Cradle-to-gate assessment and optimization of sustainable geopolymer concrete

Zhang, Mingyang; Guo, Pengwei; Tan, Xiao; Du, Jiang; Meng, Weina; Bao, Yi

DOI
10.1016/j.jclepro.2025.147387

Publication date
2025

Document Version
Final published version

Published in
Journal of Cleaner Production

Citation (APA)
Zhang, M., Guo, P., Tan, X., Du, J., Meng, W., & Bao, Y. (2025). Cradle-to-gate assessment and optimization of sustainable geopolymer concrete. *Journal of Cleaner Production*, *538*, Article 147387. https://doi.org/10.1016/j.jclepro.2025.147387

Journal of Cleaner Production 538 (2026) 147387

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Cradle-to-gate assessment and optimization of sustainable geopolymer concrete

Mingyang Zhang [a], Pengwei Guo [a,b,*], Xiao Tan [c], Jiang Du [d], Weina Meng [a], Yi Bao [a]

[a] Department of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ, 07030, United States
[b] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, 2628, CN, Netherlands
[c] College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, Jiangsu, 210024, China
[d] School of Civil Engineering, Chongqing Jiaotong University, Chongqing, 400074, China

ARTICLE INFO

Keywords:
Alkaline activators
Cost-benefit analysis
Geopolymer concrete
Life cycle performance
Solid waste
Sustainability

ABSTRACT

Geopolymer concrete (GPC) is a sustainable alternative to Portland cement concrete by eliminating Portland cement. However, using alkaline activators compromises the sustainability of GPC. This paper presents a comprehensive assessment of the cradle-to-gate life cycle cost, carbon footprint, and energy consumption of 2304 GPC mixtures which represent the state-of-the-art dataset, aiming to establish a holistic understanding of the impacts of GPC design parameters on mechanical and sustainability performance. The cost-benefit characteristics of solid wastes are considered, and strength-normalized sustainability parameters are discussed. Results reveal that the formulation of GPC plays critical roles in mechanical and sustainability metrics. Inappropriate use of alkaline activators and solid wastes can largely compromise both mechanical strength and sustainability metrics, lower than their Portland cement counterparts. Based on the large dataset, this paper identifies the appropriate upper and lower bounds for various ingredients to guide the design of GPC for balanced mechanical strength and sustainability metrics.

1. Introduction

Concrete is the mostly consumed construction material, with an annual consumption more than 30 billion tons in 2022 (UnsustaInable: concrete and cement, 2022). Despite its relatively low unit cost and carbon footprint compared to other structural materials like steel, the large volume of concrete makes it a major contributor to carbon emissions (Kilgore, 2023). Among various ingredients of concrete, cement is a major contributor of emissions (Amran et al., 2022). Cement production alone accounting for approximately 8 % of global emissions (Akan et al., 2017). As governments and industries strive toward carbon neutrality by 2050 (Guterres, 2020), the development of cleaner alternatives have become critical.

Geopolymer concrete (GPC) has emerged as a sustainable material that uses alkaline activated binders to replace cement-based binders. Previous research showed that various types of industrial by-products, such as fly ash, slag, and metakaolin (Yang et al., 2024a), could be used to produce alkaline activated binders, eliminating the material cost, carbon footprint, and embodied energy associated with cement. The binders are produced in a polymerization process by activating aluminosilicate precursors with alkaline activators (e.g., NaOH, Na_2SiO_3) (Li et al., 2024). The polymerization process creates a three-dimensional network of Si–O–Al–O bonds, forming geopolymer gels (N-A-S-H or C-A-S-H) (Guo et al., 2021a), which act as binders that hold aggregates (Liu et al., 2024). Previous research showed that GPC reduced carbon footprint by up to 60 % (Zannerni et al., 2020), in comparison with counterpart Portland cement concrete. In addition to lower upfront carbon footprint, GPC also demonstrated high resistance to chemical attack (e.g., acid attack), enhancing long-term durability and lowering maintenance cost and emission of infrastructure in harsh environments (Esparham and Ghalatian, 2022).

Although eliminating cement helps reduce material cost and carbon footprint, some scholars have argued that using alkaline activators offsets sustainability metrics by introducing additional emissions (Liu et al., 2020a), because NaOH and Na_2SiO_3 were respectively responsible for 7.2 % and 39.5 % of the total emissions. The high emissions stem from the energy-intensive manufacturing processes, such as high-temperature electrolysis and chemical synthesis, which not only

* Corresponding author. Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, 2628, CN, the Netherlands.
E-mail address: P.Guo-1@tudelft.nl (P. Guo).

https://doi.org/10.1016/j.jclepro.2025.147387
Received 28 June 2025; Received in revised form 12 November 2025; Accepted 21 December 2025
Available online 24 December 2025

generate substantial emissions but also consume energy. Two knowledge gaps have been identified: (1) Given the competing effects of removing cement and introducing alkaline activators, it remains unclear whether GPC is more sustainable than Portland cement concrete counterparts. (2) The mechanical properties and sustainability of GPC are dependent on many design variables, such as the dosages of alkaline activators, aluminosilicate precursors, and aggregates. However, it is unclear how GPC can be designed to achieve optimal sustainability. These knowledge gaps have hindered wider applications of GPC in the construction and rehabilitation of infrastructure.

To fill these knowledge gaps, this research addresses three research questions: (1) How do the design variables of GPC impact sustainability metrics, such as material cost, carbon footprint, and embodied energy? (2) What are the appropriate ranges of design variables that can lead to GPC with high mechanical performance and sustainability? (3) Does GPC have better sustainability than other types of concrete when their mechanical performance is controlled the same? Motivated by these questions, the objectives of this research are to: (1) evaluate the effects of design variables on the lifecycle performance of GPC; (2) identify practical recommendations for designing GPC through optimizing design variables; and (3) compare GPC with competing materials such as ultra-high-performance concrete (UHPC) and conventional concrete in terms of lifecycle performance.

This research involves two major challenges: (1) Holistic assessment of lifecycle performance of GPC requires considering mechanical performance like compressive strength which play critical roles in infrastructure applications, as discussed in prior research (Guo et al., 2023a). Using mechanical strength-normalized metrics provides meaningful evaluations, but assessing the mechanical performance of GPC requires costly experiments. (2) Comprehensive evaluation of GPC mixtures with various materials and design variables requires extensive experiments, which involve not only high cost but also substantial time.

This study addresses the existing knowledge gaps in GPC research by establishing an integrated framework that links mechanical performance with sustainability indicators. Previous studies mainly focused on experimental characterization or partial life cycle analyses, lacking a systematic approach to simultaneously evaluate strength, cost, carbon footprint, and energy consumption. Moreover, the influence of design variables such as activator composition, precursor type, and curing condition on both mechanical and environmental performance has not been fully quantified. To bridge these gaps, this research develops a comprehensive data-driven and cradle-to-gate LCA framework that enables large-scale evaluation of GPC mixtures without extensive experimental testing. By incorporating generative AI for data augmentation, machine learning for performance prediction, and strength-normalized sustainability metrics, the study provides new insights into the trade-offs between mechanical efficiency and environmental impact. This framework not only fills the methodological gap between material design and sustainability evaluation but also establishes practical design boundaries for optimizing GPC mixtures.

The remainder of the paper is organized as follows: Section 2 presents the methodology. Section 3 presents and discusses the results. Section 4 summarizes the findings and recommends future research opportunities.

2. Methodology

The methodological framework (Fig. 1) integrates two interconnected components, which are data-driven modeling and cradle-to-gate LCA. The data-driven modeling component involves two phases, which are dataset construction and model development. Data construction is composed of data collection (from literature) and data synthesis using generative AI techniques. Generative AI techniques are adopted to generate supplemental data, which are added to the data collected from literature. The combined dataset is used to develop a data-driven model, which provides essential data used to perform cradle-to-gate LCA and evaluate the effects of parameters on the mechanical and sustainability metrics of GPC. The results are utilized to analyze the ranges of design variables, aiming to derive recommendations that are useful to guide practitioners in designing sustainable GPC in future practice.

2.1. Dataset collection and analysis

A comprehensive dataset was collected from diverse data sources, including technical papers and reports (Liu et al., 2020a, 2020b; Ganesh and Muthukannan, 2021; Bellum, 2022; Tayeh et al., 2022; Moradikhou et al., 2020; Zaid et al., 2022; Guo et al., 2024a; Eltantawi et al., 2025; Peng and Unluer, 2022; Abdellatief et al., 2024; Aisheh et al., 2022; Oyebisi and Alomayri, 2023; Albidah et al., 2021, 2022; Júnior et al., 2021; Eisa et al., 2022; Raza et al., 2024; Kathirvel and Sreekumaran, 2021; Mayhoub et al., 2021; Zheng et al., 2023; Yang et al., 2024b; Zhang et al., 2024; Ali et al., 2024; Karthik et al., 2024; Samadi et al., 2024; Kumar et al., 2020; Ganesan et al., 2015; Deng et al., 2023; Reddy et al., 2022), available in popular databases, such as ScienceDirect, Google Scholar, and Web of Science. Relevant references were searched

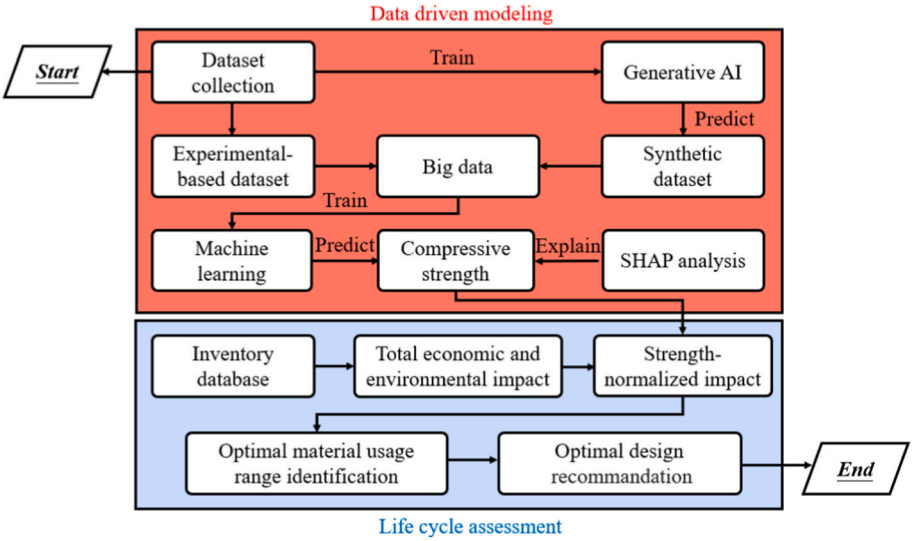

Fig. 1. Methodological framework.

using keywords “geopolymer concrete”, “mixture design”, “compressive strength”, “cost”, “carbon footprint”, and “embodied energy”. A total of 2304 GPC mixtures were identified, which include 16 design variables (features). These variables were categorized into two groups: (1) 14 variables related to materials, including the contents of fly ash (C_{FA1}), ground granulated blast furnace slag (C_{GGBS}), silica fume (C_{SF}), metakaolin (C_{MK}), fine aggregate (C_{FA2}), coarse aggregate (C_{CA}), Molarity (Mol), sodium hydroxide (C_{SH}), sodium silicate (C_{SS}), and additional water (C_{AW}), fiber volume ratio (FV), fiber diameter (FD), fiber length (FL), and fiber’s elastic modulus (FM); and (2) 2 parameters related to curing condition, including curing time (CT) and curing temperature (T).

The statistical distribution (Fig. 2) of the 16 features and compressive strength of GPC have demonstrated severe imbalance. The contents of different materials exhibit wide ranges, and the frequency results show uneven distributions. The statistics of the minimum, maximum, average values, and standard deviation of each feature and compressive strength are listed in Table A1, which corroborates the diversity and imbalance of data instances.

A correlation analysis was conducted, and the results of Pearson correlation coefficient are shown in Fig. 3. The 16 features are numbered from 1 to 16, and compressive strength is numbered 17. The 16 features exhibit low correlation (below 0.7), indicating that multicollinearity issues do not occur (Mahjoubi et al., 2021). Data normalization was performed to transform variables with different scales into the range of 0–1 using MinMaxScaler (Mahjoubi et al., 2022), aiming to improve accuracy (Guo et al., 2021b) and consistency.

2.2. Data augmentation

Generative AI provides an efficient approach for data augmentation (Mahjoubi et al., 2023a). It typically involves two components: a generator and a discriminator (Guo et al., 2024b). The generator transforms random noise into synthetic tabular data that resembles real samples, capturing both continuous and categorical features. The discriminator receives either real or generated data and determines whether it is genuine. This adversarial interaction allows the generator to gradually learn complex feature distributions and capture subtle relationships between variables. An illustration of this generative AI-based data augmentation process is shown in Fig. 4.

Three representative generative AI methods were adopted: (1) Conditional Tabular Generative Adversarial Network (CT-GAN) (Habibi et al., 2023), which extends GANs to tabular data by conditioning generation on discrete column values and using mode-specific normalization; (2) Conditional Tabular GAN with Copula-based modeling (CTAB-GAN), which enhances CT-GAN by using copula transformations to better handle mixed data types and rare categories, thereby improving the quality and diversity of generated samples (Zhao et al., 2021); and (3) Tabular Variational Autoencoder (TVAE), a variational autoencoder framework designed specifically for tabular data, which models data distribution through probabilistic latent variables and is particularly effective for generating continuous and categorical features with complex interdependencies (Tan et al., 2024).

To ensure consistency, the number of synthetic samples was kept equal to that of the original training dataset. Subsequently, the original dataset collected from literature and the synthetic dataset generated by

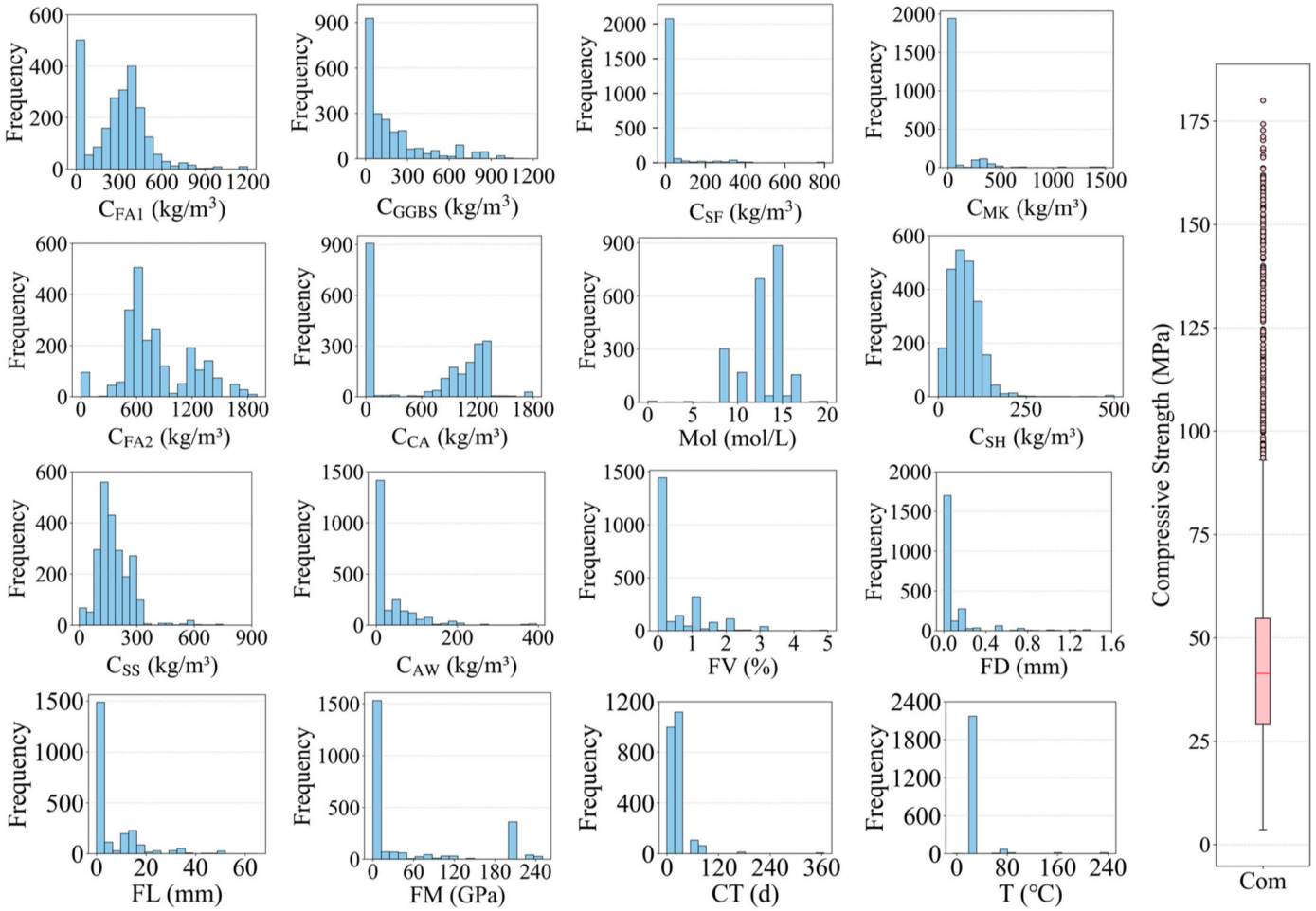

Fig. 2. Statistical distribution of input features and output variable.

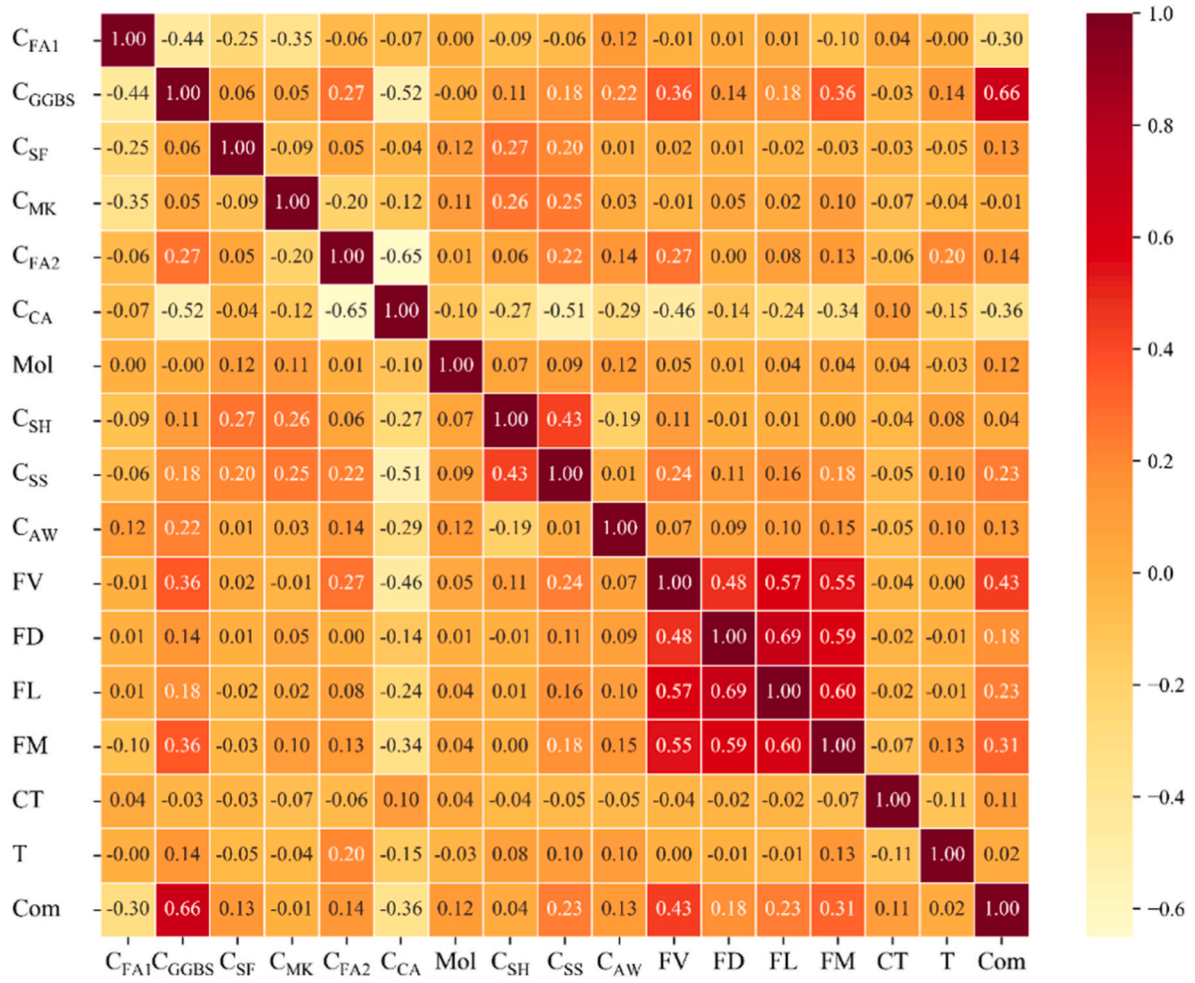

Fig. 3. Heatmap of Pearson correlation coefficient.

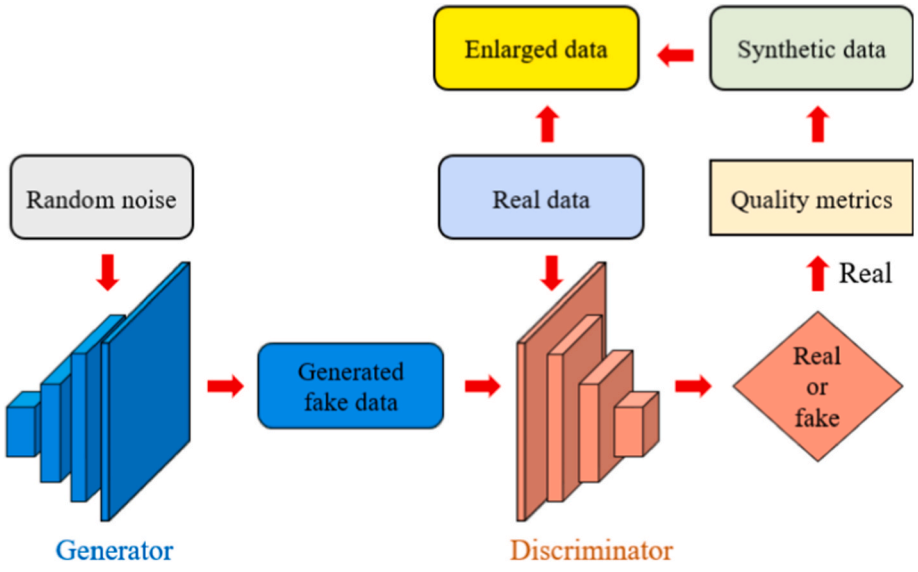

Fig. 4. Architecture of generative AI for data augmentation.

the generative AI model are merged to create a comprehensive dataset. The dataset is then randomly split, with 80 % allocated for training the machine learning models and the remaining 20 % reserved for testing.

2.3. Machine learning algorithms and performance metrics

Six previously developed machine learning algorithms, which are Linear Regression (LR) (Mahjoubi et al., 2022), Multi-layer Perceptron

(MLP) (Prasad et al., 2009), Decision Tree (DT) (Zhou et al., 2024), Random Forest (RF) (Han et al., 2019), Extreme Gradient Boosting (XGBoost) (Guo et al., 2023b), and Light Gradient Boosting Machine (LightGBM) (Guo et al., 2024c), were adopted. Their hyperparameters were tuned through a combination of grid search (Mahjoubi et al., 2023a) and k-fold cross-validation (Xu et al., 2024) to reduce the risk of overfitting. In each iteration, one-fold was used for validation while the remaining (k-1) folds were used for training. This process ensures that all data contributes to training and validation.

The performance of machine learning models was evaluated using four metrics, which are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). These metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{i,pre} - y_{i,test}\right)^2} \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{i,pre} - y_{i,test}}{y_{i,test}}\right| \times 100\% \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{i,pre} - y_{i,test}\right| \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{i,test} - y_{i,pre}\right)^2}{\sum_{i=1}^{n}\left(y_{i,test} - \bar{y}\right)^2} \tag{4}$$

where $y_{i,test}$ and $y_{i,pre}$ denote the tested and predicted compressive strength values, respectively; $\bar{y}$ is the average compressive strength; and n represents the total number of samples.

The quality of data generated using generative AI techniques was assessed using three metrics: RMSE (Mahjoubi et al., 2023b), Correlation Preservation (CP) (Xu and Veeramachaneni, 2018), and Diversity Index (DI) (Zhang et al., 2022). A lower RMSE indicates better representation of real data. CP reflects the consistency of variable relationships between the synthetic and real datasets. Higher CP values represent stronger alignment in internal structure, indicating that the synthetic data preserves key patterns from the original. The formula for CP is shown in Eq. (5).

$$CP(C_S, C_R) = \frac{1}{n}\sum_{i=1}^{n}\left|R_{s,i} - R_{r,i}\right| \tag{5}$$

where $R_{s,i}$ is the rank of variable i in the synthetic dataset C_s; $R_{r,i}$ represents rank of variable i in the real dataset; $|R_{s,i}$ - $R_{r,i}|$ denotes the absolute difference in the ranking of variable $i$between the synthetic and real datasets; and n is the total number of variables being compared.

The DI quantifies the variability of synthetic data relative to real data by calculating the average Euclidean distance between each synthetic instance ($x_{s,i}$) and its nearest real data sample (X_R). A higher DI reflects greater diversity in the synthetic dataset, supporting better model generalization and lowering the risk of overfitting. This metric, alongside others, helps ensure that synthetic data is both representative and sufficiently varied for effective model training. The formulas of DI are shown in Eq. (6):

$$DI(X_S, X_R) = \frac{1}{n}\sum_{i=1}^{n} E\left(x_{s,i}, X_R\right) \tag{6}$$

where n denotes the total number of synthetic samples; $E\left(x_{s,i}, X_R\right)$ refers to the smallest Euclidean distance between a given synthetic point and the real dataset.

$$E\left(x_{S,i}, X_R\right) = \sqrt{\sum_{j=1}^{N}\left(s_{i,j} - r_j\right)^2} \tag{7}$$

where $s_{i,j}$ refers to the value of feature j in synthetic sample i; r_j represents the reference value for feature j in the real dataset; N is the total number of features.

2.4. SHAP-based model explanation

Machine learning algorithms have demonstrated high predictive performance in the field of materials engineering; however, their complexity and lack of interpretability often raise concerns regarding their practical applicability. These models are frequently regarded as "black boxes", offering limited insight into the reasoning behind their predictions. To address this issue, Shapley Additive Explanations (SHAP) (Mahjoubi et al., 2022) is widely adopted to enhance the interpretability of machine learning models. Based on cooperative game theory, SHAP provides both global and local interpretability by quantifying the contribution of each input feature to the output of the model. In this study, the SHAP framework is employed to identify and interpret the relative importance of input variables, thereby enhancing the transparency and trustworthiness of the machine learning models used for predicting material properties. The SHAP algorithm is shown as follows:

$$\hat{y} = \psi_0 + \sum_{i=1}^{n}\psi(x_i) \tag{8}$$

where $\hat{y}$ denotes the predicted output; ψ_0 represents the average output value from the training dataset; n indicates the total number of features; and $\psi(x_i)$ corresponds to the Shapley value assigned to each feature's contribution to the output prediction, as defined by:

$$\psi(x_i) = \sum_{s\in\{x_1,\ldots,x_n\}\setminus x_i}\frac{|s|!(n-|s|-1)!}{n!}\left\{f(s\cup\{x_i\}) - f(s)\right\} \tag{9}$$

where $\{x_1, \ldots, x_n\}\setminus x_i$ denotes the set of all input features excluding the i-th feature; $|s|$ is the number of features in the subset s; $f(s)$ denotes the model prediction based on the features in subset s; and $f(s\cup\{x_i\})$ is the predicted output when the i-th feature is added to subset s.

2.5. Life cycle assessment

This study uses a Cradle-to-Gate boundary, concentrating only on the production phase of concrete while excluding construction, usage, and end-of-life stages. This approach allows for a focused evaluation of the material's early economic and environmental impacts. The analysis centers on key sustainability indicators such as cost, carbon emissions, energy consumption. It also explores the relationship between environmental impact and mechanical performance by comparing compressive strength per unit of material. This helps identify mixture designs that achieve a balance between structural performance and sustainability.

2.5.1. Inventory database

Inventory data for each raw material is collected, covering unit cost (USD), carbon emissions (kg), and embodied energy (MJ) to produce 1 kg raw materials, as shown in Table A2. The raw materials considered include fly ash, slag, silica fume, metakaolin, fine aggregate, coarse aggregate, sodium hydroxide, sodium silicate, superplasticizer, water, and steel fiber. This inventory database is used to calculate the total cost, carbon emissions, and energy consumption per cubic meter of GPC. These values are then normalized by the 28-day compressive strength to obtain strength-normalized material cost, strength-normalized carbon

footprint, and strength-normalized energy consumption metrics. This approach enables a meaningful comparison of GPC mixes in terms of both mechanical performance and environmental/economic impact.

2.5.2. Economy and environmental impacts

The total cost, carbon emissions, and energy consumption for producing 1 m^3 of GPC under standard curing conditions are calculated using Eq. (10) to Eq. (12), respectively. To assess resource efficiency, strength-normalized indicators are calculated by combining the 28-day compressive strength with the corresponding environmental and economic impacts. These indicators offer a robust basis for comparing the cost and environmental performance of GPC formulations. The developed machine learning model is initially used to predict the compressive strength of GPC, and subsequently to support strength-normalized LCA indicators.

$$M = \sum_{i=1}^{n} m_i r_i \tag{10}$$

$$C = \sum_{i=1}^{n} c_i r_i \tag{11}$$

$$E = \sum_{i=1}^{n} e_i r_i \tag{12}$$

where M, C, E denote the cost, CO_2 emission, equivalent energy consumption of producing 1 m^3 GPC, respectively; m_i, c_i, e_i are the unit cost, CO_2 emission, energy consumption used to produce the ith ingredient (i = 1, 2, …, n); and r_i is the mass of the ith ingredient.

2.5.3. Optimal material usage identification

Optimal material usage can be identified by evaluating three performance indicators: strength-normalized cost, carbon footprint, and energy consumption. For each mixture, the total cost, carbon footprint, and energy consumption are calculated by summing the contributions of individual materials based on their unit impacts. These totals are then normalized by compressive strength to produce three strength-normalized indicators. Boxplots for each indicator are used to determine their lower quartile (Q1 = 25 %) thresholds. Mixtures that fall below the Q1 thresholds for all three indicators are considered to have balanced, low environmental and economic impacts. By identifying the minimum and maximum usage values for each material across these optimal mixtures, the approach defines intersection ranges that represent the most sustainable and cost-effective usage levels for practical applications.

3. Results and discussion

3.1. Evaluation of data

To evaluate the effectiveness of data augmentation using generative AI, synthetic data is produced through two distinct approaches (Mahjoubi et al., 2023a): (1) Type I (semi-supervised): The input variables are generated by the generative models, while the output variables are predicted by the trained machine learning models. (2) Type II (unsupervised): Both the input and output variables of the synthetic data are generated by generative models. The quality of synthetic data may vary during generation, so a filtering mechanism is applied to improve reliability. This process uses the input variables from the synthetic dataset to predict the corresponding outputs. Samples with predicted outputs deviating by more than 5 % from the originally generated values are excluded, as discussed and validated in Fig. A2. This ensures only high-quality data is retained for training. As shown in Fig. 5(a), models trained with filtered (Type I) data consistently achieve lower RMSE values than those using unfiltered (Type II) data. The LightGBM model trained with CTGAN-generated data achieved the lowest RMSE, demonstrating the best predictive performance. Therefore, CTGAN with the Type I method was selected for data augmentation, and model performance was evaluated across different augmentation ratios.

The augmentation ratio was defined as the number of synthetic samples relative to the number of real samples. This ratio varied from 0 % to 100 % at an interval of 5 %, and the RMSE of the predictive models was evaluated, as illustrated in Fig. 5(b). Varying the augmentation ratio from 0 % to 100 % is a commonly adopted configuration in data augmentation studies (Jaipuria et al., 2020; Al-Bwana et al.; Yang et al., 2025). The lowest RMSE was observed at an augmentation ratio of 55 %, corresponding to 1269 synthetic samples. This indicates that a moderate level of data augmentation can enhance model accuracy. Beyond this point, RMSE increased, suggesting that excessive synthetic data may introduce noise or lead to overfitting. The optimal ratio results in a 4.87 % improvement in RMSE. Although this may not seem like a huge improvement, it is significant considering that the baseline RMSE is already low (2.707 MPa). This confirms that a moderate amount of synthetic data can improve model performance. Fig. A1 (Appendix) presents a comparison of the distributions between the generated data and the real data with optimal augmentation ratio (55 %). The results show that the generative method produces synthetic data closely aligned with the real data distribution, indicating its effectiveness for data augmentation. The synthetic samples capture the key statistical patterns and feature relationships of the original data, enhancing model generalization.

Table 1 presents a comparison of three generative models (CTGAN, CTABGAN, and VAE) using RMSE, correlation preservation, and diversity index. CTGAN achieves the lowest RMSE (13.86 MPa), indicating the highest prediction accuracy among the three models. CTGAN and CTABGAN show similar effectiveness in maintaining the correlation

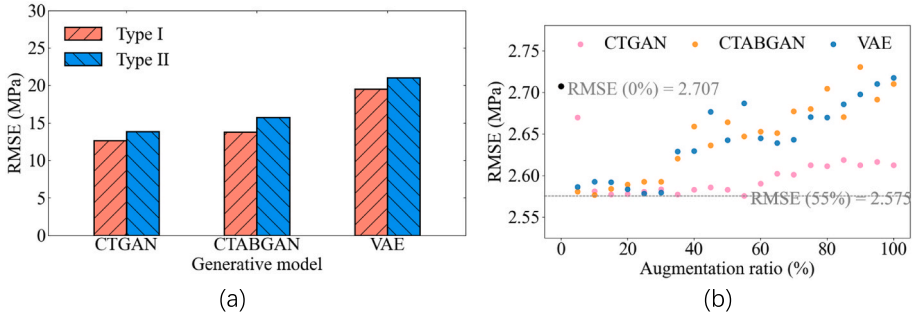

Fig. 5. Illustration of: (a) performance of the LightGBM trained on synthetic data from various generative methods; (b) effect of augmentation ratio on the prediction accuracy.

Table 1
Comparison between different generative models.

Metric	Goal	Generative model		
		CTGAN	CTABGAN	VAE
RMSE	Minimize	13.86	15.70	21.01
Correlation preservation	Minimize	0.09	0.09	0.23
Diversity index	Maximize	2.15	1.82	1.26

structure of the original data, each with a correlation difference of 0.09. In contrast, VAE exhibits a higher deviation at 0.23, indicating weaker correlation preservation. In terms of diversity, CTGAN outperforms the others, achieving the highest diversity index (2.15), suggesting a broader data distribution. These results collectively highlight CTGAN as the most effective model across all metrics.

3.2. Evaluation of machine learning models

The hyperparameters of machine learning models (LR, MLP, DT, RF, XGBoost, LightGBM) are listed in Table A3. These hyperparameters were determined using Grid Search combined with 10-fold cross-validation. Each parameter is briefly described, covering aspects such as learning strategies. For example, in the XGBoost model, a maximum tree depth of 6 and a learning rate of 0.2 were selected to balance model complexity and convergence speed. After determining the optimal hyperparameters, the model is validated using the testing dataset to assess its performance.

The prediction performance of six models is listed in Table 2. The LR model shows the lowest performance, with high MAE (>15 MPa), high RMSE (>20 MPa), and lowest R^2 (0.47). The MLP model demonstrates low accuracy ($R^2 = 0.87$) on the testing set. Tree-based models include DT, RF, XGBoost, and LightGBM exhibit strong predictive power, all achieving R^2 values above 0.95 on the testing set. Notably, LightGBM consistently delivers the best overall performance, with the lowest MAE (3.62 MPa), RMSE (5.54 MPa), and MAPE (0.10 %) on the test data, and the highest R^2 (0.97), confirming its robustness and accuracy in predicting compressive strength. These results underscore the strength of tree-based models, especially LightGBM, in capturing complex patterns.

Fig. 6 compares the predicted and actual results from the six models on both training and testing datasets. LR and MLP models show noticeable scatter in their predictions. This indicates lower accuracy, especially on the testing dataset. For DT, RM, XGBoost, and LightGAM models, the predictions closely align with the actual values, indicating strong model accuracy.

3.3. Parametric analysis

3.3.1. Impact of input variables on compressive strength

This section presents the SHAP analysis for the best-performing model, LightGBM. SHAP reveals the influence of each input feature on the predicted compressive strength, both for individual samples and across the full dataset, as shown in Fig. 7. SHAP analysis highlights ground granulated blast furnace slag as the most influential factor on the compressive strength of GPC, followed by curing time and fly ash content, contributing 36.2 %, 14.0 %, and 7.4 % to the overall impact, respectively. Sodium silicate and sodium hydroxide contents are also identified as significant, while fine aggregate is considered less impactful by LightGBM. Features such as fiber length, fiber diameter, and metakaolin content contribute minimally (less than 2 %).

From a mechanistic standpoint, these data-driven trends are consistent with the underlying geopolymerization chemistry. The strong positive contribution of GGBS reflects its high calcium content, which facilitates the formation of calcium–aluminosilicate–hydrate (C–A–S–H) gels that densify the matrix and enhance compressive strength (Mishra et al., 2024). Longer curing times promote polymerization and further gel development, explaining their high SHAP scores. Moderate Na_2SiO_3 and NaOH contents improve dissolution and polymerization of aluminosilicate precursors, whereas excessive activator concentrations increase viscosity and disrupt gel continuity, leading to strength reduction (Zhang et al., 2014). Fly ash and metakaolin show relatively weaker contributions due to their lower calcium content, forming N–A–S–H gels with limited strength gain (Hanifa et al., 2025). The low SHAP values of aggregates are expected, as they mainly act as inert fillers. These interpretations confirm that the SHAP-derived variable importance aligns with the physicochemical mechanisms governing geopolymer binder formation.

3.3.2. Correlation between parameters

SHAP quantifies the impact of each parameter by analyzing the correlation between input values and SHAP scores. As shown in Fig. 8 (b–g, k, o), higher slag content, molarity, fiber volume, and curing time exhibit clear positive linear trends with compressive strength. The increasing value of such variables corresponds to larger SHAP values. Specifically, $C_{GGBS} > 250$ kg/m^3, Mol >12 mol/L, and FV > 0.5 % enhance compressive strength. Silica fume and metakaolin show negative correlations with the compressive strength. Using less than 300 kg/m^3 of silica fume and 500 kg/m^3 of metakaolin achieve higher strength, as shown in Fig. 8(c) and (d), respectively. Fly ash content of 700–850 kg/m^3 and curing temperatures of 70–100 °C enhance compressive strength, as indicated by elevated SHAP values, as shown in Fig. 8(a) and (p), respectively. Fig. 8 (e, f, i, j) shows that aggregate, sodium silicate, and extra water contents have minimal impact on compressive strength. In contrast, Fig. 8(h) reveals that excessive sodium hydroxide clearly reduces strength, likely due to disrupted geopolymerization from high alkalinity (Asghar et al., 2023). Fiber characteristics (FM, FL, FD) in Fig. 8(l–n) also affect strength, with moderate sizes being favorable. Fibers with modulus above 190 GPa, such as steel fibers, contribute positively (Aisheh et al., 2022).

The mechanistic interpretation of these correlations further supports the data-driven findings. Higher slag content increases calcium availability, which promotes C–A–S–H gel formation and accelerates early strength gain (Mishra et al., 2024). Elevated molarity enhances precursor dissolution, yielding a more interconnected gel network (Shilar et al., 2022); however, extremely high NaOH content causes excessive alkalinity, leading to microcracking and pore coarsening that reduce strength (Zhang et al., 2014). Moderate sodium silicate dosage provides additional soluble silica, balancing polymerization kinetics and gel structure (Shilar et al., 2022). Increased curing temperature and time facilitate dehydration and condensation reactions, thus improving matrix densification (Du et al., 2021). The role of fibers, particularly those with higher modulus, is to bridge microcracks and enhance post-peak behavior (Du et al., 2021), consistent with their modest but positive

Table 2
Prediction performance of various machine learning models.

Metrics	LR		MLP		DT		RF		XGBoost		LightGBM	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
MAE	15.33	15.10	6.39	7.58	0.46	5.10	2.21	4.99	1.44	4.11	1.21	3.62
RMSE	20.80	20.50	8.50	10.48	2.23	7.35	3.74	7.30	2.68	6.49	2.84	5.54
MAPE	0.46	0.52	0.19	0.22	0.01	0.14	0.06	0.15	0.04	0.12	0.03	0.10
R^2	0.57	0.47	0.93	0.87	0.99	0.95	0.09	0.95	0.99	0.96	0.99	0.97

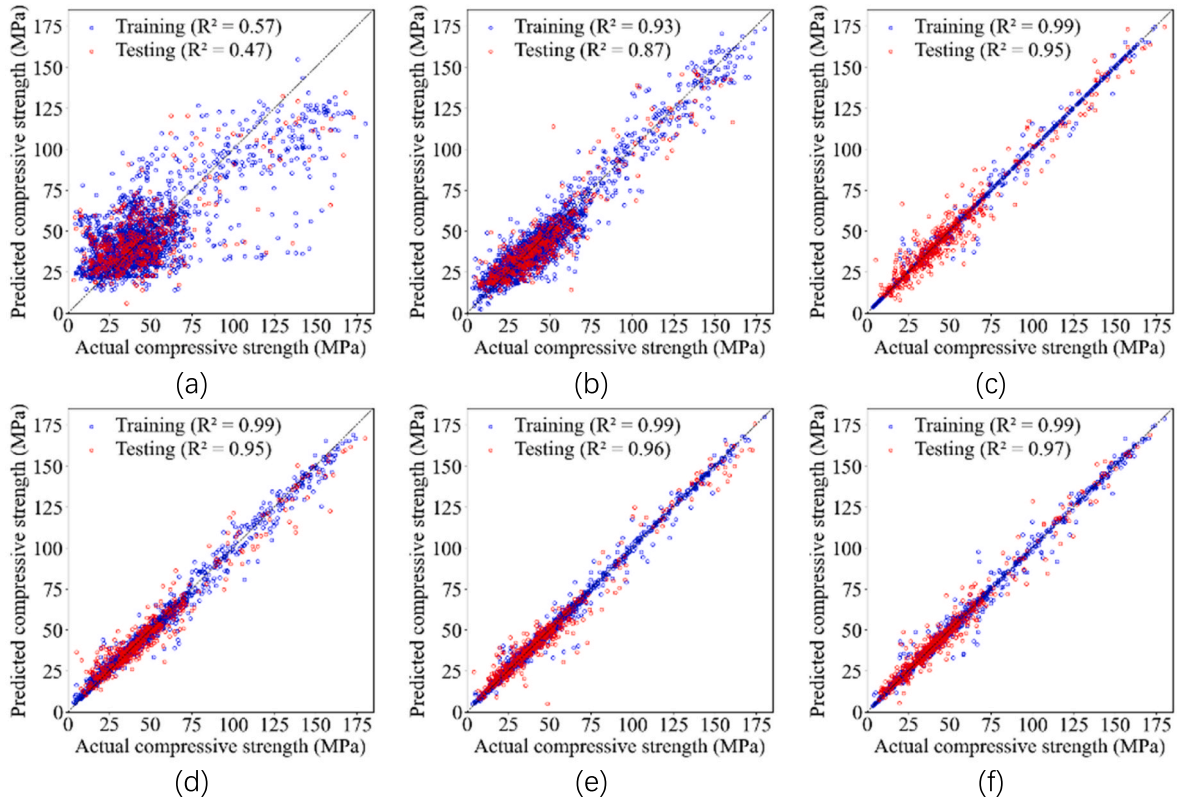

Fig. 6. Performance of machine learning models on training and testing datasets: (a) LR; (b) MLP; (c) DT; (d) RF; (e) XGBoost; and (f) LightGBM.

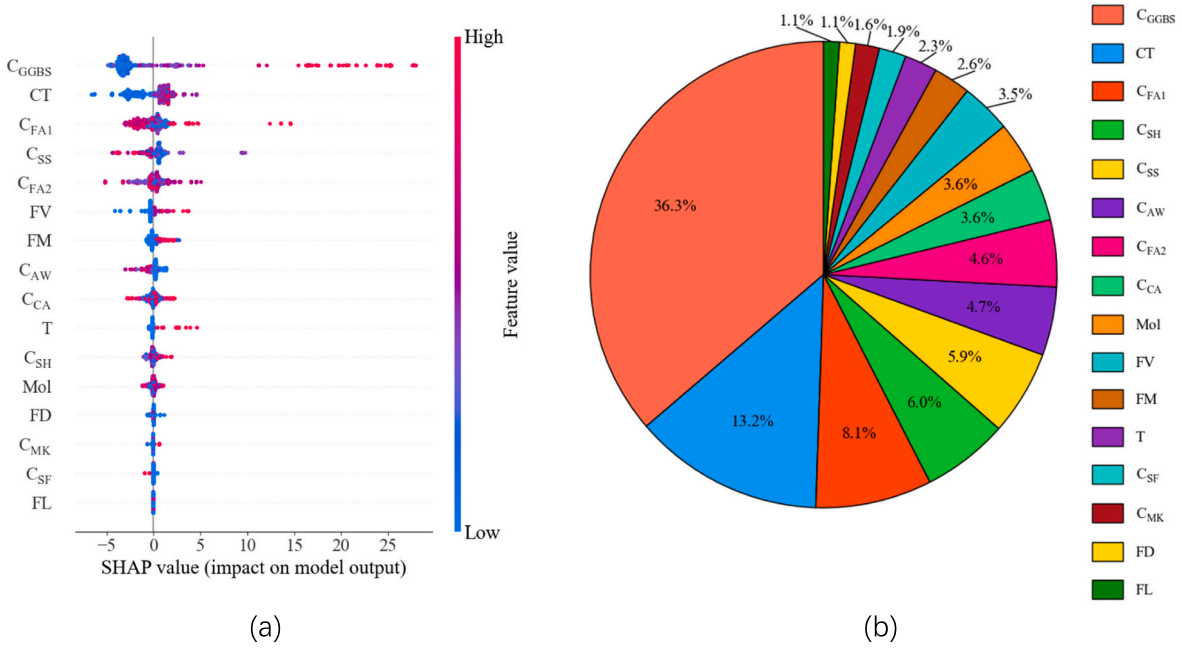

Fig. 7. Impact of input variables on compressive strength: (a) SHAP plot, and (b) pie chart.

SHAP contributions.

3.4. Life cycle assessment results

3.4.1. Cost analysis

Fig. 9(a) shows the total unit cost (from 35.75 USD/m^3 to 1861.53 USD/m^3) of rubberized GPC, with an average of 271.18 USD/m^3. Most mixtures cost less than 300 USD/m^3. This suggests that although a few high-cost outliers exist, likely due to the use of expensive materials such as steel fibers or elevated binder contents, most mixtures remain at low costs. Fig. 9(b) shows the strength-normalized cost, which ranges from 0.57 USD/MPa to 53.4 USD/MPa, with an average of 5.88 USD/MPa. Most mixtures fall below 6 USD/MPa, suggesting reasonable cost-efficiency for achieving strength.

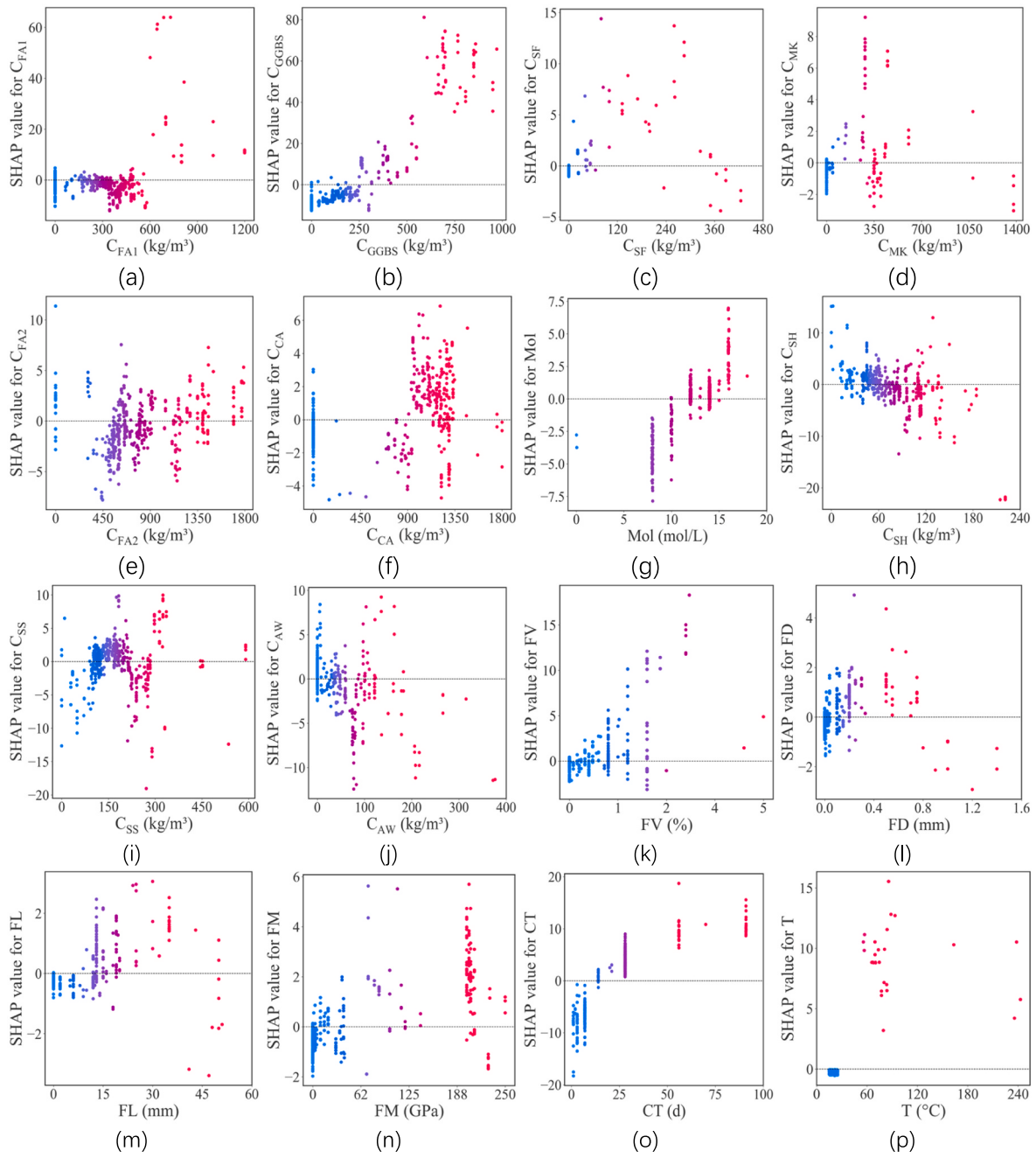

Fig. 8. SHAP distribution of different variables: (a) C_{FA1}; (b) C_{GGBS}; (c) C_{SF}; (d) C_{MK}; (e) C_{FA2}; (f) C_{CA}; (g) Mol; (h) C_{SH}; (i) C_{SS}; (j) C_{AW}; (k) FV; (l) FD; (m) FL; (n) FM; (o) CT; and (p) T.

Fig. 9(c) presents the proportion of each raw material's cost relative to the total cost of GPC. Fly ash, slag, fine and coarse aggregate contribute less than 20 %. Due to their high unit costs, NaOH and Na_2SiO_3 contribute more to the total cost, reaching up to 30 % in certain mixtures. The contributions of silica fume and metakaolin vary widely (0–90 %). Their unit costs are significantly higher than those of fly ash and slag. Steel fiber contributes the highest share to the total cost among all materials, accounting for approximately 20 %–95 %. This due to its high unit cost, which impacts on the total cost of GPC.

3.4.2. *Carbon emission*

Fig. 10(a) presents the total carbon footprint. Emissions range from 52.59 kg/m^3 to 1729.5 kg/m^3, with an average of 476.35 kg/m^3. Most GPC mixtures emit less than 500 kg/m^3, indicating moderate carbon footprint across most designs. Fig. 10(b) shows the strength-normalized carbon footprint for mixtures cured for 28 days, representing the carbon emissions required to achieve 1 MPa of compressive strength. This metric ranges from 1.7 kg/MPa to 98.1 kg/MPa, with an average of 12.4 kg/MPa. Mixtures with very low compressive strengths (e.g., 3–5 MPa) exhibit disproportionately high carbon intensity, often exceeding 100 kg/MPa. However, the majority remain below 15 kg/MPa. Fig. 10(c) shows the contribution of each raw material to the total carbon emissions. Fly ash, slag, silica fume, water, fine aggregate, and coarse aggregate exhibit relatively low contributions, typically accounting for less than 10 %, due to their lower emission factors or moderate usage levels in most mixtures. In contrast, metakaolin and steel fiber display significant variability, with their carbon shares up to 60 % and 90 %, respectively. This wide range reflects differences in mixture design strategies, while some mixtures exclude these components entirely. Notably, NaOH and Na_2SiO_3 contribute substantially to the total carbon

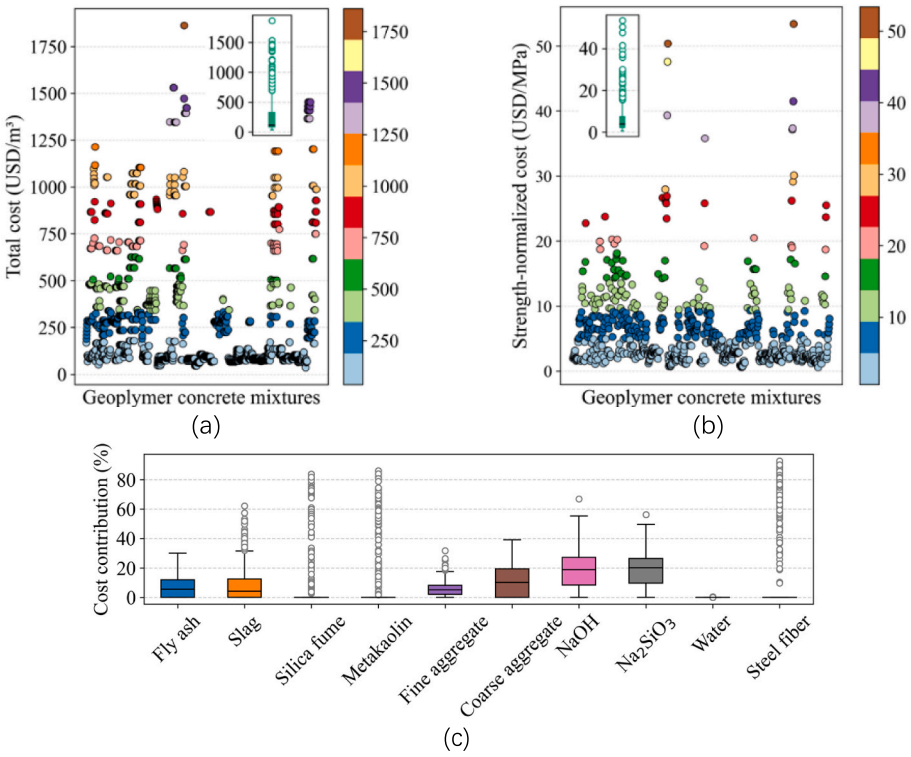

Fig. 9. Cost analysis: (a) total cost; (b) strength-normalized cost; and (c) cost contribution of individual materials in GPC mixtures.

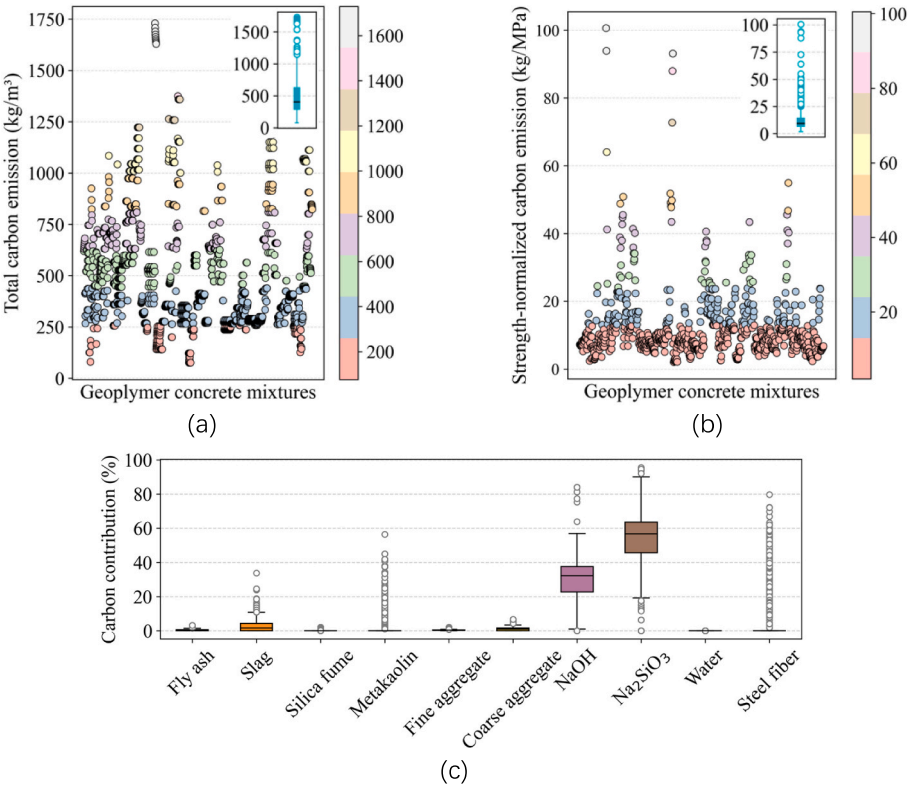

Fig. 10. Carbon analysis: (a) total carbon emission; (b) strength-normalized carbon emission; and (c) carbon contribution of individual materials in GPC mixtures.

emissions, reaching up to 40 % and 65 % in certain mixtures. This is primarily due to their high emission factors and relatively large dosages required for alkali activation in GPC. Their dominance in the carbon profile underscores a critical trade-off between mechanical performance and environmental impact.

3.4.3. *Energy efficiency*

Fig. 11(a) presents the total energy consumption associated with

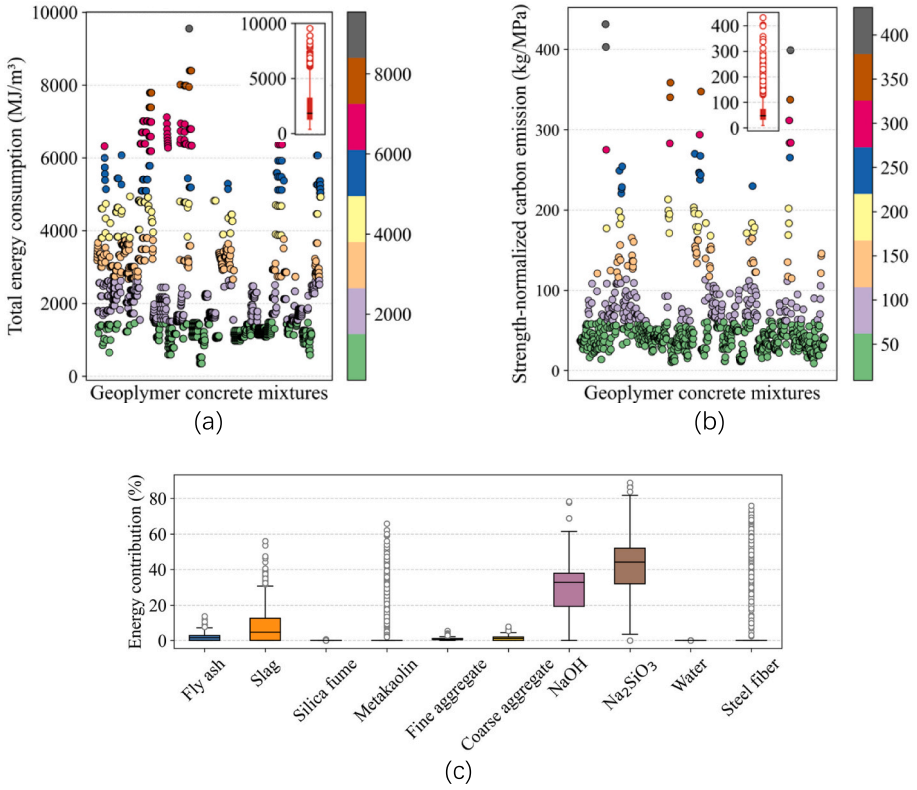

Fig. 11. Energy consumption analysis: (a) total energy consumption; (b) strength-normalized energy consumption; and (c) energy contribution of individual materials in GPC mixtures.

producing 1 m^3 of GPC, as determined from the LCA results. Energy demands range from 320.9 MJ/m^3 to 9519.6 MJ/m^3, with an average of 2491.7 MJ/m^3. The wide range in total energy consumption is primarily attributed to significant variability in material composition across the GPC mixtures. Mixtures with minimal use of high-energy materials, such as steel fiber or chemical activators, tend to fall on the lower end of the spectrum. In contrast, mixes with high steel fiber or alkali activator content show significantly higher energy demand. Most of mixtures fall below 3000 MJ/m^3, suggesting that most designs maintain moderate energy requirements. Fig. 11(b) shows the strength-normalized energy consumption for mixtures cured for 28 days, which represents the energy input required to achieve 1 MPa of compressive strength. This metric varies significantly, ranging from 8.7 MJ/MPa to 429.6 MJ/MPa, with an average of 60.7 MJ/MPa. Most mixtures remain under 100 MJ/MPa, indicating a reasonable balance between energy use and mechanical performance in most cases.

Fig. 11(c) shows the contribution of each raw material to the total energy consumption for GPC. The trend mirrors that of carbon emissions: fly ash, slag, silica fume, water, fine aggregates, and coarse aggregate contribute less than 10 %, while metakaolin and steel fiber vary widely, reaching up to 70 % and 80 %, respectively. NaOH and Na_2SiO_3 show the high contributions (up to 40 % and 55 %) due to their high emission factors and required dosages.

3.4.4. *Optimal material usage range*

The optimal material usage based on Q1 values is shown in Fig. 12. This figure shows material ranges yielding the lowest 25 % of strength-normalized cost, carbon, and energy across all mixtures. The intersection ranges identify material usage linked to low environmental and economic impacts. Fly ash and slag are effective up to 859 kg/m^3 and

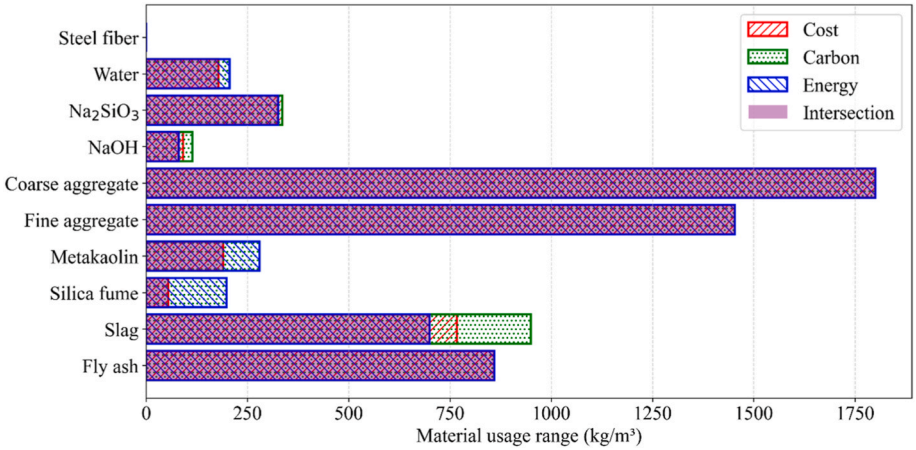

Fig. 12. Optimal material usage for enhanced environmental and economic performance.

700 kg/m^3, while silica fume and metakaolin are limited to 55 kg/m^3 and 190 kg/m^3. Aggregates show the widest ranges, up to 1453 kg/m^3 (fine) and 1800 kg/m^3 (coarse). Steel fiber is absent from the optimal zone, indicating its high cost and environmental load. Sodium hydroxide and sodium silicate are acceptable up to 80 kg/m^3 and 325 kg/m^3, requiring careful dosage to balance strength and sustainability.

3.4.5. Recommendations

The intersection ranges in Section 4.3 provide design constraints for selecting the optimal mixture. Design recommendations are based on 3D plots (Fig. 13) showing trade-offs among strength-normalized cost, carbon emissions, and energy consumption. Each point represents a mix, enabling direct comparison across economic and environmental metrics. All three indicators were normalized using MinMaxScaler and weighted equally (0.33) to reflect balanced importance. Mixes with low values in all three metrics cluster near the origin. Based on 28-day compressive strength, the optimal GPC mixture exhibits a cost of 0.74 USD/MPa, a carbon footprint of 1.84 kg CO_2/MPa, and an energy demand of 10.32 MJ/MPa, combined impact score of 0.0071. This mixture comprises 264.0 kg fly ash, 264.0 slag, 1453.0 kg fine aggregate, 13.4 kg NaOH solution, 72.3 kg Na_2SiO_3 solution, 178.3 kg water. The compressive strength of the optimal mixture is 88.0 MPa. Its optimal performance is a result of using low-impact materials, including the exclusion of fibers and low dosages of NaOH and Na_2SiO_3, along with high compressive strength.

The top 5 % optimal designs shown in Fig. 13 are analyzed for their material composition and performance characteristics. Notably, fly ash ranges from 0 to 859 kg/m^3, and slag varies up to 550 kg/m^3, indicating their flexible role as primary binders. Metakaolin appears in limited quantities, with a maximum of 46.2 kg/m^3, while silica fume is absent across all top-performing mixtures. Fine and coarse aggregates show considerable variation, ranging from 308 to 1453 kg/m^3 and 0–1253 kg/m^3 respectively, reflecting adjustments in total aggregate content. Among the alkaline activators, NaOH and Na_2SiO_3 are present in some mixtures, with maximum values of 61.6 kg/m^3 and 183 kg/m^3, respectively. Water content varies from 82.0 to 178.3 kg/m^3, depending on the mix proportions. Steel fiber is not used in any of the top 5 % mixtures, suggesting that its inclusion may not contribute positively under the selected optimization criteria.

3.4.6. Comparison between GPC and other concrete

Fig. 14 compares the cost, carbon emissions, and energy consumption of GPC, UHPC, and CC at each 10 MPa strength interval, with 10 randomly sampled data per level. At lower strength levels, GPC shows a wider variation and higher dispersion than CC, suggesting greater formulation diversity and sensitivity to precursor composition and activator dosage. The mean cost and energy of GPC rise steadily with strength, accompanied by expanding upper limits, indicating an increasingly resource-intensive production process. In contrast, CC maintains relatively stable mean values with narrow upper and lower bounds across all strength ranges, reflecting its standardized production and limited variability in mixture proportions. UHPC, though having the highest energy and carbon intensity, shows a steadier rise and smaller fluctuation at high strengths, implying a more consistent performance-oriented design strategy. Comparatively, GPC bridges the gap between CC and UHPC: it demonstrates superior sustainability potential over UHPC, especially at moderate strength levels, but with higher environmental loads than CC.

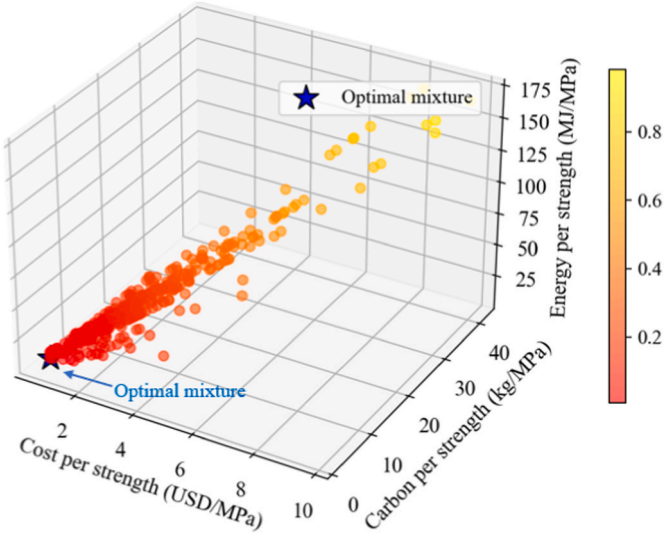

Fig. 13. Evaluation of optimal design based on strength-normalized cost, carbon emission, and energy consumption for each mixture. The color bar indicates the combined effect.

Fig. 15(a) to 15(f) compares the total cost, carbon emission, energy consumption, as well as their strength-normalized counterparts between GPC and conventional concrete (CC). The comparison includes randomly sampled 100 GPC mixtures and 100 CC mixtures, all with compressive strength below 100 MPa. Statistical analysis shows that the lower bound of GPC performance aligns with that of CC, indicating that some GPC mixtures can achieve comparable sustainability. However, most GPCs exhibit higher cost, carbon emissions, and energy consumption than CC. Many GPC mixtures have a higher environmental impact due to the large amounts of alkaline activators, which require energy and carbon-intensive production. Since the cement content in CC is not particularly high, replacing it with large amounts of industrial waste does not lead to a significant reduction in cost and environmental impact.

Fig. 16(a)–16(f) compares GPC and UHPC across total cost, carbon emission, energy consumption, and their strength-normalized counterparts. The comparison includes randomly sampled 100 GPC mixtures and 100 UHPC mixtures (Guo et al., 2023b), all with compressive strength exceeding 100 MPa. Contrary to common assumptions, GPC demonstrates slightly improved sustainability performance overall. Statistical analysis indicates that while a few GPC mixtures show marginally higher impacts, the general trend reveals better sustainability compared to UHPC. In GPC, the amount of steel fiber used is comparable to that in UHPC (Liu et al., 2025), making this component's contribution to cost and environmental impact effectively equivalent. Although GPC extensively incorporates industrial by-products such as fly ash and slag to replace a significant portion of carbon and energy-intensive Portland cement, the use of alkaline activators like NaOH and Na_2SiO_3 contributes substantially to its overall environmental impact. Strength-normalized indicators confirm similar sustainability performance between GPC and UHPC.

3.5. Graphic user interface

The GUI was developed using the PyQt5 framework in Python (Fig. 17). It provides an interactive platform for users to input mixture design parameters and then uses a pre-trained machine learning model to predict compressive strength and sustainability indicators (cost, carbon emissions, energy consumption). Strength-normalized metrics are also provided, enabling users to evaluate both mechanical performance and environmental efficiency. The pie charts show the contributions of different raw materials to total cost, carbon emissions, and energy consumption.

3.6. Discussion

3.6.1. Advantages of machine learning–LCA framework

Evaluating the mechanical and sustainability performance of geopolymer concrete (GPC) mixtures traditionally requires extensive experimental programs due to the large number of design variables,

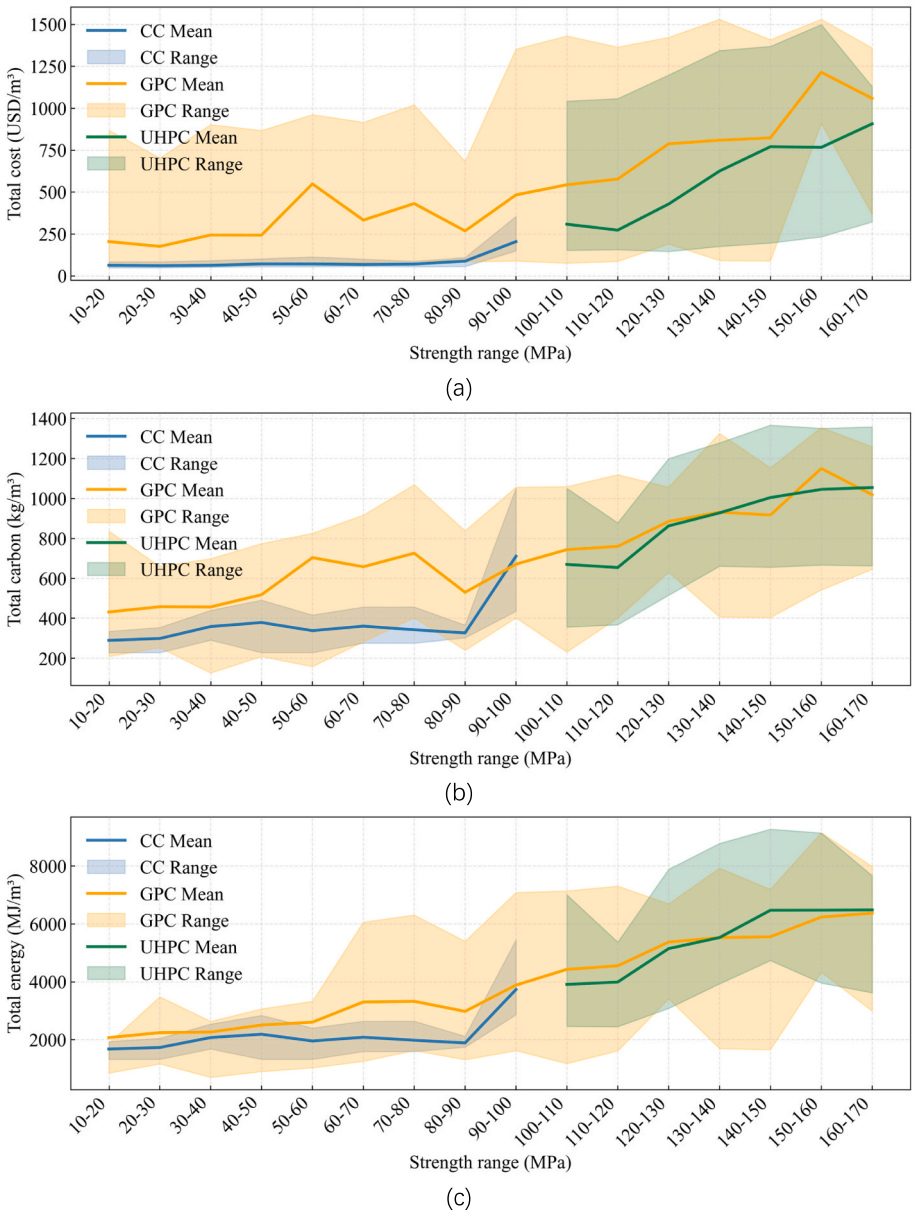

Fig. 14. Comparison of (a) total cost; (b) total carbon emission and (c) total energy consumption of GPC, UHPC, and CC mixtures across strength ranges.

including precursor types, activator compositions, and curing conditions. Such experimental efforts are not only time-consuming and costly but also constrained by the limited availability of testing facilities and the complexity of mix interactions. Moreover, incorporating economic and environmental considerations adds additional complexity to the design process.

Currently, data-driven approaches provide a more economical and efficient way to evaluate the mechanical performance of GPC. By leveraging existing experimental data and advanced machine learning algorithms, these approaches can accurately predict compressive strength and other key mechanical properties without the need for extensive laboratory testing. This not only reduces the cost and duration of experimentation but also enables rapid screening of a wide range of mixture compositions.

In this context, the integration of data-driven modeling and LCA provides an effective alternative to conventional trial-and-error experimentation. By combining generative AI for data augmentation, machine learning for predictive modeling, and cradle-to-gate LCA for sustainability evaluation, this framework enables reliable prediction and optimization of GPC performance without exhaustive laboratory testing. This approach not only accelerates the discovery of sustainable and high-performance mixtures but also provides a scalable foundation for data-informed material design.

3.6.2. *Representativeness of synthetic data*

The reliability of synthetic data generated by generative AI is essential for achieving robust model performance. To evaluate data quality and representativeness, three complementary indicators were employed: RMSE, CP, and DI. RMSE assesses how realistically the synthetic samples reproduce the statistical characteristics of the real dataset. CP measures the consistency of inter-feature relationships between real and synthetic data, while DI quantifies the diversity of generated samples, ensuring that data augmentation enriches the dataset without redundancy. As summarized in Table 1, the CTGAN-generated samples exhibited the lowest RMSE, highest CP, and largest DI, indicating that the synthetic data are realistic, structurally consistent, and sufficiently diverse. Fig. A1 further confirms that the distribution of CTGAN-generated samples closely aligns with that of the real data,

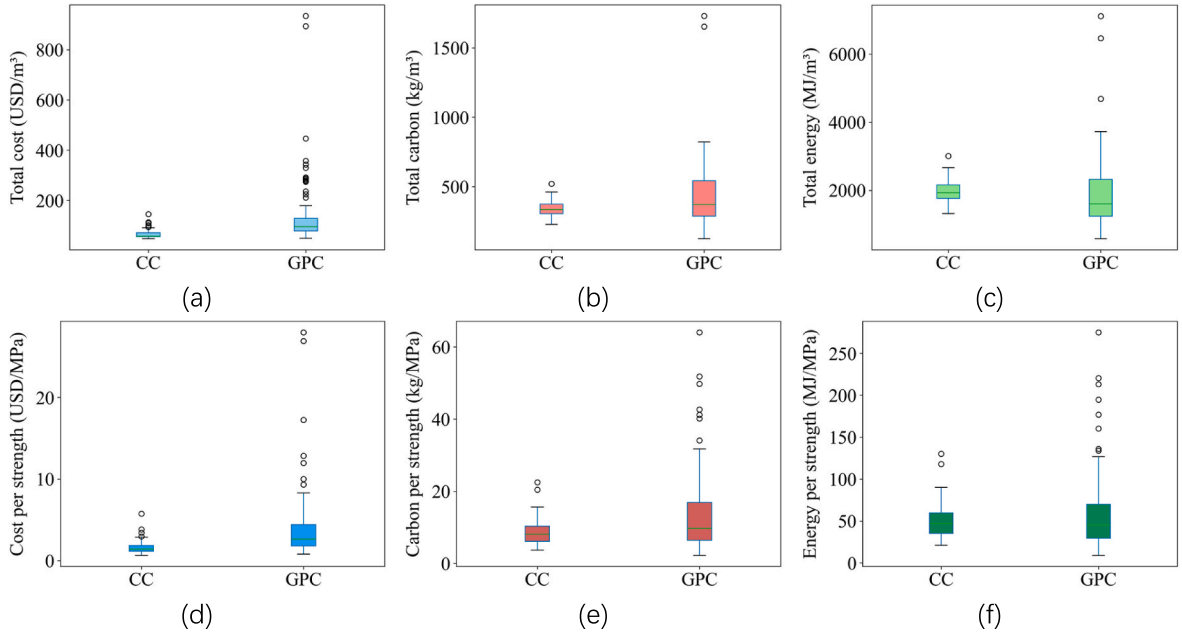

Fig. 15. Comparison of GPC and CC in terms of: (a) cost; (b) carbon emission; (c) energy consumption; (d) strength-normalized cost; strength-normalized carbon emission; strength-normalized energy consumption.

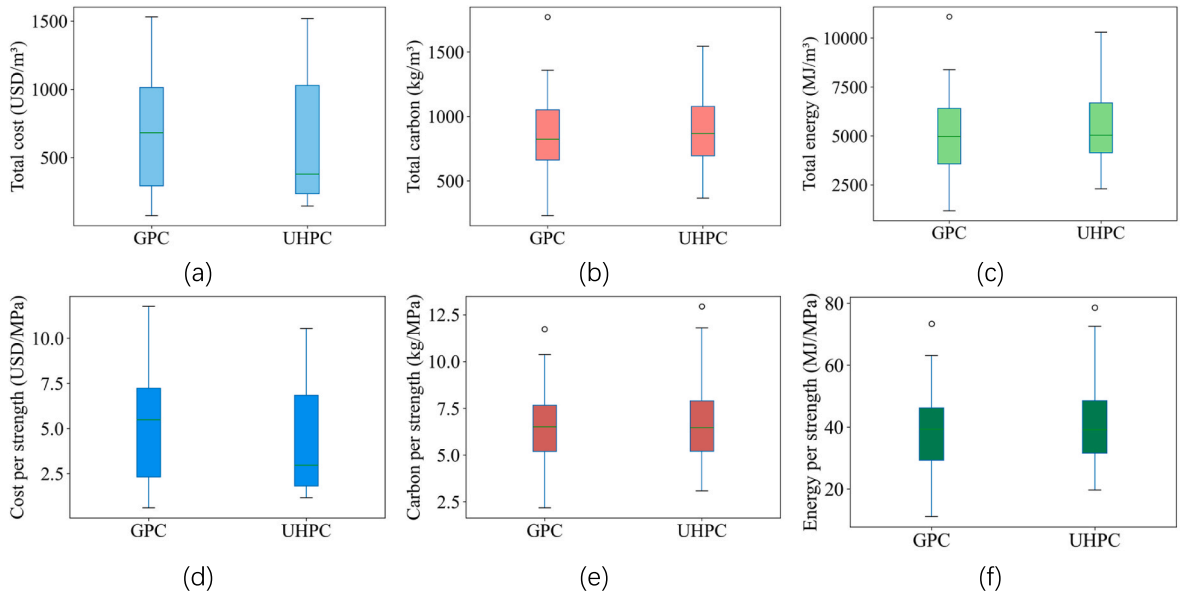

Fig. 16. Comparison of GPC and UHPC in terms of: (a) cost; (b) carbon emission; (c) energy consumption; (d) strength-normalized cost; (e) strength-normalized carbon emission; and (f) strength-normalized energy consumption.

demonstrating that the model effectively captures key statistical characteristics. Additionally, a comparison between models trained on augmented datasets and those trained solely on real data (Fig. 5) shows a 4.87 % reduction in RMSE, verifying that synthetic data enhance model generalization rather than introducing bias. Together, these results confirm that the generated data are both representative and reliable for model training, forming a solid foundation for improving the generalization and stability of predictive models.

3.6.3. *Generalization of machine learning*

The generalization capability of the best-performing machine learning model, LightGBM, was evaluated through 10-fold cross-validation, a widely adopted approach to mitigate overfitting and assess model robustness. Although no independent external dataset was available for validation, this method effectively ensures that each subset of data contributes to both training and testing. The consistently high R^2 values (>0.94) across all folds demonstrate that the model achieves strong generalization performance without signs of overfitting, as shown in Fig. 18. This confirms that the model captures stable and consistent relationships between inputs and strength.

In addition, the influence of synthetic data proportion on model performance was systematically investigated by varying the augmentation ratio from 0 % to 100 % (Fig. 5 b). The lowest RMSE was achieved at an augmentation ratio of 55 %, while higher ratios resulted in degraded performance, confirming that excessive synthetic data may lead to overfitting. This analysis demonstrates that model performance

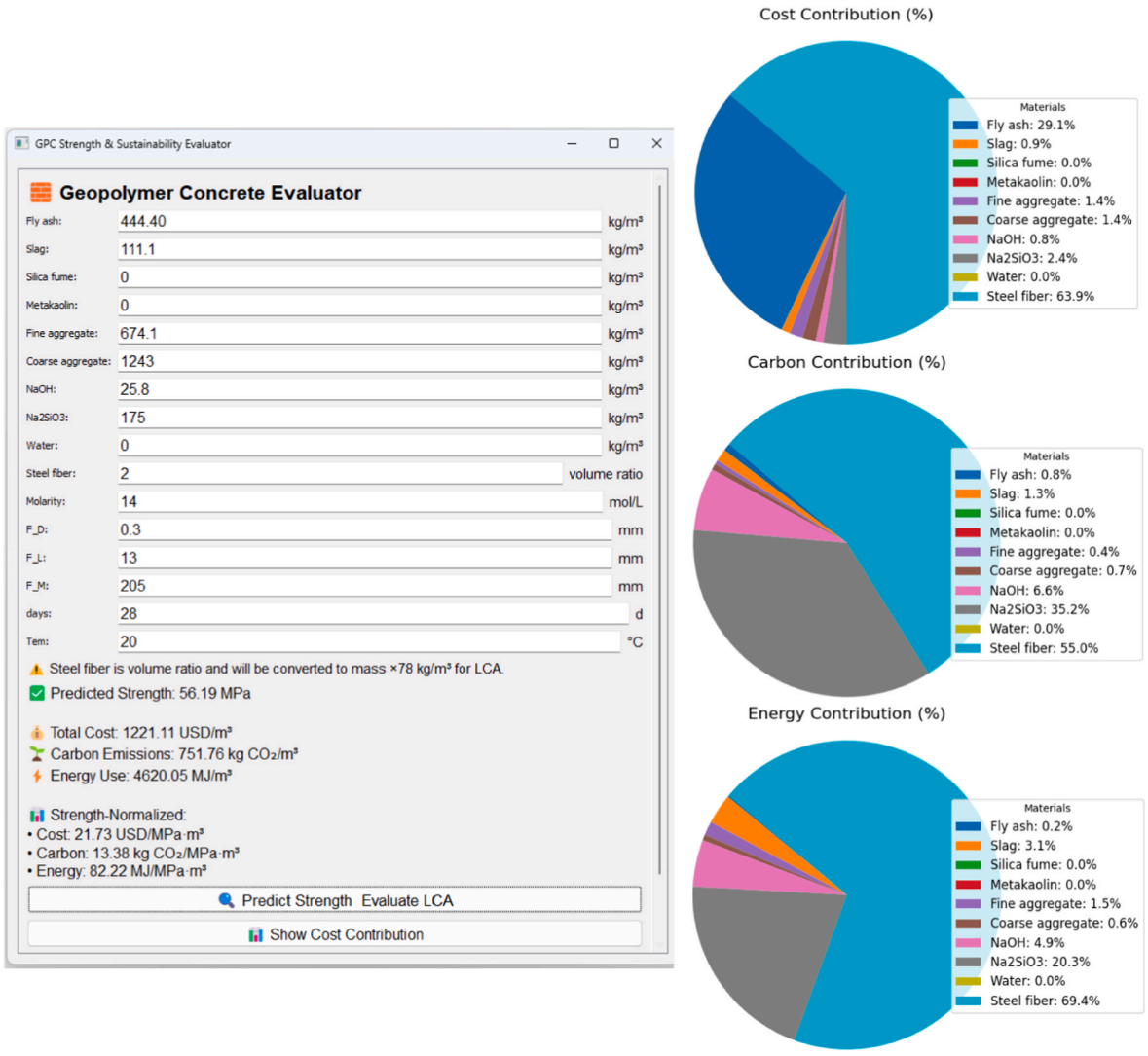

Fig. 17. GUI example for predicting compressive strength and assessing GPC sustainability.

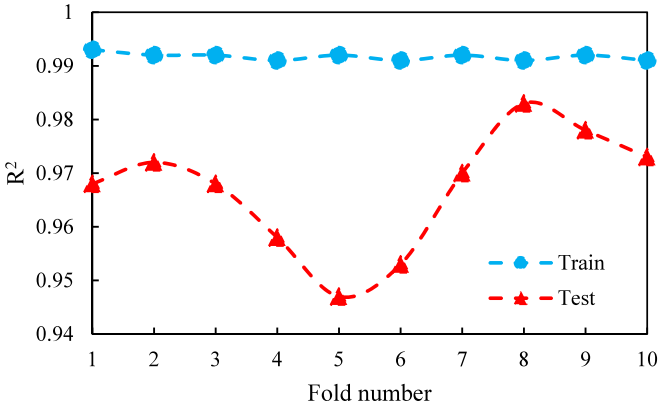

Fig. 18. Generalization performance on 10-fold cross-validation results of the LightGBM model.

improves only when the generated samples enrich data diversity rather than dominate the dataset. Therefore, the combined use of cross-validation and controlled augmentation ratio effectively mitigates overfitting and enhances model generalization to unseen GPC mixtures.

3.6.4. *Benefits of strength-normalized sustainability metrics*

This research incorporates strength-normalized sustainability metrics. These indicators include cost per unit strength (USD/MPa), carbon emission per unit strength (kg CO_2/MPa), and energy consumption per unit strength (MJ/MPa), representing the eco-efficiency of a mixture, reflecting how effectively mechanical performance is achieved relative to environmental and economic burdens. Within the current dataset, strength-normalized sustainability metrics reveal trade-offs between strength and resource efficiency. For instance, mixtures with excessive use of NaOH and Na_2SiO_3 exhibit disproportionately high cost- and energy-normalized values despite achieving high compressive strength, indicating diminishing sustainability returns at higher activator concentrations. Conversely, certain mixtures with moderate binder contents and optimized curing conditions achieve a balanced performance, with normalized carbon footprint and energy demand comparable to those of conventional concrete. Such mixtures highlight the importance of mix design optimization for achieving practical sustainability rather than merely reducing absolute emissions or cost. In practical terms, these normalized metrics can guide performance-based material selection. For example, GPC mixtures with strength-normalized carbon footprints below 10 kg CO_2/MPa and costs below 6 USD/MPa can be considered both structurally and environmentally efficient for medium-strength infrastructure applications. For ultra-high-strength GPC, while absolute impacts are higher, strength-normalized indicators remain

competitive with UHPC, underscoring their potential for structural applications requiring superior durability and load resistance.

3.6.5. Challenges and future research

The analysis in Section 3.4 is not fully comprehensive, as the current dataset only includes conventional fiber-reinforced geopolymer mixtures. Within this scope, fibers and alkali activators are identified as the main contributors to cost and environmental impact. However, in certain cases, their use may still be justified for improved mechanical performance and durability. Although steel fibers were excluded due to their higher cost and environmental footprint, they offer notable structural advantages in applications requiring high tensile strength, cracking control, or impact resistance (Du et al., 2024). Therefore, the selection of fiber type should balance environmental, economic, and structural considerations.

Compared with conventionally manufactured chemical activators, those derived from industrial or agricultural wastes show strong potential to replace traditional water glass (Wang et al., 2025; Rajan and Kathirvel, 2021). Solid waste such as rice husk and waste glass can be calcined and blended with NaOH to produce a low-carbon solid activator (Nassar et al., 2025a, 2025b; Rodriguez-Morales and Escalante-Garcia, 2024). Likewise, almond shell ash rich in KOH and K_2CO_3 can completely replace conventional activators while reducing carbon emissions by up to 56 % (Fan et al., 2025). A low-carbon activator synthesized from waste glass powder via a hydrothermal process with NaOH (6 M, 60–80 °C, 48 h) demonstrated comparable reactivity to commercial waterglass with up to 40 % lower carbon footprint (Navarro et al., 2025). In conclusion, future studies should focus on developing cost-effective, low-carbon activators using solid waste to replace alkali solutions and water glass.

The optimal upper and lower bounds of the material volumes were determined from the current dataset. Within this range, the results provide valuable guidance for mix optimization and simplify the design process of geopolymer concrete. Nevertheless, the dataset remains limited in scope, and these bounds may not fully represent the variability of materials and mix proportions in real-world applications. Future studies should aim to expand the dataset to cover a wider range of mixture designs and experimentally validate the upper and lower bounds to enhance their reliability and practical relevance. The accuracy of the LCA results is influenced by the quality and consistency of the inventory data. The data used were compiled from multiple literature sources, which may differ in terms of material composition, geographical origin, manufacturing process, and reporting methodology. Such inconsistencies can introduce uncertainties and potential biases into the life cycle results, affecting the comparability and reliability of the assessment. Therefore, future studies should prioritize the development of a standardized, region-specific inventory database for geopolymer concrete. Establishing such a database would significantly improve the consistency and transparency of LCA evaluations, enabling more reliable and data-informed decision-making in sustainable material design and selection.

As the dataset developed in this study primarily focuses on two-part GPC mixtures activated by liquid alkalis, the proposed data-driven and LCA-based framework is equally applicable to one-part GPC systems. To extend this approach, new datasets specific to one-part GPC incorporating parameters such as solid activator dosage and precursor composition need to be established. Similarly, a corresponding life cycle inventory database should be developed to capture the environmental impacts associated with solid activator production and use. Future research will focus on expanding the framework to one-part GPC, thereby enhancing its general applicability across different types of geopolymer systems.

4. Conclusion

This paper presents a comprehensive assessment of the cradle-to-gate life cycle cost, carbon footprint, and energy consumption of 2304 GPC mixtures which represent the state-of-the-art dataset, aiming to establish a holistic understanding of the impacts of GPC design parameters on mechanical and sustainability performance. Based on the above investigations, the following conclusions can be drawn:

- Based on existing GPC, among three categories of ingredients (fibers, alkaline activators, and granular ingredients), steel fibers exhibited the highest impacts on material cost (95 %), carbon footprint (90 %), and energy consumption (80 %); NaOH and Na_2SiO_3 demonstrated the second highest impacts: up to 30 % for cost, up to 65 % for carbon footprint, and up to 55 % for energy consumption; and granular ingredients had the lowest contributions to the cost, carbon footprint, and energy consumption, suggesting that an effectiveness ranking of steel fibers > alkaline activators > granular ingredients including solid wastes. However, despite their higher cost and environmental impacts, steel fibers offer superior structural performance in terms of tensile strength, crack control, and toughness, and their inclusion may still be justified in applications where such properties are critical. Future studies should focus on developing cost-effective, low-carbon activators using solid wastes to partially or fully replace alkali solutions and water glass, thereby further enhancing the sustainability of geopolymer production.
- The upper limits of the contents of various materials are recommended based on the LCA, aiming to minimize strength-normalized cost and environmental impacts: Fine and coarse aggregates contents are no higher than 1453 kg/m^3 and 1800 kg/m^3, respectively. Fly ash, slag, silica fume, metakaolin, NaOH, and Na_2SiO_3 contents are no higher than 859 kg/m^3, 700 kg/m^3, 55 kg/m^3, 190 kg/m^3, 80 kg/m^3, and 325 kg/m^3, respectively.
- Ultra-high-strength GPC (compressive strength >100 MPa) and ultra-high-performance concrete (compressive strength >120 MPa) exhibited comparable cost, carbon footprint, and embodied energy. Overall, normal-strength GPC (compressive strength <100 MPa) had higher cost, carbon footprint, and embodied energy than the Portland cement concrete counterpart due to the high dosages of alkali activators.

The data adopted in this study were sourced from available references. The data contained inconsistencies in materials, geographical origins, manufacturing processes, and reporting formats. These inconsistencies reduce the reliability of assessments. As a result, there is a critical need for dedicated research to develop a standardized, region-specific database for GPC. Establishing such a database will enhance the consistency and accuracy of assessments, ultimately supporting data-informed decision-making in material selection and formulation design.

CRediT authorship contribution statement

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

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Weina Meng received funding from the National Science Foundation.

Acknowledgement

This research was funded by the United States National Science Foundation [award number: CMMI-2046407].

Appendix

Table A1
Distribution of mixture design variables and compressive strength of GPC

No.	Symbol	Unit	Minimum	Maximum	Mean	STD
1	C_{FA1}	kg/m^3	0.0	1199.0	286.8	203.3
2	C_{GGBS}	kg/m^3	0.0	1175.0	184.8	235.8
3	C_{SF}	kg/m^3	0.0	800.0	22.1	84.4
4	C_{MK}	kg/m^3	0.0	1500.0	64.7	185.9
5	C_{FA2}	kg/m^3	0.0	1900.0	816.7	377.9
6	C_{CA}	kg/m^3	0.0	1800.0	676.4	571.9
7	Mol	mol/L	0.0	20.0	12.4	2.4
8	C_{SH}	kg/m^3	0.0	500.0	77.0	45.3
9	C_{SS}	kg/m^3	0.0	750.0	180.9	89.2
10	C_{AW}	kg/m^3	0.0	396.0	33.9	58.2
11	*FV*	%	0.0	5.0	0.4	0.7
12	*FD*	mm	0.0	1.4	0.09	0.04
13	*FL*	mm	0.0	65.0	6.0	10.1
14	*FM*	GPa	0.0	250.0	47.6	80.9
15	CT	days	0.0	365.0	23.5	27.1
16	T	°C	7.0	240.0	25.2	25.3
17	f_c	MPa	3.6	180.0	48.1	31.1

Table A2
Inventory data for each raw ingredient used in LCA

No.	Materials	Cost (USD/kg)	Carbon footprint (kg/kg)	Energy consumption (MJ/kg)
1	Fly ash	0.026 Guo et al., 2023b	0.005 Guo et al., 2023b	0.10 Mahjoubi et al., 2025
2	Slag	0.100 Guo et al., 2023b	0.085 Guo et al., 2023b	1.3 Mahjoubi et al., 2025
3	Silica fume	0.800 Guo et al., 2023b	0.014 Guo et al., 2023b	0.018 Mahjoubi et al., 2025
4	Metakaolin	0.500 Guo et al., 2023b	0.332 Guo et al., 2023b	3.48 Mahjoubi et al., 2025
5	Fine aggregate	0.010 Guo et al., 2023b	0.002 Guo et al., 2023b	0.022 Mahjoubi et al., 2025
6	Coarse aggregate	0.014 Guo et al., 2023b	0.004 Guo et al., 2023b	0.022 Mahjoubi et al., 2025
7	Sodium hydroxide	0.380 Guo et al., 2024a	1.915 Guo et al., 2024a	8.75 Jamieson et al., 2015
8	Sodium silicate	0.170 Guo et al., 2024a	1.514 Guo et al., 2024a	5.37 Kaplan et al., 2023
10	Water	0.001 Guo et al., 2023b	0.0003 Guo et al., 2023b	0.00574 Chiaia et al., 2014
11	Steel fiber	5.0 Guo et al., 2023b	2.65 Guo et al., 2023b	20.56 Mahjoubi et al., 2025

Table A3
Optimal hyperparameter for various machine learning models

Algorithms	Hyperparameters	Descriptions
LR	*fit_intercept* = *True*	Whether to calculate the intercept for this model
MLP	*normalize* = *False*	Whether to normalize the input features before fitting
	$S_1 = 100$	Number of neurons in first hidden layer
	$S_2 = 50$	Number of neurons in second hidden layer
	$\alpha = 0.00003$	Strength of the L2 regularization term
	$l_r = 0.001$	Learning rate
	batch_size = *'auto'*	Size of minibatches used for optimization
DT	*Imax* = *None*	Maximum number of leaf nodes
	$n_f = 2$	Number of features to consider for the best split
	S = *'best'*	Strategy for splitting nodes
	criterion: squared error	Squared error is adopted as the splitting criterion
RF	$n_e = 90$	Whether to calculate the intercept for this model
	$d_{\max}$ = *None*	Maximum depth of the tree
	$L_{\min} = 1$	Minimum number required to be at a leaf node
	$n_f = 2$	Number of features to consider for the best split
	criterion: mse	The function to measure the quality of a split
XGBoost	$n_e = 200$	Number of estimators
	$l_r = 0.2$	Learning rate
	$\gamma_0 = 0$	Minimum loss reduction for a further leaf partition

(*continued on next page*)

Table A3 (*continued*)

Algorithms	Hyperparameters	Descriptions
	$d_{max} = 6$	Maximum depth of each tree
	reg_alpha = 0	L1 regularization term coefficient
	reg_lambda = 1	L2 regularization term coefficient
LightGBM	$n_b = 2$	Number of boosting iterations
	$n_e = 69$	Number of boosting rounds
	$l_r = 0.32$	Learning rate
	$d_{max} = -1$	Maximum depth of the tree
	$b_f = 1.0$	Fraction of training data in each boosting iteration

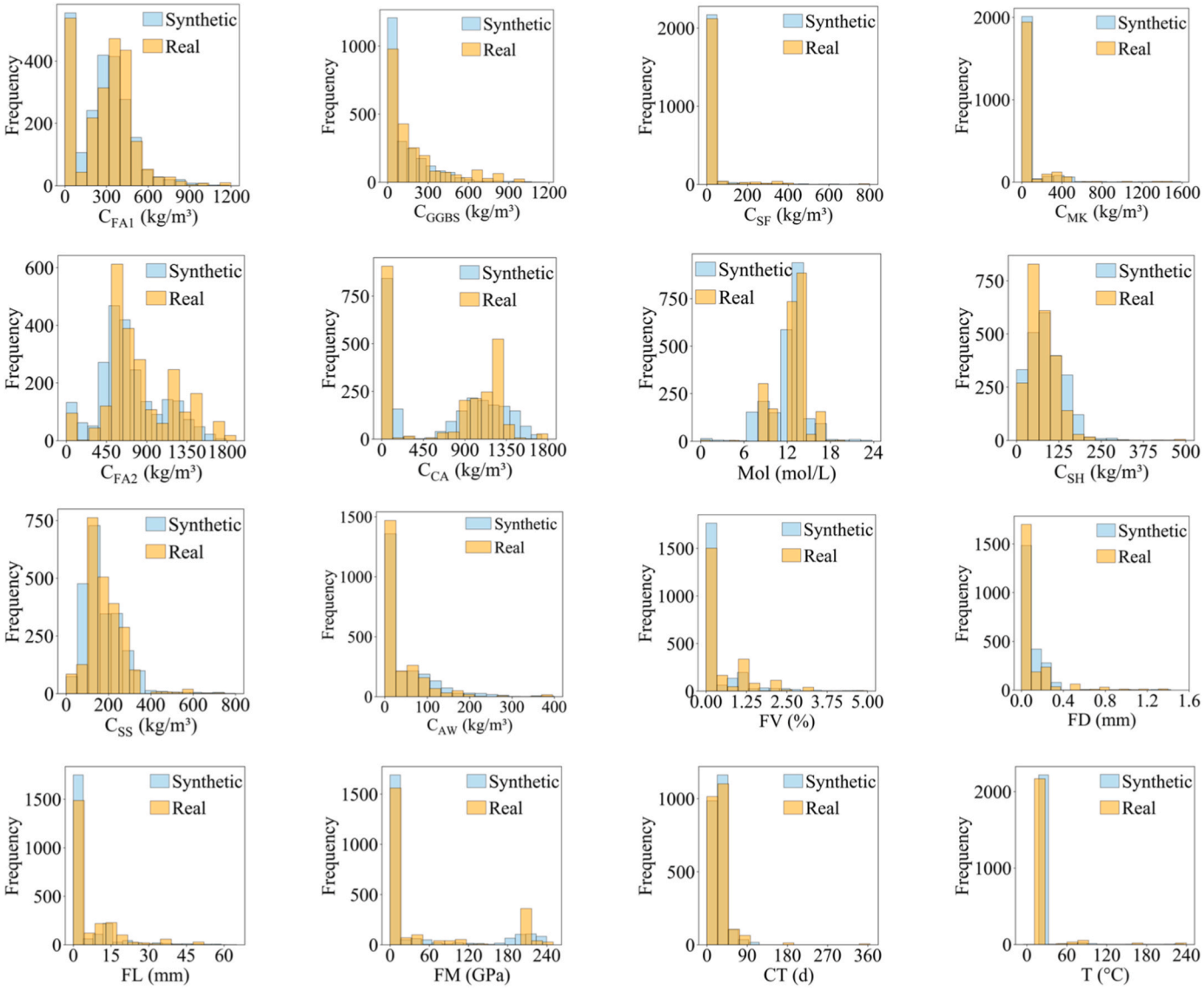

Fig. A1. Comparison of real and synthetic dataset distributions.

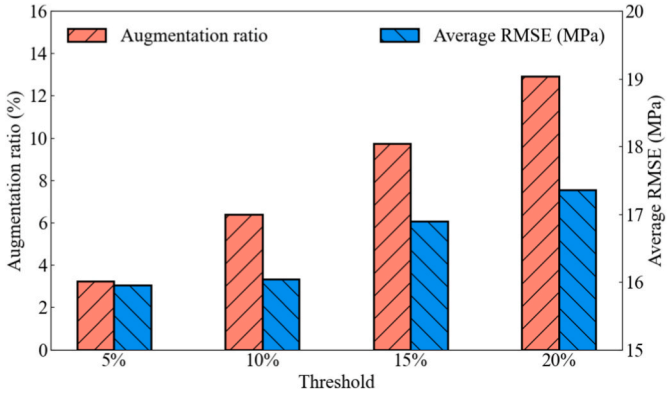

Fig. A2. Effect of deviation threshold on data filtering performance in generative data augmentation. As the threshold increases from 5 % to 20 %, the augmentation ratio rises steadily, indicating that more synthetic samples are retained. However, the average RMSE also increases, suggesting reduced data quality and model accuracy at higher thresholds. A 5 % threshold achieves the lowest RMSE and a reasonable augmentation ratio, representing the best trade-off between data quality and sample quantity.

Data availability

Geopolymer concrete datasets are available at: https://github.com/2740267974/Geopolymer-Concrete-data.

References

Abdellatief, M., Hassan, Y.M., Elnabwy, M.T., Wong, L.S., Chin, R.J., Mo, K.H., 2024. Investigation of machine learning models in predicting compressive strength for ultra-high-performance geopolymer concrete: a comparative study. Constr. Build. Mater. 436, 136884. https://doi.org/10.1016/j.conbuildmat.2024.136884.

Aisheh, Y.I.A., Atrushi, D.S., Akeed, M.H., Qaidi, S., Tayeh, B.A., 2022. Influence of steel fibers and microsilica on the mechanical properties of ultra-high-performance geopolymer concrete (UHP-GPC). Case Stud. Constr. Mater. 17, e01245. https://doi.org/10.1016/j.cscm.2022.e01245.

Akan, M.Ö.A., Dhavale, D.G., Sarkis, J., 2017. Greenhouse gas emissions in the construction industry: an analysis and evaluation of a concrete supply chain. J. Clean. Prod. 167, 1195–1207. https://doi.org/10.1016/j.jclepro.2017.07.225.

Al-Bwana, E.K., Alauthman, M., Sayahi, I., and Mahjoub, M.A., TGAN and CTGAN: a comparative analysis for augmenting COVID 19 tabular data. DOI: https://doi.org/10.5220/0013483200003929.

Albidah, A., Alghannam, M., Abbas, H., Almusallam, T., Al-Salloum, Y., 2021. Characteristics of metakaolin-based geopolymer concrete for different mix design parameters. J. Mater. Res. Technol. 10, 84–98. https://doi.org/10.1016/j.jmrt.2020.11.104.

Albidah, A., Alqarni, A.S., Abbas, H., Almusallam, T., Al-Salloum, Y., 2022. Behavior of metakaolin-based geopolymer concrete at ambient and elevated temperatures. Constr. Build. Mater. 317, 125910. https://doi.org/10.1016/j.conbuildmat.2021.125910.

Ali, S., Sheikh, M.N., Hadi, M.N., 2024. Effect of glass fibre on slag-fly ash based geopolymer concrete. Procedia Struct. Integr. 64, 1394–1401. https://doi.org/10.1016/j.prostr.2024.09.379.

Amran, M., Makul, N., Fediuk, R., Lee, Y.H., Vatin, N.I., Lee, Y.Y., Mohammed, K., 2022. Global carbon recoverability experiences from the cement industry. Case Stud. Constr. Mater. 17, e01439. https://doi.org/10.1016/j.cscm.2022.e01439.

Asghar, R., Khan, M.A., Alyousef, R., Javed, M.F., Ali, M., 2023. Promoting the green construction: scientometric review on the mechanical and structural performance of geopolymer concrete. Constr. Build. Mater. 368, 130502. https://doi.org/10.1016/j.conbuildmat.2023.130502.

Bellum, R.R., 2022. Influence of steel and PP fibers on mechanical and microstructural properties of fly ash-GGBFS based geopolymer composites. Ceram. Int. 48 (5), 6808–6818. https://doi.org/10.1016/j.ceramint.2021.11.232.

Chiaia, B., Fantilli, A.P., Guerini, A., Volpatti, G., Zampini, D., 2014. Eco-mechanical index for structural concrete. Constr. Build. Mater. 67, 386–392. https://doi.org/10.1016/j.conbuildmat.2013.12.090.

Deng, Z., Zhang, S., Deng, Z., 2023. PVA fiber-reinforced geopolymer mortar made with hybrid recycled aggregates: toward thermal insulation, lightweight and improved durability. J. Clean. Prod. 426, 139200. https://doi.org/10.1016/j.jclepro.2023.139200.

Du, J., Meng, W., Khayat, K.H., Bao, Y., Guo, P., Lyu, Z., Abu-Obeidah, A., Nassif, H., Wang, H., 2021. New development of ultra-high-performance concrete (UHPC). Compos. B Eng. 224, 109220. https://doi.org/10.1016/j.compositesb.2021.109220.

Du, J., Wang, Y., Guo, P., Meng, W., 2024. Tailoring of steel fiber surface by coating cellulose nanocrystal for enhanced flexural properties of UHPC. Cement Concr. Compos. 154, 105773. https://doi.org/10.1016/j.cemconcomp.2024.105773.

Eisa, M., Basiouny, M., Fahmy, E., 2022. Drying shrinkage and thermal expansion of metakaolin-based geopolymer concrete pavement reinforced with biaxial geogrid. Case Stud. Chem. Environ. Eng. 17, e01415. https://doi.org/10.1016/j.cscm.2022.e01415.

Eltantawi, I., Sheikh, M.N., Hadi, M.N., 2025. Influence of stainless steel fibres on the mechanical properties of self-compacting ultra-high-performance geopolymer concrete. Constr. Build. Mater. 463, 140092. https://doi.org/10.1016/j.conbuildmat.2025.140092.

Esparham, A., Ghalatian, F., 2022. The features of geopolymer concrete as a novel approach for utilization in green urban structures. J. Compos. Comp. 4 (11), 89–96. https://doi.org/10.52547/jcc.4.2.4.

Fan, J., Yeo, B.-L., Markandeya, A., Martínez, A., Harvey, J., Miller, S.A., Nassiri, S., 2025. Transforming almond shell waste into high-value activators for low-carbon concrete: life cycle assessment and technoeconomic analysis. J. Build. Eng. 105, 112547. https://doi.org/10.1016/j.jobe.2025.112547.

Ganesan, N., Abraham, R., Raj, S.D., 2015. Durability characteristics of steel fibre reinforced geopolymer concrete. Constr. Build. Mater. 93, 471–476. https://doi.org/10.1016/j.conbuildmat.2015.06.014.

Ganesh, A.C., Muthukannan, M., 2021. Development of high performance sustainable optimized fiber reinforced geopolymer concrete and prediction of compressive strength. J. Clean. Prod. 282, 124543. https://doi.org/10.1016/j.jclepro.2020.124543.

Guo, P., Bao, Y., Meng, W., 2021a. Review of using glass in high-performance fiber-reinforced cementitious composites. Cement Concr. Compos. 120, 104032. https://doi.org/10.1016/j.cemconcomp.2021.104032.

Guo, P., Meng, W., Xu, M., Li, V.C., Bao, Y., 2021b. Predicting mechanical properties of high-performance fiber-reinforced cementitious composites by integrating micromechanics and machine learning. Materials 14 (12), 3143. https://doi.org/10.3390/ma14123143.

Guo, P., Meng, W., Du, J., Han, B., Bao, Y., 2023a. Lightweight ultra-high-performance concrete (UHPC) with expanded glass aggregate: development, characterization, and life-cycle assessment. Constr. Build. Mater. 371, 130441. https://doi.org/10.1016/j.conbuildmat.2023.130441.

Guo, P., Mahjoubi, S., Liu, K., Meng, W., Bao, Y., 2023b. Self-updatable AI-assisted design of low-carbon cost-effective ultra-high-performance concrete (UHPC). Case Stud. Constr. Mater. 19, e02625. https://doi.org/10.1016/j.cscm.2023.e0262.

Guo, P., Meng, W., Bao, Y., 2024a. Knowledge-guided data-driven design of ultra-high-performance geopolymer (UHPG). Cement Concr. Compos. 153, 105723. https://doi.org/10.1016/j.cemconcomp.2024.105723.

Guo, P., Meng, W., Bao, Y., 2024b. Intelligent characterization of complex cracks in strain-hardening cementitious composites based on generative computer vision. Constr. Build. Mater. 411, 134812. https://doi.org/10.1016/j.conbuildmat.2023.134812.

Guo, P., Meng, W., Bao, Y., 2024c. Knowledge graph-guided data-driven design of ultra-high-performance concrete (UHPC) with interpretability and physicochemical reaction discovery capability. Constr. Build. Mater. 430, 136502. https://doi.org/10.1016/j.conbuildmat.2024.136502.

Guterres, A., 2020. Carbon neutrality by 2050: the world's most urgent mission. https://www.un.org/sg/en/content/sg/articles/2020-12-11/carbon-neutrality-2050-the-world%E2%80%99s-most-urgent-mission.

Habibi, O., Chemmakha, M., Lazaar, M., 2023. Imbalanced tabular data modelization using CTGAN and machine learning to improve IoT botnet attacks detection. Eng. Appl. Artif. Intell. 118, 105669. https://doi.org/10.1016/j.engappai.2022.105669.

Han, Q., Gui, C., Xu, J., Lacidogna, G., 2019. A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. Constr. Build. Mater. 226, 734–742. https://doi.org/10.1016/j.conbuildmat.2019.07.315.

Hanifa, M., Sharma, U., Thapliyal, P., Singh, L., 2025. Study on CASH to NASH gel conversion in fly ash-based alkali activated aggregates via accelerated carbonation: enhancing performance and inhibiting efflorescence. J. Sustain. Cement-Based Mater. 14 (1), 178–197. https://doi.org/10.1080/21650373.2024.2431512.

Jaipuria, N., Zhang, X., Bhasin, R., Arafa, M., Chakravarty, P., Shrivastava, S., Manglani, S., Murali, V.N., 2020. Deflating dataset bias using synthetic data augmentation. Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 772–773.
Jamieson, E., McLellan, B., Van Riessen, A., Nikraz, H., 2015. Comparison of embodied energies of ordinary Portland cement with Bayer-derived geopolymer products. J. Clean. Prod. 99, 112–118. https://doi.org/10.1016/j.jclepro.2015.03.008.
Júnior, N.S.A., Neto, J.S.A., Santana, H.A., Cilla, M.S., Ribeiro, D.V., 2021. Durability and service life analysis of metakaolin-based geopolymer concretes with respect to chloride penetration using chloride migration test and corrosion potential. Constr. Build. Mater. 287, 122970. https://doi.org/10.1016/j.conbuildmat.2021.122970.
Kaplan, G., Öz, A., Bayrak, B., Aydın, A.C., 2023. The effect of geopolymer slurries with clinker aggregates and marble waste powder on embodied energy and high-temperature resistance in prepacked concrete: ANFIS-Based prediction model. J. Build. Eng. 67, 105987. https://doi.org/10.1016/j.jobe.2023.105987.
Karthik, S., Mohan, K.S.R., Murali, G., Abid, S.R., Dixit, S., 2024. Impact of various fibers on mode I, III and I/III fracture toughness in slag, fly Ash, and silica fume-based geopolymer concrete using edge-notched disc bend specimen. Theor. Appl. Fract. Mech. 134, 104751. https://doi.org/10.1016/j.tafmec.2024.104751.
Kathirvel, P., Sreekumaran, S., 2021. Sustainable development of ultra high performance concrete using geopolymer technology. J. Build. Eng. 39, 102267. https://doi.org/10.1016/j.jobe.2021.102267.
Kilgore, G., 2023. Carbon footprint of building materials. https://8billiontrees.com/carbon-offsets-credits/carbon-footprint-of-building-materials.
Kumar, R., Suman, S.K., Sharma, M., 2020. Laboratory investigation on the synthesis and mechanical characterization of fiber reinforced geopolymer concrete. Mater. Today Proc. 32, 268–273. https://doi.org/10.1016/j.matpr.2020.01.360.
Li, H., Zhang, Z., Deng, Y., Xu, F., Hu, J., Zhu, D., Yu, Q., Shi, C., 2024. Geopolymer composites for marine application: structural properties and durability. Cement Concr. Compos. 152, 105647.
Liu, Y., Zhang, Z., Shi, C., Zhu, D., Li, N., Deng, Y., 2020a. Development of ultra-high performance geopolymer concrete (UHPGC): influence of steel fiber on mechanical properties. Cement Concr. Compos. 112, 103670. https://doi.org/10.1016/j.cemconcomp.2020.103670.
Liu, Y., Shi, C., Zhang, Z., Li, N., Shi, D., 2020b. Mechanical and fracture properties of ultra-high performance geopolymer concrete: effects of steel fiber and silica fume. Cement Concr. Compos. 112, 103665. https://doi.org/10.1016/j.cemconcomp.2020.103665.
Liu, Y., Hu, X., Du, Y., Nematollahi, B., Shi, C., 2024. A review on high-temperature resistance of geopolymer concrete. J. Build. Eng., 111241 https://doi.org/10.1016/j.jobe.2024.111241.
Liu, Y., Ren, Y., Li, Q., Shi, C., 2025. Influence of steel fiber hybridization on the static mechanical performance of ultra-high performance geopolymer concrete (UHPGC). Cement Concr. Compos. 160, 106064. https://doi.org/10.1016/j.cemconcomp.2025.106064.
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. https://doi.org/10.1016/j.jclepro.2021.129665.
Mahjoubi, S., Meng, W., Bao, Y., 2022. Auto-tune learning framework for prediction of flowability, mechanical properties, and porosity of ultra-high-performance concrete (UHPC). Appl. Soft Comput. 115, 108182. https://doi.org/10.1016/j.asoc.2021.108182.
Mahjoubi, S., Barhemat, R., Meng, W., Bao, Y., 2023a. AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC). Resour. Conserv. Recycl. 189, 106741. https://doi.org/10.1016/j.resconrec.2022.106741.
Mahjoubi, S., Barhemat, R., Meng, W., Bao, Y., 2023b. AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC). Resour. Conserv. Recycl. 189, 106741. https://doi.org/10.1016/j.resconrec.2022.106741.
Mahjoubi, S., Barhemat, R., Meng, W., Bao, Y., 2025. Review of AI-assisted design of low-carbon cost-effective concrete toward carbon neutrality. Artif. Intell. Rev. 58 (8), 225. https://doi.org/10.1007/s10462-025-11182-1.
Mayhoub, O.A., Nasr, E.-S.A., Ali, Y., Kohail, M., 2021. Properties of slag based geopolymer reactive powder concrete. Ain Shams Eng. J. 12 (1), 99–105. https://doi.org/10.1016/j.asej.2020.08.013.
Mishra, J., Nanda, B., Patro, S.K., Krishna, R., 2024. A comprehensive review on compressive strength and microstructure properties of GGBS-based geopolymer binder systems. Constr. Build. Mater. 417, 135242. https://doi.org/10.1016/j.conbuildmat.2024.135242.
Moradikhou, A.B., Esparham, A., Avanaki, M.J., 2020. Physical & mechanical properties of fiber reinforced metakaolin-based geopolymer concrete. Constr. Build. Mater. 251, 118965. https://doi.org/10.1016/j.conbuildmat.2020.118965.
Nassar, A.K., Sivanandam, S., Saran, D., Sundar, S.S., Kathirvel, P., Murali, G., Dixit, S., 2025a. Durability assessment of sustainable one-part alkali activated concrete produced from agricultural and industrial waste activators under aggressive environmental conditions. Constr. Build. Mater. 486, 141823. https://doi.org/10.1016/j.conbuildmat.2025.141823.
Nassar, A.K., Kathirvel, P., Murali, G., Krishna, A., 2025b. Development and performance evaluation of novel sustainable one-part alkali-activated fibrous concrete subjected to drop weight impact loading: an experimental study. Constr. Build. Mater. 459, 139754. https://doi.org/10.1016/j.conbuildmat.2024.139754.
Navarro, R., Alcocel, E., Sánchez, I., Aguirre, M., Zornoza, E., 2025. Waste glass powder as silica source for the activator in the preparation of alkali activated SiMn slag mortar. Constr. Build. Mater. 475, 141237. https://doi.org/10.1016/j.conbuildmat.2025.141237.
Oyebisi, S., Alomayri, T., 2023. Artificial intelligence-based prediction of strengths of slag-ash-based geopolymer concrete using deep neural networks. Constr. Build. Mater. 400, 132606. https://doi.org/10.1016/j.conbuildmat.2023.132606.
Peng, Y., Unluer, C., 2022. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. Constr. Build. Mater. 316, 125785. https://doi.org/10.1016/j.conbuildmat.2021.125785.
Prasad, B.R., Eskandari, H., Reddy, B.V., 2009. Prediction of compressive strength of SCC and HPC with high volume fly ash using ANN. Constr. Build. Mater. 23 (1), 117–128. https://doi.org/10.1016/j.conbuildmat.2008.01.014.
Rajan, H.S., Kathirvel, P., 2021. Sustainable development of geopolymer binder using sodium silicate synthesized from agricultural waste. J. Clean. Prod. 286, 124959. https://doi.org/10.1016/j.jclepro.2020.124959.
Raza, A., Ahmed, B., El Ouni, M.H., Chen, W., 2024. Mechanical, durability and microstructural characterization of cost-effective polyethylene fiber-reinforced geopolymer concrete. Constr. Build. Mater. 432, 136661. https://doi.org/10.1016/j.conbuildmat.2024.136661.
Reddy, Y.B.S., Praburanganathan, S., Mishra, M., 2022. Experimental investigation on the fiber reinforced ash-based geopolymer concrete with Musa basjoo fibers. Mater. Today Proc. 65, 3700–3706. https://doi.org/10.1016/j.matpr.2022.06.297.
Rodriguez-Morales, J., Escalante-Garcia, J.I., 2024. Recycling pulverized concrete as cementitious precursor in alkaline one-part cements, with a waste glass-based sodium silicate. J. Clean. Prod. 482, 144229. https://doi.org/10.1016/j.jclepro.2024.144229.
Samadi, M., Wong, L.S., Murali, G., Lim, N.H.A.S., Abdulkadir, I., Tan, S.Q., Chan, Y.T., 2024. Sodium metasilicate-activated one-part geopolymer concrete: impact strength assessment with bottom ash substitution and fiber reinforcement. Case Stud. Constr. Mater. 21, e03794. https://doi.org/10.1016/j.cscm.2024.e03794.
Shilar, F.A., Ganachari, S.V., Patil, V.B., Khan, T.Y., Khadar, S.D.A., 2022. Molarity activity effect on mechanical and microstructure properties of geopolymer concrete: a review. Case Stud. Constr. Mater. 16, e01014. https://doi.org/10.1016/j.cscm.2022.e01014.
Tan, Y., Zhu, H., Wu, J., Chai, H., 2024. DPTVAE: data-driven prior-based tabular variational autoencoder for credit data synthesizing. Expert Syst. Appl. 241, 122071. https://doi.org/10.1016/j.eswa.2023.122071.
Tayeh, B.A., Akeed, M.H., Qaidi, S., Bakar, B.A., 2022. Influence of microsilica and polypropylene fibers on the fresh and mechanical properties of ultra-high performance geopolymer concrete (UHP-GPC). Case Stud. Constr. Mater. 17, e01367. https://doi.org/10.1016/j.cscm.2022.e01367.
UnsustaInable: concrete and cement. https://2150-vc.medium.com, 2022.
Wang, J., Shi, D., Liu, M., Yu, K., Zhao, Y., Xia, Y., 2025. Low-carbon alkali-activated materials for Cu (II)-contaminated soil stabilization: reaction kinetics, immobilization mechanisms, and leaching behavior. Case Stud. Constr. Mater., e04912 https://doi.org/10.1016/j.cscm.2025.e04912.
Xu, L., Veeramachaneni, K., 2018. Synthesizing tabular data using generative adversarial networks. ArXiv Preprint. https://doi.org/10.48550/arXiv.1811.11264.
Xu, L., Fan, D., Liu, K., Xu, W., Yu, R., 2024. A machine learning framework for intelligent development of ultra-high performance concrete (UHPC): from dataset cleaning to performance predicting. Expert Syst. Appl. 242, 122790. https://doi.org/10.1016/j.eswa.2023.122790.
Yang, F., Liu, J., Jia, H., Chen, Z., Liu, L., Zhu, J., Li, S., Lai, C., Shi, C., 2024a. Corrosion mechanism of alkali-activated slag/metakaolin materials under carbonic acid solution. Cement Concr. Compos. 154, 105779. https://doi.org/10.1016/j.cemconcomp.2024.105779.
Yang, Z., Lu, F., Zhan, X., Zhu, H., Zhang, B., Chen, Z., Zhang, H., 2024b. Mechanical properties and mesoscopic damage characteristics of basalt fibre-reinforced seawater sea-sand slag-based geopolymer concrete. J. Build. Eng. 84, 108688. https://doi.org/10.1016/j.jobe.2024.108688.
Yang, J., Fan, Y., Hang, Z., Ni, Z., Liu, H., Feng, C., 2025. Data augmentation-aided machine learning prediction of 28-day compressive strength of CNT/cement composites. Neural Comput. Appl. 37 (5), 4009–4033. https://doi.org/10.1007/s00521-024-10875-y.
Zaid, O., Martínez-García, R., Abadel, A.A., Fraile-Fernández, F.J., Alshaikh, I.M., Palencia-Coto, C., 2022. To determine the performance of metakaolin-based fiber-reinforced geopolymer concrete with recycled aggregates. Arch. Civ. Mech. Eng. 22 (3), 114. https://doi.org/10.1007/s43452-022-00436-2.
Zannerni, G.M., Fattah, K.P., Al-Tamimi, A.K., 2020. Ambient-cured geopolymer concrete with single alkali activator. Sustain. Mater. Technol. 23, e00131. https://doi.org/10.1016/j.susmat.2019.e00131.
Zhang, Z., Provis, J.L., Reid, A., Wang, H., 2014. Geopolymer foam concrete: an emerging material for sustainable construction. Constr. Build. Mater. 56, 113–127. https://doi.org/10.1016/j.conbuildmat.2014.01.081.
Zhang, Y., Deng, X., 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. https://doi.org/10.1016/j.enbuild.2022.112185.
Zhang, H., Sarker, P.K., Wang, Q., He, B., Jiang, Z., 2024. Comparative fracture properties of ambient-cured geopolymer concrete containing four different fibers in mono and hybrid combinations using digital image correlation. J. Build. Eng. 89, 109288. https://doi.org/10.1016/j.jobe.2024.109288.

Zhao, Z., Kunar, A., Birke, R., Chen, L.Y., 2021. Ctab-gan: effective table data synthesizing. Asian Conference on Machine Learning, pp. 97–112. URL: https://proceedings.mlr.press/v157/zhao21a.

Zheng, J., Qi, L., Zheng, Y., Zheng, L., 2023. Mechanical properties and compressive constitutive model of steel fiber-reinforced geopolymer concrete. J. Build. Eng. 80, 108161. https://doi.org/10.1016/j.jobe.2023.108161.

Zhou, R., Tang, Y., Li, H., Liu, Z., 2024. Predicting the compressive strength of ultra-high-performance concrete using a decision tree machine learning model enhanced by the integration of two optimization meta-heuristic algorithms. J. Eng. Appl. Sci. 71 (1), 43. https://doi.org/10.1186/s44147-023-00350-1.