Cement and Concrete Composites 165 (2026) 106363

Contents lists available at ScienceDirect

Cement and Concrete Composites

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

A transformer-based machine learning model for optimizing the design of cementitious mixtures with mine tailings as supplementary cementitious materials

Chathuranga Balasooriya Arachchilage [a], Jian Zhao [a], Nimila Dushyantha [b], Wei Victor Liu [a,*]

[a] School of Mining and Petroleum Engineering and Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada
[b] Department of Applied Earth Sciences, Faculty of Applied Sciences, Uva Wellassa University, Passara Road, Badulla, 90000, Sri Lanka

ARTICLE INFO

Keywords:
Mine tailings
Cement
Uniaxial compressive strength
Machine learning
Transformer model
Multi-objective optimization
Waste reuse

ABSTRACT

Realizing the full potential of incorporating mine tailings as supplementary cementitious materials (SCMs) to replace ordinary Portland cement (OPC) requires carefully balancing the benefits—such as cost reduction and emissions mitigation—while ensuring the mixtures achieve the required strength. Given the demonstrated effectiveness of combining machine learning (ML) with optimization algorithms in similar multi-objective optimization (MOO) problems, for the first time, this study employed a novel tabular prior data fitted network (TabPFN) model to forecast the uniaxial compressive strength (UCS) of those mix designs. The TabPFN model outperformed traditional boosting ML models, achieving an R^2 of 0.973 and a low prediction error of 2.115 MPa. Notably, its pre-trained architecture reduced computational time by 1045 s. Building on this, a MOO case study was developed using the TabPFN model to predict UCS as the first objective, alongside separate equations used as objective functions to calculate cost and total emissions. This MOO problem was tackled using the non-dominated sorting genetic algorithm-II (NSGA-II). The optimized mixture designs achieved better balances between strength, cost, and emissions than those obtained through experimental methods, validating the use of this ML-based method for mixture design. Finally, a software tool—GreenMix AI—was developed to provide integrated access to the entire framework, translating advanced research into practical application. In essence, this research supports the reuse of mine tailings as SCMs and provides a practical pathway to developing more economical and sustainable cementitious mixtures.

1. Introduction

The reuse of mine tailings as supplementary cementitious materials (SCMs) to partially replace ordinary Portland cement (OPC) in cementitious mixtures has gained increasing attention in recent years. This strategy not only helps reduce tailings waste but also offers several key advantages. First, mine tailings can enhance strength by reacting with calcium hydroxide ($Ca(OH)_2$), a by-product of OPC hydration, to form additional hydration products [1]. Second, as a waste material with no economic value, tailings help reduce overall mixture costs [2]. Third, tailings lower greenhouse gas (GHG) emissions under the zero-burden concept, which attributes impacts only to the extraction of valuable minerals, not to waste [3].

Although higher OPC replacement with mine tailings offers greater benefits, it can reduce mixture strength. This is because less OPC leads to lower formation of primary hydration products and $Ca(OH)_2$, leaving much of the tailings unreacted and weakening the matrix [4]. This limitation has been documented in several studies. For example, Kara [5] reported a 15.7 % decrease in uniaxial compressive strength (UCS) when the copper tailings replacement level was increased from 10 % to 15 %. Similarly, Ince [6] found that replacing more than 30 % of OPC with gold tailings led to a reduction in strength, despite notable improvements in cost and emissions. These findings highlight the need to strike a careful balance among trade-offs—strength retention, cost reduction, and emissions mitigation. Traditionally, researchers have relied on experimental testing to handle this multi-objective optimization (MOO) problem. However, this process is time-consuming, expensive, and impractical for evaluating all possible mixture designs [2]. As a result, researchers often limit their studies to a narrow set of mixture

* Corresponding author.
E-mail address: victor.liu@ualberta.ca (W.V. Liu).

https://doi.org/10.1016/j.cemconcomp.2025.106363
Received 8 July 2025; Received in revised form 28 August 2025; Accepted 8 October 2025
Available online 10 October 2025

Nomenclature

SCM	Supplementary cementitious material
OPC	Ordinary Portland cement
ML	Machine learning
MOO	Multi-objective optimization
TabPFN	Tabular prior data fitted network
UCS	Uniaxial compressive strength
NSGA-II	Non-dominated sorting genetic algorithm-II
SHAP	SHapley Additive exPlanations
GBR	Gradient boosting regressor
XGBR	Extreme gradient boosting regressor
LGBR	Light gradient boosting regressor
RMSE	Root mean squared error
MAE	Mean absolute error
SSA	Specific surface area
PSD	Particle size distribution
LCA	Life cycle analysis
C-S-H	Calcium silicate hydrate
C-A-S-H	Calcium aluminate silicate hydrate
$Ca(OH)_2$	Calcium hydroxide
C_3S	Tricalcium silicate
C_2S	Dicalcium silicate
CO_2 eq	Carbon dioxide equivalent
PCE	Polycarboxylate ether
TSF	Tailings storage facility
TPE	Tree-structured Parzen estimator
MOEA/D	Multi-objective evolutionary algorithm based on decomposition
RoW	Rest of the world

designs. For example, although Ince [6] examined OPC replacement ratios from 10 % to 40 %, tests were conducted only at 10 % intervals, leaving many intermediate ratios unexplored. Due to these limitations, it remains unclear whether better-performing designs could exist that offer improved trade-offs. Therefore, a more efficient and practical method is needed to identify optimal designs for these novel cementitious mixtures and to uncover their full potential.

Data-driven intelligent design approach, which combines machine learning (ML) and optimization algorithms, has emerged as a promising alternative for efficiently optimizing cementitious mixture designs. Unlike traditional experimental approaches, which are limited to testing a finite number of designs due to practicality, the intelligent design framework utilizes the patterns learned by ML models and applies them in optimization algorithms like the non-dominated sorting genetic algorithm II (NSGA-II) [7] to explore a broader space of potential mixtures. Such approaches have revealed a vast number of superior designs not tested experimentally. For example, Cao et al. [8] applied a combined ML and optimization approach, identifying 165 alternative optimal mixture designs. In this study, one selected design—with a 26 % replacement of OPC using waste slag—achieved a cost reduction of 31.64 Chinese Yuan and a decrease in GHG emissions of 31.04 kg per cubic meter, compared to the best-performing design obtained through experimental methods. These studies demonstrate that data-driven intelligent design approach offers a highly efficient and practical solution for addressing MOO problems in cementitious mixture design.

However, existing intelligent designs have primarily relied on conventional ML models, which depend on large, high-quality datasets and often require frequent tuning to maintain accuracy on new data [2]. In contrast, recent advances in transformer-based ML models offer a promising alternative, due to their ability to transfer knowledge from prior training to new datasets [9]. This leads to better predictive performance and eliminates the need for extensive retraining [9]. The tabular prior data fitted network (TabPFN) model [10], is a recent transformer-based model, pre-trained on millions of synthetic datasets, that has showed enhanced performance relative to standard ML techniques. For instance, Yu et al. [11] used TabPFN to predict urban air temperature and reported a 17.4 % reduction in prediction error compared to boosting models. Owing to their pre-trained architecture and strong generalization capabilities with limited data, transformer-based ML models have a significant potential to yield more accurate predictions and, consequently, improve optimization outcomes compared to conventional models. Despite their potential, the application of transformer-based ML models with optimization algorithms to address the MOO challenge of designing mixtures with mine tailings as SCMs remains largely unexplored.

To this end, this research employed a new data-driven intelligent design, which used a transformer-based TabPFN model to forecast the UCS of cementitious materials incorporating mine tailings as SCMs. To compare performance and computational efficiency, conventional boosting models were also constructed alongside the TabPFN model. Next, the predictions of the TabPFN model were interpreted using SHapley Additive exPlanations (SHAP) to enhance the transparency of results. Afterwards, a MOO case study was performed using three objective functions—the TabPFN model for strength prediction, along with equations for cost and emissions—to assess the feasibility of applying this intelligent approach. Finally, GreenMix AI—a user-friendly software tool—was developed to provide access to the trained ML models and optimization algorithms.

This study serves as a trailblazer in advancing sustainable cement-based materials, making four key contributions. First, it exploits a pre-trained model architecture—TabPFN—to accurately predict UCS using limited data, offering a powerful solution for data-scarce domains and eliminating the need for time-consuming model training. Second, this work breaks new ground by utilizing intelligent approaches to uncover underlying relationships between critical factors—such as tailings characteristics, mix proportions, and curing conditions—and UCS of mixtures using tailings as SCMs. This critical information empowers engineers with deeper insights for performance-driven mixture design. Third, this work is a pioneer in applying MOO to optimize these mixtures, delivering designs that are not only cost-effective and sustainable but also difficult to achieve through experimental trials alone. Fourth, the newly developed software tool—GreenMix AI—bridges the gap between research and practice by making data-driven intelligent design methods easily accessible for practical, real-world applications. In summary, this research advances the reuse of mine tailings and supports the development of an eco-friendly and economically resilient building sector.

2. Methodology

2.1. Overview of the research framework

Fig. 1 provides an overview of the research framework adopted in this study. Initially, experimental data were gathered from published studies as discrete data points. The compiled dataset was complete, with no missing values, and included detailed information on mixture designs and materials through multiple input features, along with the corresponding UCS values as the output. The dataset was subsequently split at random into two parts: 80 % for training and 20 % for testing. Using the training dataset, three boosting algorithms—gradient boosting regressor (GBR), extreme gradient boosting regressor (XGBR), and light gradient boosting regressor (LGBR)— were employed to build predictive models. These models were further optimized using the Optuna hyperparameter tuning library (version 4.2). The transformer-based model, TabPFN, did

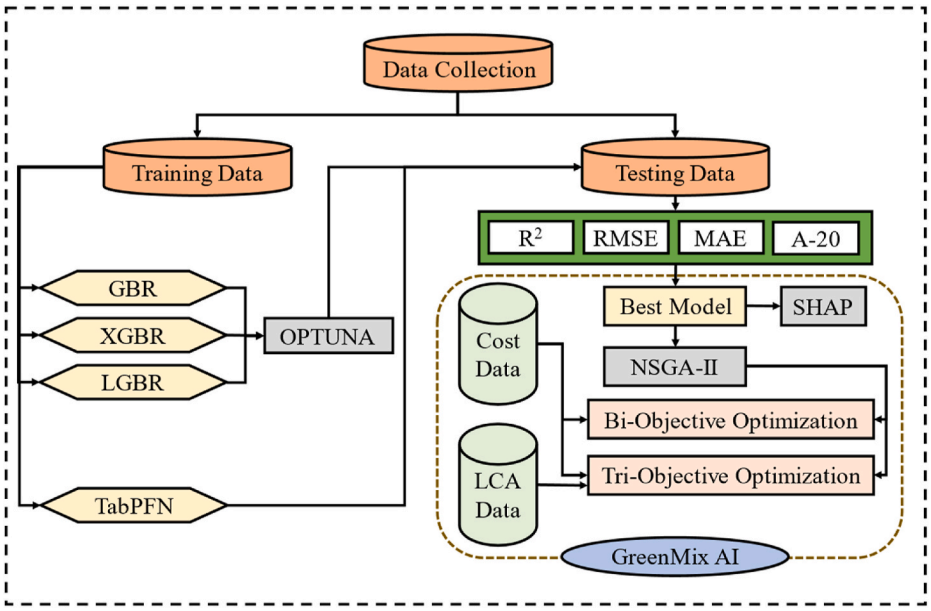

Fig. 1. Overview of the research framework.

not require model training due to its pre-trained architecture [10]. Each model was then fitted using the training data to create prediction models. Their efficacy was assessed on the test data portion using four metrics: root mean squared error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), and the A-20 score. Subsequently, the best-performing model was interpreted using SHAP analysis and incorporated into MOO case studies using the NSGA-II algorithm. The cost and life cycle analysis (LCA) data of constituent materials used in the MOO study were extracted from the literature and OpenLCA 2.4.1 software analysis (integrated with Ecoinvent 3.11 database), respectively. Finally, GreenMix AI, a user-friendly software was developed providing easy access to the models and algorithms. The entire methodology was implemented in the Google Colab Jupyter notebook environment, using python 3.11.13 and scikit-learn 1.6.1, with access to an NVIDIA T4 Tensor Core GPU.

Table 1
Research articles selected for data extraction.

	Tailings category	Reference		Tailings category	Reference
1	Copper	[14]	14	Coal	[15]
2	Copper	[16]	15	Iron	[17]
3	Copper	[18]	16	Zinc	[19]
4	Copper	[20]	17	Marble	[21]
5	Zinc, Copper, Gold	[22]	18	Gold	[23]
6	Iron	[24]	19	Iron	[25]
7	Iron	[26]	20	Iron	[25]
8	Tungsten	[27]	21	Lead, Zinc	[28]
9	Copper	[5]	22	Phosphorous	[29]
10	Molybdenum	[30]	23	Iron	[31]
11	Phosphate	[32]	24	Molybdenum	[33]
12	Gold	[34]	25	Gold	[4]
13	Iron	[35]			

2.2. Dataset

2.2.1. Dataset construction

An initial review of the literature was systematically performed to identify research articles reporting UCS results for cementitious materials utilizing tailings as SCMs. The search was performed on reputable research indexing platforms, comprising Web of Science, Compendex and Scopus, using terms like cement, mine tailings, mine waste, supplementary cementitious materials and uniaxial compressive strength. A database of experimental results was compiled from the selected articles, containing detailed information on various inputs and their associated UCS values as the output. It is important to note that some articles were excluded due to missing information, such as material properties or chemical compositions. Additionally, this study focused solely on mixtures containing mine tailings, and articles involving other SCMs (e.g., fly ash, slag) in combination with mine tailings were not considered. This decision was made because such data points are limited in number, and their inclusion could introduce bias into the dataset due to their underrepresentation.

Ultimately, a total of 399 unique data points were collected from 25 research articles published between 1950 and 2025. Table 1 summarizes the research articles referenced for data extraction in this study, alongside the specific category of mine tailings reported in each article. This dataset represents a significant advancement over those reported in previous studies, which included only 24 data points (later combined with a concrete dataset containing no mine tailings) [12] and 148 data points [13], respectively. Moreover, these datasets were limited to a single type of mine tailings—copper—whereas the dataset compiled in this study incorporates data from research on 11 different types of tailings. This increased diversity enhances the versatility and generalizability of the models developed using this dataset.

2.2.2. Description and statistical summary of the dataset

The collected dataset consists of 14 input features, with UCS as the sole output feature, serving as the target variable for model training. The input features represent a combination of chemical and physical properties of both OPC and mine tailings, material proportions, and the curing time of the samples. Specific surface areas (SSAs) of OPC and mine tailings were included as input features in place of conventional particle size distribution (PSD) parameters because SSA is a more reliable indicator of material reactivity, which reflects the available surface area for hydration reactions [28]. Since PSD parameters are strongly correlated with SSA—smaller particle sizes typically result in higher SSA and increased reactivity—the exclusion of PSD metrics, such as median particle size, helps reduce redundancy and prevents unnecessary complexity in the models [28]. In addition, CaO, SiO_2, and Al_2O_3 percentages of OPC were selected to represent the type of OPC used across different studies. While the OPC type could alternatively be included as a categorical feature (e.g., based on name or grade), oxide composition

data were preferred, as continuous variables tend to perform better in ML models [36]. The other possible option was to consider the standard strength of cement paste (e.g., UCS at 28 days); however, oxide composition data were consistently reported across all the selected studies, whereas information on standard strength was missing in some cases. Similarly, oxide compositions were used as input features to represent the type of mine tailings, but were limited to Al_2O_3 and SiO_2, as their ions are the primary contributors to forming supplementary hydration materials—calcium alumino silicate hydrate (C-A-S-H) gels—which can improve the UCS of the mixtures [37]. In addition, this decision helped reduce the complexity of the dataset by focusing only on the most relevant input features, which is especially important when working with a limited dataset. In addition, to represent the material proportions used in each study, five input features were included: OPC, mine tailings, fine aggregates, coarse aggregates, and superplasticizer, all expressed as mass per unit volume. It should be noted that the dataset contains mixtures of both mortar and concrete. For mortar mixtures, which do not contain coarse aggregates, a value of zero was assigned to the coarse aggregate field. Similarly, for mixtures that did not include any superplasticizer, a value of zero was used for that feature. Besides the chosen inputs, other parameters like curing temperature and humidity were not included in this study, as their values remained consistent (i.e., ambient temperature and relative humidity above 95 %) across all the studies and did not contribute any variability to the dataset.

Table 2 provides summary statistics for both the input variables and UCS. Additionally, Fig. 2 displays data spread histograms for each feature. According to the statistical summary and visualizations, the dataset shows wide value ranges and relatively high standard deviations across input features, which supports the development of more generalized and robust prediction models.

2.3. Machine learning models

2.3.1. Gradient boosting regressor models

The gradient boosting regressor (GBR) models are based on a sequential boosting approach, which is more advanced than traditional decision tree or random forest architectures [38]. Although GBR relies on multiple trees, it contrasts with the random forest method, which simply averages predictions from many trees [39]. Instead, GBR improves prediction performance iteratively by targeting the errors from the preceding model at each step [40]. This iterative process allows GBR to gradually reduce overall prediction error and build a more accurate and robust model. In this study, in addition to the original GBR model, two other variations, namely the XGBR [41] and LGBR [42] were also used to construct prediction models.

2.3.2. Tabular prior data fitted network regressor

TabPFN is a transformer-based ML model specifically designed for tabular data. Unlike traditional models such as GBR, XGBR, and LGBR, TabPFN does not require model training, hyperparameter tuning, or weight adjustments (typically associated with deep learning) [10]. The architecture of TabPFN consists of three main stages: (1) synthetic data generation, (2) pre-training, (3) real world prediction (i.e., final application) [10]. Instead of relying on limited real-world data, TabPFN is trained on a very large number of synthetic datasets. These datasets are created to resemble real tabular problems, with many possible relationships between input features (x) and target values (y) [10]. The process begins by choosing basic parameters such as dataset size, number of features, and complexity. Random input values are then generated assuming normal, uniform, or mixed probability distributions [10]. Next, these inputs are passed through a causal graph that applies different transformations, such as small neural networks or decision tree rules, while Gaussian noise is added to simulate uncertainty [10]. Finally, feature and target node values are extracted, yielding samples of input–output pairs. By repeating this process millions of times, more than 100 million synthetic datasets are created for training. This synthetic approach avoids issues such as data scarcity, privacy concerns, and contamination from using real-world data during training [10].

In the pre-training step, each dataset can be seen as being generated by a different **hypothesis** (φ) about how inputs and outputs are related, where $\varphi \in \varnothing$ (set of hypotheses). During training, part of each synthetic dataset is treated as training data (x feature values and corresponding y values), while the remaining part is used as test data (x feature values with held out y values). TabPFN is trained to predict the test values given the training examples as context [43]. Its transformer architecture allows TabPFN to capture relationships both between features and between data samples through the two-way attention mechanism [43]. To do this, TabPFN learns to approximate the posterior predictive distribution (PPD) (Equation (1)), which represents the probabilities of possible outcomes for test inputs after considering many different hypotheses and how well they explain the training data [43].

$$\mathrm{p}(y_{test}|x_{test}, D_{train}) \propto \int p(y|x,\varphi)p(D|\varphi)p(\varphi)d\varphi \tag{1}$$

where $\mathrm{p}(y_{test}|x_{test}, \mathrm{D})$ is the PPD of test data, $p(y|x,\varphi)$ is the probability distribution of output y under the hypothesis φ for x, $p(D|\varphi)$ is the probability distribution of given data under the selected hypothesis, and $p(\varphi)$ is the prior probability of the selected hypothesis before seeing the training data. TabPFN conducts the prior fitting of training data by sampling available hypotheses and synthetic data. This process is repeated until its parameters (θ) are optimized for the testing data, conditioned using the training data [43]. This is achieved through calculating the cross-entropy loss between model prediction and actual target values (held out). This loss is given by Equation (2).

Table 2
Summary statistics of the input features and the output.

Input Feature	Minimum	Q1 (25 %)	Q2 (50 %)	Q3 (75 %)	Maximum	Mean	Standard Deviation
SSA of OPC (cm^2/kg)	2870	3231	3500	3580	5420	3421	268.56
CaO % in OPC	17.6	20.22	21.3	21.88	28.16	21.65	2.65
SiO_2 % in OPC	3.3	4.41	5.07	5.62	8.07	5.16	1.23
Al_2O_3 % in OPC	54.8	61.99	63	64.49	67.22	62.17	3.3
SSA of Mine Tailings (cm^2/kg)	670	1814	2840	5776	12666	3793	2742.7
SiO_2 % in Mine Tailings	1.12	32.2	52.56	63.3	92.9	47.98	26.79
Al_2O_3 % in Mine Tailings	0	2.16	6.85	13.68	25.1	8.13	6.97
OPC (kg/m^3)	164	320.5	410.1	527.3	742.2	416.9	124.63
Mine Tailings (kg/m^3)	8.9	53.3	80	175.8	466.6	114.4	87.8
Water (kg/m^3)	103.5	160	254.6	293	390.6	243.1	78
Fine Aggregates (kg/m^3)	585.9	785	1757.8	1757.8	2343.8	1380.1	528.5
Coarse Aggregates (kg/m^3)	0	0	0	857	1503.8	346.5	503.1
Superplasticizer (kg/m^3)	0	0	0	3.13	6.7	1.88	2.5
Curing Time (days)	3	7	28	42	90	32.1	31.26
UCS (MPa)	3	24	33.9	42.85	59.6	33.5	12.4

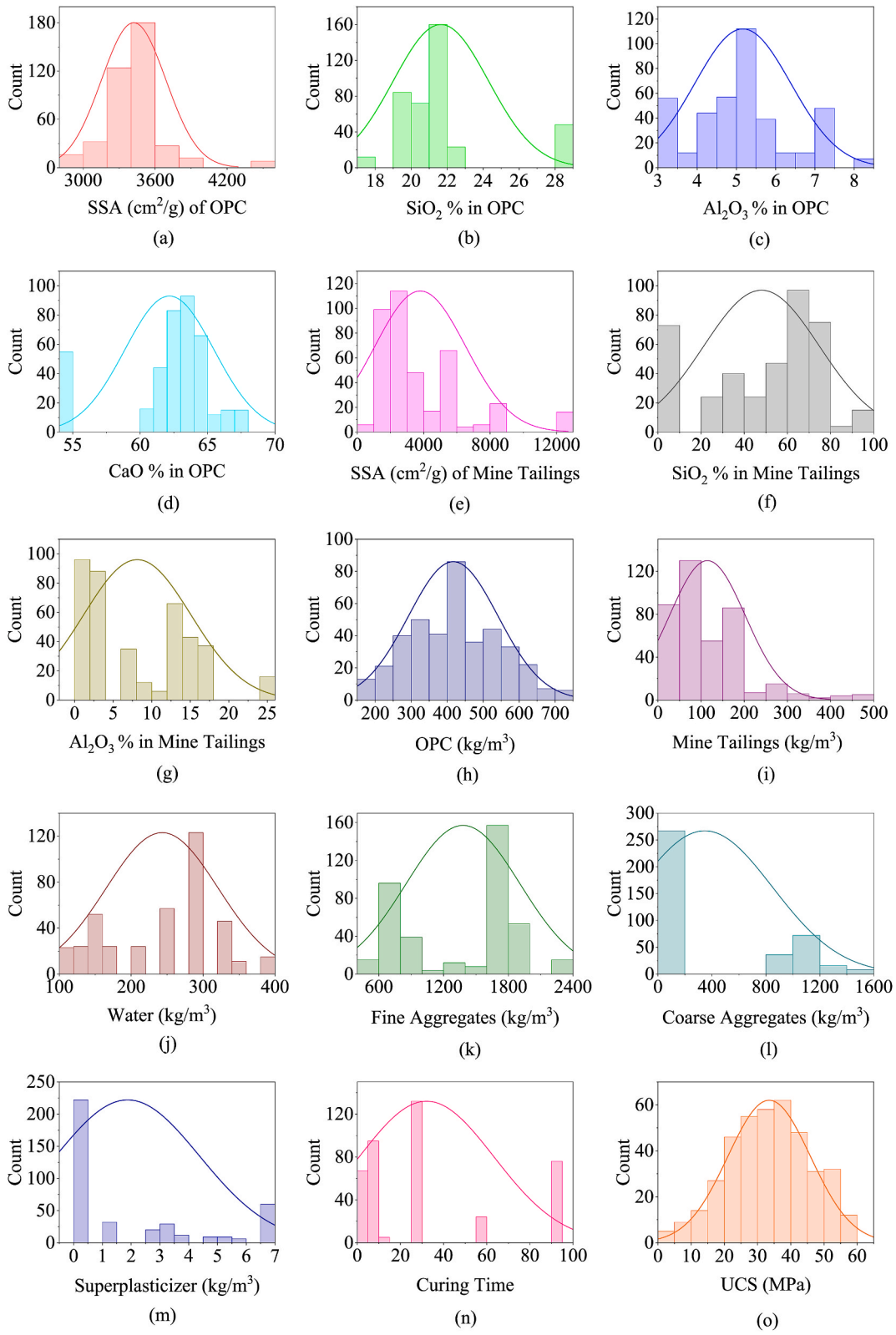

Fig. 2. Data distribution of input features and the UCS.

$$L_{PFN} = E_{\{(x_{test}, y_{test}) \cup D_{train}\} \sim p(D)} [-\log q_\theta (y_{test} | x_{test}, D_{train})] \quad (2)$$

where $\{(x_{test}, y_{test}) \cup D_{train}\} \sim p(D)$ is the selected dataset comprising testing and training data, $-\log q_\theta (y_{test} | x_{test}, D_{train})$ is the cross entropy loss for true y_{test} after comparing it with the predicted y_{test}.

Because TabPFN is pretrained on millions of synthetic datasets, it can be applied directly to new tasks without additional training. A user only needs to provide the training data for the task at hand, which the model treats as context. Using this context, TabPFN predicts the labels of unseen inputs in a single forward pass [10]. On OpenML benchmark datasets, TabPFN has shown superior performance compared to

conventional boosting models, achieving speedups of up to 230 × while maintaining or even improving predictive accuracy [43].

2.4. Optuna framework

Optuna is an open-source optimization framework designed specifically for hyperparameter tuning in ML models. It formulates the tuning process as an optimization task, aiming to either maximize or minimize a defined objective function [44]. In this context, the hyperparameters serve as inputs to the function, and the output is the validation score—calculated using a chosen evaluation metric for a given model [44]. Unlike traditional options such as grid or random search, which rely on predefined or randomly chosen hyperparameter combinations, Optuna uses a Bayesian optimization approach [45]. This allows it to dynamically select new hyperparameter sets based on the results of previous trials, enabling it to explore more encouraging areas of the search domain and significantly improve efficiency [44]. Additionally, Optuna incorporates a pruning mechanism that monitors the intermediate performance of trials. If a trial is unlikely to yield good results, it can be terminated early based on a predefined threshold, saving computational time and resources [44]. In this study, when using Optuna, the tree-structured Parzen estimator (TPE) [46] was used as the sampling algorithm, while the Hyperband pruner [44] was employed to terminate underperforming trials. The optimal hyperparameter combination was selected based on the minimum average RMSE obtained through repeated five-fold cross-validation, performed three times for different splits of the training data. This strategy is particularly important for limited datasets to ensure sufficient evaluation coverage and reduce the influence of outliers [47].

2.5. Performance evaluation metrics

This study used four well-established performance evaluation metrics for regression tasks: R^2, MAE, RMSE, and A-20. Among these, MAE and RMSE are error-based metrics, while R^2 and A-20 are unitless scores [48–52]. A model is considered to perform well when its MAE and RMSE values are close to zero, and its R^2 and A-20 scores are close to one [53–55]. Their respective formulas are shown in Equations (3)–(6).

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \tag{3}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} \left| y_i - \hat{y}_i \right| \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{5}$$

$$A-20\ index = \frac{N20}{N} \tag{6}$$

where N is the number of samples, $N20$ is the number of samples where the predicted to actual value ratio is inside the range of 0.8–1.20, y_i is the actual UCS, $\hat{y}_i$ is the predicted UCS, and $\bar{y}_i$ is the mean of the UCS values.

2.6. SHAP analysis

Although non-linear ML models perform well on tabular regression tasks, they lack interpretability [56]. This issue is particularly relevant for the models used in this study—boosting trees and transformer-based architectures—as their complex internal structures make it difficult to understand how predictions are made. However, the SHAP analysis, as an effective tool for interpreting ML models, can be used to unearth the importance of each input feature as well as their impact on model predictions [57]. SHAP analysis adapts the game theory concept by treating the prediction process as a collaborative game, where each input feature is treated as player and the model's output represents the game's result [58]. The method quantifies the impact of each input by examining the change in predictions when the feature is included versus when it is excluded, across all possible combinations of input features [58]. By aggregating these differences, SHAP assigns a value—known as the SHAP value—that reflects the individual influence of each feature on the model's prediction [58].

2.7. Multi-objective optimization

Although the advantages of using mine tailings as SCMs in cementitious mixtures are well recognized, it is essential to explore the trade-offs involved—particularly in balancing the UCS, cost, and emissions—to identify optimal mixture designs. To address this, the current study formulates two MOO problems, denoted as: $F_2(q)$ and $F_3(q)$, which are explained comprehensively in section 2.7.1.

2.7.1. Definitions of the objective functions

This study examined both bi-objective and tri-objective optimizations, as separate problems. The bi-objective optimization (i.e., improving UCS and reducing the cost) problem is formulated as shown in Equation (7).

$$\min F_2(q) = [\, -f_1(c, q, 28), f_2(q) \,] \tag{7}$$

For tri-objective optimization, Equation (7) can be extended to include emissions reduction as an additional objective. This is shown in Equation (8).

$$\min F_3(q) = [\, -f_1(c, q, 28), f_2(q), f_3(q) \,] \tag{8}$$

where $f_1(c, q, 28)$ represents the optimal ML model used to forecast the UCS at 28 days, and $f_2(q)$ and $f_3(q)$ denote the linear functions used to calculate the total material cost and total emissions (as carbon dioxide equivalent (CO_2 eq.)) of the mixtures, respectively. The UCS prediction model is expressed as a negative function to align with the minimization framework of the MOO problem. Here, c and q are arrays of input features used in the ML model, where c represents the features held constant and q represents the features that are varied during the optimization process. These arrays are further expressed using Equations (9) and (10).

$$c = \left(OPC_{SSA}, OPC_{CaO}, OPC_{SiO_2}, OPC_{Al_2O_3}, MT_{SSA}, MT_{SiO_2}, MT_{Al_2O_3}\right) \tag{9}$$

$$q = (Q_{OPC}, Q_{MT}, Q_W, Q_{FA}, Q_{CA}, Q_{SP}) \tag{10}$$

In this formulation, OPC_{SSA} *and* MT_{SSA} represent the specific surface areas of OPC and mine tailings, respectively, whereas the terms $OPC_{CaO}, OPC_{SiO_2}, OPC_{Al_2O_3}, MT_{SiO_2}, MT_{Al_2O_3}$ denote the respective oxide compositions of OPC and mine tailings. Moreover, the terms Q_{OPC}, $Q_{MT}, Q_W, Q_{FA}, Q_{CA}, Q_{SP}$ represent the masses (in kg) of OPC, mine tailings, water, fine aggregates, coarse aggregates, and superplasticizer, respectively, used per one cubic meter of the cementitious mixture. It is important to highlight that the features held constant (array c) do not possess practical significance when treated as optimization variables. For instance, although the optimization algorithm could theoretically identify an optimal chemical composition of tailings, such a result lacks practical relevance since tailoring tailings with a specific composition is infeasible. Therefore, only the material proportioning variables in array q are used in the optimization process. The values in array c are kept constant for a selected study based on the specific types of OPC and mine tailings used. However, if the type of OPC or tailings changes in a different study, the corresponding c values must also be updated accordingly.

The functions for total material cost $f_2(q)$ and emissions (CO_2 eq.) $f_3(q)$ are defined using Equation (11) and Equation (12), respectively.

$$f_2(q) = C_{OPC}Q_{OPC} + C_{MT}Q_{MT} + C_WQ_W + C_{FA}Q_{FA} + C_{CA}Q_{CA} + C_{SP}Q_{SP} \quad (11)$$

$$f_3(q) = E_{OPC}Q_{OPC} + E_{MT}Q_{MT} + E_WQ_W + E_{FA}Q_{FA} + E_{CA}Q_{CA} + E_{SP}Q_{SP} \quad (12)$$

where set of coefficients $C_{OPC}, C_{MT}, C_w, C_{FA}, C_{CA}, C_{SP}$ and $E_{OPC}, E_{MT}, E_w, E_{FA}, E_{CA}, E_{SP}$ are the unit material costs (US$/per kg) and unit emissions (kg CO_2 eq./per kg) of OPC, mine tailings, water, fine aggregates, coarse aggregates, and superplasticizer, respectively. It is important to select appropriate coefficient values based on the geographical and economic context of a selected study. This flexibility ensures that the proposed MOO framework remains applicable and adaptable across different regions and case studies.

2.7.2. Constraints

Constraints play a crucial role formulating a MOO study by ensuring the generated results are both practical and feasible for real-world implementation [2]. This study considered both range and ratio constraints, which are represented mathematically according to Equation (13).

$$C_{min} \leq C_i \leq C_{max} \quad (13)$$

where C_i denotes a selected input feature or ratio, and C_{min} and C_{max} represent its lower and upper bounds, respectively. Range constraints were applied to all variables in array q thereby restricting the optimization process to values within experimentally validated limits. Additionally, ratio constraints—water-to-cement ratio, fine aggregate-to-cement ratio, and coarse aggregate-to-cement ratio—were also imposed. By applying these constraints, the optimization respects established relationships between material proportions and performance characteristics, thereby helping ensure that the predicted mixtures are both constructible and reliable. In this study, a case study was performed to explore the effectiveness of intelligent design in solving the formulated MOO problem. Comprehensive details on the selection of values for arrays c and q, along with the calculations of associated coefficients and constraints, are provided in Appendix A.

2.7.3. NSGA-II algorithm

The NSGA-II algorithm utilizes a combined approach that integrates two key concepts: non-dominated sorting and genetic operations. In the context of MOO, a solution is said to dominate another if it performs at least equally across all objectives and outperforms in one or more of them [59]. Based on this principle, the algorithm identifies a pareto front composed of solutions that are non-dominated with respect to each other, yet dominate all other solutions in the population [7]. Subsequent fronts are formed by removing the previous front and repeating the dominance evaluation on the remaining solutions [59]. To evolve the population toward better solutions, NSGA-II applies genetic mechanisms, including selection, crossover, and mutation [59]. Crossover uses predecessor solutions (parents) and join them to find new solutions (children), exploring the search domain. Mutation modifies these solutions marginally to further exploit better performing regions [7]. This process is repeated iteratively, and after each generation, both parent and children solutions are considered together and sorted using non-dominated sorting and a crowding distance metric [7]. The best individuals are then selected for the next generation. This cycle repeats until a pre-determined number of steps or a specific convergence scenario is attained.

3. Results and discussion

3.1. Calibration and performance evaluation of ML models

Unlike the TabPFN model, the conventional tree-based boosting models selected in this study require proper calibration of hyperparameters to achieve optimal performance [57]. To accomplish this, the Optuna framework was employed to tune key hyperparameters for each boosting model. Table 3 the optimal hyperparameter combinations that minimizes the average RMSE of model predictions when using Optuna.

In the next step, the base and tuned boosting models, along with the TabPFN model, were evaluated on the whole training dataset as well as on the unseen testing dataset to determine the most successful ML model. Table 4 lists the performance of base models on both training and testing datasets, whereas Table 5 summaries the results of tuned models and the TabPFN with their respective computational times. In these base models, default hyperparameters assigned by scikit-learn python package are used, which are different from the values listed in Table 3 for tuned models.

Based on the results, the base GBR model performs well on the training data yet shows a significant drop in prediction accuracy on the testing portion, as indicated by a 66.3 % increase in RMSE and a 61.1 % increase in MAE. Similarly, the R^2 and A-20 scores decrease by 3.5 % and 6.8 %, respectively, further indicating reduced performance on unseen data. This performance drop can be attributed to the use of default hyperparameters, which are not tuned for the specific particularities of the dataset [57]. As a result, the GBR model likely overfits the training data and lacks generalization. XGBR and LGBR base models follow the same undesirable pattern, highlighting the importance of calibrating the ML models.

Compared with the performance of base models on testing data, the tuned models achieved lower prediction errors, despite some overfitting still being present. For example, the tuned GBR model reduced RMSE and MAE by 30 % and 31.1 %, respectively, while R^2 and A-20 scores improved by 3.1 % and 5.6 %, respectively. Similarly, the other tuned boosting models—XGBR and LGBR—also outperformed their base counterparts by achieving lower errors and improved R^2 and A-20 scores on the testing data. Accordingly, while tuning did not fully eliminate overfitting (i.e., due to the limitation of data), it led to boosting models that generalize more effectively to unseen data. Among the tuned boosting models, the GBR model achieved the best overall performance, attaining the smallest RMSE (2.166 MPa) and MAE (1.573 MPa), as well as the highest R^2 (0.972) and A-20 (0.963) scores on the testing data. The TabPFN model slightly outperformed even the best tuned boosting model (GBR) in terms of prediction accuracy, achieving a lower RMSE

Table 3
Optimal hyperparameters of boosting models.

ML Model	Hyperparameter	Optimal Value
GBR	Learning rate	0.1675
	Max depth	3
	Min samples split	4
	Min samples leaf	6
	Subsample	0.8405
	N estimators	709
XGBR	Learning rate	0.4775
	Max depth	4
	Min child weight	3.3387
	Subsample	0.6647
	Colsample by tree	0.7669
	N estimators	956
	Reg alpha	0.6095
	Reg lambda	172
LGBR	Learning rate	0.1701
	Max depth	5
	Num leaves	29
	Min data leaf	17
	Feature fraction	0.8710
	Bagging fraction	0.7440
	Bagging frequency	2
	N estimators	902
	Reg alpha	0.8860
	Reg lambda	16.8168

Table 4
Performance of base ML models.

Model	Dataset	RMSE (MPa)	R^2	MAE (MPa)	A-20
GBR	Training	1.861	0.977	1.418	0.978
	Testing	3.095	0.943	2.285	0.912
XGBR	Training	0.155	1.000	0.080	1.000
	Testing	2.829	0.953	2.049	0.925
LGBR	Training	1.682	0.981	1.256	0.987
	Testing	2.939	0.949	2.126	0.938

Table 5
Performance of tuned ML models.

Model	Computational time (s)	Dataset	RMSE (MPa)	R^2	MAE (MPa)	A-20
GBR	1052	Training	0.642	0.997	0.486	0.997
Tuned		Testing	2.166	0.972	1.573	0.963
XGBR	600	Training	0.812	0.996	0.578	0.997
Tuned		Testing	2.176	0.972	1.652	0.950
LGBR	176	Training	0.914	0.994	0.672	0.997
Tuned		Testing	2.288	0.969	1.689	0.963
TabPFN	7	Training	0.917	0.994	0.661	1.000
		Testing	2.115	0.973	1.440	0.963

(2.115 MPa) and MAE (1.440 MPa), along with a marginally higher R^2 (0.973) and the same A-20 (0.963). The superior performance of TabPFN can be attributed to its architecture, which leverages Bayesian inference over multiple hypotheses, weighting each by how well it fits the training data [10]. In addition, unlike conventional models trained only on task-specific data, TabPFN is pretrained on millions of synthetic datasets, giving it prior knowledge for similar prediction tasks [11]. Moreover, its two-way attention mechanism captures complex feature interactions that tree-based models often miss, as these models combine features sequentially rather than simultaneously [43]. These characteristics enable TabPFN to typically outperform conventional boosting models, especially when data are limited. Although TabPFN outperforms other models, its performance on testing data still falls short of that on the training data. This is because the model has access to the true labels in the training set, which allows it to be well-conditioned for those points. However, due to the data limitations, TabPFN may not fully capture the complete context, leading to slightly lower performance on unseen data. In addition to its superior predictive performance, the TabPFN model did not require training or hyperparameter tuning, taking only 7 s for internal data preprocessing. This represents a substantial improvement in computational efficiency compared to the 1052 s required by the tuned GBR model for hyperparameter optimization and training. The speed advantage of TabPFN is due to its architecture as a pre-trained transformer model, which eliminates model training and tuning requirements on new datasets [10].

Fig. 3 shows the agreement between the true and estimated values for both base ((a), (c), (e)) and tuned boosting models ((b), (d), (f)) as well as the TabPFN model (g). The visualizations further confirm the impact of tuning, as fewer testing data points (i.e., red points) fall outside the ±20 % error bounds in the tuned models compared to their base counterparts. Furthermore, in the TabPFN plot, most testing data are closely located with the X = Y line, denoting superior predictive performance over all the tuned boosting models. Similar trends are evident in Fig. 4, which presents the regression error characteristic (REC) curves. These plots show the percentage of predictions (i.e., accuracy in y-axis) falling within a given error tolerance (i.e., residual error in x-axis). The ideal model line, representing perfect prediction accuracy, reaches 100 % at zero error. In Fig. 4 (a), the tuned boosting models outperform TabPFN on the training data, as their curves lie closer to the ideal model line, indicating a higher proportion of accurate predictions at lower error thresholds. Conversely, in Fig. 4 (b), TabPFN demonstrates the best performance on the testing data, with a steeply rising curve (i.e., closest to the ideal model), suggesting that most of its predictions fall within a small error range. These results further confirm the TabPFN model's strong generalization capability than the conventional boosting models, producing more accurate predictions on unseen data. Given its superior predictive performance and significantly lower computational time, the TabPFN model was selected as the most suitable ML model for estimating the UCS of cementitious mixtures examined in this study.

3.2. Explanation of the TabPFN model

The SHAP analysis was employed to explain the best ML model, TabPFN, to ensure its results are transparent and trustworthy. Fig. 5 depicts the order of input features based on their significance for model predictions. The x-axis of the plot shows the mean absolute SHAP values, while the y-axis represents input features used in the ML model. Based on the results, curing time stands out as the most important feature for model's output, UCS of the mixtures. This is likely due to the strong dependence of UCS on the quantity of hydration products, which varies significantly with the duration of the curing period [60]. Following curing time, the proportions of water and OPC in the mixture ranked next in importance, with values of 5.22 and 5.20, respectively. This can be explained by the critical role both components play in the hydration process. OPC serves as the primary binding agent that holds the other materials—such as aggregates and additives—together, while water activates the chemical reaction with OPC [61]. The CaO content in OPC ranks next in importance, highlighting the influence of OPC composition on the strength development of cementitious mixtures. CaO is directly associated with the key clinker phases—tricalcium silicate (C_3S) and dicalcium silicate (C_2S)—which govern the formation of calcium silicate hydrate (C-S-H) gels, the primary hydration product responsible for strength [62]. This highlights the TabPFN model's capability to differentiate between OPC varieties used across studies and to adjust its UCS predictions accordingly. Another key component, the amount of fine aggregates, received the next highest importance score of 1.25. This result is consistent with their known role in reducing porosity by filling voids within the mixture, which in turn contributes to improved UCS [62]. Following fine aggregates, the amount of mine tailings was ranked next in importance. In the referenced studies, mine tailings were primarily utilized as a partial substitution for OPC, serving as a SCM. When reactive, mine tailings can interact with $Ca(OH)_2$ to form additional hydration products, such as calcium aluminate silicate hydrate (C–A–S–H) gels [16]. These products can contribute to improving the UCS of the mixtures [16]. However, the contribution of OPC to hydration and strength development is significantly greater, which explains why the importance of the mine tailings feature is ranked lower than that of OPC [16]. The next most important features are associated with the chemical and physical properties of mine tailings. The SiO_2 and Al_2O_3 contents directly influence the formation of C-A-S-H gels, while the SSA of the tailings is linked to their reactivity [63]. The higher importance ranking of SiO_2 compared to Al_2O_3 is likely due to the predominantly silicate nature (i.e., higher percentage of SiO_2) of most mine tailings [63].

The remaining features considered by the TabPFN model—namely the SiO_2 and Al_2O_3 contents of OPC, the SSA of OPC, superplasticizer content, and coarse aggregates—received the lowest importance scores, which are collectively represented in the plot. In the case of OPC, the SiO_2 and Al_2O_3 contents are lower compared to CaO and exhibit limited variability across different samples. Similarly, OPC is a standardized material, and its SSA tends to remain relatively consistent, contributing to the lower importance of these features. In contrast, the SSA of mine tailings shows considerable variation, especially in studies where mechanical activation was applied [30,19], likely explaining its higher importance ranking relative to OPC SSA. The contribution of superplasticizer to UCS is also minimal, which aligns with findings from previous studies [57]. Lastly, the coarse aggregate content was ranked

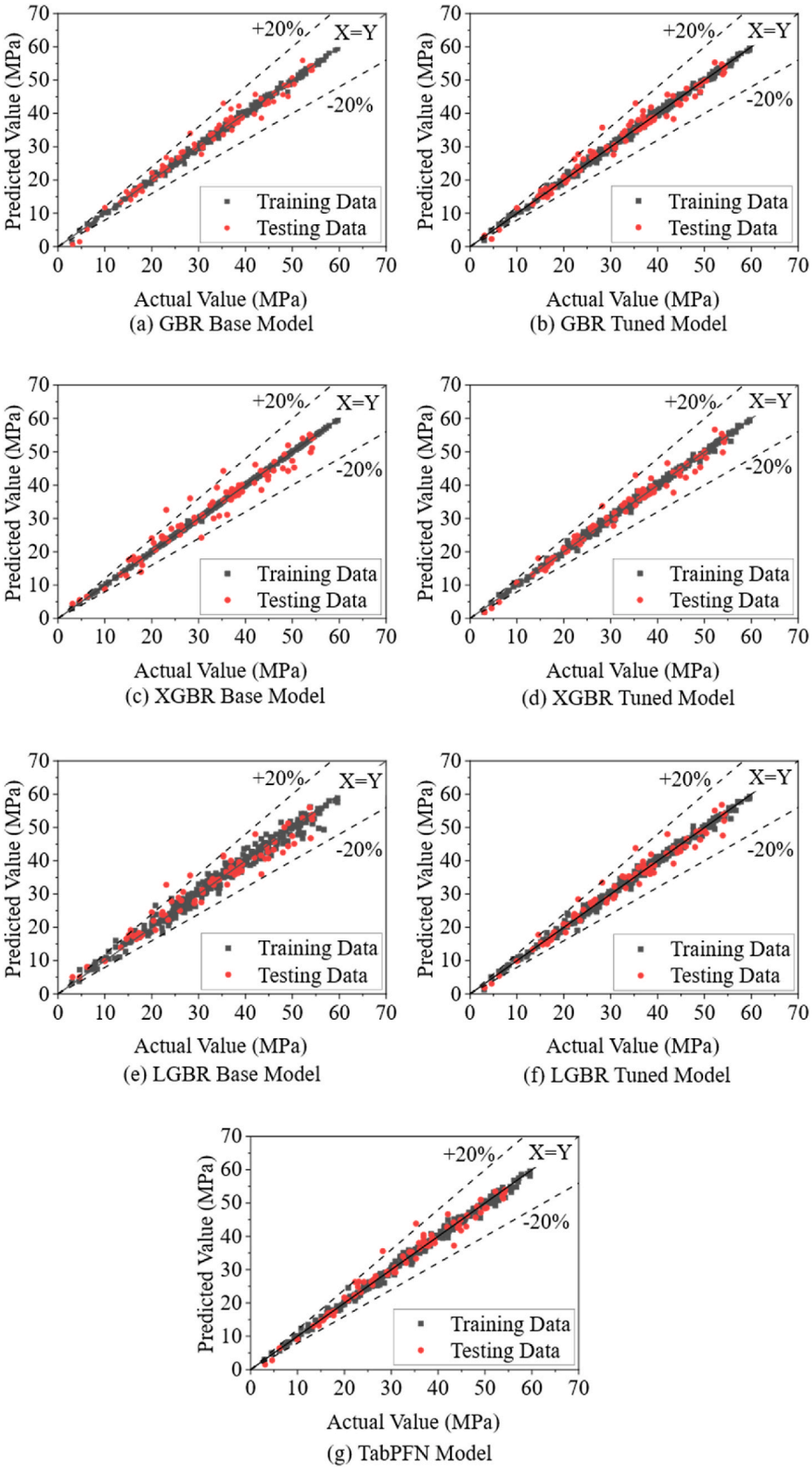

Fig. 3. Actual vs predicted value plots of different models.

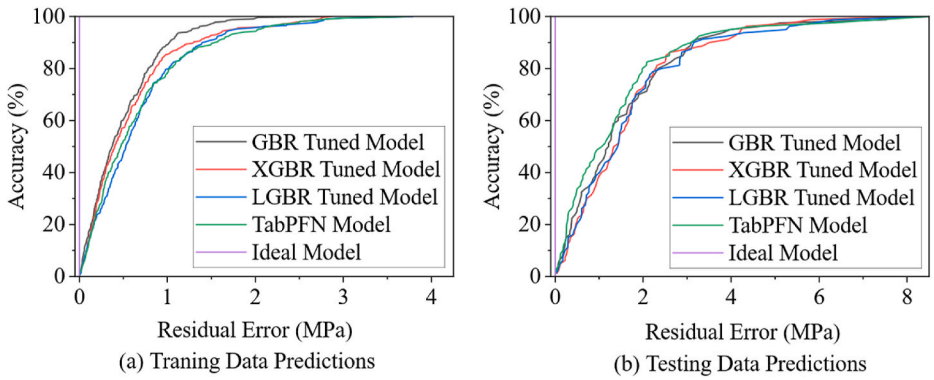

Fig. 4. Regression error characteristic curves of tuned booting models and TabPFN model.

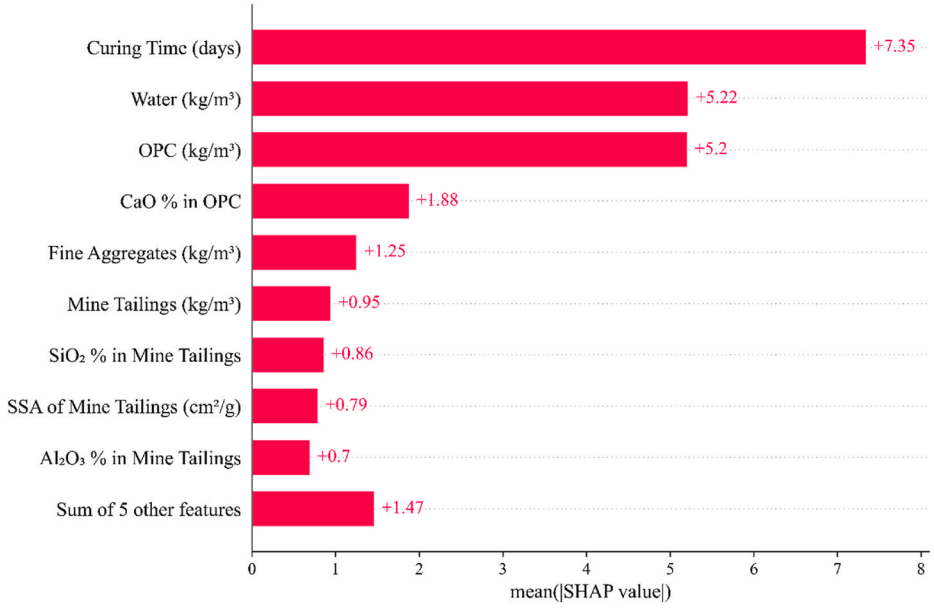

Fig. 5. TabPFN model feature importance analysis for UCS prediction.

lowest, primarily because the majority of the datasets in this study involved mortar mixtures, in which coarse aggregates were not used (i.e., their value was zero).

In contrast to the feature importance plot, the impact analysis shown in Fig. 6 provides extensive information about each input feature's relationship with the output. In this plot, a SHAP value of zero represents the mean UCS prediction made by the TabPFN model. Each dot symbolizes a unique data point, with its position on the x-axis indicating its SHAP value—that is, the contribution of that feature to the prediction. The color of each dot corresponds to the feature's actual value, based on the gradient bar: red indicates higher feature values, while blue represents lower values. Curing time has a positive contribution to UCS, as indicated by the red data points (i.e., higher curing times) being associated with positive SHAP values. This aligns with existing literature, which shows that longer curing durations result in more hydration products, thereby enhancing the strength of the mixture [61]. Next, the contents of water and OPC show negative and positive correlations with UCS, respectively, which is consistent with experimental findings. High water content tends to lower the strength of cementitious mixtures by increasing porosity, as the excess water evaporates and leaves behind voids upon drying [2]. In contrast, mixtures with higher OPC content typically generate more C-S-H gels as hydration products, leading to increased strength [62]. The relationship between CaO content in OPC and UCS is complex. While higher CaO levels generally promote early-age strength development through accelerated hydration and greater C–S–H formation, excessive CaO—especially when not effectively utilized in pozzolanic reactions—can lead to a buildup of unreacted $Ca(OH)_2$. This may negatively affect long-term strength and microstructural integrity [20]. The UCS exhibits a negative response to excessive fine aggregate content in the mixture. This can be ascribed to the reduced amount of OPC per unit volume, which limits the formation of hydration products and weakens the bonding between particles [62]. Literature suggests that excessive replacement of OPC with mine tailings can reduce the availability of $Ca(OH)_2$ due to cement dilution [16]. Although higher amounts of mine tailings introduce additional silica and alumina into the mixture, the limited presence of reactive $Ca(OH)_2$ hinders pozzolanic reactions and hydration products, ultimately leading to a decrease in UCS [16]. The trend observed for mine tailings in Fig. 6 aligns with these findings in the literature, as red data points—representing higher tailings content—are associated with negative SHAP values. The trends for SiO_2 and Al_2O_3 contents in mine tailings further support the reliability of the TabPFN model, as they confirm observations from literature. Although tailings with higher Si and Al are typically crystalline and less reactive, mechanical activation can alter their crystalline structure, making them more reactive [31]. Under such conditions, tailings with higher Si and Al contents can release more reactive ions into the matrix, promoting the formation of additional C-A-S-H gels as secondary hydration products [16]. The

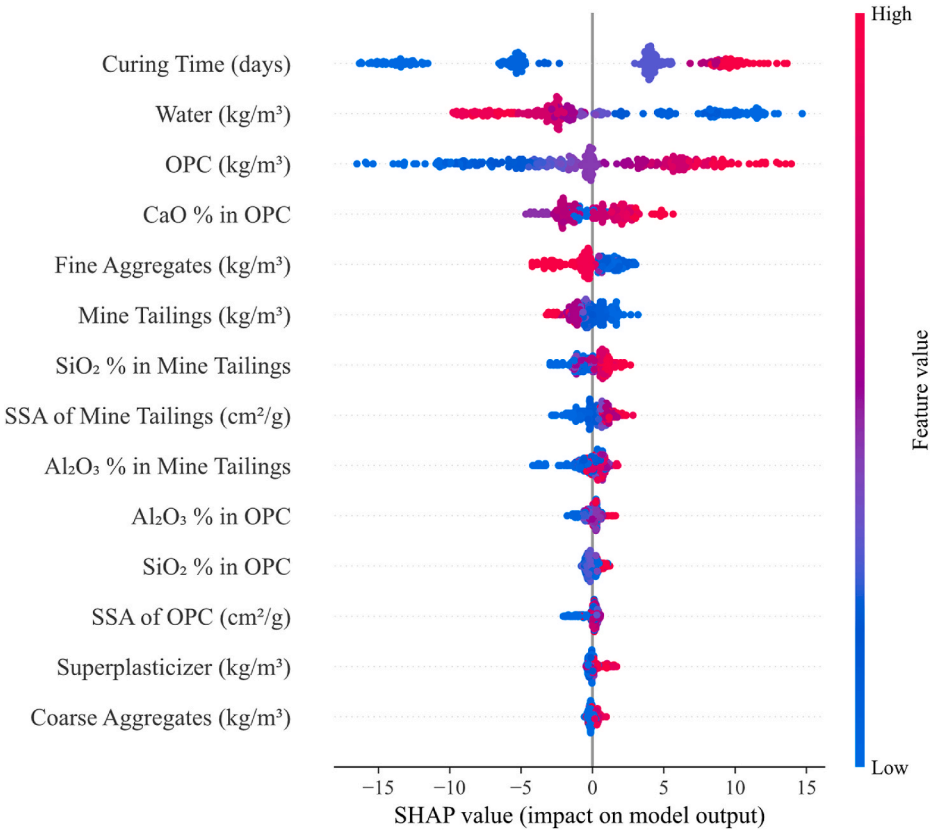

Fig. 6. Feature impact analysis for UCS prediction based on SHAP values from the TabPFN model.

observation in this study is consistent with these findings because the dataset is dominated by mechanically activated tailings, whose enhanced reactivity explains the positive relationship between higher Si and Al contents and improved strength. Likewise, the trend for SSA of mine tailings is consistent with literature, showing that increased SSA enhances reactivity, leading to higher UCS through the generation of more hydration products [16]. For OPC, the trends observed for SiO_2 and Al_2O_3 contents, as well as SSA, mirror those of mine tailings and show strong agreement with experimental evidence [61]. The plot for superplasticizer content indicates that increasing its dosage leads to improved UCS. This is likely due to the superplasticizer's ability to effectively disperse cement particles, promoting faster and more complete hydration reactions [18]. Lastly, coarse aggregate content shows a positive correlation with UCS, which can be attributed to the inclusion of concrete samples in the dataset. These samples generally exhibit higher strength compared to mortar mixtures, which lack coarse aggregates. Overall, the SHAP analysis reinforces the reliability of the TabPFN model, as the feature impact trends closely align with established findings in the literature, providing interpretable and consistent explanations for the model's predictions.

3.3. Multi-objective mixture design optimization

This section presents the results of the two separate multi-objective optimization analyses—bi-objective and tri-objective—conducted on the selected case study described in Appendix A.

3.3.1. Bi-objective optimization results

Fig. 7 illustrates the results of bi-objective optimization considering the cost and the UCS of different cementitious mixtures at 28 days. In the scatter plot, experimental designs and pareto-optimal solutions (i.e., pareto front) are represented by purple and orange points, respectively. The distribution reveals two distinct groups among the experimental

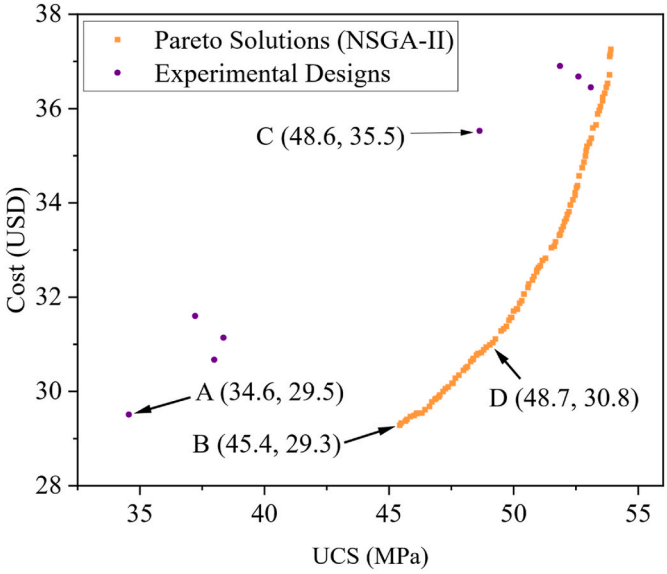

Fig. 7. Pareto front and experimental designs of the bi-objective optimization problem.

designs: one characterized by lower UCS and cost, and the other associated with higher UCS and cost. Compared to the first group, the higher UCS designs incorporate increased aggregate content (e.g., from 684 kg/m^3 to 725 kg/m^3), which enhances strength but also leads to higher overall mixture costs. Additionally, within each group, variations in UCS and cost can be attributed to differences in tailings content (ranging from 20 kg/m^3 to 60 kg/m^3). Since tailings are less expensive than OPC, mixtures with higher tailings content tend to be more cost-effective.

Fig. 7 shows the pareto front consisting of 100 unique, non-dominated optimal solutions. In each case, improving one objective would lead to a decline in the other, highlighting the trade-offs involved in the optimization process. Appendix B presents a detailed summary of these solutions, including the optimal proportions of each constituent material along with their corresponding UCS and cost values. As indicated by the results, the pareto-optimal solutions offer improved trade-offs between the two objectives, identifying mixture designs that either achieved higher UCS or reduced cost more effectively than the experimental designs. For example, point D on the pareto front (48.7 MPa, US$ 30.8/m^3) demonstrates that the cost can be reduced by US$ 4.7/m^3 compared to the experimental design at point C (48.6 MPa, US$ 35.5/m^3), while maintaining a nearly identical UCS through a different mixture composition. Likewise, the comparison between points A (34.6 MPa, US$ 29.5/m^3) and B (45.4 MPa, US$ 29.3/m^3) reveals that for a similar expenditure, the UCS of the mixtures can be significantly improved by 10.8 MPa. Furthermore, the well-distributed pareto front provides the flexibility to select an optimal mixture design based on specific engineering requirements, with each solution offering a guaranteed improvement over the experimental designs in at least one objective. This approach ultimately addresses the challenge of identifying the optimal proportions of mine tailings to replace OPC, along with the appropriate quantities of other constituent materials, aiming to reduce the overall cost of the mixtures without compromising their mechanical performance.

3.3.2. Tri-objective optimization results

Fig. 8 shows the results for extended analysis from the bi-objective problem, which considers an additional objective: emissions, alongside cost and UCS. Each scatter point represents a solution defined by its UCS, cost, and emissions values plotted along the x, y, and z axes, respectively. The experimental designs are represented by blue scatter points, whereas red points represent 100 unique pareto optimal solutions. These solutions are summarized in Appendix C. The differences in the z-axis (emissions) values among the points can be attributed to the varying OPC replacement ratios with mine tailings, as well as the substantial disparity in emission factors between OPC and mine tailings (0.919 kg/m^3 and 0.16424 kg/m^3, respectively). Unlike the bi-objective case where the pareto front appears as a 2D curve, the tri-objective pareto front can be visualized as a surface distributed across a 3D space.

The pareto front from the tri-objective optimization aligns with the results of the previous bi-objective analysis, consistently outperforming the experimental designs with improved mixture compositions. These optimized solutions achieve better trade-offs by enhancing at least one of the objectives. For instance, compared to the experimental design at point A (37.2, 31.6, 383), the pareto-optimal solution at point B (45.4, 29.3, 349.5) achieves a substantial reduction in both emissions (by 33.5 kg CO_2 eq./m^3) and cost (by US$ 2.3/m^3), while also improving UCS by 8.2 MPa. Similarly, the comparison between points C (experimental) and D (pareto) reveals marginal improvement in UCS by 0.7 MPa, along with reductions in cost and emissions by US$ 0.6/m^3 and 8.9 kg CO_2 eq./m^3, respectively. Although the cost and emissions savings may appear marginal per unit volume of mixture, their impact becomes significantly more substantial when considered in the context of real-world construction projects which use large volumes of concrete. Overall, the results from both bi-objective and tri-objective optimization demonstrate the effectiveness of the combined TabPFN and NSGA-II approach in guiding the mixture design of mine tailings-based cementitious materials. The generated pareto fronts provide engineers with a valuable decision-making tool to select optimal mixture designs that meet specific performance, cost, and emission targets. Importantly, the multi-objective optimization framework is adaptable to different regional contexts by updating cost and emissions coefficients, as well as modifying relevant constraints, making it broadly applicable across diverse construction settings.

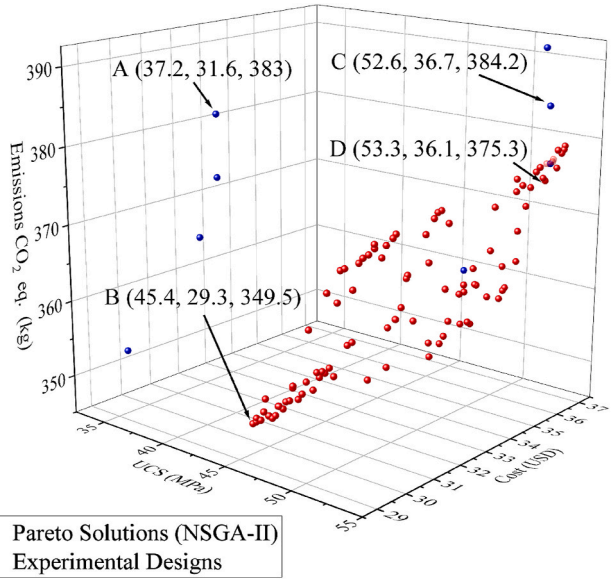

Fig. 8. Pareto front and experimental designs of the tri-objective optimization problem.

3.4. GreenMix AI: a software tool

The newly developed software tool, GreenMix AI, provides user-friendly access to the ML models and optimization algorithms developed in this study. It features three main functions: (1) updating the existing model with new data, (2) predicting the UCS of mixture designs, and (3) performing multi-objective optimization. Each function is organized into a dedicated tab within the software interface for ease of use.

Fig. 9 shows a snapshot of the first tab, which enables users to update the current TabPFN model with new data. Section (a) displays the available tabs, allowing easy navigation between different functionalities. When new experimental data—such as that obtained from a different type of mine tailings—is available, users can upload it using button (b), which opens a file dialog to select a locally stored file in comma separate values (CSV) format. Button (c) initiates data validation, checking for missing values and verifying that the input features match the required format. Once the data is validated, button (d) triggers the update process, retraining the TabPFN model with the new data and saving the updated model for future use. A dedicated notifications area provides real-time feedback, guiding users through each step and alerting them to any issues encountered.

Fig. 10 illustrates the second tab of the GreenMix AI software, which allows users to predict the UCS of a specific mixture design. In section (a), users are required to input all necessary parameters related to the mixture. Once the inputs are provided, the prediction can be initiated by clicking button (b), which triggers the trained ML model embedded in the backend of the software. The predicted strength is then displayed in the output box (c), with additional guidance provided through prompt messages displayed below the output area.

Fig. 11 presents the third and final tab of the GreenMix AI software, which enables users to perform MOO of mixture designs. In the backend, the software formulates the optimization problems based on the methodology described in section 2.7.1. In area (a) in Fig. 11, users can choose between bi-objective and tri-objective optimization modes using radio buttons. Section (b) guides users through a step-by-step input process, where they provide all necessary information, including values for constant variables, cost and emissions coefficients, range constraints, and ratio constraints. Section (c) allows users to specify optimization parameters (e.g., population size, number of generations), and the optimization process is initiated by clicking button (d). Once executed,

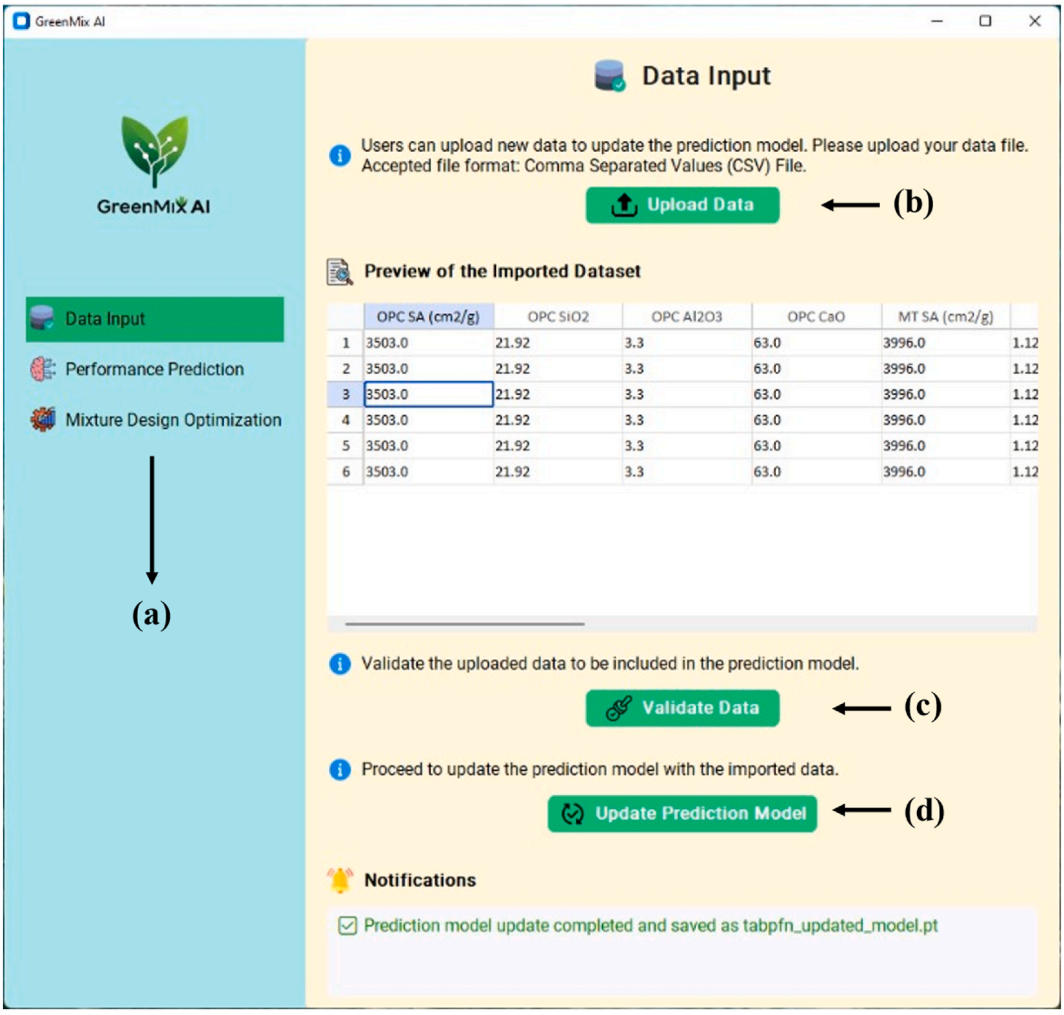

Fig. 9. GreenMix AI Table 1 for data input.

the algorithm runs in the background, and the optimized mixture designs are saved as an Excel (.xlsx) file locally on the user's computer. To enhance user experience, the software provides real-time notifications to flag issues such as missing values or illogical inputs (e.g., upper bounds being smaller than lower bounds), and to suggest corrective actions or next steps. Additionally, a continuously updated progress bar displays the optimization status, offering users a clear view of ongoing computational progress.

GreenMix AI enables easy customization, allowing users to tailor their own MOO studies based on different types of mine tailings and OPC, experimental conditions, and cost and emissions data (i.e., changes depending on the regional and transport-specific factors)—simply by modifying input values within the program. This flexibility makes the software a universally applicable tool, with the potential to significantly advance sustainable and economically resilient practices in the construction industry globally.

4. Research limitations and recommendations for future studies

Although the dataset used in this study is significantly larger and more diverse than those found in the literature, continuously updating the dataset remains important to further enhance model performance. Furthermore, incorporating additional input features—such as curing time, humidity level, cement mixing parameters (e.g., duration and speed) and UCS testing conditions (e.g., loading rate and sample size)—could enhance the robustness of the current model. The TabPFN model is particularly well-suited for this, as it requires minimal adjustments when new data are added—unlike many traditional models that require extensive re-tuning. The developed software tool, GreenMix AI, further streamlines this process, allowing users with minimal coding experience to easily update the model through a simple, user-friendly interface. In the future, concepts such as federated learning could be integrated into the software, enabling collaborative training of the TabPFN model across multiple users or institutions without directly sharing raw data—thereby preserving data privacy. In addition, as transformer-based pre-trained models continue to evolve, it will be important to compare the performance of newer models against TabPFN as they become available. This will help ensure that the most accurate and efficient tools are used for future applications. Beyond the potential technical improvements to GreenMix AI, an important research direction will be the evaluation of its usability. Future work will involve assessing the software with end-users to examine its practicality, user-friendliness, and ability to integrate into real-world workflows.

Additionally, this study employed only the NSGA-II algorithm for the MOO case study. Future work should explore other MOO algorithms, such as multi-objective evolutionary algorithm based on decomposition (MOEA/D) [64] and NSGA-III [65], to evaluate their applicability and compare results. In addition to the above limitations, the authors acknowledge the absence of experimental validation for the optimized mixture designs. While the data-driven results obtained in this study are promising, confirming their reliability through dedicated laboratory experiments is essential. Future work will therefore focus on conducting

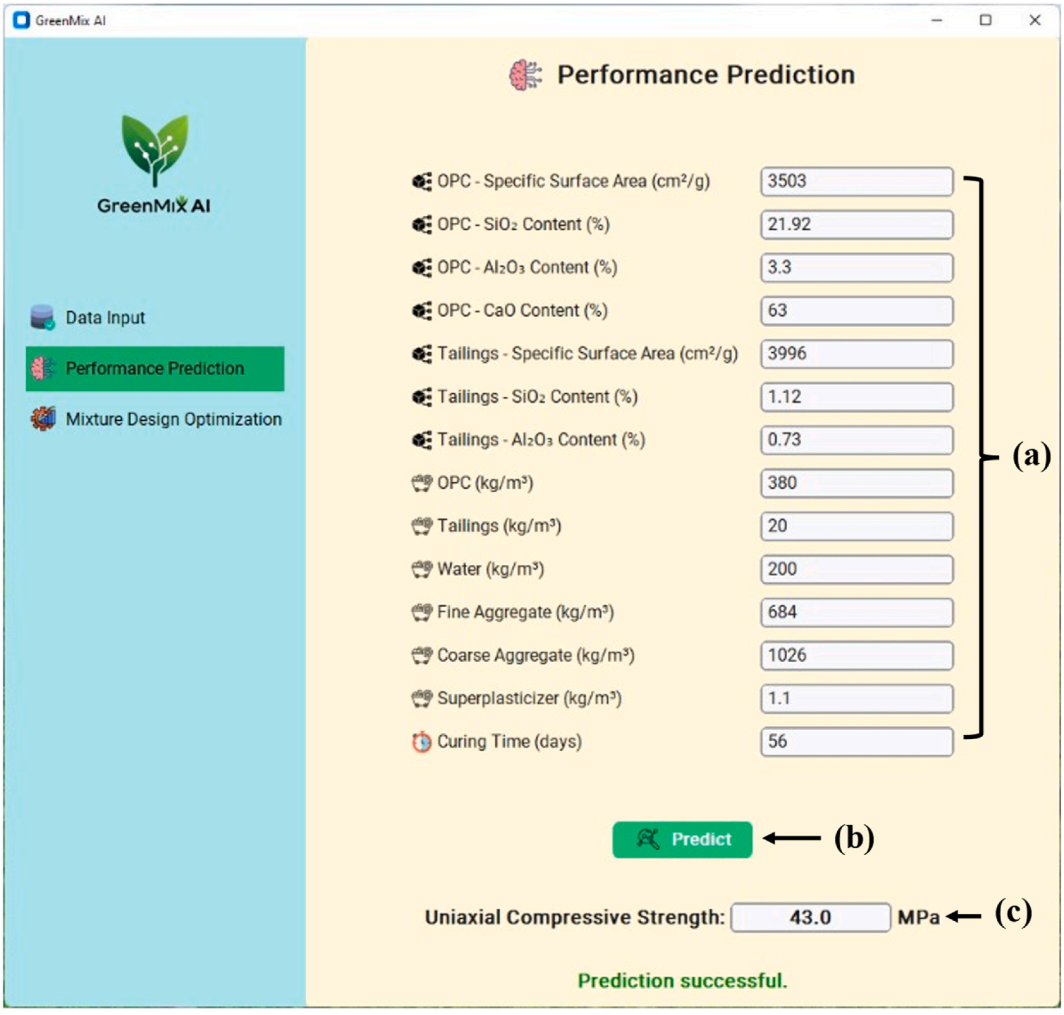

Fig. 10. GreenMix AI Table 2 for performance prediction.

such experiments to validate the optimized designs. Finally, while this study focused solely on UCS as the primary objective, other critical performance indicators—such as setting time, workability (flow), and tensile strength—can also be considered as additional objectives, depending on specific design requirements.

5. Conclusion

This research introduces a new data-driven intelligent design that applies machine learning (ML) techniques to the multi-objective optimization (MOO) of cementitious mixtures incorporating mine tailings as supplementary cementitious materials (SCMs). A transformer-based tabular prior data fitted network (TabPFN) model—shown to outperform conventional boosting ML models—was used to predict uniaxial compressive strength (UCS) and further interpreted using SHapley Additive exPlanations (SHAP). In the MOO case study, the TabPFN model was integrated with cost and emissions objectives to optimize the mixture design. The following conclusions can be reached based on the findings of this study.

1. The Optuna framework effectively optimized the hyperparameters of the selected boosting ML models, resulting in enhanced performance over their respective baseline versions, with some models achieving up to a 30 % reduction in UCS prediction errors on the testing data. Among all models evaluated, the transformer-based TabPFN model exhibited the best overall performance, outperforming even the top-tuned boosting models. Importantly, the TabPFN model required no model-specific training or hyperparameter tuning, significantly reducing computational time by 1045 s. Given its predictive accuracy, computational efficiency, and adaptability, TabPFN is particularly well-suited for regression tasks involving small, literature-based datasets that are frequently updated.
2. SHAP analysis identified curing time as the most prominent input in the TabPFN model, followed by key constituent materials such as water, OPC, fine aggregates, and mine tailings. The results emphasized the importance of controlling water content, aggregate proportions, and the OPC replacement ratio with mine tailings, as each showed a strong relationship with UCS. Additionally, the analysis revealed that the reactivity of mine tailings—reflected by their specific surface area—plays a critical role in strength development. Tailings with higher SiO_2 and Al_2O_3 contents were also associated with greater strength gains, highlighting the need to select appropriate materials based on performance requirements. Overall, the SHAP-derived model interpretations were aligned with experimental evidence, reinforcing the model's reliability and interpretability. These insights provide engineers with practical guidance for designing sustainable mixtures tailored to specific strength targets.
3. The MOO case study demonstrated the critical value of optimizing mine tailings based-mixture designs to achieve maximum benefits of waste reutilization. All pareto-optimal solutions generated by the TabPFN model, in combination with the non-dominated sorting genetic algorithm-II (NSGA-II), outperformed the original experimental

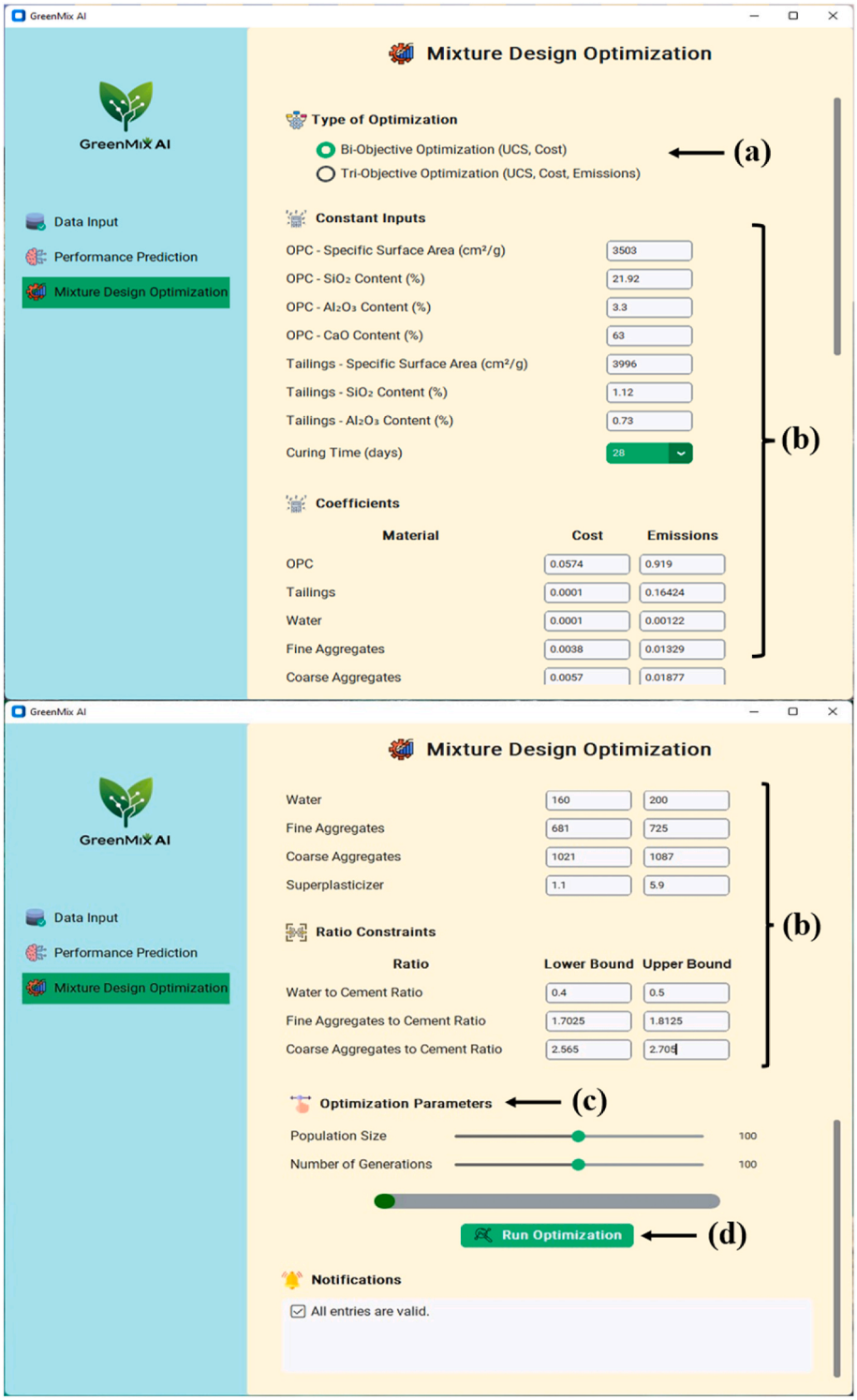

Fig. 11. GreenMix AI Table 3 for mixture design optimization.

designs in both bi-objective and tri-objective optimization scenarios. These optimized solutions offer engineers a practical framework for tailoring mixtures to meet specific performance requirements while minimizing cost and environmental impact.

4. The developed software tool, GreenMix AI, serves as a critical bridge between advanced research methodologies and practical implementation. It provides streamlined access to the ML models and MOO algorithms developed in this study, without requiring users to have expertise in coding, ML, or optimization. Through an intuitive interface, users can update the model with their own experimental data, predict the UCS of new mixture designs, and perform customized optimization studies to identify designs that balance strength, cost, and emissions. Overall, this research showcases the potential of integrating ML and optimization techniques to promote the use of mine tailings as SCMs, advancing both engineering efficiency and sustainability in the mining industry.

CRediT authorship contribution statement

Chathuranga Balasooriya Arachchilage: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jian Zhao:** Writing – original draft, Methodology, Formal analysis. **Nimila Dushyantha:** Writing – original draft, Software, Resources. **Wei Victor Liu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Wei Victor Liu reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN-2024-04712).

Appendix A

A.1. Cost and LCA data

In this study, a previously published experimental investigation from Egypt [21], which utilized marble tailings as SCMs, was used as the basis for the case study formulation. This study was selected as it provides experimental results for a wide range of mixture conditions—such as varying water-to-cement ratios, aggregate-to-cement ratios, and tailings contents—resulting in multiple data points. Accordingly, the coefficients used in Equations (11) and (12) in the manuscript were determined based on values reported in the literature and supplemented with data from the Ecoinvent database accessed through the OpenLCA software. These values are listed in Table A.1.

Table A.1
Values of cost and emission coefficients used in the case study

Material	Cost coefficients (US$/kg)		Emission coefficients (CO_2 eq./kg)	
OPC	C_{OPC}	0.0574	E_{OPC}	0.919
Mine tailings	C_{MT}	0.0001	E_{MT}	0.16424
Water	C_W	0.0001	E_w	0.00122
Fine aggregates	C_{FA}	0.0038	E_{FA}	0.01329
Coarse aggregates	C_{CA}	0.0057	E_{CA}	0.01877
Superplasticizer	C_{SP}	1.2	E_{SP}	1.67377

The cost coefficients for each material were derived from the latest available reports on the Statista online database, which provide average construction material prices in Egypt. The cost of OPC was reported as US$ 2.87 per 50 kg sack, resulting in a unit price of approximately US$ 0.0574 per kg. The prices of fine and coarse aggregates were given as US$ 6.08/m^3 and US$ 10.49/m^3, respectively. These values were converted to cost per kg by assuming average bulk densities of 1600 kg/m^3 for fine aggregates and 1850 kg/m^3 for coarse aggregates. Similarly, the commercial rate for water in Egypt, estimated at US$ 0.0001 per kilogram, was used to ensure consistency with data representative of an industrial-scale project. Furthermore, commercial bulk pricing was considered for estimating the unit cost of polycarboxylate ether (PCE) superplasticizer (40 % active substance), reflecting typical rates for large-scale procurement. Regarding mine tailings, a zero-burden approach was adopted, meaning the upstream costs associated with mining, mineral processing, and tailings generation were excluded. It was assumed that marble tailings were dry and readily available at the tailings storage facility (TSF), requiring no additional energy for drying. However, the transport costs were considered non-negligible, and it was assumed that the TSF and the concrete mixing plant are located within 1.5 km distance, consistent with assumptions made by Zhou et al. (2024) [66]. Finally, based on an average transport cost of US$ 0.066 per tonne per kilometer in Egypt and a distance of 1.5 km, the final cost of mine tailings was estimated at US$ 0.001 per kg.

The emission coefficients presented in Table A.1 were derived using data from version 3.11 of the Ecoinvent database. For all constituent materials, market activity processes were selected, because they account for emissions associated with the production of each material as well as transportation from producers to consumers (i.e., averaged) within the selected region. Since Egypt is not explicitly represented in the database, the region labeled rest of the world (RoW) was used for the analysis. All the life cycle impact analysis data were generated using the ReCiPe 2016 (v1.03), midpoint (H) method. Consistent with the cost calculations, the emissions associated with mine tailings were based solely on transportation. Specifically, the market process "freight lorry, 32-ton, EURO 6" from the Ecoinvent database was used, assuming a transport distance of 1.5 km. Emissions from upstream processes were excluded in accordance with the zero-burden approach. It is important to emphasize the influence of transport distance, as studies have shown that transporting materials over distances greater than 15 km can negate the environmental benefits of using more sustainable alternatives [67]. Therefore, the feasibility of utilizing mine tailings as SCMs in cementitious mixtures largely depends on the proximity of the TSF to the concrete plant.

A.2. Constraints

Constraints play a crucial role in a MOO problem by ensuring that the generated solutions are both practical and feasible for real-world implementation. In this study, both range constraints and ratio constraints were applied, as shown in Table A.2 and Table A.3. These constraints were derived from the mixture designs reported in the referenced case study on marble tailings [21]. It is important to note that the specific values of these

constraints should be adjusted based on the mixture designs and requirements of other studies.

Table A.2
Range constraints

	Lower bound (kg/m^3)	Upper bound (kg/m^3)
OPC	340	380
Mine Tailings	20	60
Water	160	200
Fine aggregates	681	725
Coarse aggregates	1021	1087
Superplasticizer	1.1	5.9

Table A.3
Ratio constraints

	Lower bound (kg/m^3)	Upper bound (kg/m^3)
Water-to-cement ratio (W/C)	0.4	0.5
Fine aggregates-to-cement ratio (FA/C)	1.7025	1.8125
Coarse aggregate-to-cement ratio (CA/C)	2.565	2.705

Appendix B

Table B.1
Optimal mixture designs and their respective UCS and cost values identified by bi-objective optimization

OPC	Tailings	Water	Fine Aggregates	Coarse Aggregates	SP	UCS (MPa)	Cost ($/m^3)
372.90	31.09	162.03	690.09	1064.33	5.87	53.88	37.16
374.18	31.12	162.43	691.82	1065.76	5.88	53.90	37.26
373.22	31.19	161.97	688.60	1064.37	5.82	53.86	37.11
365.27	32.74	160.00	681.92	1025.16	4.08	52.52	34.32
352.95	32.80	160.03	681.79	1021.96	1.73	48.51	30.77
364.97	33.07	160.05	681.37	1066.77	5.20	53.37	35.88
360.49	33.18	160.01	681.08	1022.86	3.08	51.28	32.82
365.57	33.19	160.00	681.14	1064.26	3.92	52.57	34.36
365.57	33.20	160.00	681.14	1064.26	5.88	53.85	36.72
365.63	33.20	160.00	681.13	1031.69	5.88	53.77	36.53
365.69	33.21	160.00	681.11	1064.26	3.79	52.47	34.21
364.97	33.22	160.09	681.13	1024.58	5.70	53.57	36.24
365.91	33.22	160.08	681.08	1028.91	3.48	52.11	33.66
361.93	33.22	160.04	681.65	1029.37	1.93	49.88	31.57
360.40	33.24	160.00	681.09	1022.88	1.46	48.79	30.88
360.46	33.24	160.08	681.09	1022.88	1.41	48.71	30.82
361.93	33.25	160.04	681.65	1029.37	1.93	49.92	31.57
364.93	33.25	160.04	681.08	1023.56	1.76	49.82	31.51
364.12	33.26	160.02	681.90	1022.86	4.88	52.95	35.20
363.42	33.26	160.02	681.75	1023.54	4.84	52.91	35.11
361.30	33.29	160.04	681.79	1030.96	5.30	53.18	35.59
361.67	33.34	160.07	681.00	1022.44	1.78	49.60	31.33
364.96	33.35	160.08	681.15	1028.91	3.48	52.04	33.60
364.96	33.35	160.00	681.15	1025.15	3.41	51.99	33.50
365.61	33.35	160.04	681.91	1025.13	4.94	53.12	35.37
363.42	33.41	160.03	681.75	1023.49	2.77	51.00	32.62
364.11	33.41	160.00	681.95	1025.13	2.44	50.60	32.28
363.36	33.41	160.04	681.26	1023.78	3.88	52.28	33.96
363.43	33.41	160.03	682.09	1024.52	3.43	51.91	33.43
364.48	33.41	160.04	681.02	1031.30	5.87	53.71	36.45
363.41	33.41	160.02	681.89	1030.88	3.93	52.38	34.07
363.38	33.47	160.04	681.62	1024.05	5.55	53.43	35.96
364.31	33.78	160.04	681.79	1023.27	1.61	49.50	31.28
364.95	33.85	160.04	681.26	1023.78	4.56	52.82	34.86
364.95	33.85	160.04	681.26	1023.78	4.67	52.88	34.99
363.37	33.85	160.09	681.78	1023.32	2.14	50.25	31.87
363.37	33.85	160.07	681.78	1023.23	2.19	50.33	31.92
364.86	33.86	160.07	687.14	1029.23	2.19	50.41	32.07
364.85	33.86	160.07	687.14	1029.24	2.19	50.41	32.06
364.22	33.86	160.02	681.02	1023.39	4.35	52.63	34.57
361.61	33.87	160.05	681.90	1029.08	2.48	50.58	32.20
365.58	33.89	160.03	681.75	1029.21	2.77	51.16	32.78

(*continued on next page*)

Table B.1 (*continued*)

OPC	Tailings	Water	Fine Aggregates	Coarse Aggregates	SP	UCS (MPa)	Cost ($/m^3)
364.12	33.89	160.00	681.90	1033.37	4.90	53.04	35.29
363.01	33.90	160.02	681.98	1022.59	1.38	48.89	30.94
351.24	33.94	160.03	681.79	1023.82	1.83	48.57	30.81
363.41	33.97	160.03	681.79	1022.73	2.73	50.93	32.57
364.96	33.98	160.09	681.24	1023.84	5.65	53.57	36.17
364.23	33.98	160.02	682.08	1034.31	3.62	52.15	33.75
351.24	33.98	160.02	681.80	1021.98	1.54	47.98	30.45
351.24	33.98	160.03	681.80	1029.07	1.72	48.39	30.70
350.55	33.98	160.04	681.80	1023.27	1.72	48.28	30.63
351.46	33.98	160.03	681.80	1029.18	1.69	48.36	30.68
364.22	33.98	160.02	682.08	1034.31	3.67	52.22	33.81
364.31	33.98	160.00	681.24	1023.50	3.08	51.52	33.05
364.24	33.98	160.04	681.62	1030.06	5.55	53.49	36.04
364.80	33.99	160.02	681.75	1029.21	2.52	50.81	32.44
364.10	33.99	160.02	681.75	1023.51	2.52	50.75	32.37
364.79	33.99	160.03	681.07	1025.19	2.62	50.92	32.53
364.25	33.99	160.04	681.75	1023.51	2.76	51.07	32.67
351.68	33.99	160.04	681.40	1022.40	1.59	48.14	30.53
364.79	33.99	160.02	681.12	1034.31	3.93	52.47	34.16
364.92	34.00	160.00	681.14	1024.59	5.78	53.65	36.32
365.06	34.03	160.03	681.33	1031.32	3.23	51.84	33.31
365.27	34.03	160.03	681.46	1031.32	3.23	51.87	33.33
364.27	34.04	160.04	681.87	1025.13	4.91	53.04	35.26
365.11	34.10	160.03	682.01	1027.32	4.43	52.75	34.74
360.46	34.39	160.02	681.79	1021.88	1.55	49.04	30.99
361.25	34.39	160.02	681.77	1022.49	1.55	49.16	31.03
349.23	34.40	160.02	681.77	1022.49	1.55	47.80	30.34
365.05	34.80	160.04	681.40	1025.69	5.20	53.31	35.65
363.12	34.87	160.04	681.46	1033.63	3.11	51.63	33.08
365.04	34.94	160.03	681.78	1066.03	1.72	50.00	31.70
363.15	34.95	160.00	681.95	1049.35	3.11	51.68	33.17
351.49	35.04	160.08	681.67	1022.38	1.57	48.08	30.50
362.61	35.05	160.05	681.13	1025.03	1.77	49.71	31.38
361.84	35.05	160.08	681.64	1021.64	1.59	49.26	31.11
363.06	35.07	160.00	682.11	1029.21	2.02	50.12	31.75
363.06	35.14	160.00	682.15	1029.21	2.02	50.15	31.75
343.19	39.01	160.02	681.65	1024.20	1.69	47.56	30.17
343.19	39.19	160.02	681.01	1024.20	1.69	47.53	30.17
341.31	39.21	160.03	681.12	1026.11	1.56	47.03	29.92
341.31	39.21	160.02	681.01	1024.40	1.54	46.99	29.88
341.31	39.21	160.03	681.59	1022.88	1.50	46.84	29.83
341.31	39.21	160.03	681.59	1022.88	1.52	46.89	29.85
341.94	39.23	160.02	681.77	1022.54	1.35	46.62	29.68
341.98	39.24	160.02	681.65	1022.42	1.69	47.37	30.09
340.39	39.32	160.04	682.09	1022.74	1.30	46.32	29.54
341.56	39.40	160.03	681.12	1026.11	1.61	47.17	30.00
340.77	39.61	160.09	681.66	1023.92	1.14	45.66	29.37
340.77	39.61	160.09	681.66	1023.92	1.71	47.26	30.06
340.92	39.61	160.09	681.12	1024.01	1.27	46.14	29.54
340.92	39.61	160.09	681.12	1022.76	1.27	46.11	29.53
340.84	39.66	160.04	681.56	1023.70	1.34	46.45	29.61
340.92	39.71	160.08	681.08	1022.81	1.24	46.01	29.49
351.55	39.73	160.02	681.98	1022.61	1.38	47.66	30.28
340.16	40.94	160.14	681.35	1022.68	1.20	45.72	29.41
340.56	41.01	160.03	681.38	1024.41	1.23	45.86	29.47
340.07	41.45	160.00	681.08	1022.82	1.10	45.43	29.28
341.98	41.78	160.03	681.31	1022.75	1.42	46.72	29.78
340.07	41.87	160.00	681.06	1025.09	1.13	45.48	29.33

Appendix C

Table C.1
Optimal mixture designs and their respective UCS, cost and emission values identified by tri-objective optimization

OPC	Tailings	Water	Fine Aggregates	Coarse Aggregates	SP	UCS (MPa)	Cost ($/m3)	CO_2 eq. (kg)
351.92	31.99	160.19	681.72	1029.30	3.87	51.27	33.32	363.72
363.82	32.81	160.13	681.80	1021.16	2.97	51.21	32.88	373.14
364.52	32.81	160.13	681.80	1023.54	2.99	51.31	32.96	373.86
364.62	32.81	160.13	681.80	1023.54	3.05	51.43	33.04	374.05
351.92	32.85	160.13	681.97	1023.50	2.99	50.16	32.24	362.29
349.38	33.11	160.01	681.64	1029.63	4.82	51.69	34.31	363.16

(*continued on next page*)

Table C.1 (*continued*)

OPC	Tailings	Water	Fine Aggregates	Coarse Aggregates	SP	UCS (MPa)	Cost ($/m3)	CO_2 eq. (kg)
349.57	33.11	160.01	681.64	1024.40	4.82	51.62	34.30	363.24
363.76	33.77	160.03	681.66	1024.44	2.17	50.33	31.93	371.95
363.35	33.78	160.04	681.67	1069.00	5.83	53.70	36.55	378.54
363.76	33.78	160.04	681.67	1069.00	5.86	53.76	36.62	378.97
361.31	34.64	160.03	681.67	1024.43	5.69	53.33	36.02	375.74
361.31	34.64	160.03	681.49	1022.06	5.04	52.80	35.23	374.61
361.23	34.65	160.04	681.52	1028.07	1.60	49.29	31.12	368.89
360.81	34.65	160.04	681.52	1028.39	4.45	52.53	34.52	373.28
364.59	34.67	160.04	681.68	1024.89	5.86	53.67	36.41	379.05
361.99	34.71	160.07	681.52	1065.94	1.68	49.70	31.48	370.44
364.10	34.72	160.03	681.68	1068.06	5.87	53.78	36.64	379.44
360.84	34.73	160.17	681.80	1024.29	2.04	49.95	31.61	369.22
363.35	34.74	160.13	681.57	1028.05	5.83	53.51	36.32	377.93
363.22	34.74	160.03	681.80	1029.75	1.76	49.76	31.44	371.03
345.32	34.76	160.03	681.52	1027.93	1.35	47.02	29.91	353.86
352.39	34.77	160.02	681.50	1024.59	4.52	51.82	34.11	365.61
363.32	34.79	160.13	681.62	1025.00	5.48	53.33	35.88	377.26
363.32	34.79	160.00	681.62	1025.00	5.65	53.44	36.09	377.55
345.94	34.80	160.02	681.63	1024.74	5.79	52.12	35.25	361.81
345.54	34.80	160.02	681.66	1023.89	5.79	52.03	35.23	361.43
343.59	34.80	160.08	682.06	1021.56	1.35	46.79	29.78	352.18
343.59	34.80	160.03	682.06	1021.52	1.41	46.92	29.85	352.27
361.60	34.81	160.03	681.66	1021.20	2.91	51.02	32.67	371.31
357.37	34.84	160.00	682.01	1021.52	1.76	49.01	31.05	365.52
344.13	34.84	160.03	682.01	1021.66	1.76	47.70	30.29	353.36
353.80	34.84	160.02	682.01	1021.32	4.88	52.16	34.60	367.46
349.28	34.90	160.03	682.08	1024.77	2.85	49.82	31.92	359.98
348.91	34.91	160.08	681.66	1024.44	3.33	50.58	32.48	360.45
360.96	34.98	160.02	681.66	1024.30	5.35	53.17	35.59	374.91
360.99	34.98	160.01	681.71	1070.16	5.87	53.52	36.48	376.67
360.48	34.98	160.01	681.71	1028.08	5.80	53.31	36.12	375.28
360.97	35.00	160.03	681.62	1024.55	1.41	48.80	30.86	368.32
357.00	35.12	160.03	681.65	1028.30	3.17	51.10	32.76	367.70
355.98	35.65	160.03	682.07	1024.89	1.55	48.51	30.75	364.09
363.15	35.65	160.04	681.63	1027.94	4.88	52.91	35.17	376.31
354.44	35.65	160.02	682.05	1025.45	5.49	52.69	35.39	369.29
343.47	35.65	160.04	681.63	1025.00	4.88	51.12	34.02	358.17
349.10	35.67	160.03	681.64	1024.89	4.41	51.44	33.78	362.55
342.46	35.97	160.13	681.69	1021.15	1.12	45.88	29.43	350.93
342.46	35.97	160.00	681.69	1021.15	1.36	46.72	29.72	351.32
361.68	36.02	160.03	681.49	1024.42	5.12	53.02	35.36	375.36
342.08	36.09	160.13	681.54	1021.15	2.88	49.19	31.52	353.53
356.00	36.09	160.04	682.07	1024.77	2.86	50.43	32.32	366.38
355.60	36.09	160.04	682.11	1023.93	2.86	50.39	32.30	366.00
349.10	36.12	160.04	681.68	1024.89	5.86	52.42	35.53	365.06
362.13	36.13	160.02	681.52	1064.45	5.25	53.25	35.77	376.75
361.59	36.19	160.08	681.49	1021.14	1.68	49.43	31.20	369.47
362.12	36.20	160.13	681.51	1029.55	2.03	50.09	31.70	370.71
344.74	36.21	160.13	681.49	1029.55	2.04	48.33	30.72	354.76
360.63	36.24	160.01	681.67	1024.53	1.36	48.66	30.78	368.14
343.39	36.25	160.03	681.47	1021.58	4.09	50.62	33.05	356.81
361.59	36.26	160.08	682.06	1024.55	3.34	51.65	33.21	372.34
346.37	36.27	160.02	681.71	1027.30	1.88	48.24	30.61	355.97
346.37	36.27	160.02	681.70	1027.30	1.67	47.87	30.35	355.60
349.66	36.30	160.00	681.68	1024.65	5.45	52.23	35.06	364.90
350.08	36.30	160.00	681.62	1024.65	4.46	51.54	33.90	363.64
345.59	36.37	160.04	682.04	1026.09	5.37	51.74	34.74	361.08
345.38	36.42	160.19	681.50	1021.14	1.59	47.48	30.16	354.47
357.45	36.42	161.39	682.00	1021.52	5.59	52.78	35.66	372.27
345.45	36.42	160.19	681.50	1021.14	1.80	47.92	30.42	354.88
349.69	36.60	160.00	681.97	1024.67	2.99	49.99	32.11	360.87
349.69	36.60	160.13	681.97	1024.67	2.99	50.02	32.11	360.87
343.20	36.77	160.08	681.48	1022.08	5.04	51.35	34.19	358.32
362.63	36.78	160.04	681.52	1029.03	5.83	53.51	36.29	377.63
362.63	36.78	160.03	681.51	1029.02	2.15	50.23	31.87	371.47
348.91	36.88	160.03	681.66	1024.44	2.18	48.89	31.09	358.84
357.60	36.88	160.03	681.67	1024.36	1.32	48.14	30.56	365.38
348.71	36.89	160.13	681.54	1026.38	4.55	51.50	33.94	362.66
345.13	36.89	160.03	681.50	1024.36	4.43	51.06	33.57	359.12
343.91	36.90	160.03	681.45	1024.36	5.82	51.86	35.17	360.33
345.70	36.91	160.03	681.66	1024.36	1.86	48.09	30.52	355.35
343.00	36.91	160.04	681.71	1024.39	1.55	47.23	30.00	352.36
345.27	36.95	160.03	681.66	1024.36	1.36	47.05	29.90	354.13
348.70	36.97	160.13	681.56	1026.38	2.09	48.76	30.99	358.55
343.69	37.00	160.00	682.00	1021.08	1.59	47.37	30.07	353.02
341.14	37.00	160.00	682.02	1021.34	1.14	45.78	29.38	349.91

(*continued on next page*)

Table C.1 (*continued*)

OPC	Tailings	Water	Fine Aggregates	Coarse Aggregates	SP	UCS (MPa)	Cost ($/m3)	CO_2 eq. (kg)
346.11	37.22	160.03	681.51	1021.07	1.80	48.03	30.46	355.63
361.51	37.49	160.03	682.01	1025.08	1.76	49.60	31.31	369.82
341.96	37.49	160.04	682.03	1021.33	4.45	50.77	33.40	356.30
340.54	37.49	160.04	681.71	1021.33	4.32	50.48	33.16	354.77
342.83	37.59	160.02	681.55	1028.30	4.44	50.91	33.48	357.23
341.14	40.67	160.02	681.49	1021.16	5.37	51.07	34.45	357.59
340.03	40.78	160.12	681.68	1021.15	1.10	45.42	29.27	349.45
340.67	40.78	160.12	681.68	1025.08	1.10	45.56	29.33	350.11
340.54	40.78	160.01	681.49	1028.56	1.26	46.02	29.54	350.33
342.02	40.78	160.01	681.49	1021.45	1.32	46.56	29.65	351.66
340.54	40.79	160.01	681.70	1028.56	1.27	46.06	29.54	350.34
342.02	40.79	160.01	681.70	1021.26	1.32	46.51	29.65	351.65
340.02	40.79	160.00	681.68	1028.55	1.12	45.52	29.33	349.61
340.54	40.79	160.01	681.70	1028.55	1.33	46.37	29.61	350.44
352.10	40.79	160.01	681.70	1023.64	1.31	47.51	30.23	360.94
343.32	40.83	160.04	682.04	1021.33	1.12	45.90	29.48	352.52
340.10	41.77	160.01	681.67	1025.00	1.32	46.25	29.56	350.12
341.40	41.79	160.08	682.06	1025.01	3.30	49.58	32.01	354.63

Data availability

Data will be made available on request.

References

[1] M. Gou, L. Zhou, N.W.Y. Then, Utilization of tailings in cement and concrete: a review, Sci. Eng. Compos. Mater. 26 (1) (2019) 449–464.

[2] C.B. Arachchilage, G. Huang, J. Zhao, C. Fan, W.V. Liu, Hybrid extreme gradient boosting regressor models for the multi-objective mixture design optimization of cementitious mixtures incorporating mine tailings as fine aggregates, Cement Concr. Compos. 154 (2024) 105787.

[3] L.R. Adrianto, L. Ciacci, S. Pfister, S. Hellweg, Toward sustainable reprocessing and valorization of sulfidic copper tailings: scenarios and prospective LCA, Sci. Total Environ. 871 (2023) 162038.

[4] X. Lyu, G. Yao, Z. Wang, Q. Wang, L. Li, Hydration kinetics and properties of cement blended with mechanically activated gold mine tailings, Thermochim. Acta 683 (2020) 178457.

[5] İ. Bekem Kara, Characterization of copper tailings in Murgul Copper Plant, Turkey, and its utilization potential in cement mortar with nano- and micro-silica, Environ. Sci. Pollut. Control Ser. 29 (24) (2022) 36938–36950.

[6] C. Ince, Reusing gold-mine tailings in cement mortars: mechanical properties and socio-economic developments for the Lefke-Xeros area of Cyprus, J. Clean. Prod. 238 (2019) 117871.

[7] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.

[8] Y. Cao, F. Su, M.F. Antwi-Afari, J. Lei, X. Wu, Y. Liu, Enhancing mix proportion design of low carbon concrete for shield segment using a combination of Bayesian optimization-NGBoost and NSGA-III algorithm, J. Clean. Prod. 465 (2024) 142746.

[9] L. Torrey, J. Shavlik, Transfer Learning, Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques, IGI global2010, pp. 242-264.

[10] N. Hollmann, S. Müller, K. Eggensperger, F. Hutter, Tabpfn: a Transformer that Solves Small Tabular Classification Problems in a Second, 2022 arXiv preprint arXiv:2207.01848.

[11] Z. Yu, R. Yu, X. Ge, J. Fu, Y. Hu, S. Chen, Tabular prior-data fitted network for urban air temperature inference and high temperature risk assessment, Sustain. Cities Soc. 128 (2025) 106484.

[12] E.O. Kurniati, H. Zeng, M.I. Latypov, H.J. Kim, Machine learning for predicting compressive strength of sustainable cement paste incorporating copper mine tailings as supplementary cementitious materials, Case Stud. Constr. Mater. 21 (2024) e03373.

[13] S. Surehali, T. Han, J. Huang, A. Kumar, N. Neithalath, On the use of machine learning and data-transformation methods to predict hydration kinetics and strength of alkali-activated mine tailings-based binders, Constr. Build. Mater. 419 (2024) 135523.

[14] J. Esmaeili, H. Aslani, O. Onuaguluchi, Reuse potentials of copper mine tailings in mortar and concrete composites, J. Mater. Civ. Eng. 32 (5) (2020) 04020084.

[15] M. Soltaninejad, M. Soltaninejad, F.S. K, M.K. Moshizi, V. Sadeghi, P. Jahanbakhsh, Environmental-friendly mortar produced with treated and untreated coal wastes as cement replacement materials, Clean Technol. Environ. Policy 23 (10) (2021) 2843–2860.

[16] J. Esmaeili, H. Aslani, Use of copper mine tailing in concrete: strength characteristics and durability performance, J. Mater. Cycles Waste Manag. 21 (3) (2019) 729–741.

[17] C. Yun-hong, Y. Si-hui, Z. Jing-yu, S. Xiao-hui, Test research on hydration process of cement-iron tailings powder composite cementitious materials, Powder Technol. 399 (2022) 117215.

[18] A. Barzegar Ghazi, A. Jamshidi-Zanjani, H. Nejati, Utilization of copper mine tailings as a partial substitute for cement in concrete construction, Constr. Build. Mater. 317 (2022) 125921.

[19] Y. Agrawal, T. Gupta, Examining the effect of the particle modification on zinc tailing waste as a cementitious material in concrete: performance and toxicity assessment, J. Mater. Cycles Waste Manag. 24 (3) (2022) 1156–1171.

[20] K.P. Arunachalam, S. Avudaiappan, N. Maureira, F. Da Costa Garcia Filho, S. N. Monteiro, I.D. Batista, A.R.G. de Azevedo, Innovative use of copper mine tailing as an additive in cement mortar, J. Mater. Res. Technol. 25 (2023) 2261–2274.

[21] A.A. Aliabdo, A.E.M. Abd Elmoaty, E.M. Auda, Re-use of waste marble dust in the production of cement and concrete, Constr. Build. Mater. 50 (2014) 28–41.

[22] N.M. Sigvardsen, M.R. Nielsen, C. Potier, L.M. Ottosen, P.E. Jensen, Utilization of mine tailings as partial cement replacement, Modern Environ. Sci. Eng. 4 (4) (2018) 365–374.

[23] B. Chen, L. Pang, Z. Zhou, Q. Chang, P. Fu, Study on the hydration properties of a ternary cementitious material system containing activated gold tailings and granulated blast furnace slag, J. Build. Eng. 63 (2023) 105574.

[24] C.-R. Wu, Z.-Q. Hong, Y.-H. Yin, S.-C. Kou, Mechanical activated waste magnetite tailing as pozzolanic material substitute for cement in the preparation of cement products, Constr. Build. Mater. 252 (2020) 119129.

[25] B. Liu, H. Meng, G. Pan, H. Zhou, D. Li, Relationship between the fineness and specific surface area of iron tailing powder and its effect on compressive strength and drying shrinkage of cement composites, Constr. Build. Mater. 357 (2022) 129421.

[26] L. Cui, P. Chen, L. Wang, Y. Xu, H. Wang, Reutilizing waste iron tailing powders as filler in mortar to realize cement reduction and strength enhancement, Materials (2022).

[27] Y.J. Kim, Y.J. Kim, Y.W. Choi, An experimental research on self-consolidating concrete using tungsten Mine Tailings, KSCE J. Civ. Eng. 20 (4) (2016) 1404–1410.

[28] H. Wang, C. Ju, M. Zhou, F. zheng, Y. Dong, H. Hou, S. Liu, Grinding kinetics of lead–zinc tailing powders and its optimal particle size as a pozzolanic admixture in cement mortar, Adv. Powder Technol. 33 (9) (2022) 103730.

[29] Y. Liu, L. Jiang, J. Li, Q. Zhang, L. Yang, J. Cao, Study on the performance and hydration mechanism of concrete incorporating phosphorus tailings, J. Mater. Civ. Eng. 36 (9) (2024) 04024281.

[30] T. Luo, Y. Yi, Q. Sun, L.G. Li, L. Tang, C. Hua, The effects of adding molybdenum tailings as cementitious paste replacement on the fluidity, mechanical properties and micro-structure of concrete, J. Build. Eng. 62 (2022) 105377.

[31] X. Li, X. Gu, B. Liu, Z. Li, W. Zhang, J. Liu, M.L. Nehdi, Evaluation of waste powder from open pit mines as supplementary cementitious material: crystal structure and hydration characteristics, J. Build. Eng. 71 (2023) 106514.

[32] K. Zheng, J. Zhou, M. Gbozee, Influences of phosphate tailings on hydration and properties of Portland cement, Constr. Build. Mater. 98 (2015) 593–601.

[33] T. Luo, Y. Yi, F. Liu, Q. Sun, X. Pan, C. Hua, Early-age hydration and strength formation mechanism of composite concrete using molybdenum tailings, Case Stud. Constr. Mater. 16 (2022) e01101.

[34] K. Kunt, M. Yıldırım, F. Dur, E. Derun, S. Pişkin, Utilization of bergama gold tailings as an additive in the mortar, Celal Bayar Univer. J. Sci. 11 (3) (2015), 0.

[35] Y. Cheng, F. Huang, W. Li, R. Liu, G. Li, J. Wei, Test research on the effects of mechanochemically activated iron tailings on the compressive strength of concrete, Constr. Build. Mater. 118 (2016) 164–170.

[36] W. Zhu, R. Qiu, Y. Fu, Comparative Study on the Performance of Categorical Variable Encoders in Classification and Regression Tasks, 2024 09682. ArXiv abs/2401.

[37] S. Debbarma, G.D. Ransinchung, R. N, M. Dhaka, Effects of a Portland cement additive rich in SiO2 and Al2O3 in microstructure densification of RAP incorporated RCCP mixes, Constr. Build. Mater. 258 (2020) 119626.
[38] B. De Ville, Decision trees, Wiley Interdisciplinary Reviews, Comput. Stat. 5 (6) (2013) 448–455.
[39] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[40] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[41] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, Xgboost: extreme gradient boosting, R package version 0.4-2 1 (4) (2015) 1–4.
[42] J. Fan, X. Ma, L. Wu, F. Zhang, X. Yu, W. Zeng, Light Gradient Boosting machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data, Agric. Water Manag. 225 (2019) 105758.
[43] N. Hollmann, S. Müller, L. Purucker, A. Krishnakumar, M. Körfer, S.B. Hoo, R. T. Schirrmeister, F. Hutter, Accurate predictions on small data with a tabular foundation model, Nature 637 (8045) (2025) 319–326.
[44] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: a next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery, Anchorage, AK, USA, 2019, pp. 2623–2631.
[45] S. Shekhar, A. Bansode, A. Salim, A comparative study of hyper-parameter optimization tools. 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2021, pp. 1–6.
[46] S. Watanabe, Tree-structured parzen estimator: understanding its algorithm components and their roles for better empirical performance, arXiv preprint arXiv: 2304 (2023) 11127.
[47] I.K. Nti, O. Nyarko-Boateng, J. Aning, Performance of machine learning algorithms with different K values in K-fold cross-validation, Int. J. Inf. Technol. Comput. Sci. 13 (6) (2021) 61–71.
[48] C. Fan, N. Zhang, B. Jiang, W.V. Liu, Preprocessing large datasets using gaussian mixture modelling to improve prediction accuracy of truck productivity at mine sites, Arch. Min. Sci. 67 (No 4) (2022) 661–680.
[49] C. Fan, Z. Na, J. Bei, W.V. Liu, Prediction of truck productivity at mine sites using tree-based ensemble models combined with Gaussian mixture modelling, Int. J. Min. Reclamat. Environ. 37 (1) (2023) 66–86.
[50] C. Fan, N. Zhang, B. Jiang, W.V. Liu, Weighted ensembles of artificial neural networks based on Gaussian mixture modeling for truck productivity prediction at open-pit mines, Mining, Metallurg. Explorat. 40 (2) (2023) 583–598.
[51] C. Fan, N. Zhang, B. Jiang, W.V. Liu, Using deep neural networks coupled with principal component analysis for ore production forecasting at open-pit mines, J. Rock Mech. Geotech. Eng. 16 (3) (2024) 727–740.
[52] C.B. Arachchilage, C. Fan, J. Zhao, G. Huang, W.V. Liu, A machine learning model to predict unconfined compressive strength of alkali-activated slag-based cemented paste backfill, J. Rock Mech. Geotech. Eng. 15 (11) (2023) 2803–2815.
[53] C. Fan, A.C. Balasooriya, Z. Na, J. Bei, W.V. Liu, Machine learning with SHapley additive exPlanations for evaluating mine truck productivity under real-site weather conditions at varying temporal resolutions, Int. J. Min. Reclamat. Environ. 38 (10) (2024) 810–832.
[54] O.F. Ugurlu, C. Fan, B. Jiang, W.V. Liu, Deep neural network models for improving truck productivity prediction in open-pit mines, Mining, Metallurg. Explorat. 41 (2) (2024) 619–636.
[55] J. Zhao, C. Fan, G. Huang, Y. Guo, C.B. Arachchilage, R. Gupta, W.V. Liu, Machine learning-assisted characterization of the thermal conductivity of cement-based grouts for borehole heat exchangers, Constr. Build. Mater. 449 (2024) 138506.
[56] I.U. Ekanayake, D.P.P. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP), Case Stud. Constr. Mater. 16 (2022) e01059.
[57] C. Balasooriya Arachchilage, G. Huang, C. Fan, W.V. Liu, Forecasting unconfined compressive strength of calcium sulfoaluminate cement mixtures using ensemble machine learning techniques integrated with shapely-additive explanations, Constr. Build. Mater. 409 (2023) 134083.
[58] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[59] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. International Conference on Parallel Problem Solving from Nature, Springer, 2000, pp. 849–858.
[60] J. Zhao, G. Huang, Y. Guo, Z. Feng, R. Gupta, W.V. Liu, Development of a novel cement-based grout with enhanced thermal and sealing performance for borehole heat exchangers, Energy Build. 302 (2024) 113754.
[61] H.F. Taylor, Cement Chemistry, Thomas Telford London1997.
[62] W. Kurdowski, Cement and Concrete Chemistry, Springer Science & Business2014.
[63] E. Yilmaz, B. Koohestani, S. Cao, Chapter 13 - recent practices in mine tailings' recycling and reuse, in: C. Qi, C.H. Benson (Eds.), Managing Mining and Minerals Processing Wastes, Elsevier2023, pp. 271-304.
[64] Q. Zhang, H. Li, MOEA/D: a multiobjective evolutionary Algorithm based on decomposition, IEEE Trans. Evol. Comput. 11 (6) (2007) 712–731.
[65] Z. Cui, Y. Chang, J. Zhang, X. Cai, W. Zhang, Improved NSGA-III with selection-and-elimination operator, Swarm Evol. Comput. 49 (2019) 23–33.
[66] H. Zhou, H. Basarir, T. Poulet, W. Li, R.A. Kleiv, A. Karrech, Life cycle assessment of recycling copper slags as cement replacement material in mine backfill, Resour. Conserv. Recycl. 205 (2024) 107591.
[67] M. Sabău, D.V. Bompa, L.F.O. Silva, Comparative carbon emission assessments of recycled and natural aggregate concrete: environmental influence of cement content, Geosci. Front. 12 (6) (2021) 101235.