

Modelling the compressive strength of geopolymer recycled aggregate concrete using ensemble machine learning

Emad Golafshani^{a,*}, Nima Khodadadi^b, Tuan Ngo^a, Antonio Nanni^b, Ali Behnood^c

^a Department of Infrastructure Engineering, University of Melbourne, Melbourne, Australia

^b Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA

^c Department of Civil Engineering, The University of Mississippi, Mississippi, USA

ARTICLE INFO

Keywords:

Geopolymer recycled aggregate concrete
Fly ash
Slag cement
Compressive strength
Artificial intelligence
Ensemble models

ABSTRACT

In the quest to reduce the environmental impact of the construction sector, the adoption of sustainable and eco-friendly materials is imperative. Geopolymer recycled aggregate concrete (GRAC) emerges as a promising solution by substituting supplementary cementitious materials, including fly ash and slag cement, for ordinary Portland cement and utilizing recycled aggregates from construction and demolition waste, thus significantly lowering carbon emissions and resource consumption. Despite its potential, the widespread implementation of GRAC has been hindered by the lack of an effective mix design methodology. This study seeks to bridge this gap through a novel machine learning (ML)-based approach to accurately model the compressive strength (CS) of GRAC, a critical parameter for ensuring structural integrity and safety. By compiling a comprehensive database from existing literature and enhancing it with synthetic data generated through a tabular generative adversarial network, this research employs eight ensemble ML techniques, comprising three bagging and five boosting methods, to predict the CS of GRAC with high precision. The boosting models, notably extreme gradient boosting, light gradient boosting, gradient boosting, and categorical gradient boosting regressors, demonstrated superior performance, achieving a mean absolute percentage error of less than 6 %. This precision in prediction underscores the viability of ML in optimizing GRAC formulations for enhanced structural applications. The identification of testing age, natural fine aggregate content, and recycled aggregate ratio as pivotal factors offers valuable insights into the mix design process, facilitating more informed decisions in material selection and proportioning. Moreover, the development of a user-friendly graphical interface for CS prediction exemplifies the practical application of this research, potentially accelerating the adoption of GRAC in mainstream construction practices. By enabling the practical use of GRAC, this research contributes to the global effort to promote sustainable development within the construction industry.

1. Introduction

The construction industry plays a vital role in the global economy and population welfare, providing infrastructure, buildings, and facilities essential for modern society. However, this industry is also known for its significant environmental impact, resource consumption, and waste generation [1]. As environmental concerns become increasingly important worldwide, there is a growing imperative to adopt sustainable and eco-friendly practices in construction. Geopolymer concrete stands as an interesting innovation in the realm of construction materials, driven by a strong commitment to sustainability [2]. The key components of geopolymer concrete are aggregates, an aluminosilicate source

material (e.g., silica fume, fly ash, or slag cement), and an alkaline activator solution (e.g., sodium hydroxide and sodium silicate) [3]. Additionally, activators are chemicals or solutions that initiate and accelerate the geopolymerization process. They are mixed with the binder precursor materials to promote the formation of geopolymer gels.

One of the key advantages of geopolymer concrete is its reduced environmental impact. It requires less energy to produce because it does not rely on the high-temperature production of Portland cement. Additionally, it utilizes industrial by-products like fly ash and slag cement, which can reduce waste and landfill disposal [4]. Geopolymer concrete can have a significantly lower carbon footprint compared to conventional concrete due to the reduced CO₂ emissions associated with

* Corresponding author.

E-mail address: emad.mohammadigolafshani@unimelb.edu.au (E. Golafshani).

<https://doi.org/10.1016/j.advengsoft.2024.103611>

Received 6 January 2024; Received in revised form 17 February 2024; Accepted 20 February 2024

Available online 1 March 2024

0965-9978/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

its production [5]. This makes it an attractive choice for sustainable construction and environmentally conscious projects. Geopolymer concrete often exhibits better resistance to chemical attack, corrosion, and abrasion compared to traditional concrete. It can withstand harsh environments and is particularly suitable for infrastructure in aggressive chemical or marine settings [6]. Moreover, geopolymer concrete has good fire-resistant properties. It can withstand high temperatures without significant loss of structural integrity [7]. This makes it suitable for fire-resistant construction elements like tunnels and fireproofing applications.

The use of recycled aggregates, often sourced from construction and demolition wastes, in this type of concrete, generates geopolymer recycled aggregate concrete (GRAC); an environmentally friendly construction material that not only reduces the burden on overflowing landfills but also optimizes the efficient use of construction resources [8, 9]. However, the use of recycled aggregates in GRAC can decrease its compressive strength [10], fracture energy [10], acid resistance [11], durability performance [12], and fire resistance [13] depending on the proportion of recycled aggregate used. This reduction is mainly due to the increased porosity of the concrete and weaker interfacial transition zone [10]. Eliminating traditional Portland cement and using fly ash and slag cement, obtained from by-products of industrial processes, as binders of GRAC, provides a sustainable and eco-friendly way to significantly diminish the carbon footprint [14]. Furthermore, the incorporation of fly ash, slag cement, and silica fume enhances the

mechanical and durability properties of GRAC. This improvement is due to their interaction with calcium silicate hydrate, calcium aluminate silicate hydrate, and sodium aluminate silicate hydrate gels, which offsets the negative effects associated with using 100 % recycled aggregate [15].

The complexity observed in the behaviour of GRAC arises from multiple sources [16]. One significant factor is the intricate chemical reactions involving aluminosilicate source materials, which interact with activators. The behaviour of GRAC exhibits notable variations based on the specific characteristics and composition of these source materials, as well as the activator formulation employed [16]. Furthermore, the properties of recycled aggregates used in the mix can exert a substantial influence on the concrete's performance [17]. Additionally, the conditions during the curing process, encompassing factors like temperature and duration, play a pivotal role in the geopolymerization reactions, rendering GRAC highly responsive to variations in curing methods [18]. Consequently, the absence of a generalized standardized design method tailored to the intricacies of GRAC becomes apparent. Given these multifaceted challenges and the imperative to optimize GRAC for practical applications, the integration of machine learning (ML) techniques emerges as a valuable approach. ML has the capacity to model the intricate relationships between input parameters and the mechanical properties of concrete, thus facilitating more precise predictions and streamlined mix designs [19].

In the past five years, as summarized in Table 1, several research

Table 1
Studies in the past five years regarding the prediction of CS of geopolymers using ML.

Used SCMs	Input features	Input number	Used ML methods	Database size	Refs
FA, Sl	FA, Sl, NCA, NFA, W, SP, SH, SS	8	GPR, ANN, RFR, GBR	177	[20]
Sl, SF, Zeolite	Sl, SF, Zeolite, SH, Age	5	DT, BR, LSBR	351	[21]
FA	FA, AA/FA, NCA, NFA, SS/SH, SHC, CT, Age	8	LR, ANN, ABR	154	[22]
FA, Sl, MK	FA, Sl, MK, NCA, NFA, W, SP, SH, SHC, CT, CD	11	ANN, RFR, MARS, HEML	1123	[23]
FA	FA, NCA, NFA, W, (SiO ₂ , Na ₂ O, CaO, Al ₂ O ₃ , Fe ₂ O ₃ , LOI, SSA) in FA, AA/FA, SH, SHC, SS, CT, HCD	17	GEP, ANFIS, ANN	245	[24]
Sl, Volcanic ash	Volcanic ash/B, Sl/B, Na/Al, Si/Al, two features related to room curing duration, two features related to oven drying duration	8	LR, Poly2, Elastic Net, Bayesian Ridge, SVR, RFR, BR, GBR, XGBR , Stacking	80	[25]
FA	(SiO ₂ , Al ₂ O ₃ , Fe ₂ O ₃ , CaO, K ₂ O, SO ₃ , Na ₂ O, P ₂ O ₅ , MgO, TiO ₂ , LOI) in FA, NCA/FA, NFA/FA, SP/FA, W/FA, SH/FA, SHC, AA/FA, CT, HCD, CD	21	ANN	896	[26]
FA, Sl	FA, Sl, NCA, NFA, W, Reactivity modulus, Hydraulic modulus, SiO ₂ modulus, Al ₂ O ₃ modulus, Lime modulus, SH, SS, SS- Na ₂ O, SS- SiO ₂ , SS-W, CT, RH, Age	18	RFR, ERT, GBR , XGBR	676	[27]
Sl, Corncob ash	Sl, Corncob ash, NCA, NCA, W, SH pellets, SHC, SS, CD, Concrete grade	10	DNN	288	[28]
FA	SiO ₂ modulus, Na ₂ O/AA, SiO ₂ /Al ₂ O ₃ , Na ₂ O/SiO ₂ , L/S, Pretreatment CT, HCD, Age	8	RFR, GBR, XGBR	492	[29]
FA, Sl, MK	FA/B, Sl/B, MK/B, W/B, L/S, SH, SS, SS/AA, SH/AA, CT	10	ANN , RFR, KNN	191	[30]
Sl, Perlite	Perlite, SHC, CT, HCD	4	LR, ANN , MARS, GBR	180	[31]
FA, Sl	FA, Sl, NCA, NFA, W, SP, SH, SS, CT, CD, RH, Age	12	ANN	1178	[32]
FA, Sl, Red mud	n(H ₂ O)/n(Na ₂ O), Na ₂ O, n(SiO ₂)/n(Na ₂ O), n(SiO ₂)/n(Al ₂ O ₃), n(SiO ₂)/n(CaO), L/S, CT, CD	8	KNN, SVR, BR, RFR, ETR, GBR, XGBR , DNN	557	[33]
FA, Sl, SF, RHA, Calcined clay	FA, Sl, SF, RHA, Calcined clay, NCA, NFA, W, SP, SH, SHC, SS, SiO ₂ /Na ₂ O	13	ANN	381	[34]
FA	FA, NCA, NFA, W, SP, SiO ₂ , Al ₂ O ₃ , SH, SHC, SS, Na ₂ SiO ₃ /NaOH, AA/FA, CT, HCD	14	ANN , SVR , ELM	110	[35]
FA	FA, NCA, NFA, SH, SHC, SS, SiO ₂ , Na ₂ O, Age	9	DTR, BR , ABR	154	[36]
FA	FA, NFA, NCA, SH, SHC, SS, SS/SH, SiO ₂ /Al ₂ O ₃ , AA/FA, CT, HCD, Age	12	ANN , DNN , ResNet	860	[37]
FA	SS/SH, AA/FA, SHC, CT	4	ANFIS	90	[38]
FA	FA, NCA, NFA, W, SP, SH, SHC, SS, SiO ₂ /SS, Na ₂ O/SS, CT, HCD	12	BLR	162	[39]
FA	W/FA, SP/FA, Initial CT, CD, Age, TAV, SHC, SiO ₂ /Na ₂ SiO ₃	8	GEP , MEP	311	[40]
FA	FA, NCA, NFA, W, SH, SHC, SS, CT, HCD	9	DNN , ResNet	335	[41]
FA, Sl	FA/B, Sl/B, Sodium ion/AA, Silicon ion/AA, Boron ion/AA	5	ANN , GP	114	[42]

AA: Alkaline activators, ABR: Adaptive boosting regressor, Age: Testing age, Al₂O₃: Aluminium oxide, ANFIS: Adaptive neuro-fuzzy inference system, ANN: Artificial neural network, B: Binder, BLR: Bayesian linear regression, BR: Bagging regressor, CaO: Calcium oxide, CD: Curing duration, CT: Curing temperature, DTR: Decision tree regressor, ELM: Extreme learning machine, ETR: Extremely randomized trees regressor, FA: Fly ash, Fe₂O₃: Iron trioxide, GBR: Gradient boosting regressor, GEP: Gene expression programming, GPR: Gaussian process regression, HCD: High-temperature curing duration, HEML: Hybrid ensemble machine learning model, KNN: K-nearest neighbour, K₂O: Potassium oxide, LR: Linear regression, LOI: Loss on ignition, L/S: Liquid to solid ratio, LSBR: Least-squares boosting regressor, MARS: Multivariate regression spline, MEP: Multi-expression programming, MgO: Magnesium oxide, MK: Metakaolin, Na₂O: Sodium oxide, NCA: Natural coarse aggregate, NFA: natural fine aggregate, P₂O₅: Phosphorus pentoxide, Poly2: 2nd-degree polynomial regressor, RFR: Random forest regressor, ResNet: Deep residual network, RH: Relative humidity, RHA: Rice husk ash, SH: Sodium Hydroxide, SHC: Sodium Hydroxide concentration, SiO₂: Silicon dioxide, Sl: Slag cement, SO₃: Sulphur trioxide, SP: Superplasticizer, SS: Sodium Silicate, SSA: Specific surface area, SVR: Support vector regression, TAV: Total aggregate volume, TiO₂: Titanium oxide, W: Water, XGBR: Extreme gradient boosting regressor.

efforts have aimed to model the CS of geopolymer mortars and concrete using ML techniques. The best-developed ML models are bolded in this table. Researchers have primarily focused on investigating the utilization of fly ash and slag cement in geopolymers due to their widespread availability and distinctive chemical properties. These studies have employed a diverse range of input features, with the number of features in models ranging from 4 to 21. The size of the datasets used for developing ML models has exhibited significant variability, ranging from smaller datasets with 80 samples to more extensive datasets comprising 1178 samples. The ratio of dataset size to the number of input features across these 23 studies has fluctuated between 7.6 and 102.1, with an average ratio of 39.2. This diversity underscores the importance of tailoring dataset size to specific modelling contexts, considering factors such as problem complexity, choice of regression models, data quality, and inherent noise. The average ratio of 39.2 from these studies can serve as a valuable benchmark, suggesting that, on average, these investigations have found datasets approximately 39 times larger than the number of input features to be suitable for their regression modelling tasks within the scope of this study. In addition, researchers have employed various ML techniques in these studies, with ensemble models being a prevalent choice, featured in nearly half of the conducted investigations. Remarkably, ensemble models have demonstrated superior performance compared to other ML model types in approximately one-third of the studies conducted. These insights not only provide valuable guidance regarding dataset sizing in this study's context but also underscore the effectiveness of ensemble techniques in accurately modelling the CS of geopolymers.

The existing body of research has revealed a notable deficiency in the realm of developing ML models specifically tailored to predict the CS of GRAC. In light of this gap, a comprehensive CS database for GRAC was meticulously curated by drawing upon available literature sources. Recognizing the paramount significance of database size during the model development phase, proactive measures were taken to expand the size of the database, achieved through the utilization of the tabular

generative adversarial network (TGAN) technique [43] as a means to generate synthetic data. A series of rigorous analyses were subsequently executed with the primary objective of determining the optimal size for this synthetic database. Furthermore, a diverse array of ensemble techniques, encompassing both bagging and boosting methodologies, were harnessed for the purpose of modelling the CS of GRAC. Ensemble techniques were chosen for their ability to deliver superior predictive performance, particularly in complex scenarios with non-linear data relationships. These methods effectively prevent overfitting, ensuring models perform well on unseen data. They also increase model diversity, which improves prediction accuracy by aggregating insights from multiple models. Additionally, ensemble methods enhance interpretability, offering insights into the importance of different predictors. After finding the best ML model, sensitivity analysis and a parametric study were executed to elucidate the significance of input features and comprehend the CS trends resulting from variations in these inputs. The rest of the paper is organized as follows: Section 2 presents the study framework. Detailed explanations about data preparation are given in Section 3. Section 4 discusses the ML techniques used in this study. Section 5 presents the results and corresponding discussions, followed by Section 6, which outlines the key findings and conclusions of the study.

2. Study framework

The various stages of this study are illustrated in Fig. 1, offering an overview of the research framework. The study initiates with a crucial data preparation phase, encompassing essential steps such as data gathering, data preprocessing, and data partitioning. Subsequently, a synthetic database is generated, with an examination of numerous parameters conducted to ensure the reliability of this synthetic database. Following this, the study delves into the development of eight ensemble models, comprising three bagging ensemble models and five boosting ensemble models. These models aim to predict the CS of GRAC. Several

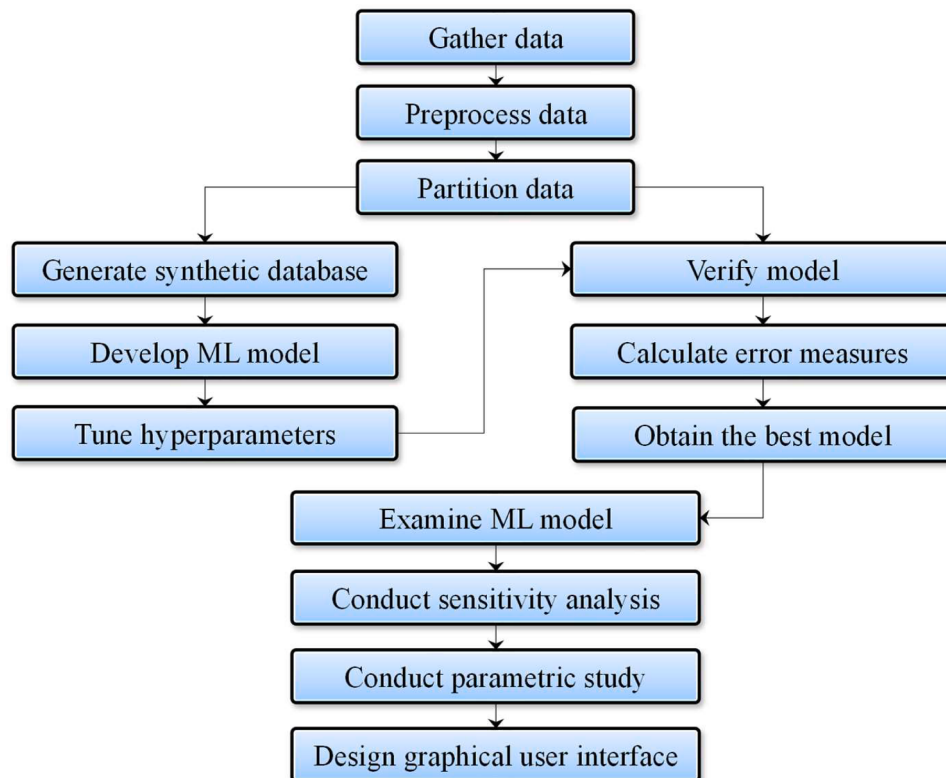


Fig. 1. The study framework.

statistical analyses are performed, and various error measures are computed for the ensemble models using the testing dataset during the verification phase. These analyses serve to assess the effectiveness of the developed models, ultimately leading to the selection of the best-performing machine learning model among the ensembles. Advancing further, a critical phase dedicated to sensitivity analysis and parametric study is undertaken. This phase investigates the roles and effects of input features on the CS of GRAC. Finally, to enhance user interaction and practical utility, a user-friendly graphical user interface (GUI) is created. More details regarding the different steps are given in the following section.

3. Data preparation

3.1. Data gathering and preprocessing

A comprehensive database of the CS of GRAC, including 314 data samples, was gathered from 14 peer-reviewed papers in the literature. The collected database included the contents of fly ash (FA), slag cement (SI), natural coarse aggregate (NCA), recycled coarse aggregate (RCA), natural fine aggregate (NFA), sodium hydroxide (SH), sodium silicate (SS), superplasticizer (SP), as well as recycled aggregate water absorption (RWA), sodium hydroxide concentration (SHC), curing temperature (CT), high-temperature curing duration (HCD), and testing age (Age) as candidate predictors, and the CS of GRAC as the dependent feature. The information on the studies used in this research is given in Table 2.

The Variance Inflation Factor (VIF) serves as a metric for measuring the extent of multicollinearity. Within this research, a VIF score exceeding 10 is regarded as an indication of significant multicollinearity, which has the potential to influence the outcomes of the model [58]. Elevated VIF values signal that a predictor variable exhibits strong correlations with other predictor variables within the model, which can result in potentially unreliable interpretations. As shown in Fig. 2(a), three input features, including NFA, RCA, and NCA, exhibit VIF scores exceeding 10, with the latter two features registering VIF scores surpassing 100. To address this issue, a new input feature, recycled aggregate ratio (RAR), was introduced as a replacement for both RCA and NCA. The VIF scores for the input features considering the newly introduced input feature were then computed, as illustrated in Fig. 2(b). In this revised scenario, all input features demonstrate VIF

scores of approximately six or less, effectively mitigating the risk of multicollinearity within this study.

3.2. Synthetic database generation

Tabular generative adversarial network (TGAN) is a method used to create synthetic data samples that mimic the statistical properties and patterns found in the main database [43]. This serves the purpose of expanding the database, which can potentially enhance the performance of ML models. Generating synthetic samples through TGAN proves particularly beneficial when dealing with limited data, a scenario commonly encountered in complex regression tasks. By introducing synthetic data, additional variability is injected into the main dataset, effectively acting as a regularization technique to prevent overfitting and improve the generalization capabilities of regression models.

TGAN operates by training two networks, a generator, and a discriminator, in a competitive manner. The generator network learns to map random noise vectors to the main data samples, generating synthetic data that closely resembles real data. In the context of TGAN, the use of a Long Short-Term Memory (LSTM) network as the generator allows the model to capture sequential dependencies within the data. Each time step in the LSTM corresponds to a feature in the tabular data, and the LSTM layer processes these features sequentially. Simultaneously, the discriminator network's role is to distinguish between real and synthetic data. It undergoes training using a combination of real data from the main dataset and synthetic data produced by the generator. The discriminator's objective is to classify data samples as either real or synthetic, and it often employs a Multi-Layer Perceptron (MLP), a feedforward neural network capable of capturing intricate relationships in tabular data. The MLP architecture in the discriminator enables it to scrutinize features for discerning real data from synthetically generated data [43].

When comparing a synthetic database generated by TGAN with the main database, it is essential to assess how accurately the synthetic data replicates the underlying patterns, distributions, and relationships inherent in the main dataset. In this study, the non-parametric Kolmogorov-Smirnov test was employed to compare the overall distributions between the input features of the main and synthetic databases [59]. Various synthetic database sizes were considered, represented by ratios to the main database size, including 0.7, 0.8, 0.9, and 1. Fig. 3 visually represents the p-values of input features obtained from the Kolmogorov-Smirnov test for different synthetic database sizes. Notably, for synthetic database size ratios of 0.7 and 0.8, all input features yielded p-values greater than 0.05, signifying that the distribution of all input features in the main and synthetic databases closely aligns. Consequently, this study selects a synthetic database size ratio of 0.8 for further analysis.

Histograms and probability distributions serve as valuable plots for evaluating the resemblance between synthetic and main databases. These plots enable the depiction of data attributes and have the potential to unveil whether the generated data closely mirrors the tendencies, arrangements, and distributions found in the main database. As demonstrated in Fig. 4, the histograms and probability distributions of the input features in both the main and synthetic databases display a remarkable likeness in their shapes and configurations. This resemblance strongly indicates that the generated data effectively captures the main data's distributional and probabilistic traits, successfully replicating the statistical characteristics inherent in the main database.

Pairwise feature correlation analysis involves calculating Pearson correlation coefficients between pairs of features. This analysis helps understand the relationships and dependencies between different features in the dataset. The correlation matrix of input and output features for the combined database is shown in Fig. 5. In general, the absolute correlation coefficient between two features above 0.8 indicates a high degree of linear relationship between them. This could imply that changes in one variable are often associated with predictable changes in

Table 2

Information from studies used in the database of this research.

Sample ID	Data number	Data percentage (%)	Ref
S-0.8, S-0.7, S-0.6, S-0.5, S-0.4, S-0.3	24	7.64	[44]
GPC-FG30, GPC-FG40, GPC-FG50, GPC-FG60	8	2.55	[45]
M20, M30, M40, M50, M60	10	3.18	[46]
M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12	44	14.01	[47]
F60G40R0, F40G60R0, F20G80R0, F40G60R25, F40G60R50, F40G60R75, F40G60R100	21	6.69	[48]
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	24	7.64	[49]
R0, R25, R50, R100, r-R0S15, r-R25S15, r-R50S15, r-R100S15, r-R0S30, r-R25S30, r-R50S30, r-R100S30, a-R0S15, a-R25S15, a-R50S15, a-R100S15, a-R0S30, a-R25S30, a-R50S30, a-R100S30	80	25.48	[50]
GNAC/B0, GRAC/B0	6	1.91	[51]
S-0.40, S-0.45, S-0.50	24	7.64	[52]
S00, S10, S20, S30, S00R50, S10R50, S20R50, S30R50, S00R100, S10R100, S20R100, S30R100	36	11.46	[53]
1, 2, 3, 4, 5, 6, 7, 8	8	2.55	[54]
GL8, GL12, GL16, GR8, GR12, GR16	6	1.91	[55]
GPC0, GPC15, GPC30, GPC50	8	2.55	[56]
GRC0, GRC50, GRC100	15	4.78	[57]

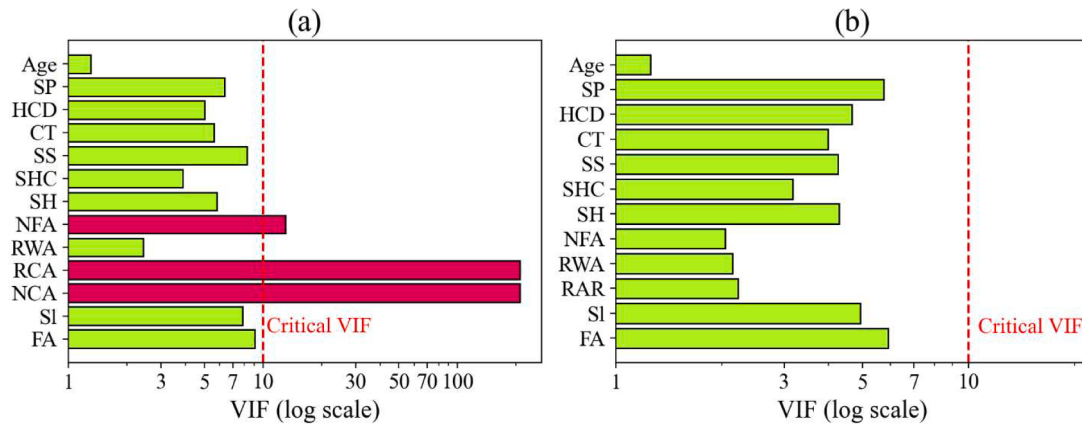


Fig. 2. The VIF scores for the a) initial input features and b) final input features.

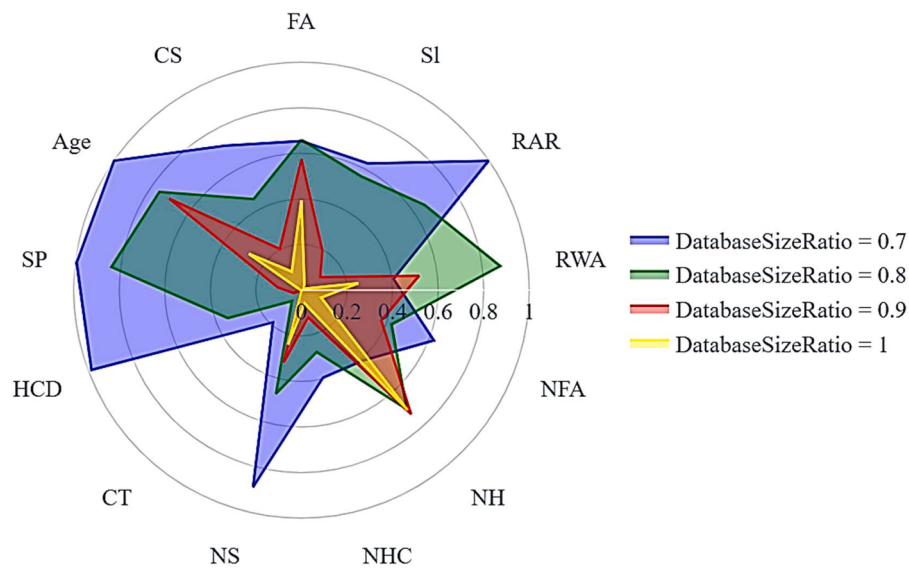


Fig. 3. The p-values of the Kolmogorov-Smirnov test of all input features for different synthetic database sizes.

the other. Since the VIF scores of FA, SI, CT, and HCD features are below six, their impact on multicollinearity is not substantial, even though the absolute correlation coefficients between SI-FA and CT-HCD are high. This is because the increase in standard errors due to multicollinearity is relatively small, and thereby the impact on the ML model's predictions and interpretations is not significant.

Table 3 presents the statistical characteristics of both input and output features for the main, synthetic, and combined databases. A detailed examination of these statistical parameters for the main and synthetic databases reveals a notable degree of alignment, suggesting an acceptable agreement between the two. This observation implies that the features within the synthetic database have been effectively distributed, adhering closely to the underlying feature distributions of the main database. It is worth mentioning that every data sample incorporates fly ash, underscoring its prevalence in the database. Moreover, approximately 63 % of the data samples include slag cement, while an even higher proportion, approximately 78 %, comprises natural coarse aggregate. Additionally, about 37 % of the data samples underwent curing at elevated temperatures, ranging from 40 to 80 °Celsius. Delving into the CS of GRAC, it was found that the average CS across all ages stands at 43.4 MPa. However, when explicitly considering a testing age of 28 days, the average CS significantly elevates to 49.2 MPa. This substantial increase highlights the potential of GRAC to be categorized as high-strength concrete.

4. ML techniques used in this study

In this research, a set of eight ensemble methodologies was employed, comprising three bagging and five boosting ensemble techniques, to create predictive models for the CS of GRAC. The rationale behind selecting the decision tree regressor (DTR) as the fundamental building block for our ensemble models stemmed from its favorable attributes, including interpretability, robustness, and capacity to effectively handle non-linear relationships. In this section, the pivotal role of the DTR as the foundational component of ensemble techniques is elucidated. Subsequently, three specific bagging ensemble techniques are explained, namely the random forest regressor (RFR), bagging regressor (BR), and extremely randomized trees regressor (ETR). Following that, five distinct boosting ensemble techniques are expounded: the adaptive boosting regressor (ABR), gradient boosting regressor (GBR), extreme gradient boosting regressor (XGBR), categorical boosting regressor (CBR), and light gradient boosting regressor (LBR), providing valuable insights into the role each played in enhancing the predictive capabilities of ML models.

4.1. DTR

A DTR is constructed in a hierarchical structure, including nodes and leaves, that aims to predict a continuous target feature. Nodes represent

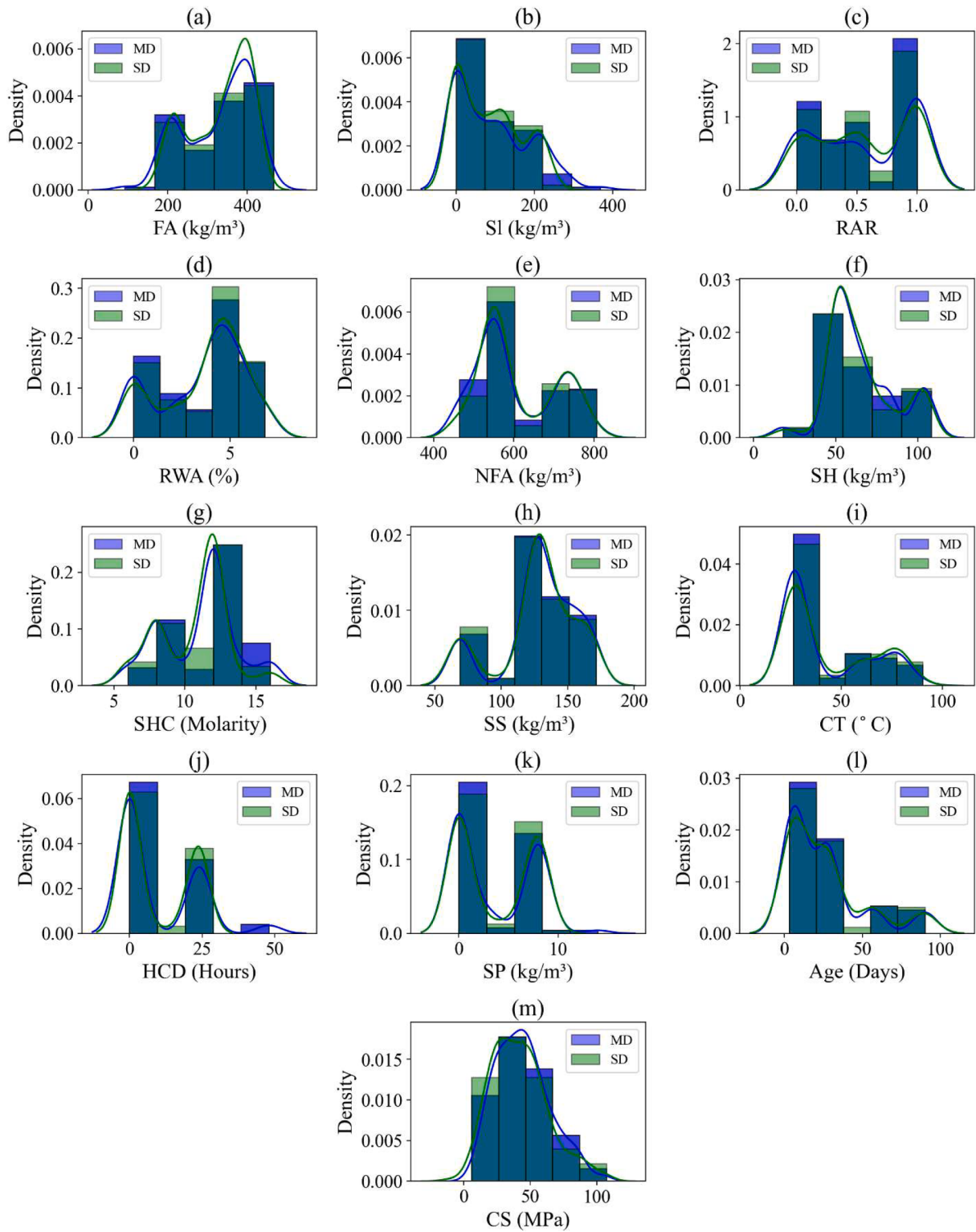


Fig. 4. The distribution histograms of the main database (MD) and synthetic database (SD).

decision points based on features, while leaves represent the predicted numerical values [60]. At each node, the DTR algorithm selects a feature and a threshold to split the data into two subsets. The splitting criterion, such as Mean Squared Error (MSE), aims to minimize the variance of the target feature within each subset. The tree-building process is recursive. It starts at the root node with the entire dataset and recursively splits it into subsets at each node, moving down the tree. The splitting process continues until a stopping condition is met, which may include a

maximum tree depth, a minimum number of data points per leaf, or a minimum reduction in variance. When the tree-building process is complete, each leaf node contains a constant value (typically the mean of the target values in the leaf), which is the predicted numerical value for the target feature within that leaf's region.

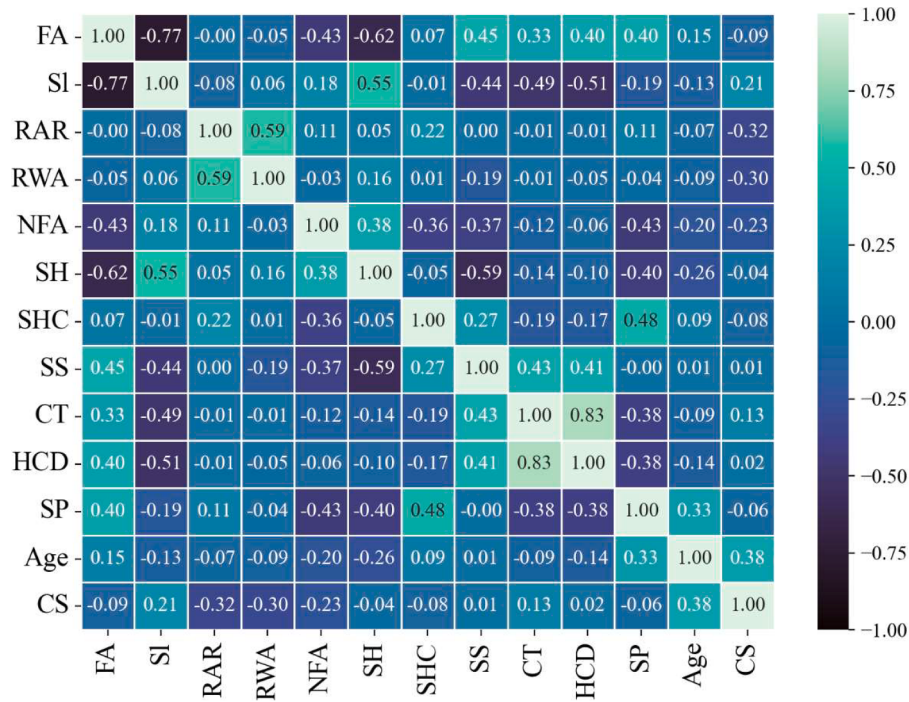


Fig. 5. Correlation matrixes of input and output features for the combined database.

Table 3

Statistical parameters of input and output features for the main, synthetic, and combined databases.

Features	Mean*	Standard deviation	min	25 %	50 %	75 %	Max
FA (kg/m ³)	327.42/332.73/ 329.76	85.71/75.34/81.27	92/175/92	256/280/276	350/350/350	400/400/400	466/428/466
SI (kg/m ³)	92.49/85.38/89.35	92.96/79.62/87.32	0/0/0	0/0/0	60/65.68/60	165/132.97/ 145.55	368/263.74/368
RAR	0.56/0.56/0.56	0.41/0.39/0.40	0/0/0	0.20/0.25/0.25	0.5/0.5/0.5	1/1/1	1/1/1
RWA (%)	3.56/3.68/3.61	2.19/2.14/2.17	0/0/0	2.12/2.12/2.12	4.22/4.22/4.22	4.99/4.99/4.99	6.78/6.78/6.78
NFA (kg/m ³)	605.19/610.09/ 607.35	100.31/96.19/ 98.45	465/465/465	540/550/550	560/560/560	720/720/720	806/806/806
SH (kg/m ³)	65.85/64.95/65.46	20.72/20.14/20.46	18/18/18	51.5/51.5/51.5	61/61/61	80/75.64/80	108/103/108
SHC (Molarity)	11.04/10.67/10.88	2.63/2.41/2.54	6/6/6	8/8/8	12/12/12	12/12/12	16/16/16
SS (kg/m ³)	128.87/127.10/ 128.09	28.83/29.27/29.01	69/69/69	120/121.8/120	128.5/128.5/ 128.5	150/150/150	171.4/171.4/171.4
CT (oC)	41.91/43.93/42.80	21.23/21.57/21.39	26.37/26.37/ 26.37	27/27/27	27/27/27	60/60.21/60	90/89/90
HCD (Hours)	9.48/9.18/9.35	13.45/11.34/12.55	0/0/0	0/0/0	0/0/0	24/24/24	48/24/48
SP (kg/m ³)	3.60/3.69/3.64	3.94/3.80/3.88	0/0/0	0/0/0	1.42/1.52/1.42	8/8/8	13.986/8.99/13.99
Age (Days)	24.67/26.10/25.30	24.69/25.13/24.87	3/3/3	7/7/7	14/23.5/20	28/28/28	90/90/90
CS (MPa)	44.37/42.23/43.43	20.36/21.09/20.70	5.97/5.97/5.97	28.96/25.63/ 26.90	42.28/40.36/ 41.94	56.89/54.03/ 55.02	107.15/107.15/ 107.15

* The first, second, and third values in (-/-/-) show the statistical parameters of the main, synthetic, and combined databases, respectively.

4.2. RFR

An RFR is an ensemble ML technique that enhances prediction accuracy and robustness by combining the forecasts of multiple base DTRs. In this approach, each RFR is constructed using a random subset of the training data (known as bootstrapping) and a random subset of the features. This randomness introduces diversity among the constituent trees. The RFR aggregates the predictions made by all the individual DTRs to arrive at the final prediction which helps mitigate overfitting when compared to using single DTRs. The DTR offers several hyperparameters that can be fine-tuned to optimize model performance [61]. These include the number of trees (NT), the maximum tree depth (MTD), as well as parameters like the minimum samples required for a split (MSS), the minimum samples required for a leaf (MSL), and the

maximum number of features considered at each split (MNF).

4.3. BR

BR seeks to enhance the precision and reliability of regression models by combining forecasts generated by multiple base DTRs. BR and RFR share similarities in their approaches. However, the principal distinction lies in the fact that BR does not encompass feature selection, typically utilizing all available features when constructing each base DTR [62].

4.4. ETR

An ETR is a bagging ensemble method similar in behavior to an RDR.

What sets the ETR apart is the degree of randomness injected during the tree construction process [63]. Like the RFR, it employs bootstrap sampling to generate multiple subsets of the training data. However, the ETR takes randomness to a greater extent by introducing extra randomness when determining split points for each node within the decision trees. Unlike the traditional DTR, which selects the optimal split from a subset of features at each node, the ETR opts for random split points without evaluating all potential splits. This adds an extra layer of diversity and randomness among the trees.

4.5. ABR

ABR is a boosting ensemble technique crafted to enhance the accuracy and robustness of regression models. It operates sequentially, constructing a series of base DTRs to rectify the errors made by preceding models. In this algorithm, each data point is assigned a weight [64]. Initially, all data points carry equal weights. However, as the algorithm advances, these weights are adjusted to assign greater importance to data points that were associated with higher prediction errors in earlier base models. This approach employs weighted voting when making the ultimate prediction, effectively weighting predictions based on the performance of each base model. Models that perform better are granted higher weights in the final prediction. ABR offers hyperparameters that can be fine-tuned to optimize model performance, including parameters like the number of trees (NT) and the learning rate (LR), which govern the contribution of each base DTR to the final prediction.

4.6. GBR

GBR is a boosting ensemble method engineered to craft precise and resilient regression models. Similar to ABR, GBR builds an ensemble model by progressively introducing base DTRs to the ensemble [65]. During each iteration, this algorithm identifies the errors in the existing ensemble and trains a new base DTR to rectify these errors. It achieves this by fitting the new DTR to the residual errors of the prior ensemble. To do this, the technique employs a gradient descent optimization approach to minimize the loss function. This entails computing the gradient of the loss function concerning the ensemble's predictions and adjusting the predictions in the direction that minimizes the loss. Much like ABR, GBR employs weighted voting for computing the final prediction. The hyperparameters governing this algorithm are akin to those used in the RFR. Additionally, there exists a learning rate (LR) parameter that regulates the step size within the gradient descent process.

4.7. XGBR

XGBR represents an improved and fine-tuned version of the conventional GBR, delivering outstanding predictive precision and resilience. In its approach, it integrates L_1 (Lasso) and L_2 (Ridge) regularization techniques into the objective function to curb overfitting and increase the model's capacity for generalization [66,67]. This algorithm retains the hyperparameters present in GBR, and additionally, it introduces L_1 and L_2 regularization parameters. As a result, XGBR is more robust to noisy data and outliers due to its regularization terms and the ability to penalize extreme values.

4.8. CBR

CBR is especially adept at handling tabular data that includes a mix of numerical and categorical features. This algorithm incorporates built-in techniques to mitigate overfitting, reducing the sensitivity to hyperparameter tuning. This property can result in more robust models, particularly in scenarios involving noisy data [68]. In this study, several key hyperparameters were regarded as influential for CBR, including the number of trees (NT), maximum tree depth (MTD), learning rate (LR), L_2

Regularization term, random strength (RS), and bagging temperature (BT).

4.9. LBR

LBR is recognized for its efficiency, speed, and adeptness at managing extensive datasets. It sequentially introduces trees to rectify the errors committed by prior trees, progressively refining the model's performance using a gradient-based optimization approach [69]. LBR employs a histogram-based algorithm to identify optimal node splits, thereby reducing memory consumption and accelerating training. This algorithm embraces a leaf-wise tree growth strategy, where it extends the tree by dividing nodes that yield the most substantial reduction in the loss function. While this approach may result in deeper trees and potentially enhanced model performance, it does raise the prospect of overfitting. Furthermore, it leverages regularization methods like L_1 and L_2 regularization to prevent overfitting and elevate model generalization.

4.10. Hyperparameters' tuning

The hyperparameters of ensemble models are predefined parameters that are not learned from data but significantly influence how the ensemble model performs. In this research, both k -fold cross-validation and the Bayesian optimization algorithm are employed to fine-tune these hyperparameters. The k -fold cross-validation offers each data point an opportunity to contribute to both the training and validation phases, effectively mitigating the risk of overfitting [70]. It operates by dividing the development dataset into k roughly equal-sized folds. The model undergoes k rounds of training and validating, with each round employing a different fold as the validation set while the remaining $(k-1)$ folds serve for training. The ensemble model's overall performance is determined by calculating the average validation error across these k evaluations.

Bayesian optimization is a powerful global optimization technique [71] used to discover the optimal hyperparameters for ensemble models efficiently. It initiates with an initial set of hyperparameters and assesses the average validation error associated with these settings. Based on these observations, Bayesian optimization constructs a probabilistic surrogate model of the error function. It then selects the next set of hyperparameters to evaluate, aiming to minimize this error function. Importantly, it strikes a balance between exploration (sampling in regions with high uncertainty) and exploitation (sampling in regions with anticipated high performance). This iterative process continues until convergence towards the optimal hyperparameter set or the satisfaction of predefined stopping criteria.

4.11. Model assessment

Assessing the performance of ML models using different error metrics is necessary for evaluating their predictive capabilities. In this research, a range of error metrics was employed to comprehensively evaluate the ensemble models created. These metrics encompass the root mean squared error (RMSE), normalized root mean squared error (NRMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2), defined as follows:

$$RMSE = \sqrt{\frac{1}{SN} \sum_i (CS_i^E - CS_i^P)^2} \quad (1)$$

$$NRMSE = \frac{RMSE}{CS} \quad (2)$$

$$MAE = \frac{1}{SN} \sum_i |CS_i^E - CS_i^P| \quad (3)$$

$$MAPE = \frac{100}{SN} \sum_i \left| \frac{CS_i^E - CS_i^P}{CS_i^E} \right| \quad (4)$$

$$R^2 = 1 - \frac{\sum_i (CS_i^E - CS_i^P)^2}{\sum_i (CS_i^E - \bar{CS})^2} \quad (5)$$

In the above equations, CS_i^E represents the experimental CS value of the i th data sample, and CS_i^P signifies the predicted CS value of the same data sample; SN denotes the total number of data samples; and $\bar{CS}$ represents the average CS across the entire database. RMSE serves as an indicator of the average magnitude of discrepancies between predicted and experimental CS values. It quantifies how closely the model's predictions align with the observed data, with smaller values indicating better performance. MAE, on the other hand, represents the average absolute differences between predicted and experimental CS values. It provides a straightforward measure of prediction accuracy, with lower values signifying improved model performance. MAPE calculates the average percentage disparity between predicted and experimental CS values, offering insight into the relative error in predictions, particularly useful for assessing prediction accuracy. R^2 evaluates the goodness of fit between the model predictions and the experimental results, indicating the proportion of variance in the dependent feature explained by the independent features. A R^2 value closer to 1 suggests a strong fit, while a lower value indicates a weaker fit. Lastly, NRMSE assesses the dispersion of predicted CS values around the actual CS values, helping gauge the spread of predictions. It can be categorized into specific intervals, such as ($NRMSE \leq 0.1$), ($0.1 < NRMSE \leq 0.2$), ($0.2 < NRMSE \leq 0.3$), and ($0.3 < NRMSE$), each corresponding to "Excellent", "Good", "Fair", and "Poor" predictions, respectively.

5. Results and discussion

In this research, eight different ensemble techniques were employed to model the CS of GRAC. During the development phase, 80 % of the combined database, which includes both the main and synthetic databases, was utilized, leaving the remaining 20 % for the testing phase. The development dataset was used to set the hyperparameters of the developed models, while the testing dataset was employed to examine the performance of the developed model against unknown data. For hyperparameters' tuning, k -fold cross-validation with five folds was conducted and the Bayesian optimization algorithm was employed over the development dataset for 50 iterations. During the development phase, four folds were randomly selected for training an ensemble model, with the remaining fold used for model validation. By varying the order of folds and utilizing all folds in the training and validation

phases, a total of five ensemble models were developed and the average performance of these ensemble models in the validation phase indicates the error associated with the preset hyperparameter values. Choosing five folds for k -fold cross-validation was based on computational efficiency, an optimal bias-variance trade-off, and alignment with other studies in the literature [72,73]. Additionally, ten different initial conditions (seeds) during the model development phase of the ensemble models were considered. Fig. 6 presents both the average RMSE values and the best RMSE values from ten independent repetitions of ensemble algorithms in both the development and testing phases. Generally, the GBR, XGBR, CBR, and LBR models exhibited superior average and best performance compared to the other four ensemble models. On the contrary, the ABR model consistently performed the poorest among all the developed ensemble models in both the development and testing phases. It's worth noting that the ETR model displayed a larger performance gap between the development and testing phases, suggesting a comparatively lower generalization capability for this algorithm.

Table 4 provides a comprehensive overview of the optimal hyperparameters that have been derived for the top-performing ensemble models. Notably, the XGBR stands out as having a relatively straightforward configuration, employing 244 trees with a maximum depth of 4. This simplicity in the ensemble model design contributes significantly to mitigating the risk of overfitting when compared to the other models. On the contrary, the bagging ensemble models with greater tree depth tend to produce intricate decision trees, a factor that could potentially escalate the likelihood of overfitting. Therefore, it is essential to strike a balance between model complexity and performance to ensure robust and accurate predictions.

Table 5 provides a comprehensive overview of the error metrics for the top-performing ensemble models during both the development and testing phases, with the best error metrics bolded in the table. In the

Table 4

The optimal hyperparameters of the best-developed ensemble models.

Models	Optimal values
RFR	NT=955, MTD=15, MSS=2, MSL=1, MNF=3
BR	NT=374, MTD=19
ETR	NT=664, MTD=18, MSS=2, MSL=1, MNF=7
ABR	NT=423, LR=1.9912
GBR	NT=695, MTD=12, MSS=3, MSL=4, MNF=2, LR=0.0206
XGBR	NT=244, MTD=4, MNF=7, MSR*=0.8337, LR=0.0800, $L_1=0.3034$, $L_2=0.3671$
CBR	NT=500, MTD=9, BT=2.0343, $L_2=0.0126$, LR=0.1987
LBR	NT=528, MTD=12, MSL=2, MNF=8, MSR=0.5436, LR=0.0994, $L_1=0.4100$, $L_2=0.6612$

MSR: Maximum sample ratio used in each node.

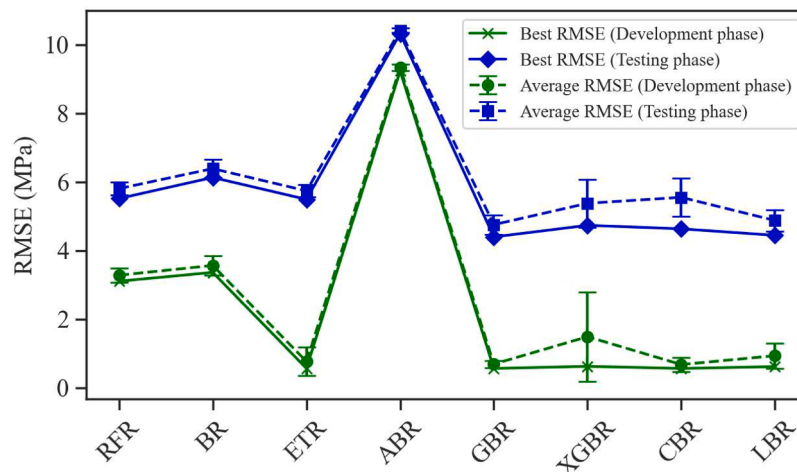


Fig. 6. The average and best performance of the developed ensemble models in the development and testing phases.

Table 5
Error metrics of the developed ensemble models for the CS of GRAC.

Models	Development phase					Testing phase				
	RMSE (MPa)	MAE (MPa)	MAPE (%)	R ²	NRMSE	RMSE (MPa)	MAE (MPa)	MAPE (%)	R ²	NRMSE
RFR	3.1067	1.5882	5.8993	0.9763	0.0725	5.6617	3.1906	9.5231	0.9361	0.1240
BR	3.3630	1.8885	6.9003	0.9722	0.0785	6.3605	3.5136	10.4146	0.9193	0.1393
ETR	0.5632	0.0656	0.1272	0.9992	0.0131	5.6599	2.3650	6.6237	0.9361	0.1240
ABR	9.2359	7.8239	27.4928	0.7906	0.2155	10.4708	8.5151	27.1727	0.7814	0.2293
GBR	0.5633	0.0730	0.1494	0.9992	0.0131	5.3894	1.9313	5.6639	0.9421	0.1180
XGBR	0.6225	0.2476	0.7038	0.9990	0.0145	4.7515	1.8565	5.3076	0.9550	0.1041
CBR	0.5630	0.0609	0.1132	0.9992	0.0131	6.0431	2.0053	5.2951	0.9272	0.1323
LBR	0.6146	0.2299	0.6423	0.9991	0.0143	4.8976	1.9246	5.3042	0.9522	0.1073

development phase, the CBR model stands out as the leader, with the lowest RMSE, MAE, and MAPE values, closely trailed by the ETR and GBR models. These three models consistently exhibit significantly superior performance compared to the other models. Furthermore, with the exception of the ABR model, all models achieve an “Excellent” rating in NRMSE, indicating an excellent fit to the development data. The R² values, which signify the goodness of fit, approach one for all models in the development phase, except for the ABR model, highlighting their strong fitting capabilities. Transitioning to the testing phase, it is reasonable to observe that all models experience higher error values when confronted with unseen data. Notably, the NRMSE values for all models, except the ABR model, maintain a “Good” performance level when applied to new data samples. Among these models, the XGBR, LBR, and GBR models excel in the testing phase, delivering outstanding results in terms of RMSE, NRMSE, MAE, and R². Furthermore, the MAPE values for the GBR, XGBR, CBR, and LBR models consistently remain below 6 % in their predictions during the testing phase. The R² values for all models, except the ABR model, continue to register as high (above

0.9), underscoring their ability to elucidate a substantial portion of the variance in the CS, even in this unseen data scenario. The superiority of the XGBR and LBR models in modelling the CS of GRAC during the development phase can be attributed to 1) their efficacy in handling various types of data, especially their effectiveness in managing sparse data; 2) the use of regularization techniques to penalize complex models, thus preventing overfitting; and 3) their high speed compared to the GBR model, which can lead to better performance on datasets with complex relationships and interactions among features.

The Taylor diagram offers a visual and comprehensive means to assess and compare multiple ML models against a reference dataset, aiding in the identification of the top-performing models in replicating observed data patterns and statistics. Within this diagram, the “Actual” point signifies the reference dataset. At the same time, each model is represented as a data point, with its position indicating how effectively it reproduces the statistical attributes of the reference data. The distance from the central point corresponds to the model’s RMSE concerning the reference data, and the azimuthal angle reflects the correlation between

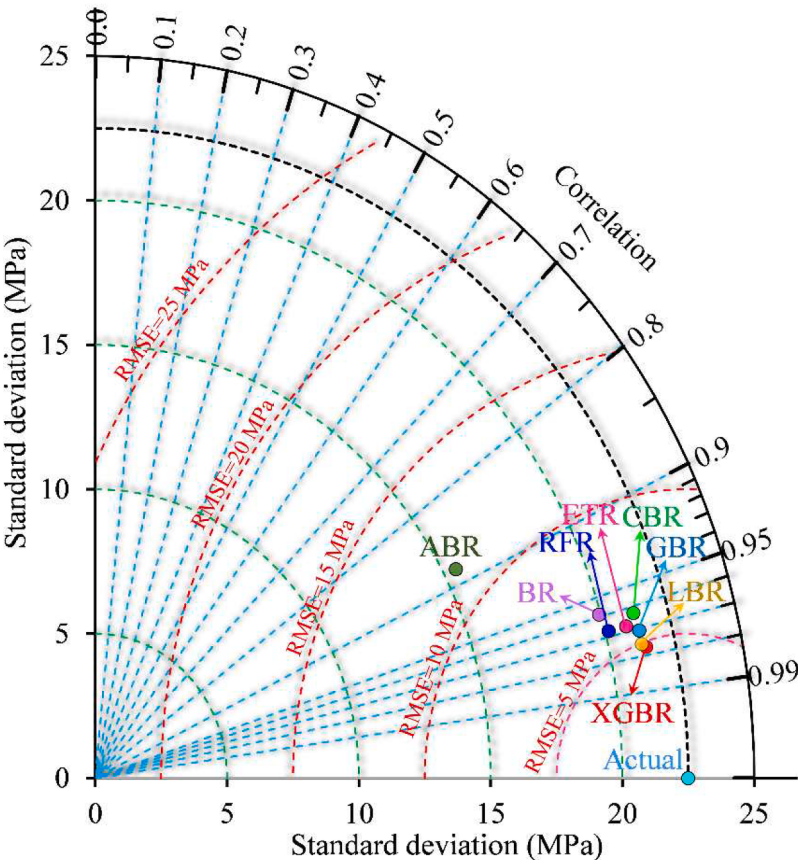


Fig. 7. Taylor diagram of the developed ensemble models.

the model's predictions and the reference dataset. In the context of this study, the testing phase Taylor diagram (depicted in Fig. 7) demonstrates that the XGBR model outperforms the others, displaying the shortest distance to the "Actual" point and showcasing the best performance. The LBR and GBR models trail closely behind, whereas the ABR model shows a significantly lower performance. This indicates that the XGBR model adeptly strikes a balance between accuracy and generalization despite its relative simplicity compared to the other ensemble models.

Fig. 8 presents a comprehensive view of the experimental CS alongside the CS predicted by the ensemble models for each testing data sample, including the ratio of predicted CS to experimental CS across these samples. The figure demonstrates a strong alignment between the experimental CS values and the predictions made by the ensemble models, with the exception of the ABR model. In most cases across the ensemble models, there is only one testing data sample displaying a notably high ratio of predicted to experimental CS, hinting at the possibility of it being an outlier or noisy data point. On average, the predicted CS tends to slightly overestimate the actual CS of GRAC, as

indicated by the predicted to experimental CS ratios exceeding one for all ensemble models. Notably, among the ensemble models developed, the CBR, LBR, and XGBR models stand out with predicted to experimental CS ratios of 1.0177, 1.0183, and 1.0188, signifying their superior performance in terms of predictive accuracy.

SHAP values, originating from cooperative game theory and applied in ML, serve as a technique to gauge the significance of individual input features within predictive models. They hold great value in comprehending intricate model behaviors and find extensive application in enhancing the interpretability of ML models [74]. The method operates by introducing random permutations to feature values while maintaining the stability of others, consequently calculating the mean disparity in model predictions. A higher positive SHAP value indicates a feature's propensity to elevate predictions, whereas a lower negative value signifies its inclination to lower predictions. These values offer comprehensive insights on both a global and local scale concerning feature importance. Fig. 9 demonstrates the mean SHAP values attributed to input features, as acquired from the three top-performing ensemble models, i.e., the XGBR, LBR, and GBR models. Notably, there exists a

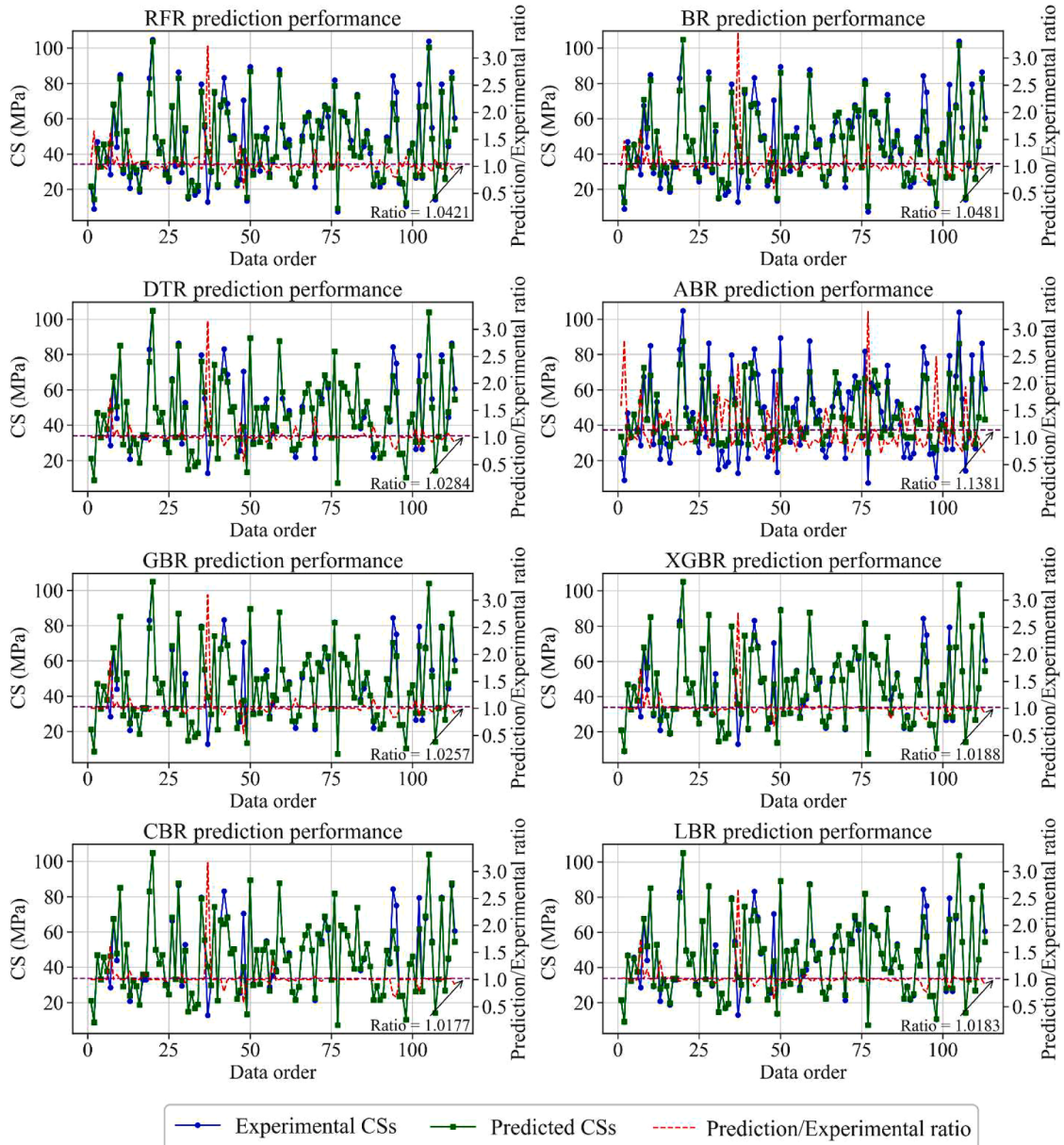


Fig. 8. Comparison between the experimental and the predicted CSs.

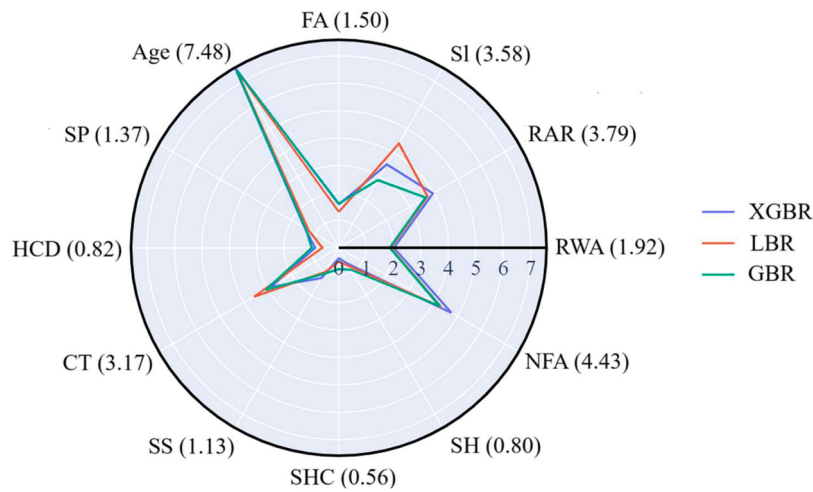


Fig. 9. Input features' importance obtained by the three best ensemble models.

robust alignment in the feature ranking across various ML models, underscoring the stability of these rankings. The average absolute SHAP values are enclosed in parentheses within the figure. Concrete age, boasting a SHAP value of 7.48, emerges as the most influential feature by a substantial margin in its impact on the CS of GRAC. Following concrete age is the content of natural fine aggregate, with a SHAP value of 4.43, securing the second position. Moreover, the recycled aggregate ratio, slag cement content, and curing temperature, exhibiting SHAP values ranging from 3.79 to 3.17, occupy subsequent positions in terms of influence.

Due to the interplay of input features on the CS of GRAC, bivariate effects of these inputs were examined through a counterplot displayed in Fig. 10. In general, increasing the binder content, including fly ash and slag cement contents, in the mixture enhances the CS of GRAC which is in line with findings of [44]. These SCMs are rich in silica and alumina, crucial components for geopolymerization. When mixed together, they experience a synergistic geopolymerization reaction, producing a more robust geopolymer binder that enhances particle adhesion and increases CS. However, the highest CS was achieved when the fly ash content ranged from 260 kg/m³ to 330 kg/m³, and the slag cement content exceeded 190 kg/m³ within the database's limitations. Notably, the binder amount within the database falls between 250 kg/m³ and 666 kg/m³, rendering the results in the extreme upper right and lower left regions of Fig. 10(a) inconclusive. According to Fig. 10(b), an increase in recycled aggregate water absorption diminishes the CS of GRAC due to the recycled aggregate's tendency to absorb a significant amount of mixing water, which aligns with the findings of [14]. This leaves insufficient water for the essential chemical reactions required for geopolymerization. Optimal CS is achieved with a recycled aggregate ratio of less than 0.3, and ratios between 0.65 and 0.8 also yield promising CS, aligning with prior research [47]. Based on Fig. 10(c), sodium hydroxide concentrations exceeding approximately 12 expedite geopolymerization reactions, resulting in faster and more complete geopolymer bond formation, which aligns with the findings of [55]. This yields a denser and stronger concrete matrix, contributing to higher CS. According to Fig. 10(d), optimal results are observed when sodium hydroxide ranges between 70 kg/m³ to 100 kg/m³ and sodium silicate falls between 125 kg/m³ and 135 kg/m³. These ranges strike a balance between providing sufficient hydroxide ions for initiating reactions while maintaining the necessary silicate and alumina content for geopolymer formation. Based on Fig. 10(e), elevated curing temperatures above 40 Celsius, combined with curing durations between 8 and 20 h, enhance the CS of GRAC. Higher temperatures can expedite geopolymerization reactions, resulting in faster and more extensive binder formation. The results align with the findings of [49]. Longer curing durations allow for better

development of the geopolymer matrix, enhancing the bond between aggregates and the binder.

On the contrary, prolonged curing durations exceeding 20 h can negatively impact CS due to various factors, including excessive moisture loss, microstructural changes, delayed strength gain, increased porosity, internal stresses, and microcracking, all contributing to reduced CS. As observed in Fig. 10(f), increasing the proportion of natural fine aggregate decreases the CS of GRAC due to dilution effects. A higher concentration of natural fine aggregate dilutes the geopolymer binder, reducing the availability of reactive materials needed to form the geopolymer matrix, which is crucial for binding aggregate particles. From Fig. 10(g), it can be observed that as the testing age increases, the CS of GRAC also rises since geopolymerization is a process that depends on time. As seen in Fig. 10(h), CS benefits when SP is below about 8 kg/m³, enhancing mix stiffness and minimizing bleeding. Conversely, excessive SP dosage beyond 8 kg/m³ can lead to over-fluidity and segregation in the mix, potentially weakening the concrete and diminishing CS.

Fig. 11 showcases the graphical user interface (GUI) meticulously crafted for the purpose of modelling the CS of GRAC. This user-friendly interface empowers individuals to interact directly with the research outcomes, simplifying the process of calculating the CS values for GRAC. This strategic and user-centric design goes beyond scientific rigor and ensures that the research findings can be readily utilized in practical, real-world situations. The designed GUI serves as a bridge between the complexities of the research and its practical implementation, making valuable insights more accessible and actionable for a broader audience. For the given example in Fig. 11, the CS predicted by the XGBR and LGBR models are close, which is consistent with the error metrics provided in Table 5. The relatively small variance between these model predictions indeed suggests lower uncertainty in the predicted CS for GRAC for the given input values.

To bridge the gap between the theoretical aspects of our research and its practical application in the field, the developed ML tool for predicting the CS of GRAC can be applied by engineers in several impactful ways: 1) Engineers can use the ML tool to rapidly assess and optimize GRAC mix designs for various construction projects; 2) The tool can assist in selecting the most suitable materials based on their availability and the environmental impact of their production; 3) The tool can serve as an educational resource for engineers and researchers to understand the behavior of GRAC under different conditions.

6. Conclusions

Geopolymer binders, comprised of fly ash and slag cement, along

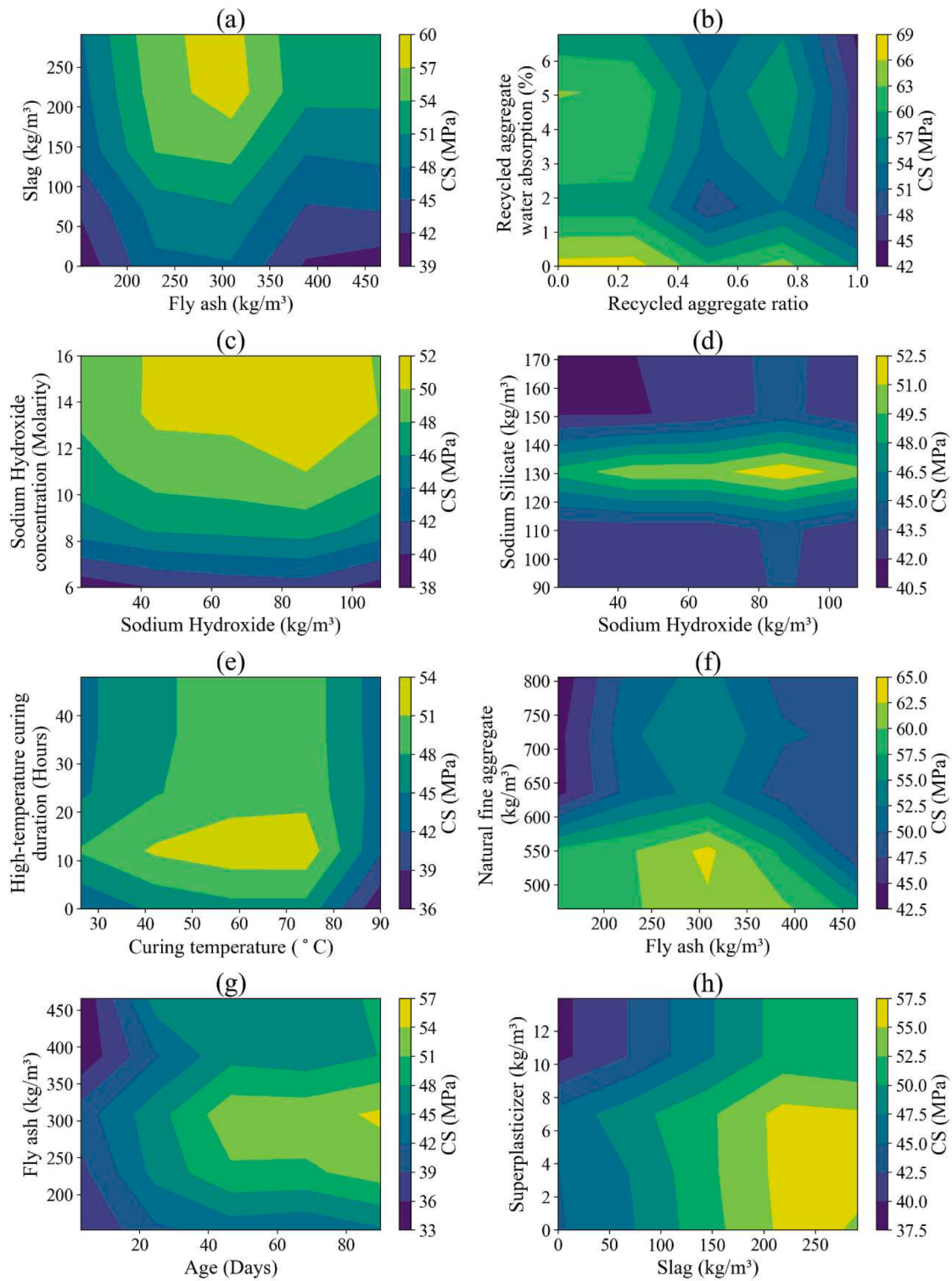


Fig. 10. The effect of a) SI-FA, b) RWA-RAR, c) SHC-SH, d) SS-SH, e) HCD-CT, f) NFA-FA, g) FA-Age, and h) SP-SI on the CS of GRAC.

with recycled aggregate, are environmentally friendly materials, aligning with sustainable development principles. This research focused on the development of several ensemble machine learning (ML) models to predict the compressive strength (CS) of geopolymer recycled aggregate concrete (GRAC). Various factors, including the quantities of fly ash, slag cement, natural fine aggregate, sodium hydroxide, sodium silicate, and superplasticizer, as well as the ratio of recycled coarse aggregate, water absorption of recycled aggregate, concentration of sodium

hydroxide, curing temperature, high-temperature curing duration, and testing age, were considered as predictors for GRAC's CS. The key findings of this study can be summarized as follows:

- Creating a synthetic GRAC database, which was 80 % the size of the main GRAC database, proved to be the optimal size for effectively increasing the database size while maintaining its statistical

Fig. 11. The graphical user interface designed for modelling the CS of GRAC.

properties. This ratio was selected after examining different ratios of 0.7, 0.8, 0.9, and 1.

- Among the eight developed ensemble models, the extreme gradient boosting regressor (XGBR) stood out with its simplicity, featuring 244 trees and a tree depth of 4. It exhibited the lowest error measures, including a root mean squared error of 4.7515 MPa, mean absolute percentage error of 5.3076 %, coefficient of determination of 0.9550, and a predicted-to-experimental CS ratio of 1.02. This highlights the effectiveness of regularization techniques in simplifying the model and reducing overfitting risks.
- The sensitivity analysis, conducted by calculating SHAP values, identified the testing age as the most influential feature on the CS of GRAC, with a SHAP value of 7.48. Natural fine aggregate followed as the second most influential feature, with a SHAP value of 4.43. Additionally, the recycled aggregate ratio, content of slag cement, and curing temperature were ranked next in terms of influence, with SHAP values ranging from 3.79 to 3.17.
- In a bivariate parametric study, optimal CS was achieved when the fly ash content ranged from 260 kg/m³ to 330 kg/m³, and the slag cement quantity exceeded 190 kg/m³, within the limitations of the database. An optimal CS was observed with a recycled aggregate ratio of less than 0.3, and ratios between 0.65 and 0.8 yielded promising CS. Sodium hydroxide concentrations exceeded approximately 12 expedited geopolymerization reactions, resulting in faster and more complete geopolymer bond formation. Optimal CS was achieved when sodium hydroxide ranged between 70 kg/m³ to 100 kg/m³, and sodium silicate fell between 125 kg/m³ and 135 kg/m³. Elevated curing temperatures above 40 Celsius, combined with curing durations between 8 and 20 h, enhanced the CS of GRAC. Increasing the proportion of natural fine aggregate decreased the CS of GRAC due to dilution effects. Additionally, increasing the testing age continually boosted the CS of GRAC, reflecting the time-dependent nature of the geopolymerization process. CS benefited when superplasticizer content was below about 8 kg/m³, improving mix stiffness and reducing bleeding.

The CS predictions using the ML models in this study are reliable for

predictive tasks within the range of the training data and the reliability decreases with out-of-range input data. Future work could involve expanding the training dataset to cover a broader range of conditions, which would enhance the model's generalization capabilities and further minimize the uncertainty associated with predictions near or beyond the data boundaries. While the current study focuses on the CS, future work could extend to evaluating the long-term durability of GRAC under various environmental conditions, including exposure to chloride penetration, carbonation depth, sulfate attack, and freeze-thaw cycles, by employing ML techniques. This would enhance understanding of GRAC's lifecycle performance and its suitability across different climatic zones. Another potential area of study is the optimization of the GRAC mix design, with a focus on cost-effectiveness and sustainability. Additionally, it is suggested that future studies model the CS of GRAC incorporating other supplementary cementitious materials, such as silica fume and rice husk ash, as well as exploring various types of alkaline activator solutions using ML algorithms.

CRediT authorship contribution statement

Emad Golafshani: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data curation, Writing – original draft, Visualization, Project administration. **Nima Khodadadi:** Investigation, Methodology, Validation, Writing – review & editing. **Tuan Ngo:** Investigation, Supervision, Writing – review & editing. **Antonio Nanni:** Investigation, Writing – review & editing. **Ali Behnood:** Investigation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

References

- [1] Liang W, Yin W, Zhong Y, Tao Q, Li K, Zhu Z, et al. Mixed artificial intelligence models for compressive strength prediction and analysis of fly ash concrete. *Adv Eng Software* 2023;185. <https://doi.org/10.1016/j.advengsoft.2023.103532>.

- [2] Liang G, Yao W, Wei Y. A green ultra-high performance geopolymer concrete containing recycled fine aggregate: mechanical properties, freeze-thaw resistance and microstructure. *Sci Total Environ* 2023;895:165090. <https://doi.org/10.1016/j.scitotenv.2023.165090>.
- [3] Zhang P, Sun X, Wang F, Wang J. Mechanical properties and durability of geopolymer recycled aggregate concrete: a review. *Polymers (Basel)* 2023;15. <https://doi.org/10.3390/polym15030615>.
- [4] Asghar R, Khan MA, Alyousef R, Javed MF, Ali M. Promoting the green Construction: scientometric review on the mechanical and structural performance of geopolymer concrete. *Constr Build Mater* 2023;368. <https://doi.org/10.1016/j.conbuildmat.2023.130502>.
- [5] Qaidi SMA, Sulaiman Atrushi D, Mohammed AS, Unis Ahmed H, Faraj RH, Emad W, et al. Ultra-high-performance geopolymer concrete: a review. *Constr Build Mater* 2022;346. <https://doi.org/10.1016/j.conbuildmat.2022.128495>.
- [6] Muthu M, Sadowski L. Performance of permeable concrete mixes based on cement and geopolymer in aggressive aqueous environments. *J Build Eng* 2023;76. <https://doi.org/10.1016/j.jobe.2023.107143>.
- [7] Shahmansouri AA, Nematzadeh M, Behnood A. Mechanical properties of GGBFS-based geopolymer concrete incorporating natural zeolite and silica fume with an optimum design using response surface method. *J Build Eng* 2021;36. <https://doi.org/10.1016/j.jobe.2020.102138>.
- [8] Wang J, Xie J, Wang C, Zhao J, Liu F, Fang C. Study on the optimum initial curing condition for fly ash and GGBS based geopolymer recycled aggregate concrete. *Constr Build Mater* 2020;247. <https://doi.org/10.1016/j.conbuildmat.2020.118540>.
- [9] Xie J, Wang J, Rao R, Wang C, Fang C. Effects of combined usage of GGBS and fly ash on workability and mechanical properties of alkali activated geopolymer concrete with recycled aggregate. *Compos B Eng* 2019;164:179–90. <https://doi.org/10.1016/j.compositesb.2018.11.067>.
- [10] Zheng Z, Deng P. Mechanical and fracture properties of slag/steel slag-based geopolymer fully recycled aggregate concrete. *Constr Build Mater* 2024;413. <https://doi.org/10.1016/j.conbuildmat.2023.134533>.
- [11] Liu H, Liu X, Wang X, Zhu P, Yang L, Yan X. The impact of original aggregate and attached mortar types of recycled aggregates on the sulfuric acid resistance of geopolymer recycled concrete. *J Build Eng* 2024;82. <https://doi.org/10.1016/j.jobe.2023.108273>.
- [12] Manhanpally N, Nagarajan P, Saha S, Yunus Raja I. Mechanical and durability characteristics of GGBS-Dolomite geopolymer concrete using recycled coarse aggregates. *Mater Today Proc* 2023. <https://doi.org/10.1016/j.matpr.2023.04.217>.
- [13] Yu M, Wang T, Chi Y, Li D, Yuan Li L, Shi F. Residual mechanical properties of GGBS-FA-SF blended geopolymer concrete after exposed to elevated temperatures. *Constr Build Mater* 2024;411. <https://doi.org/10.1016/j.conbuildmat.2023.134378>.
- [14] Xu F, Li X, Xiong Q, Li Y, Zhu J, Yang F, et al. Influence of aggregate reinforcement treatment on the performance of geopolymer recycled aggregate permeable concrete: from experimental studies to PFC 3D simulations. *Constr Build Mater* 2022;354. <https://doi.org/10.1016/j.conbuildmat.2022.129222>.
- [15] Singh RP, Vanapalli KR, Jadda K, Mohanty B. Durability assessment of fly ash, GGBS, and silica fume based geopolymer concrete with recycled aggregates against acid and sulfate attack. *J Build Eng* 2024;82:108354. <https://doi.org/10.1016/j.jobe.2023.108354>.
- [16] Mesgari S, Akbarnezhad A, Xiao JZ. Recycled geopolymer aggregates as coarse aggregates for Portland cement concrete and geopolymer concrete: effects on mechanical properties. *Constr Build Mater* 2020;236. <https://doi.org/10.1016/j.conbuildmat.2019.117571>.
- [17] Hasnaoui A, Ghorbel E, Wardeh G. Performance of metakaolin/slag-based geopolymer concrete made with recycled fine and coarse aggregates. *J Build Eng* 2021;42. <https://doi.org/10.1016/j.jobe.2021.102813>.
- [18] Arulrajah A, Mohammadinia A, Phummiphan I, Horpibulsuk S, Samingthong W. Stabilization of recycled demolition aggregates by geopolymers comprising calcium carbide residue, fly ash and slag precursors. *Constr Build Mater* 2016;114: 864–73. <https://doi.org/10.1016/j.conbuildmat.2016.03.150>.
- [19] Behnood A, Golafshani EM. Artificial intelligence to model the performance of concrete mixtures and elements: a review. *Arch Comput Meth Eng* 2022;29: 1941–64. <https://doi.org/10.1007/s11831-021-09644-0>.
- [20] Li Y, Shen J, Lin H, Li Y. Optimization design for alkali-activated slag-fly ash geopolymer concrete based on artificial intelligence considering compressive strength, cost, and carbon emission. *J Build Eng* 2023;75. <https://doi.org/10.1016/j.jobe.2023.106929>.
- [21] Kina C, Tanyildizi H, Turk K. Forecasting the compressive strength of GGBFS-based geopolymer concrete via ensemble predictive models. *Constr Build Mater* 2023; 405:133299. <https://doi.org/10.1016/j.conbuildmat.2023.133299>.
- [22] Shamim Ansari S, Muhammad Ibrahim S, Danish Hasan S. Conventional and ensemble machine learning models to predict the compressive strength of fly ash based geopolymer concrete. *Mater Today Proc* 2023. <https://doi.org/10.1016/j.matpr.2023.04.393>.
- [23] Parhi SK, Patro SK. Prediction of compressive strength of geopolymer concrete using a hybrid ensemble of grey wolf optimized machine learning estimators. *J Build Eng* 2023;71. <https://doi.org/10.1016/j.jobe.2023.106521>.
- [24] Nazar S, Yang J, Amin MN, Khan K, Ashraf M, Aslam F, et al. Machine learning interpretable-prediction models to evaluate the slump and strength of fly ash-based geopolymer. *J Mater Res Technol* 2023;24:100–24. <https://doi.org/10.1016/j.jmrt.2023.02.180>.
- [25] Nofalah MH, Ghadir P, Hasanazadehshooili H, Aminpour M, Javadi AA, Nazem M. Effects of binder proportion and curing condition on the mechanical characteristics of volcanic ash- and slag-based geopolymer mortars; machine learning integrated experimental study. *Constr Build Mater* 2023;395:132330. <https://doi.org/10.1016/j.conbuildmat.2023.132330>.
- [26] Dong W, Huang Y, Cui A, Ma G. Mix design optimization for fly ash-based geopolymer with mechanical, environmental, and economic objectives using soft computing technology. *J Build Eng* 2023;72. <https://doi.org/10.1016/j.jobe.2023.106577>.
- [27] Huang Y, Huo Z, Ma G, Zhang L, Wang F, Zhang J. Multi-objective optimization of fly ash-slag based geopolymer considering strength, cost and CO₂ emission: a new framework based on tree-based ensemble models and NSGA-II. *J Build Eng* 2023; 68. <https://doi.org/10.1016/j.jobe.2023.106070>.
- [28] Oyebeisi S, Alomayri T. Artificial intelligence-based prediction of strengths of slag-ash-based geopolymer concrete using deep neural networks. *Constr Build Mater* 2023;400. <https://doi.org/10.1016/j.conbuildmat.2023.132606>.
- [29] Shen J, Li Y, Lin H, Li H, Lv J, Feng S, et al. Prediction of compressive strength of alkali-activated construction demolition waste geopolymers using ensemble machine learning. *Constr Build Mater* 2022;360. <https://doi.org/10.1016/j.conbuildmat.2022.129600>.
- [30] Kuang F, Long Z, Kuang D, Liu X, Guo R. Application of back propagation neural network to the modeling of slump and compressive strength of composite geopolymers. *Comput Mater Sci* 2022;206. <https://doi.org/10.1016/j.commatsci.2022.111241>.
- [31] Alakara EH, Nacar S, Sevim O, Korkmaz S, Demir I. Determination of compressive strength of perlite-containing slag-based geopolymers and its prediction using artificial neural network and regression-based methods. *Constr Build Mater* 2022; 359. <https://doi.org/10.1016/j.conbuildmat.2022.129518>.
- [32] Shobeiri V, Bennett B, Xie T, Visintin P. A generic framework for augmented concrete mix design: optimisation of geopolymer concrete considering environmental, financial and mechanical properties. *J Clean Prod* 2022;369. <https://doi.org/10.1016/j.jclepro.2022.133382>.
- [33] Huo W, Zhu Z, Sun H, Ma B, Yang L. Development of machine learning models for the prediction of the compressive strength of calcium-based geopolymers. *J Clean Prod* 2022;380. <https://doi.org/10.1016/j.jclepro.2022.135159>.
- [34] Rehman F, Khokhar SA, Khushnood RA. ANN based predictive mimicker for mechanical and rheological properties of eco-friendly geopolymer concrete. *Case Stud Construct Mater* 2022;17. <https://doi.org/10.1016/j.cscm.2022.e01536>.
- [35] Peng Y, Unluer C. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. *Constr Build Mater* 2022;316. <https://doi.org/10.1016/j.conbuildmat.2021.125785>.
- [36] Ahmad A, Ahmad W, Aslam F, Joyklad P. Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. *Case Stud Construct Mater* 2022;16. <https://doi.org/10.1016/j.cscm.2021.e00840>.
- [37] Emarah DA. Compressive strength analysis of fly ash-based geopolymer concrete using machine learning approaches. *Results Mater* 2022;16. <https://doi.org/10.1016/j.rinma.2022.100347>.
- [38] Ahmad M, Rashid K, Tariq Z, Ju M. Utilization of a novel artificial intelligence technique (ANFIS) to predict the compressive strength of fly ash-based geopolymer. *Constr Build Mater* 2021;301. <https://doi.org/10.1016/j.conbuildmat.2021.124251>.
- [39] Toufigh V, Jafari A. Developing a comprehensive prediction model for compressive strength of fly ash-based geopolymer concrete (FAGC). *Constr Build Mater* 2021; 277. <https://doi.org/10.1016/j.conbuildmat.2021.122241>.
- [40] Chu HH, Khan MA, Javed M, Zafar A, Ijaz Khan M, Alabduljabbar H, et al. Sustainable use of fly-ash: use of gene-expression programming (GEP) and multi-expression programming (MEP) for forecasting the compressive strength geopolymer concrete. *Ain Shams Engineering Journal* 2021;12:3603–17. <https://doi.org/10.1016/j.asej.2021.03.018>.
- [41] Nguyen KT, Nguyen QD, Le TA, Shin J, Lee K. Analyzing the compressive strength of green fly ash based geopolymer concrete using experiment and machine learning approaches. *Constr Build Mater* 2020;247. <https://doi.org/10.1016/j.conbuildmat.2020.118581>.
- [42] Bagheri A, Nazari A, Sanjayan J. The use of machine learning in boron-based geopolymers: function approximation of compressive strength by ANN and GP. *Measurement (Lond)* 2019;141:241–9. <https://doi.org/10.1016/j.measurement.2019.03.001>.
- [43] Xu L, Veeramachaneni K. Synthesizing Tabular Data using Generative Adversarial Networks 2018.
- [44] Gopalakrishna B, Dinakar P. Mix design development of fly ash-GGBS based recycled aggregate geopolymer concrete. *J Build Eng* 2023;63. <https://doi.org/10.1016/j.jobe.2022.105551>.
- [45] Singh RP, Vanapalli KR, Cheela VRS, Peddiredy SR, Sharma HB, Mohanty B. Fly ash, GGBS, and silica fume based geopolymer concrete with recycled aggregates: properties and environmental impacts. *Constr Build Mater* 2023;378. <https://doi.org/10.1016/j.conbuildmat.2023.131168>.
- [46] Vinay Kumar V, Bhikshma V, Vijaya Prasad B. Study on fresh and mechanical properties for different grades of geopolymer concrete with recycled coarse aggregate. *Mater Today Proc* 2022;60:708–14. <https://doi.org/10.1016/j.matpr.2022.02.326>.
- [47] Moulya HV, Chandrashekhara A. Experimental investigation of effect of recycled coarse aggregate properties on the mechanical and durability characteristics of geopolymer concrete. *Mater Today Proc* 2022;59:1700–7. <https://doi.org/10.1016/j.matpr.2022.03.403>.
- [48] Kumar G, Mishra SS. Effect of recycled concrete aggregate on mechanical, physical and durability properties of GGBS-fly ash-based geopolymer concrete. *Innov Infrastruct Solut* 2022;7. <https://doi.org/10.1007/s41062-022-00832-w>.

- [49] Pawluczuk E, Kalinowska-Wichrowska K, Jiménez JR, Fernández-Rodríguez JM, Suescum-Morales D. Geopolymer concrete with treated recycled aggregates: macro and microstructural behavior. *J Build Eng* 2021;44. <https://doi.org/10.1016/j.jobbe.2021.103317>.
- [50] Saloni Parveen, Lim YY, Pham TM. Effective utilisation of ultrafine slag to improve mechanical and durability properties of recycled aggregates geopolymer concrete. *Clean Eng Technol* 2021;5. <https://doi.org/10.1016/j.clet.2021.100330>.
- [51] Waqas RM, Butt F, Danish A, Alqurashi M, Mosaberpanah MA, Masood B, et al. Influence of bentonite on mechanical and durability properties of high-calcium fly ash geopolymer concrete with natural and recycled aggregates. *Materials (Basel)* 2021;14. <https://doi.org/10.3390/ma14247790>.
- [52] Le HB, Bui QB, Tang L. Geopolymer recycled aggregate concrete: from experiments to empirical models. *Materials (Basel)* 2021;14:1–22. <https://doi.org/10.3390/ma14051180>.
- [53] Hu Y, Tang Z, Li W, Li Y, Tam VWY. Physical-mechanical properties of fly ash/GBFS geopolymer composites with recycled aggregates. *Constr Build Mater* 2019; 226:139–51. <https://doi.org/10.1016/j.conbuildmat.2019.07.211>.
- [54] Krishnan T, Purushothaman R. Optimization and influence of parameter affecting the compressive strength of geopolymer concrete containing recycled concrete aggregate: using full factorial design approach. *IOP Conf Ser Earth Environ Sci*, 80. Institute of Physics Publishing; 2017. <https://doi.org/10.1088/1755-1315/80/1/012013>.
- [55] Nuaklong P, Sata V, Chindaprasit P. Influence of recycled aggregate on fly ash geopolymer concrete properties. *J Clean Prod* 2016;112:2300–7. <https://doi.org/10.1016/j.jclepro.2015.10.109>.
- [56] Shaikh FUA. Mechanical and durability properties of fly ash geopolymer concrete containing recycled coarse aggregates. *Int J Sustainable Built Environ* 2016;5: 277–87. <https://doi.org/10.1016/j.ijsbe.2016.05.009>.
- [57] Shi XS, Wang QY, Zhao XL, Frank C. Discussion on properties and microstructure of geopolymer concrete containing fly ash and recycled aggregate. *Adv Mat Res* 2012; 450-451:1577–83. <https://doi.org/10.4028/www.scientific.net/AMR.450-451.1577>.
- [58] Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 2013;36:27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- [59] Lehmann EL, Joseph PR, George C. Testing statistical hypotheses, 3. New York: Springer; 1986.
- [60] Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1:81–106. <https://doi.org/10.1007/bf00116251>.
- [61] Breiman L. Random forests. *Mach Learn* 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
- [62] Bühlmann P, Yu B. Analyzing bagging. *Ann Stat* 2002;30:927–61.
- [63] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn* 2006;63: 3–42. <https://doi.org/10.1007/s10994-006-6226-1>.
- [64] Freund Y, Schapire Robert E. A decision-theoretic generalization of on-line learning and an application to boosting. In: *Proceedings of the European conference on computational learning theory*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1995. p. 23–37.
- [65] Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002;38: 367–78.
- [66] Chen T, He T. xgboost: eXtreme Gradient Boosting. 2023.
- [67] Tianqi C, Carlos G. XGBoost: a Scalable Tree Boosting System. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*; 2016. p. 785–94.
- [68] Dorogush A.V., Ershov V., Gulin A. CatBoost: gradient boosting with categorical features support 2018.
- [69] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Adv Neural Inf Process Syst*. 30. New York: Curran Associates; 2017. p. 3146–54.
- [70] Anguita D, Ghelardoni L, Ghio A, Oneto L, Ridella S. The “K” in K-fold Cross Validation. In: *ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*; 2012.
- [71] Pelikan M, Goldberg DE, Cantú-Paz E. BOA: the Bayesian optimization algorithm. In: *Proceedings of the genetic and evolutionary computation conference GECCO-99*; 1999.
- [72] Mohammadi Golafshani E, Kim T, Behnood A, Ngo T, Kashani A. Sustainable mix design of recycled aggregate concrete using artificial intelligence. *J Clean Prod* 2024;442:140994. <https://doi.org/10.1016/j.jclepro.2024.140994>.
- [73] Mohammadi Golafshani E, Kashani A, Kim T, Arashpour M. Concrete chloride diffusion modelling using marine creatures-based metaheuristic artificial intelligence. *J Clean Prod* 2022;374. <https://doi.org/10.1016/j.jclepro.2022.134021>.
- [74] Lundberg SM, Lee Su-In. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;30.