mechanical properties and superior resistance to various effects such as chemical attacks [6], corrosion [7], and fire [8], making it a competitive solution for structures exposed to harsh environments. This makes geopolymer more promising considering climate change on human habitats [9]. Moreover, the rapid hardening character of geopolymer promotes accelerated construction [10].

Extensive research has been conducted on the development of geopolymer, primarily based on experiments which provide reliable original data that not only advance fundamental knowledge but also promote engineering designs of geopolymer. Experiment-based design and evaluation of geopolymer have been elaborated in previous papers [11–21]. An important lesson learned from previous research is that the mixture design and processing methods must be tailored to achieve desired performance in specific projects because the fresh and hardened properties of geopolymer are closely related to the physical and

^{*} Corresponding author.

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

<https://doi.org/10.1016/j.cemconcomp.2024.105723>

Received 10 January 2024; Received in revised form 9 July 2024; Accepted 19 August 2024

Available online 23 August 2024

0958-9465/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

chemical properties of the raw ingredients, mix proportions, and processing methods. For example, when the particle size gradation and chemical composition of fly ash were changed, the mechanical strengths of geopolymer were largely changed [21].

In real practices, the change of the physical and chemical properties of the raw ingredients is unavoidable for different reasons. First, various types of solid wastes such as fly ash and slag have been utilized to prepare geopolymer. The physical and chemical properties of those ingredients are different. Second, for each type of solid waste such as fly ash, the physical and chemical properties still vary largely in different plants and even different batches in the same plant. With the change of the physical and chemical properties of the raw ingredients of geopolymer, it is often necessary to modify the mix proportion and processing method via repeated time-consuming experiments, posing a significant challenge in the utilization of geopolymer in time-limited projects.

Recently, artificial intelligence (AI) techniques have been utilized to develop time-efficient methods for designing and characterizing concrete efficiently [22–30]. For example, ultra-high-performance concrete (UHPC) [22] and strain-hardening cementitious composites (SHCC) [23] have been designed to achieve low material cost and low carbon emission while retaining superior mechanical properties and durability, simultaneously, by strategically utilizing locally-available solid wastes. The basic idea of AI-assisted design of concrete is to integrate a data-driven predictor and an optimizer:

- The predictor is a machine learning model that correlates the concrete design variables, such as the mixture design, processing methods, and testing methods, with the interested concrete properties, such as the flowability, mechanical strengths, and durability [22], through a training process based on available experimental data.
- The optimizer is an optimization algorithm that maximizes or minimizes certain concrete properties, such as the maximization of mechanical strengths and the minimization of cost and carbon emissions [23]. Multi-objective optimization and decision-making methods have been proposed to achieve multiple objectives simultaneously [23].

Although the efficacy and efficiency of AI-assisted design methods have been verified using experiments [31–35], concrete experts are still concerned about the reliability of those methods because they are based on data analysis rather than concrete knowledge. There are multiple facts that have limited a wider acceptance of machine learning methods in the concrete industry:

- The machine learning prediction models are black-box models that cannot explain the prediction results of concrete properties. The process of generating the prediction results is based on complex mathematical computation, which is different from experiments that can be easily checked to assess the rationality and the quality of the generated data.
- The property prediction of machine learning models is based on models which may violate concrete principles. The physical and chemical properties of raw ingredients and their effects on physical and chemical reactions in concrete were considered in recent study [36], but most studies did not consider the physical and chemical properties of ingredients.
- The sources and quality of data include uncertainties and errors. While both the quantity and the quality of data are important because data is the source of knowledge for machine learning predictive models, many existing papers do not provide detailed information about how the adopted datasets have been generated and how the mixture design variables have been selected. It is difficult to assess the quality of data by concrete experts.

In addition to the above problems, there are also concerns about the generalizability of the AI methods for different types of concrete materials. These concerns have become important obstacles that hinder the application of AI-assisted design methods in the concrete industry, although AI-assisted design methods have shown advantages in previous research.

To address these challenges, this paper presents an approach to incorporate concrete domain knowledge into AI-assisted material design, aimed at an integrated knowledge-guided data-driven design paradigm. In the proposed approach, the domain knowledge about geopolymer concrete is imparted in the form of a knowledge graph, which has been used in large AI systems for various domains such as medical, financial, education, and cyber security [37–41]. Knowledge graph has not been constructed in the context of AI-assisted material design.

The overarching goal of this study is to develop a knowledge-guided interpretable AI-design approach for auto-discovery of sustainable geopolymer concrete, which represents an alternative solution to not only the concrete industry but also the waste management community. To achieve this goal, this research has been conducted for the following three objectives:

- (1) To create a knowledge graph which embodies the domain knowledge about geopolymer.
- (2) To develop the interpretable machine learning approach along with a new framework that integrates a knowledge graph, machine learning, and multi-objective optimization.
- (3) To implement the interpretable machine learning approach and knowledge graph into the design of ultra-high-performance geopolymer (UHPG).

The proposed approach is developed in the context of geopolymer concrete, and its efficacy is assessed by the mechanical properties, material cost, and carbon emission of designed UHPG. This approach has three important novelties:

- First, a knowledge graph is constructed and integrated into AI-assisted design of UHPG, making the machine learning model interpretable and compliant with domain knowledge via imparting geopolymer domain knowledge. Specifically, the knowledge graph is used to achieve three abilities: (a) Guide the selection of design variables for machine learning predictors through identifying relevant variables, enabling concrete experts to assess the rationality of the adopted design variables. (b) Interpret the results from machine learning predictors via explicitly revealing the underlying mechanisms such as physicochemical reactions. (c) Facilitate the identification of existing knowledge gaps and the generation of new knowledge for enriching the knowledge graph.
- Second, the physical and chemical properties of raw materials are considered in the proposed approach using an artificial language, thereby enabling the AI-designer to handle various solid wastes that involve different physical and chemical properties.
- Third, by integrating the knowledge graph into the design framework, a novel explainable prediction and optimization framework is developed to achieve the ability to consider relevant domain knowledge about geopolymer concrete.

The remainder of the paper is organized as follows: Section 2 presents the methods adopted to achieve these objectives, including the new framework and its components, with an emphasis placed on the rationality and the construction of the knowledge graph. Section 3 presents the results from the proposed approach. Section 4 summarizes the key findings of this research.

2. Methods

A knowledge-guided data-driven geopolymers concrete designer is presented (Fig. 1). The designer includes four main components: (1) Construction of a knowledge graph. The knowledge graph is constructed to guide variable selection and explain the prediction results of machine learning models. (2) Construction of a dataset. The dataset is constructed by considering crucial factors, including mixing design variables, physiochemical properties of raw materials, and processing methods identified by the knowledge graph. Different forms of data are integrated into numerical data via feature engineering. (3) Prediction by machine learning. Model selection and evaluation are performed to select a model for predicting compressive strength of geopolymers with high accuracy. (4) Multi-objective optimization. The aim is to maintain high strength while reducing cost and carbon emissions when designing geopolymers concrete. The optimization algorithm uses the results predicted by machine learning, along with carbon emissions and costs of raw materials, to generate a series of solutions. A multi-criteria decision-making (MCDM) method is applied to find high-performance, low-carbon, and cost-effective geopolymers concrete.

2.1. Overview

The core idea for knowledge graph-guided AI design of geopolymers concrete is shown in Fig. 2. The framework integrates a knowledge graph, a machine learning predictor, and a multi-objective optimizer, which play three critical roles in the design team: (1) The knowledge graph acts as a domain expert who imparts domain knowledge to the team. (2) The machine learning predictor provides the quantitative prediction of geopolymers concrete properties based on the input variables, which include the mixture design variables, processing methods, testing methods, and so on, depending on the specific application. (3) The multi-objective optimizer outputs the optimal geopolymers design based on the design objectives that are dependent on the specific use case.

The integration of knowledge graph and machine learning predictor enable the AI design team to achieve the qualitative reasoning ability via the knowledge graph and the quantitative regression ability via the machine learning predictor, generating a knowledgeable AI designer that is able to predict UHPG properties and interpret the prediction results. Further, a multi-objective optimizer is incorporated to achieve the capability of optimizing the mechanical properties, cost, and carbon emission of UHPG simultaneously.

The construction of the knowledge graph involves two main tasks, which are knowledge collection and graphic representation. Knowledge collection refers to the collection of domain knowledge, which is the knowledge about geopolymers in this research. Graphic representation refers to the construction of graphs that describe the knowledge, based on knowledge graph techniques [42]. More details about knowledge graphs are available in section 2.2.

The establishment of machine learning predictor can be performed based on a procedure well-developed in previous research [43]. The procedure has several steps, such as data collection, the selection of input variable, data pre-processing, model selection, hyperparameter tuning, and model training. In previous research, data collection and variable selection were performed in an empirical or arbitrary way that could not be generalized for handling different types of concrete. In this research, a knowledge graph is utilized to guide the selection of input variables and data collection, and the selection of input variables is promoted before the step of data collection, aimed at facilitating the time-consuming data collection effort. Once the machine learning predictor is established, the geopolymers properties can be predicted by specifying the input variables. In this, the knowledge graph is utilized to interpret the prediction results, making the AI designer interpretable. More details about the machine learning predictor are available in section 2.3.

The multi-objective optimizer utilizes the data provided by the machine learning predictor to perform multi-objective optimization based on specific design objectives and output the optimal design of geopolymers. In this paper, a multi-objective optimizer is utilized to discover new UHPG mixtures that can achieve high mechanical strengths, low material cost, and low carbon emission, simultaneously. More details about the multi-objective optimizer are available in section 2.4.

The interconnection of the knowledge graph, machine learning predictor, and multi-objective optimizer is demonstrated by using the predictor and optimizer to discover new geopolymers mixtures. This discovery leads to new knowledge and the expansion of the knowledge graph. More details on this process are in section 3.

2.2. Knowledge graph

A knowledge graph is a structured graphic representation of knowledge, and it can be utilized to organize, store, and represent the relationships between different pieces of information in a way that is understandable and operable for computers and human. A knowledge graph typically has two types of components, which are nodes and

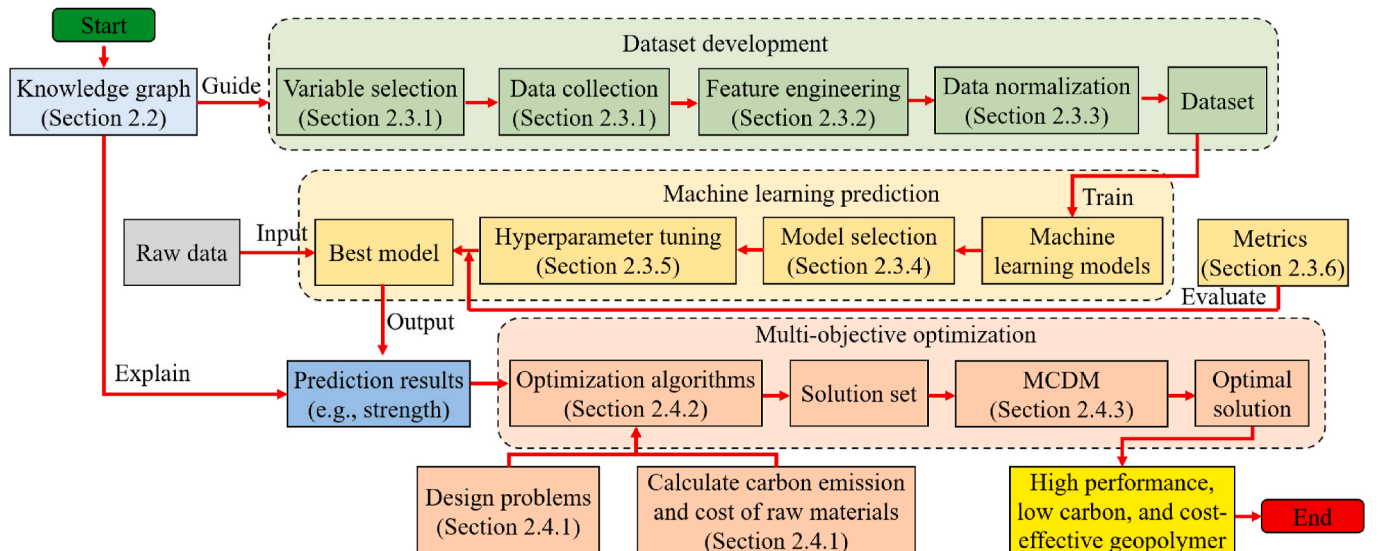


Fig. 1. Illustration of the proposed knowledge-guided data-driven geopolymers concrete designer.

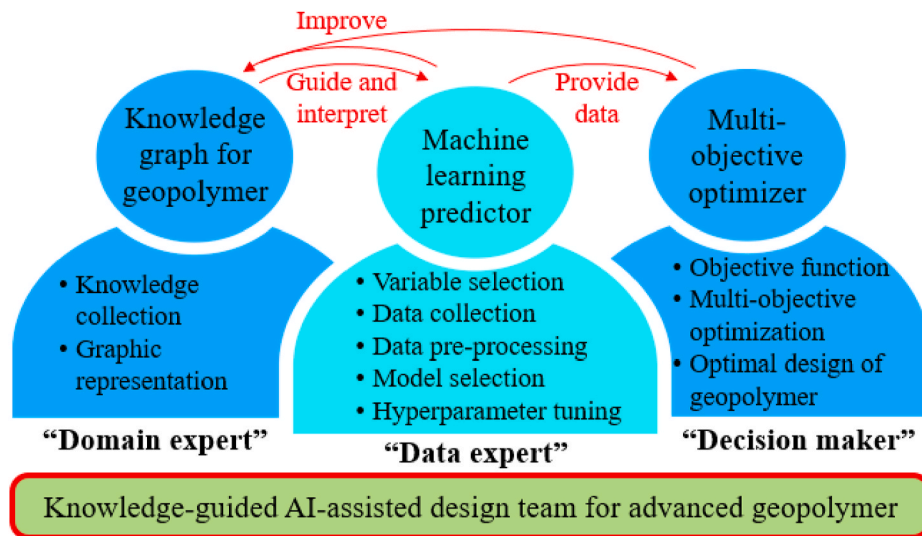


Fig. 2. Analogical description of the proposed framework as a three-member cooperative team.

edges, as shown in Fig. 3 by text boxes and arrows, respectively. Nodes represent entities, concepts, or data points. Edges represent the relationships between entities. Each box represents a node containing entities. For example, the entity on the top level is "Geopolymer concrete". Each arrow is used to indicate the relationship between two entities. For example, the relationship between "Geopolymer concrete" and "Raw ingredients" is "has", meaning geopolymer concrete has raw ingredients. It is noted that Fig. 3 is a simplified example of the knowledge graph for geopolimer. More domain knowledge can be added to enrich the knowledge graph, as shown in Fig. A1 in the appendix.

To construct a knowledge graph for geopolimer, three important aspects must be considered, including the available raw ingredients, processing methods, and concerned properties:

- (1) The raw ingredients include the precursors, aggregates, activators, water, and fibers. It is essential to understand the mechanisms of different ingredients that affect the concerned properties of geopolimer. Geopolimer is produced via geopolimerization reactions between aluminosilicate sources (or calcium-silicate source) and alkaline activators, forming a three-dimensional

network with mechanical strengths. Initially, activators create an alkaline environment that facilitates the dissolution of aluminosilicate materials, releasing reactive SiO_2 and Al_2O_3 , which are essential for geopolimerization. Next, a three-dimensional polymeric network is formed via polycondensation reactions, with chemical bond aggregation by reactive SiO_2 and Al_2O_3 . Then, the network is further solidified via curing, yielding a hardened product. The compressive strength of geopolimer is influenced by the dissolution of Si/Al (or Ca/Si) compositions in the raw materials, as depicted in Table 1. Particle size also affects the compressive strength of geopolimers by influencing reactivity. Smaller particles enhancing reactivity by seeding effect during geopolimerization.

- (2) The processing methods include the pre-treatment of raw ingredients, mixing protocol, curing scheme, and testing methods. In this research, the considered factors are the curing scheme and testing methods [44]. While the variation of the pre-treatment of raw ingredients and mixing protocol are not considered in this research, but they can be considered using the same method in future studies. It is essential to include the knowledge about how

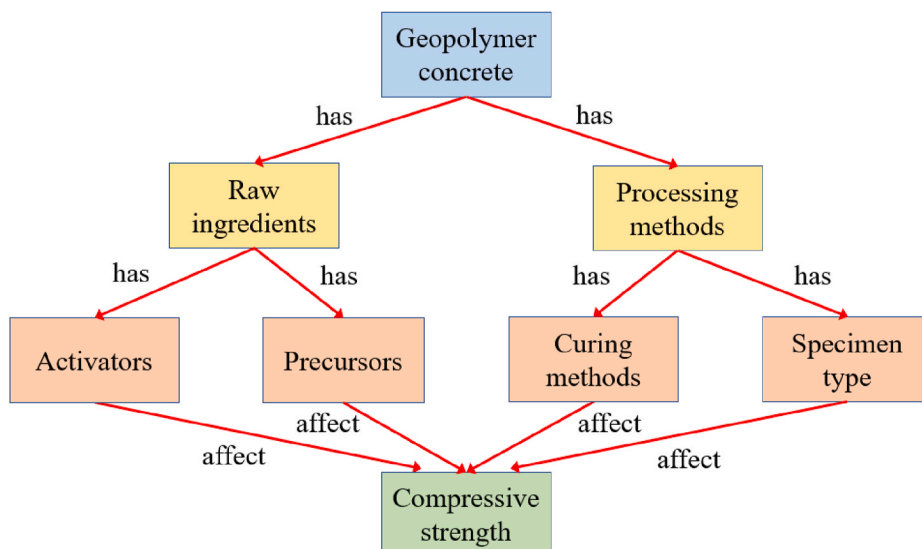
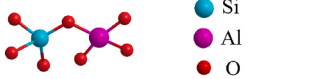
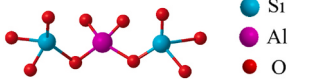
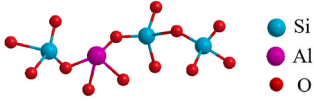
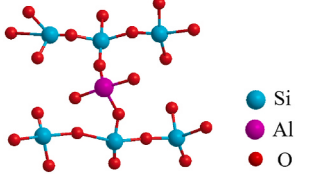


Fig. 3. Simple example of a knowledge graph for geopolimer considering compressive strengths.

Table 1
Effect of different Si-to-Al ratios on geopolymerization.

Si-to-Al ratio	Chemical formulation	Illustration
Si:Al = 1	-Si-O-Al-O-	
Si:Al = 2	-Si-O-Al-O-Si-O-	
Si:Al = 3	-Si-O-Al-O-Si-O-Si-O-	
Si:Al > 3	$\begin{array}{c} \text{O} \quad \text{O} \quad \text{O} \\ \quad \quad \\ \text{O}-\text{Si}-\text{O}-\text{Si}-\text{O}-\text{Si}-\text{O} \\ \quad \quad \\ \text{O} \quad \text{O} \quad \text{O} \\ \\ \text{O}-\text{Al}-\text{O} \\ \\ \text{O} \quad \text{O} \quad \text{O} \\ \quad \quad \\ \text{O}-\text{Si}-\text{O}-\text{Si}-\text{O}-\text{Si}-\text{O} \\ \quad \quad \\ \text{O} \quad \text{O} \quad \text{O} \end{array}$	

the curing schemes (e.g., standard and steam curing) and testing methods (e.g., specimen type and loading rate) impact the concerned properties of geopolymer.

- (3) The concerned properties can include the fresh and hardened properties. In this research, the concerned properties are the compressive strength, material cost, and carbon emission [32]. In practices, it is also important to consider the fresh properties such as the flowability and setting time because they play important roles in the mechanical properties and durability as well as the construction quality [45]. More properties can be considered using the same methods.

2.3. Machine learning

The machine learning predictor is an essential role in an AI design team (Fig. 2) and provides the prediction of material properties. The following subsections introduce variable selection, data collection, data pre-processing, model selection, hyperparameter tuning, and performance metrics.

2.3.1. Variable selection and data collection

Knowledge graphs play a crucial role in guiding variable selection by qualitatively identifying relevant factors affecting the mechanical properties of geopolymer. Based on the knowledge graph, relevant variables of geopolymer are selected and then utilized as the input variables of machine learning models. In this research, the key variables were identified and classified into numerical variables (e.g., mixture design variables, solution concentrations, curing temperatures, and curing times), textual variables (e.g., physical and chemical properties of raw ingredients), and categorical variables (e.g., curing methods, specimen type).

It is noted that the physical and chemical properties of raw ingredients are considered in this research. This is different from the use of engineering names such as Class C or Class F fly ash for representing raw ingredients since an engineering name corresponds to ingredients with different physical and chemical properties. The use of the physical and chemical properties enables machine learning to consider various wastes and knowledge graph to interpret the machine learning results. In this study, the dataset includes 676 different geopolymer mixtures in the literature. The data analysis for numerical data, textual data, and

categorical data are presented as follows.

2.3.1.1. Numerical data. Numerical data include various mixture design variables such as the mass contents of binders (slag, fly ash, silica fume, and metakaolin), aggregates (fine and coarse aggregate), liquid materials (NaOH solution, Na₂SiO₃ solution, and water), as well as NaOH concentration, steel fiber content (by volume), curing temperature, and curing time. The output variable is the 28-day compressive strength of geopolymer concrete. The statistics of the numerical data are shown in Table 2.

Multicollinearity analysis is a crucial aspect of regression analysis that involves multiple independent variables. Multicollinearity refers to a situation when independent variables are highly correlated [46], which can lead to misleading interpretations of variable importance. To analyze multicollinearity, a correlation matrix was built to determine the Pearson correlation coefficients of the numerical input variables [25]. When the maximum Pearson correlation coefficient is greater than 0.7, multicollinearity occurs, and the dataset must be modified to eliminate multicollinearity [24]. The results of correlation analysis are shown in Fig. 4. The Pearson correlation coefficient of the input variables was up to 0.65, indicating the absence of multicollinearity. Therefore, it is appropriate to use this dataset to train a machine learning model.

2.3.1.2. Textual data. In geopolymer, the chemical compositions such as the CaO, SiO₂, and Al₂O₃ percentages in binders play significant roles in strength development. However, different batches of solid waste usually exhibit distinct physicochemical characteristics. It is inappropriate to use the engineering names of solid wastes to represent the materials in the development of machine learning models without considering the physicochemical properties of ingredients. The importance of utilizing the physicochemical properties as input variables in machine learning models has been proven in a recent study on predicting concrete properties when various solid wastes are used [36]. A ternary diagram of the chemical compositions of various solid waste is shown in Fig. 5. Slag, fly ash, and metakaolin are popular aluminosilicate and calcium-silicate sources for geopolymer [13,47].

The strength of geopolymer is influenced by the median particle size (D₅₀) of raw ingredients. D₅₀ provides a central measure of the particle

Table 2
Description of selected mixture design variables of geopolymer concrete.

Number	Variables	Range	Mean	S.D.	Skewness	Kurtosis
1	Slag content	0–1.00	0.42	0.37	0.10	−1.56
2	Fly ash content	0–1.00	0.35	0.39	0.59	−1.27
3	Metakaolin content	0–1.00	0.18	0.37	1.71	1.01
4	Silica fume content	0–0.30	0.06	0.10	1.44	0.66
5	Fine aggregate content	0.16–3.30	1.43	0.63	1.01	0.34
6	Coarse aggregate content	0–5.09	1.49	1.49	0.34	−1.35
7	NaOH content	0–0.63	0.15	0.10	1.23	1.91
8	NaOH concentration (M)	0–16	12.47	2.28	−2.26	11.83
9	Na ₂ SiO ₃ content	0–0.94	0.32	0.16	0.76	1.20
10	Extra water content	0–0.34	0.05	0.10	1.73	1.61
11	Steel fiber content (vol%)	0–3.0	0.38	0.78	1.97	2.78
12	Curing temperature (°C)	20–100	26.47	16.12	2.59	5.45
13	Curing age (days)	1–91	19.58	14.53	1.80	6.96
14	Compressive strength (MPa)	5.6–171.2	64.8	43.95	0.83	−0.56

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.

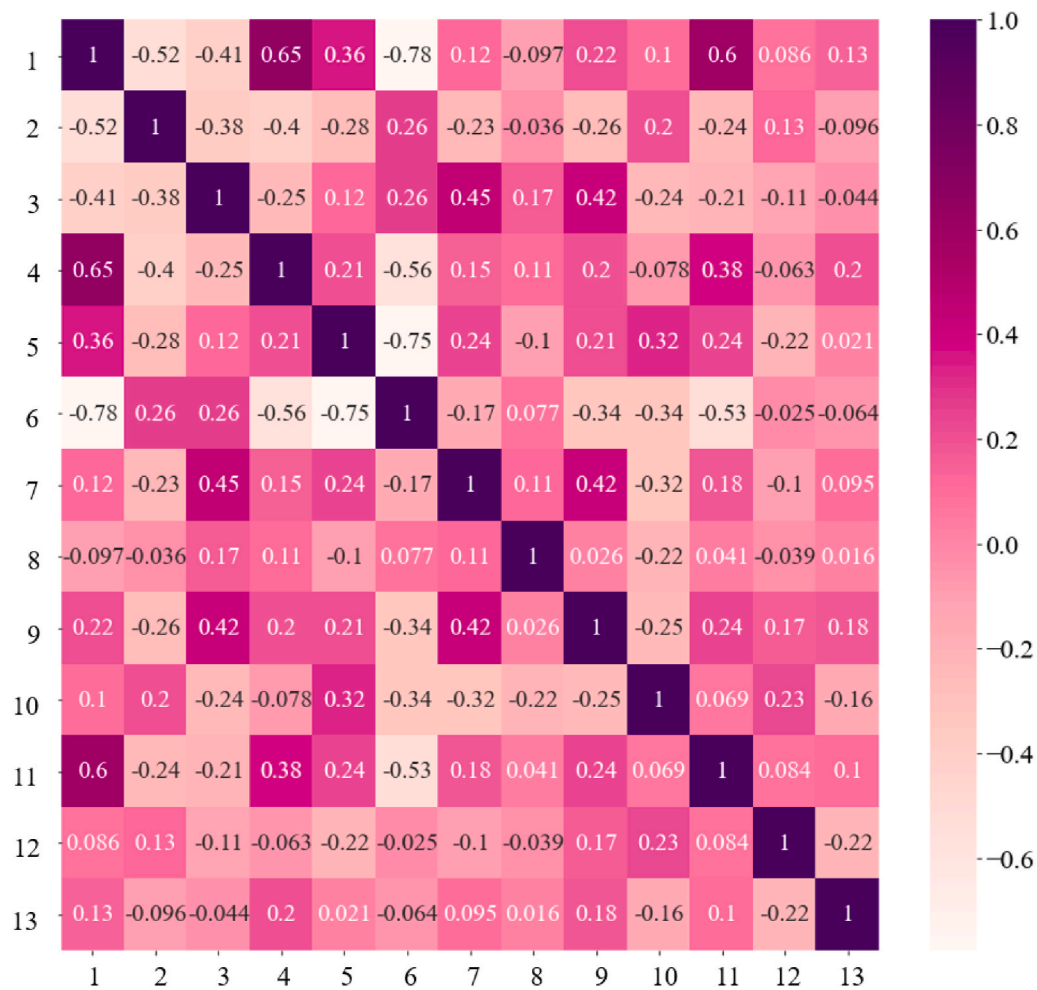


Fig. 4. Correlation matrix of the input numerical data. Numbers 1–13 in the heatmap correspond to the variable numbers in Table 1.

size distribution. The selection of physicochemical properties also considers the accessibility of the data, the chemical elements CaO, SiO₂, Al₂O₃, and D₅₀ are widely available in the existing publications. The statistics of the physicochemical properties that affect the compressive strength of geopolymer are listed in Table 3, including the types, ranges, mean values, kurtosis, and skewness of the physicochemical properties.

2.3.1.3. Categorical data. The compressive strength of geopolymer concrete depends on the curing method and sample type. For example,

steam curing can promote geopolymerization and enhance the compressive of geopolymer [48]. The test results of the compressive strength are also associated with the size of specimens. Typically, the compressive strength decreases with the increase of the specimen size [49], known as the size effect. The curing methods and specimen types are shown in Table 4. The curing methods include: (1) Standard curing, at room temperature and relative humidity higher than 95 %; (2) heat curing; (3) steam curing, at high temperature and high humidity; and (4) air curing. Four different types of specimens for compressive testing

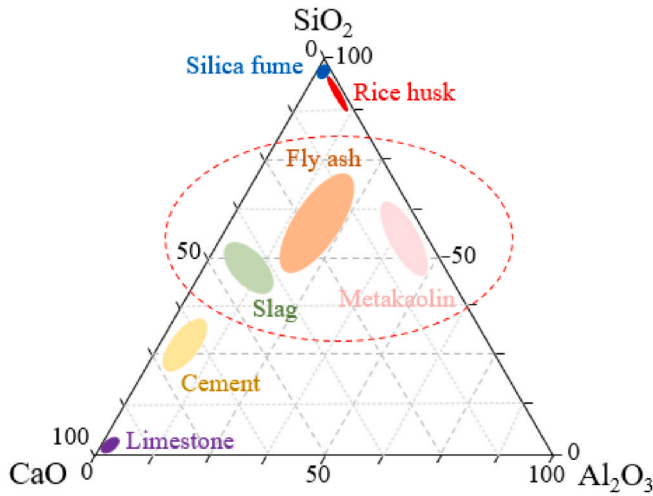


Fig. 5. Ternary diagram (CaO, SiO₂, and Al₂O₃) of various types of solid wastes for geopolymers.

were considered: (1) cubic specimens measuring 51 mm in side length [50]; (2) cubic specimens measuring 71 mm in side length [51]; (3) prism specimens measuring 160 mm × 40 mm × 40 mm [52]; and (4) cylinder specimens measuring 51 mm in diameter and 101 mm in height [53].

2.3.2. Feature engineering

One-hot encoding is a technique used to represent categorical data as vectors or numerical data, which is often utilized in machine learning and data pre-processing tasks. It is useful when a machine learning algorithm utilizes categorical variables but require numerical input data, such as artificial neural networks [54]. In this study, categorical data were converted into numerical data using the one-hot encoding method and then utilized to train the machine learning model. An example of one-hot encoding is shown in Fig. 6.

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 raw ingredients in geopolymers was stored in a dictionary, as shown in Eq. (1). DictVectorizer was used to extract the categorical and numerical features from the dictionary [55]. 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.

$$\{\text{'Type': Metakaolin, 'CaO': 0.9, 'SiO}_2\text{: 51.3, 'Al}_2\text{O}_3\text{: 37.4, 'D50': 2.3}\} \quad (1)$$

2.3.3. Data normalization

The significant discrepancy of the numeric values of different variables (Table 2) may highly affect the results of the machine learning models. It is often beneficial to normalize the input and output data of machine learning models to prevent overfitting between different numerical scales. In this research, the values of the input variables were normalized to the range of −1 to 1, as described in Eq. (2):

$$x^* = \frac{x - \mu}{\sigma} \quad (2)$$

where x is the original data; x^* is the normalized data; μ is the mean value; and σ is the standard deviation. The distribution of data was kept the same before and after the application of the data normalization method [56].

2.3.4. Selection of machine learning algorithms

Eight different machine learning algorithms have been investigated and compared to achieve the optimal machine learning model in terms of the prediction accuracy and generalability. These machine learning algorithms include linear regression, ridge regression, support vector regressor, K neighbors regressor, and four ensemble methods which are random forest, extreme gradient boosting (XGBoost), light gradient Boosting machine (LightGBM), and CatBoost regressor, as shown in Table 5.

Three high-performnace algorithms evaluated based on prediction accuracy are selected for future research. Since there are many

Table 4

Description of curing method and specimen type of geopolymers concrete.

Number	1	2	3	4
Curing method	Standard curing	Heat curing	Steam curing	Air curing
Specimen type	Cube 51	Cube 71	Prism	Cylinder

Note: “Cube 51”: cubic specimens measuring 51 mm × 51 mm × 51 mm; “Cube 70.7”: cubic specimens measuring 70.7 mm × 70.7 mm × 70.7 mm; “Prism”: prism specimens measuring 160 mm × 40 mm × 40 mm; and “Cylinder”: cylindrical specimens with heights of 101 mm and diameters of 51 mm.

Table 3

Statistical data of physicochemical information of raw ingredients.

Number	Materials	Properties	Range	Mean	S.D.	Kurtosis	Skewness
1	Slag	CaO (%)	34.6–46.3	40.1	3.6	−0.63	0.22
		SiO ₂ (%)	25.3–36.7	33.8	2.9	6.08	−2.27
		Al ₂ O ₃ (%)	11.2–17.2	14.4	1.8	−1.00	0.02
		D ₅₀ (μm)	1.0–167.0	27.9	47.1	9.81	3.08
2	Fly ash	CaO (%)	1.2–14.2	3.8	3.3	5.44	2.31
		SiO ₂ (%)	36.2–56.9	50.4	5.8	1.01	−1.21
		Al ₂ O ₃ (%)	19.9–31.9	27.4	3.6	0.20	−0.91
		D ₅₀ (μm)	1.0–350.0	51.2	91.1	10.11	3.10
3	Metakaolin	CaO (%)	0.1–1.8	1.1	0.6	1.92	−1.13
		SiO ₂ (%)	45.3–51.0	49.0	2.5	−0.26	−1.04
		Al ₂ O ₃ (%)	37.1–42.6	40.5	2.6	−2.48	−0.71
		D ₅₀ (μm)	0.15–4.3	1.4	1.9	2.22	1.61
4	Silica fume	CaO (%)	0.7–1.8	1.4	0.5	−1.66	0.06
		SiO ₂ (%)	92.1–96.4	94.4	1.9	−1.42	0.16
		Al ₂ O ₃ (%)	0–0.8	0.2	0.4	0.05	1.03
		D ₅₀ (μm)	0.2–4.3	1.4	1.9	2.22	1.61

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.

Curing method	Standard curing	Heat curing	Steam curing	Air curing
Standard curing	1	0	0	0
Standard curing	1	0	0	0
Seam curing	0	1	0	0
Heat curing	0	0	1	0
Air curing	0	0	0	1

Fig. 6. Example of converting curing methods into numerical data using one-hot encoding.

candidate models and the hyperparameter-tuning process is time-consuming, the machine learning models are initially tested without hyperparameter tuning. This model selection method is consistent with the methods in Refs. [24,57].

2.3.5. Hyperparameters tuning

Bayesian hyperparameter tuning is used for automatic hyperparameter tuning. This technique is particularly useful when the machine learning model is complex and has many hyperparameters. Bayesian optimization is applied to automatically and iteratively optimize the hyperparameters following the Gaussian process [58]. Each iteration is based on the Gaussian function fitted in the previous iteration, aimed at identifying better hyperparameters compared with the prior iteration.

In this research, cross-validation was performed in the optimization process to improve the prediction 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 [58], as shown in Eq. (3):

$$x_p = \underset{x \in X}{\operatorname{argmax}} f(x) \quad (3)$$

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.

2.3.6. Performance metrics

To evaluate the performance of machine learning models, four typical performance metrics were adopted, which are the mean absolute error (MAE), root mean squared error (RMSE), Mean absolute percentage error (MAPE), and coefficient of determination (R^2), respectively. The mathematical definitions of the three metrics are shown in Eq. (4) to Eq. (7).

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - A_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - A_i)^2} \quad (5)$$

Table 5
Machine learning algorithms.

Number	Model	Category
1	Linear regression	Single model
2	Ridge regression	Single model
3	Support vector regressor	Single model
4	K neighbors regressor	Single model
5	Random forest	Ensemble learning
6	Extreme gradient boosting	Ensemble learning
7	Light gradient Boosting machine	Ensemble learning
8	CatBoost regressor	Ensemble learning

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - A_i}{A_i} \right| \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - \bar{A})^2}{\sum_{i=1}^n (A_i - \bar{A})^2} \quad (7)$$

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.

2.4. Multi-objective optimization

2.4.1. Design objectives

To design sustainable UHPG, three objective functions are considered: (i) minimization of the cost ($F1$), (ii) minimization of carbon emission, ($F2$), and (iii) maximization of the compressive strength ($F3$), as shown in Eq. (8) and Eq. (9). In this study, the power consumption of high-temperature curing was assumed to be 3 kW; the cost of electricity in New Jersey is \$0.19 per kWh [59], and the equivalent CO₂ emission is 0.371 kg [60].

$$F1 = \sum_{i=1}^n m_i \times r_i + P \times t \times r_c \quad (8)$$

$$F2 = \sum_{i=1}^n m_i \times C_i + P \times t \times C_c \quad (9)$$

where n is the number of raw materials; m_i is the mass of the i th ingredient; r_i is the unit price of the i th ingredient; and C_i is the carbon emission of manufacturing the i th ingredient; P is the powder consumption for high-temperature curing; t is the curing time, in days; r_c is the unit price of electricity for curing; C_c is the CO₂ emission produced by consuming electricity during curing.

The inventory of the unit cost and carbon emission of the raw ingredients are listed in Table 6, including different precursors, fine and coarse aggregate, activators, water, and steel fiber.

2.4.2. Multi-objective optimization algorithm

In this research, a multi-objective optimization method based on the AGE-MOEA algorithm was used to optimize the compressive strength, cost, and carbon emission, simultaneously. AGE-MOEA was designed for multi-objective optimization using genetic algorithms and often used in decision-making problems where multiple conflicting objectives are considered simultaneously. The population and generation are set at 100 and 300, respectively [74]. Constraint functions are set to ensure reasonable design of geopolymer, as shown in Eq. (10), which respectively enforce the total volume fraction of binder, aggregate, fiber, and liquid.

$$(V_b + V_{Fa} + V_{Ca} + V_f + V_l - 1) < 0.0001 \quad (10)$$

where X_i is the i th design variable in Table 1; V_b , V_{Fa} , V_{Ca} , V_f , and V_l represent the volume of binder, fine aggregate, coarse aggregate, fibers, and liquid, respectively.

Table 6
Inventory of representative raw material.

Number	Materials	Specific gravity	Cost (\$/kg)	Carbon emission (kg/kg)
1	Slag	2.90 [22]	0.100 [61]	0.085 [62]
2	Fly ash	2.70 [22]	0.026 [63]	0.005 [64]
3	Metakaolin	2.62 [22]	0.500 [61]	0.332 [65]
4	Silica fume	2.20 [22]	0.800 [61]	0.014 [66]
5	Fine aggregate	2.64 [22]	0.162 [67]	0.020 [68]
6	Coarse aggregate	2.74 [69]	0.014 [70]	0.004 [64]
7	NaOH	1.57 [29]	0.380 [29]	1.915 [29]
8	Na ₂ SiO ₃	2.10 [29]	0.170 [29]	1.514 [29]
9	Water	1.00 [22]	0.001 [71]	0.0003 [72]
10	Steel fiber	7.80 [22]	5.000 [61]	2.650 [73]

2.4.3. Multi-criteria decision making

Multi-criteria decision-making is performed to choose the optimal solutions while considering the mutual impact of different objective functions. This study utilizes the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) for determining the optimal solutions. The basic mechanism of TOPSIS involves identifying the optimal and worst solutions from the normalized solution matrix, calculating the distance of each evaluation object to these solutions, and determining the closeness of each object to the optimal solution for decision making [75].

TOPSIS is utilized to help decision-makers to strike a balance between quantitative data and their subjective preferences by allowing them to assign weights to criteria. However, it is important to note that the quality of the results in TOPSIS depends on the accurate assignment of criteria weights and the correct normalization of data. In addition, TOPSIS assumes that the criteria are independent and that the ideal and anti-ideal solutions represent the true best and worst outcomes, which may not always be the case in real-world scenarios.

3. Results and discussion

3.1. Property prediction

3.1.1. Selection of machine learning methods

The results of the performance metrics of eight machine learning methods for predicting the 28-day compressive strength of geopolymer are listed in Table 7. Default hyperparameters are used. The XGBoost model shows the highest accuracy, as indicated by the lowest MAE, RMSE, and MAPE, and the highest R^2 . The random forest and CatBoost regressor models also achieved reasonable accuracy ($R^2 > 0.80$). Therefore, they are selected for further investigations in this research.

Table 7
Pre-selection of machine learning algorithms.

Model	MAE (MPa)	RMSE (MPa)	MAPE	R^2
Support vector regressor	33.21	50.93	0.512	0.629
K neighbors regressor	29.67	40.91	0.489	0.644
Ridge regression	22.84	24.79	0.443	0.714
Linear regression	22.86	24.79	0.442	0.714
LightGBM	15.89	20.13	0.387	0.789
CatBoost regressor	14.63	20.13	0.343	0.802
Random forest	14.81	19.89	0.343	0.817
XGBoost	14.06	19.06	0.315	0.831

The MAE and R^2 results from the different models are compared in Fig. 7.

3.1.2. Hyperparameter tuning

The hyperparameters of the random forest, CatBoost, and XGBoost algorithms have been tuned to improve the accuracy. The optimal hyperparameters were obtained through Bayesian optimization and 5-fold cross-validation in terms of the smallest MAE. The iterative curves of the MAE values during hyperparameter tuning are shown in Fig. 8.

In each iteration, the hyperparameters obtained from 5-fold cross-validation were averaged and used to evaluate the predictive performance of the models. All MAE curves converged within 100 iterations, revealing the excellent performance of Bayesian optimization in hyperparameter tuning. The XGBoost model achieved the lowest MAE after hyperparameter tuning. The optimal hyperparameters and corresponding search spaces are shown in Table 8.

3.1.3. Prediction accuracy

The prediction performance of the machine learning models was re-evaluated, as shown in Table 9. The three machine learning models achieved excellent prediction accuracy on the training and testing datasets. Among the three models, the XGBoost model showed the highest accuracy, with the lowest MAE, MAPE and RMSE and the highest R^2 , followed by the CatBoost model.

The predicted results from the three machine learning models are compared with the actual results for both the training and testing datasets, as shown in Fig. 9. Overall, the predicted results are consistent with the actual results.

In the performance evaluation, the metrics were determined using the testing dataset, which are experimental data that are unseen to the trained machine learning models. In other words, the prediction performance of the three trained machine learning models has been evaluated using extensive experimental data, as shown by the dense red dots in Fig. 9. The high accuracy indicates that it is reasonable to utilize the trained machine learning models to predict geopolymer properties because that can largely mitigate the experimental testing effort.

The performance of the XGBoost model was further validated using a set of experimental data extracted from two references [21,47]. Those experimental data were not used in the training or the testing datasets, so they were new data for the predictive models. The comparison of the predicted results from the machine learning models and the experimental results are shown in Fig. 10. When the slag replacement percentage increases from 0 to 100 % or when the steel fiber content increases from 0 to 3 %, the change of the compressive strength of geopolymer is evaluated. The maximum absolute error between the experimental results and the predicted results was less than 5 MPa, indicating the high predictive accuracy of the XGBoost model.

3.2. Interpretation of prediction results

3.2.1. Case study 1

3.2.1.1. Raw materials and mixture design. In case study 1, the effect of silica fume on the compressive strength of geopolymer is studied. The physiochemical properties of binders are listed in Table 10. Slag was rich in CaO, SiO₂, and Al₂O₃, while silica fume was rich in SiO₂. The slag and silica fume had comparable particle sizes.

The analysis of chemical compositions of binders is shown in Fig. 11. When the silica fume content in the binder system increases from 0 % to 15 %, the SiO₂ content increases from 31 % to 51.3 %, the CaO content decreases from 36.8 % to 25.8 %, and the Al₂O₃ content decreases from 17.4 % to 12.3 %.

The investigated mixing proportions by mass are listed in Table 11. Silica fume was used to partially replace slag at 0, 15 %, and 30 % by the

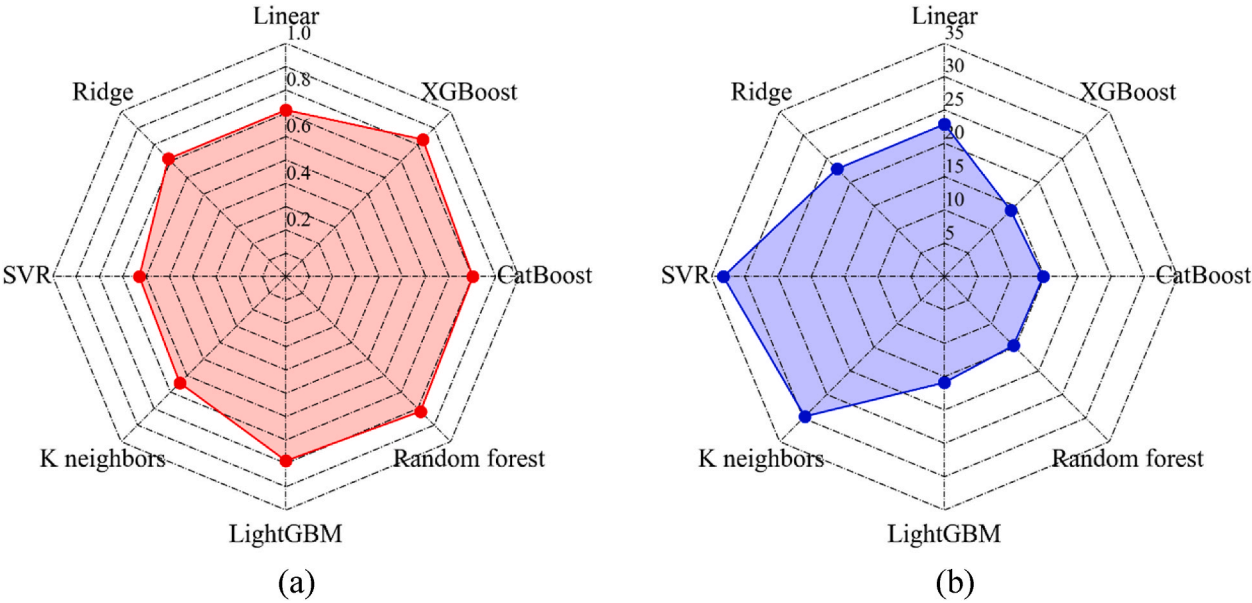


Fig. 7. Comparison of eight different machine learning models: (a) R^2 , and (b) MAE (unit: MPa).

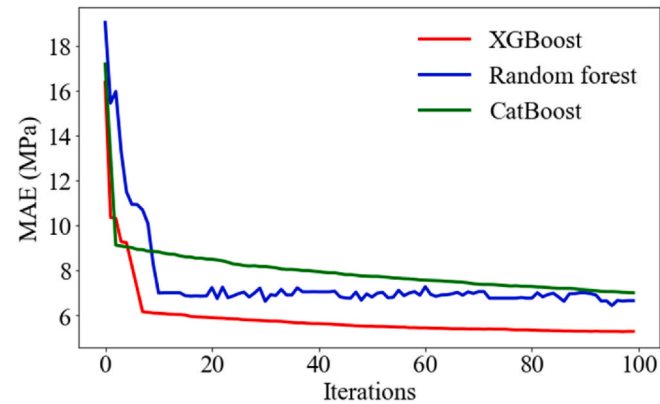


Fig. 8. Convergence of the prediction error of different machine learning models.

Table 8
Hyperparameters of the different machine learning models.

Model	Name	Searching space	Optimal hyperparameters
CatBoost	Learning rate	0.05–1.0	0.092
	Bagging temperature	0.05–1.0	0.341
	Border count	10–100	50
	Colsample_bylevel	0.1–1.0	0.2
	Depth	4–20	8
Random forest	Maximum depth	1–20	8
	Maximum features	1–20	19
	Maximum sample split	1–10	8
	Maximum sample leaf	1–10	9
XGBoost	Booster	Gbtree, Gblinear, Dart	Gbtree
	Number of estimators	10–200	120
	Learning rate	0.01–0.20	0.012
	Maximum depth	1–20	4
	Colsample_bynode	0–10	1
	Number of leaves	2–10	2

Table 9
Comparison of the performance of different machine learning models.

Model	Dataset	MAE (MPa)	RMSE (MPa)	MAPE	R^2
Random forest	Training	3.17	6.69	0.021	0.98
	Testing	7.20	8.76	0.075	0.94
CatBoost	Training	3.29	5.31	0.013	0.99
	Testing	5.68	8.68	0.064	0.95
XGboost	Training	2.49	4.69	0.010	0.99
	Testing	5.35	7.33	0.027	0.97

mass of slag [66]. A mixture of sodium hydroxide and sodium silicate solutions was used as the activator. The liquid-to-binder ratio was set at 0.30, and the concentration of NaOH was 12 M. The steel fiber content was 2 vol%. The specimens were demolded after 24 h and stored at room temperature until testing. The compressive strength test was conducted using cube specimens measuring 71 mm in the side length at 28 days.

3.2.1.2. Prediction results. The XGBoost model was used to predict the compressive strength of geopolymer mixtures. The prediction results of 28-day compressive strengths are shown in Fig. 12. The maximum error is 2.7 MPa, indicating that the XGBoost predictive model can provide reasonable predictions.

3.2.1.3. Interpretation of prediction results. The knowledge graph has been utilized to interpret the results from the XGBoost predictive model in terms of the underlying mechanisms of the effects of silica fume content on compressive strength. The mechanisms are identified from the knowledge graph (Fig. A2). The influencing pathways of the effects of silica fume and slag contents on compressive strength are revealed by the red arrows between the nodes. With the knowledge graph (Fig. A2), the influencing pathways can be determined automatically and utilized to list the following explanation:

Silica fume content $\rightarrow$ Ca/Si ratio $\rightarrow$ Compressive strength
Si/Al ratio

According to the machine learning model, when the replacement percentage of silica fume increases from 0 to 30 %, the 28-day compressive strength increases from 101.3 MPa to 129.7 MPa. The above explanation indicates that the increase of silica fume content

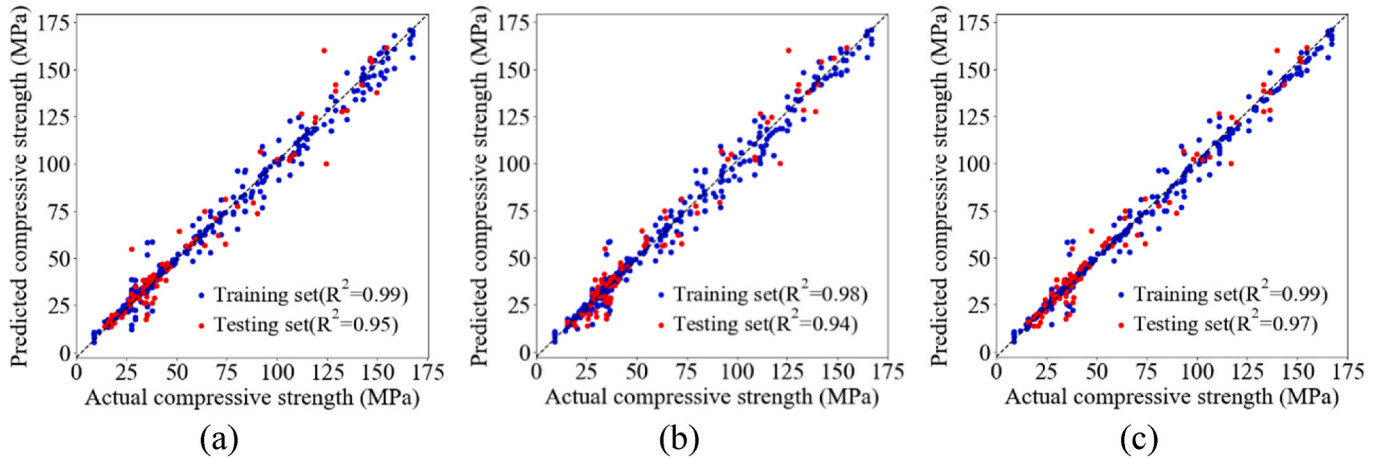


Fig. 9. Comparison of the prediction results versus the actual results for: (a) CatBoost; (b) random forest; and (c) XGBoost.

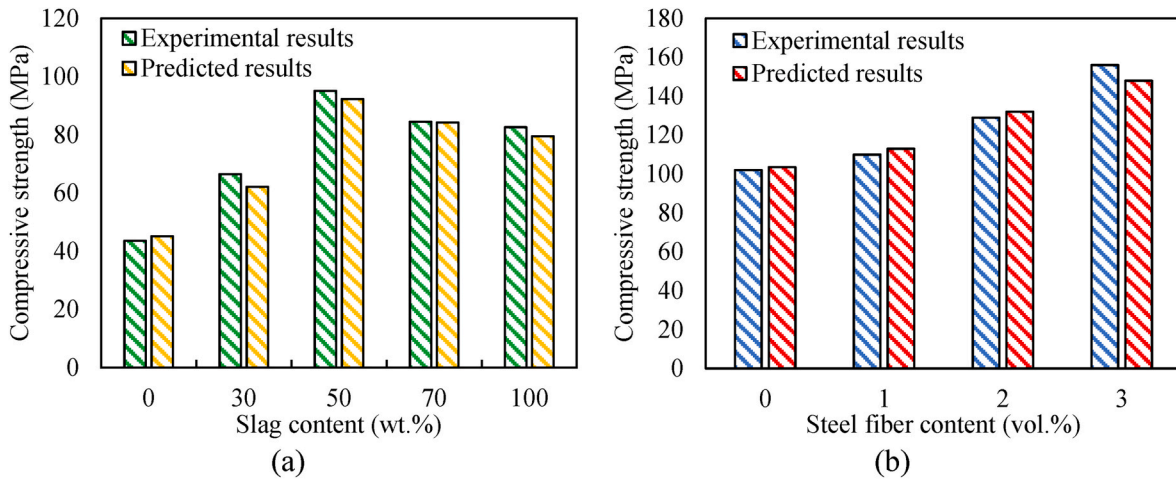


Fig. 10. Effects of different variables on the compressive strength: (a) slag content by the mass of geopolimer, and (b) steel fiber content by the volume of geopolimer.

Table 10

Physiochemical information of binder materials [66].

Oxides (%)	CaO	SiO ₂	Al ₂ O ₃	MgO	Fe ₂ O ₃	SO ₃	D ₅₀ (μm)
Slag	36.77	30.97	17.41	9.01	1.03	1.82	0.70
Silica fume	0.29	98.70	0.29	0.11	0.03	0.03	0.90

modifies the Ca/Si and Si/Al ratios in the binder system, thus increasing the compressive strength, since the modification of the ratios promotes the generation of C-(A)-S-H gel [76].

3.2.2. Case study 2

3.2.2.1. *Raw materials and mixture design.* In case study 2, geopolimer mixtures were prepared with fly ash and slag activated by NaOH solution. The physiochemical properties of the adopted fly ash and slag are shown in Table 12.

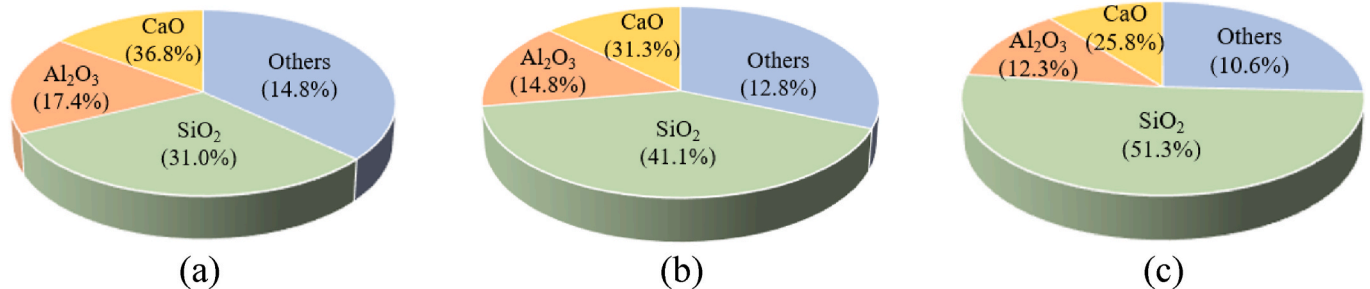


Fig. 11. Composition of the binders for different silica fume contents: (a) no silica fume; (b) 15 % silica fume by the mass of binders; and (c) 30 % silica fume by the mass of binders.

Table 11
Mixing proportions in case study 1.

Mixtures	Slag	Silica fume	Fine aggregate	NaOH	Na ₂ SiO ₃
1	1.00	0	1.20	0.12	0.18
2	0.85	0.15	1.20	0.12	0.18
3	0.70	0.30	1.20	0.12	0.18

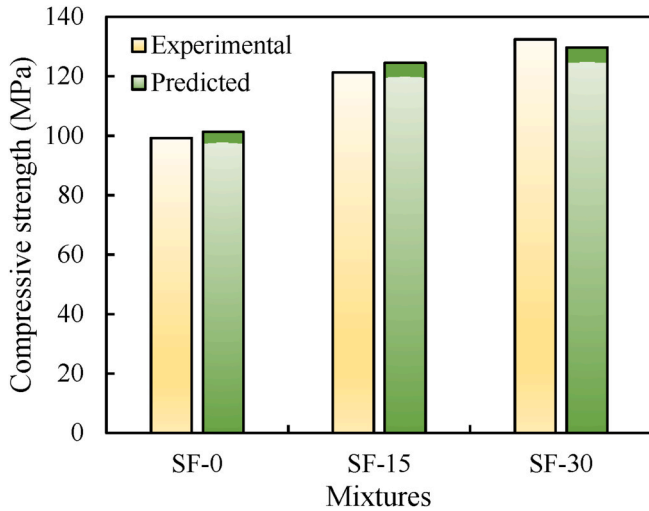


Fig. 12. Comparison of the predicted and experimental results of the compressive strength of geopolymer at 28 days. SF-15: silica fume accounts for 15 % of the binder by weight.

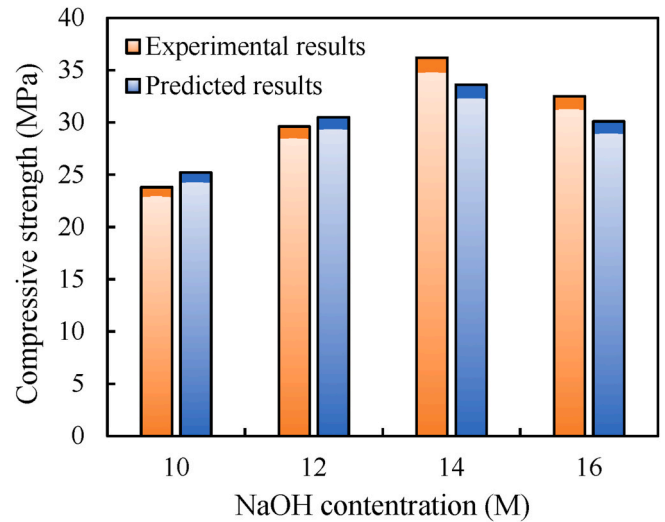


Fig. 13. Comparison of the predicted 28-day compressive strength and the experimental results of 28-day compressive strength for geopolymer concrete with different NaOH concentration.

3.2.2.3. Interpretation of prediction results. The interpretation of influencing pathways of NaOH concentration is shown in Fig. A3. The red solid lines indicate the primary factors that affect compressive strength. The directed arrows identified from the knowledge graph (Fig. A3) explicitly reveal the underlying mechanisms for the change of the compressive strength of geopolymer concrete. With the knowledge graph, the influencing pathways is determined automatically and utilized to list the following explanation:

NaOH concentration $\rightarrow$ $\begin{matrix} \text{High concentration} \\ \text{Excessive concentration} \end{matrix} \rightarrow \begin{matrix} \text{More C-(A)-S-H} \\ \text{Less C-(A)-S-H} \end{matrix} \rightarrow \text{Compressive strength}$

Regarding the mixing proportion, the fly ash and slag occupied 67 wt % and 33 wt% of the binder, respectively; the fine aggregate-to-binder ratio was 1.65; the coarse aggregate-to-binder ratio was 2.35; and the liquid-to-binder ratio was 0.35. The concentration of NaOH solution ranged from 10 M to 16 M to investigate the effect of NaOH concentration. Cylindrical geopolymer samples were cast, demolded after 24 h, and then cured in air at room temperature until testing. More details about the experiments are available in Ref. [15].

3.2.2.2. Prediction results. The XGBoost model was used to predict the compressive strength of geopolymer concrete. The comparison between the prediction and the experimental results is shown in Fig. 13. The results show that the prediction results agree with the experimental results, overall, indicating that the machine learning model can provide reasonable predictions. When the NaOH concentration increases from 10 M to 14 M, the 28-day compressive strength of geopolymer concrete increases from 25.2 MPa to 33.6 MPa. When the NaOH concentration increases from 14 M to 16 M, the 28-day compressive strength decreases from 33.6 MPa to 30.1 MPa.

Table 12
Physiochemical information of fly ash and slag [15].

Oxides (%)	CaO	SiO ₂	Al ₂ O ₃	MgO	Fe ₂ O ₃	SO ₃	D ₅₀ (μm)
Fly ash	3.54	54.76	26.41	0.78	8.48	1.20	17.2
Slag	33.23	32.26	16.35	8.29	3.53	1.32	12.3

When the NaOH concentration is increased from 12 M to 14 M, the compressive strength is improved because the dissolution of silica and aluminum in the binder system is promoted by the increase of pH value [13]. The higher dissolution rate of amorphous silica and aluminum in turn promotes geopolymerization, therefore producing more C-(A)-S-H gels and in turn increasing the compressive strength. When the NaOH is increased from 14 M to 16 M, the compressive strength is reduced because polycondensation is retarded, therefore producing less C-(A)-S-H gels and decreasing the compressive strength [77].

3.2.2.4. Generation of new knowledge. When machine learning results are beyond the scope of the knowledge graph, the knowledge graph cannot be utilized to interpret the results, but it can provide information to support the further development of the knowledge graph and generate new knowledge. In this research, the knowledge graph has been utilized for two other purposes besides the interpretation of machine learning data:

- (1) The knowledge graph has been used to guide the machine learning model to consider new mixtures, motivated by the compressive strength lower than 40 MPa (Fig. 13). To increase the compressive strength, the knowledge graph is used to identify the influencing factors (Fig. 14), such as the curing method and mixing proportions, which are important input variables of the machine learning model. The input variables are justified according to the knowledge graph, aiming at high compressive strengths and the influencing mechanisms.

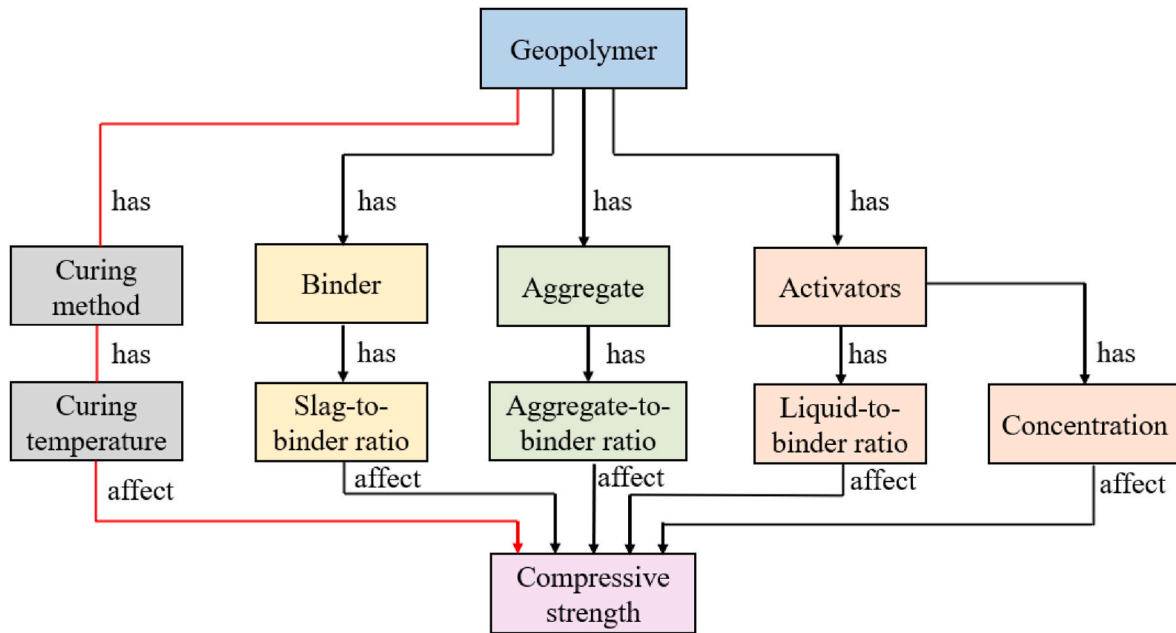


Fig. 14. Identification of key influencing factors for the compressive strengths of geopolymer.

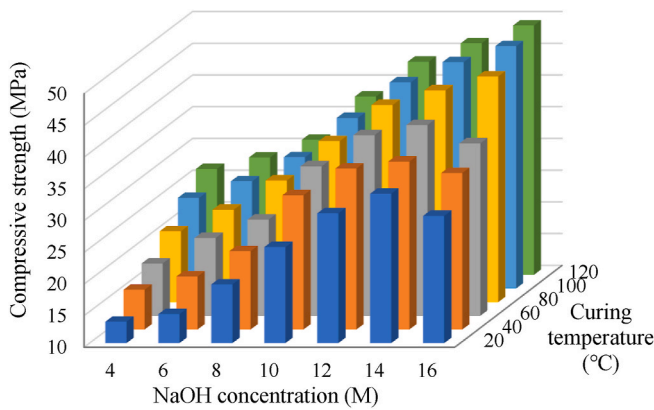


Fig. 15. Interaction of curing temperature and NaOH concentration on the compressive strength.

(2) The knowledge graph can be used to guide the generation of new knowledge. According to Fig. 14, both curing temperature and NaOH concentration influence the compressive strength. However, it is unknown whether the two variables have any coupling effect or not. This informs the need for evaluating the coupling effect of them using the machine learning model. The results indicate that the two variables have significant interactions, as shown in Fig. 15. This finding can be utilized to enrich the knowledge graph.

It should be noted that the underlying mechanism of the coupling effect between the curing temperature and NaOH concentration still needs to be uncovered through experimental research, and the research results can be added to the knowledge graph. The knowledge graph promotes the generation of new knowledge by facilitating the identification of uninterpretable observations or results from machine learning predictive models.

3.3. Multi-objective optimization

Multi-objective optimization has been performed to optimize the design of UHPG in this research. The method and results are provided

Table 13

Physiochemical properties of available binder materials [66].

Oxides (%)	CaO	SiO ₂	Al ₂ O ₃	MgO	Fe ₂ O ₃	SO ₃	D ₅₀ (μm)
Fly ash	1.24	41.22	31.18	3.40	2.15	0.05	17.2
Slag	38.81	33.81	14.78	9.50	0.76	0.06	12.3
Silica fume	1.84	95.38	0.05	0.50	0.30	0.00	0.15

based on a case study with available raw materials, which are fly ash, slag, and silica fume. Their physical and chemical properties are shown in Table 13. The curing condition is standard curing at room temperature (20 °C), and the size of specimen is the cube with a side length of 51 mm. To achieve UHPG, the 28-day compressive strength is defined as a design constraint, meaning the non-dominated solutions with 28-day compressive strengths lower than 120 MPa are not included in the solution set.

In this design optimization, three design objectives have been considered simultaneously, which are the material cost (F1), carbon

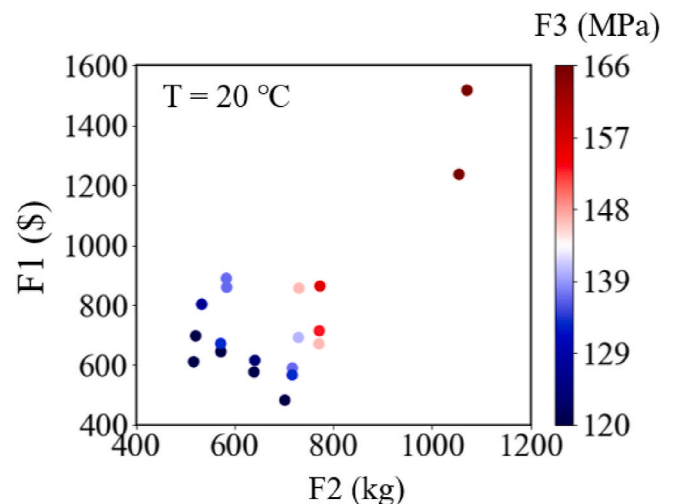


Fig. 16. Results of the nondominated solution set for the multi-objective optimization of UHPG.

emission (F2), and 28-day compressive strength (F3). The material cost and carbon emission are calculated based on the inventory data (Table 6) and the mixing proportion of each mixture, and the 28-day compressive strength is predicted using the trained machine learning model (section 3.1). With the multi-objective optimizer, the Pareto plot of the optimal solutions is shown in Fig. 16.

Multi-objective optimization often delivers a set of solutions, as shown in Fig. 17. TOPSIS has been used to recommend the best solution for specific applications, such as a solution which has low cost, low carbon emission, and high compressive strength. The mixture design is compared with representative UHPG mixtures, as shown in Table 14. When the 28-day compressive strength is sustained, the UHPG recommended by the AI-designer saves the cost by up to 19.8 % and reduces the carbon emission by up to 27.3 %, compared with typical UHPG mixtures [47].

The material cost and carbon emission of the different UHPG mixtures have been normalized by their compressive strengths at 28 days, as shown in Fig. 18. Compared with representative UHPG mixtures in references, the mixture designed by the proposed approach achieved lower cost and lower carbon emission normalized by the compressive strength, showing the efficacy of the proposed approach in designing UHPG.

3.4. Discussion on construction of knowledge graphs

In this study, the knowledge graph was manually constructed based on available references. Three challenges have been identified: (1) The manual construction process is time-consuming, and the constructed knowledge graph has a limited scope for particular applications. In this study, we primarily considered the particle size and chemical composition of the raw ingredients employed to produce geopolymer concrete. (2) The manually constructed knowledge graphs are influenced by historical data and human biases, which can compromise accuracy and reliability. As domain knowledge and available data continue growing, updating knowledge graphs becomes increasingly complex and challenging. (3) The knowledge graph is used for variable selection and interpreting prediction results. However, due to limitations in data availability, some minor factors that could affect material properties cannot be included in the knowledge graph. For instance, attributes, such as particle shape and surface morphology, influence compressive strength, but there is insufficient data to train machine learning models considering these attributes. The selection of information for

Table 14

Comparison of UHPG mixture designs.

Raw ingredients	AI-designer	UHPG-1 [47]	UHPG-2 [78]	UHPG-3 [66]
Slag (kg/m ³)	759	688	1000	665
Fly ash (kg/m ³)	123	172	160	0
Silica fume (kg/m ³)	72	45	100	0285
Fine aggregate (kg/m ³)	670	905	938	1140
Coarse aggregate (kg/m ³)	430	0	0	0
NaOH (kg/m ³)	55	45	74	114
Na ₂ SiO ₃ (kg/m ³)	178	314	480	171
Water (kg/m ³)	108	97	74	0
Steel fiber (kg/m ³)	117	156	236	156
NaOH concentration (M)	14	12	12	12
Material cost (\$/m ³)	887.4	1106.4	1625.9	1331.6
Carbon emission (kg/m ³)	766.1	1053.1	1599.8	973.9
Compressive strength (MPa)	156.1	149.7	151.2	132.4

constructing the knowledge graph is a challenge.

4. Conclusions

This paper presents a knowledge graph-guided AI approach for automatic discovery or design of geopolymer concrete, and the implementation into the design of UHPG is demonstrated. This research has been motivated by the utilization of domain knowledge and data in achieving efficient material discovery. Based on the above investigations, the following conclusions can be drawn:

- The incorporation of knowledge graph into the prediction-optimization framework of AI designer enables the AI designer to achieve interpretability of the prediction results from the machine learning model for geopolymer properties. The influencing pathways are automatically determined in the knowledge graph, explaining the underlying mechanisms of the effects of key design variables of geopolymer. This capability transforms black-box machine learning models into explainable machine learning models that can be assessed by geopolymer or concrete domain experts.
- The knowledge graph constructed by geopolymer domain experts can be utilized to guide the development of machine learning

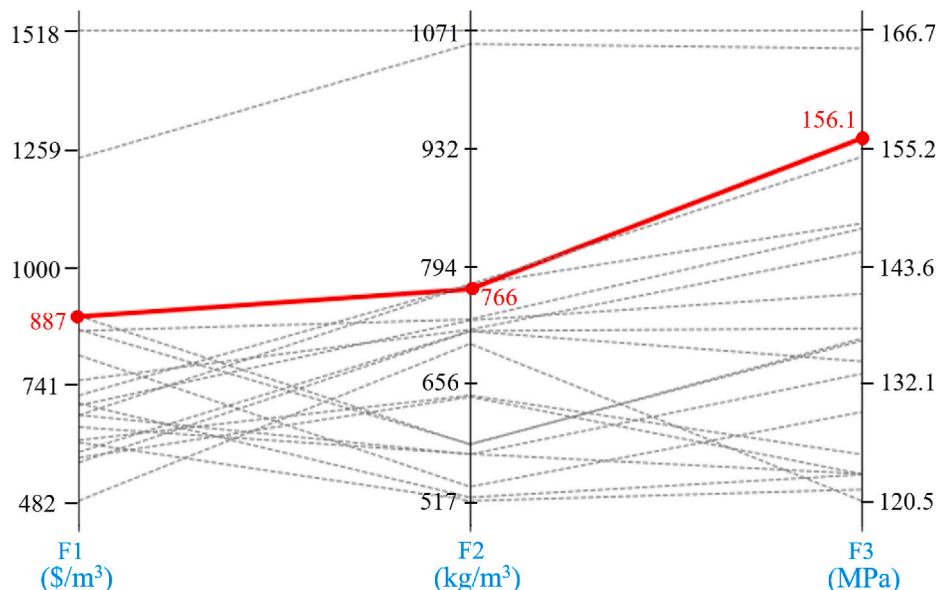


Fig. 17. Non-dominated solutions (gray color) and a solution recommended by TOPSIS (red color).

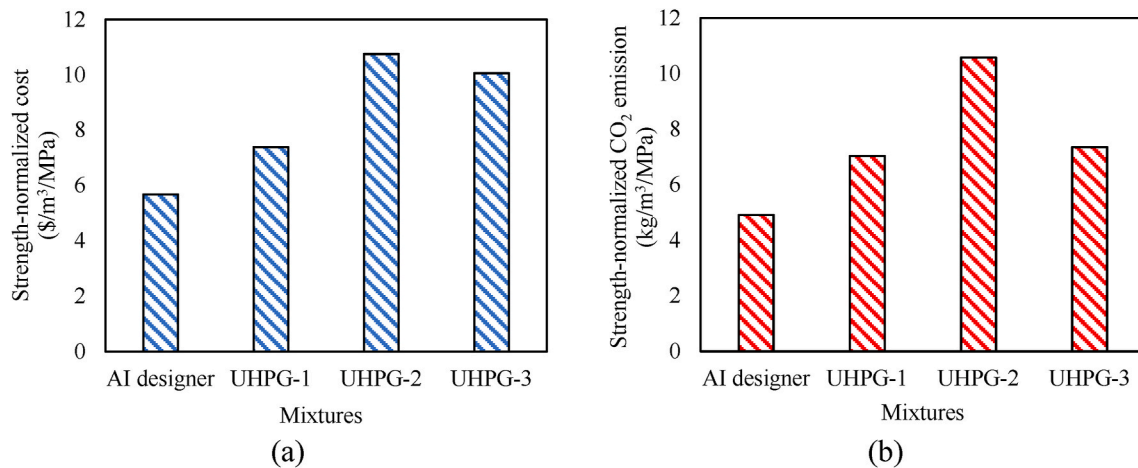


Fig. 18. UHPG results: (a) strength-normalized cost; and (b) strength-normalized carbon emission.

models. Particularly, the knowledge graph can be used to identify key design variables of geopolymer, ensuring that the key design variables are fully considered in establishing the machine learning models. This capability enhances the rationality and reliability of machine learning predictive models.

- The integration of knowledge graph and machine learning models enables the growth of the knowledge graph and offers a new pathway to generate new knowledge. Important unknown knowledge can be identified from the knowledge graph, and machine learning models can provide quantitative predictions to facilitate the generation of new knowledge via further experimental research.
- The proposed approach has been implemented and utilized to discover the promising design of UHPG with high compressive strength, low material cost, and low carbon emission. Other geopolymer mixtures can be designed following the same procedure while adjusting the design objectives for different applications.
- The physical and chemical properties of raw materials are important information that must be considered in the machine learning predictive model, aimed at handling the significant variations of solid wastes. The consideration of the physical and chemical properties of raw materials imparts unprecedented reliability, generalizability, and transferability to the machine learning model, facilitating the utilization of various solid wastes.

Although the proposed approach has demonstrated important advantages in material discovery, this research is still at the feasibility study stage with a relatively low technology readiness level. The following limitations have been identified to promote future research toward the maturation of knowledge-guided AI design of geopolymer concrete:

- Developing advanced computing techniques to automate the construction of knowledge graphs will largely improve the efficiency and application scope of the proposed approach. The achievement of automation can mitigate the influence of historical data and human biases, enhancing the reliability of knowledge graphs. Additionally, methods for selecting relevant information from literature are useful to identify and exclude minor factors.
- As domain knowledge and available data continue growing, innovative solutions for managing and updating knowledge graphs are

crucial for scalable applications. These advances will enable the construction of comprehensive knowledge graphs for the design of geopolymer concrete and other types of materials, and the knowledge graphs are useful for developing more capable AI designers for geopolymer concrete and other materials.

- The quantity of available data is limited, which imposes challenges to the evaluation of the performance of the proposed approach, such as the evaluation of the predicted results and the optimization results of geopolymer design. It is important to develop innovative techniques to automate the collection of data from available references. More experiments can be conducted to generate high-quality data to validate and improve the approach.

CRediT authorship contribution statement

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

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.

Data availability

Data will be made available on request.

Acknowledgement

This research was funded by the United States National Science Foundation [award numbers: CMMI-2046407 and CPS-2305882], the United States Department of Transportation [award number: 693JK32310008POTA], and the United States National Oceanic and Atmospheric Administration [award number: NA23OAR4170169].

16

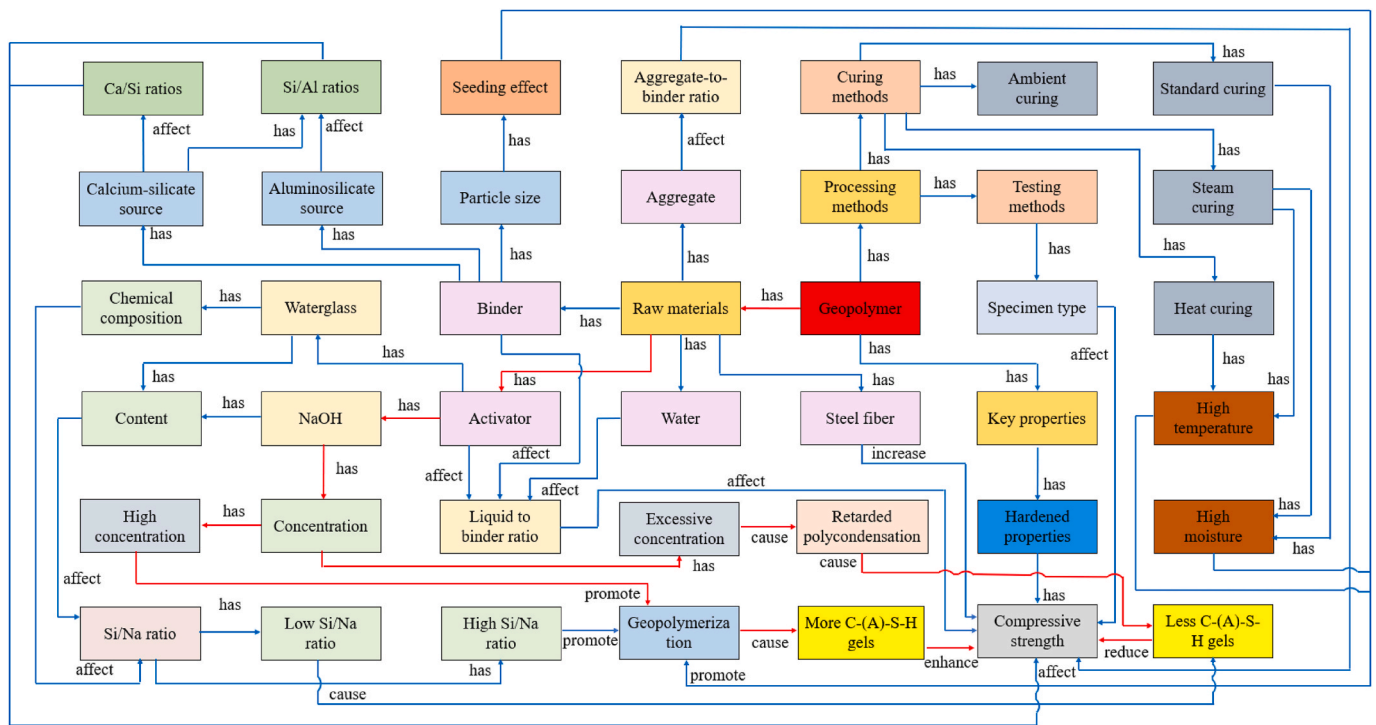


Fig. A3. Knowledge graph used to interpret the prediction results in case study 2.

References

- [1] Unsustainable: concrete and cement. <https://2150-vc.medium.com>, 2022.
- [2] G. Kilgore, Carbon Footprint of Building Materials, 2023. <https://8billiontrees.co/m/carbon-offsets-credits/carbon-footprint-of-building-materials>.
- [3] A. Guterres, Carbon neutrality by 2050: the world's most urgent mission. <https://www.un.org/sg/en/content/sg/articles/2020-12-11/carbon-neutrality-2050-the-world%E2%80%99s-most-urgent-mission>, 2020.
- [4] Y. Liu, C. Lu, X. Hu, C. Shi, Effect of silica fume on rheology of slag-fly ash-silica fume-based geopolymer pastes with different activators, *Cement Concr. Res.* 174 (2023) 107336, <https://doi.org/10.1016/j.cemconres.2023.107336>.
- [5] G.M. Zannerni, K.P. Fattah, A.K. Al-Tamimi, Ambient-cured geopolymer concrete with single alkali activator, *Sustain. Mater. Tech.* 23 (2020) e00131, <https://doi.org/10.1016/j.susmat.2019.e00131>.
- [6] A. Esparham, F. Ghalatian, The features of geopolymer concrete as a novel approach for utilization in green urban structures, *J. Compos. Comp.* 4 (11) (2022) 89–96, <https://doi.org/10.52547/jcc.4.2.4>.
- [7] M. Zhang, H. Xu, A.L.P. Zeze, X. Liu, M. Tao, Coating performance, durability and anti-corrosion mechanism of organic modified geopolymer composite for marine concrete protection, *Cement Concr. Compos.* 129 (2022) 104495, <https://doi.org/10.1016/j.cemconcomp.2022.104495>.
- [8] M. Amran, S.-S. Huang, S. Debbarma, R.S. Rashid, Fire resistance of geopolymer concrete: a critical review, *Construct. Build. Mater.* 324 (2022) 126722, <https://doi.org/10.1016/j.cscm.2021.e00661>.
- [9] R. Bajpai, K. Choudhary, A. Srivastava, K.S. Sangwan, M. Singh, Environmental impact assessment of fly ash and silica fume based geopolymer concrete, *J. Clean. Prod.* 254 (2020) 120147, <https://doi.org/10.1016/j.jclepro.2020.120147>.
- [10] A.K. Sinha, S. Talukdar, Enhancement of the properties of silicate activated ultrafine-slag based geopolymer mortar using retarder, *Construct. Build. Mater.* 313 (2021) 125380, <https://doi.org/10.1016/j.conbuildmat.2021.125380>.
- [11] H.H. Nguyễn, P.H. Nguyễn, Q.-H. Lương, W. Meng, B.Y. Lee, Mechanical and autogenous healing properties of high-strength and ultra-ductility engineered geopolymer composites reinforced by PE-PVA hybrid fibers, *Cement Concr. Compos.* 142 (2023) 105155, <https://doi.org/10.1016/j.cemconcomp.2023.105155>.
- [12] D. Hardjito, S.E. Wallah, D.M. Sumajouw, B.V. Rangan, On the development of fly ash-based geopolymer concrete, *ACI Mater. J.* 101 (6) (2004) 467–472, <https://doi.org/10.14359/13485>.
- [13] N.B. Singh, Fly ash-based geopolymer binder: a future construction material, *Miner. Eng.* 8 (7) (2018) 299, <https://doi.org/10.3390/min8070299>.
- [14] R. Asghar, M.A. Khan, R. Alyousef, M.F. Javed, M. Ali, Promoting the green construction: scientometric review on the mechanical and structural performance of geopolymer concrete, *Construct. Build. Mater.* 368 (2023) 130502, <https://doi.org/10.1016/j.conbuildmat.2023.130502>.
- [15] S.V. Patankar, Y.M. Ghugal, S.S. Jamkar, Effect of concentration of sodium hydroxide and degree of heat curing on fly ash-based geopolymer mortar, *Indian J. Mater. Sci.* 2014 (2014) 1–6, <https://doi.org/10.1155/2014/938789>.
- [16] K. Somna, C. Jaturapitakkul, P. Kajitvichyanukul, P. Chindaprasit, NaOH-activated ground fly ash geopolymer cured at ambient temperature, *Fuel* 90 (6) (2011) 2118–2124, <https://doi.org/10.1016/j.fuel.2011.01.018>.
- [17] B. Singh, M. Rahman, R. Paswan, S. Bhattacharyya, Effect of activator concentration on the strength, ITZ and drying shrinkage of fly ash/slag geopolymer concrete, *Construct. Build. Mater.* 118 (2016) 171–179, <https://doi.org/10.1016/j.conbuildmat.2016.05.008>.
- [18] S. Witzleben, Minimizing the global warming potential with geopolymer-based insulation material with Miscanthus fiber, *Polymers* 14 (15) (2022) 3191, <https://doi.org/10.3390/polym14153191>.
- [19] X. Ge, Y. Liu, Y. Mao, X. Hu, C. Shi, Characteristics of fly ash-based geopolymer concrete in the field for 4 years, *Construct. Build. Mater.* 382 (2023) 131222, <https://doi.org/10.1016/j.conbuildmat.2023.131222>.
- [20] X. Ge, X. Hu, C. Shi, Impact of micro characteristics on the formation of high-strength Class F fly ash-based geopolymers cured at ambient conditions, *Construct. Build. Mater.* 352 (2022) 129074, <https://doi.org/10.1016/j.conbuildmat.2022.129074>.
- [21] L.N. Assi, E.E. Deaver, P.J.C. Ziehl, Effect of source and particle size distribution on the mechanical and microstructural properties of fly Ash-Based geopolymer concrete, *Construct. Build. Mater.* 167 (2018) 372–380, <https://doi.org/10.1016/j.conbuildmat.2018.01.193>.
- [22] 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>.
- [23] 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>.
- [24] 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>.
- [25] 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>.
- [26] 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>.

- [27] P. Guo, W. Meng, Y. Bao, Automatic identification and quantification of dense microcracks in high-performance fiber-reinforced cementitious composites through deep learning-based computer vision, *Cement Concr. Res.* 148 (2021) 106532, <https://doi.org/10.1016/j.cemconres.2021.106532>.
- [28] P. Guo, X. Meng, W. Meng, Y. Bao, Monitoring and automatic characterization of cracks in strain-hardening cementitious composite (SHCC) through intelligent interpretation of photos, *Compos. B Eng.* 242 (2022) 110096, <https://doi.org/10.1016/j.compositesb.2022.110096>.
- [29] 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.jobbe.2023.106070>.
- [30] H. Zhu, X. Wu, Y. Luo, Y. Jia, C. Wang, Z. Fang, X. Zhuang, S. Zhou, Prediction of early compressive strength of ultrahigh-performance concrete using machine learning methods, *Int. J. Comput. Methods* (2022) 2141023, <https://doi.org/10.1016/j.cmm.2022.214527>.
- [31] 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).
- [32] 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>.
- [33] Z. Wu, C. Shi, K.H. Khayat, Investigation of mechanical properties and shrinkage of ultra-high performance concrete: influence of steel fiber content and shape, *Compos. B Eng.* 174 (2019) 107021, <https://doi.org/10.1016/j.compositesb.2019.107021>.
- [34] V. Corinaldesi, G. Moriconi, Mechanical and thermal evaluation of ultra high performance fiber reinforced concretes for engineering applications, *Construct. Build. Mater.* 26 (1) (2012) 289–294, <https://doi.org/10.1016/j.conbuildmat.2011.06.023>.
- [35] W. Meng, K.H. Khayat, Effect of hybrid fibers on fresh properties, mechanical properties, and autogenous shrinkage of cost-effective UHPC, *ACI 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).
- [36] 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>.
- [37] L. Li, P. Wang, J. Yan, Y. Wang, S. Li, J. Jiang, Z. Sun, B. Tang, T.-H. Chang, S. Wang, Real-world data medical knowledge graph: construction and applications, *Artif. Intell. Med.* 103 (2020) 101817, <https://doi.org/10.1016/j.artmed.2020.101817>.
- [38] 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.
- [39] N. Kertkeidkachorn, R. Ichise, An automatic knowledge graph creation framework from natural language text, *IEICE Trans. Info Syst.* 101 (1) (2018) 90–98, <https://doi.org/10.1587/transinf.2017SWP0006>.
- [40] Y. Jia, Y. Qi, H. Shang, R. Jiang, A. Li, A practical approach to constructing a knowledge graph for cybersecurity, *Engineering* 4 (1) (2018) 53–60, <https://doi.org/10.1016/j.eng.2018.01.004>.
- [41] X. Zou, A survey on application of knowledge graph, *J. Phys. Conf.* 1487 (1) (2020) 012016, <https://doi.org/10.1088/1742-6596/1487/1/012016>.
- [42] P. Guo, W. Meng, Y. Bao, Knowledge graph-guided data-driven design of ultra-high-performance concrete (UHPC) with interpretability and physicochemical reaction discovery capability, *Construct. Build. Mater.* 430 (2024) p.136502, <https://doi.org/10.1016/j.conbuildmat.2024.136502>.
- [43] Q. Song, R. Yu, Z. Shui, X. Wang, S. Rao, Z. Lin, Optimization of fibre orientation and distribution for a sustainable ultra-high performance fibre reinforced concrete (UHPFRC): experiments and mechanism analysis, *Construct. Build. Mater.* 169 (2018) 8–19, <https://doi.org/10.1016/j.conbuildmat.2018.02.130>.
- [44] 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, *Arabian J. Geosci.* 15 (4) (2022) 345, <https://doi.org/10.1007/s12517-022-09617-y>.
- [45] J. Du, P. Guo, W. Meng, Effect of water-based nanoclay and ambient temperature on rheological properties of UHPC pastes, *Construct. Build. Mater.* 370 (2023) 130733, <https://doi.org/10.1016/j.conbuildmat.2023.130733>.
- [46] 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. Hydraulic Eng.* 27 (sup1) (2021) 343–353, <https://doi.org/10.1080/09715010.2019.1653799>.
- [47] Y. Liu, Z. Zhang, C. Shi, D. Zhu, N. Li, Y. Deng, Development of ultra-high performance geopolymer concrete (UHPGC): influence of steel fiber on mechanical properties, *Cement Concr. Compos.* 112 (2020) 103670, <https://doi.org/10.1016/j.cemconcomp.2020.103670>.
- [48] A.M. Tahwia, M. Abd Ellatif, A.M. Heneigel, M. Abd Elrahman, Characteristics of eco-friendly ultra-high-performance geopolymer concrete incorporating waste materials, *Ceram. Int.* 48 (14) (2022) 19662–19674, <https://doi.org/10.1016/j.ceramint.2022.03.103>.
- [49] 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>.
- [50] ASTM C109/C109M-20, Standard test method for compressive strength of hydraulic cement mortars, https://doi.org/10.1520/C0109_C0109M-20, 2020.
- [51] BS EN 12390-3:2002 Testing hardened concrete. Compressive strength of test specimens DOI: <https://doi.org/10.3403/BSEN12390>.
- [52] GBT 17671-1999, Method of testing cements - determination of strength, <https://www.chinesestandard.net/PDF.aspx/GBT17671-1999>, 1999.
- [53] ASTM C39/C39M-21, Standard Test Method for Compressive Strength of Cylindrical Concrete Specimens, 2023, https://doi.org/10.1520/C0039_C0039M-21.
- [54] 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>.
- [55] A.H. Patoary, M.J.B. Kibria, A. Kaium, Implementation of automated Bengali parts of speech tagger: an approach using deep learning algorithm, in: 2020 IEEE Region 10 Symposium (TENSYP), 2020, pp. 308–311, <https://doi.org/10.1109/TENSYP50017.2020.9230907>.
- [56] N. Bjorck, C.P. Gomes, B. Selman, K.Q. Weinberger, Understanding batch normalization, *Adv. Neural Inf. Process. Syst.* 31 (2018), <https://arxiv.org/abs/1806.02375>.
- [57] L. Xu, D. Fan, K. Liu, W. Xu, R. Yu, A machine learning framework for intelligent development of Ultra-High performance concrete (UHPC): from dataset cleaning to performance predicting, *Expert Syst. Appl.* 242 (2024) 122790, <https://doi.org/10.1016/j.eswa.2023.122790>.
- [58] A.H. Victoria, G. Maragatham, Automatic tuning of hyperparameters using Bayesian optimization, *Evolving Systems* 12 (2021) 217–223, <https://doi.org/10.1007/s12530-020-09345-2>.
- [59] Cost of electricity in New Jersey, <https://www.energysage.com/local-data/electricity-cost/nj/>, 2023.
- [60] How much carbon dioxide is produced per kilowattour of U.S. electricity generation? (2021), <https://www.eia.gov/tools/faqs/faq.php?id=74&t=11>.
- [61] A. Alsaman, C.N. Dang, J.R. Marti-Vargas, W.M. Hale, Mixture-proportioning of economical UHPC mixtures, *J. Build. Eng.* 27 (2020) 100970, <https://doi.org/10.1016/j.jobbe.2019.100970>.
- [62] S.A. Miller, Supplementary cementitious materials to mitigate greenhouse gas emissions from concrete: can there be too much of a good thing? *J. Clean. Prod.* 178 (2018) 587–598, <https://doi.org/10.1016/j.jclepro.2018.01.008>.
- [63] D. Zhang, B. Jaworska, H. Zhu, K. Dahlquist, V.C. Li, Engineered Cementitious Composites (ECC) with limestone calcined clay cement (LC3), *Cement Concr. Compos.* 114 (2020) 103766, <https://doi.org/10.1016/j.cemconcomp.2020.103766>.
- [64] G. Habert, J.D.E. De Lacaillerie, N. Roussel, An environmental evaluation of geopolymer based concrete production: reviewing current research trends, *J. Clean. Prod.* 19 (11) (2011) 1229–1238, <https://doi.org/10.1016/j.jclepro.2011.03.012>.
- [65] X. Sun, Y. Zhao, Y. Tian, P. Wu, Z. Guo, J. Qiu, J. Xing, G. Xiaowei, Modification of high-volume fly ash cement with metakaolin for its utilization in cemented paste backfill: the effects of metakaolin content and particle size, *Powder Technol.* 393 (2021) 539–549, <https://doi.org/10.1016/j.powtec.2021.07.067>.
- [66] P. Kathirvel, S. Sreekumaran, Sustainable development of ultra high performance concrete using geopolymer technology, *J. Build. Eng.* 39 (2021) 102267, <https://doi.org/10.1016/j.jobbe.2021.102267>.
- [67] K. Wille, C. Boisvert-Cotulio, Material efficiency in the design of ultra-high performance concrete, *Construct. Build. Mater.* 86 (2015) 33–43, <https://doi.org/10.1016/j.conbuildmat.2015.03.087>.
- [68] J. Chen, P.-L. Ng, R. Jaskulski, W. Kubissa, Use of quartz sand to produce low embodied energy and carbon footprint plaster, *J. Sustain. Architect. Civ. Eng.* 21 (4) (2017) 75–81, <https://doi.org/10.5755/J01.sace.21.4.20005>.
- [69] M.S. Hameed, A. Sekar, Properties of green concrete containing quarry rock dust and marble sludge powder as fine aggregate, *J. Eng. Appl. Sci.* 4 (4) (2009) 83–89, <https://doi.org/10.17577/IJERTCONV4IS23062>.
- [70] Z. Dong, H. Tan, J. Yu, F. Liu, A feasibility study on Engineered cementitious Composites mixed with coarse aggregate, *Construct. Build. Mater.* 350 (2022) 128587, <https://doi.org/10.1016/j.conbuildmat.2022.128587>.
- [71] M. Adamu, B.S. Mohammed, M.S. Liew, Mechanical properties and performance of high volume fly ash roller compacted concrete containing crumb rubber and nano silica, *Construct. Build. Mater.* 171 (2018) 521–538, <https://doi.org/10.1016/j.conbuildmat.2018.03.138>.
- [72] G. Long, Y. Gao, Y. Xie, Designing more sustainable and greener self-compacting concrete, *Construct. Build. Mater.* 84 (2015) 301–306, <https://doi.org/10.1016/j.conbuildmat.2015.02.072>.
- [73] R. Chan, M.A. Santana, A.M. Oda, R.C. Paniguel, L.B. Vieira, A.D. Figueiredo, I. Galobardes, Analysis of potential use of fibre reinforced recycled aggregate concrete for sustainable pavements, *J. Clean. Prod.* 218 (2019) 183–191, <https://doi.org/10.1016/j.jclepro.2019.01.221>.
- [74] 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>.
- [75] D. Ozturk, F. Batuk, Technique for order preference by similarity to ideal solution (TOPSIS) for spatial decision problems, *Proceedings ISPRS* 1 (4) (2011), https://doi.org/10.1007/978-3-030-64765-0_4.
- [76] M. Jin, Y. Ma, W. Li, J. Huang, Y. Yan, H. Zeng, C. Lu, J. Liu, Multi-scale investigation on composition-structure of C-(A)-SH with different Al/Si ratios under attack of decalcification action, *Cement Concr. Res.* 172 (2023) 107251, <https://doi.org/10.1016/j.cemconres.2023.107251>.
- [77] K. Pimraksa, P. Chindaprasit, A. Rungeth, K. Sagoe-Crentsil, T. Sato, Lightweight geopolymer made of highly porous siliceous materials with various Na₂O/Al₂O₃ and SiO₂/Al₂O₃ ratios, *Mater. Sci. Eng., A* 528 (21) (2011) 6616–6623, <https://doi.org/10.1016/j.msea.2011.04.044>.
- [78] S. Xu, P. Yuan, J. Liu, Z. Pan, Z. Liu, Y. Su, J. Li, C. Wu, Development and preliminary mix design of ultra-high-performance concrete based on geopolymer, *Construct. Build. Mater.* 308 (2021) 125110, <https://doi.org/10.1016/j.conbuildmat.2021.125110>.