



AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC)

Soroush Mahjoubi, Rojyar Barhemat, Weina Meng, Yi Bao *

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

ARTICLE INFO

Keywords:

Carbon footprint
Embodied energy
Machine learning
Many-objective optimization
Semi-supervised learning
Synthetic tabular data
Ultra-high performance concrete (UHPC)

ABSTRACT

This paper presents an AI-guided approach to automatically discover low-carbon cost-effective ultra-high performance concrete (UHPC). The presented approach automates data augmentation, machine learning model generation, and mixture selection by integrating advanced techniques of generative modeling, automated machine learning, and many-objective optimization. New data are synthesized by generative modeling and semi-supervised learning to enlarge datasets for training machine learning models that are automatically generated to predict the compressive strength, flexural strength, mini-slump spread, and porosity of UHPC. The proposed approach was used to explore new UHPC mixtures in two design scenarios with different objectives. The first scenario maximizes the compressive and flexural strengths and minimizes porosity while retaining self-consolidation. The second scenario minimizes the life-cycle carbon footprint, embodied energy, and material cost, besides the objectives of the first scenario. The life-cycle carbon footprint, embodied energy, and material cost of the UHPC in the second scenario are respectively reduced by 73%, 71%, and 80%, compared with the UHPC in the first scenario. This research advances the capability of developing cementitious composites using AI-guided approaches.

1. Introduction

Ultra-high performance concrete (UHPC) is a family of advanced cementitious composites featuring self-consolidation, superior mechanical properties, and long-term durability (Meng and Khayat, 2018). The 28-day compressive strength of UHPC exceeds 120 MPa under standard curing conditions (Du et al., 2021; Meng et al., 2018). The high strength is associated with the dense microstructure due to high particle packing density and low porosity. UHPC also has high tensile and flexural strengths due to the use of chopped fibers dispersed in the cementitious matrix, and the fibers provide crack-bridging effects (Karim and Shafei, 2021). Given the dense microstructure and discontinuous pore network, UHPC possesses exceptional durability (Ahlborn et al., 2008; Lu et al., 2021). Due to the extraordinary properties, UHPC attracted increasing attention in the past decade. UHPC has been successfully used in new construction (Gaudillière et al., 2018; Muttoni et al., 2013) and rehabilitation of existing civil engineering structures (Aaleti et al., 2013; Doiron, 2016; Hain et al., 2019; Kennedy et al., 2015).

A main drawback that hinders wider acceptance of UHPC in engineering practices is the high material cost and carbon footprint, due to

use of costly and high-carbon raw materials such as steel fibers, cement, and organic admixtures. Many studies were conducted to develop low-carbon cost-effective UHPC mixtures by using alternative materials. For example, Meng et al. (2018) developed four UHPC mixtures using local river sand, masonry sand, and supplementary cementitious materials (SCMs) such as fly ash and slag. Test results showed that the developed UHPC mixtures delivered high mechanical properties and low cost as well as low carbon footprint and low embodied energy. Consistent test results of UHPC were obtained in other studies (Meng et al., 2016; Mosaberpanah et al., 2019; Wille et al., 2011). The tests were designed based on extensive experimental efforts and long time for those experiments that involved necessary time for curing concrete. trial-and-error experiments of UHPC mixtures are usually time-consuming for two reasons. First, the standard tests of mechanical properties such as the compressive strength and the flexural properties are performed after the UHPC specimens are cured for 28 days according to ASTM C109 and ASTM C1609. The durability tests such as external sulfate attack test can take more than 6 months according to ASTM C1012. Second, there are many mixture design variables such as the water-to-binder ratio and the sand-to-binder ratio. Those variables have

* Corresponding author.

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

<https://doi.org/10.1016/j.resconrec.2022.106741>

Received 30 January 2022; Received in revised form 29 August 2022; Accepted 27 October 2022

Available online 2 November 2022

0921-3449/© 2022 Elsevier B.V. All rights reserved.

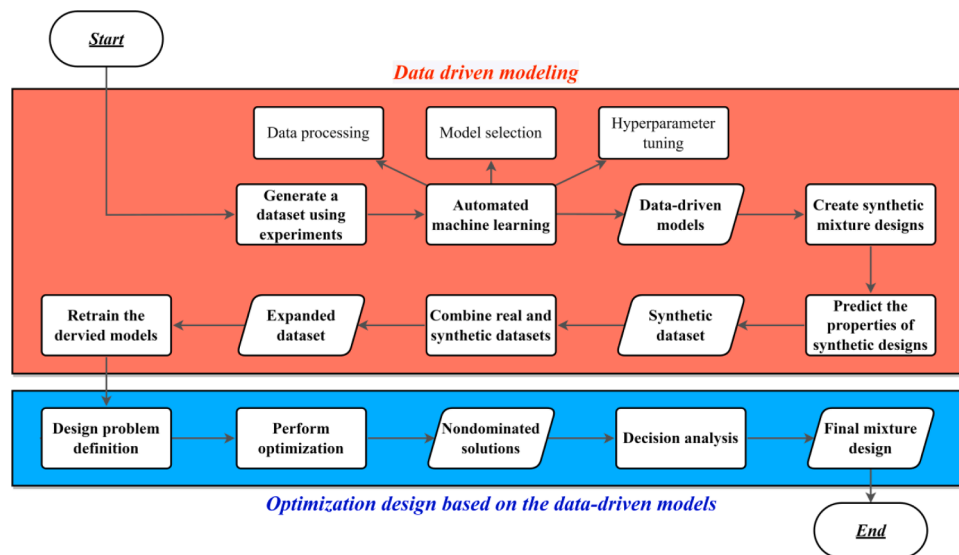


Fig. 1. Flowchart of the proposed method for auto-discovery of low-carbon cost-effective UHPC.

significant effects on the key properties such as the fresh and the mechanical properties of UHPC. It is time-consuming to test the effects of those variables on the different properties. There will be more experiments required when the coupling effects between different variables are considered. Considering tradeoffs between different properties and material cost, developing new UHPC is relevant in engineering projects with different desired properties, while the low efficiency and high cost of developing new UHPC based on intensive experiments have stalled their adoption in engineering applications.

An alternative approach to develop UHPC is to use data-driven machine learning models that are trained to predict UHPC properties based on calibrated relationship between design variables and properties (Mahjoubi et al., 2022; Sun et al., 2021). With the predictive model, an optimization algorithm can be incorporated to optimize the design variables such as the water-to-cement and sand-to-binder ratios (Mahjoubi et al., 2021). A few studies have been conducted to develop UHPC based on machine learning. Fan et al. (2021a) trained a second-order polynomial regression model using 19 experimental data to predict the packing density of UHPC. Ghafari et al. (2015) trained artificial neural networks using 53 experimental data to predict the compressive strength and flowability of UHPC. Abellán-García and Guzmán-Guzmán (2021) trained two random forest models using 600 experimental data to predict the energy absorption capacity and ultimate strain capacity of UHPC. Sadrossadat et al. (2021) trained artificial neural networks using 53 experimental data to predict the compressive and flowability of UHPC. Fan et al. (2020, 2021b) trained an artificial neural network to predict the particle packing density, compressive strength, and flexural strength of UHPC. The dataset had 80 experimental data in reference (Fan et al., 2021b) and 26 experimental data in reference (Fan et al., 2020). A concrete mixture was designed using the Modified Andreasen and Andersen dry particle packing model (Funk and Dinger, 2013) and the least-square method (Björck, 1990). The limestone powder content of the UHPC mixture was optimized, aiming to maximize the particle packing density.

Previous research showed that data-driven models were promising to predict properties, and optimization algorithms minimized the material cost of UHPC and other types of concrete. Four major limitations were identified from previous studies: (1) There is lack of data for training machine learning models. This is a common challenge for AI-guided material discovery. The dilemma is that (i) when the dataset is small, machine learning models have low accuracy and generalizability; and (ii) when the dataset is large, meaning that development of the material is mature, the significance of machine learning-based new material

discovery becomes limited. (2) Simple machine learning models were used to regress high-dimensional relations of different types of concrete with many variables. Each model considered limited types of variables, making the model inapplicable to other cases with different raw materials and variables. (3) It is difficult for people who do not have expertise in machine learning to generate a machine learning model with good performance. The tasks, such as data preprocessing, model selection, and hyperparameter tuning, of developing a machine learning model requires special knowledge in machine learning. (4) Most existing efforts of optimizing UHPC focused on maximizing the mechanical properties or minimizing the material cost of UHPC by optimizing one or two mixture design variables while the other variables were kept constant. It is unclear how to discover low-carbon low-cost UHPC mixtures while retaining the desired mechanical properties, workability, and durability. These limitations represent major technical challenges of machine learning methods for prediction of concrete properties.

This study intends to address these limitations by developing an approach that integrates machine intelligence and evolutionary many-objective optimization for auto-discovery of UHPC. There are four main research objectives: (1) to establish a framework to automatically optimize the mechanical, flowability, durability, economic, and ecological properties; (2) to develop a data synthesis method based on generative modeling and semi-supervised learning to enlarge datasets for improving accuracy and generalizability. Semi-supervised learning is involved to derive output variables by regression models instead of generative models. (3) to develop high-fidelity machine learning models to predict the compressive strength, flexural strength, mini-slump spread, and porosity of UHPC; and (4) to automatically discover new UHPC mixtures with desired properties for many objectives.

This research is novel in four aspects: (1) An automated machine learning (AutoML) approach is presented based on the Microsoft Azure Copeland et al., 2015; Fusi et al., 2018) which performs data preprocessing, model selection, hyperparameter tuning, automatically. (2) An approach is developed to synthesize artificial, yet reasonable data based on a conditional generative adversarial network (Mirza and Osindero, 2014; Xu et al., 2019), Copula generative adversarial network (Kamthe et al., 2021), Gaussian Copula (Patki et al., 2016), and variational autoencoder (Kingma and Welling, 2013). The synthetic data are used to supplement test data and boost the accuracy and generalizability of predictive models. (3) Key mechanical properties (i.e., 28-day compressive strength and tensile strength), workability (i.e., mini-slump spread), and porosity were considered in the AI-guided design process of UHPC. (4) The predictive models are integrated with

Table 1

Summary statistics of the variables corresponding to the compressive strength dataset.

Number	Variable	Unit	Range	Mean	Skew. ¹	Kurt. ²
1	Cement-to-cm ^{3,4}	1	0.17–1	0.76	-1.02	1.12
2	Cement type	MPa	42.5, 52.5	48.1	-0.23	-1.96
3	Fly ash-to-cm	1	0–0.55	0.05	2.07	3.60
4	Slag-to-cm	1	0–0.70	0.04	2.92	9.00
5	Silica fume-to-cm	1	0–0.25	0.13	-0.24	-1.38
6	Metakaolin-to-cm	1	0–0.29	0.01	4.19	19.37
7	Nano silica-to-cm	1	0–0.17	0.01	3.53	20.24
8	Limestone-to-cm	1	0–2.64	0.08	6.40	63.39
9	Quartz powder-to-cm	1	0–0.46	0.06	1.49	1.10
10	Sand-to-cm	1	0–2.90	1.18	0.90	1.83
11	Maximum aggregate size	mm	0.10–5	1.76	1.28	1.49
12	Water-to-cm	1	0.12–0.5	0.20	2.19	6.95
13	Superplasticizer-to-cm	1	0–0.15	0.04	2.31	5.61
14	Steel fiber volume	%	0–6.2	1.84	2.14	6.55
15	Aspect ratio of fibers	1	30–83	39	-0.30	-1.75
16	Size of specimen	mm	40–110	55	1.28	-0.02
17	28-day compressive strength	MPa	57–180	125	-0.36	-0.44

¹ “Skew.” stands for skewness.² “Kurt.” stands for kurtosis.³ “cm” stands for cementitious materials.⁴ ratios are by weight.

a many-objective optimization method, namely Adaptive Geometry Estimation-based Many-Objective Evolutionary Algorithm (AGE-MOEA) (Panichella, 2019), and a decision-making approach, called Pareto optimal solutions using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (Hwang and Yoon, 1981), for many-objective optimization in two design scenarios of UHPC. This research advances the capability for efficient discovery of new UHPC.

2. Methodology

Fig. 1 shows the flowchart of the framework. There are eight main steps: (1) Four datasets are established using experimental data of compressive strength, flexural strength, porosity, and mini-slump spread of UHPC. Each dataset is divided into training and test sets as elaborated in Section 2.1. (2) Automated machine learning generates predictive models for the four datasets, as shown in Section 2.2. (3) Generative techniques synthesize artificial data to enlarge the training datasets, as discussed in Section 2.3. (4) The predictive models are re-trained using the enlarged datasets composed of experimental data and synthetic data, and compared against eight state-of-the-art methods, as elaborated in Section 2.4. (5) Six objective functions along with design constraints are formulated to optimize the mechanical properties, eco-efficiency, and cost-efficiency of UHPC, as elaborated in Section 2.5. (6) Mixture optimization problems of UHPC are defined, as elaborated in Section 2.6. (7) An evolutionary many-objective optimization algorithm (AGE-MOEA) is utilized to solve the optimization problems, as elaborated in 2.7. (8) A decision-making method is applied to select the most preferable optimal solutions, as elaborated in Section 2.8.

2.1. Dataset

Four datasets with 785 experimental data from 49 references were established (Abbas et al., 2015; Ahmed et al., 2021; Ashkezari et al., 2020; Bonneau et al., 2000; Chan and Chu, 2004; Charron et al., 2007; Chu and Kwan, 2019; Corinaldesi, 2012; Corinaldesi and Moriconi, 2012; Gesoglu et al., 2016; Ghafari et al., 2014, 2016; Graybeal, 2007; Graybeal and Hartmann, 2003; (Guo et al., 2021); Guvensoy et al., 2004; Habel et al., 2006; Hassan et al., 2012; Hassan, 2013; Huang et al., 2017;

Table 2

Summary statistics of the variables corresponding to the flexural strength dataset.

Number	Variable	Unit	Range	Mean	Skew. ¹	Kurt. ²
1	Cement-to-cm ^{3,4}	1	0.24–1	0.77	-1.01	0.60
2	Cement type	MPa	42.5, 52.5	48.04	-0.22	-1.98
3	Fly ash-to-cm	1	0–0.55	0.07	1.67	1.79
4	Slag-to-cm	1	0–0.45	0.03	3.04	7.92
5	Silica fume-to-cm	1	0–0.25	0.11	0.24	-1.52
6	Metakaolin-to-cm	1	0–0.286	0.02	3.34	11.99
7	Nano silica-to-cm	1	0–0.062	0.01	1.56	0.87
8	Limestone -to-cm	1	0–0.6	0.07	2.09	2.77
9	Quartz powder-to-cm	1	0–0.4	0.04	2.48	4.97
10	Sand-to-cm	1	0–2.897	1.24	0.71	1.94
11	Maximum aggregate size	mm	0.1–5	1.79	1.25	1.61
12	Water-to-cm	1	0.125–0.4	0.21	1.64	2.69
13	Superplasticizer-to-cm	1	0–0.1	0.03	1.28	1.25
14	Steel fiber volume	%	0–6.2	1.65	2.79	11.17
15	Aspect ratio of fibers	1	30–81.25	35.78	-0.14	-1.91
16	Length of specimen	mm	100–500	172.32	3.47	14.57
17	28-day flexural strength	MPa	6.6–39.4	21.83	0.35	-0.58

¹ “Skew.” stands for skewness.² “Kurt.” stands for kurtosis.³ “cm” stands for cementitious materials.⁴ ratios are by weight.

Jiang et al., 2015; Kang et al., 2018; Li et al., 2020; Liu and Guo, 2018; Liu et al., 2018; Ma et al., 2002; Mo et al., 2020; Pourbaba et al., 2018; Prem et al., 2015; Rajasekar et al., 2019; Richard and Cheyrezy, 1995; Roberti et al., 2021; Sadrmomtazi et al., 2018; Seps et al., al.; Song et al., 2018a, 2018b; Staquet and Espion, 2004; Tafroui et al., 2009; Voo Yen et al., 2010; Wang et al., 2012; Wu et al., 2017a, 2017b, 2016a; Wu et al., 2016b; Yang et al., 2009; Yu et al., 2017, 2014a, (Yu et al., 2014b), 2015a, (Yu et al., 2015b)). The datasets included the compressive strength, flexural strength, and porosity of UHPC at 28 days, as well as the mini-slump spread. The number of test data for these four properties was 379, 146, 152, and 108, respectively. In total, 481 unique UHPC mixtures were considered in the datasets. Those mixtures adopted Class F fly ash, ordinary Portland cement, and straight steel fibers. Standard curing was applied to the specimens used in evaluating the mechanical properties and porosity of UHPC at 28 days. The compressive strengths were evaluated using cubic specimens according to code ASTM C109. The flexural strengths were evaluated using beam specimens through three-point bending tests according to ASTM C1609.

Each dataset was divided into training and test sets, with 80% data randomly selected for the training set and 20% data for the test set. In the datasets, 16 design variables were considered for compressive strength and flexural strength, and 15 design variables were considered for the porosity and mini-slump spread. Tables 1 to 4 list design variables and statistics of the mean, range, skewness, and kurtosis.

In Tables 1 to 4, all the ratios are by mass. Although the compressive strength of UHPC should be greater than 120 MPa, several mixture designs are involved in the developed dataset with compressive strength lower than the specified value. The reason is that some studies investigated the effects of the mixture design variables such as the water-to-cement ratio and the fiber content on the compressive strength. In those studies, some mixture designs reached 120 MPa, but some mixture designs did not reach 120 MPa.

Skewness reflects the asymmetry of distribution (Cain et al., 2017). Kurtosis indicates the outlier-prone extent of distribution. According to reference (Mallery and George, 2000), when the skewness and the kurtosis of a distribution are in the range of -2 to 2, the distribution is considered as a normal distribution. Most of the variables did not follow

Table 3

Summary statistics of the variables corresponding to the mini-slump spread dataset.

Number	Variable	Unit	Range	Mean	Skew. ¹	Kurt. ²
1	Cement-to-cm ^{3,4}	1	0.2–1.0	0.72	-0.58	-0.23
2	Cement type	MPa	42.5, 52.5	49.34	-0.80	-1.38
3	Fly ash-to-cm	1	0–0.6	0.07	1.72	2.63
4	Slag-to-cm	1	0–0.5	0.07	1.90	2.46
5	Silica fume-to-cm	1	0–0.3	0.12	-0.02	-1.61
6	Metakaolin-to-cm	1	0–0.3	0.01	6.70	49.00
7	Nano silica-to-cm	1	0–0.1	0.01	1.36	0.39
8	Limestone-to-cm	1	0–0.6	0.10	1.55	1.01
9	Quartz powder-to-cm	1	0–0.4	0.05	2.04	3.47
10	Sand-to-cm	1	0–2.9	1.16	0.75	1.75
11	Maximum aggregate size	mm	0.1–4.8	1.48	1.39	3.22
12	Water-to-cm	1	0.2–0.4	0.21	1.80	3.16
13	Superplasticizer-to-cm	1	0–0.2	0.05	1.80	3.10
14	Steel fiber volume	%	0–6.2	1.35	1.50	3.61
15	Aspect ratio of fibers	1	30–75.0	38.67	-0.43	-1.78
16	Mini-slump spread	mm	100.8–353.6	231.61	-0.11	-0.85

¹ “Skew.” stands for skewness.² “Kurt.” stands for kurtosis.³ “cm” stands for cementitious materials.⁴ ratios are by weight.**Table 4**

Summary statistics of the variables corresponding to the porosity dataset.

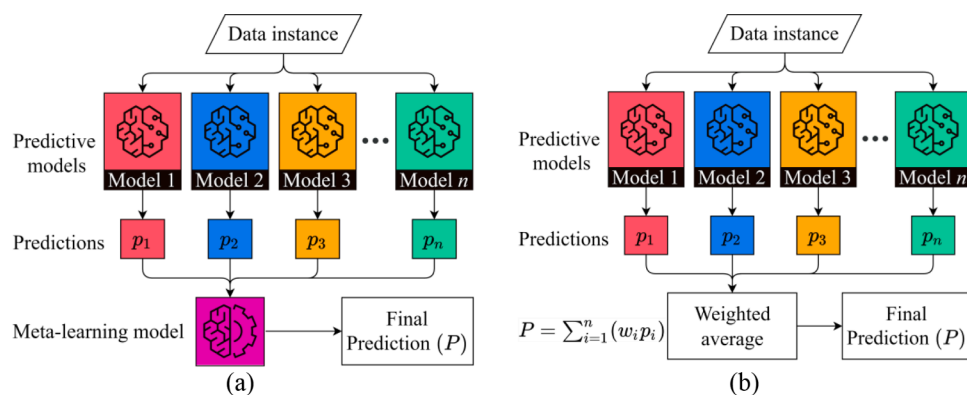
Number	Variable	Unit	Range	Mean	Skew. ¹	Kurt. ²
1	Cement-to-cm ^{3,4}	1	0.67–1.00	0.72	-0.58	-0.23
2	Cement type	MPa	42.50–52.50	49.34	-0.80	-1.38
3	Fly ash-to-cm	1	0–0.22	0.07	1.72	2.63
4	Slag-to-cm	1	0–0.30	0.07	1.90	2.46
5	Silica fume-to-cm	1	0–0.19	0.12	-0.02	-1.61
6	Metakaolin-to-cm	1	0–0.29	0.01	6.70	49.00
7	Nano silica-to-cm	1	0–0.06	0.01	1.36	0.39
8	Limestone-to-cm	1	0–3.56	0.10	1.55	1.01
9	Quartz powder-to-cm	1	0–0.40	0.05	2.04	3.47
10	Sand-to-cm	1	0–3.71	1.16	0.75	1.75
11	Maximum aggregate size	mm	0.50–2.00	1.48	1.39	3.22
12	Water-to-cm	1	0.13–0.4	0.21	1.80	3.16
13	Superplasticizer-to-cm	1	0.01–0.10	0.05	1.80	3.10
14	Steel fiber volume	%	0–6.2	1.35	1.50	3.61
15	Aspect ratio of fibers	1	30–60	231.61	-0.11	-0.85
16	28-day porosity	%	4.53–17	10.69	-0.01	2.24

¹ “Skew.” stands for skewness.² “Kurt.” stands for kurtosis.³ “cm” stands for cementitious materials.⁴ ratios are by weight.

normal distributions.

2.2. Automated machine learning

To address the challenges of high-dimensional regression and model configuration, this study proposes to use Azure AutoML to automate the

**Fig. 2.** The flowcharts for the ensemble learning: (a) stacking and (b) voting ensemble methods.

efforts of data preprocessing, model configuration, and hyperparameter tuning (Copeland et al., 2015; Mukunthu and Gillett, 2018). Azure AutoML automatically designs the optimal machine learning pipeline, which is a sequence of steps from data preprocessing to property prediction, based on probabilistic matrix factorization (Lawrence and Urtasun, 2009), Bayesian optimization, and ensemble learning (Fusi et al., 2018). Probabilistic matrix factorization explores high-performance machine learning models by leveraging prior knowledge obtained by training more than 1000 machine learning pipelines with various datasets. Probabilistic matrix factorization embeds different pipelines in a latent space based on their performance across various datasets (Fusi et al., 2018). The idea is that if the performance of a few pipelines is similar for two datasets, the performance of the remaining pipelines is perhaps similar for the two datasets. A set of pipelines are selected based on predicted performance and tested using the given dataset.

The performance of the machine learning models during the training process is assessed using k-fold cross-validation. This process continues until the target accuracy of probabilistic matrix factorization is achieved. Bayesian optimization is then performed to optimize hyperparameters of the selected pipelines. Finally, the predictive models determined by the designed pipelines are combined using voting and stacking ensemble methods to improve the accuracy (Lawrence and Urtasun, 2009). A voting ensemble model predicts based on the weighted average of predictions made by heterogeneous models. Stacking trains a meta-model based on the predictions of individual models, as shown in Fig. 2.

2.3. Generative modeling

Synthetic data, as implied by its name, are artificial data, but they are not random data. Synthetic data are generated using generative models trained using real test data, and they are used to supplement the test data because machine learning models are prone to overfitting issues when the models are trained using a small dataset.

Generative models have been widely used in computer vision tasks for image augmentation, aiming to circumvent the problem of overfitting (Xue et al., 2021). This study tailored the generative models, which were developed for synthesizing images in computer vision tasks, to synthesize tabular data based on semi-supervised learning (Li et al., 2018). The input mixture design variables of UHPC were synthesized using generative models trained using real test data, and the properties of UHPC mixtures were determined using the predictive models trained in Section 2.2. The synthetic data and real experimental data comprise the new training set.

Four generative approaches were used to synthesize the data: (1) GM1: conditional generative adversarial network (Mirza and Osindero, 2014; Xu et al., 2019), (2) GM2: Copula generative adversarial network (Kamthe et al., 2021), (3) GM3: Gaussian Copula (Patki et al., 2016), and (4) GM4: variational autoencoder (Kingma and Welling, 2013). The quality of the synthetic data was assessed by the machine learning efficacy (Xu and Veeramachaneni, 2018), the correlation preservation (Xu and Veeramachaneni, 2018), and the diversity index (Zhang et al., 2022), while the number of synthetic data was kept the same as that of the real training set. Machine learning efficacy evaluates the representation degree of synthetic data against real data. Machine learning models were trained using the synthetic data and real data, respectively. Machine learning efficacy was calculated as $RMSE_S/RMSE_R$, denoting the ratio of the root mean square errors (RMSE) of models trained using the synthetic data and real data, respectively. Correlation preservation (CP) assesses the difference between the correlation coefficients of input and output variables from the real dataset and the correlation coefficients of variables from the synthetic dataset:

$$CP(C_S, C_R) = \frac{1}{n} \sum_{i=1}^n |R_{s,i} - R_{r,i}| \quad (1)$$

Table 5

The investigated regression methods.

Number	Method	Category
1	Ridge	Linear
2	Passive aggressive	Linear
3	Multi-layer perceptron	Artificial neural network
4	Support vector machine	Nonlinear
5	Partial least squares	Cross decomposition
6	Random forest	Ensemble
7	Light gradient boosting machine	Ensemble
8	Azure AutoML	Ensemble
9	The proposed method	Ensemble

where C_S and C_R are two arrays containing the Pearson correlation coefficients for the synthetic and real data, respectively; $R_{s,i}$ and $R_{r,i}$ are the Pearson correlation coefficients of the i th input variable and derived by real and synthetic data, respectively; and n is the number of input variables. Pearson correlation coefficient (R) between two variables x and y can be calculated as:

$$R(x, y) = \frac{\sum_{i=1}^{\varphi} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{\varphi} (x_i - \bar{x})^2 \sum_{i=1}^{\varphi} (y_i - \bar{y})^2}} \quad (2)$$

where x_i is the i th value of x ; $\bar{x}$ is the mean of x ; y_i is the i th value of y ; and $\bar{y}$ is the mean of y ; φ is the number of data instances; R is in the range of -1 to 1: value -1 means a total negative linear relation; zero means no correlation; and 1 means a total positive relation.

The diversity index quantifies the diversity between the synthetic data and real data. The diversity index (DI) is the average of Euclidean distance between real and synthetic data instances:

$$DI(X_S, X_R) = \frac{1}{\varphi} \sum_{i=1}^{\varphi} E(x_{s,i}, X_R) \quad (3a)$$

$$E(x_{s,i}, D_R) = \sqrt{\sum_{j=1}^N (s_{ij} - r_j)^2} \quad (3b)$$

where X_S and X_R are the synthetic and real datasets; φ is the number of data instances; $x_{s,i}$ is the i th synthetic data instance; $E(x_{s,i}, D_R)$ is the minimum Euclidean distance between the i th synthetic data instance and real data instances; N is the number of variables; s_{ij} is the j -th variable of the i th synthetic data instance; r_j is the j -th variable of the real data instances with minimum Euclidean distance from the i th synthetic data instance.

The optimal number of added synthetic data was obtained by a parametric study. Augmentation ratio is defined as the ratio of the number of synthetic data to the number of real data. In the parametric study, the augmentation ratio was changed from 0 to 200%. The augmentation ratio that led to minimum RMSE was selected as an optimal augmentation ratio.

2.4. Comparison of regression models

With the new training data, nine predictive models were trained based on different methods, including (i) two linear regression methods, which are ridge (Hoerl and Kennard, 1970) and passive aggressive (Crammer et al., 2006), (ii) a deep neural network, which is multi-layer perceptron, (iii) a nonlinear machine learning method, which is support vector machine (Cortes and Vapnik, 1995), (iv) a cross decomposition method, which is partial least squares (Geladi and Kowalski, 1986), (v) two ensemble learning methods, which are random forest methods (Breiman, 2001), (vi) a light gradient boosting machine (Ke et al., 2017), and (vii) a model obtained using the proposed method based on Azure AutoML, as listed in Table 5. The nine models were compared in term of five metrics, which are the RMSE, mean absolute error (MAE), mean

Table 6
Inventory of the raw materials of UHPC.

Number	Material	Specific gravity(unitless)	Carbon footprint(kg CO ₂ -eq/kg)	Embodied energy(MJ/kg)	Material cost(USD/kg)
1	Portland cement	3.14 (Meng et al., 2018)	0.83 (Long et al., 2015)	5.8 (Müller et al., 2014)	0.082 (Alsaman et al., 2020)
2	Fly ash	2.70 (Meng et al., 2018)	0.027 (Vincent et al., 2021)	0.83 (Long et al., 2015)	0.04 (Alsaman et al., 2020)
3	Slag	2.90 (Yalçinkaya and Yazıcı, 2017)	0.052 (Vincent et al., 2021)	1.59 (Long et al., 2015)	0.1 (Alsaman et al., 2020)
4	Silica fume	2.20 (Meng et al., 2018)	0.0039 (Ghavami et al., 2021)	0.036 (Kathirvel and Sreekumaran, 2021)	0.8 (Alsaman et al., 2020)
5	Metakaolin	2.62 (Alharbi et al., 2021)	0.4 (Long et al., 2015)	3.48 (Long et al., 2015)	0.5 (Alsaman et al., 2020)
6	Nano silica	2.30 (Alharbi et al., 2021)	1.69 (Ghavami et al., 2021)	71.36 (Ghavami et al., 2021)	2.5 (Adamu et al., 2018)
7	Limestone powder	2.73 (Abellán-García, 2020)	0.019 (Chiaia et al., 2014)	0.76 (Chiaia et al., 2014)	0.12 (Alsaman et al., 2020)
8	Quartz powder	2.67 (Vaitkevicius et al., 2014)	0.023 (Kathirvel and Sreekumaran, 2021)	0.85 (Kathirvel and Sreekumaran, 2021)	0.8 (Alsaman et al., 2020)
9	Fine sand	2.64 (Meng et al., 2018)	0.01 (Shi et al., 2019)	0.11 (Müller et al., 2014)	0.025 (Zhang et al., 2020)
10	Tap water	1.00 (Meng et al., 2018)	0.0003 (Long et al., 2015)	0.006 (Long et al., 2015)	0.001 (Adamu et al., 2018)
11	Superplasticizer	1.05 (Meng et al., 2018)	0.72 (Chiaia et al., 2014)	18.3 (Chiaia et al., 2014)	3.4 (Alsaman et al., 2020)
12	Steel fiber	7.80 (Meng et al., 2018)	1.50 (Chiaia et al., 2014)	20.56 (Chiaia et al., 2014)	5.0 (Alsaman et al., 2020)

absolute percentage error (MAPE), mean absolute deviation (MAD), and R²:

$$MAE(P, A) = \frac{1}{\varphi} \sum_{i=1}^{\varphi} |p_i - a_i| \quad (4a)$$

$$MAPE(P, A) = \frac{1}{\varphi} \sum_{i=1}^{\varphi} \left| \frac{p_i - a_i}{a_i} \right| \quad (4b)$$

$$MAD(P, A) = \text{median}(|p_1 - a_1|, |p_2 - a_2|, \dots, |p_n - a_n|) \quad (4c)$$

$$R^2(P, A) = 1 - \frac{\sum_{i=1}^{\varphi} (p_i - a_i)^2}{\sum_{i=1}^{\varphi} [a_i - \text{mean}(a_i)]^2} \quad (4d)$$

2.5. Objective functions

To achieve desired properties of UHPC, six objective functions are considered: (1) maximization of the compressive strength, (2) maximization of the flexural strength, (3) minimization of the porosity, (4) minimization of the carbon footprint (CF), (5) minimization of the embodied energy (EE), and (6) minimization of the material cost (MC), calculated as:

$$CF = \sum_{i=0}^n m_i \times CO_2 - eq_i \quad (5a)$$

$$EE = \sum_{i=0}^n m_i \times ee_i \quad (5b)$$

$$MC = \sum_{i=0}^n m_i \times c_i \quad (5c)$$

where n is the number of raw materials; m_i is the mass of i th raw material in a unit mass of UHPC; $CO_2 - eq_i$ is the carbon dioxide equivalent of a unit mass of the i th raw material; ee_i is the amount of embodied energy of a unit mass of the i th raw material; c_i is the unit price of the i th raw material.

Table 6 lists the inventory for the unit carbon footprint, embodied energy, and unit price of the raw materials of UHPC. Six types of SCMs were considered, which are Portland cement, fly ash, slag, silica fume, metakaolin, and nano silica. Two different types of fillers were considered, which are limestone powder and quartz powder.

Table 7

The objectives of the two investigated design optimization problems.

Design objective	Design scenario DS1	DS2
Compressive strength	Included	Included
Flexural strength	Included	Included
Porosity	Included	Included
Carbon footprint	Excluded	Included
Embodied energy	Excluded	Included
Material cost	Excluded	Included

2.6. Design problems

Table 7 lists two design scenarios, designated as DS1 and DS2: DS1 is a three-objective optimization problem for the compressive strength, flexural strength, and porosity. DS2 is a six-objective optimization problem for all the six objectives. DS2 is a many-objective optimization problem as it has more than three objectives (Ishibuchi et al., 2008). Two design constraints are imposed. First, the mini-slump spread is in the range of 260 mm to 300 mm to ensure adequate flowability and fiber dispersion according to reference (Liu et al., 2020). The mini-slump spread is predicted using the developed machine learning model. Second, the total ratio of cementitious materials is equal to 1, expressed as:

$$(s-280)^2 - 20^2 < \varepsilon_1 \quad (6a)$$

$$\left(\sum_{i=1}^6 r_i - 1 \right)^2 < \varepsilon_2 \quad (6b)$$

where s is the mini-slump spread; ε_1 and ε_2 are tolerance levels, which are set to 1 and 0.01, respectively; and r_i is the ratio of the i th cementitious material.

2.7. Evolutionary optimization

This study adopts the AGE-MOEA (Panichella, 2019) to solve the optimization problems formulated in Section 2.6. AGE-MOEA aims to determine the optimal solutions that are non-dominated to each other but are superior to the rest of solutions. The objectives of non-dominated solutions cannot be improved without compromising the other objectives. The projection of actual non-dominated solutions is known as Pareto front. AGE-MOEA estimates the Pareto front to ensure diversity and proximity of non-dominated solutions. AGE-MOEA was compared with three other evolutionary algorithms: Two-Archive Evolutionary

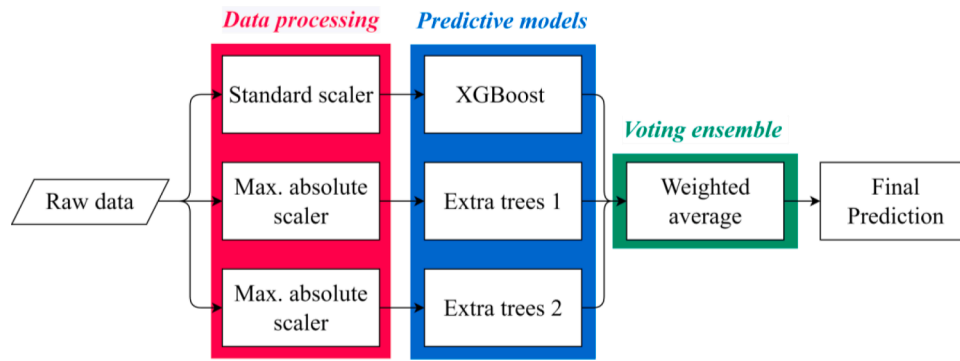


Fig. 3. The machine learning pipeline designed by the proposed automated machine learning.

Algorithm for Constrained Multi-objective Optimization (C-TAEA) (Li et al., 2019), Non-dominated Sorting Genetic Algorithm II (NSGA-II), and Unified non-dominated Sorting Genetic Algorithm III (UNSGA-III) (Seada and Deb, 2014). The number of generations and population size for the algorithms were set to 300 and 100, respectively. The hyper-volume indicator was used to evaluate the quality of solutions (Zitzler and Thiele, 1998). The reference point for the two design scenarios is set based on the minimum values of compressive strength, flexural strength, mini-slump spread, and porosity, and maximum values of carbon footprint, embodied energy, and material cost. The hypervolume indicator was normalized by the maximum of hypervolume.

2.8. Multi-criteria decision-making

Multi-criteria decision-making is a type of decision-making that explicitly considers multiple conflicting criteria. This study adopts multi-criteria decision-making to select the most preferable mixture design solutions according to the conflicting design objectives, as discussed in Section 2.5. When there are six or more objectives, the multi-objective optimization problems are called many-objective optimization problems. With the multiple solutions obtained from AGE-MOEA in Section 2.7, a multi-criteria decision-making method called TOPSIS was utilized to select the ultimate optimal solutions based (Hwang and Yoon, 1981; Mahjoubi et al., 2021). The main idea of TOPSIS is that a preferable solution among all the determined solutions is the one that has the least distance to the ideal solution and the most distance from the nadir solution in the objective space. The ideal solution is the combination of the best values for the objective functions, while the nadir solution is the combination of the worst values for the objective functions. The objective space is a multi-dimensional space where each dimension indicates the magnitude of an objective function. TOPSIS incorporates relative weights of criterion importance according to the design preference. Therefore, new UHPC mixtures are discovered automatically for different applications with various design preferences. In this study, the relative weight for each criterion is set to 1.

3. Results and discussion

Section 3 presents the results including the machine learning pipeline and hyperparameters of machine learning models designed by Azure AutoML (Section 3.1), performance evaluation of the synthetic data (Section 3.2), comparison of predictive models obtained by the proposed method and other regression methods (Section 3.3), and the discovered low-carbon UHPC mixtures (Section 3.4).

3.1. Machine learning pipeline

Fig. 3 shows the machine learning pipeline automatically designed by Azure AutoML using the compressive strength dataset. The machine learning pipeline consists of three main modules, which are data

Table 8
The optimal hyperparameters of the predictive models.

Predictive models	Hyperparameters	Optimal results
XGBoost	Grow policy	Loss guide
	L1 regularization parameter (α)	0
	L2 regularization parameter (λ)	0.625
	Learning rate	0.2
	Loss function	squared error
	Maximum depth	7
	Maximum number of bins	1023
	Maximum number of leaves	31
	Minimum split loss	0
	Number of estimators	100
Extra trees 1	Subsample ratio	0.7
	Tree construction algorithm	Histogram-based
	Bootstrapping	False
	Fraction of variables at each split	0.8
	Loss function	Mean squared error
Extra trees 2	Minimum number of samples in each leaf	4
	Number of estimators	10
	Bootstrapping	True
	Fraction of variables at each split	0.7
	Loss function	Mean squared error
	Minimum number of samples in each leaf	3
	Number of estimators	25

preprocessing, prediction, and voting ensemble. In data preprocessing, two scaler transforms were applied, which were standard scaler (z_1) and maximum absolute scaler (z_2), defined as:

$$z_1 = \frac{x - m}{\sigma} \quad (7)$$

$$z_2 = \frac{x - m}{\max(x)} \quad (8)$$

where x is an input variable; m and σ are the average and standard deviation of x ; and $\max(x)$ is the maximum value of x .

In the prediction module, three models were employed for prediction, which are an XGBoost (Chen and Guestrin, 2016) and two extra trees (Geurts et al., 2006). Their hyperparameters were determined by Azure AutoML (Table 8). In the voting ensemble, predictions from the models were combined with assigned weights 0.364, 0.545, and 0.091, respectively, to obtain the final prediction.

3.2. Evaluation of synthetic data

To evaluate the effect of the semi-supervised learning method, synthetic data are generated in two ways: (1) Type I: Synthetic data are

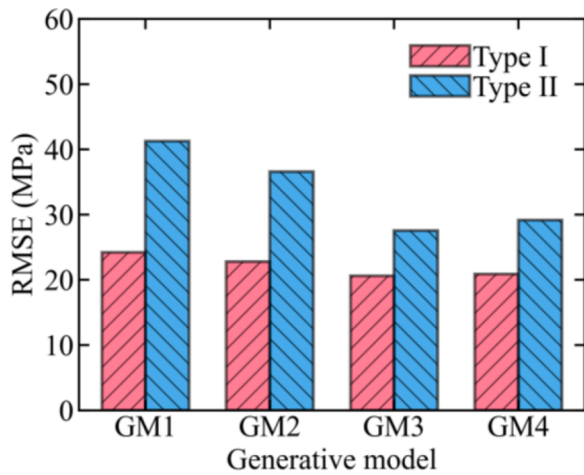


Fig. 4. The performance of machine learning models trained with the synthetic data obtained by different generative models; vertical axis shows the RMSE of the developed machine learning models trained by the synthetic data obtained by the four generative models on the real test set.

Table 9
Evaluation results of the synthetic data.

Metric	Goal	Generative model			
		GM1	GM2	GM3	GM4
Machine learning efficacy	Minimize	1.48	1.34	1.11	1.14
Correlation preservation	Minimize	0.11	0.10	0.09	0.08
Diversity index	Maximize	13.4	10.7	14.0	4.27

generated by semi-supervised learning and generative modeling. The input variables are obtained by generative models, and the output variables are predicted by the trained models. (2) Type II: Both the input and output variables of synthetic data are generated by generative models. Fig. 4 shows the RMSE of compressive strength predicted by models trained using the two types of synthetic data. The RMSE of Type I models are lower than that of Type II models, indicating the benefits of the proposed semi-supervised learning method. The predictive model trained using the synthetic data from GM3 achieved the highest accuracy.

Table 9 lists the metrics of the synthetic datasets. The synthetic data determined by GM3 has the best performance among the investigated generative models. The quality of synthetic data is further evaluated in Appendix A based on the differences between the Pearson correlation coefficients calculated by the real and synthetic data, and the Euclidean distances between real and synthetic data instances. Kernel density estimation, which is a non-parametric approach to estimate the probability density function of a variable [115], is utilized to study the distribution of variables. In addition, the cumulative sum of differences and Euclidean distances are investigated in Appendix A. The results show that the range of differences is the smallest for the synthetic data generated by GM3, and GM3 generates the most diverse synthetic data. It can be concluded from the results that GM3 is superior among other investigated generative models.

To investigate the effect of the relative number of synthetic data, an augmentation ratio is defined as the ratio of the number of synthetic data to the number of real data. As the augmentation ratio varies from 0 to 200%, the RMSE of the trained predictive models evaluated using the test set are plotted in Fig. 5. The optimal augmentation ratios that achieve the minimum RMSE values for the compressive strength, flexural strength, mini-slump spread, and porosity datasets are 20%, 20%, 85%, and 110%, respectively, which correspond to 268, 168, 159, and

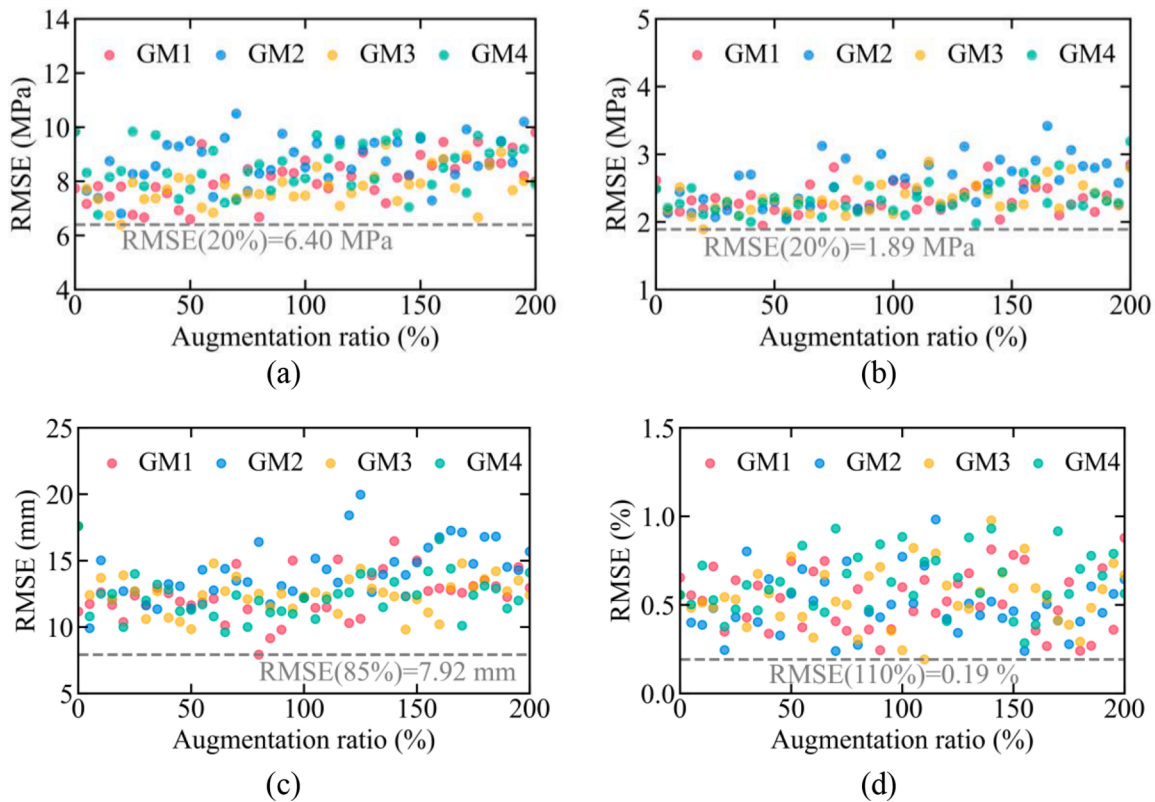


Fig. 5. The effect of augmentation ratio on the performance of machine learning models to predict (a) compressive strength, (b) flexural strength, (c) mini-slump spread, and (d) porosity of UHPC; each dot represent a machine learning model trained with the combination of real training data and synthetic data, while each dashed line shows the minimum of RMSE.

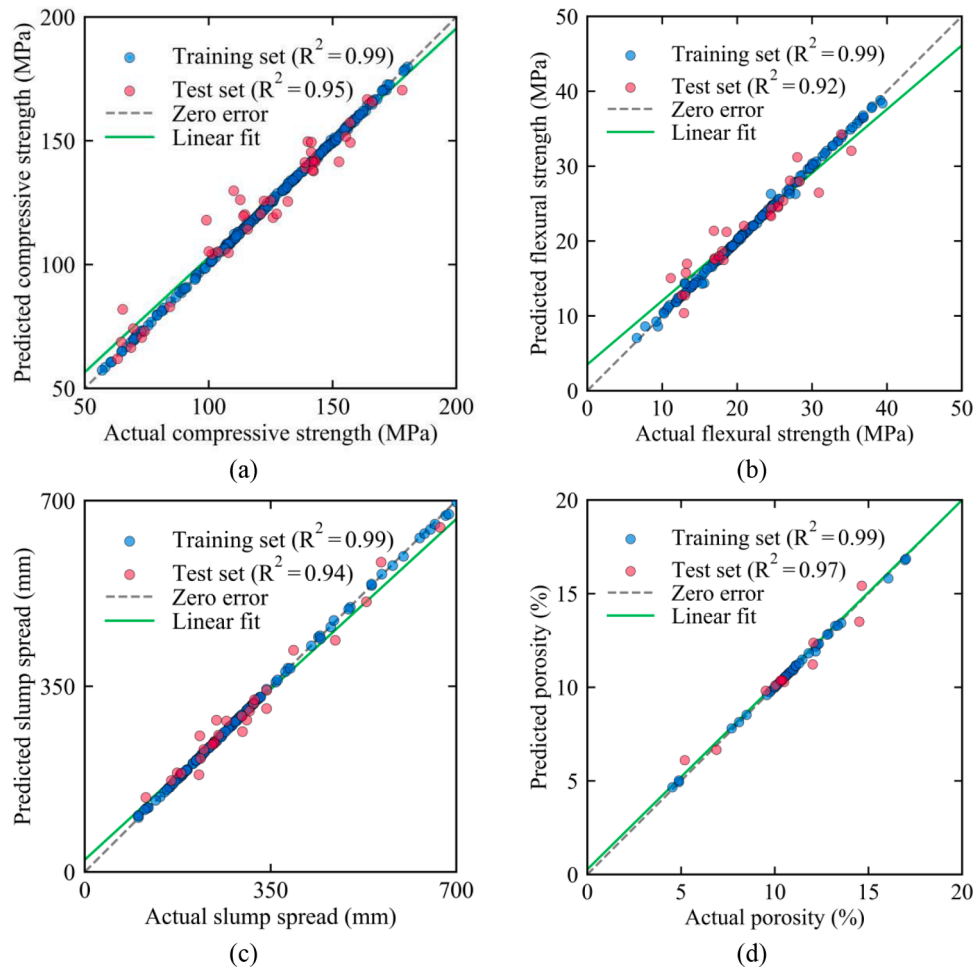


Fig. 6. Actual experimental results versus the predicted results for (a) compressive strength, (b) flexural strength, (c) mini-slump spread, and (d) porosity of UHPC.

102 data instances in the compressive strength, flexural strength, mini-slump spread, and porosity datasets. The ranges of the synthetic data were within the range of real data, as elaborated in Appendix C.

3.3. Predictive performance

Fig. 6 compares the prediction results against the actual values in the test datasets. The test datasets are unseen to the predictive models. These plots show that the predictions are in a good agreement with the actual experimental observations. The R^2 values of the predictive models on the test datasets are higher than 0.92. The results indicate that the predictive models have high prediction accuracy and generalization performance. The distributions of the residual errors are investigated to demonstrate the robustness of predictions made by the four models (Appendix B).

Fig. 7 shows the Taylor diagrams (Taylor et al., 2012) for the predictions made by the proposed method and the state-of-the-art regression methods in Section 2.7. A Taylor diagram provides a visual representation of the comparison between predictions and observed data. Each dot represents a predictive model. The model closest to the experimental data (observed) shows the highest prediction accuracy.

Fig. 7 shows that the proposed method has the highest prediction accuracy and precision. Both the RMSE and Pearson correlation coefficient are improved by the incorporation of the synthetic data. The RMSE values of the compressive strength, flexural strength, mini-slump spread, and porosity models are reduced by 31%, 7%, 3%, and 73%, respectively. The proposed approach improves the prediction performance, especially when the dataset is small, such as the porosity dataset with

only about 100 data instances.

The performance metrics defined in Section 2.6 are calculated for the trained models to predict the compressive strength, flexural strength, mini-slump spread, and porosity, as listed in Table 10. The R^2 values of the predictive models evaluated using the test datasets are higher than 0.92, indicating that the predictive models have satisfactory accuracy and generalizability, ensuring that the data-driven models can be utilized to design UHPC mixtures.

3.4. Many-objective optimization

Fig. 8(a) and Fig 8 (b) show the normalized hypervolume indicator of the solutions obtained for DS1 and DS2 by the four investigated methods during the optimization process. The normalized hypervolume indicator value increases with the number of iterations, indicating that the algorithms iteratively improve their solutions. Regarding DS1, the normalized hypervolume indicators of the solutions obtained by the proposed optimization methods at the last iteration process are comparable with each other. On the contrary, the normalized hypervolume indicator of the solutions determined for DS2 by AGE-MOEA during the optimization process is significantly higher than that of the other methods. Noted that DS1 is a multi-objective optimization problem with three objectives, while DS2 is a many-objective optimization problem with six objectives. Therefore, it can be said that the four evolutionary optimization methods have comparable performance for multi-objective optimization problems, but AGE-MOEA has better performance for many-objective optimization problems.

Table 11 lists two UHPC mixtures discovered using the proposed

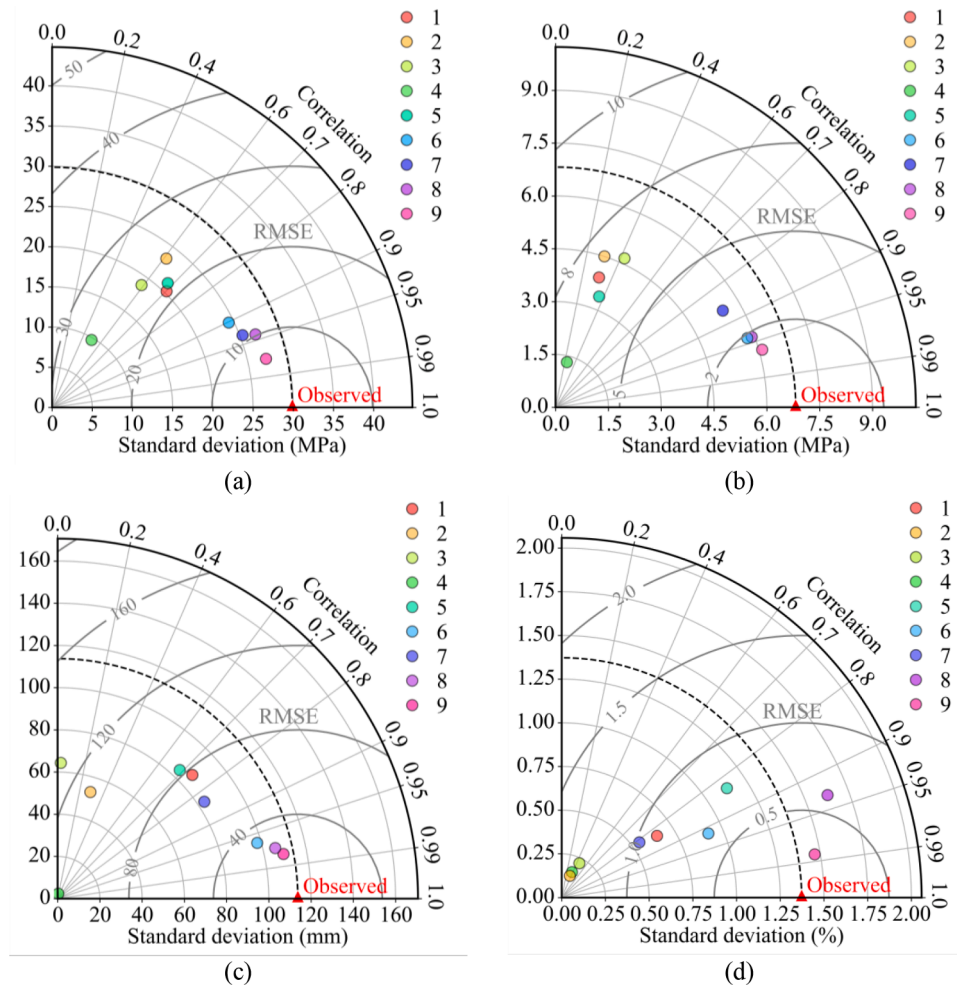


Fig. 7. Taylor diagrams of the machine learning models for: (a) compressive strength, (b) flexural strength, (c) mini-slump spread, and (d) porosity. The labels are defined in Table 5. Both the horizontal and vertical axes show the magnitude of standard deviation.

Table 10
Performance metrics of the predictive models for UHPC properties.

Dataset	Metric	Compressive strength	Flexural strength	Mini-slump spread	Porosity
Training	MAE	0.35	0.18	1.34	0.17
	MAPE	0.00	0.01	0.02	0.01
	MAD	0.22	0.08	0.75	0.02
	RMSE	0.51	0.35	2.15	0.30
	R ²	0.99	0.99	0.99	0.97
Test	MAE	5.00	1.61	12.56	0.17
	MAPE	0.04	0.09	0.04	0.02
	MAD	3.50	0.94	7.65	0.22
	RMSE	6.40	1.89	9.62	0.16
	R ²	0.95	0.92	0.94	0.97

approach for the two design scenarios and compares the two mixtures with three representative cost-effective UHPC mixtures that were developed through step-by-step experimental tests. The mixture discovered for DS1 has higher mechanical properties than the mixture discovered for DS2. The mixture discovered for DS2 has low carbon footprint, embodied energy, and material cost while satisfying the requirements of mechanical properties, workability, and porosity. The compressive strength and flexural strength of the UHPC mixture for DS1 are 22% and 40% higher than those of the UHPC mixture for DS2. The carbon footprint, embodied energy, and material cost of the UHPC mixture for DS2 are 73%, 71%, and 80% lower than those of DS1.

Compared with three UHPC mixtures, which are designated as FAC60 (Meng et al., 2016), UHPC-2 (Wille and Boisvert-Cotulio, 2015), and UHPFRC (Yu et al., 2017) developed through experiments, the discovered UHPC mixture for DS2 has comparable mechanical properties, workability, and porosity while highly reducing the carbon footprint, embodied energy, and material cost.

In addition to the two UHPC mixtures, more non-dominated solutions were obtained by AGE-MOEA. Although the other solutions are listed in a table as Supplementary Data. While the other solutions are not selected based on the two design scenarios, they are possibly the optimal solutions in other design scenarios.

With the UHPC mixtures designed for DS1 and DS2, the roles of different ingredients (see the code from Table 6) on the carbon footprint, embodied energy, and material cost of UHPC can be quantitatively evaluated, as shown in Fig. 9. These rectangular treemaps show the share of each ingredient of the UHPC mixtures.

In Fig. 9(a), the cement is responsible for more than 58% the carbon footprint of the UHPC mixture for DS1, followed by steel fibers which are responsible for more than 37% the carbon footprint. In Fig. 9(b), cement is responsible for more than 74% the carbon footprint of the UHPC mixture for DS2, followed by limestone powder which is responsible for about 5% of the carbon footprint. The UHPC mixture for DS2 utilizes higher volumes of low-carbon raw materials such as limestone powder and slag while minimizes the volumes of high-carbon raw materials, thus achieving the low carbon footprint. In Fig. 9(c), the steel fibers are responsible for about 50% of the embodied energy of the UHPC mixture for DS1, followed by cement which is responsible for

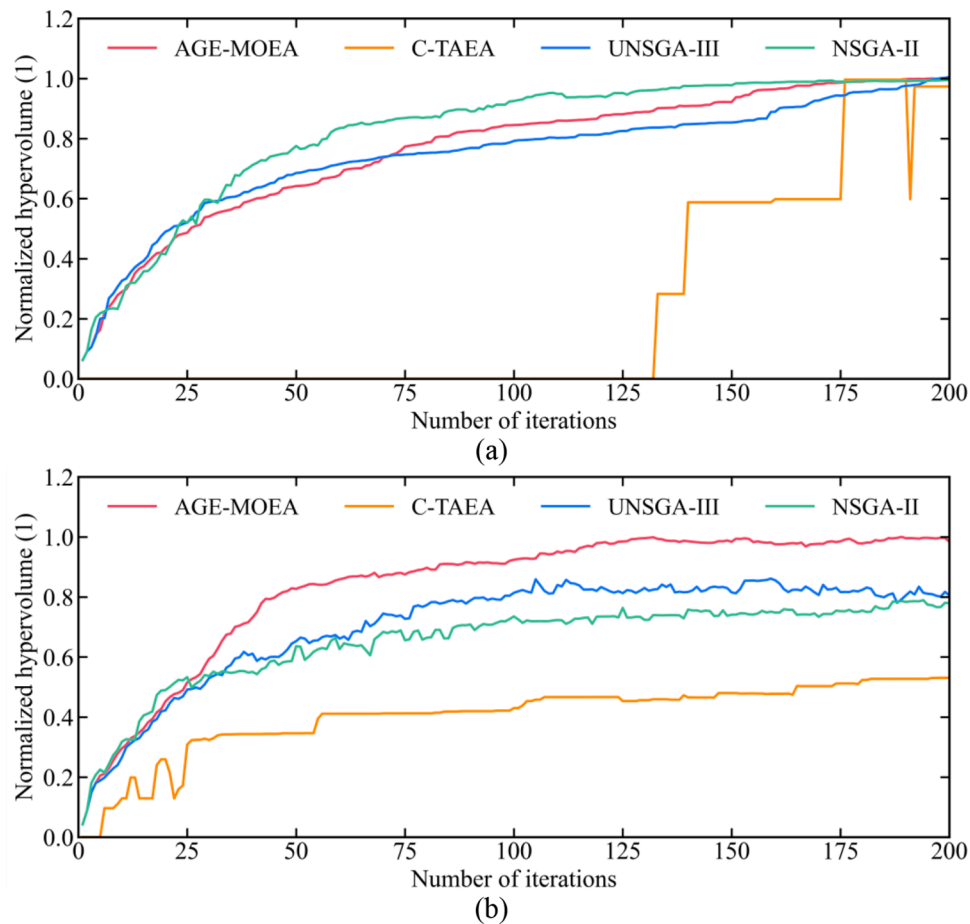


Fig. 8. Evolution of the normalized hypervolume of the optimal solutions for (a) DS1 and (b) DS2.

Table 11

UHPC mixtures discovered for different design scenarios.

Number	Design variable	Unit	Discovered mixtures		Existing mixtures		
			DS1	DS2	FAC60	UHPC-2	UHPFRC
1	Cement-to-cm	1	0.74	0.31	0.47	0.67	0.82
2	Cement type	MPa	52.50	52.50	52.5	52.5	52.5
3	Fly ash-to-cm	1	0.00	0.14	0.53	0.16	0.18
4	Slag-to-cm	1	0.22	0.29	0	0	0
5	Silica fume-to-cm	1	0.11	0.17	0	0.17	0
6	Metakaolin-to-cm	1	0.00	0.04	0	0	0
7	Nano silica-to-cm	1	0.00	0.00	0	0	0
8	Limestone-to-cm	1	0.24	0.95	0	0	0
9	Quartz powder-to-cm	1	0.37	0.01	0	0	0
10	Sand-to-cm	1	0.86	1.47	1.07	1.00	1.18
11	Maximum aggregate size	mm	0.43	0.41	4.75	1.20	2
12	Water-to-cm	1	0.19	0.15	0.18	0.15	0.21
13	Superplasticizer-to-cm	1	0.02	0.02	0.01	0.02	0.04
14	Steel fiber volume	%	2.98	0.05	2	0	2.5
15	Aspect ratio of fiber	1	40.5	74.2	65	N/A	65
Number	UHPC property	Unit	Discovered mixtures		Existing mixtures		
			DS1	DS2	FAC60	UHPC-2	UHPFRC
1	Compressive strength	MPa	171.82	133.3	120	166	160
2	Flexural strength	MPa	32.24	19.22	20.1	18.5	20.0
3	Mini-slump spread	mm	260.2	271.9	285	265	283
4	Porosity	%	8.88	12.13	13.9	9.64	14.0
5	Carbon footprint	kg CO ₂ -eq/m ³	922.5	247.1	673	652	984
6	Embodied energy	MJ/m ³	9631	2750	6819	5014	9507
7	Material cost	\$/m ³	1664	329.1	905	472	1203

more than 38% of the embodied energy. In Fig. 9(d), cement is responsible for more than 46% of the embodied energy of the UHPC mixture for DS2, followed by limestone powder which is responsible for

about 18% of the embodied energy. In Fig. 9(e), the steel fibers are responsible for about 70% of the material cost of the UHPC mixture for DS1, followed by quartz powder which is responsible for 15% the

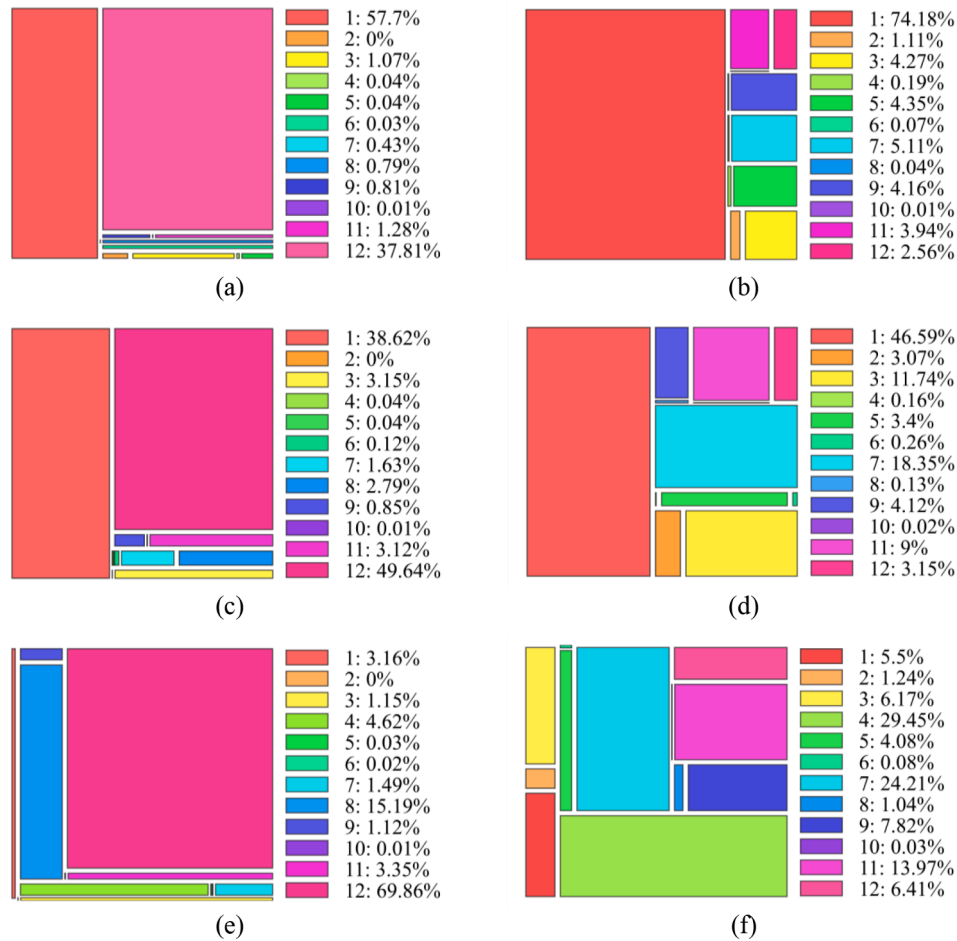


Fig. 9. Depiction of the share of raw materials: (a) carbon footprint of DS1; (b) carbon footprint of DS2; (c) embodied energy of DS1; (d) embodied energy of DS2; (e) material cost of DS1; and (f) material cost of DS2. The labels 1 to 12 of the raw materials are defined in Table 6.

material cost. In Fig. 9(f), silica fume is responsible for about 29% of the material cost of the UHPC mixture for DS2, followed by limestone powder which is responsible for more than 24% of the material cost. The UHPC mixture for DS2 utilizes higher volumes of low-energy low-cost raw materials and minimizes the volumes of energy-intensive costly raw materials, thus achieving the low embodied energy and material cost.

4. Conclusions

This study develops an AI-guided method to automatically discover low-carbon cost-effective UHPC. The presented approach integrates advanced techniques of generative modeling, automated machine learning, and many-objective optimization, aiming to automate data augmentation, machine learning model generation, and UHPC mixture selection. Artificial yet sound data were synthesized to supplement experimental data for training four machine learning models used to predict compressive strength, flexural strength, mini-slump spread, and porosity of UHPC. The quality of synthetic data is assessed using machine learning efficacy, diversity, and correlation preservation. Based on the predictive models and specified objectives, two UHPC mixtures were discovered for two design scenarios. Based on the above investigations, the following conclusions are drawn:

- The developed approach is effective for auto-discovery of low-carbon cost-effective UHPC mixtures in specified design scenarios with particular objectives. The two discovered UHPC mixtures achieved desired compressive and flexural strengths, workability, and porosity. The carbon footprint, embodied energy, and cost of the

UHPC mixture discovered in the second scenario are respectively reduced by 73%, 71%, and 80%, compared with those of the first scenario. The carbon footprint, embodied energy, and cost of the discovered UHPC mixtures are lower than those of existing UHPC mixtures with comparable mechanical properties.

- The presented method for synthesis of data is effective in augmenting training data for new material discovery problems where there is lack of sufficient data. The incorporation of the semi-supervised learning technique into the generative modeling improves the performance of the synthetic data in terms of efficacy. The improvement is likely because semi-supervised learning integrates information embodied in experimental data.
- The automated machine learning method processes the training data, determines the machine learning pipeline, and performs hyperparameter tuning for machine learning models without human intervention. The generated machine learning models achieve high accuracy and generalization performance in predicting the four considered material properties of UHPC.
- The proposed methods for many-objective optimization problems based on AGE-MOEA and TOPSIS demonstrate excellent performance. The adopted evolutionary optimization method achieves a higher hypervolume indicator than the other state-of-the-art optimization methods. The optimal UHPC mixture designs selected by TOPSIS are consistent with the defined objectives and design constraints.

The proposed approach provides an alternative solution for efficient development of low-carbon cost-effective UHPC in different

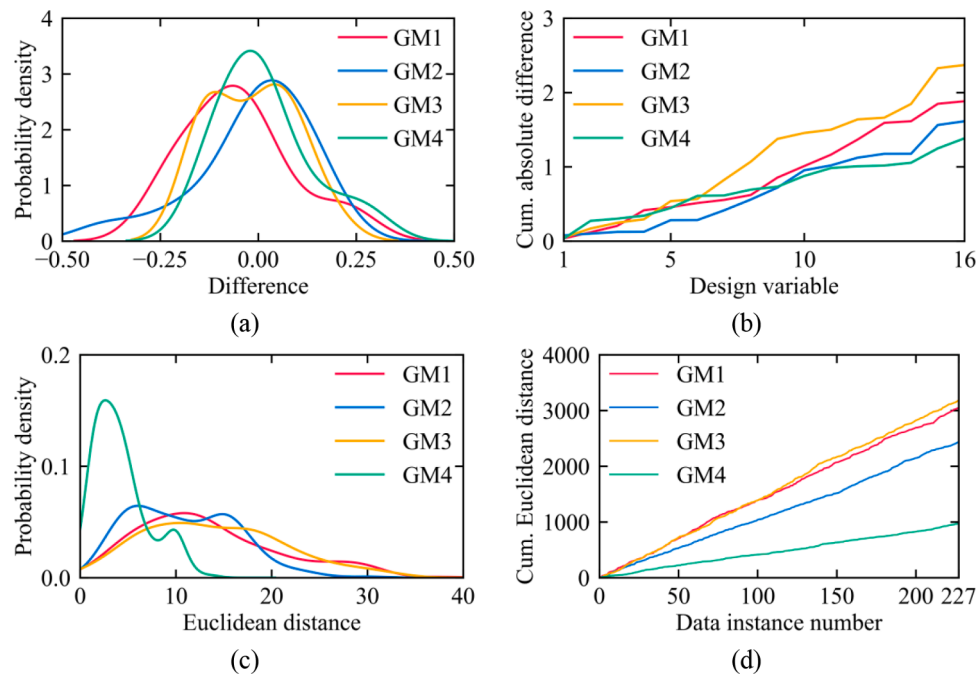


Fig. A.1. An investigation on correlation preservation and diversity of synthetic data: (a) kernel density estimation of the differences between the correlation coefficients derived by real and synthetic data; (b) cumulative summation of differences between correlation coefficients derived by real and synthetic data; (c) kernel density estimation of Euclidean distances of synthetic data; and (d) cumulative sum of euclidean distances.

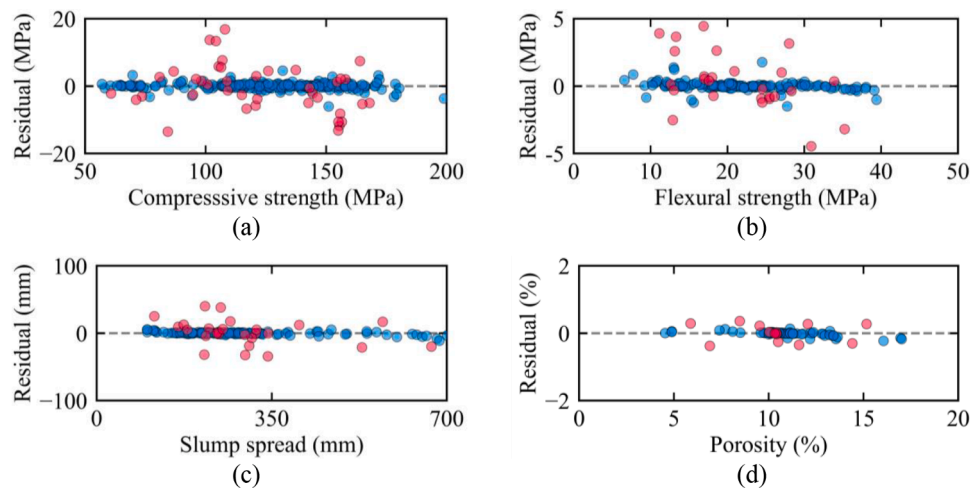


Fig. B.1. Residual errors versus the experimental results for (a) compressive strength, (b) flexural strength, (c) mini-slump spread, and (d) porosity of UHPC.

applications. The proposed methods for data synthesis, automated machine learning, and many-objective optimization are promising for development of other types of materials with minimal human intervention. It is envisioned that different objectives and relative weights can be assigned to tailor the proposed approach for different design scenarios. Further research can be performed to experimentally test the proposed approach in different applications.

Data availability

The datasets are available online at: www.doi.org/10.17632/dd62d5hyzr.3 (Mahjoubi and Bao, 2022).

CRediT authorship contribution statement

Soroush Mahjoubi: Data curation, Formal analysis, Investigation,

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

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have shared the link to our data in the revised manuscript.

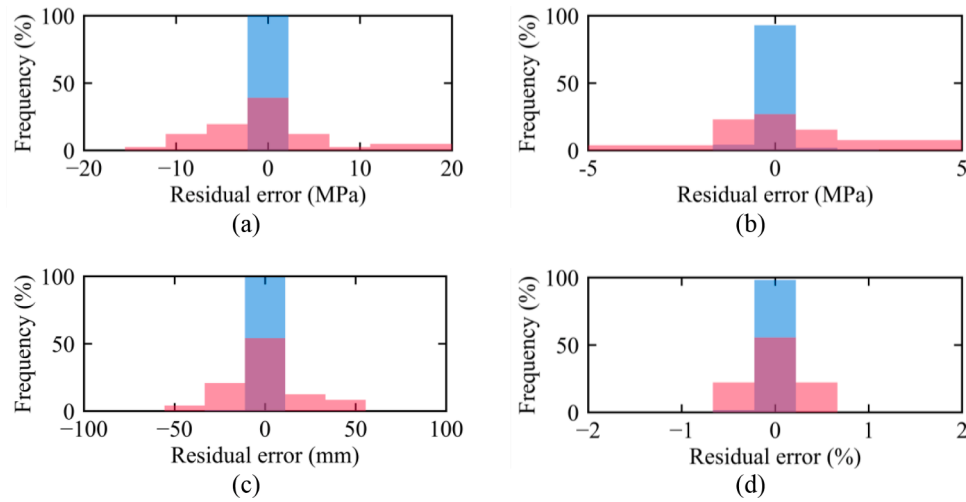


Fig. B.2. Histogram of residual errors in predicting (a) compressive strength, (b) flexural strength, (c) mini-slump spread, and (d) porosity of UHPC.

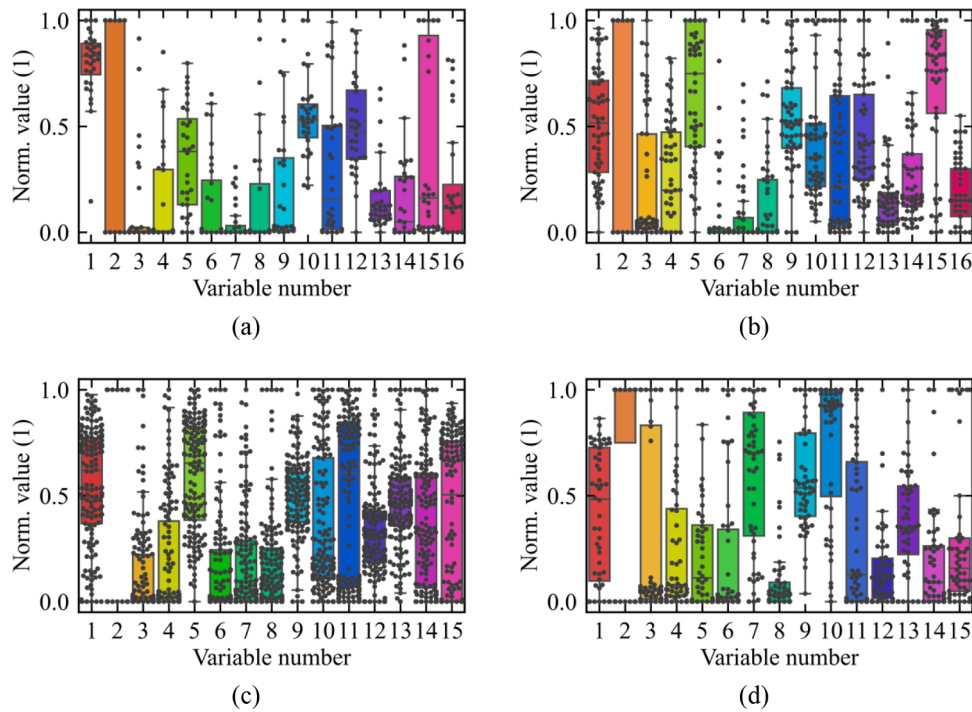


Fig. C.1. Box plots indicating the range of variables of the synthetic data generated for the (a) compressive strength, (b) flexural strength, (c) mini-slump spread, and (d) porosity datasets; variable numbers are defined in Tables 1-4; each dot represents a synthetic data point; in vertical axis, 0 and 1 represent the maximum and minimum values of the variables for the data points in the real dataset.

Appendix A

Fig. A.1 shows the investigation on the correlation preservation and diversity of the synthetic data. Fig. A.1(a) shows the kernel density estimation and cumulative sum of the difference between the Pearson correlation coefficients obtained by the real dataset. Fig. A.1(b) shows the Pearson correlation coefficients obtained by the synthetic data. The range of differences is the narrowest for the synthetic data obtained by GM3. The total sum of difference is the greatest for GM3. The total sum of difference for GM4 is the lowest.

Fig. A.1(c) and Fig. A.1(d) show the kernel density estimation and cumulative sum of the Euclidean distances between synthetic data instances and real data instances. The probability density function of the distances for GM4 has a right-skewed distribution. The total Euclidean distance for GM3 is the largest. The total Euclidean distance for GM4 is the smallest. The data instances obtained by GM3 are more diverse than the data instances determined by the other generative models.

Appendix B

Fig. B.1 shows the distribution of residual errors. The residuals are scattered around the zero-error line. There is no apparent indication of a nonlinear relationship between the observed and predicted values. Fig. B.2 shows the histogram of residuals corresponding to the predictions made by the four developed models. The distributions of residuals are approximately symmetrical and centered at zero. It can be concluded that the predictions made by the four data-driven models are accurate over the entire range of output variables.

Appendix C

Fig. C.1 shows the range of the variables for the synthetic data. To facilitate the comparison of the range of synthetic data with that of the real data, the values of variables are normalized using Eq. (C.1):

$$z_{n,i} = \frac{x_{n,i} - \min(X_i)}{\max(X_i) - \min(X_i)} \quad (\text{C.1})$$

where $z_{n,i}$ and $x_{n,i}$ are the normalized and the original values of the i th variable of the n -th data point; and X_i is an array containing the values of the i th variable for all the data points in the real dataset. Fig. C.1 indicates that the ranges of the variables of the synthetic data are within the ranges of the real data.

References

- Aaleti, S., Petersen, B., Sritharan, S., 2013. Design Guide For Precast UHPC Waffle Deck Panel system, Including Connections. Federal Highway Administration.
- Abbas, S., Soliman, A.M., Nehdi, M.L., 2015. Exploring mechanical and durability properties of ultra-high performance concrete incorporating various steel fiber lengths and dosages. *Construction and Build. Mater.* 75, 429–441.
- Abellán-García, J., 2020. Four-layer perceptron approach for strength prediction of UHPC. *Construction and Build. Mater.* 256, 119465.
- Abellán-García, J., Guzmán-Guzmán, J.S., 2021. Random forest-based optimization of UHPFRC under ductility requirements for seismic retrofitting applications. *Construction and Build. Mater.* 285, 122869.
- Adamu, M., Mohammed, B.S., Shahir Liew, M., 2018. Mechanical properties and performance of high volume fly ash roller compacted concrete containing crumb rubber and nano silica. *Construction and Build. Mater.* 171, 521–538.
- Ahlborn, T.M., Peuse, E.J., Misson, D.L., 2008. Ultra-high Performance Concrete For Michigan bridges, Material performance: Phase I. Michigan Department of Transportation.
- Ahmed, T., Elchalakani, M., Karrech, A., Dong, M., Mohamed Ali, M.S., Yang, H., 2021. ECO-UHPC with high-volume class-F fly ash: new insight into mechanical and durability properties. *J. Mater. Civil Eng.* 33 (7), 04021174.
- Alharbi, Y.R., Abadel, A.A., Mayhoub, O.A., Kohail, M., 2021. Effect of using available metakaoline and nano materials on the behavior of reactive powder concrete. *Construction and Build. Mater.* 269, 121344.
- Alsalmán, A., Dang, C.N., Martí-Vargas, J.R., Micah Hale, W., 2020. Mixture-proportioning of economical UHPC mixtures. *J. Build. Eng.* 27, 100970.
- Ashkezari, G.D., Fotouhi, F., Razmara, M., 2020. Experimental relationships between steel fiber volume fraction and mechanical properties of ultra-high performance fiber-reinforced concrete. *J. Build. Eng.* 32, 101613.
- Björck, Å., 1990. Least squares methods. *Handbook of Numerical Anal.* 1 465–652.
- Bonneau, O., Vernet, C., Moranville, M., Aitcin, P.-C., 2000. Characterization of the granular packing and percolation threshold of reactive powder concrete. *Cement and Concrete Res.* 30 (12), 1861–1867.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Cain, M.K., Zhang, Z., Yuan, K.-H., 2017. Univariate and multivariate skewness and kurtosis for measuring nonnormality: prevalence, influence and estimation. *Behav. Res. Methods* 49 (5), 1716–1735.
- Chan, Y.-W., Chu, S.-H., 2004. Effect of silica fume on steel fiber bond characteristics in reactive powder concrete. *Cement and Concrete Res.* 34 (7), 1167–1172.
- Charron, J.-P., Denarié, E., Brühwiler, E., 2007. Permeability of ultra high performance fiber reinforced concretes (UHPFRC) under high stresses. *Mater. Structures* 40 (3), 269–277.
- Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.
- Chiaia, B., Fantilli, A.P., Guerini, A., Volpatti, G., Zampini, D., 2014. Eco-mechanical index for structural concrete. *Construction and Build. Mater.* 67, 386–392.
- Chu, S.H., Kwan, A.K.H., 2019. Mixture design of self-levelling ultra-high performance FRC. *Construction and Build. Mater.* 228, 116761.
- Copeland, M., Soh, J., Puca, A., Manning, M., Gollob, D., 2015. *Microsoft Azure and Cloud Computing*. Microsoft Azure. Springer, pp. 3–26.
- Corinaldesi, V., 2012. The study of using fly ash to produce ultra high performance fibre reinforced concrete. *Adv. Mat. Res.* 535, 1889–1892.
- Corinaldesi, V., Moriconi, G., 2012. Mechanical and thermal evaluation of ultra high performance fiber reinforced concretes for engineering applications. *Construction and Build. Mater.* 26 (1), 289–294.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach Learn* 20 (3), 273–297.
- Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., Singer, Y., 2006. Online passive aggressive algorithms. *J. Machine Learn. Res.* 551–585.
- Doiron, G., 2016. In: *Pier repair/retrofit Using UHPC—Examples of completed projects in north america*, International Interactive Symposium on Ultra-High Performance Concrete. Iowa State University Digital Press.
- Du, J., Meng, W., Khayat, K.H., Bao, Y., Guo, P., Lyu, Z., Abu-oheidah, A., Nassif, H., Wang, H., 2021. New development of ultra-high-performance concrete (UHPC). *Compos. Part B: Eng.* 224, 109220.
- Fan, D., Rui, Y., Liu, K., Tan, J., Shui, Z., Wu, C., Wang, S., Guan, Z., Hu, Z., Su, Q., 2021a. Optimized design of steel fibres reinforced ultra-high performance concrete (UHPC) composites: towards to dense structure and efficient fibre application. *Construction and Build. Mater.* 273, 121698.
- Fan, D., Yu, R., Fu, S., Yue, L., Wu, C., Shui, Z., Liu, K., Song, Q., Sun, M., Jiang, C., 2021b. Precise design and characteristics prediction of ultra-high performance concrete (UHPC) based on artificial intelligence techniques. *Cement and Concrete Compos.* 122, 104171.
- Fan, D., Yu, R., Shui, Z., Wu, C., Wang, J., Su, Q., 2020. A novel approach for developing a green Ultra-High Performance Concrete (UHPC) with advanced particles packing meso-structure. *Construction and Build. Mater.* 265, 120339.
- Funk, J.E., Dinger, D.R., 2013. *Predictive Process Control of Crowded Particulate Suspensions: Applied to Ceramic Manufacturing*. Springer Science & Business Media.
- Fusi, N., Sheth, R., Elibol, M., 2018. Probabilistic matrix factorization for automated machine learning. *Adv. Neural Inf. Process Syst.* 31, 3348–3357.
- Gaudillière, N., Duballet, R., Bouyssou, C., Mallet, A., Roux, P., Zakeri, M., Dirrenberger, J., 2018. Large-scale Additive Manufacturing of Ultra-High-Performance Concrete of Integrated Formwork For Truss-Shaped pillars, Robotic Fabrication in Architecture, Art and Design. Springer, pp. 459–472.
- Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* 185, 1–17.
- Gesoglu, M., Güneysi, E., Asaad, D.S., Muhyaddin, G.F., 2016. Properties of low binder ultra-high performance cementitious composites: comparison of nanosilica and microsilica. *Construction and Build. Mater.* 102, 706–713.
- Geurts, P., Ernst, D., Wehenkel, L.J.M.I., 2006. Extremely Randomized Trees 63 (1), 3–42.
- Ghafari, E., Costa, H., Júlio, E., 2015. Statistical mixture design approach for eco-efficient UHPC. *Cement and Concrete Compos.* 55, 17–25.
- Ghafari, E., Costa, H., Júlio, E., Portugal, A., Durães, L., 2014. The effect of nanosilica addition on flowability, strength and transport properties of ultra high performance concrete. *Mater. Des.* 59, 1–9.
- Ghafari, E., Ghafari, S.A., Costa, H., Júlio, E., Portugal, A., Durães, L., 2016. Effect of supplementary cementitious materials on autogenous shrinkage of ultra-high performance concrete. *Construction and Build. Mater.* 127, 43–48.
- Ghavami, S., Naseri, H., Jahanbakhsh, H., Moghadas Nejad, F., 2021. The impacts of nano-SiO₂ and silica fume on cement kiln dust treated soil as a sustainable cement-free stabilizer. *Construction and Build. Mater.* 285, 122918.
- Graybeal, B.A., 2007. Compressive Behavior of ultra-high-performance fiber-reinforced concrete. *ACI Mater. J.* 104 (2).
- Graybeal, B.A., Hartmann, J.L., 2003. Strength and durability of ultra-high performance concrete. In: *Concrete Bridge Conference*, pp. 1–20.
- Guo, P., Meng, W., Xu, M., Li, V.C., Bao, Y., 2021. Predicting mechanical properties of high-performance fiber-reinforced cementitious composites by integrating micromechanics and machine learning. *Materials (Basel)* 14 (12), 3143.
- Guvensoy, G., Bayramov, F., Ilki, A., Sengul, C., Tasdemir, A., Kocatürk, N., Yerlikaya, M., 2004. Mechanical behavior of high performance steel fiber reinforced cementitious composites under cyclic loading condition. In: *Proceedings of the International Symposium on UHPC*, pp. 649–660.
- Habel, K., Viviani, M., Denarié, E., Brühwiler, E., 2006. Development of the mechanical properties of an ultra-high performance fiber reinforced concrete (UHPFRC). *Cement and Concrete Res.* 36 (7), 1362–1370.
- Hain, A., Zaghi, A.E., Fields, T., Barakat, R., Cardinali, A., Culmo, M.P., Lopata, T., 2019. Implementation of UHPC for the repair of a steel bridge with corrosion damage in

- CT. In: International Interactive Symposium on Ultra-High Performance Concrete, 1 ed. Iowa State University Digital Press.
- Hassan, A., Jones, S., Mahmud, G., 2012. Experimental test methods to determine the uniaxial tensile and compressive behaviour of ultra high performance fibre reinforced concrete (UHPRFC). *Construction and Build. Mater.* 37, 874–882.
- Hassan, A.M., 2013. Ultra High Performance Fibre Reinforced Concrete For Highway Bridge Applications. University of Liverpool.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12 (1), 55–67.
- Huang, W., Kazemi-Kamyab, H., Sun, W., Scrivener, K., 2017. Effect of cement substitution by limestone on the hydration and microstructural development of ultra-high performance concrete (UHPC). *Cement and Concrete Compos.* 77, 86–101.
- Hwang, C.-L., Yoon, K., 1981. Methods for multiple attribute decision making. In: Hwang, C.-L., Yoon, K. (Eds.), *Methods for multiple attribute decision making. Multiple Attribute Decision Making: Methods and Applications A State-Of-The-Art Survey* 58–191.
- Ishibuchi, H., Tsukamoto, N., Nojima, Y., 2008. Evolutionary Many-Objective optimization: A short Review. *IEEE*, pp. 2419–2426.
- Jiang, G., Rong, Z., Sun, W., 2015. Effects of metakaolin on mechanical properties, pore structure and hydration heat of mortars at 0.17w/b ratio. *Construction and Build. Mater.* 93, 564–572.
- Kamthe, S., Assefa, S., Deisenroth, M., 2021. Copula flows for synthetic data generation. *arXiv preprint arXiv:2101.00598*.
- Kang, S.-H., Jeong, Y., Tan, K.H., Moon, J., 2018. The use of limestone to replace physical filler of quartz powder in UHPRFC. *Cement and Concrete Compos.* 94, 238–247.
- Karim, R., Shafei, B., 2021. Flexural response characteristics of ultra-high performance concrete made with steel microfibers and macrofibers. *Structural Concrete* 22 (6), 3476–3490.
- Kathirvel, P., Sreekumar, S., 2021. Sustainable development of ultra high performance concrete using geopolymer technology. *J. Build. Eng.* 39, 102267.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* 3146–3154.
- Kennedy, D., Habel, K., Fraser, G., 2015. Ultra high-performance concrete column jacket retrofit for the mission bridge. In: 11th Canadian Conference on Earthquake Engineering.
- Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Lawrence, N.D., Urtasun, R., 2009. Non-linear matrix factorization with Gaussian processes. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. Montreal, Quebec, Canada. Association for Computing Machinery, pp. 601–608.
- Li, K., Chen, R., Fu, G., Yao, X., 2019. Two-archive evolutionary algorithm for constrained multiobjective optimization. *IEEE Trans. Evolutionary Computation* 23 (2), 303–315.
- Li, S., Cheng, S., Mo, L., Deng, M., 2020. Effects of steel slag powder and expansive agent on the properties of ultra-high performance concrete (UHPC): based on a case study. *Materials (Basel)* 13 (3), 683.
- Li, T., Liu, X., Su, S., 2018. Semi-supervised text regression with conditional generative adversarial networks. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 5375–5377.
- Liu, J., Guo, R., 2018. Applications of steel slag powder and steel slag aggregate in ultra-high performance concrete. *Adv. Civil Eng.* 2018, 1426037.
- Liu, J., Khayat, K.H., Shi, C., 2020. Effect of superabsorbent polymer characteristics on rheology of ultra-high performance concrete. *Cement and Concrete Compos.* 112, 103636.
- Liu, Z., El-Tawil, S., Hansen, W., Wang, F., 2018. Effect of slag cement on the properties of ultra-high performance concrete. *Construction and Build. Mater.* 190, 830–837.
- Long, G., Gao, Y., Xie, Y., 2015. Designing more sustainable and greener self-compacting concrete. *Construction and Build. Materials* 84, 301–306.
- Lu, Z., Feng, Z.-g., Yao, D., Li, X., Ji, H., 2021. Freeze-thaw resistance of Ultra-High performance concrete: dependence on concrete composition. *Construction and Build. Mater.* 293, 123523.
- Ma, J., Dietz, J., Dehn, F., 2002. Ultra high performance self compacting concrete. *Lacer* 7, 33–42.
- Mahjoubi, S., Bao, Y., 2022. The key material properties of ultra-high-performance concrete (UHPC). *Mendeley Data*, 3 ed. www.doi.org/10.17632/dd62d5hyzr.3.
- Mahjoubi, S., Barhemat, R., Guo, P., Meng, W., Bao, Y., 2021. 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, 129665.
- Mahjoubi, S., Meng, W., Bao, Y., 2022. Auto-tuning learning framework for prediction of flowability, mechanical properties, and porosity of ultra-high-performance concrete (UHPC). *Appl. Soft Comput.* 115, 108182.
- Mallery, P., George, D., 2000. SPSS For Windows Step By Step. Allyn & Bacon, Inc.
- Meng, W., Khayat, K.H., 2018. Effect of hybrid fibers on fresh properties, mechanical properties, and autogenous shrinkage of cost-effective UHPC. *J. Mater. Civil Eng.* 30 (4), 04018030.
- Meng, W., Samaranyake, V.A., Khayat, K.H., 2018. Factorial design and optimization of ultra-high-performance concrete with lightweight sand. *ACI Mater. J.* 115 (1).
- Meng, W., Valipour, M., Khayat, K.H., 2016. Optimization and performance of cost-effective ultra-high performance concrete. *Mater. Structures* 50 (1), 29.
- Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
- Mo, Z., Wang, R., Gao, X., 2020. Hydration and mechanical properties of UHPC matrix containing limestone and different levels of metakaolin. *Construction and Build. Mater.* 256, 119454.
- Mosaberpanah, M.A., Eren, O., Tarassoly, A.R., 2019. The effect of nano-silica and waste glass powder on mechanical, rheological, and shrinkage properties of UHPC using response surface methodology. *J. Mater. Res. Technol.* 8 (1), 804–811.
- Mukunthu, D., Gillett, S., 2018. Announcing Automated ML Capability in Azure Machine Learning. Microsoft Azure.
- Müller, H.S., Haist, M., Vogel, M., 2014. Assessment of the sustainability potential of concrete and concrete structures considering their environmental impact, performance and lifetime. *Construction and Build. Mater.* 67, 321–337.
- Muttoni, A., Brauen, U., Jaquier, J.-L., Moullet, B., 2013. A new roof for the Olympic museum at Lausanne, Switzerland. In: *Proceedings of International Symposium on Ultra-High Performance Fiber-Reinforced Concrete*, pp. 69–76.
- Panichella, A., 2019. An adaptive evolutionary algorithm based on non-euclidean geometry for many-objective optimization. pp. 595–603.
- Patki, N., Wedge, R., Veeramachaneni, K., 2016. The synthetic data vault. In: *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 399–410.
- Pourbaba, M., Asefi, E., Sadaghian, H., Mirmiran, A., 2018. Effect of age on the compressive strength of ultra-high-performance fiber-reinforced concrete. *Construction and Build. Mater.* 175, 402–410.
- Prem, P.R., Murthy, A.R., Bharatkumar, B.H., 2015. Influence of curing regime and steel fibres on the mechanical properties of UHPC. *Magazine of Concrete Res.* 67 (18), 988–1002.
- Rajasekar, A., Arunachalam, K., Kottaisamy, M., 2019. Assessment of strength and durability characteristics of copper slag incorporated ultra high strength concrete. *J. Clean. Prod.* 208, 402–414.
- Richard, P., Cheyrezy, M., 1995. Composition of reactive powder concretes. *Cement and Concrete Res.* 25 (7), 1501–1511.
- Roberti, F., Cesari, V.F., de Matos, P.R., Pelisser, F., Pilar, R., 2021. High- and ultra-high-performance concrete produced with sulfate-resisting cement and steel microfiber: autogenous shrinkage, fresh-state, mechanical properties and microstructure characterization. *Construction and Build. Mater.* 268, 121092.
- Sadrnontazi, A., Tajasosi, S., Tahmouresi, B., 2018. Effect of materials proportion on rheology and mechanical strength and microstructure of ultra-high performance concrete (UHPC). *Construction and Build. Mater.* 187, 1103–1112.
- Sadrossadat, E., Basarir, H., Karrech, A., Elchalakani, M., 2021. Multi-objective mixture design and optimisation of steel fiber reinforced UHPC using machine learning algorithms and metaheuristics. *Eng. Comput.*
- Seada, H., Deb, K., 2014. U-NSGA-III: a unified evolutionary algorithm for single, multiple, and many-objective optimization. *COIN Report 2014022*.
- Šeps, K., Broukalová, I., Chylik, R., Cement Substitutions in Uhp and Their Influence On Principal Mechanical-Physical Properties, 1 ed. IOP Publishing, p. 012009.
- Shi, Y., Long, G., Ma, C., Xie, Y., He, J., 2019. Design and preparation of ultra-high performance concrete with low environmental impact. *J. Clean. Prod.* 214, 633–643.
- Song, Q., Yu, R., Shui, Z., Wang, X., Rao, S., Lin, Z., 2018a. Optimization of fibre orientation and distribution for a sustainable Ultra-High Performance Fibre Reinforced Concrete (UHPRFC): experiments and mechanism analysis. *Construction and Build. Mater.* 169, 8–19.
- Song, Q., Yu, R., Wang, X., Rao, S., Shui, Z., 2018b. A novel self-compacting ultra-high performance fibre reinforced concrete (SCUHPRFC) derived from compounded high-active powders. *Construction and Build. Mater.* 158, 883–893.
- Staquet, S., Espion, B., 2004. Early-age autogenous shrinkage of UHPC incorporating very fine fly ash or metakaolin in replacement of silica fume. In: *International Symposium on UHPC*, pp. 587–599.
- Sun, J., Ma, Y., Li, J., Zhang, J., Ren, Z., Wang, X., 2021. Machine learning-aided design and prediction of cementitious composites containing graphite and slag powder. *J. Build. Eng.* 43, 102544.
- Tafraoui, A., Escadeillas, G., Lebaill, S., Vidal, T., 2009. Metakaolin in the formulation of UHPC. *Construction and Build. Mater.* 23 (2), 669–674.
- Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. *Bullet. Am. Meteorological Soc.* 93 (4), 485–498.
- Vaitkevicius, V., Serelis, E., Hilbig, H., 2014. The effect of glass powder on the microstructure of ultra high performance concrete. *Construction and Build. Mater.* 68, 102–109.
- Vincent, T., Gholampour, A., Ozbakkaloglu, T., Ngo, T.D., 2021. Waste-based alkali-activated mortars containing low- and high-halloysite kaolin nanoparticles. *J. Clean. Prod.* 129428.
- Voo Yen, L., Poon Wai, K., Foster Stephen, J., 2010. Shear strength of steel fiber-reinforced ultrahigh- performance concrete beams without stirrups. *J. Structural Eng.* 136 (11), 1393–1400.
- Wang, C., Yang, C., Liu, F., Wan, C., Pu, X., 2012. Preparation of ultra-high performance concrete with common technology and materials. *Cement and Concrete Compos.* 34 (4), 538–544.
- Wille, K., Boisvert-Cotulio, C., 2015. Material efficiency in the design of ultra-high performance concrete. *Construction and Build. Mater.* 86, 33–43.
- Wille, K., Naaman, A.E., Parra-Montesinos, G.J., 2011. Ultra-high performance concrete with compressive strength exceeding 150MPa (22 ksi): a simpler way. *ACI Mater. J.* 108 (1).
- Wu, Z., Shi, C., He, W., 2017a. Comparative study on flexural properties of ultra-high performance concrete with supplementary cementitious materials under different curing regimes. *Construction and Build. Mater.* 136, 307–313.
- Wu, Z., Shi, C., He, W., Wang, D., 2017b. Static and dynamic compressive properties of ultra-high performance concrete (UHPC) with hybrid steel fiber reinforcements. *Cement and Concrete Compos.* 79, 148–157.

- Wu, Z., Shi, C., He, W., Wu, L., 2016a. Effects of steel fiber content and shape on mechanical properties of ultra high performance concrete. *Construction and Build. Mater.* 103, 8–14.
- Wu, Z., Shi, C., Khayat, K.H., 2016b. Influence of silica fume content on microstructure development and bond to steel fiber in ultra-high strength cement-based materials (UHSC). *Cement and Concrete Compos.* 71, 97–109.
- Xu, L., Skoularidou, M., Cuesta-Infante, A., Veeramachaneni, K., 2019. Modeling tabular data using conditional gan. *arXiv preprint arXiv:1907.00503*.
- Xu, L., Veeramachaneni, K., 2018. Synthesizing tabular data using generative adversarial networks. *arXiv preprint arXiv:1811.11264*.
- Xue, Y., Ye, J., Zhou, Q., Long, L.R., Antani, S., Xue, Z., Cornwell, C., Zaino, R., Cheng, K. C., Huang, X., 2021. Selective synthetic augmentation with HistoGAN for improved histopathology image classification. *Med. Image Anal.* 67, 101816.
- Yalçınkaya, Ç., Yazıcı, H., 2017. Effects of ambient temperature and relative humidity on early-age shrinkage of UHPC with high-volume mineral admixtures. *Construction and Build. Mater.* 144, 252–259.
- Yang, S.L., Millard, S.G., Soutsos, M.N., Barnett, S.J., Le, T.T., 2009. Influence of aggregate and curing regime on the mechanical properties of ultra-high performance fibre reinforced concrete (UHPRFC). *Construction and Build. Mater.* 23 (6), 2291–2298.
- Yu, R., Song, Q., Wang, X., Zhang, Z., Shui, Z., Brouwers, H.J.H., 2017. Sustainable development of Ultra-High Performance Fibre Reinforced Concrete (UHPFRC): towards to an optimized concrete matrix and efficient fibre application. *J. Clean. Prod.* 162, 220–233.
- Yu, R., Spiesz, P., Brouwers, H.J.H., 2014a. Effect of nano-silica on the hydration and microstructure development of Ultra-High Performance Concrete (UHPC) with a low binder amount. *Construction and Build. Mater.* 65, 140–150.
- Yu, R., Spiesz, P., Brouwers, H.J.H., 2014b. Mix design and properties assessment of ultra-high performance fibre reinforced concrete (UHPFRC). *Cement and Concrete Res.* 56, 29–39.
- Yu, R., Spiesz, P., Brouwers, H.J.H., 2015a. Development of an eco-friendly ultra-high performance concrete (UHPC) with efficient cement and mineral admixtures uses. *Cement and Concrete Compos.* 55, 383–394.
- Yu, R., Spiesz, P., Brouwers, H.J.H., 2015b. Development of ultra-high performance fibre reinforced concrete (UHPFRC): towards an efficient utilization of binders and fibres. *Construction and Build. Mater.* 79, 273–282.
- Zhang, D., Jaworska, B., Zhu, H., Dahlquist, K., Li, V.C., 2020. Engineered cementitious composites (ECC) with limestone calcined clay cement (LC3). *Cement and Concrete Compos.* 114, 103766.
- Zhang, Y., Deng, X., Zhang, Y., Zhang, Y., 2022. Generation of sub-item load profiles for public buildings based on the conditional generative adversarial network and moving average method. *Energy Build.* 268, 112185.
- Zitzler, E., Thiele, L., 1998. Multiobjective optimization using evolutionary algorithms — a comparative case study. In: Eiben, A.E., Bäck, T., Schoenauer, M., Schwefel, H.-P. (Eds.), *Multiobjective optimization using evolutionary algorithms — a comparative case study. Parallel Problem Solving from Nature* 292–301.