



---

*Research article*

## **Decision tree models for the estimation of geo-polymer concrete compressive strength**

**Ji Zhou<sup>1</sup>, Zhanlin Su<sup>2</sup>, Shahab Hosseini<sup>3</sup>, Qiong Tian<sup>1</sup>, Yijun Lu<sup>4</sup>, Hao Luo<sup>4</sup>, Xingquan Xu<sup>5</sup>, Chupeng Chen<sup>4,5,\*</sup> and Jiandong Huang<sup>4,\*</sup>**

<sup>1</sup> College of Civil and Environmental Engineering, Hunan University of Science and Engineering, Yongzhou 425199, China

<sup>2</sup> Shandong Energy Group Xinwen Mining Co., Ltd., Taian 271233, China

<sup>3</sup> Faculty of the Engineering, Tarbiat Modares University, Jalal AleAhmad, Nasr, Tehran, Iran

<sup>4</sup> School of Civil Engineering, Guangzhou University, Guangzhou 510006, China

<sup>5</sup> Guangdong Hualu Transport Technology Co., Ltd, Guangzhou, China

\* **Correspondence:** Email: 1112216034@e.gzhu.edu.cn, jiandong.huang@hotmail.com.

**Abstract:** The green concretes industry benefits from utilizing gel to replace parts of the cement in concretes. However, measuring the compressive strength of geo-polymer concretes (CSGPoC) needs a significant amount of work and expenditure. Therefore, the best idea is predicting CSGPoC with a high level of accuracy. To do this, the base learner and super learner machine learning models were proposed in this study to anticipate CSGPoC. The decision tree (DT) is applied as base learner, and the random forest and extreme gradient boosting (XGBoost) techniques are used as super learner system. In this regard, a database was provided involving 259 CSGPoC data samples, of which four-fifths of is considered for the training model and one-fifth is selected for the testing models. The values of fly ash, ground-granulated blast-furnace slag (GGBS), Na<sub>2</sub>SiO<sub>3</sub>, NaOH, fine aggregate, gravel 4/10 mm, gravel 10/20 mm, water/solids ratio, and NaOH molarity were considered as input of the models to estimate CSGPoC. To evaluate the reliability and performance of the decision tree (DT), XGBoost, and random forest (RF) models, 12 performance evaluation metrics were determined. Based on the obtained results, the highest degree of accuracy is achieved by the XGBoost model with mean absolute error (MAE) of 2.073, mean absolute percentage error (MAPE) of 5.547, Nash–Sutcliffe (NS) of 0.981, correlation coefficient (R) of 0.991, R<sup>2</sup> of 0.982, root mean square error (RMSE) of 2.458, Willmott's index (WI) of 0.795, weighted mean absolute percentage error (WMAPE) of 0.046, Bias of 2.073, square index (SI) of 0.054, p of 0.027, mean relative error (MRE) of -0.014,

and  $a^{20}$  of 0.983 for the training model and MAE of 2.06, MAPE of 6.553, NS of 0.985, R of 0.993,  $R^2$  of 0.986, RMSE of 2.307, WI of 0.818, WMAPE of 0.05, Bias of 2.06, SI of 0.056, p of 0.028, MRE of -0.015, and  $a^{20}$  of 0.949 for the testing model. By importing the testing set into trained models, values of 0.8969, 0.9857, and 0.9424 for  $R^2$  were obtained for DT, XGBoost, and RF, respectively, which show the superiority of the XGBoost model in CSGPoC estimation. In conclusion, the XGBoost model is capable of more accurately predicting CSGPoC than DT and RF models.

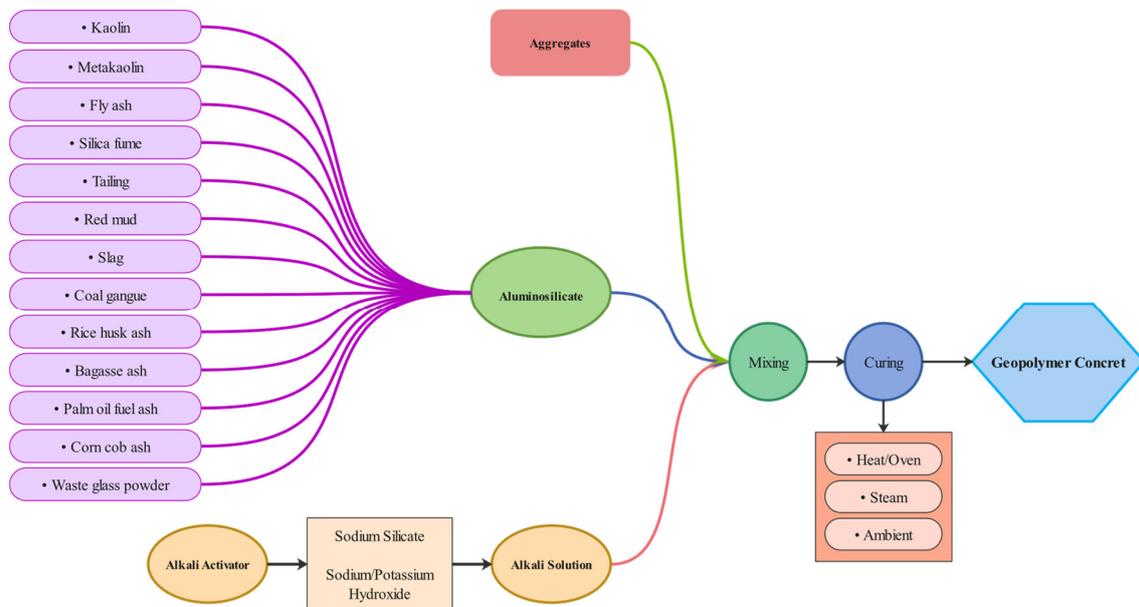
**Keywords:** geo-polymer concrete; compressive strength; super learner; extreme gradient boosting; decision tree; random forest

## 1. Introduction

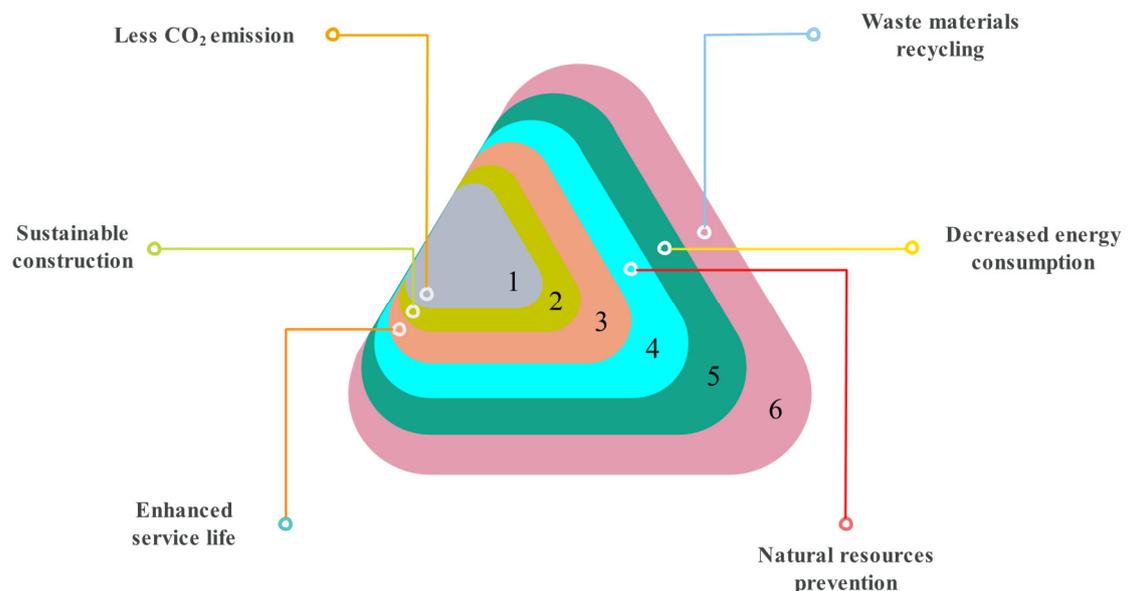
Cement concretes are the construction component that is employed on a global scale with the highest rate [1,2]. Ordinary Portland cement is one of the most common forms of binding agent used in cement concretes. Additional kinds of aggregate, water, and binding agents are other components of cement concretes. After aluminum and steel, oxidized polycyclic aromatic hydrocarbon is regarded as the third most incredibly energy-demanding chemical in the world. Ordinary Portland cement is responsible for seven percent of the overall energy that is needed by industries [3]. Unfortunately, the manufacturing of ordinary Portland cement results in the emission of enormous quantities of greenhouse gases like carbon dioxide, which has a significant role in the progression of global warming [4–6]. It is anticipated that the production of ordinary Portland cement would result in the release of 1,400,000,000 tons of greenhouse gases on a yearly basis [7,8]. Because of this, scientists have focused their attention on finding ways to lessen the amount of ordinary Portland cement that is used thanks to the development of alternative binders. There is some evidence that suggests that alkali-activated components, including geo-polymers, are preferred to cement concrete [9–12]. The reaction of precursor and activator results in the formation of alkali-activated compounds. In accordance with the amount of calcium present in the products of the chemical process, these were divided into two categories: 1) those that are high in calcium and have a  $Ca/(Si + Al)$  proportion that is higher than one (geo-polymers) and 2) those that are weak in calcium [13].

A geo-polymer is an innovative kind of binder produced for the production of concretes instead of ordinary Portland cement [13–17]. This was performed in order to improve the efficiency of production. The objective is to develop construction supplies that are sustainable-based, eco-friendly, and do not contain ordinary Portland cement. There is a significant increase in the number of distinct kinds of waste products that are being produced and deposited in landfills as a direct result of the ongoing expansion of industry and population. Rice husk ash, waste glass powder, ground granulated glass furnace slag, silica fume, fly ash, etc., are included in this category of wastes. It is detrimental to the environment to dispose of the mentioned wastes in landfills since they contribute to contamination in the environment [18,19]. Because geo-polymer concrete (GPoC) requires basic components with higher aluminum silicate concentrations present in waste materials, recycling these types of materials to produce GPoC would reduce the volume of pollutants that are released into the atmosphere [20,21]. Figure 1 provides a visual representation of the GPoC manufacturing procedure, in which a variety of different kinds of components and curing regimens are shown to be utilized during the making of GPoC. As can be seen in Figure 2, the utilization of such kinds of waste materials will be beneficial to

both the natural environment and the economic system. This is because these materials are abundant and the need for reasonably priced housing is expected to increase in tandem with the growth of populations [8,22–24]. In general, the use of GPoC for studies is becoming more common, and it has the potential to overtake other environmentally friendly construction materials [25,26]. Despite this, GPoC has an opportunity to produce a substantial contribution to the continued existence of cement concrete technologies as well as the construction sector in the years to come.



**Figure 1.** Schematic illustration of the geo-polymer construction strategy [27].



**Figure 2.** Advantages of geo-polymer concrete containing waste materials [28].

Recent advancements in artificial intelligence (AI) have provided an explanation for the widespread use of artificial intelligence techniques for anticipating the properties of a variety of

materials in civil engineering [29–35]. Also, varying AI techniques have been employed to predict the mechanical properties of engineering materials [36–41]. In a study that was conducted by Huang et al. [42], a comparison was conducted between three different artificial intelligence techniques known as decision trees (DT), AdaBoost, and bagging regressor in order to predict the compressive strength of GPoC (CSGPoC) that included fly ash materials. In comparison to the other systems examined, it proved that the bagging regressor approach showed the highest level of accuracy. In a separate study conducted by Ahmad et al. [43], artificial neural networks and gene expression programming (GEP) models were used to generate an estimate of the compressive strength of concretes that included recycled aggregates. In the study, the GEP model provided a more accurate forecast than the artificial neural network. A study conducted by Song et al. [44] used an artificial neural network approach to explore the compressive strength of concretes including waste materials, and they were able to correctly anticipate the needed conclusion. According to the findings of the study, it is possible to effectively use machine learning methods to anticipate any kind of mechanical feature that is associated with concretes. The tensile and compressive strengths of concretes with high performance were predicted using a number of artificial intelligence methods, as detailed by Nguyen et al. [45]. They concluded that the approaches of combined artificial intelligence were more accurate than the methods of artificial intelligence used in standalone form. This is due to the fact that the various machine learning techniques, in order to produce a more accurate model, frequently utilize the abilities of weak learners including decision trees and multi-layer perception neural networks. Therefore, several researchers have documented different artificial intelligence systems that have better degrees of precision in their evaluation of the attributes of materials. In light of this, it is absolutely necessary to carry out more in-depth research in order to shed light on this particular issue. Some literature models for predicting different characteristics of concretes are reported in Table 1.

**Table 1.** Using artificial intelligence techniques to predict various characteristics of concretes.

Author	Year	Technique	Number of data
Huang et al. [46]	2021	SVM	114
Sarir et al. [47]	2019	GEP	303
Balf et al. [48]	2021	DEA	114
Ahmad et al. [49]	2021	GEP, ANN, DT	642
Azimi-Pour et al. [50]	2020	SVM	-
Saha et al. [51]	2020	SVM	115
Hahmansouri et al. [52]	2020	GEP	351
Hahmansouri et al. [53]	2019	GEP	54
Aslam et al. [54]	2020	GEP	357
Farooq et al. [55]	2020	RF and GEP	357
Asteris and Kolovos [56]	2019	ANN	205
Selvaraj and Sivaraman [57]	2019	IREMSVM-FR with RSM	114
Zhang et al. [58]	2019	RF	131
Kaveh et al. [59]	2018	M5MARS	114
Sathyan et al. [60]	2018	RKSA	40
Vakhshouri and Nejadi [61]	2018	ANFIS	55
Belalia Douma et al. [62]	2017	ANN	114

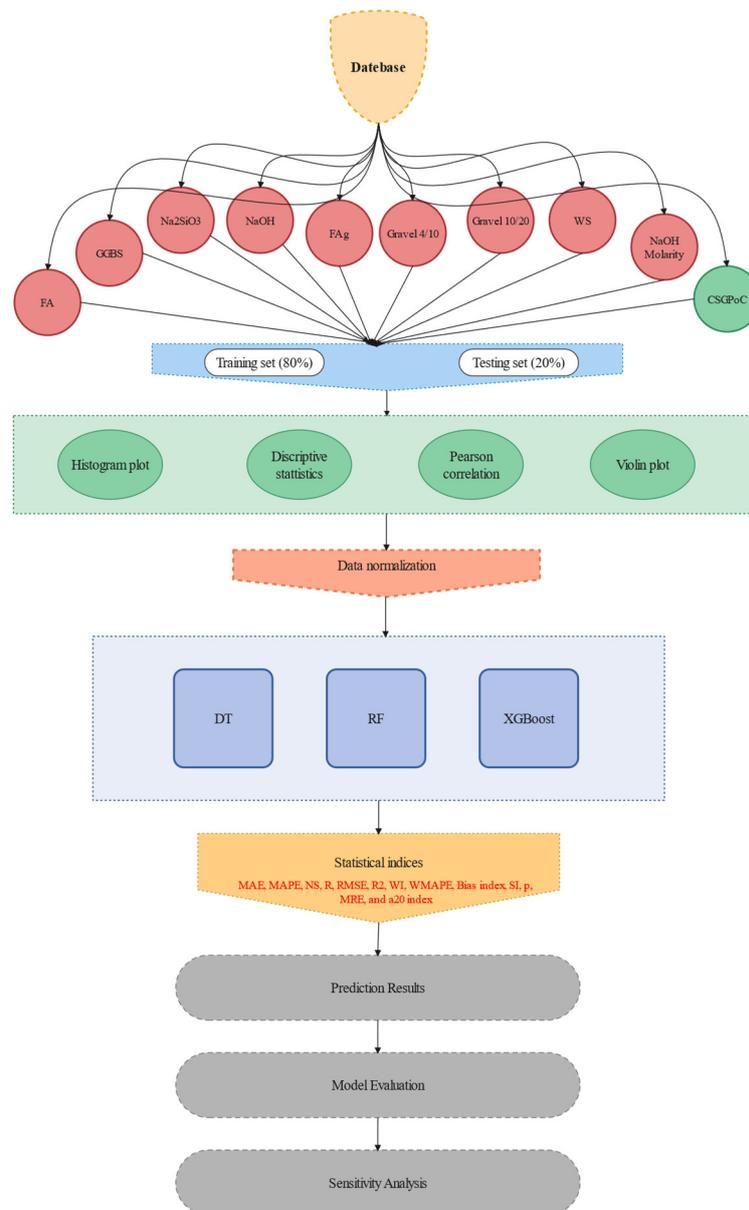
*Continued on next page*

Author	Year	Technique	Number of data
Abu Yaman et al. [63]	2017	ANN	69
Ahmad et al. [64]	2021	GEP, DT and Bagging	270
Farooq et al. [65]	2021	ANN, bagging and boosting	1030
Bušić et al. [66]	2020	MV	21
Javad et al. [67]	2020	GEP	277
Nematzadeh et al. [68]	2020	RSM, GEP	108
Güçlüer et al. [69]	2021	ANN, SVM, DT	100
Ahmad et al. [70]	2021	ANN, DT, GB	207
Asteris et al. [71]	2021	ANN, GPR, MARS	1030
Emad et al. [72]	2022	ANN, M5P,	306
Shen et al. [73]	2022	XGBoost, AdaBoost, and Bagging	372
Kuma et al. [74]	2022	GPR, SVMR	194
Jaf et al. [75]	2023	NLR, MLR, ANN	236
Mahmood et al. [76]	2023	NLR, M5P, ANN	280
Ali et al. [77]	2023	LR, MLR, NLR, PQ, IA, FQ	420

SVM: Support vector machine; GEP: Gene expression programming; ANN: Artificial neural network; DT: Decision tree; RF: Random Forest; DEA: Data envelopment analysis; RSM: Response surface methodology; ANFIS: Adaptive neuro fuzzy inference system; MV: Micali-Vazirani algorithm; RKSA: Retina key scheduling algorithm; GB: gradient boosting; GPR: Gaussian Process Regression; MARS: Multivariate Adaptive Regression Splines; SVMR: Support Vector Machine Regression; NLR: Nonlinear regression; MLR: Multi-linear regression; LR: linear regression; PQ: pure quadratic; IA: interaction; FQ: full quadratic.

This research differs from experimentation-based research in that it examines the CSGPoC using both base models of artificial intelligence methods as well as their ensemble form for predicting the CSGPoC. Experiment-based studies require considerable quantities of personal effort in addition to costly and lengthy experiments. By tackling the aforementioned challenges, using advanced technology such as artificial intelligence will help the building industry [13,78,79]. It is challenging to determine how several factors, such as precursor materials, activator solution, aggregates amount, and others, affect the strength of GPoC utilizing experimental procedures. Machine learning approaches may quickly and easily determine the combined impact of its constituent parts. Given that numerous studies have been accomplished to ascertain the determination of CSGPoC, machine learning models need a dataset, which could have been acquired from previous studies. Following data gathering, machine learning models can be trained to predict material attributes. Recent research has used machine learning techniques with a constrained set of effective parameters and databases to determine the intensity of GPoC. For instance, Dao et al.'s [80] use of machine learning approaches to forecast the CSGPoC employing three inputs and 210 data rows. Similar to this research, [81] employed 210 data rows and 4 inputs. In order to examine the effectiveness of various machine learning approaches used to anticipate the CSGPoC, the current study used nine effective parameters on CSGPoC and 295 data points based on literature review. The results of this study are also contrasted with those of related earlier investigations. The superior accuracy of machine learning approaches is anticipated to come from employing more input parameters and data points. The main goal of the present work is to identify the best machine learning method for calculating the CSGPoC using anticipated results and the impact

of different parameters on GPoC strength. The computational flowchart of the study is depicted in Figure 3.

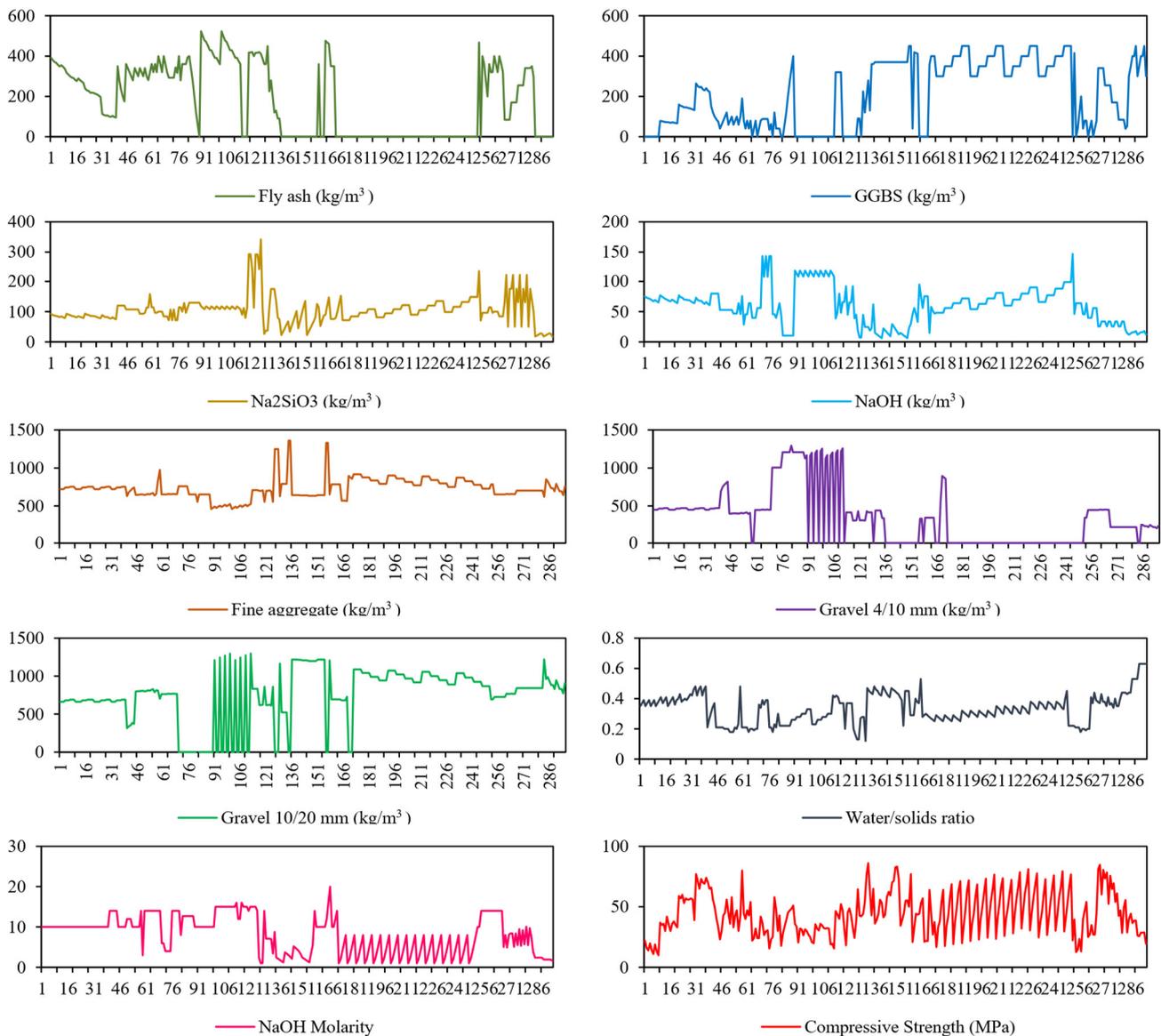


**Figure 3.** The flowchart of the study.

## 2. Data analysis

The composition of the various alkali excitation components and solid waste from the industry that are used to make gels are the basic components, and the amount of these initial components used in the production of gels have an effect on the efficiency of the gels [24,79,82,83]. It is necessary to have adequate Na<sup>+</sup> and OH in order to finish every step of the polymerization of gels, and the amount of both of these ions has a direct bearing on the amount of force that can be exerted by the gels [84]. In light of the two aforementioned explanations, as well as the impact that the properties and ratios of the initial materials have on the compressive strength of concretes, the authors of the present study

came to the conclusion that GGBS, sodium silicate, fly-ash, gravel stones (4–10 mm to 10–20 mm), water/solids proportion, sodium hydroxide, sodium hydroxide molarity, and fine aggregates are the effective parameters to determine and predict the CSGPoC to complete a dataset comprising 295 data points. The required CSGPoC data of developing models is gathered from a study conducted by Yong et al. [28]. This data includes nine parameters as inputs and CSGPoC as the output of the models. Statistical analysis of the inputs and output is reported in Table 2. Furthermore, a better view for variations of inputs as well as CSGPoC in 295 various points are demonstrated in Figure 4.



**Figure 4.** The variation of input variables and CSGPoC for 295 data points.

**Table 2.** Statistical analysis of effective parameters datasets.

Parameter	Symbol	Unit	Median	Min	Mean	Max	StD	
Fly ash	FA	kg/m <sup>3</sup>	170	0	178.265	523	173.979	
GGBS	GGBS	kg/m <sup>3</sup>	225	0	209.831	450	163.271	
Na <sub>2</sub> SiO <sub>3</sub>	Na <sub>2</sub> SiO <sub>3</sub>	kg/m <sup>3</sup>	100	18	104.059	342	44.9000	
NaOH	NaOH	kg/m <sup>3</sup>	64	6.300	60.042	147	30.391	
Inputs	Fine aggregate	FAg	kg/m <sup>3</sup>	721	459	731.209	1360	138.078
	Gravel 4–10 mm	Gravel 4–10	kg/m <sup>3</sup>	309	0	335.828	1293.400	373.884
	Gravel 10–20 mm	Gravel 10–20	kg/m <sup>3</sup>	815	0	741.556	1298	361.336
	Water/solids ratio	WS	N/A	0.330	0.120	0.330	0.630	0.095
	NaOH molarity	NaOH molarity	N/A	10	1	8.193	20	4.596
Output	Compressive strength of geo-polymer concrete	CSGPoC	MPa	43	10	44.474	86.080	18.010

Before model developing, the correlation coefficient between parameters should be evaluated [13,85–87]. If the correlation between two parameters is high, the multicollinearity problem appears in the model. Therefore, the Spearman correlation coefficient between CSGPoC and effective parameters is calculated as shown in Figure 5. This figure is a heatmap of Spearman correlation coefficient that can be determined by following equation [88–91]:

$$r = \frac{\sum_{i=1}^n (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_{i=1}^n (x_i - x_m)^2 \times \sum_{i=1}^n (y_i - y_m)^2}} \quad (1)$$

where  $n$ ,  $x_m$ , and  $y_m$  stand for the number of datasets, average value across all  $x$  data, and the average value across all  $y$  data, respectively. When the  $r$  value is  $r > 0$ ,  $r = 0$ ,  $r \approx 1$ ,  $r < 0$ , or  $r \approx -1$ , then there is positive linear correlation, no correlation, stronger positive linear correlation, negative linear correlation, or stronger negative linear correlation, respectively [8,87,92]. From Figure 5, there exist medium negative and positive linear correlations between CSGPoC and FA and GGBS at -0.43 and 0.46, respectively. Moreover, the correlation between CSGPoC and NaOH, Gravel 4–10 mm, and NaOH molarity is a weak negative linear correlation with  $r$  equal to -0.22, -0.27, and -0.12, respectively, and the correlation between CSGPoC and other parameters is a weak positive linear correlation. Based on these results, the investigation of developed models impact will not be influenced by the occurrence of multicollinearity between effective parameters on CSGPoC. In the following, the violin plot of parameters is showed in Figure 6. In this figure, the median, Q1, Q2, Q3, minimum, and maximum values of parameters are presented.

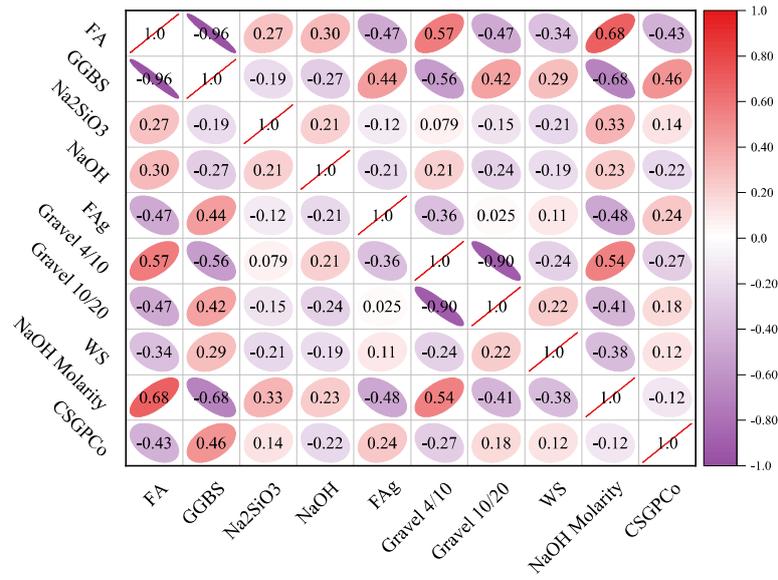
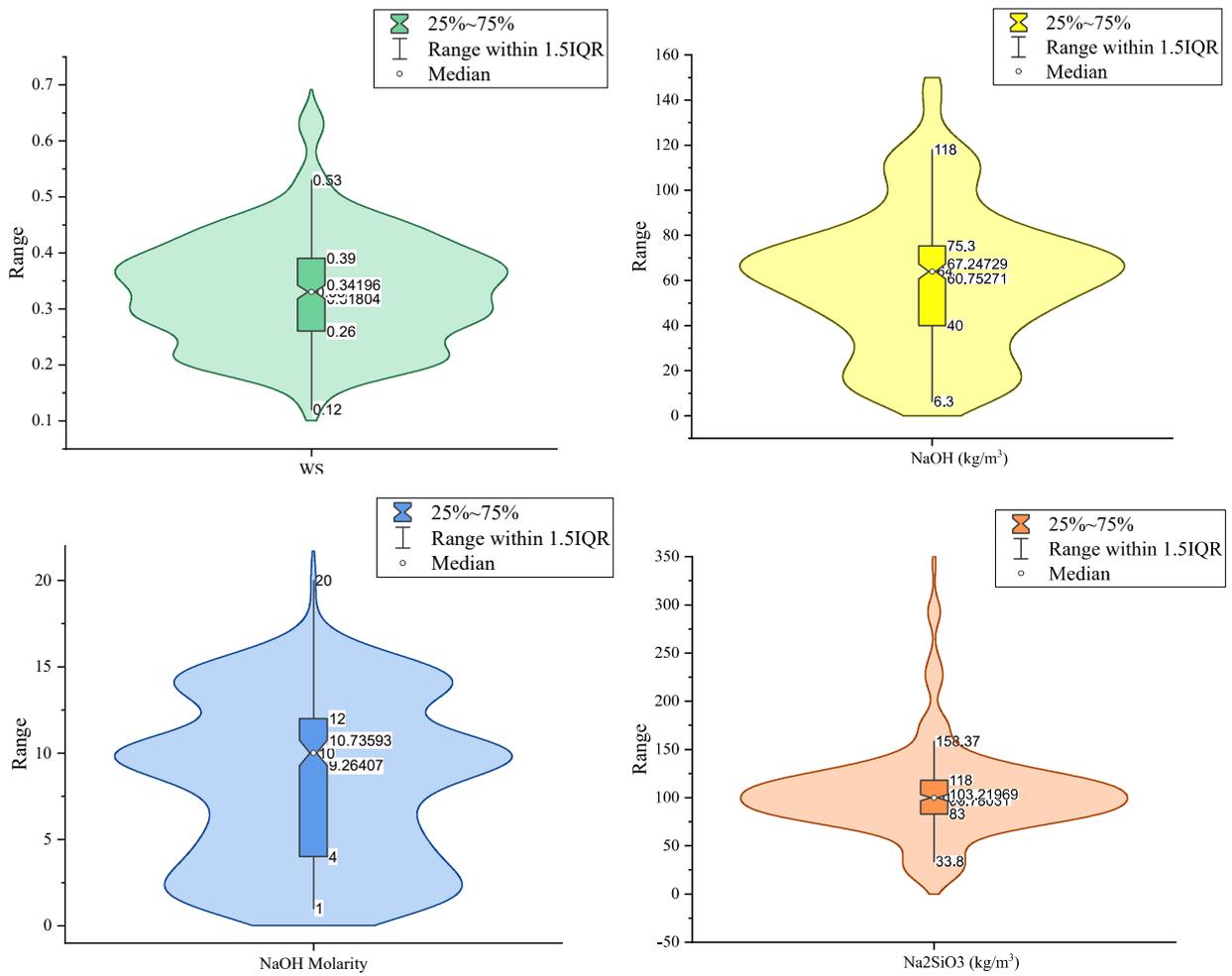
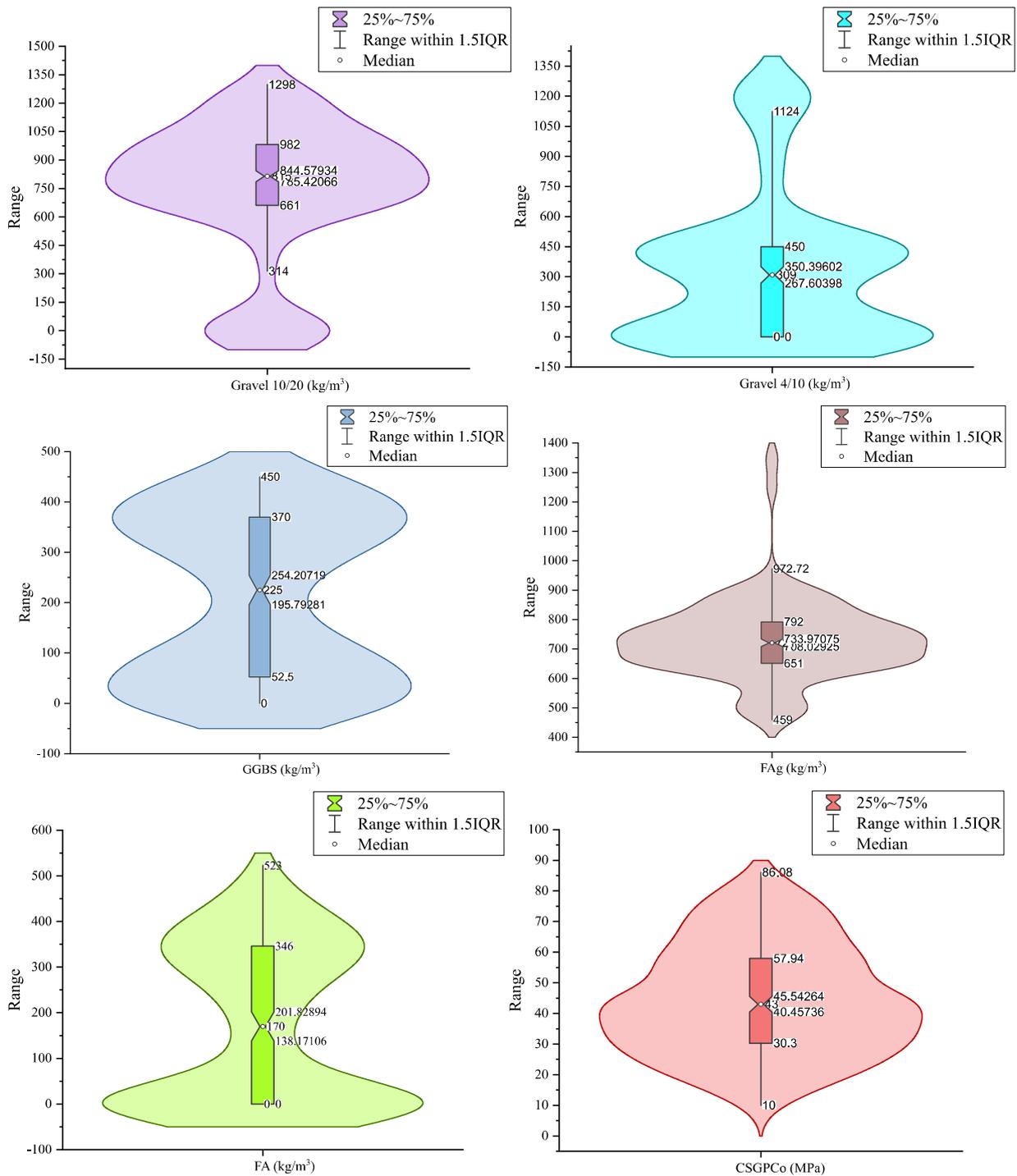


Figure 5. Heatmap plot of effective parameters.



Continued on next page



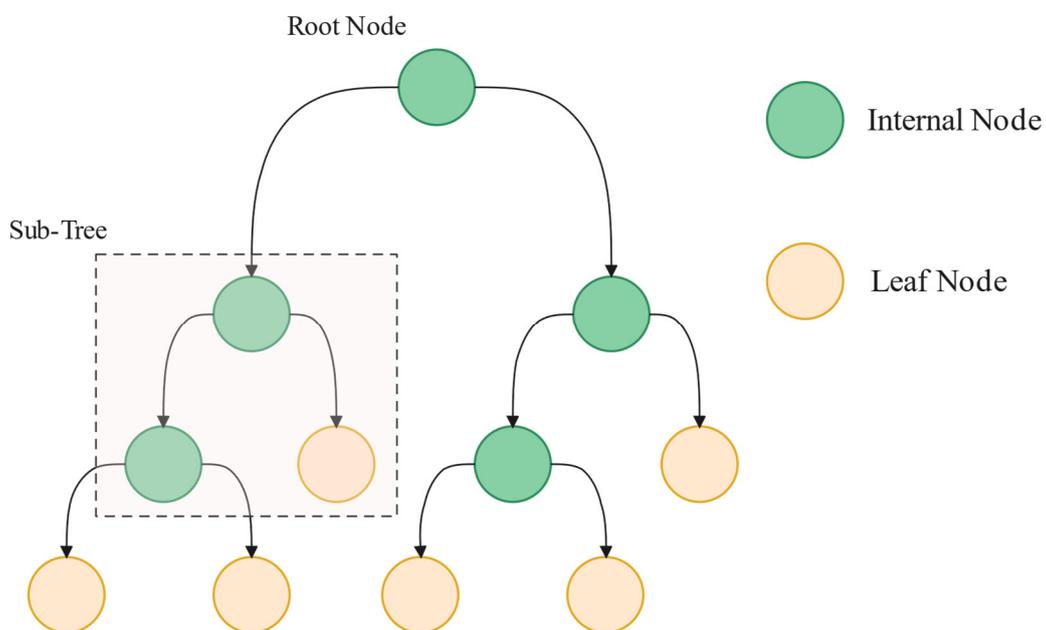
**Figure 6.** Violin plot of effective parameters.

### 3. Materials and methods

#### 3.1. Decision tree

The DT is an artificial intelligence method widely employed for classifying issues, including those involving regression. Classes are included inside the trees. On the other hand, if there is not already a

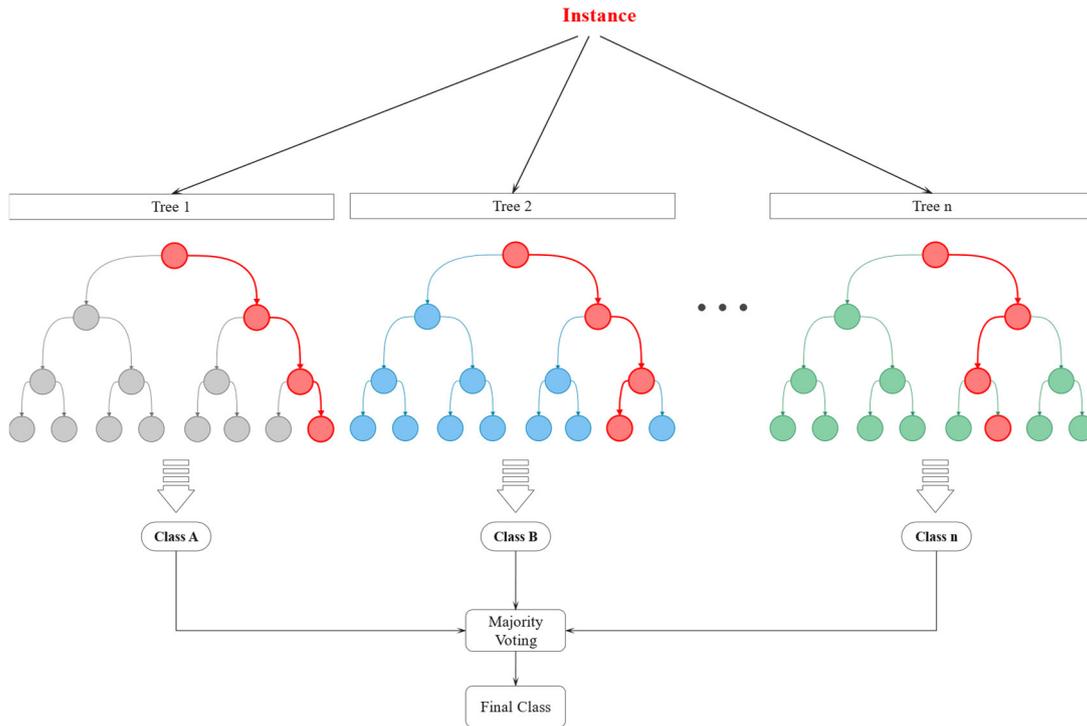
class for the data, researchers can employ the regression approach to develop predictions about the result based on the effective parameters [13,87,93]. A DT is a hierarchical classification algorithm, and the internal node of a DT is equivalent to databases properties. The branches of the tree reflect the results of the regulations, and the leaf nodes stand for different outcomes. A DT is constructed from two different nodes: the decision nodes and the leaf nodes. Leaf nodes do not possess branches and are regarded as the decision's result. In contrast, a decision node is capable of making a choice since they contain multiple decision-making branches. As its name implies, a DT is a kind of data architecture of trees, with a root node and it increases in size based on the number of branches [94]. The DT splits the data points into different sections. The target and the projected numbers are compared at every splitting point, and the difference is determined by the procedure being carried out. The error values are calculated at each division point, and the parameter with the least fit function is selected as a division point. The operation is then iterated as necessary. Figure 7 presents the DT flowchart.



**Figure 7.** An example of DT architecture.

### 3.2. Random forest

Breiman [95] first put out the idea of using the RF technique, which is a common approach to soft computing. The RF method relies on decision-trees computations, and it has the ability to assemble numerous decision trees into a complicated structure in order to arrive at a conclusion on the classifications or regressions that have been presented to them. Throughout this phase of computation, the DTs that create the RF architecture receive training by randomly choosing parameters and data points from the primary CSGPoC database. Breiman [95] and Liaw and Wiener [96] both provide thorough overviews of the RF technique. The conceptual view of the RF model is demonstrated in Figure 8.



**Figure 8.** A schematic representation of RF model.

### 3.3. eXtreme-Gradient Boosting (XGBoost)

XGBoost is a recommendation system developed by Chen et al. [97]. The lifting technique consists of training many of the base models (learners) using specific approaches, such as the simple decision trees that have low depth, and then combining the forecasting outcomes of these weak base learners using specific techniques in order to significantly enhance the estimation impact [8,98]. As its weak learners, XGBoost employs regressive trees with a short depth. Let the learners that were acquired in the first phase be  $y;0'$ . This will apply to the initial shallow regression tree models that were created during the training phase. It is assumed that the resulting model appears as  $F_0(t)$  and that  $t$  denotes the instance vector in the space of features. XGBoost continues by computing and obtaining the first and second derivatives, which are  $h_i$  and  $g_i$ , of the loss function of the errors among the classifiers. This is done after the function has been evaluated. The value that was anticipated in the process before this one, which was  $m - 1$ , the objective function of  $F_m(t)$ , may be calculated using the second-order expansions of the Taylor function as follows:

$$Obj^{(m)} = \sum_{i=1}^N \left[ g_i F_m(t_i) + \frac{1}{2} h_i F_m^2(t_i) \right] + \Omega(F_m) \quad (2)$$

The regularizing component is denoted by the symbol  $\Omega(F_m)$ , and its purpose is to prevent the technique from unnecessarily enhancing the degree of complexity of the models in an effort to enhance its precision, which would result in the overfitting problem.  $(F_m)$  is equivalent to the following expressions:

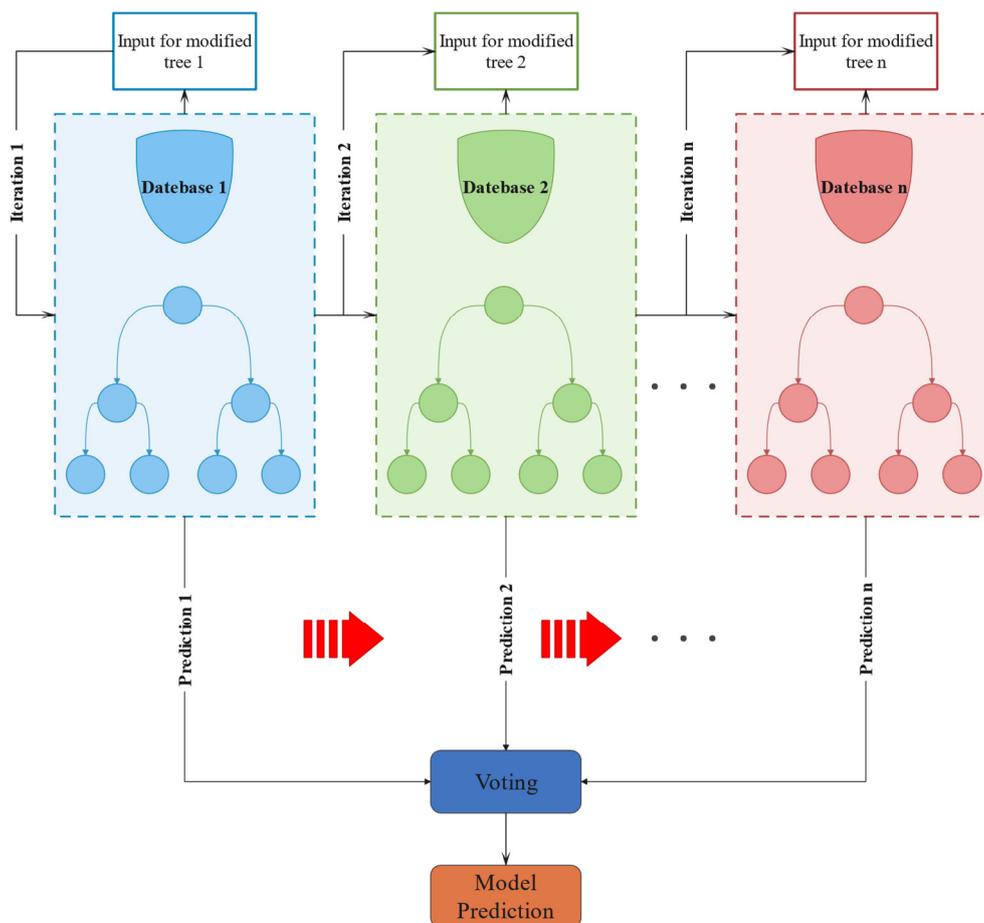
$$\Omega(F_m) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

In which  $\gamma$  and  $\lambda$  stand for penalty coefficients,  $w$  represents the weighting of the regression leaves node,  $T$  indicates the number of regression leaves node, and  $\|w\|^2$  shows the effect that the weight of

the regression leaves node has on the level of complexity of the models. Equation (3) reveals that the fitted objective in each repetition of the XGBoost objective function is the difference between the value that was anticipated and the actual value of the data. This can be observed by comparing the projected values to the actual values. The goal of the training procedure is to reduce  $obj(m)$  to the smallest possible value. The MSE value may be used by the regression trees node dividing to choose which dividing features to use. Subsequently, it is possible to generate an additional shallower trees model called  $F_m(t)$ , and the learner may be updated as follows:

$$y'_m(t) = y'_{m-1}(t) + F_m(t) \quad (4)$$

The XGBoost regression flowchart is illustrated in Figure 9.



**Figure 9.** Voting process in the XGBoost structure.

#### 4. Date preparation

Before developing machine learning models and predicting CSGPoC, two main steps are implemented. The closer the distance between the inputs and CSGPoC parameters is, the machine learning techniques can better learn the relationships among parameters. Also, the training of machine learning techniques is only performed based on the parameters' values, not their unit. Therefore, the input values should be normalized in the range  $[0,1]$  in the first step of pre-analysis using Eq (5) to achieve a rational output [99–101].

$$x_{norm} = \frac{(x_i - x_{maximum})}{(x_{maximum} - x_{minimum})} \quad (5)$$

In which  $x_{norm}$ ,  $x_i$ ,  $x_{maximum}$ , and  $x_{minimum}$  signify the standardized value, actual value, maximum value, and minimum value, respectively [102].

In the second step of pre-analysis, the data points were randomly categorized into two main phases: training samples (80% of whole CSGPoC, 236 data out of 295 CSGPoC data) and testing samples (20% of whole CSGPoC, 59 data out of 295 CSGPoC data). Then the train samples were applied for models learning and the testing samples were applied for the evaluation of models performance.

Overfitting is a common challenge in machine learning models, where a model performs exceptionally well on the training data but struggles with unseen data. To address this concern, the study has meticulously incorporated a safeguard in its model development process. They employed rigorous cross-validation techniques, ensuring that the model's performance is evaluated on diverse subsets of the dataset. Additionally, the authors applied regularization methods, such as dropout or weight decay, to prevent the model from becoming overly complex and fitting noise in the data. Furthermore, the use of a diverse and representative dataset, along with extensive hyperparameter tuning, contributes to the generalization capability of the model [103].

## 5. Results and discussion

The construction of the base learner (DT) and super learner (RF and XGBoost) predictive models on 295 CSGPoC data is highlighted and discussed in this section. The performance and efficiency of the developed DT, XGBoost, and RF models were evaluated utilizing 12 statistical metrics involving mean absolute error (MAE), mean absolute percentage error (MAPE), Nash–Sutcliffe (NS), correlation coefficient (R), root mean square error (RMSE),  $R^2$ , Willmott's index (WI), weighted mean absolute percentage error (WMAPE), bias index, square index (SI), p, mean relative error (MRE), and a20 index [98,104–109]. The determination of these metrics can be performed by the following equations. It should be noted that the performance of the developed predictive models is analyzed and described using scatter plots, ribbon charts, violin plots, Taylor diagrams, and error plots.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O - P)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(P - O)| \quad (7)$$

$$R^2 = \frac{\sum_{i=1}^r (O - \bar{O})^2 - \sum_{i=1}^r (O - P)^2}{\sum_{i=1}^r (O - \bar{O})^2} \quad (8)$$

$$R = \frac{\sum (O_i - \bar{P})(P_i - \bar{P})}{\sqrt{\sum (O_i - \bar{P})^2 \sum (P_i - \bar{P})^2}} \quad (9)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{O - P}{O} \right| * 100 \quad (10)$$

$$WMAPE = \frac{\sum_{i=1}^n \left| \frac{O - P}{O} \right| * \alpha}{\sum_{i=1}^n O} \quad (11)$$

$$NS = 1 - \frac{\sum_{i=1}^n (O - P)^2}{\sum_{i=1}^n (O - \bar{P})^2} \quad (12)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (P - O)^2 \quad (13)$$

$$SI = \frac{RMSE}{\frac{1}{n} \sum_{i=1}^n O} \quad (14)$$

$$\rho = \frac{SI}{1+R} \quad (15)$$

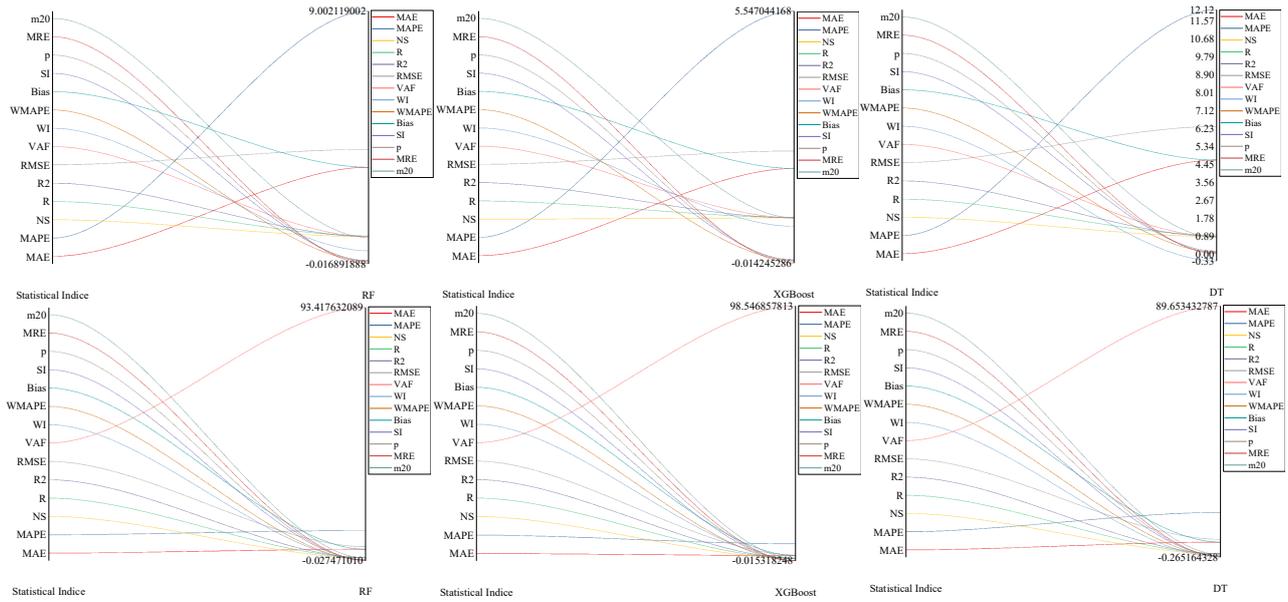
$$MRE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|O-P|}{O} \right) \quad (16)$$

$$a^{20} - index = \frac{m^{20}}{M} \quad (17)$$

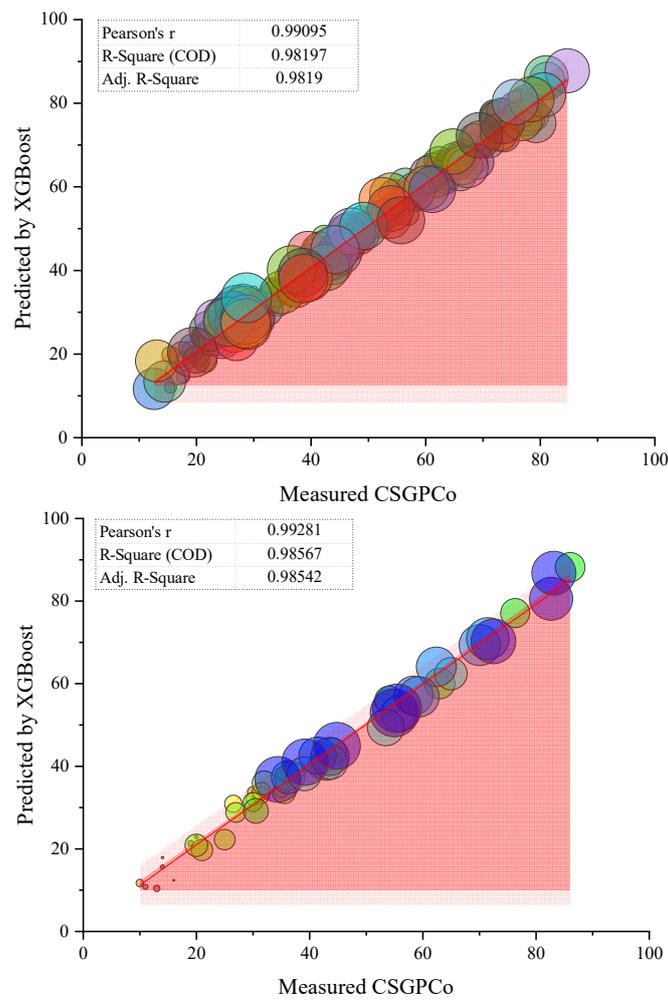
where  $O_i$  and  $P_i$  are the measured and anticipated  $i$ th CSGPoC values respectively,  $n$  stands the number data points,  $\bar{O}$  is the mean of the measured CSGPoC, and  $\bar{P}$  is the mean of the predicted CSGPoC. Moreover,  $m^{20}$  and  $M$  indicate the total number of data points and the number of points with a ratio “measured value” over “predicted value” between 0.80 and 1.20 respectively [110–112].

Some evaluation metrics including error indices of MRE, RSME, MAPE, MAE, SI,  $p$ , WI, bias, and WMAPE are applied to error analysis and evaluation of the relationships between the measured CSGPoC and predicted one with base learner and super learner models. The minimum value for error indices reveals the highest prediction capability. The  $R^2$ , NS, and  $R$  determine model precision within a range of 0–1, and an amount higher than 0.95 for these metrics indicates that the proposed models present a highly reliable and accurate prediction. The obtained performance evaluation metrics of all the developed RF, XGBoost, and RF models are summarized in Figure 10. The predictive model can be specified as the most accurate system when the errors of MRE, RSME, MAPE, MAE, SI,  $p$ , WI, bias, and WMAPE are the lowest, and the values of accuracy of  $R^2$ , NS, and  $R$  are higher. From Figure 10, the XGBoost model presents the highest performance prediction level based on the evaluation metrics. The highest degree of accuracy yielded by the XGBoost model achieved MAE of 2.073, MAPE of 5.547, NS of 0.981,  $R$  of 0.991,  $R^2$  of 0.982, RMSE of 2.458, WI of 0.795, WMAPE of 0.046, bias of 2.073, SI of 0.054,  $p$  of 0.027, MRE of -0.014, and  $a^{20}$  of 0.983 for the training model and MAE of 2.06, MAPE of 6.553, NS of 0.985,  $R$  of 0.993,  $R^2$  of 0.986, RMSE of 2.307, WI of 0.818, WMAPE of 0.05, bias of 2.06, SI of 0.056,  $p$  of 0.028, MRE of -0.015, and  $a^{20}$  of 0.949 for the testing model. Furthermore, the highest  $R^2$  (0.9819 for training part and 0.9857 for testing part) is achieved by the XGBoost model, while the lowest  $R^2$  (0.8859 for training part and 0.8969 for testing part) is obtained by the DT model. Therefore, the DT model has the worst performance and accuracy concerning the evaluation indices with MAE of 4.666, MAPE of 12.117, NS of 0.871,  $R$  of 0.941,  $R^2$  of 0.886, RMSE of 6.333, WI of -0.335, WMAPE of 0.103, bias of 4.666, SI of 0.143,  $p$  of 0.074, MRE of 0.029, and  $a^{20}$  of 0.826 for the training set and MAE of 5.103, MAPE of 15.736, NS of 0.896,  $R$  of 0.947,  $R^2$  of 0.897, RMSE of 6.111, WI of -0.265, WMAPE of 0.125, bias of 5.103, SI of 0.149,  $p$  of 0.076, MRE of -0.036, and  $a^{20}$  of 0.729 for the testing set.

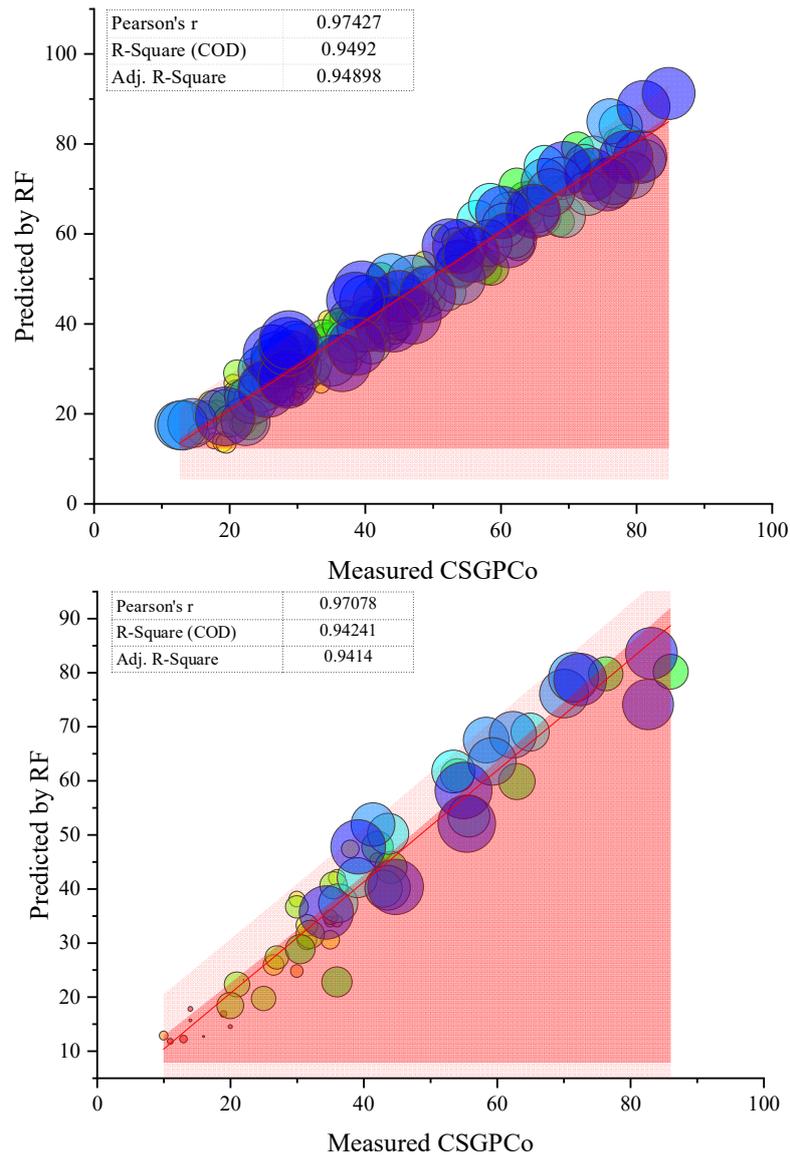
The obtained  $R^2$  and correlation between measured and predicted CSGPoC by XGBoost, RF, and DT is illustrated in Figures 11–13, respectively. As can be seen, the highest  $R^2$  value is relevant to the XGBoost model in both the training and testing sets.



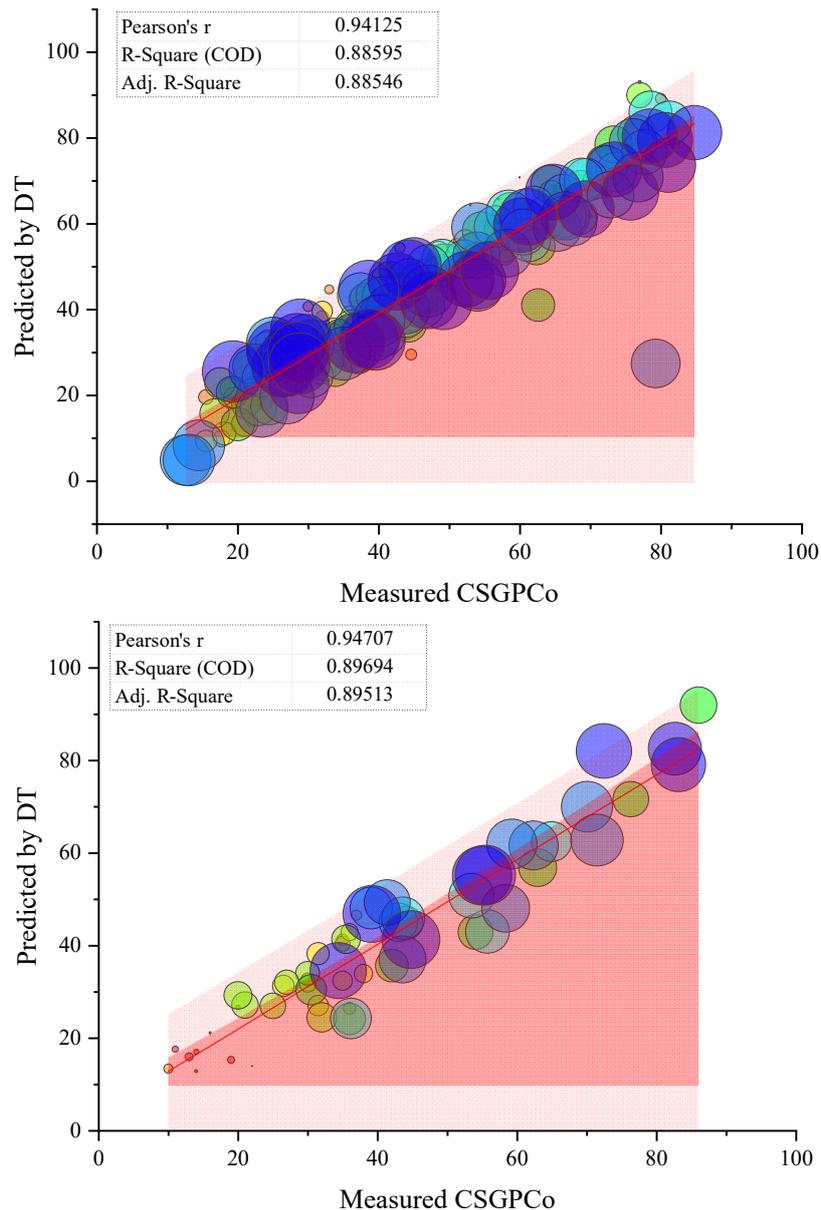
**Figure 10.** Ribbon chart for training (above) and testing (below) models.



**Figure 11.** The obtained  $R^2$  and correlation between measured and predicted CSGPoC by the XGBoost model.

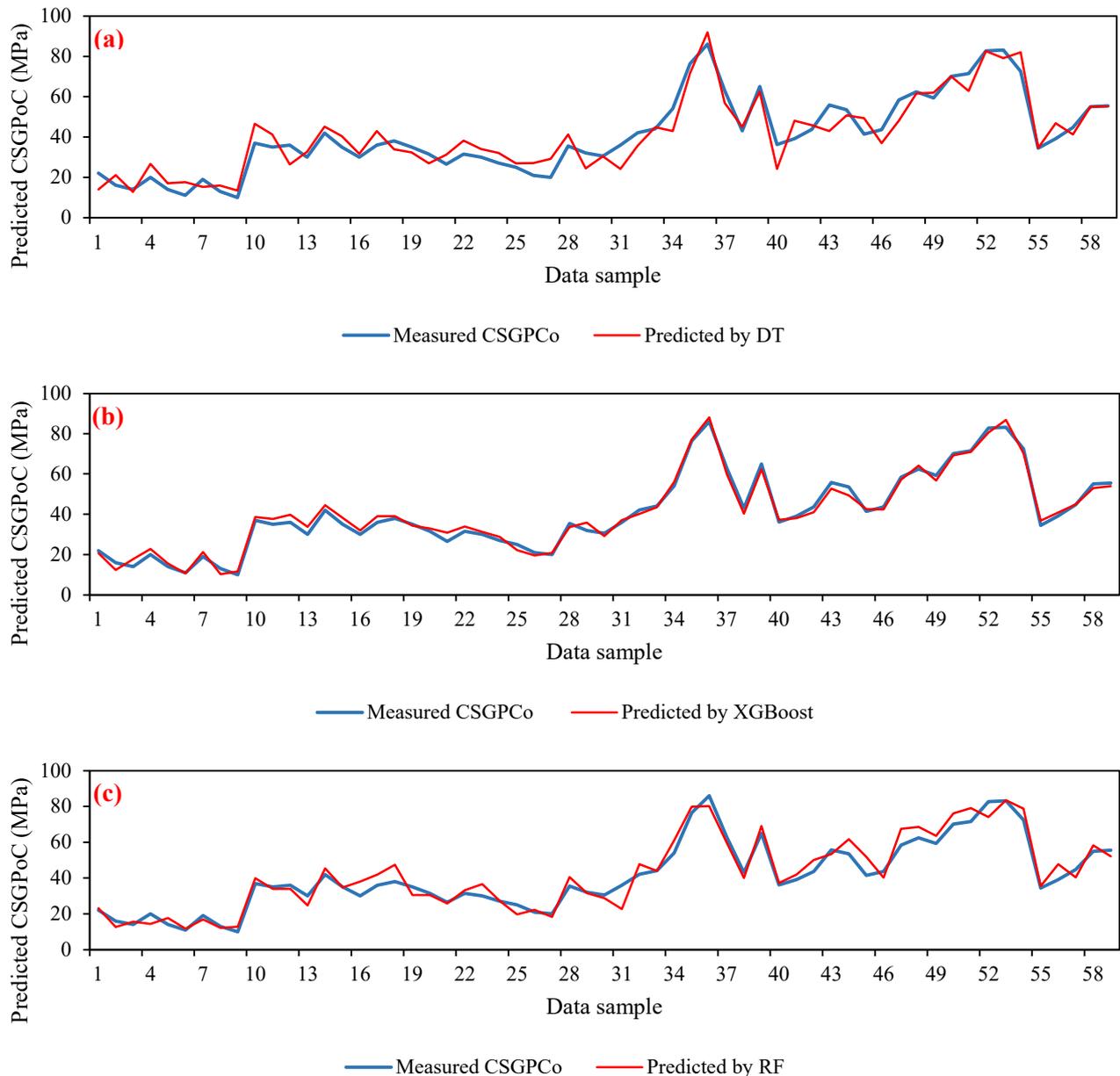


**Figure 12.** The obtained  $R^2$  and correlation between measured and predicted CSGPoC by the RF model.



**Figure 13.** The obtained  $R^2$  and correlation between measured and predicted CSGPoC by the DT model.

In the following, the performance of the models is evaluated by adding testing sets to trained models. The capability of the prediction performance is specified by testing data. An acceptable value for statistical metrics in the training phase does not mean that the models are predicting correctly, and the efficiency of the models should be evaluated using test data. If the model error is reduced by importing the test data into modeling process, it can be concluded that the model training is not performed well and the model is not able to predict the CSGPoC values in the real world. The testing results are revealed in Figure 14. It can be seen that the prediction of CSGPoC is conducted correctly, as the results of the testing phase are approximately close to the results of the training phase for all DT, RF, and XGBoost models. Specially, the results of the XGBoost model are acceptable and its accuracy is confirmed for predicting CSGPoC in other projects.



**Figure 14.** Testing results of predicted CSGPoC by (a) DT, (b) XGBoost, and (c) RF models.

The distribution of error for all models is revealed from the violin plot, which is demonstrated in Figure 15. The violin diagram shows the range of errors relevant to the DT, XGBoost, and RF models. From Figure 15, it can be found that the range of errors in the XGBoost model is lower for both the training and testing phases compared to the RF and DT models. Of the models, the DT model has the lowest accuracy due to its higher error ranges on the violin plot. The schematic demonstration of the standard deviation, coefficient determination, and RMSE values can be shown using Taylor's diagram, which is displayed for the training and testing parts of the models in Figure 16. This figure assigns the best-fitted model. In this figure, the red dashed line represents the standard deviation of the data. As depicted in Figure 16, both the standard deviation and  $R^2$  are close to 1 for reference point. The XGBoost, DT, and RF models are illustrated in this figure using green, orange and purple colored squares. The XGBoost symbol is very close to the reference symbol (red square), which shows that

the XGBoost model reflects the reality-based results and presents better results in terms of performance and precision values. It can be concluded that although the accuracy of the DT and RF models is acceptable, XGBoost is the superior model for predicting CSGPoC.

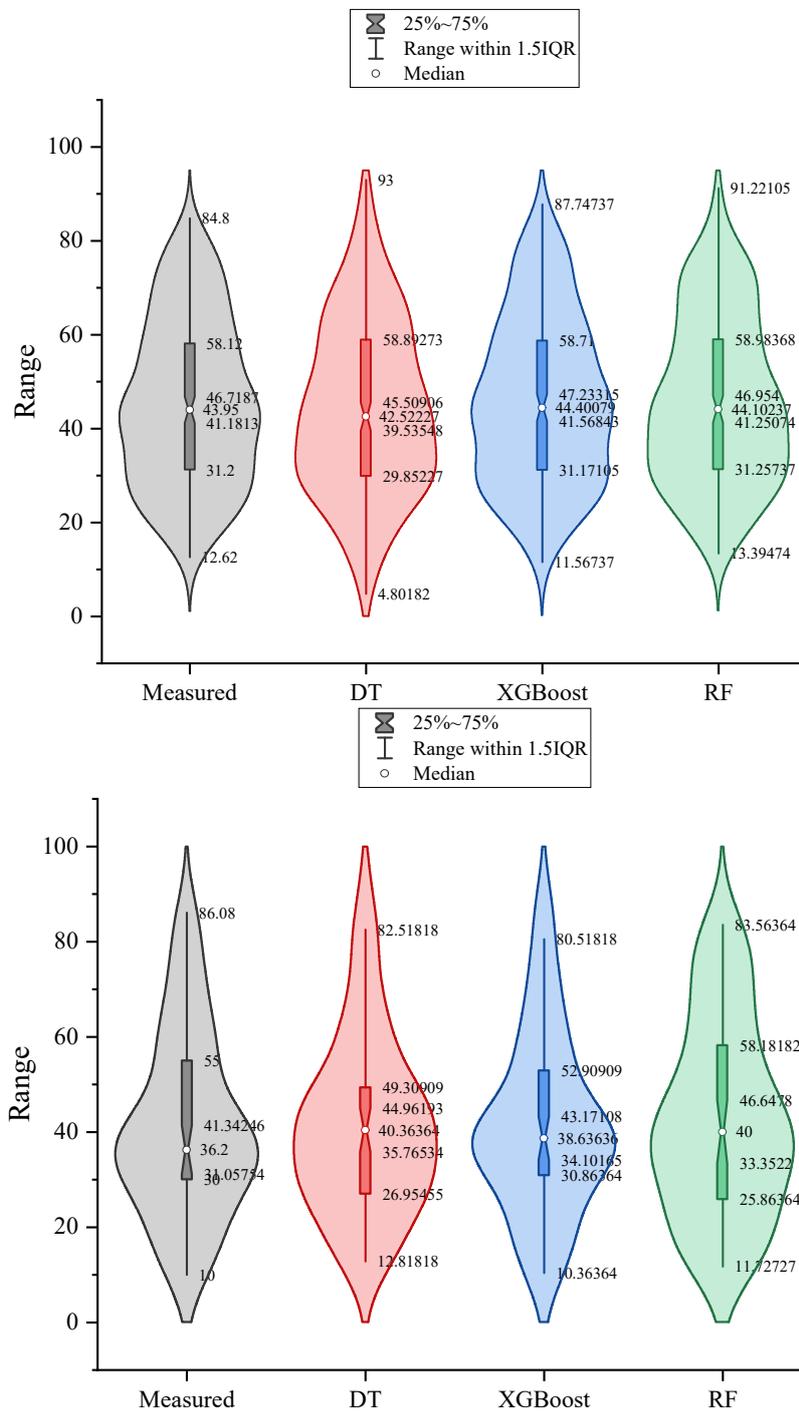
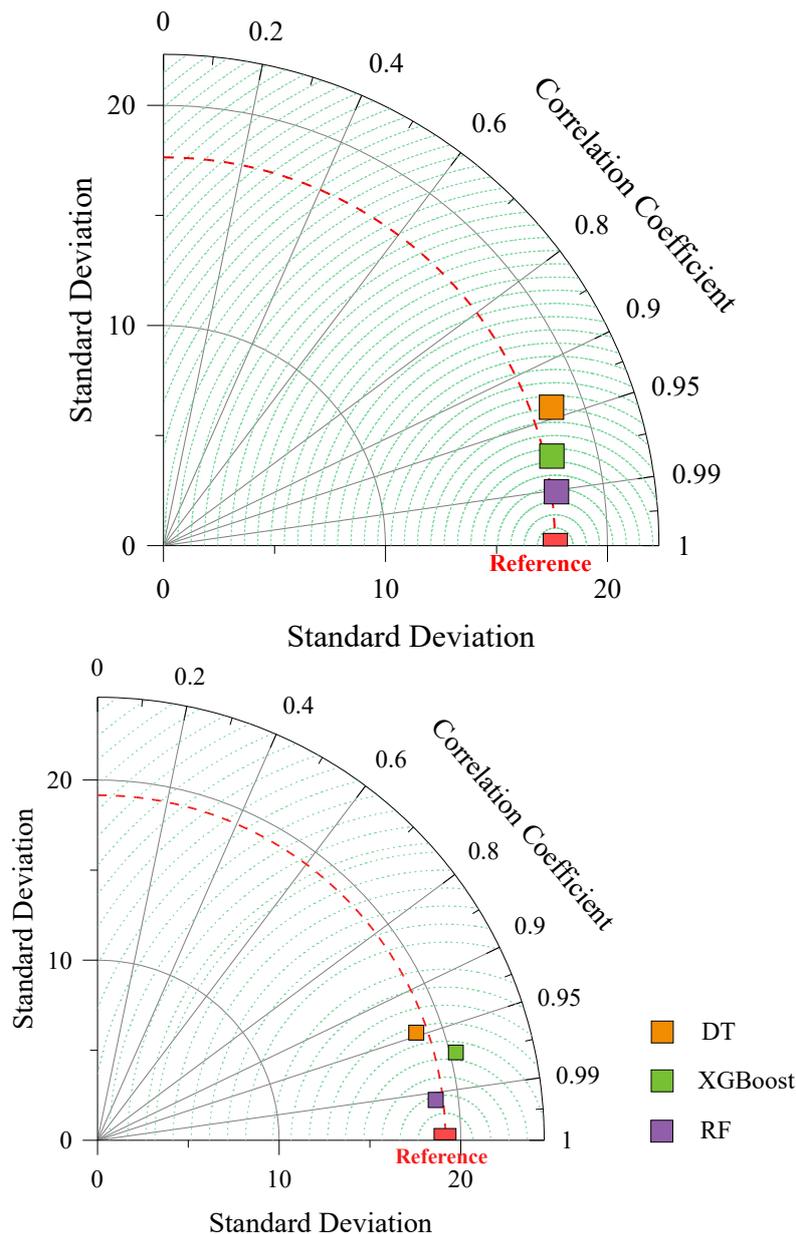


Figure 15. Testing results of predicted CSGPoC by DT, XGBoost, and RF models.



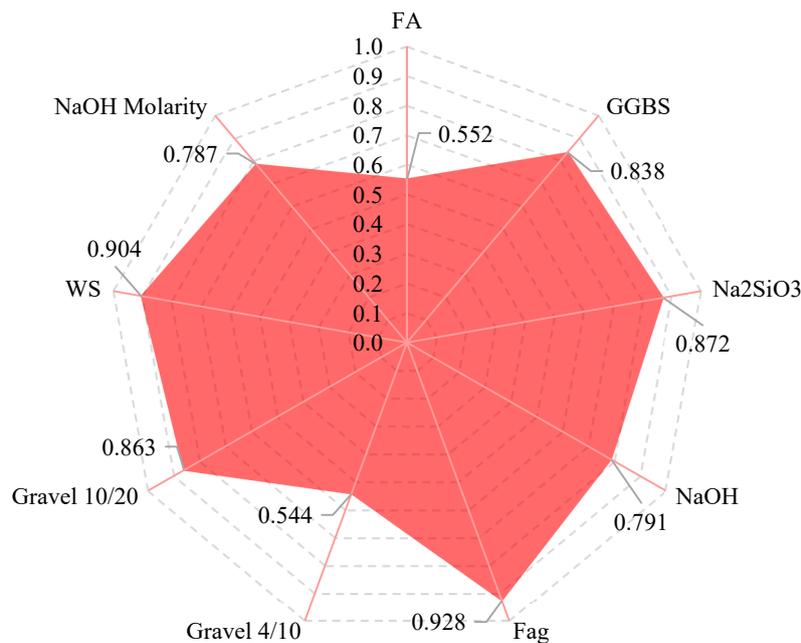
**Figure 16.** Testing results of predicted CSGPoC by DT, XGBoost, and RF models.

In the last step of modeling processes, a sensitivity analysis on the effective parameters is performed. Sensitivity analysis techniques, such as the cosine amplitude method (CAM), evaluate the impact of input parameters or assumptions on the output of a model or system. This method involves systematically varying individual input parameters while keeping other factors constant and measuring the resulting changes in the model's output. By applying this method, researchers can quantify the model's sensitivity to specific input variations and identify the parameters that had the most critical impacts on the model's behaviors. Through this analysis, researchers can identify critical parameters that contribute the most to output variability, allowing for the prioritization of resources and efforts toward addressing and optimizing these influential factors. Also, this method evaluates how sensitive the model is to small or large fluctuations in input parameters [113], as follows:

$$S_{ij} = \frac{\sum_{k=1}^m x_{ik} \cdot x_{jk}}{\sqrt{\left(\sum_{k=1}^m x_{ik}^2\right) \cdot \left(\sum_{k=1}^m x_{jk}^2\right)}} \quad (18)$$

where  $x_{ik}$  and  $x_{jk}$  represent the inputs and output variables, and  $m$  stands the number of data points.

In accordance with the devised methodology of CAM, emphasis was placed on assessing the sensitivity of output variables to input variables. As illustrated in Figure 17, the influence of input parameters (X) on objective functions (outputs) was investigated. A higher value of  $r_{ij}$ , closer to 1, signifies a more pronounced impact of the input parameters on the objectives (outputs). The outcomes depicted in Figure 17 reveal that a majority of the input parameters exhibit significant effects on CSGPoC. Specifically, the parameters Fag and WS demonstrated the most substantial impacts on CSGPoC, with strengths of 0.928 and 0.904, respectively. In the second rank, Na<sub>2</sub>SiO<sub>3</sub>, gravel 10/20, and GGBS parameters exhibited comparable strengths of 0.872, 0.863, and 0.838, indicating approximately similar levels of influence. Additionally, the parameters NaOH and NaOH molarity exerted notable effects on CSGPoC, with strengths of 0.791 and 0.787, respectively. Also, the impact of FA on CSGPoC was moderate, as evidenced by a strength value of 0.552.



**Figure 17.** The strength of the impact of effective parameters on CSGPoC.

## 6. Limitations and future works

Despite the promising results obtained from the XGBoost model in predicting CSGPoC, it is essential to acknowledge certain limitations within this study. First, the generalization of the developed models may be constrained by the specific composition and characteristics of the dataset used for training and testing. The reliance on a singular dataset, consisting of 259 CSGPoC samples, may not fully encapsulate the diverse range of conditions and materials encountered in real-world scenarios.

Additionally, the study focuses on a specific set of input parameters, and the exclusion of other potentially influential factors could limit the model's applicability to a broader spectrum of geo-polymer concrete formulations. Furthermore, the current research primarily addresses the prediction aspect, and considering various external factors, the practical implementation of the proposed models in an industrial setting remains a subject for future exploration.

To advance the field of geo-polymer concrete compressive strength estimation, future research endeavors should aim to address the identified limitations and explore new avenues. First, expanding the dataset to include a more extensive variety of geo-polymer concrete formulations and considering additional influential factors could enhance the robustness and generalization of the developed models. The incorporation of real-world complexities, such as environmental conditions and variations in raw materials, would contribute to the models' reliability in practical applications. Moreover, a comparative analysis with other advanced machine learning algorithms and the integration of hybrid models could provide further insights into optimizing the accuracy and efficiency of CSGPoC prediction. The scalability and adaptability of the models for different scales of construction projects and manufacturing processes should also be investigated. Hence, validation through large-scale field trials would validate the models' effectiveness and facilitate their seamless integration into the decision-making processes of the green concretes industry.

## 7. Conclusions

The main purpose of the current research is to establish a robust predictive system for predicting CSGPoC. The creation of environmentally friendly building supplies may be helped along by the increased use of geo-polymer concrete in the construction sector, which is also helping to popularize this material. This research has a beneficial effect on advancing the use of geopolymer concrete by boosting its use. An efficient super learner technique for predicting the CSGPoC was proposed with the use of the XGBoost and RF models, which allowed for the development of a high-performance model. A database comprised of 259 CSGPoC data points was gathered from literature for developing DT, RF, and XGBoost models as well as accurately predicting CSGPoC. For developing models, nine effective parameters, including FA, GGBS, Na<sub>2</sub>SiO<sub>3</sub>, NaOH, FAg, Gravel 4/10, Gravel 10/20, WS, and NaOH molarity, were considered. The obtained results clarified that the highest R<sup>2</sup> was determined by the XGBoost model as 0.9819 and 0.9857 for the training and testing parts, respectively. Hence, the XGBoost outperformed DT with R<sup>2</sup> of 0.8859 and 0.8969 and RF with R<sup>2</sup> of 0.9492 and 0.9424 for training and testing phases, respectively. It can be concluded that the XGBoost super learner model is significantly more efficient in establishing estimation models of CSGPoC than the DT and RF models.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

This research was supported by the Guangdong provincial science and technology plan project (Grant No. 2021B1111610002), Natural Science Foundation of Hunan (Grant No. 2023JJ50418) and Hunan Provincial transportation technology project (Grant No. 202109). The writers are grateful for this support.

## Conflict of interest

The authors declare no conflict of interest.

## References

1. S. H. Chu, H. Ye, L. Huang, L. G. Li, Carbon fiber reinforced geopolymer (FRG) mix design based on liquid film thickness, *Constr. Build. Mater.*, **269** (2021), 121278. <https://doi.org/10.1016/j.conbuildmat.2020.121278>
2. J. Huang, M. Zhou, J. Zhang, J. Ren, N. I. Vatin, M. M. S. Sabri, The use of GA and PSO in evaluating the shear strength of steel fiber reinforced concrete beams, *KSCE J. Civ. Eng.*, **26** (2022), 3918–3931. <https://doi.org/10.1007/s12205-022-0961-0>
3. K. V. Teja, P. P. Sai, T. Meena, Investigation on the behaviour of ternary blended concrete with scba and sf, *IOP Conf. Ser.: Mater. Sci. Eng.*, **263** (2017), 032012. <https://doi.org/10.1088/1757-899X/263/3/032012>
4. Z. Bayasi, J. Zhou, Properties of silica fume concrete and mortar, *Int. Concr. Abstr. Portal*, **90** (1993), 349–356. <https://doi.org/10.14359/3892>
5. X. Li, D. Qin, Y. Hu, W. Ahmad, A. Ahmad, F. Aslam, et al., A systematic review of waste materials in cement-based composites for construction applications, *J. Build. Eng.*, **45** (2022), 103447. <https://doi.org/10.1016/j.jobe.2021.103447>
6. J. Huang, M. Zhou, J. Zhang, J. Ren, N. I. Vatin, M. M. S. Sabri, Development of a new stacking model to evaluate the strength parameters of concrete samples in laboratory, *Iran. J. Sci. Technol. Trans. Civ. Eng.*, **46** (2022), 4355–4370. <https://doi.org/10.1007/s40996-022-00912-y>
7. A. Cleetus, R. Shibu, P. M. Sreehari, V. K. Paul, B. Jacob, Analysis and study of the effect of GGBFS on concrete structures, *Int. Res. J. Eng. Technol.*, **5** (2018), 3033–3037.
8. Y. Qiu, J. Zhou, Short-term rockburst damage assessment in burst-prone mines: an explainable XGBOOST hybrid model with SCSO algorithm, *Rock Mech. Rock Eng.*, **56** (2023), 8745–8770. <https://doi.org/10.1007/s00603-023-03522-w>
9. V. K. Nagaraj, D. L. V. Babu, Assessing the performance of molarity and alkaline activator ratio on engineering properties of self-compacting alkaline activated concrete at ambient temperature, *J. Build. Eng.*, **20** (2018), 137–155. <https://doi.org/10.1016/j.jobe.2018.07.005>
10. W. P. Zakka, N. H. A. S. Lim, M. C. Khun, A scientometric review of geopolymer concrete, *J. Cleaner Prod.*, **280** (2021), 124353. <https://doi.org/10.1016/j.jclepro.2020.124353>
11. P. Zhang, K. Wang, Q. Li, J. Wang, Y. Ling, Fabrication and engineering properties of concretes based on geopolymers/alkali-activated binders—a review, *J. Cleaner Prod.*, **258** (2020), 120896. <https://doi.org/10.1016/j.jclepro.2020.120896>
12. A. Ahmad, W. Ahmad, F. Aslam, P. Joyklad, Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques, *Case Stud. Constr. Mater.*, **16** (2022), e00840. <https://doi.org/10.1016/j.cscm.2021.e00840>
13. Y. Qiu, J. Zhou, Short-term rockburst prediction in underground project: insights from an explainable and interpretable ensemble learning model, *Acta Geotech.*, **18** (2023), 6655–6685. <https://doi.org/10.1007/s11440-023-01988-0>

14. F. Farooq, X. Jin, M. F. Javed, A. Akbar, M. I. Shah, F. Aslam, et al., Geopolymer concrete as sustainable material: a state of the art review, *Constr. Build. Mater.*, **306** (2021), 124762. <https://doi.org/10.1016/j.conbuildmat.2021.124762>
15. A. Hosan, S. Haque, F. Shaikh, Compressive behaviour of sodium and potassium activators synthesized fly ash geopolymer at elevated temperatures: a comparative study, *J. Build. Eng.*, **8** (2016), 123–130. <https://doi.org/10.1016/j.jobe.2016.10.005>
16. H. L. Muttashar, M. A. M. Ariffin, M. N. Hussein, M. W. Hussin, S. B. Ishaq, Self-compacting geopolymer concrete with spend garnet as sand replacement, *J. Build. Eng.*, **15** (2018), 85–94. <https://doi.org/10.1016/j.jobe.2017.10.007>
17. K. Z. Farhan, M. A. M. Johari, R. Demirboğa, Assessment of important parameters involved in the synthesis of geopolymer composites: a review, *Constr. Build. Mater.*, **264** (2020), 120276. <https://doi.org/10.1016/j.conbuildmat.2020.120276>
18. C. Herath, C. Gunasekara, D. W. Law, S. Setunge, Long term mechanical performance of nano-engineered high volume fly ash concrete, *J. Build. Eng.*, **43** (2021), 103168. <https://doi.org/10.1016/j.jobe.2021.103168>
19. W. Ahmad, A. Ahmad, K. A. Ostrowski, F. Aslam, P. Joyklad, P. Zajdel, Application of advanced machine learning approaches to predict the compressive strength of concrete containing supplementary cementitious materials, *Materials*, **14** (2021), 5762. <https://doi.org/10.3390/ma14195762>
20. H. K. Tchakouté, C. H. Rüscher, S. Kong, N. Ranjbar, Synthesis of sodium waterglass from white rice husk ash as an activator to produce metakaolin-based geopolymer cements, *J. Build. Eng.*, **6** (2016), 252–261. <https://doi.org/10.1016/j.jobe.2016.04.007>
21. A. Ahmad, W. Ahmad, K. Chaiyasarn, K. A. Ostrowski, F. Aslam, P. Zajdel, et al., Prediction of geopolymer concrete compressive strength using novel machine learning algorithms, *Polymers*, **13** (2021), 3389. <https://doi.org/10.3390/polym13193389>
22. B. B. Jindal, Investigations on the properties of geopolymer mortar and concrete with mineral admixtures: a review, *Constr. Build. Mater.*, **227** (2019), 116644. <https://doi.org/10.1016/j.conbuildmat.2019.08.025>
23. M. S. Reddy, P. Dinakar, B. H. Rao, Mix design development of fly ash and ground granulated blast furnace slag based geopolymer concrete, *J. Build. Eng.*, **20** (2018), 712–722. <https://doi.org/10.1016/j.jobe.2018.09.010>
24. J. Zhou, X. Shen, Y. Qiu, X. Shi, K. Du, Microseismic location in hardrock metal mines by machine learning models based on hyperparameter optimization using bayesian optimizer, *Rock Mech. Rock Eng.*, **56** (2023), 8771–8788. <https://doi.org/10.1007/s00603-023-03483-0>
25. C. L. Wong, K. H. Mo, U. J. Alengaram, S. P. Yap, Mechanical strength and permeation properties of high calcium fly ash-based geopolymer containing recycled brick powder, *J. Build. Eng.*, **32** (2020), 101655. <https://doi.org/10.1016/j.jobe.2020.101655>
26. S. K. John, Y. Nadir, K. Girija, Effect of source materials, additives on the mechanical properties and durability of fly ash and fly ash-slag geopolymer mortar: a review, *Constr. Build. Mater.*, **280** (2021), 122443. <https://doi.org/10.1016/j.conbuildmat.2021.122443>
27. Y. Zou, C. Zheng, A. Alzahrani, W. Ahmad, A. Ahmad, A. Mohamed, et al., Evaluation of artificial intelligence methods to estimate the compressive strength of geopolymers, *Gels*, **8** (2022), 271. <https://doi.org/10.3390/gels8050271>

28. J. Huang, M. Zhou, H. Yuan, M. M. S. Sabri, X. Li, Prediction of the compressive strength for cement-based materials with metakaolin based on the hybrid machine learning method, *Materials*, **15** (2022), 3500. <https://doi.org/10.3390/ma15103500>
29. B. Indraratna, D. J. Armaghani, A. G. Correia, H. Hunt, T. Ngo, Prediction of resilient modulus of ballast under cyclic loading using machine learning techniques, *Transp. Geotech.*, **38** (2023), 100895. <https://doi.org/10.1016/j.trgeo.2022.100895>
30. S. Medawela, D. J. Armaghani, B. Indraratna, R. K. Rowe, N. Thamwattana, Development of an advanced machine learning model to predict the pH of groundwater in permeable reactive barriers (PRBs) located in acidic terrain, *Comput. Geotech.*, **161** (2023), 105557. <https://doi.org/10.1016/j.compgeo.2023.105557>
31. D. J. Armaghani, P. G. Asteris, A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength, *Neural Comput. Appl.*, **33** (2021), 4501–4532. <https://doi.org/10.1007/s00521-020-05244-4>
32. A. D. Skentou, A. Bardhan, A. Mamou, M. E. Lemonis, G. Kumar, P. Samui, et al., Closed-form equation for estimating unconfined compressive strength of granite from three non-destructive tests using soft computing models, *Rock Mech. Rock Eng.*, **56** (2023), 487–514. <https://doi.org/10.1007/s00603-022-03046-9>
33. W. Mahmood, A. S. Mohammed, P. G. Asteris, R. Kurda, D. J. Armaghani, Modeling flexural and compressive strengths behaviour of cement-grouted sands modified with water reducer polymer, *Appl. Sci.*, **12** (2022), 1016. <https://doi.org/10.3390/app12031016>
34. P. G. Asteris, P. B. Lourenço, P. C. Roussis, C. E. Adami, D. J. Armaghani, L. Cavaleri, et al., Revealing the nature of metakaolin-based concrete materials using artificial intelligence techniques, *Constr. Build. Mater.*, **322** (2022), 126500. <https://doi.org/10.1016/j.conbuildmat.2022.126500>
35. M. S. Barkhordari, D. J. Armaghani, P. G. Asteris, Structural damage identification using ensemble deep convolutional neural network models, *CMES-Comp. Model. Eng. Sci.*, **134** (2023), 835–855. <https://doi.org/10.32604/cmesc.2022.020840>
36. L. Cavaleri, M. S. Barkhordari, C. C. Repapis, D. J. Armaghani, D. V. Ulrikh, P. G. Asteris, Convolution-based ensemble learning algorithms to estimate the bond strength of the corroded reinforced concrete, *Constr. Build. Mater.*, **359** (2022), 129504. <https://doi.org/10.1016/j.conbuildmat.2022.129504>
37. M. Koopialipoor, P. G. Asteris, A. S. Mohammed, D. E. Alexakis, A. Mamou, D. J. Armaghani, Introducing stacking machine learning approaches for the prediction of rock deformation, *Transp. Geotech.*, **34** (2022), 100756. <https://doi.org/10.1016/j.trgeo.2022.100756>
38. M. Parsajoo, D. J. Armaghani, A. S. Mohammed, M. Khari, S. Jahandari, Tensile strength prediction of rock material using non-destructive tests: a comparative intelligent study, *Transp. Geotech.*, **31** (2021), 100652. <https://doi.org/10.1016/j.trgeo.2021.100652>
39. D. J. Armaghani, Y. Y. Ming, A. S. Mohammed, E. Momeni, H. Maizir, Effect of SVM kernel functions on bearing capacity assessment of deep foundations, *J. Soft Comput. Civ. Eng.*, **7** (2023), 111–128. Available from: [https://www.jssoftcivil.com/article\\_169219.html](https://www.jssoftcivil.com/article_169219.html).
40. E. Momeni, B. He, Y. Abdi, D. J. Armaghani, Novel hybrid XGBoost model to forecast soil shear strength based on some soil index tests, *CMES-Comp. Model. Eng. Sci.*, **136** (2023), 2527–2550. <https://doi.org/10.32604/cmesc.2023.026531>

41. J. Zhou, Z. Wang, C. Li, W. Wei, S. Wang, D. J. Armaghani, et al., Hybridized random forest with population-based optimization for predicting shear properties of rock fractures, *J. Comput. Sci.*, **72** (2023), 102097. <https://doi.org/10.1016/j.jocs.2023.102097>
42. J. Huang, M. Zhou, H. Yuan, M. M. Sabri, X. Li, Towards sustainable construction materials: a comparative study of prediction models for green concrete with metakaolin, *Buildings*, **12** (2022), 772. <https://doi.org/10.3390/buildings12060772>
43. A. Ahmad, K. Chaiyasarn, F. Farooq, W. Ahmad, S. Suparp, F. Aslam, Compressive strength prediction via gene expression programming (Gep) and artificial neural network (ann) for concrete containing rca, *Buildings*, **11** (2021), 324. <https://doi.org/10.3390/buildings11080324>
44. H. Song, A. Ahmad, K. A. Ostrowski, M. Dudek, Analyzing the compressive strength of ceramic waste-based concrete using experiment and artificial neural network (Ann) approach, *Materials*, **14** (2021), 4518. <https://doi.org/10.3390/ma14164518>
45. H. Nguyen, T. Vu, T. P. Vo, H. T. Thai, Efficient machine learning models for prediction of concrete strengths, *Constr. Build. Mater.*, **266** (2021), 120950. <https://doi.org/10.1016/j.conbuildmat.2020.120950>
46. J. Huang, Y. Sun, J. Zhang, Reduction of computational error by optimizing SVR kernel coefficients to simulate concrete compressive strength through the use of a human learning optimization algorithm, *Eng. Comput.*, **38** (2022), 3151–3168. <https://doi.org/10.1007/s00366-021-01305-x>
47. P. Sarir, J. Chen, P. G. Asteris, D. J. Armaghani, M. M. Tahir, Developing GEP tree-based, neuro-swarm, and whale optimization models for evaluation of bearing capacity of concrete-filled steel tube columns, *Eng. Comput.*, **37** (2021), 1–19. <https://doi.org/10.1007/s00366-019-00808-y>
48. F. R. Balf, H. M. Kordkheili, A. M. Kordkheili, A new method for predicting the ingredients of self-compacting concrete (SCC) including fly ash (FA) using data envelopment analysis (DEA), *Arabian J. Sci. Eng.*, **46** (2021), 4439–4460. <https://doi.org/10.1007/s13369-020-04927-3>
49. A. Ahmad, F. Farooq, K. A. Ostrowski, K. Śliwa-Wieczorek, S. Czarnecki, Application of novel machine learning techniques for predicting the surface chloride concentration in concrete containing waste material, *Materials*, **14** (2021), 2297. <https://doi.org/10.3390/ma14092297>
50. M. Azimi-Pour, H. Eskandari-Naddaf, A. Pakzad, Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete, *Constr. Build. Mater.*, **230** (2020), 117021. <https://doi.org/10.1016/j.conbuildmat.2019.117021>
51. P. Saha, P. Debnath, P. Thomas, Prediction of fresh and hardened properties of self-compacting concrete using support vector regression approach, *Neural Comput. Appl.*, **32** (2020), 7995–8010. <https://doi.org/10.1007/s00521-019-04267-w>
52. A. A. Shahmansouri, H. A. Bengar, E. Jahani, Predicting compressive strength and electrical resistivity of eco-friendly concrete containing natural zeolite via GEP algorithm, *Constr. Build. Mater.*, **229** (2019), 116883. <https://doi.org/10.1016/j.conbuildmat.2019.116883>
53. J. Huang, M. M. Sabri, D. V. Ulrikh, M. Ahmad, K. A. Alsaffar, Predicting the compressive strength of the cement-fly ash–slag ternary concrete using the firefly algorithm (FA) and random forest (RF) hybrid machine-learning method, *Materials*, **15** (2022), 4193. <https://doi.org/10.3390/ma15124193>
54. F. Aslam, F. Farooq, M. N. Amin, K. Khan, A. Waheed, A. Akbar, et al., Applications of gene expression programming for estimating compressive strength of high-strength concrete, *Adv. Civ. Eng.*, **2020** (2020), 8850535. <https://doi.org/10.1155/2020/8850535>

55. F. Farooq, M. N. Amin, K. Khan, M. R. Sadiq, M. F. Javed, F. Aslam, et al., A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC), *Appl. Sci.*, **10** (2020), 7330. <https://doi.org/10.3390/app10207330>
56. P. G. Asteris, K. G. Kolovos, Self-compacting concrete strength prediction using surrogate models, *Neural Comput. Appl.*, **31** (2019), 409–424. <https://doi.org/10.1007/s00521-017-3007-7>
57. S. Selvaraj, S. Sivaraman, Prediction model for optimized self-compacting concrete with fly ash using response surface method based on fuzzy classification, *Neural Comput. Appl.*, **31** (2019), 1365–1373. <https://doi.org/10.1007/s00521-018-3575-1>
58. J. Zhang, G. Ma, Y. Huang, J. Sun, F. Aslani, B. Nener, Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression, *Constr. Build. Mater.*, **210** (2019), 713–719. <https://doi.org/10.1016/j.conbuildmat.2019.03.189>
59. A. Kaveh, T. Bakhshpoori, S. M. Hamze-Ziabari, M5' and mars based prediction models for properties of selfcompacting concrete containing fly ash, *Period. Polytech., Civ. Eng.*, **62** (2018), 281–294. <https://doi.org/10.3311/PPci.10799>
60. D. Sathyan, K. B. Anand, A. J. Prakash, B. Premjith, Modeling the fresh and hardened stage properties of self-compacting concrete using random kitchen sink algorithm, *Int. J. Concr. Struct. Mater.*, **12** (2018), 24. <https://doi.org/10.1186/s40069-018-0246-7>
61. B. Vakhshouri, S. Nejadi, Prediction of compressive strength of self-compacting concrete by ANFIS models, *Neurocomputing*, **280** (2018), 13–22. <https://doi.org/10.1016/j.neucom.2017.09.099>
62. O. B. Douma, B. Boukhatem, M. Ghrici, A. Tagnit-Hamou, Prediction of properties of self-compacting concrete containing fly ash using artificial neural network, *Neural Comput. Appl.*, **28** (2017), 707–718. <https://doi.org/10.1007/s00521-016-2368-7>
63. M. A. Yaman, M. A. Elaty, M. Taman, Predicting the ingredients of self compacting concrete using artificial neural network, *Alexandria Eng. J.*, **56** (2017), 523–532. <https://doi.org/10.1016/j.aej.2017.04.007>
64. A. Ahmad, F. Farooq, P. Niewiadomski, K. Ostrowski, A. Akbar, F. Aslam, et al., Prediction of compressive strength of fly ash based concrete using individual and ensemble algorithm, *Materials*, **14** (2021), 794. <https://doi.org/10.3390/ma14040794>
65. F. Farooq, W. Ahmed, A. Akbar, F. Aslam, R. Alyousef, Predictive modeling for sustainable high-performance concrete from industrial wastes: a comparison and optimization of models using ensemble learners, *J. Cleaner Prod.*, **292** (2021), 126032. <https://doi.org/10.1016/j.jclepro.2021.126032>
66. R. Buši, M. Benšić, I. Miličević, K. Strukar, Prediction models for the mechanical properties of self-compacting concrete with recycled rubber and silica fume, **13** (2020), 1821. <https://doi.org/10.3390/ma13081821>
67. M. F. Javed, F. Farooq, S. A. Memon, A. Akbar, M. A. Khan, F. Aslam, et al., New prediction model for the ultimate axial capacity of concrete-filled steel tubes: an evolutionary approach, *Crystals*, **10** (2020), 741. <https://doi.org/10.3390/cryst10090741>
68. M. Nematzadeh, A. A. Shahmansouri, M. Fakoor, Post-fire compressive strength of recycled PET aggregate concrete reinforced with steel fibers : optimization and prediction via RSM and GEP, *Constr. Build. Mater.*, **252** (2020), 119057. <https://doi.org/10.1016/j.conbuildmat.2020.119057>

69. K. Güçlüer, A. Özbeyaz, S. Göymen, O. Günaydın, A comparative investigation using machine learning methods for concrete compressive strength estimation, *Mater. Today Commun.*, **27** (2021), 102278. <https://doi.org/10.1016/j.mtcomm.2021.102278>
70. A. Ahmad, K. A. Ostrowski, M. Maślak, F. Farooq, I. Mehmood, A. Nafees, Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature, *Materials*, **14** (2021), 4222. <https://doi.org/10.3390/ma14154222>
71. P. G. Asteris, A. D. Skentou, A. Bardhan, P. Samui, K. Pilakoutas, Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models, *Cem. Concr. Res.*, **145** (2021), 106449. <https://doi.org/10.1016/j.cemconres.2021.106449>
72. W. Emad, A. S. Mohammed, R. Kurda, K. Ghafor, L. Cavaleri, S. M. A. Qaidi, et al., Prediction of concrete materials compressive strength using surrogate models, *Structures*, **46** (2022), 1243–1267. <https://doi.org/10.1016/j.istruc.2022.11.002>
73. Z. Shen, A. F. Deifalla, P. Kamiński, A. Dyczko, Compressive strength evaluation of ultra-high-strength concrete by machine learning, *Materials*, **15** (2022), 3523. <https://doi.org/10.3390/ma15103523>
74. A. Kumar, H. C. Arora, N. R. Kapoor, M. A. Mohammed, K. Kumar, A. Majumdar, et al., Compressive strength prediction of lightweight concrete: machine learning models, *Sustainability*, **14** (2022), 2404. <https://doi.org/10.3390/su14042404>
75. D. K. I. Jaf, P. I. Abdulrahman, A. S. Mohammed, R. Kurda, S. M. A. Qaidi, P. G. Asteris, Machine learning techniques and multi-scale models to evaluate the impact of silicon dioxide (SiO<sub>2</sub>) and calcium oxide (CaO) in fly ash on the compressive strength of green concrete, *Constr. Build. Mater.*, **400** (2023), 132604. <https://doi.org/10.1016/j.conbuildmat.2023.132604>
76. W. Mahmood, A. S. Mohammed, P. G. Asteris, H. Ahmed, Soft computing technics to predict the early-age compressive strength of flowable ordinary Portland cement, *Soft Comput.*, **27** (2023), 3133–3150. <https://doi.org/10.1007/s00500-022-07505-x>
77. R. Ali, M. Muayad, A. S. Mohammed, P. G. Asteris, Analysis and prediction of the effect of Nanosilica on the compressive strength of concrete with different mix proportions and specimen sizes using various numerical approaches, *Struct. Concr.*, **24** (2023), 4161–4184. <https://doi.org/10.1002/suco.202200718>
78. J. Huang, J. Zhang, X. Li, Y. Qiao, R. Zhang, G. S. Kumar, Investigating the effects of ensemble and weight optimization approaches on neural networks' performance to estimate the dynamic modulus of asphalt concrete, *Road Mater. Pavement Des.*, **24** (2023), 1939–1959. <https://doi.org/10.1080/14680629.2022.2112061>
79. J. Zhou, S. Huang, Y. Qiu, Optimization of random forest through the use of MVO, GWO and MFO in evaluating the stability of underground entry-type excavations, *Tunnelling Underground Space Technol.*, **124** (2022), 104494. <https://doi.org/10.1016/j.tust.2022.104494>
80. D. V. Dao, S. H. Trinh, H. B. Ly, B. T. Pham, Prediction of compressive strength of geopolymers concrete using entirely steel slag aggregates: novel hybrid artificial intelligence approaches, *Appl. Sci.*, **9** (2019), 1113. <https://doi.org/10.3390/app9061113>
81. Q. A. Wang, J. Zhang, J. Huang, Simulation of the compressive strength of cemented tailing backfill through the use of firefly algorithm and random forest model, *Shock Vib.*, **2021** (2021), 5536998. <https://doi.org/10.1155/2021/5536998>

82. X. Liang, X. Yu, G. Ding, Y. Jing, J. Huang, Environmental and mechanical effects of rubberised open-graded asphalt mixtures incorporating with titanium dioxide: a laboratory investigation, *Int. J. Pavement Eng.*, **24** (2023), 2241604. <https://doi.org/10.1080/10298436.2023.2241604>
83. X. Liang, X. Yu, B. Xu, C. Chen, G. Ding, Y. Jin, et al., Storage stability and compatibility in foamed warm-mix asphalt (FWMA) containing recycled asphalt pavement (RAP) binder, *J. Mater. Civ. Eng.*, 2024. <https://doi.org/10.1061/JMCEE7/MTENG-16468>
84. G. Gu, F. Xu, X. Huang, S. Ruan, C. Peng, J. Lin, Foamed geopolymer: the relationship between rheological properties of geopolymer paste and pore-formation mechanism, *J. Cleaner Prod.*, **277** (2020), 123238. <https://doi.org/10.1016/j.jclepro.2020.123238>
85. J. Huang, J. Xue, Optimization of SVR functions for flyrock evaluation in mine blasting operations, *Environ. Earth Sci.*, **81** (2022), 434. <https://doi.org/10.1007/s12665-022-10523-5>
86. J. Huang, J. Zhang, Y. Gao, H. Liu, Intelligently predict the rock joint shear strength using the support vector regression and firefly algorithm, *Lithosphere*, **2021** (2021), 2467126. <https://doi.org/10.2113/2021/2467126>
87. J. Zhou, X. Shen, Y. Qiu, X. Shi, M. Khandelwal, Cross-correlation stacking-based microseismic source location using three metaheuristic optimization algorithms, *Tunnelling Underground Space Technol.*, **126** (2022), 104570. <https://doi.org/10.1016/j.tust.2022.104570>
88. X. Wang, S. Hosseini, D. J. Armaghani, E. T. Mohamad, Data-driven optimized artificial neural network technique for prediction of flyrock induced by boulder blasting, *Mathematics*, **11** (2023), 2358. <https://doi.org/10.3390/math11102358>
89. S. Hosseini, R. Poormirzaee, M. Hajihassani, Application of reliability-based back-propagation causality-weighted neural networks to estimate air-overpressure due to mine blasting, *Eng. Appl. Artif. Intell.*, **115** (2022), 105281. <https://doi.org/10.1016/j.engappai.2022.105281>
90. S. Hosseini, R. Pourmirzaee, D. J. Armaghani, M. M. S. Sabri, Prediction of ground vibration due to mine blasting in a surface lead–zinc mine using machine learning ensemble techniques, *Sci. Rep.*, **13** (2023), 6591. <https://doi.org/10.1038/s41598-023-33796-7>
91. S. Hosseini, M. Monjezi, E. Bakhtavar, A. Mousavi, Prediction of dust emission due to open pit mine blasting using a hybrid artificial neural network, *Nat. Resour. Res.*, **30** (2021), 4773–4788. <https://doi.org/10.1007/s11053-021-09930-5>
92. S. Hosseini, A. Mousavi, M. Monjezi, Prediction of blast-induced dust emissions in surface mines using integration of dimensional analysis and multivariate regression analysis, *Arabian J. Geosci.*, **15** (2022), 163. <https://doi.org/10.1007/s12517-021-09376-2>
93. H. I. Erdal, Two-level and hybrid ensembles of decision trees for high performance concrete compressive strength prediction, *Eng. Appl. Artif. Intell.*, **26** (2013), 1689–1697. <https://doi.org/10.1016/j.engappai.2013.03.014>
94. W. B. Chaabene, M. Flah, M. L. Nehdi, Machine learning prediction of mechanical properties of concrete: critical review, *Constr. Build. Mater.*, **260** (2020), 119889. <https://doi.org/10.1016/j.conbuildmat.2020.119889>
95. L. Breiman, Random forests, *Mach. Learn.*, **45** (2001), 5–32. <https://doi.org/10.1023/A:1010933404324>
96. A. Liaw, M. Wiener, Classification and regression by randomForest, *R News*, **2** (2002), 18–22.
97. T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016), 785–794. <https://doi.org/10.1145/2939672.2939785>

98. J. Zhao, S. Hosseini, Q. Chen, D. J. Armaghani, Super learner ensemble model: a novel approach for predicting monthly copper price in future, *Resour. Policy*, **85** (2023), 103903. <https://doi.org/10.1016/j.resourpol.2023.103903>
99. Q. Wang, J. Qi, S. Hosseini, H. Rasekh, J. Huang, ICA-LightGBM algorithm for predicting compressive strength of geo-polymer concrete, *Buildings*, **13** (2023), 2278. <https://doi.org/10.3390/buildings13092278>
100. E. Bakhtavar, S. Hosseini, K. Hewage, R. Sadiq, Green blasting policy: simultaneous forecast of vertical and horizontal distribution of dust emissions using artificial causality-weighted neural network, *J. Cleaner Prod.*, **283** (2021), 124562. <https://doi.org/10.1016/j.jclepro.2020.124562>
101. E. Bakhtavar, S. Hosseini, K. Hewage, R. Sadiq, Air pollution risk assessment using a hybrid fuzzy intelligent probability-based approach: mine blasting dust impacts, *Nat. Resour. Res.*, **30** (2021), 2607–2627. <https://doi.org/10.1007/s11053-020-09810-4>
102. S. Hosseini, R. Poormirzaee, M. Hajihassani, R. Kalatehjari, An ANN-Fuzzy cognitive map-based Z-number theory to predict flyrock induced by blasting in open-pit mines, *Rock Mech. Rock Eng.*, **55** (2022), 4373–4390. <https://doi.org/10.1007/s00603-022-02866-z>
103. R. Ali, M. Muayad, A. S. Mohammed, P. G. Asteris, Analysis and prediction of the effect of Nanosilica on the compressive strength of concrete with different mix proportions and specimen sizes using various numerical approaches, *Struct. Concr.*, **24** (2022), 4161–4184. <https://doi.org/10.1002/suco.202200718>
104. S. Hosseini, A. Mousavi, M. Monjezi, M. Khandelwal, Mine-to-crusher policy: planning of mine blasting patterns for environmentally friendly and optimum fragmentation using Monte Carlo simulation-based multi-objective grey wolf optimization approach, *Resour. Policy*, **79** (2022), 103087. <https://doi.org/10.1016/j.resourpol.2022.103087>
105. S. Hosseini, R. Poormirzaee, M. Hajihassani, An uncertainty hybrid model for risk assessment and prediction of blast-induced rock mass fragmentation, *Int. J. Rock Mech. Min. Sci.*, **160** (2022), 105250. <https://doi.org/10.1016/j.ijrmms.2022.105250>
106. S. Hosseini, J. Khatti, B. O. Taiwo, Y. Fissaha, K. S. Grover, H. Ikeda, et al., Assessment of the ground vibration during blasting in mining projects using different computational approaches, *Sci. Rep.*, **13** (2023), 18582. <https://doi.org/10.1038/s41598-023-46064-5>
107. S. Hosseini, S. Javanshir, H. Sabeti, P. Tahmasebizadeh, Mathematical-based gene expression programming (GEP): a novel model to predict zinc separation from a bench-scale bioleaching process, *J. Sustainable Metall.*, **9** (2023), 1601–1619. <https://doi.org/10.1007/s40831-023-00751-9>
108. S. Hosseini, R. Pourmirzaee, Green policy for managing blasting induced dust dispersion in open-pit mines using probability-based deep learning algorithm, *Expert Syst. Appl.*, **240** (2023), 122469. <https://doi.org/10.1016/j.eswa.2023.122469>
109. A. I. Lawal, S. Hosseini, M. Kim, N. O. Ogunsola, S. Kwon, Prediction of factor of safety of slopes using stochastically modified ANN and classical methods: a rigorous statistical model selection approach, *Nat. Hazard.*, 2023. <https://doi.org/10.1007/s11069-023-06275-5>
110. M. Apostolopoulou, P. G. Asteris, D. J. Armaghani, M. G. Douvika, P. B. Lourenço, L. Cavaleri, et al., Mapping and holistic design of natural hydraulic lime mortars, *Cem. Concr. Res.*, **136** (2020), 106167. <https://doi.org/10.1016/j.cemconres.2020.106167>
111. P. G. Asteris, M. Koopialipour, D. J. Armaghani, E. A. Kotsonis, P. B. Lourenço, Prediction of cement-based mortars compressive strength using machine learning techniques, *Neural Comput. Appl.*, **33** (2021), 13089–13121. <https://doi.org/10.1007/s00521-021-06004-8>

112. S. Hosseini, M. Monjezi, E. Bakhtavar, Minimization of blast-induced dust emission using gene-expression programming and grasshopper optimization algorithm: a smart mining solution based on blasting plan optimization, *Clean Technol. Environ. Policy*, **24** (2022), 2313–2328. <https://doi.org/10.1007/s10098-022-02327-9>
113. Y. H. Jong, C. I. Lee, Influence of geological conditions on the powder factor for tunnel blasting, *Int. J. Rock Mech. Min. Sci.*, **41** (2004), 533–538. <https://doi.org/10.1016/j.ijrmms.2004.03.095>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)