Journal of Cleaner Production 329 (2021) 129665

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

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

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

Department of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ, USA

ARTICLE INFO

Handling Editor: Jiri Jaromir Klemeš

Keywords:
Automated machine learning
Carbon footprint
Evolutionary algorithm
Multi-objective optimization
Strain-hardening cementitious composite (SHCC)
Tree-based pipeline optimization

ABSTRACT

This study develops a framework for property prediction and multi-objective optimization of strain-hardening cementitious composites (SHCC) based on automated machine learning. Three machine learning models are developed to predict the compressive strength, tensile strength, and ductility of SHCC. A tree-based pipeline optimization method is enhanced and used to enable automatic configuration of machine learning models, which are trained using three datasets considering 14 mix design variables and achieve reasonable prediction accuracy. With the predictive models, five objective functions are formulated for mechanical properties, life-cycle cost, and carbon footprint of SHCC, and the five objective functions are optimized in six design scenarios. The objective functions are optimized using innovative optimization and decision-making techniques (Unified Non-dominated Sorting Genetic Algorithm III and Technique for Order of Preference by Similarity to Ideal Solution). This research will promote efficient development and applications of high-performance SHCC in concrete and construction industry.

1. Introduction

Strain-hardening cementitious composite (SHCC) is a family of high-performance fiber-reinforced cementitious composites that are capable of resisting higher tensile loads after they are cracked. SHCC is also known as engineered cementitious composite and bendable concrete in different contexts. SHCC features high crack resistance, ductility, toughness, and control of crack width (Li, 2003). SHCC attracts increasing interest in concrete and construction industry since it outperforms conventional fiber-reinforced concrete in terms of post-cracking properties (Nguyễn et al., 2021; Xu et al., 2020, 2021; Meng and Khayat, 2016) and durability (Liu et al., 2017) under various threats (Jun and Mechtcherine, 2010). After cracked, SHCC has retained load-carrying capabilities with controlled crack widths usually narrower than 100 μm (Li, 2019). The fine cracks benefit long-term durability because microcracks do not highly compromise the impermeability of SHCC (Li, 2019). Besides the exceptional mechanical properties and durability, SHCC can be prepared using high-volume waste materials to reduce material cost and carbon footprint. For example, high-volume fly ash is usually used to prepare SHCC with improved sustainability and improve fresh and hardened properties (Bao et al., 2019). The use of supplementary cementitious materials also promoted self-healing of microcracks in SHCC, which further improved the long-term durability (Li and Herbert, 2012). The use of polymeric fibers such as polyvinyl alcohol (PVA) and polypropylene (PP) fibers in SHCC was capable of significantly enhancing the fire resistance and mitigating explosive spalling of SHCC exposed to high temperatures (Li et al., 2017a). Due to the superior material properties of SHCC, many successful structural applications were reported. SHCC was used to construct and repair engineering structures to enhance the resilience to various threats such as earthquake, impact, and fire, as well as service life while reducing maintenance costs (Li et al., 2017b, c; Liu et al., 2019).

Since the desired material properties of SHCC vary in different applications, the mix design of SHCC should be tailored for different use cases. In the literature, theoretical models based on micromechanics were developed to design SHCC by mechanistically tuning the matrix, fiber, and fiber-matrix interface (Kanda and Li, 1999). The micromechanics models intrinsically link the key mechanical properties of

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

https://doi.org/10.1016/j.jclepro.2021.129665
Received 15 August 2021; Received in revised form 10 October 2021; Accepted 7 November 2021
Available online 10 November 2021

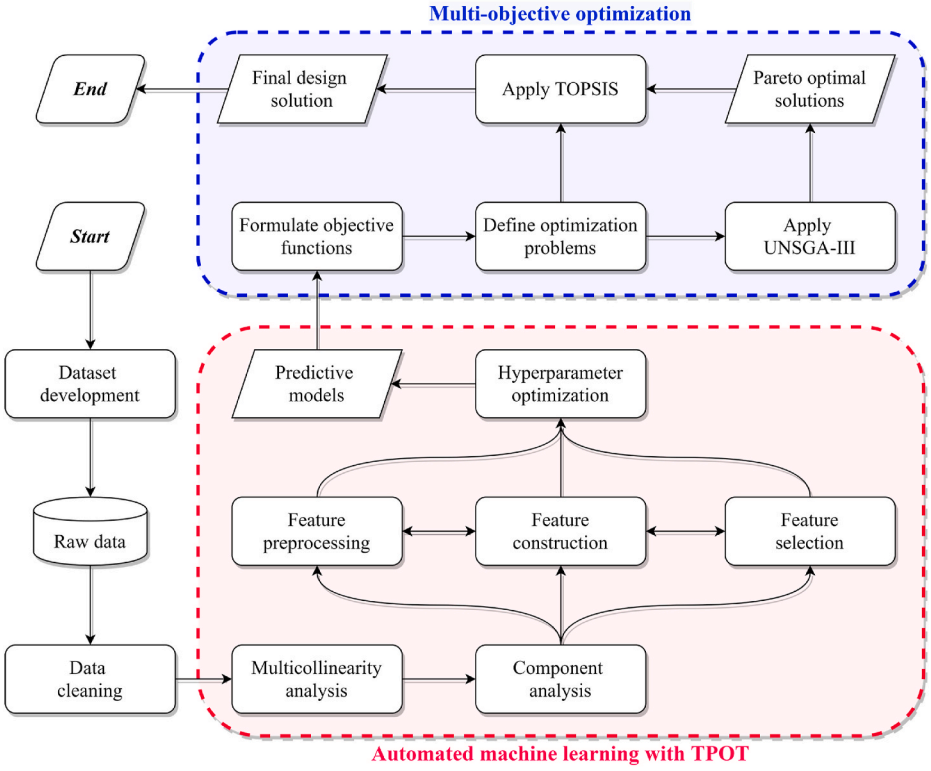

Fig. 1. Flowchart of the proposed method for prediction and multi-objective optimization.

SHCC with the material parameters that can be quantitatively evaluated through experiments such as flexural tests for evaluating the toughness of cementitious matrix and single fiber pullout tests for evaluating the fiber-matrix interface bond strength and toughness. Other important methods included lattice models and extended finite element models (Kang and Bolander, 2016; Huang and Zhang, 2016). However, the design of SHCC for practical applications still involves extensive experimental efforts that are time consuming and costly because multiple parameters must be determined through tests (Kanda and Li, 1999). The demand for efforts is exacerbated by the fact that the mechanical properties and sustainability of SHCC are sensitive to many mix design variables related to adopted raw materials (Lepech et al., 2008; Tosun-Felekoğlu et al., 2017; Zhang et al., 2016). In a nutshell, there is a lack of effective and efficient methods to simultaneously optimize the mechanical properties, cost, and sustainability of SHCC for different intended industrial applications.

Besides the methods based on theoretical models, data-driven intelligent prediction and optimization methods received increasing attention in the development of materials. For prediction of material properties, machine learning models were developed to link mix design with properties (Huang et al., 2021; Ke and Duan, 2021; Lu et al., 2021a–c; Sun et al., 2021). Trained machine learning models were used to predict compressive strengths and elastic moduli of various types of concrete (Cook et al., 2019; Han et al., 2020; Marani et al., 2020). Recently, micromechanics models were integrated with machine learning models to enable reasonable prediction of tensile strength and ductility (Guo et al., 2021a). On the other hand, advanced optimization techniques such as metaheuristic optimization were developed to optimize the mix design of concrete (Sun et al., 2021; Zhang et al., 2020), while the mechanical properties of concrete were predicted by high-fidelity predictive models. Inspired by the previous studies, it is hypothesized that predictive models based on machine learning can be integrated with optimization techniques for property prediction and mix design optimization of SHCC. It is further posited that multiple objective functions can be considered simultaneously to optimize the mix design of SHCC.

With the above hypothesis, this research aims to develop a new framework for property prediction and multi-objective optimization of SHCC, to promote the efficacy and efficiency in development and structural applications of SHCC. To this end, this research has four objectives and contributions: (1) to establish a framework integrating capabilities of property prediction and multi-objective optimization for SHCC; (2) to develop high-fidelity predictive models to predict the compressive strength, tensile strength, and ductility of SHCC; (3) to present objective functions for optimizing the compressive strength, tensile strength, ductility, material cost, and carbon footprint of SHCC; and (4) to demonstrate the framework for multi-objective optimization of SHCC mixtures in six design scenarios.

Compared with previous research on prediction and optimization of concrete with data-driven methods, this research is novel in three aspects: (1) This research develops a method to automate the configuration of machine learning models and hyperparameter tuning based on a Tree-based Pipeline Optimization Tool (TPOT) (Olson and Moore, 2019). The TPOT combines a tree representation of machine learning models with a genetic algorithm to optimize the prediction accuracy. (2) This research considers a wide range of raw materials and mix design variables in datasets used to train machine learning models and integrates multiple data processing methods to improve the quality of datasets. (3) This research utilizes Unified Non-dominated Sorting Genetic Algorithm III (UNSGA-III) (Seada and Deb, 2014) for multi-objective optimization of SHCC and selects Pareto optimal solutions using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (Hwang and Yoon, 1981). This study is expected to promote development of SHCC for the concrete and construction industry.

The remainder of the paper is organized as follows: Section 2 presents the methodology of surrogate modeling-based optimization. Section 3 discusses the developed machine learning models. Section 4 discusses the optimization process and the optimal design solutions of SHCC. Section 5 summarizes the conclusions.

Table 1
Summary statistics of the design variables.

Number	Design variable	Unit	Range	Mean	Median	Standard deviation
1	Cement-to-binder ratio	1	0.152–1	0.474	0.413	0.218
2	Fly ash-to-binder ratio	1	0–0.848	0.350	0.494	0.306
3	Slag-to-binder ratio	1	0–0.808	0.128	0.000	0.209
4	Rice husk-to-binder ratio	1	0–0.36	0.005	0.000	0.032
5	Limestone-to-binder ratio	1	0–0.577	0.025	0.000	0.083
6	Metakaolin-to-binder ratio	1	0–0.094	0.002	0.000	0.014
7	Silica fume-to-binder ratio	1	0–0.206	0.017	0.000	0.038
8	Sand-to-binder ratio	1	0–1.4	0.406	0.360	0.205
9	Water-to-binder ratio	1	0.11–0.80	0.267	0.260	0.082
10	Superplasticizer content	%	0–2.7	0.807	0.500	0.614
11	Fiber volume	%	0–3	1.890	2.000	0.434
12	Fiber length	mm	6–27	12.01	12.00	3.604
13	Fiber diameter	μm	12–39	33.49	39.00	8.611
14	Fiber Young's modulus	GPa	4–200	60.03	42.80	37.71

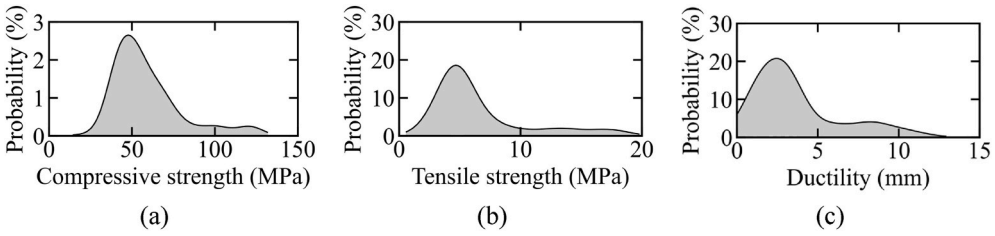

Fig. 2. Normal density functions of the investigated mechanical properties of SHCC: (a) compressive strength, (b) tensile strength, and (c) ductility.

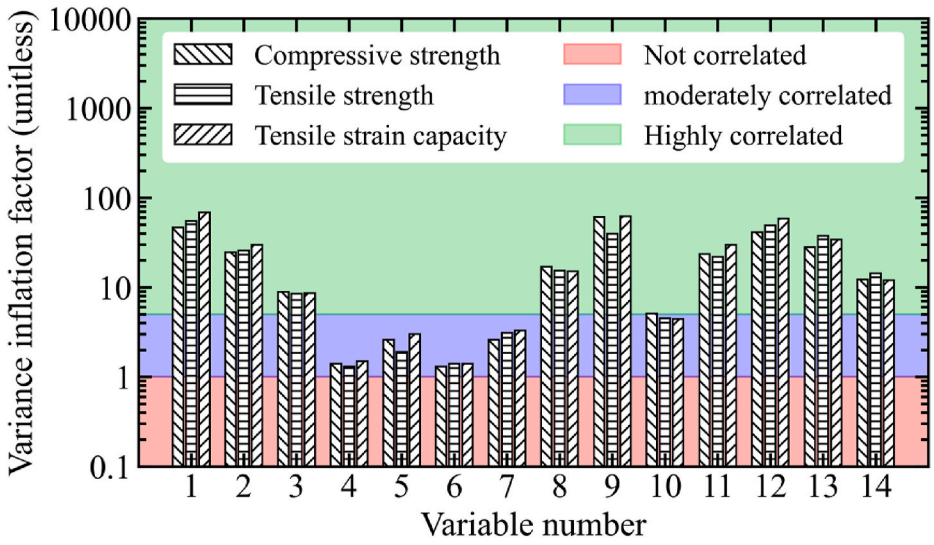

Fig. 3. The variance inflation factors corresponding to the developed datasets. The horizontal axis is in logarithmic scale.

2. Methodology

2.1. Overview

Fig. 1 shows the presented framework for data-driven prediction and design optimization of SHCC through integrating automated machine learning, data processing, and multi-objective optimization. The framework includes seven steps: (1) Three datasets are established by collecting test data from previous publications such as papers and reports, as elaborated in Section 2.2. (2) Data cleaning is performed using the isolation forest method to identify and remove anomalous data from the three datasets, as elaborated in Section 2.3. (3) High-fidelity predictive models are developed using the processed datasets and the TPOT for automated machine learning and hyperparameter tuning, as elaborated in Section 2.4. (4) Different design objectives are defined using the surrogate models, as elaborated in Section 2.5. (5) Multi-objective optimization problems are defined, as elaborated in Section 2.6. (6) The UNSGA-III is applied to solve the defined multi-objective optimization problems, as elaborated in Section 2.7. (7) The TOPSIS is used to select the optimal solution, as elaborated in Section 2.8.

2.2. Development of dataset

Datasets are essential for development of predictive models based on machine learning. This study develops three large datasets for the compressive strength, tensile strength, and ductility of SHCC by extracting 745 test data from references (Ding et al., 2018a, 2018b; Guo et al., 2021b; Kim et al., 2007; Lepech et al., 2008; Li et al., 2002, 2017a,

Table 2
Description of the design objectives.

Design objective	Description	Goal
F_1	Compressive strength	Maximization
F_2	Tensile strength	Maximization
F_3	Ductility	Maximization
F_4	Carbon footprint	Minimization
F_5	Material cost	Minimization

2019, 2019; Lin et al., 2020; Said and Razak, 2015; Turk and Nehdi, 2018; Wang et al., 2020; Xu et al., 2019a, 2019b; Yu et al., 2017, 2018, 2019, 2020, 2019; Zheng et al., 2018; Zhou et al., 2010, 2019; Zhu et al., 2012). The numbers of data points for the compressive strength, tensile strength, and ductility are 264, 244, and 237, respectively. A total of 14 design variables are selected according to reference (Guo et al., 2021a). Table 1 shows the design variables and their varying ranges, the mean and median values, as well as the standard deviation. The table indicates large variations of the units, distributions, and ranges of the variables.

Statistical analysis was performed to evaluate the distributions of mechanical properties of SHCC in the established datasets. Fig. 2 shows the normal density functions of the compressive strength, tensile strength, and ductility. In this research, ductility refers to the ultimate tensile strain corresponding to peak tensile stress in the constitutive relationship of a SHCC mixture. The average values of the compressive strength, tensile strength, and ductility are 50 MPa, 5 MPa, and 3%, respectively.

Multicollinearity is a common problem in multivariate regression analysis and compromises the prediction accuracy when the input variables are highly correlated (Chatterjee and Hadi, 2015). This study diagnoses existence of multicollinearity using the variance inflation factor as elaborated in reference (Moré, 1978). Overall, the variance inflation factor increases with the multicollinearity extent. When there is no multicollinearity, the variance inflation factor is equal to one, which is the smallest value. A variance inflation factor that is larger than five indicates significant multicollinearity (James et al., 2013). Fig. 3 shows that 9 out of 14 variables are highly correlated, indicating that it is necessary to process the data to eliminate multicollinearity.

2.3. Data cleaning

Datasets developed using previous publications generally contain anomalous data. Anomalies can be generated by different causes such as the operation errors in experiments, data entry, and data processing (Zimek and Schubert, 2017). This study performs data cleaning to identify and remove anomalous data using a decision tree algorithm based on the isolated forest method (Liu et al., 2008). The idea is that anomalous data have different features and can be separated from the normal data. The isolated forest method has two steps: (1) to train an ensemble of binary decision trees based on a dataset; and (2) to evaluate the anomaly score of each data point based on the ensemble model.

2.4. Tree-based pipeline optimization (TPOT)

The prediction accuracy of a machine learning model depends on its architecture, dataset, and hyperparameters (Pappa, 2020). It is tedious and unreliable to manually optimize the architecture, dataset, and hyperparameters. This study proposes to maximize the prediction accuracy through employing the TPOT that automatically configures machine learning models, processes dataset, and tunes hyperparameters based on a genetic algorithm (Olson and Moore, 2019; Willis et al., 1997). In this study, the TPOT considers various types of data processing and machine learning operators. Data processing is conducted to prepare the input data for training of machine learning models. In data processing, multiple algorithms are considered for feature preprocessing and construction. Feature preprocessing is applied to make the variables more meaningful and representative, while feature construction is applied to generate new variables and eliminate old variables, aiming to improve the performance of machine learning model. Multiple machine learning operators, including linear methods such as logistic regression and Ridge, individual tree-based methods such as decision tree, and ensemble tree-based models such as XGBoost and random forest (Olson and Moore, 2019) are considered by TPOT. Combination of machine learning operators is also considered using a stacking technique, which applies multiple machine learning operators to the dataset and uses the predictions of operators as the inputs of a machine learning operator, called meta-learner (Sikora, 2015).

Since Section 2.2 indicates occurrence of multicollinearity, two methods are applied to eliminate multicollinearity, which are principal component analysis (Wold et al., 1987) and independent component analysis (Comon, 1994). Although TPOT considers the two component analysis methods for configuration, it is possible that neither of the methods is selected by TPOT. Therefore, TPOT is modified to select a component analysis method if multicollinearity is existed. TPOT algorithm follows the following four sequential steps: (1) A set of 100 machine learning models with random configurations are generated, and the cross-validation accuracy of the models is evaluated using the established dataset. (2) The models are evaluated based on two criteria, which are cross-validation accuracy and number of operators. The top 20 machine learning models are selected by constructing a Pareto-front based on the two criteria according to NSGA-II (Deb et al., 2002). (3) New machine learning models are generated by performing three genetic operations (crossover, mutation, and selection) (Olson and Moore, 2019; Willis et al., 1997) with the 20 selected models. (4) Steps 2 and 3 are repeated for 100 times. The configuration of the machine learning model and the hyperparameters determined from the above steps are elaborated in Section 3.

2.5. Design objectives

Five objectives are considered for optimization of SHCC, as listed in

Table 3
The density, carbon footprint, and cost of materials.

Number	Material	Density (kg/m^3)	CO_2-e (kg/kg)	Cost ($/kg)	References
1	Cement	3130	0.832	0.11	(Chiaia et al., 2014; Heirman et al., 2009)
2	Fly ash	2350	0.009	0.046	(Heirman et al., 2009; Purnell, 2013)
3	Slag	2880	0.019	0.10	(Aydın and Baradan, 2013; Purnell, 2013)
4	Rice husk ash	2140	0.013	0.113	(Gursel et al., 2016; Hossain and Elsayed, 2018; Sathawane et al., 2013)
5	Limestone powder	2700	0.017	0.122	(Heirman et al., 2009; Müller et al., 2014)
6	Metakaolin	2620	0.42	0.50	(Heath et al., 2014; Poon et al., 2001)
7	Silica fume	2170	0.0003	0.50	(Habert et al., 2013; Nepomuceno et al., 2012)
8	Sand	2640	0.0025	0.014	(Chiaia et al., 2014; Heirman et al., 2009)
9	Water	1000	0.0003	0	(Chiaia et al., 2014; Heirman et al., 2009)
10	Superplasticizer	1080	0.720	3.0	(Heirman et al., 2009; Long et al., 2015)
11	PVA fiber	1300	3.43	8.0	(Pakravan and Ozbakkaloglu, 2019; Yew et al., 2015)
12	PE fiber	970	4.08	11.0	(Pakravan and Ozbakkaloglu, 2019; Ranade, 2014; Yu et al., 2020)
13	PP fiber	900	2.70	1.75	(Ahangari et al., 2013; Chen et al., 2021; Pakravan and Ozbakkaloglu, 2019)
14	Steel fiber	7800	2.75	4.50	(Chen et al., 2021; Chiaia et al., 2014; Pakravan and Ozbakkaloglu, 2019)

Table 4
The physical and mechanical properties of fibers.

Fiber type	Length (mm)	Diameter (μm)	Young's modulus (GPa)
PVA	8–12	39	42.8
PE	12–18	20–25	100
PP	10–12	12–18	6
Steel	27	40	200

Table 2. Three objective functions (F_1, F_2, and F_3) are defined for the mechanical properties, including compressive strength, tensile strength, and ductility. The fourth objective function (F_4) is defined to enhance sustainability by minimizing carbon footprint. The unit carbon footprint (CF) for 1 m^3 of SHCC is defined as:

$$CF = \sum_{k=0}^{n} M_k \times CO_{2-e,k} \tag{1}$$

where n is the number of the ingredient types; M_k is the mass of k-th ingredient; $CO_{2-e,k}$ is the carbon dioxide equivalent of the k-th ingredient.

The fifth objective function (F_5) is defined to improve economic benefits by minimizing material cost. The unit material cost for 1 m^3 of SHCC is defined as:

$$C = \sum_{k=0}^{n} M_k \times c_k \tag{2}$$

where c_k is the cost of 1 kg of the k-th ingredient.

Table 3 lists the inventory for the unit carbon footprint and cost as well as density of ingredients of SHCC. Seven types of binder are considered, including cement and six types of supplementary cementitious materials such as slag, fly ash, and rice husk ash, which were otherwise landfilled as industrial solid wastes. Four types of fibers are considered, which are PVA fiber, PP fiber, polyethylene (PE) fiber, and steel fiber. The unit cost of each ingredient is average of different values reported in references. The physical and mechanical properties of the fibers are listed in Table 4.

2.6. Multi-objective design scenarios

Based on the five design objectives, this study investigates six multi-objective design problems for optimization of SHCC to accommodate different applications and scenarios, as listed in Table 5. DS_1 and DS_4 are defined to maximize the mechanical properties of SHCC; DS_2 and DS_5 are defined to maximize the mechanical performance and minimize the carbon footprint; DS_3 and DS_6 are defined to maximize the mechanical performance and minimize the material cost and carbon footprint.

2.7. Multi-objective optimization

This study adopts an evolutionary algorithm UNSGA-III for multi-objective optimization (Seada and Deb, 2014). UNSGA-III was developed based on the Non-dominated Sorting Genetic Algorithm III (NSGA-III) (Deb and Jain, 2014), which was previously developed and commonly used for multi-objective optimization (Seada and Deb, 2014). In general, a multi-objective optimization problem involves conflicting objectives, so in general, the optimal solutions of different objectives are different. UNSGA-III is employed to generate a set of optimal solutions, called Pareto optimal solutions. With a Pareto optimal solution, none of the objective functions can be further improved without degrading the other objective functions, as elaborated in reference (Seada and Deb, 2014). The numbers of iterations and population size are set to 300 and 100. The two values are selected by trial and error to obtain near-optimal solutions and avoid premature convergence. A design constraint is imposed to ensure that the mass of binder is equal to the sum of mass of cement and supplementary cementitious materials:

Table 5
The proposed multi-objective design scenarios.

Design scenarios	Design objective				
	F_1	F_2	F_3	F_4	F_5
DS_1	×	✓	✓	×	×
DS_2	×	✓	✓	✓	×
DS_3	×	✓	✓	✓	✓
DS_4	✓	✓	✓	×	×
DS_5	✓	✓	✓	✓	×
DS_6	✓	✓	✓	✓	✓

Note: the symbol "×" means "not included", and the symbol "✓" means "included".

$$\left(\sum_{i=1}^{7} V_i - 1\right)^2 < \varepsilon \tag{3}$$

where V_i is the i-th design variable in Table 1; and ε is the tolerance, which is set to 0.001. The upper and lower bounds of the variables are defined in Table 1.

2.8. Selection of optimal solutions

This study adopts TOPSIS for decision-making of multi-objective optimization through ranking the Pareto optimal solutions obtained from UNSGA-III (Hwang and Yoon, 1981). The Pareto optimal solution ranked as the highest is selected as the final design solution of SHCC. The basic mechanism of TOPSIS is to determine the distance from each solution to the ideal solution and the distance from each solution to the worst solution in a n-dimensional objective space, where n is the number of objective functions. The solution with the first rank has the shortest distance from the ideal solution and the longest distance from the worst solution. The process of TOPSIS to rank m number of Pareto optimal solutions of a multi-objective problem with n number of objective functions is as follows:

- The objective function values of the Pareto optimal solutions are normalized by:

$$\eta_{i,j} = \frac{a_{i,j}}{\sqrt{\sum_{k=1}^{m} a_{k,j}^2}} \tag{4}$$

where $a_{i,j}$ and $\eta_{i,j}$ are the actual and normalized values of the j-th objective function value corresponding to the i-th Pareto optimal solution.

- The ideal and worst solutions can be determined by:

$$S^+ = \left\{ \left(Min\left(\eta_{i,j}\right) \middle| \, i = 1,\ 2, \ldots, m \,|\, j \in J^-\right),\ \left(Max\left(\eta_{i,j}\right) \middle| \, i = 1,\ 2, \ldots, m \,|\, j \in J^+\right) \right\} = \left\{ s_1^+, s_2^+, \ldots, s_n^+ \right\} \tag{5a}$$

$$S^- = \left\{ \left(Max\left(\eta_{i,j}\right) \middle| \, i = 1,\ 2, \ldots, m \,|\, j \in J^-\right),\ \left(Min\left(\eta_{i,j}\right) \middle| \, i = 1,\ 2, \ldots, m \,|\, j \in J^+\right) \right\} = \left\{ s_1^-, s_2^-, \ldots, s_n^- \right\} \tag{5b}$$

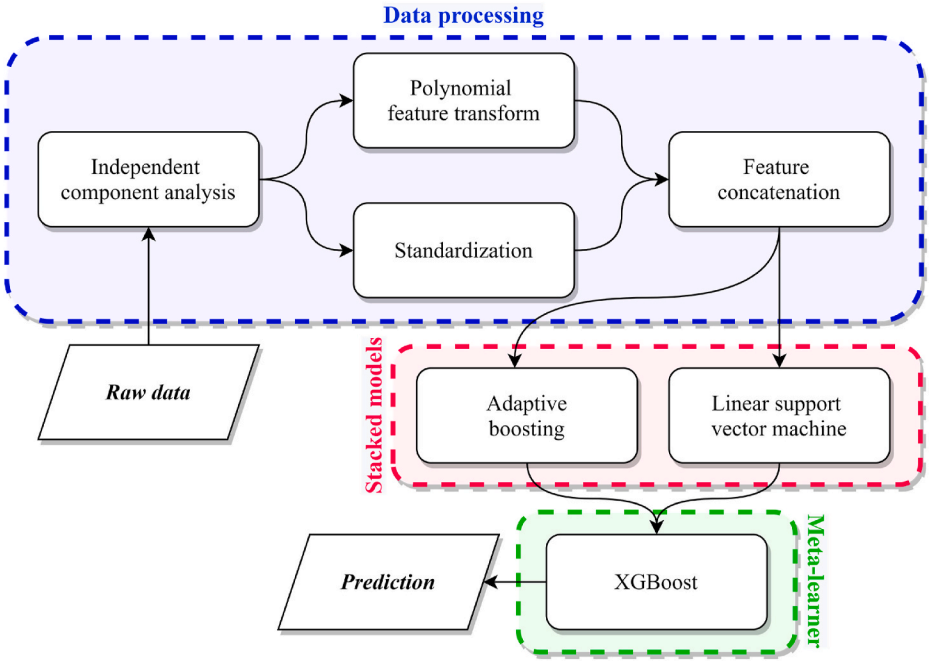

Fig. 4. Depiction of the configuration of the machine learning model determined by the TPOT.

Table 6
The hyperparameters of the machine learning models.

Machine learning model	Hyperparameter name	Optimal hyperparameter
SVM	Epsilon	1
	Loss function	L1 loss
	Regularization parameter	20
AdaBoost	Learning rate	0.01
	Loss function	Squared error
	Number of estimators	200
XGBoost	Learning rate	0.1
	Loss function	Squared error
	Maximum depth	9
	Minimum child weight	4
	Number of estimators	100
	Subsample ratio	0.9

where S^+ and S^- are the ideal and worst solutions; n and m are the number of objective functions and solutions; J^+ is associated with the objective functions to be maximized; J^- is associated with the objective functions to be minimized.

- The L^2-norm distance from the Pareto optimal solution to the ideal and worst solutions are defined as:

$$d_{i,I} = \sqrt{\sum_{j=1}^{m} \left(\eta_{i,j} - s_j^+\right)^2} \tag{6a}$$

$$d_{i,W} = \sqrt{\sum_{j=1}^{m} \left(\eta_{i,j} - s_j^-\right)^2} \tag{6b}$$

where $d_{i,I}$ and $d_{i,W}$ are the distances from the i-th Pareto optimal solution to the ideal and the worst solutions, respectively; and s_j^+ and s_j^- are the j-th components of the ideal and the worst solutions, respectively.

- The similarity score of each of the Pareto optimal solutions can be determined as:

$$\delta_i = \frac{d_{i,W}}{d_{i,I} + d_{i,W}} \tag{7}$$

- The Pareto optimal solutions are ranked based on the similarity score. The Pareto optimal solution with the highest similarity score is ranked the first and selected as the final solution.

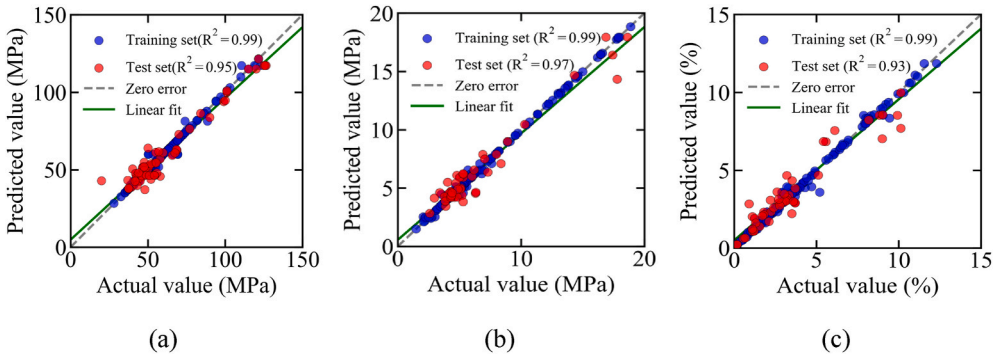

Fig. 5. Comparison of the predicted and actual values of (a) compressive strength, (b) tensile strength, and (c) ductility.

Table 7
The performance metrics of the machine learning models.

Dataset	Performance metric	Mechanical properties		
		Compressive strength	Tensile strength	Ductility
Training	R^2	0.990	0.998	0.993
	ME	9.773	0.939	1.619
	MAE	0.589	0.072	0.105
	MSE	3.614	0.031	0.047
	MAD	0.029	0.009	0.040
Testing	R^2	0.954	0.965	0.931
	ME	18.26	1.710	2.454
	MAE	4.125	0.568	0.507
	MSE	30.01	0.535	0.563
	MAD	3.165	0.467	0.287

3. Machine learning models

This section presents the machine learning models developed using the TPOT. The configuration and hyperparameter tuning of machine learning models are presented in Section 3.1. The prediction accuracy of the machine learning models is evaluated in Section 3.2.

3.1. Architecture and hyperparameters

Fig. 4 shows the configuration of machine learning model determined using the TPOT, which automates the model configuration and hyperparameter tuning processes through the steps shown in Fig. 1. Compared with TPOT in previous research, this study modifies the TPOT by incorporating a component analysis module for eliminating multicollinearity of mix design variables. With the modified TPOT, once the datasets are formed and cleaned, multiple machine learning models are combined based on a stacking technique for ensemble machine learning used to predict the properties of SHCC. The predictions from a machine learning model are used as the input variables of another predictive model, as elaborated in reference (Pavlyshenko, 2018). In this study, a linear support vector machine (SVM) (Cortes and Vapnik, 1995) and adaptive boosting (AdaBoost) (Freund and Schapire, 1996) are applied to predict the inputs of the meta-learner. Finally, an extreme gradient boosting algorithm (XGBoost) (Chen and Guestrin, 2016) is selected as the meta-learner.

Multicollinearity is eliminated through independent component analysis. New variables are generated through polynomial feature transform. The new variables include polynomial combinations of input variables. For example, if the input variables are a and b, the results of polynomial feature transform include 1, a, b, a^2, ab, and b^2 (Swamynathan, 2019). Then, the new variables are normalized by Equation (8):

$$z = \frac{x - u}{\sigma} \tag{8}$$

where x is an input variable; u is the average of the input variable; and σ is the standard deviation.

Table 6 lists the hyperparameters of machine learning models derived by the TPOT. A total of 12 hyperparameters are automatically tuned to improve the prediction accuracy of the machine learning models including SVM, AdaBoost, and XGBoost in the ensemble model.

3.2. Prediction accuracy

Five performance metrics are used to assess the developed predictive models, which are the coefficient of determination (R^2), maximum error (ME), mean absolute error (MAE), mean square error (MSE), and mean absolute deviation (MAD).

- The coefficient of determination measures the proportion of variability of actual values:

$$R^2(P,A) = 1 - \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{\sum_{i=1}^{n}[a_i - mean(a_i)]^2} \tag{9}$$

where $P = \{p_1, p_2, \ldots, p_n\}$f and $A = \{a_1, a_2, \ldots, a_n\}$ are two vectors

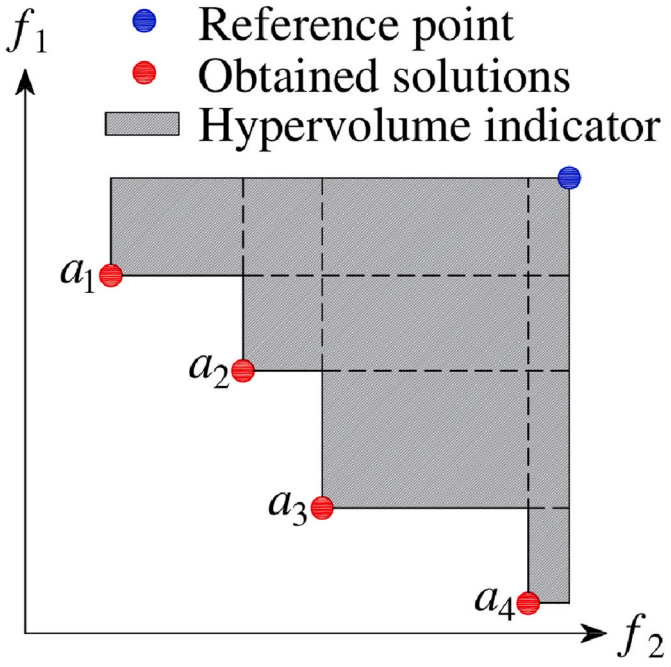

Fig. 7. Hypervolume indicator for a set of solutions determined for an optimization problem with two objective functions to be minimized, namely f_1 and f_2; a_i is the i-th solution.

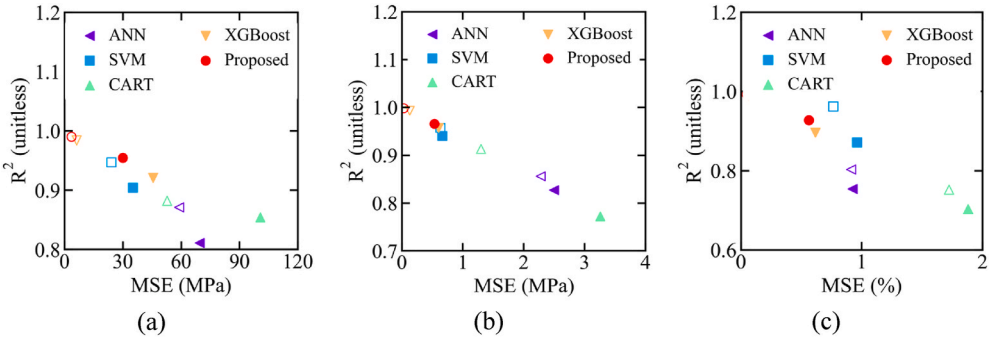

Fig. 6. Comparison of different machine learning methods for (a) compressive strength, (b) tensile strength, and (c) ductility. The filled and empty markers represent the performance of machine learning models with the test and training datasets, respectively.

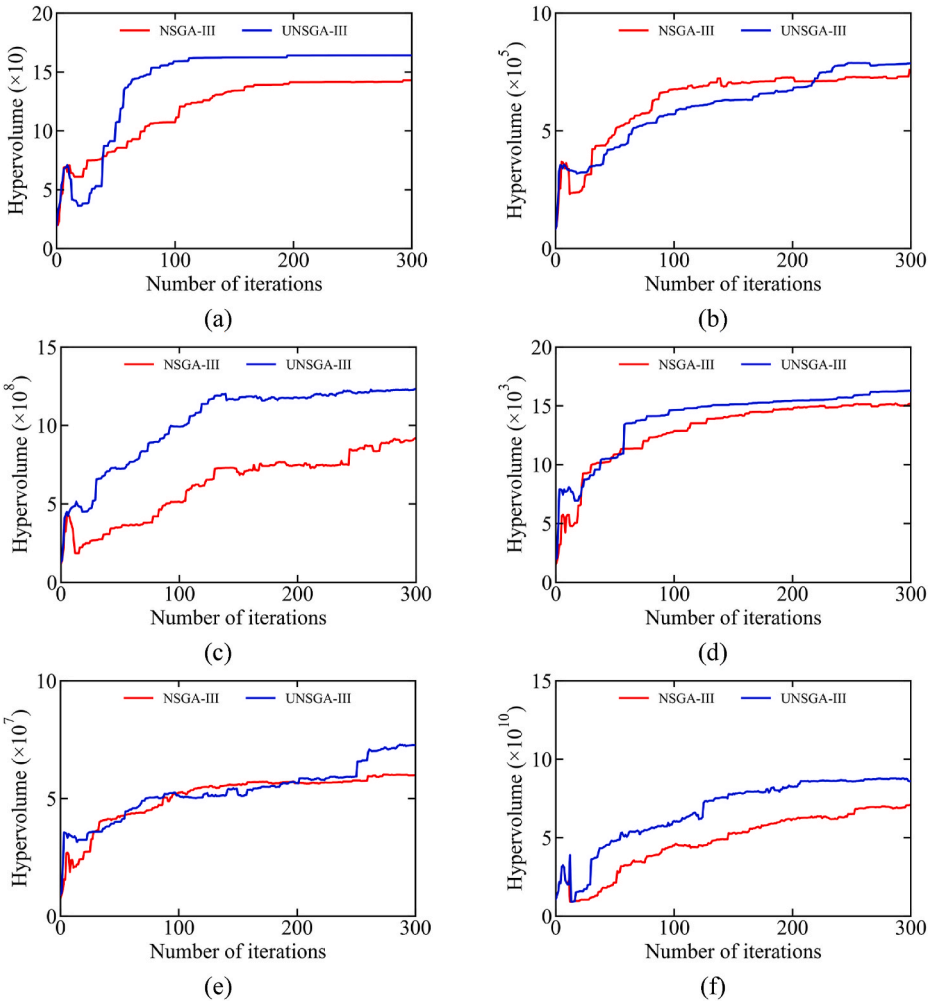

Fig. 8. Hypervolume indicator of the provided solutions for the best trials corresponding to (a) DS1, (b) DS2, (c) DS3, (d) DS4, (e) DS5, and (f) DS6.

containing the predicted and actual values; n is the number of observed data points.

- Maximum error measures the worst-case error:

$$ME(P,A) = max(|p_1 - a_1|, |p_2 - a_2|, \ldots, |p_n - a_n|) \quad (10)$$

where the notation $|.|$ denotes the absolute value.

- Mean absolute error measures the average magnitude of prediction errors:

$$MAE(P,A) = \frac{1}{n}\sum_{i=1}^{n}|p_i - a_i| \quad (11)$$

- Mean square error is the average squared difference between the predicted and actual values:

$$MSE(P,A) = \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{n} \quad (12)$$

- Mean absolute deviation (MAD) measures the variability of error:

$$MAD(P,A) = median(|p_1 - a_1|, |p_2 - a_2|, \ldots, |p_n - a_n|) \quad (13)$$

Fig. 5 compares the prediction results against the actual values. The smallest R^2 of the developed machine learning models is 0.93, implying that the developed models are capable of reasonably predicting the key mechanical properties of SHCC.

The performance metrics of the developed models for predicting the compressive strength, tensile strength, and ductility are listed in Table 7. The results corresponding to the training dataset and testing dataset show high accuracy. Therefore, the developed models are capable of predicting the mechanical properties of SHCC with reasonable accuracy.

Fig. 6 compares the prediction accuracy of the proposed method with the prediction accuracy of a previous research on predictive models based on artificial neural network (ANN), support vector machine (SVM), classification and regression trees (CART), and XGBoost as elaborated in reference (Guo et al., 2021). The proposed method demonstrates higher prediction accuracy in terms of MSE and R^2.

Although the training accuracy of the XGBoost model is comparable with the training accuracy of the proposed method, the test accuracy of the proposed method is higher than that of the XGBoost model, especially for the compressive strength. Taking the results of test dataset of the compressive strength for example, the MSE of the proposed method and the XGBoost model in the reference are 30 MPa and 46 MPa, respectively, and the R^2 of the proposed method and the XGBoost model in the reference are 0.95 and 0.92, respectively.

4. Multi-objective optimization

With the predictive models in Section 3, this section presents the multi-objective optimization of SHCC based on the five objective functions and six design scenarios in Section 2.

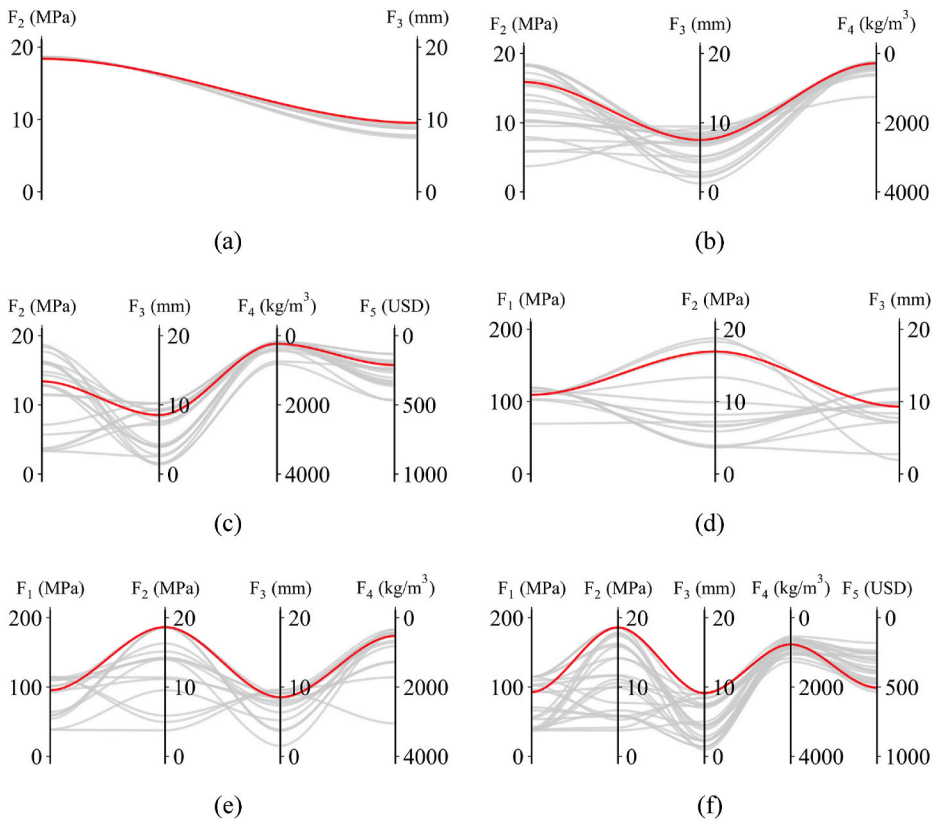

Fig. 9. Parallel coordinates plot of objective designs for the Pareto optimal solutions obtained by UNSGA-III corresponding to (a) DS1, (b) DS2, (c) DS3, (d) DS4, (e) DS5, and (f) DS6. The red lines represent the final solution selected by TOPSIS.

Table 8
Final solutions for mix design of SHCC.

Design variable	Design scenarios					
	DS_1	DS_2	DS_3	DS_4	DS_5	DS_6
Cement-to-binder ratio	0.69	0.20	0.13	0.16	0.15	0.44
Fly ash-to-binder ratio	0.07	0.27	0.62	0.63	0.62	0.15
Slag-to-binder ratio	0.04	0.24	0.07	0	0	0.12
Rice husk-to-binder ratio	0.02	0	0	0.10	0.03	0
Limestone-to-binder ratio	0.08	0.07	0.09	0.06	0.07	0.07
Metakaolin-to-binder ratio	0.02	0.06	0.02	0.02	0.08	0.08
Silica fume-to-binder ratio	0.07	0.14	0.06	0.02	0.03	0.04
Sand-to-binder ratio	0.68	1.37	1.39	0.97	1.39	1.4
Water-to-binder ratio	0.18	0.42	0.44	0.13	0.13	0.21
Superplasticizer content	2.21	2.02	1.79	1.98	2.58	2.61
Fiber type	PE	PP	PP	PE	PE	PE
Fiber volume (%)	2.53	2.05	2.63	2.08	1.96	1.82
Fiber length (mm)	17	12	11	15	17	18
Fiber diameter (μm)	21	13	18	25	24	24
Material properties						
Compressive strength (MPa)	92.27	57.45	58.34	109.24	95.36	92.75
Tensile strength (MPa)	18.68	15.31	13.37	16.92	18.64	18.19
Ductility (%)	9.53	7.51	8.52	9.16	8.73	9.12
Unit material cost ($/m^3)	1588.78	286.25	236.28	1216.87	527.87	673.83
Unit carbon footprint (kg/m^3)	733.67	289.19	212.86	534.05	580.44	504.26

4.1. Optimization process

Multi-objective optimization is performed using UNSGA-III and NSGA-III, respectively. A hypervolume indicator is used to evaluate the solutions. The hypervolume indicator is defined as the volume of optimal solutions in the objective space with respect to a reference point (Zitzler and Thiele, 1998), as shown in Fig. 7. A higher hypervolume indicator means a better solution. The reference point of each multi-objective is set by the minimum values of compressive strength, tensile strength, or ductility, and the maximum values of carbon footprint and material cost of SHCC mixtures.

Fig. 8 plots the hypervolume results corresponding to the design scenarios. The hypervolume indicators increase with the number of iterations, indicating improvement of the Pareto solutions. The hypervolume gradually increases to a certain value, showing a stabilization trend. The results indicate that the number of iterations is adequate for the optimization algorithms to converge. UNSGA-III achieves higher hypervolume indicators than NSGA-III, meaning that UNSGA-III has

better performance in multi-objective optimization of SHCC and thus is recommended.

4.2. Optimal solutions

With the multi-objective optimization solutions, TOPSIS is performed to select the optimal solutions for different design scenarios. Fig. 9 shows the Pareto optimal solutions obtained by UNSGA-III. The final solutions selected by TOPSIS are shown in red color, and the other solutions are shown in grey color. Different design objectives and their ranges are considered in different design scenarios, consistent with Tables 2 and 3.

A desired solution has large values for F_1 to F_3 and small values for F_4 and F_5, so the sequence of values of F_4 and F_5 is opposite to those of F_1 to F_3. The selected optimal solutions in different design scenarios represent the most plausible optimal solutions of mix design for SHCC. For example, the optimal solution for DS_6 has desired compressive strength, tensile strength, ductility, carbon footprint, and cost, simultaneously. Specifically, the compressive strength, tensile strength, and ductility of the optimal mixture for DS_6 are 93 MPa, 18 MPa, and 9 mm, respectively, and its unit carbon footprint and cost are 504 kg and $673 per cubic meter, respectively. It is noted that the optimal mix designs are different for different design scenarios. For example, the optimal solution for DS1, which does consider the carbon footprint and material cost, achieves a compressive strength, tensile strength, and ductility of 92 MPa, 19 MPa, and 10 mm, respectively.

Table 8 summarizes the final solutions of mix design and the corresponding material properties of SHCC for different design scenarios. The results show that the selected optimal solutions of mix design of SHCC are consistent with the defined performance objectives and design scenarios. The results reveal the complexity of optimal design of SHCC considering a wide range of mix design variables which interact with each other and have complicated coupling effects. The complexity of mix design of SHCC reflects the significance of data-driven methods for the high design efficiency.

5. Conclusions

This study proposes a new method for prediction of properties and multi-objective optimization of SHCC by integrating machine learning and evolutionary optimization algorithms. A tree-based pipeline optimization tool is adopted to automatically configure machine learning models for high accuracy in prediction of compressive strength, tensile strength, and ductility. Based on the predictive models, five objective functions and six design scenarios are considered.

Based on the above investigations, the following conclusions are drawn:

- The developed machine learning models are capable of predicting the mechanical properties of SHCC. The coefficient of determination of the predictive models for the compressive strength, tensile strength, and ductility are 0.95, 0.97, and 0.93, respectively, as evaluated using the test datasets.
- In the multi-objective optimization of SHCC, UNSGA-III demonstrates better performance than NSGA-III in terms of hypervolume indicator. In the investigated design scenarios, UNSGA-III achieves higher values of hypervolume indicator than NSGA-III. The Pareto optimal solutions are evaluated using the TOPSIS method for selection of the final optimal solutions, which are dependent on the defined performance objectives and design scenarios.
- The proposed method for prediction and multi-objective optimization provides an alternative solution for efficient development of SHCC because the method is capable of generating plausible mix design solutions for different design scenarios with particle preference of the material properties. The plausible solutions can be validated by a reduced number of tests.

Based on the developed method, it is promising to consider the other design variables such as the particle size of materials and investigate other important properties of SHCC such as flowability and durability. Extensive tests should be conducted to evaluate the performance of the proposed method. It is also interesting to test the applicability of the method for other families of high-performance fiber-reinforced cementitious composites such as ultra-high-performance concrete.

Credit author contribution statement

Soroush Mahjoubi: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft. Rojyar Barhemat: Data curation, Software, Writing – review & editing. Pengwei Guo: Data curation, Validation, Writing – review & editing. Weina Meng: Conceptualization, Funding acquisition, 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.

Acknowledgement

This research was funded by National Science Foundation [grant number CMMI-2046407], Pipeline and Hazardous Materials Safety Administration of the United States Department of Transportation [grant number 693JK31950008CAAP], and Stevens Institute of Technology.

References

Ahangari, K., Beygi, M., Rezaei, Y., 2013. Applicability of fiber reinforced self-compacting concrete for tunnel lining. Arabian Journal of Geosciences 6 (10), 3841–3846.
Aydın, S., Baradan, B., 2013. The effect of fiber properties on high performance alkali-activated slag/silica fume mortars. Compos. B Eng. 45 (1), 63–69.
Bao, Y., Xu, M., Soltan, D., Xia, T., Shih, A., Clack, H.L., Li, V.C., 2019. Three-dimensional printing multifunctional engineered cementitious composites (ECC) for structural elements. In: Wangler, T., Flatt, R.J. (Eds.), First RILEM International Conference on Concrete and Digital Fabrication – Digital Concrete 2018. Springer International Publishing, Cham, pp. 115–128.
Chatterjee, S., Hadi, A.S., 2015. Regression Analysis by Example. John Wiley & Sons.
Chen, M., Zhong, H., Chen, L., Zhang, Y., Zhang, M., 2021. Engineering properties and sustainability assessment of recycled fibre reinforced rubberised cementitious composite. J. Clean. Prod. 278, 123996.
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. Construct. Build. Mater. 67, 386–392.
Comon, P., 1994. Independent component analysis, a new concept? Signal Process. 36 (3), 287–314.
Cook, R., Lapeyre, J., Ma, H., Kumar, A., 2019. Prediction of compressive strength of concrete: critical comparison of performance of a hybrid machine learning model with standalone models. J. Mater. Civ. Eng. 31 (11), 04019255.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.
Deb, K., Jain, H., 2014. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. IEEE Trans. Evol. Comput. 18 (4), 577–601.
Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6 (2), 182–197.
Ding, Y., Yu, J., Yu, K., Xu, S., 2018a. Basic mechanical properties of ultra-high ductility cementitious composites: from 40 MPa to 120 MPa. Compos. Struct. 185, 634–645.
Ding, Y., Yu, K., Yu, J., Xu, S., 2018b. Structural behaviors of ultra-high performance engineered cementitious composites (UHP-ECC) beams subjected to bending-experimental study. Construct. Build. Mater. 177, 102–115.
Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. Proceedings of the Thirteenth International Conference on International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., Bari, Italy, pp. 148–156.

Guo, P., Meng, W., Xu, M., Li, V.C., Bao, Y., 2021a. Predicting mechanical properties of high-performance fiber-reinforced cementitious composites by integrating micromechanics and machine learning. Materials 14 (12), 3143.
Guo, P., Meng, W., Bao, Y., 2021b. Automatic identification and quantification of dense microcracks in high-performance fiber-reinforced cementitious composites through deep learning-based computer vision. Cement Concr. Res. 148, 106532.
Gursel, A.P., Maryman, H., Ostertag, C., 2016. A life-cycle approach to environmental, mechanical, and durability properties of "green" concrete mixes with rice husk ash. J. Clean. Prod. 112, 823–836.
Habert, G., Denarié, E., Šajna, A., Rossi, P., 2013. Lowering the global warming impact of bridge rehabilitations by using ultra high performance fibre reinforced concretes. Cement Concr. Compos. 38, 1–11.
Han, T., Siddique, A., Khayat, K., Huang, J., Kumar, A., 2020. An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete. Construct. Build. Mater. 244, 118271.
Heath, A., Paine, K., McManus, M., 2014. Minimising the global warming potential of clay based geopolymers. J. Clean. Prod. 78, 75–83.
Heirman, G., Hendrickx, R., Vandewalle, L., Van Gemert, D., Feys, D., De Schutter, G., Desmet, B., Vantomme, J., 2009. Integration approach of the Couette inverse problem of powder type self-compacting concrete in a wide-gap concentric cylinder rheometer: Part II. Influence of mineral additions and chemical admixtures on the shear thickening flow behaviour. Cement Concr. Res. 39 (3), 171–181.
Hossain, Z., Elsayed, A., 2018. Use of Rice Hull Ash (RHA) as a Sustainable Source of Construction Material. Transportation Consortium of South-Central States.
Huang, J.S., Liew, J.X., Liew, K.M., 2021. Data-driven machine learning approach for exploring and assessing mechanical properties of carbon nanotube-reinforced cement composites. Compos. Struct. 267, 113917.
Huang, T., Zhang, Y.X., 2016. Numerical modelling of mechanical behaviour of engineered cementitious composites under axial tension. Comput. Struct. 173, 95–108.
Hwang, C.L., Yoon, K., 1981. Methods for multiple attribute decision making. In: Hwang, C.-L., Yoon, K. (Eds.), Multiple Attribute Decision Making: Methods and Applications A State-Of-The-Art Survey. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 58–191.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer.
Jun, P., Mechtcherine, V., 2010. Behaviour of Strain-hardening Cement-based Composites (SHCC) under monotonic and cyclic tensile loading: Part 1 – experimental investigations. Cement Concr. Compos. 32 (10), 801–809.
Kanda, T., Li, V.C., 1999. New micromechanics design theory for pseudostrain hardening cementitious composite. J. Eng. Mech. 125 (4), 373–381.
Kang, J., Bolander, J.E., 2016. Multiscale modeling of strain-hardening cementitious composites. Mech. Res. Commun. 78, 47–54.
Ke, X., Duan, Y., 2021. A Bayesian machine learning approach for inverse prediction of high-performance concrete ingredients with targeted performance. Construct. Build. Mater. 270, 121424.
Kim, J.K., Kim, J.S., Ha, G.J., Kim, Y.Y., 2007. Tensile and fiber dispersion performance of ECC (engineered cementitious composites) produced with ground granulated blast furnace slag. Cement Concr. Res. 37 (7), 1096–1105.
Lepech, M.D., Li, V.C., Robertson, R.E., Keoleian, G.A., 2008. Design of green engineered cementitious composites for improved sustainability. ACI Mater. J. 105 (6), 567.
Li, V.C., 2003. On engineered cementitious composites (ECC) a review of the material and its applications. J. Adv. Concr. Technol. 1 (3), 215–230.
Li, V.C., 2019. High-performance and multifunctional cement-based composite material. Engineering 5 (2), 250–260.
Li, V.C., Herbert, E., 2012. Robust self-healing concrete for sustainable infrastructure. J. Adv. Concr. Technol. 10 (6), 207–218.
Li, V.C., Wu, C., Wang, S., Ogawa, A., Saito, T., 2002. Interface tailoring for strain-hardening polyvinyl alcohol-engineered cementitious composite (PVA-ECC). ACI Mater. J. 99 (5), 463–472.
Li, X., Bao, Y., Wu, L., Yan, Q., Ma, H., Chen, G., Zhang, H., 2017a. Thermal and mechanical properties of high-performance fiber-reinforced cementitious composites after exposure to high temperatures. Construct. Build. Mater. 157, 829–838.
Li, X., Bao, Y., Xue, N., Chen, G., 2017b. Bond strength of steel bars embedded in high-performance fiber-reinforced cementitious composite before and after exposure to elevated temperatures. Fire Saf. J. 92, 98–106.
Li, X., Wang, J., Bao, Y., Chen, G., 2017c. Cyclic behavior of damaged reinforced concrete columns repaired with high-performance fiber-reinforced cementitious composite. Eng. Struct. 136, 26–35.
Li, X., Xu, Z., Bao, Y., Cong, Z., 2019. Post-fire seismic behavior of two-bay two-story frames with high-performance fiber-reinforced cementitious composite joints. Eng. Struct. 183, 150–159.
Lin, J.-X., Song, Y., Xie, Z.-H., Guo, Y.-C., Yuan, B., Zeng, J.-J., Wei, X., 2020. Static and dynamic mechanical behavior of engineered cementitious composites with PP and PVA fibers. Journal of Building Engineering 29, 101097.
Liu, F.T., Ting, K.M., Zhou, Z.-H., 2008. Isolation forest. In: Eighth IEEE International Conference on Data Mining. IEEE, pp. 413–422.
Liu, H., Zhang, Q., Li, V., Su, H., Gu, C.J.C., Materials, B., 2017. Durability study on engineered cementitious composites (ECC) under sulfate and chloride environment. Construct. Build. Mater. 133, 171–181.
Liu, Y., Zhang, Q., Bao, Y., Bu, Y., 2019. Static and fatigue push-out tests of short headed shear studs embedded in Engineered Cementitious Composites (ECC). Eng. Struct. 182, 29–38.
Long, G., Gao, Y., Xie, Y., 2015. Designing more sustainable and greener self-compacting concrete. Construct. Build. Mater. 84, 301–306.
Lu, H., Iseley, T., Matthews, J., Liao, W., Azimi, M., 2021a. An ensemble model based on relevance vector machine and multi-objective SALP swarm algorithm for predicting burst pressure of corroded pipelines. J. Petrol. Sci. Eng. 203, 108585.
Lu, H., Xu, Z.D., Iseley, T., Matthews, J.C., 2021b. Novel data-driven framework for predicting residual strength of corroded pipelines. J. Pipeline Syst. Eng. Pract. 12 (4), 04021045.
Lu, H., Behbahani, S., Ma, X., Iseley, T., 2021c. A multi-objective optimizer-based model for predicting composite material properties. Construct. Build. Mater. 284, 122746.
Marani, A., Jamali, A., Nehdi, M.L.J.M., 2020. Predicting ultra-high-performance concrete compressive strength using tabular generative. Adversarial Networks 13 (21), 4757.
Meng, W., Khayat, K.H., 2016. Mechanical properties of ultra-high-performance concrete enhanced with graphite nanoplatelets and carbon nanofibers. Compos. B Eng. 107, 113–122.
Moré, J.J., 1978. The Levenberg-Marquardt Algorithm: Implementation and Theory, Numerical Analysis. Springer, pp. 105–116.
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. Construct. Build. Mater. 67, 321–337.
Nepomuceno, M., Oliveira, L., Lopes, S.M.R., 2012. Methodology for mix design of the mortar phase of self-compacting concrete using different mineral additions in binary blends of powders. Construct. Build. Mater. 26 (1), 317–326.
Nguyễn, H.H., Lương, Q.-H., Choi, J.-I., Ranade, R., Li, V.C., Lee, B.Y., 2021. Ultra-ductile behavior of fly ash-based engineered geopolymer composites with a tensile strain capacity up to 13.7%. Cement Concr. Compos. 122, 104133.
Olson, R.S., Moore, J.H., 2019. TPOT: a tree-based pipeline optimization tool for automating machine learning. In: Hutter, F., Kotthoff, L., Vanschoren, J. (Eds.), Automated Machine Learning: Methods, Systems, Challenges. Springer International Publishing, pp. 151–160.
Pakravan, H.R., Ozbakkaloglu, T., 2019. Synthetic fibers for cementitious composites: a critical and in-depth review of recent advances. Construct. Build. Mater. 207, 491–518.
Pappa, G.L., 2020. Fitness Landscape Analysis of Automated Machine Learning Search Spaces, Evolutionary Computation in Combinatorial Optimization: 20th European Conference. Springer Nature, p. 114.
Pavlyshenko, B., 2018. Using Stacking Approaches for Machine Learning Models. IEEE Second International Conference on Data Stream Mining & Processing, pp. 255–258.
Poon, C.S., Lam, L., Kou, S.C., Wong, Y.L., Wong, R., 2001. Rate of pozzolanic reaction of metakaolin in high-performance cement pastes. Cement Concr. Res. 31 (9), 1301–1306.
Purnell, P., 2013. The carbon footprint of reinforced concrete. Adv. Cement Res. 25, 362–368.
Ranade, R., 2014. Advanced Cementitious Composite Development for Resilient and Sustainable Infrastructure. Ph. D. Dissertation. University of Michigan, Ann Arbor, Michigan, U.S.
Said, S.H., Razak, H.A., 2015. The effect of synthetic polyethylene fiber on the strain hardening behavior of engineered cementitious composite (ECC). Mater. Des. 86, 447–457.
Sathawane, S.H., Vairagade, V.S., Kene, K.S., 2013. Combine effect of rice husk ash and fly ash on concrete by 30% cement replacement. Procedia Engineering 51, 35–44.
Seada, H., Deb, K., 2014. U-NSGA-III: A Unified Evolutionary Algorithm for Single, Multiple, and Many-Objective Optimization. COIN Report 2014022.
Sikora, R., 2015. A modified stacking ensemble machine learning algorithm using genetic algorithms, Handbook of research on organizational transformations through big data analytics. IGi Global 43–53.
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. Journal of Building Engineering 43, 102544.
Swamynathan, M., 2019. Step 3: fundamentals of machine learning. In: Swamynathan, M. (Ed.), Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python. Apress, Berkeley, CA, pp. 145–262.
Tosun-Felekoğlu, K., Gödek, E., Keskinateş, M., Felekoğlu, B., 2017. Utilization and selection of proper fly ash in cost effective green HTPP-ECC design. J. Clean. Prod. 149, 557–568.
Turk, K., Nehdi, M.L., 2018. Coupled effects of limestone powder and high-volume fly ash on mechanical properties of ECC. Construct. Build. Mater. 164, 185–192.
Wang, Y., Liu, F., Yu, J., Dong, F., Ye, J., 2020. Effect of polyethylene fiber content on physical and mechanical properties of engineered cementitious composites. Construct. Build. Mater. 251, 118917.
Willis, M.J., Hiden, H.G., Marenbach, P., McKay, B., Montague, G.A., 1997. Genetic Programming: an Introduction and Survey of Applications, Second International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications. IET, pp. 314–319.
Wold, S., Esbensen, K., Geladi, P.J.C., systems, i.l., 1987. Principal component analysis, 2 (1–3), 37–52.
Xu, M., Bao, Y., Wu, K., Shi, H., Guo, X., Li, V.C., 2019a. Multiscale investigation of tensile properties of a TiO2-doped engineered cementitious composite. Construct. Build. Mater. 209, 485–491.
Xu, M., Bao, Y., Wu, K., Xia, T., Clack, H., Shi, H., Li, V.C., 2019b. Influence of TiO2 incorporation methods on NOx abatement in engineered cementitious composites. Construct. Build. Mater. 221, 375–383.
Xu, M., Clack, H., Xia, T., Bao, Y., Wu, K., Shi, H., Li, V.C., 2020. Effect of TiO2 and fly ash on photocatalytic NOx abatement of Engineered Cementitious Composites. Construct. Build. Mater. 236, 117559.

Xu, M., Yu, J., Zhou, J., Bao, Y., Li, V.C., 2021. Effect of curing relative humidity on mechanical properties of engineered cementitious composites at multiple scales. Construct. Build. Mater. 284, 122834.
Yew, M.K., Bin Mahmud, H., Ang, B.C., Yew, M.C., 2015. Effects of low volume fraction of polyvinyl alcohol fibers on the mechanical properties of oil palm shell lightweight concrete. Advances in Materials Science and Engineering 2015, 425236.
Yu, K., Wang, Y., Yu, J., Xu, S., 2017. A strain-hardening cementitious composites with the tensile capacity up to 8. Construct. Build. Mater. 137, 410–419.
Yu, K., Dai, J., Lu, Z., Poon, C.S., 2018. Rate-dependent tensile properties of ultra-high performance engineered cementitious composites (UHP-ECC). Cement Concr. Compos. 93, 218–234.
Yu, K., Zhu, W., Ding, Y., Lu, Z., Yu, J., Xiao, J., 2019. Micro-structural and mechanical properties of ultra-high performance engineered cementitious composites (UHP-ECC) incorporation of recycled fine powder (RFP). Cement Concr. Res. 124, 105813.
Yu, K., Ding, Y., Liu, J., Bai, Y., 2020. Energy dissipation characteristics of all-grade polyethylene fiber-reinforced engineered cementitious composites (PE-ECC). Cement Concr. Compos. 106, 103459.
Zhang, J., Wang, Q., Wang, Z., 2016. Optimizing design of high strength cement matrix with supplementary cementitious materials. Construct. Build. Mater. 120, 123–136.
Zhang, J., Huang, Y., Wang, Y., Ma, G., 2020. Multi-objective optimization of concrete mixture proportions using machine learning and metaheuristic algorithms. Construct. Build. Mater. 253, 119208.
Zheng, Y., Zhang, L., Xia, L.J.C., 2018. Investigation of the behaviour of flexible and ductile ECC link slab reinforced with FRP. Construct. Build. Mater. 166, 694–711.
Zhou, J., Qian, S., Beltran, M.G.S., Ye, G., van Breugel, K., Li, V.C., 2010. Development of engineered cementitious composites with limestone powder and blast furnace slag. Mater. Struct. 43 (6), 803–814.
Zhu, Y., Yang, Y., Yao, Y., 2012. Use of slag to improve mechanical properties of engineered cementitious composites (ECCs) with high volumes of fly ash. Construct. Build. Mater. 36, 1076–1081.
Zimek, A., Schubert, E., 2017. Outlier detection. In: Liu, L., Özsu, M.T. (Eds.), Encyclopedia of Database Systems. Springer New York, New York, pp. 1–5.
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.), Parallel Problem Solving from Nature. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 292–301.