Challenges with hard-to-learn data in developing machine learning models for predicting the strength of multi-recycled aggregate concrete

Jeonghyun Kim[1*]

[1] *Faculty of Civil Engineering, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*

Corresponding author: Jeonghyun Kim (jeonghyun.kim@pwr.edu.pl)

Abstract

Research on multi-recycled aggregate concrete (MRAC), which involves reusing recycled concrete, has been actively pursued to promote sustainable practices. However, studying the properties of MRAC often requires significant time and resources. Machine learning (ML)-based predictive methods offer a promising solution to overcome these challenges. In this study, ML models were developed and evaluated to predict the compressive strength of MRAC using a dataset of 197 samples, 8 input features, grid search, cross-validation, and 9 algorithms. The results demonstrated that ML models could achieve high accuracy ($R^2 > 0.9$) even without the application of advanced techniques. However, certain data points consistently exhibited high error rates across multiple models, and the potential causes of poor performance associated with these data points were investigated. Additionally, a post-analysis of the evaluated models was conducted using Shapley Additive Explanations to assess the effect of key features, and recommendations were provided for improving MRAC properties for future research.

Keywords: artificial intelligence, machine learning, hyperparameter, multi-recycled aggregate concrete, sustainable construction materials, prediction error

1. Introduction

Concrete is the most widely used construction material worldwide, valued for its various advantages, such as durability, strength, and cost-effectiveness. However, the key properties of concrete can typically be assessed after sufficient curing, which requires a long curing time for evaluation. This prolonged timeline poses challenges in experimental workflow, emphasizing the need for predictive methods to estimate concrete properties in advance. Yet, due to the nonlinear and complex behavior of concrete, traditional statistical modeling techniques often struggle to provide accurate predictions [1,2]. This limitation becomes more pronounced when new types of materials are introduced in concrete.

Over the past few decades, aligned with global sustainability goals, concrete has been used as a recycled aggregate for new concrete. The recycled aggregates, composed of original aggregates and old mortar, possess distinct physical and chemical characteristics compared to natural aggregates [3,4]. Such differences complicate the behavior of concrete made with recycled aggregates, making traditional statistical models originally designed for natural aggregates often unsuitable for recycled aggregate concrete [5,6]. Machine learning (ML)-based models, by contrast, have shown promise in overcoming these challenges. ML models can capture the intricate relationships and diverse characteristics of concrete mixtures, offering a robust alternative for predicting the properties of recycled aggregate concrete [7,8].

An area of growing research interest is multi-recycled aggregate concrete (MRAC), which involves the successive recycling of concrete materials. Repeated recycling alters the physical and chemical properties of recycled aggregates, making it even more challenging to predict the strength and durability of MRAC compared to conventional recycled aggregate concrete [9–11]. These property changes highlight the limitations of experimental methods, which require substantial costs and time to model the properties of MRAC. The challenges become more pronounced as the number of recycling cycles increases. Data-driven predictive technologies, such as ML, can be considered promising solutions to overcome the limitations of experimental-based modeling. Despite their potential, there have been limited attempts to predict MRAC properties using ML techniques. This study aims to fill this gap by exploring the potential for developing ML models to predict the compressive strength of MRAC, providing valuable insights through model evaluation and post-analysis to effectively predict the performance of MRAC.

2. Research objectives and significance

This study aims not only to develop ML models for predicting the compressive strength of MRAC

but also to deliver actionable insights through comprehensive post-analysis. Specifically, the study aims to achieve the following objectives:

(1) ML models are developed to predict the compressive strength of MRAC based on its mix composition. A key focus is on incorporating the evolving characteristics of recycled aggregates, influenced by repeated recycling, as input features. This approach is designed to enhance predictive accuracy by accounting for the unique behaviors of multi-recycled aggregates.

(2) The performance of the developed ML models is assessed using metrics such as coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). An error analysis is conducted to identify data points with significant discrepancies between predicted and actual values. Furthermore, potential causes of these discrepancies are examined to understand the limitations and areas for improvement in the models.

(3) Post-analysis of the developed models is conducted to assess the relative importance and contribution of key input features influencing the compressive strength of MRAC. This analysis not only provides valuable insights for optimizing MRAC mix design but also highlights directions for future research aimed at improving the strength.

3. Research framework

The workflow of this study is organized into three sequential stages, each designed to systematically address the research objectives:

(1) Database construction: The first stage focuses on constructing a database to support the development of a compressive strength prediction model for MRAC. This involves collecting data from relevant literature, analyzing the statistical characteristics of the data, and executing preprocessing steps tailored for ML. Preprocessing includes splitting the dataset into training and testing subsets, as well as scaling the data to ensure compatibility with ML algorithms. Comprehensive details of this stage are presented in Section 4.

(2) Machine learning model development and evaluation: The second stage is dedicated to building various ML models based on the prepared database and evaluating their performance. Key procedures in this stage include hyperparameter optimization via grid search, cross-validation to mitigate overfitting, and performance-based model assessment. The findings of this stage are

elaborated in Section 5.

(3) Post-Analysis: In the third stage, SHAP analysis is used to assess the contributions of individual input features to the output feature, and examine whether their effects align with existing theoretical knowledge. By providing deeper interpretability of the ML results, this process generates valuable insights that can inform future improvements. Detailed discussions of this stage are presented in Section 6.

Figure 1 outlines the workflow of this study.

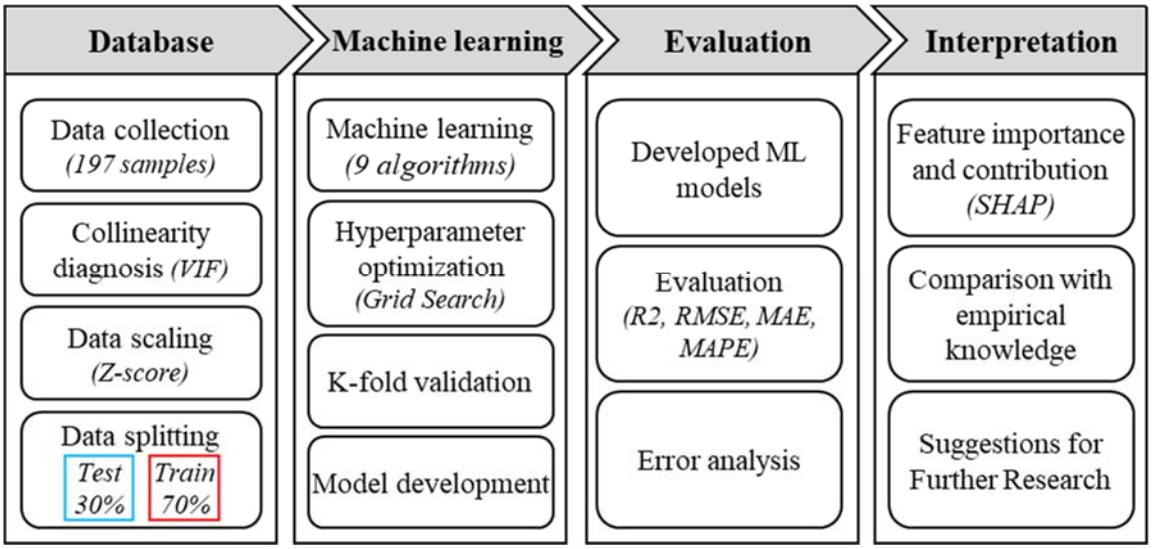

Figure 1. Research framework

4. Data processing

4.1. Data preprocessing

To develop a ML model for predicting the compressive strength of MRAC, a dataset was constructed by collecting data from previously published literature. The selection of input features for model training is a critical aspect in the dataset construction process. Various factors influence concrete strength, including ingredient properties, mix proportions, production conditions (e.g., mixing procedure and time), and environmental factors (e.g., temperature and relative humidity). However, incorporating all these parameters into the dataset is impractical, and many relevant parameters are not comprehensively reported in the literature. Consequently, this study limits the input features to mix proportions, which are both highly influential on concrete strength and widely reported [12]. Additionally, the recycling cycle and the characteristics of recycled materials were considered as input

features in alignment with the objectives of this study. The input variables selected during the initial database construction are as follows: recycling cycle (R-Cycle), water absorption of coarse aggregate (CA-WA), cement (C), slag (S), fly ash (F-ash), silica fume (SF), water (W), national coarse aggregate (NCA), recycled coarse aggregate (RCA), fine aggregate (FA), age, compressive strength of concrete (CS). The CS values obtained from specimens of various sizes were normalized to a reference 150 mm cube to account for the size effect [13].

In regression models, high linear correlations between explanatory variables can lead to unreliable results due to information redundancy [14]. To address this, collinearity diagnostics were conducted on the dataset. Collinearity can be assessed using the Variance Inflation Factor (VIF), where a VIF value greater than 10 indicates that a particular independent variable is highly correlated with other independent variables [15]. As shown in Figure 2(a), the VIF values for NCA and RCA content in the collected data were 68.8 and 54.8, respectively, both significantly exceeding the threshold. To resolve this collinearity issue, these two independent variables were combined into a single variable, coarse aggregate (CA), and the aggregate replacement ratio (Rep) was added. Moreover, considering the "curse of dimensionality," where the required data size grows exponentially with the number of dimensions, reducing dimensions was an effective way to enhance model reliability when data samples were limited [16,17]. Therefore, variables with fewer data points, such as S, F-ash, and SF, were integrated into the water-to-binder ratio (w/b). After these adjustments, the revised dataset demonstrated reduced VIF values for all variables, as shown in Figure 2(b), with no values exceeding the threshold, confirming that high collinearity had been eliminated. Through this preprocessing, the final dataset comprised 197 data points with 8 input features and 1 output feature. A statistical summary of the prepared dataset is provided in Table 1, with detailed individual data points available in the supplementary materials accompanying this paper.

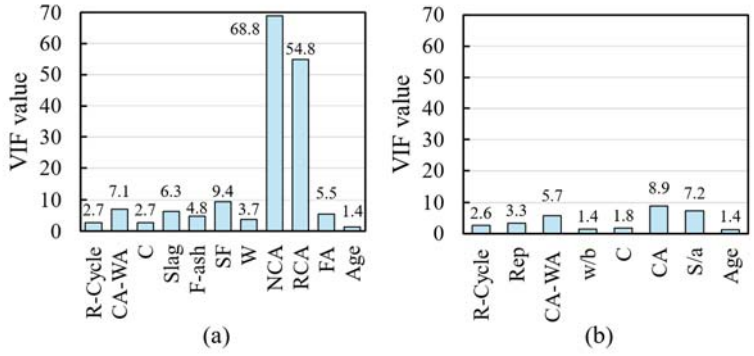

Figure 2. Variance inflation factor of input features before and after dimensionality reduction

Table 1. Descriptive statistics of the dataset used in the study

Feature	Abbreviation	Unit	Min	Max	Median	Mean	Standard deviation
Recycling cycle	R-Cycle	-	0	3	2	1.63	1.07
Replacement ratio	Rep	%	0	100	100	69.5	40.2
Water absorption of coarse aggregate	CA-WA	-	0.32	13.3	5.6	5.55	3.34
Cement	C	kg/m^3	207	527	400	388	74
Water-binder ratio	w/b	-	0.37	0.71	0.5	0.51	0.074
Coarse aggregate	CA	kg/m^3	614	1144	942	931	113
Sand-aggregate ratio	S/a	-	0.344	0.630	0.409	0.430	0.073
Age	-	Day	2	120	28	41.6	33.8
Compressive strength	CS	MPa	10.96	65.53	47.60	45.36	13.12

4.2. Data splitting and scaling

In this study, the dataset was partitioned into training and test sets to avoid overfitting of the predictive models and to evaluate their generalization ability. Specifically, the data was randomly split, with 70% designated for the training set and the remaining 30% for the test set. The training set was used for model development and hyperparameter optimization, while the test set was reserved solely for the final evaluation of model performance.

Certain ML algorithms rely on assigning weights to input features during training. When features differ significantly in scale, those with larger numerical values may disproportionately affect the model, potentially skewing its performance. To mitigate this, it is important to standardize the scale of features. In this study, z-score normalization was applied to transform each feature so that it has a mean of 0 and a standard deviation of 1. This process ensures that all features are scaled uniformly, preventing any single feature from dominating the learning process and enabling the ML algorithm to perform more effectively and equitably.

This is a preprint of a manuscript that has not been peer-reviewed. The final version of this paper was accepted and published in Applied Soft Computing by Elsevier. (https://doi.org/10.1016/j.asoc.2025.113110)

5. Machine learning model development

5.1. Model training and evaluation

In the model development, nine ML algorithms were employed: Linear Regression (LR), Lasso Regression (LSR) [18], Polynomial Regression (PR) [18], Support Vector Regression (SVR) [19], K-Nearest Neighbors (KNN) [20], Artificial Neural Networks (ANN) [21], Random Forest (RF) [22], eXtreme Gradient Boosting (XGB) [23], and Light Gradient Boosting (LGB) [24]. Comprehensive descriptions and theoretical foundations for these algorithms can be found in the referenced literature.

In the model development process, the grid search technique was employed for hyperparameter optimization. Table 2 summarizes the range of hyperparameters explored and the final values selected for each algorithm. Grid search is an exhaustive method that evaluates all possible combinations of hyperparameters within a predefined search space, helping identify the optimal configuration to maximize model performance. After hyperparameter tuning, 5-fold cross-validation was applied to assess model reliability and mitigate the risk of overfitting. In this procedure, the dataset is divided into five equal parts, with each fold serving as the validation set once, while the remaining four folds are used for training. This iterative process ensures that the model does not overfit to the training data, providing a more accurate evaluation of its generalization ability. Once cross-validation was completed, the model with the best performance was selected. This model was subsequently applied to the test set, which was not involved in the cross-validation process, to generate the final predictions.

Table 2. Hyperparameter settings and selected values

Model	Hyperparameter		
	Parameter	Range	Selected
LR	fit_intercept	True, False	True
	n_jobs	-1, 1	-1
	positive	True, False	False
LSR	alphas	0.001, 0.01, 0.1, 1, 10, 100	0.001, 0.01, 0.1, 1, 10, 100
	fit_intercept	True, False	True
	max_iter	100, 300, 500, 1000	300
	tol	0.0001, 0.001, 0.01	0.0001
	selection	cyclic, random	random
PR	poly__degree	1, 2, 3, 4	2
	model__fit_intercept	True, False	False
	model__positive	True, False	False
SVR	C	0.1, 1, 10, 100, 200, 300, 500	200
	epsilon	0.01, 0.1, 0.2, 0.5, 0.7, 0.9	0.5
	kernel	linear, poly, rbf	rbf

KNN	n_neighbors	3, 5, 7, 10, 15, 20	5
	weights	uniform, distance	distance
	metric	euclidean, manhattan, minkowski	manhattan
	p	1, 2,	1
	algorithm	auto, ball_tree, kd_tree, brute	Auto
	leaf_size	10, 20, 30, 40, 50	10
ANN	hidden_layer_sizes	128, 256, 256-128, 128-64	128-64
	activation	tanh, relu	tanh
	solver	adam, sgd, lbfgs	sgd
	alpha	0.0001, 0.001, 0.01	0.01
	learning_rate_init	0.001, 0.0005, 0.0001	0.0001
	max_iter	1000, 1500, 3000, 5000	3000
RF	n_estimators	100, 200, 300, 400	200
	max_depth	1, 3, 5, 7, 8, 9, 10, 11	10
	min_samples_split	2, 3, 4, 5	2
	min_samples_leaf	1, 2, 3, 4	1
	max_features	sqrt, log2	Log
	bootstrap	True, False	False
XGB	n_estimators	100, 200, 300, 400, 500, 700, 800, 1000	500
	max_depth	1, 2, 3, 5	5
	learning_rate	0.01, 0.05, 0.1, 0.2	0.1
	colsample_bytree	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1.0	0.3
	min_child_weight	1, 2, 3, 5	2
LGB	n_estimators	400, 600, 800, 1000	1000
	max_depth	1, 2, 3, 5	3
	learning_rate	0.01, 0.1, 0.2, 0.3	0.2
	num_leaves	2, 4, 6	4
	min_child_samples	1, 3, 5, 10	1

Model performance was evaluated using commonly used performance metrics: R-squared (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) (eqs 1-4). The R^2 value ranges from 0 to 1 and indicates the model's goodness of fit. A value closer to 1 suggests that the model better explains the variation in the dependent variable. RMSE measures the square root of the average squared differences between the predicted and actual values, providing an overall assessment of the model accuracy. A lower RMSE value indicates that the predictions are closer to the actual values. MAE calculates the average absolute differences between predicted and actual values, while MAPE calculates the percentage error between them. Both metrics assess model accuracy, with smaller values indicating fewer errors and better predictive performance.

-.Coefficient of determination (R-squared):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad \text{(eq.1)}$$

-.Root mean squared error (RMSE, MPa):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{(eq.2)}$$

This is a preprint of a manuscript that has not been peer-reviewed. The final version of this paper was accepted and published in Applied Soft Computing by Elsevier. (https://doi.org/10.1016/j.asoc.2025.113110)

-. Mean absolute error (MAE, MPa):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad \text{(eq.3)}$$

-. Mean absolute percentage error (MAPE, %)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \times 100 \qquad \text{(eq.4)}$$

where, y_i and $\hat{y}_i$ represent the actual and predicted values, while $\bar{y}$ denotes the mean of the actual values, repsectively. *n* indicates the total number of data points.

5.2. Post-hoc analysis

As ML systems become more complex, the need to not only improve prediction accuracy but also to understand and interpret prediction outcomes has grown increasingly important [25]. In this study, the SHAP technique, a method for eXplainable Artificial Intelligence (XAI), was employed. SHAP, which is based on the Shapley value from game theory, is used to provide a systematic framework for explaining the predictions of nonlinear models. The contributions of individual features to model predictions were assessed, and valuable insights were provided, along with suggested directions for future research.

6. Results and discussion

Table 3 summarizes the results of predicting CS of MRAC using various algorithms, presenting performance metrics including R^2, MAE, RMSE, and MAPE. On the training set, R^2 values ranged from 0.552 to 1.000, whereas on the test set, they ranged from 0.414 to 0.934, indicating a slight decline in performance when evaluated on unseen data.

The linear regression models, LR and LSR, showed the lowest predictive performance with R^2 values of 0.414 and 0.415, respectively, indicating that their predictive accuracy was below 50%. In contrast, models such as PR, SVR, KNN, ANN, and ensemble models like RF, XGB, and LGB demonstrated comparatively better performance [26,27]. For the training set, these advanced models achieved R^2 values between 0.956 and 1.000, with corresponding MAE, RMSE, and MAPE values ranging from 0.09 to 2.23 MPa, 0.13 to 2.75 MPa, and 0.1% to 6.1%, respectively, indicating robust model training.

On the test set, these metrics ranged from 0.75 to 0.93 for R^2, 2.23 to 5.05 MPa for MAE, 3.32 to 6.39 MPa for RMSE, and 4.9% to 12.4% for MAPE, showing a decline in prediction accuracy and an increase in error compared to the training set. These results are consistent with results typically reported in previous studies on predicting concrete properties [28,29].

Although KNN and RF models achieved 100% accuracy on the training set, their performance dropped on the test set, with accuracy values of 75.4% and 87.1%, respectively, representing decreases of about 25% and 13%. This suggests potential overfitting and a decline in generalization ability. The XGB model demonstrated the best predictive performance, achieving the highest accuracy (R^2 = 0.934) and the lowest error metrics (MAE = 2.23 MPa, RMSE = 3.32 MPa, MAPE = 4.9%). However, it is important to recognize that model performance is influenced by various factors, including the dataset, parameter settings, and data preprocessing. According to the "no free lunch" theorem, no single algorithm can consistently provide the best performance across all types of problems [30,31]. Therefore, a comprehensive analysis of multiple models is necessary. In this study, the focus was on five models (PR, SVR, ANN, XGB, LGB) that achieved R^2 values greater than 0.9 on the test set.

Table 3. Performance results of machine learning across various algorithms

Model	Training				Testing				Runtime (s)
	R2	MAE	RMSE	MAPE	R2	MAE	RMSE	MAPE	
LR	0.553	7.27	8.75	18.9	0.414	8.14	9.86	19.6	4.74
LSR	0.552	7.30	8.76	19.0	0.415	8.15	9.85	19.7	4.95
PR	**0.956**	**2.23**	**2.75**	**6.1**	**0.912**	**2.98**	**3.81**	**7.3**	**5.75**
SVR	**0.984**	**0.93**	**1.65**	**2.7**	**0.896**	**3.21**	**4.16**	**7.9**	**6.94**
KNN	1.000	0.02	0.13	0.1	0.754	5.05	6.39	12.4	6.80
ANN	**0.974**	**1.54**	**2.11**	**3.9**	**0.900**	**3.45**	**4.07**	**8.4**	**1410.93**
RF	1.000	0.12	0.24	0.3	0.871	3.63	4.63	9.0	81.42
XGB	**0.999**	**0.32**	**0.50**	**0.8**	**0.934**	**2.23**	**3.32**	**4.9**	**157.67**
LGB	**0.999**	**0.09**	**0.32**	**0.3**	**0.903**	**2.69**	**4.01**	**6.2**	**144.28**

6.1. Prediction performance

The scatter plot comparing the actual and predicted values for the five selected models is presented in Figure 3. Since only models with an R^2 greater than 0.9 were chosen, all the models exhibit a strong correlation between the experimental and predicted values, with the points closely aligned along the

equality line (x=y). Most predicted values fall within a ±20% error margin, with the XGB model showing a tighter clustering within a ±10% error range.

An interesting observation is that, across different ML models, data points that fall outside the ±20% error range, where training performance is poor or test predictions are inaccurate, are located in similar zones. Specifically, in all five models, certain training and test data points were consistently overpredicted in regions where the actual compressive strength was below approximately 15 MPa. Additionally, similar trends were observed near actual values around 51 MPa in the PR, SVR, and XGB models, and around 37 MPa in the SVR, XGB, and LGB models.

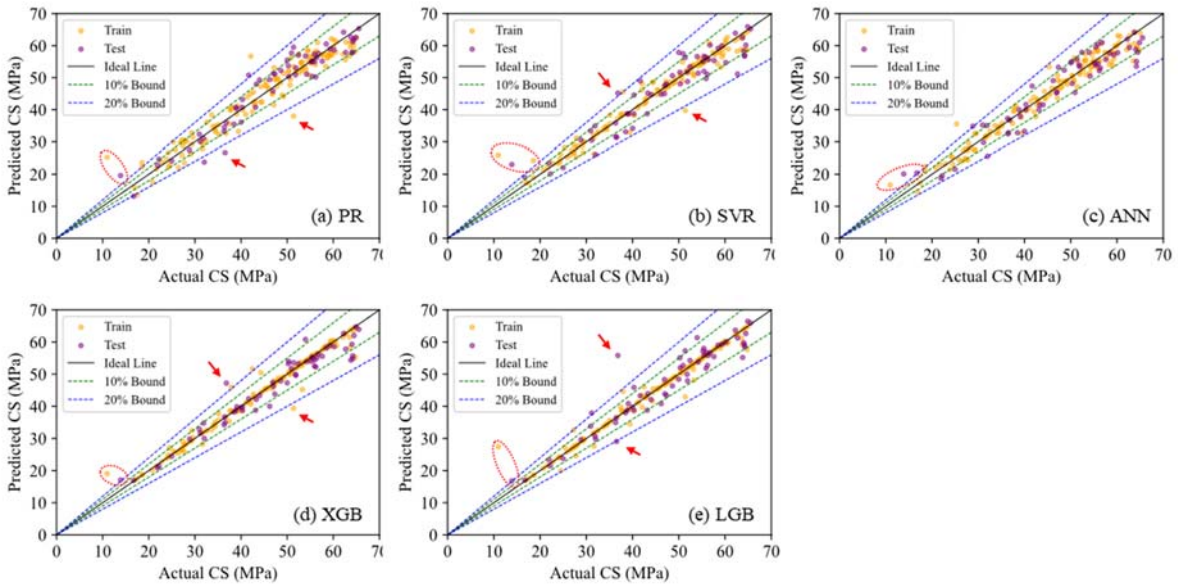

Figure 3. Scatter plot of predicted versus actual compressive strength: (a) PR; (b) SVR; (c) ANN; (d) XGB; (e) LGB

To further examine the data points with large errors, the actual values, predicted values, and residuals (i.e., the difference between the actual and predicted values) for each individual data point are shown in Figure 4. Through the error peaks, it can be seen that, except for the ANN model, the errors for data points 53 and 119 are significantly higher compared to the others across the four models. In addition, specific data points, such as 3, 15, and 100, were found to have poor learning performance depending on the ML model. For the test set, relatively high errors were observed for data points such as 152, 156, 159, and 162. These results indicate that certain data points in the dataset pose consistent challenges for accurate modeling and prediction, regardless of the algorithm employed.

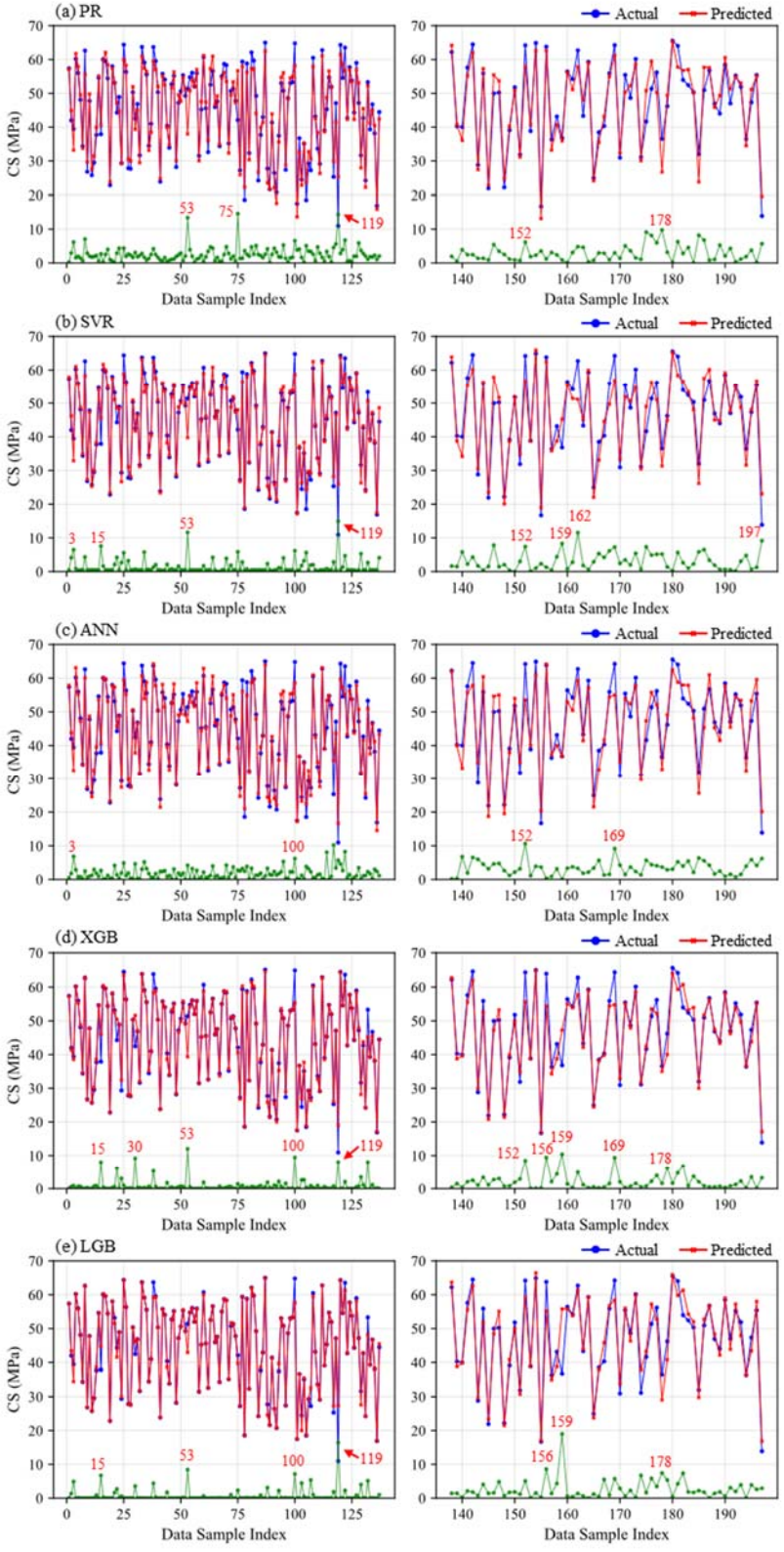

Figure 4. Actual values, predicted values, and residuals for each data point across various machine learning models: (a) PR; (b) SVR; (c) ANN; (d) XGB; (e) LGB

The output feature of the dataset used in this study, CS of MRAC, ranges from 10.96 MPa to 65.53 MPa. Expressing residuals in MPa helps quantify the absolute size of errors, but the relative impact of these errors depends on the actual value. For instance, an error of 3 MPa accounts for approximately 4.6% of the maximum compressive strength (65.53 MPa), whereas it corresponds to 27.4% when compared to the minimum compressive strength (10.96 MPa), indicating a significant error rate. Thus, to more accurately evaluate model performance, it is essential to express the difference between the actual and predicted values as a ratio.

Table 4 lists the top five data points with the highest error rates in both the training and test sets. As shown in Figure 4, data point 119 exhibited the highest error rate across all ML models, and the predicted value was overpredicted from 52.3% to a maximum of 150.0% compared to the actual value. In the test set, data point 197 was found to have the highest error rate in the PR, SVR, and ANN models and ranked as the second- and third-highest error rate in the XGB and LGB models, respectively. The error rates for this data point ranged from 21.1% (in the LGB model) to 66.0% (in the SVR model).

However, in the MPa-based error graph, the error peak for data point 197 is not particularly noticeable, which may cause this data point to be overlooked in error analysis based solely on absolute error values. The actual CS for data point 197 is 13.9 MPa, which is among the lowest in the dataset. As a result, despite significant fluctuations in predicted values, the absolute error remained relatively small, ranging from 2.9 MPa to 6.2 MPa (with the exception of the 9.1 MPa error observed in the SVR model in the MPa-based error graph). Therefore, interpreting the error only in MPa units may make it difficult to properly identify data points with high error rates. To address this, incorporating error rate analysis would provide a more refined approach to evaluating model performance.

Table 4. Data points with the highest error rates in the training and test sets

Data	Rank	PR	SVR	ANN	XGB	LGB
Train	**1**	**130.7% (119)**	**136.7% (119)**	**52.3% (119)**	**73.7% (119)**	**150.0% (119)**
	2	34.7% (75)	**30.7% (105)**	40.4% (117)	**-23.5% (53)**	19.8% (107)
	3	**27.7% (105)**	**-22.7% (53)**	**20.9% (105)**	21.7% (30)	-18.3% (103)
	4	**-26.1% (53)**	**20.1% (15)**	17.9% (114)	**21.2% (15)**	**17.8% (15)**
	5	-22.1% (101)	-16.9% (129)	-17.5% (3)	-15.1% (132)	**-16.5% (53)**
Test	**1**	**41.2% (197)**	**66.0% (197)**	**44.9% (197)**	**28.3% (159)**	**51.6% (159)**
	2	-26.8% (178)	**22.7% (159)**	22.1% (155)	**23.2% (197)**	21.7% (174)
	3	**-25.7% (185)**	-18.5% (162)	20.8% (143)	-14.8% (156)	**21.1% (197)**
	4	22.0% (175)	**-18.3% (185)**	**-20.0% (185)**	-14.6% (169)	-20.5% (178)
	5	-21.3% (155)	17.7% (175)	-16.9% (140)	-13.4% (179)	13.8% (167)

- Data points 1-137 correspond to the training set, while data points 138-197 correspond to the test set.

The behavior of data with high error rates, specifically, those reported more than three times in Table 4, including data points 15, 53, 105, and 119 from the training set, and 159, 185, and 197 from the test set, was examined in the context of the original literature. According to general knowledge, when concrete is repeatedly recycled without reinforcement, its mechanical strength and durability tend to be degraded. Each recycling cycle typically leads to an increase in the water absorption of aggregates, which in turn raises porosity and promotes the formation of microcracks. This cycle creates a negative feedback loop, which further weakens the material [9]. Building on this empirical knowledge, Figure 5 illustrates the relationship between the R-Cycle, CA-WA, and CS. The data points with high error rates and their comparison groups from the original source are shown with colored solid lines, while data points that did not have high error rates are represented by black dashed lines for comparison.

Upon analysis, no significant differences were observed between the group of high error rate data and the group with lower error rate data. In both groups, a consistent decrease in CS was observed as R-Cycle and CA-WA rate increased. This suggests that when individual data points are integrated into a dataset, they may not perform as expected. This is because data from different studies often come with varying experimental conditions. For example, variables such as temperature, humidity, measurement methods, concrete ingredient properties, and equipment specifications are often missing in the literature. These differences introduce inconsistencies across the data and may result in non-linear interactions or the influence of hidden variables. Therefore, while relationships between variables might be clear in individual studies, combining multiple datasets can complicate model training.

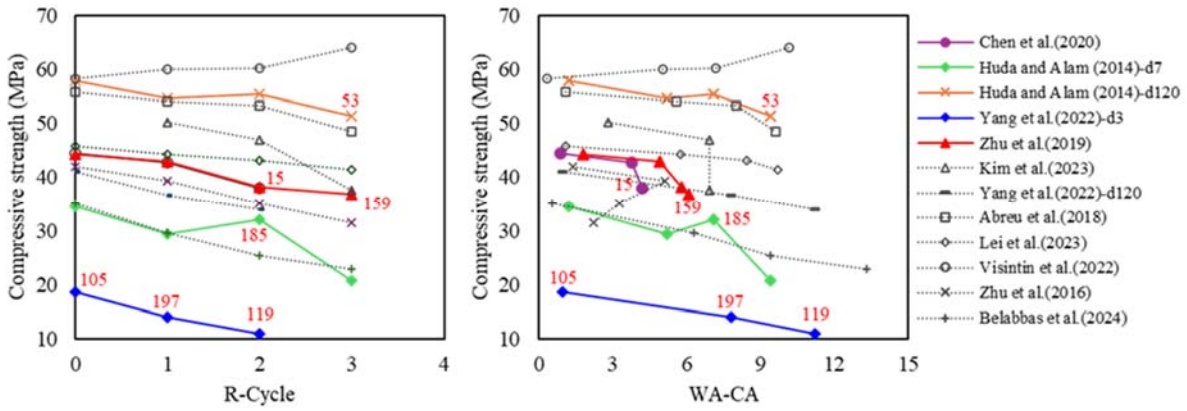

Figure 5. Compressive strength behavior in the original literature for the top 5 data points with the highest error rates and their comparison group (solid line), and for some data points not included in the top 5 (dotted line), with respect to recycling cycles (a) and aggregate absorption (b)

Creating a comprehensive database is one of the greatest hurdles in the development of ML models. As noted by Li et al. [16], testing concrete properties often requires long curing periods, making the collection of extensive datasets both time-consuming and costly. Consequently, datasets for predicting concrete properties are frequently constrained by sample size. This limitation is evident in studies on MRAC, an emerging area of research characterized by especially limited data availability. In this study, to make the most of the available data, studies incorporating supplementary cementitious materials (SCMs), such as slag, fly ash, and silica fume in MRAC, were not excluded from the dataset. Of the 197 total data samples, 14 contained slag, 31 contained fly ash, and 11 contained silica fume, including duplicates.

When the size of the data sample is limited, dimensionality reduction should be considered to maintain the reliability of the training process [16,17]. Accordingly, in this study, instead of using the slag, fly ash, and silica fume as separate input features, they were integrated into the w/b, as mentioned in the previous section. Based on this context, the dataset was backtracked to analyze potential causes of high-error data. Table 5 summarizes the data points with an error rate exceeding 20% for each model, including their index number, actual values, predicted values, error rates, and possible causes. Of the 33 data points with high error, 15 involved the inclusion of SCMs. This may be due to the contrasting effects of SCM types on concrete properties. For example, silica fume enhances early-age strength development [32–34], whereas slag contributes to strength development at later ages [35]. Consequently, merging different types of SCMs into a single variable might introduce noise into the model, adversely affecting learning and prediction performance. Therefore, special caution is needed when integrating distinct SCM types into a single feature. Moreover, the prediction errors were higher when the actual CS was significantly low (approximately 15 MPa). CS values from a study by Yang et al. [36] measured at 3 days of curing exhibited high errors across all models. Statistical analysis of the dataset in Table 1 indicates that the median value of CS is 47.60 MPa, skewed towards the maximum value (65.53 MPa), with the majority of CS data concentrated in the 40–60 MPa range. This uneven distribution likely limits the model capacity to learn and predict low-strength behavior. These findings emphasize the pivotal influence of dataset quality and composition in shaping ML model performance and reliability. They also highlight the potential risks of unintentionally compromising model reliability during data processing by individuals developing ML models.

However, since ML models inherently function as black boxes, pinpointing clear and direct causes of high-error predictions is challenging. While the two causes presented above (integration of SCMs and insufficient low-strength data) likely contributed to the high error rates, definitive conclusions cannot be drawn. Further validation of these potential causes remains an ongoing challenge for future research.

A potential solution could be to analyze SCM-related variables seperately or expand the dataset to include a broader range of variable combinations. However, challenges such as limited sample sizes, high-dimensional complexities, and the potential loss of essential information during data integration remain significant obstacles in the development of ML models.

Table 5. Information and potential causes of data points with an error rate exceeding 20%

	Reference	Data index	Actual CS	Predicted CS	Error	Remark
PR	[37]	178	36.54	26.74	26.8%	X
	[38]	53	51.39	37.97	26.1%	Fly ash 70 kg
	[38]	155	16.70	13.15	21.3%	Fly ash 70 kg
	[38]	175	41.64	50.80	22.0%	Fly ash 70 kg
	[38]	185	32.02	23.80	25.7%	Fly ash 70 kg
	[36]	105	18.54	23.67	27.7%	3 days of age
	[36]	119	10.96	25.29	130.7%	3 days of age
	[36]	197	13.88	19.60	41.2%	3 days of age
	[39]	75	42.1	56.69	34.7%	Slag 46 kg, fly ash 184 kg, silica fume 23 kg
	[40]	101	17.48	13.62	22.1%	Silica fume 40 kg
SVR	[38]	53	51.39	39.74	22.7%	Fly ash 70 kg
	[36]	105	18.54	24.23	30.7%	X
	[36]	119	10.96	25.94	136.7%	X
	[36]	197	13.88	23.05	66.0%	X
	[41]	159	36.81	45.16	22.7%	Slag 40 kg, fly ash 83 kg, silica fume 21 kg
	[42]	15	37.91	45.52	20.1%	Slag 45 kg, fly ash 90 kg, silica fume 23 kg
ANN	[38]	155	16.70	20.40	22.1%	Fly ash 70 kg
	[38]	185	32.02	25.63	20.0%	Fly ash 70 kg
	[36]	105	18.54	22.42	20.9%	3 days of age
	[36]	119	10.96	16.69	52.3%	3 days of age
	[36]	197	13.88	20.12	44.9%	3 days of age
	[40]	117	25.32	35.55	40.4%	X
XGB	[37]	30	42.46	51.65	21.7%	X
	[38]	53	51.39	39.33	23.5%	Fly ash 70 kg
	[36]	119	10.96	19.04	73.7%	3 days of age
	[36]	197	13.88	17.10	23.2%	3 days of age
	[41]	159	36.81	47.23	28.3%	Slag 40 kg, fly ash 83 kg, silica fume 21 kg
	[42]	15	37.91	45.94	21.2%	Slag 45 kg, fly ash 90 kg, silica fume 23 kg
LGB	[37]	178	36.54	29.06	20.5%	X
	[36]	119	10.96	27.41	150%	3 days of age
	[36]	197	13.88	16.81	21.1%	3 days of age
	[36]	174	31.15	37.91	21.7%	7 days of age
	[41]	159	36.81	55.82	51.6%	Slag 40 kg, fly ash 83 kg, silica fume 21 kg

6.2. SHAP analysis

This is a preprint of a manuscript that has not been peer-reviewed. The final version of this paper was accepted and published in Applied Soft Computing by Elsevier. (https://doi.org/10.1016/j.asoc.2025.113110)

6.2.1. Global interpretation

The SHAP summary plot in Figure 6 visually illustrates how input features influence the output feature in the ML models. In the plot, dots with redder colors correspond to higher values of the input feature, while bluer dots represent lower values. The SHAP values represent the contribution of each feature to the output feature.

The PR model identified CA, S/a, and w/b as the most influential features affecting CS, whereas the SVR and ANN models highlighted S/a, CA-WA, and CA as the top three significant features. For the XGB model, age, C, and CA were the primary features, while the LGB model identified age, S/a, and CA as the key contributors. Although the ranking of feature importance varied slightly across the models, the overall influence of the key features on CS was consistent. An increase in age enhanced CS, whereas increases in S/a, CA-WA, and w/b decreased CS. These findings align well with empirical knowledge [43]. While CA is traditionally considered a positive contributor to concrete strength, with higher content improving CS [44,45], this study defines CA as natural, recycled, and multi-recycled aggregates, which can experience quality degradation over multiple recycling cycles. Consequently, the observed reduction in CS with increased CA content is reasonable [46,47]. This result indicates that the ML models effectively captured the characteristics of multi-recycled aggregates. Interestingly, the Rep and R-Cycle were identified as features with low. This indicates that CS is not directly governed by these factors but rather by the degradation of aggregate quality resulting from these factors, which ultimately leads to reduced CS [48].

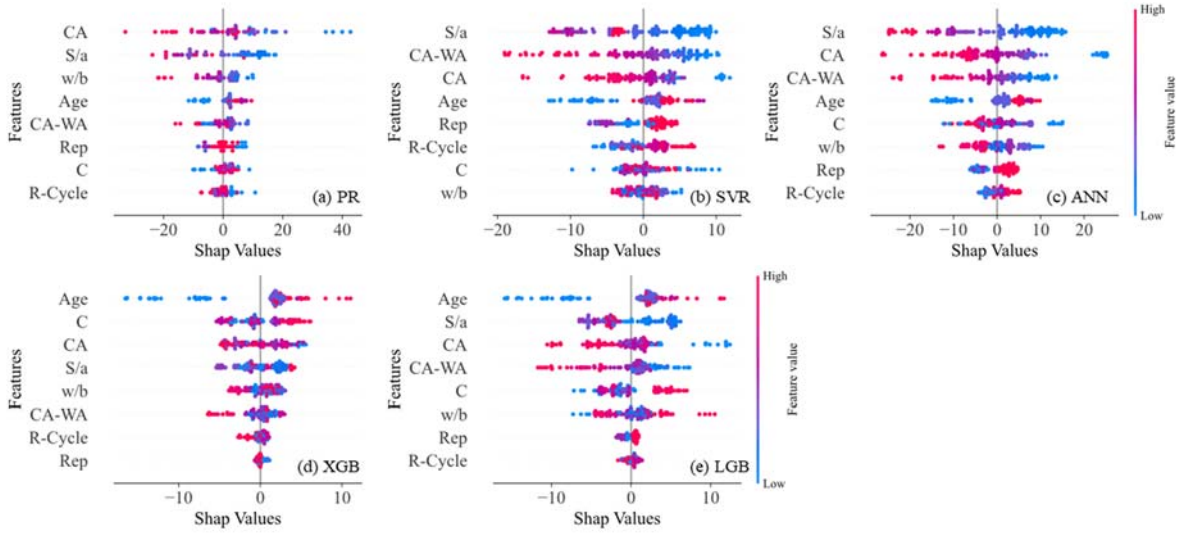

Figure 6. SHAP summary plot of various machine learning models: (a) PR; (b) SVR; (c) ANN; (d) XGB; (e) LGB

6.2.2. Local interpretation

SHAP dependence plots are used to understand the interactions between input features and their effects on the output feature. Figure 7 and Figure 8 display the SHAP dependence plots for the XGB and PR models, which demonstrated excellent accuracy in this study.

In both the XGB and PR models, CS increased with age (Figures 7 (a) and Figure 8 (a)) and cement content (Figures 7 (b) and Figure 8 (b)). For CA content, an increase within the range of 600-900 kg/m³ enhanced CS, but beyond this threshold, it led to a decrease in compressive strength (Figures 7 (c) and Figure 8 (c)). S/a reduced CS when it increased from 0.35 to around 0.50, but once it surpassed 0.5, it started to enhance CS. This behavior could be associated with the decrease in low-quality CA as S/a ratio rises (Figures 7 (d) and Figure 8 (d)). For w/b, the XGB model showed some dispersion, but a downward trend in CS was observed up to w/b of 0.55, after which a sudden increase in CS occurred. In contrast, the PR model exhibited a consistent decrease in CS with increasing w/b, indicating that the PR model better reflects empirical knowledge (Figures 7 (e) and Figure 8 (e)). The increase in CA-WA contributed to a decrease in CS, and as the replacement rate rose, so did CA-WA (Figures 7 (f) and Figure 8 (f)). For R-cycle, no significant change in CS was observed up to the second recycling cycle. However, after the third recycling cycle, CS decreased, with considerable variation (Figures 7 (g) and Figure 8 (g)).

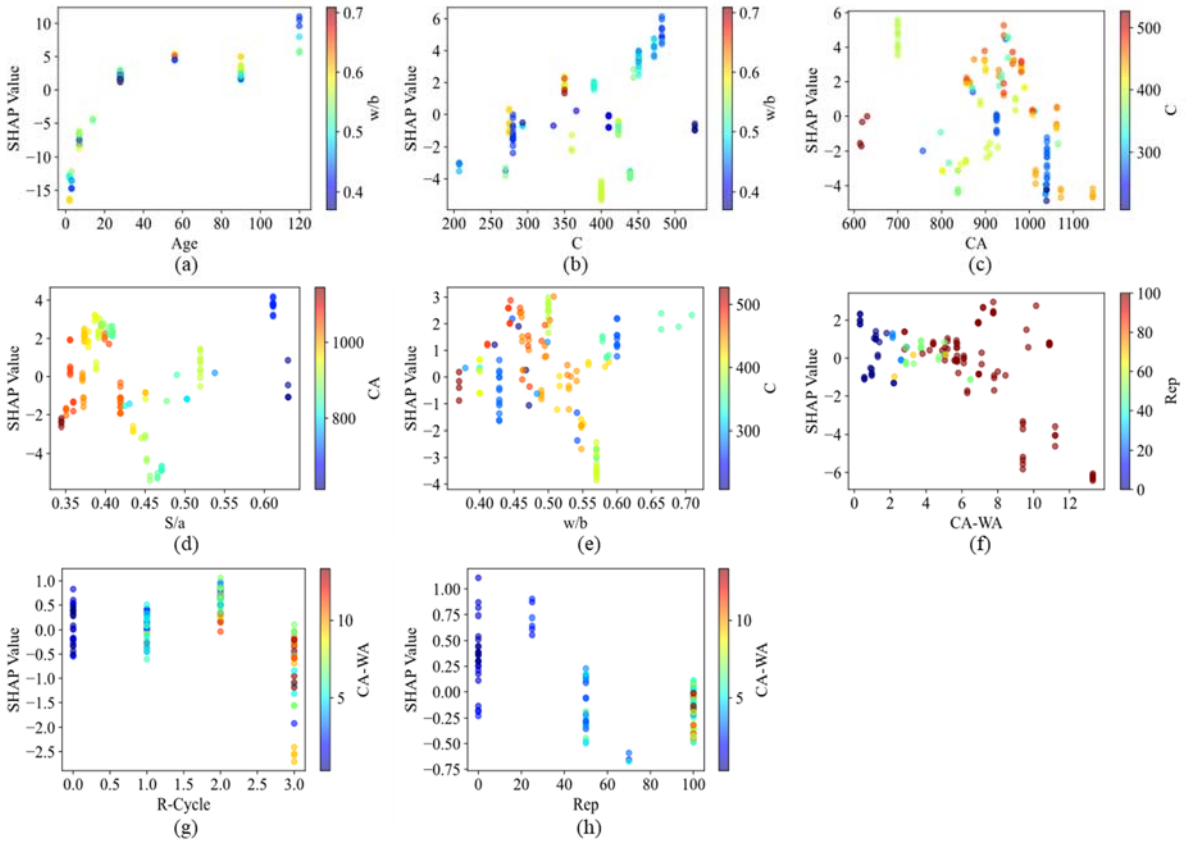

Figure 7. SHAP dependent plot of XGB model for interation of input features on out feature: (a) age; (b) cement; (c) coarse aggregate content; (d) sand-to-aggregate ratio; (e) water-to-binder ratio; (f) water absorption of coarse aggregate; (g) recycling cycle; and (h) replacement ratio.

This is a preprint of a manuscript that has not been peer-reviewed. The final version of this paper was accepted and published in Applied Soft Computing by Elsevier. (https://doi.org/10.1016/j.asoc.2025.113110)

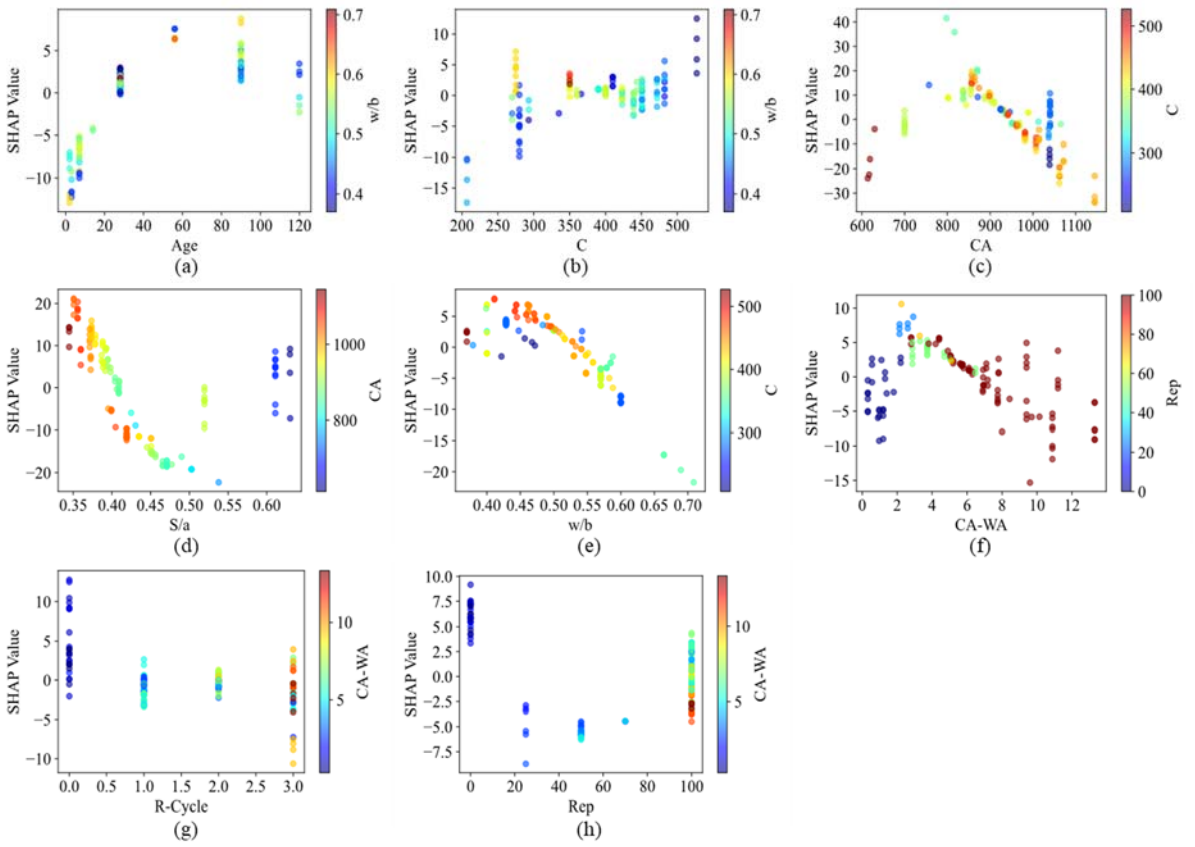

Figure 8. SHAP dependent plot of PR model for interation of input features on out feature: (a) age; (b) cement; (c) coarse aggregate content; (d) sand-to-aggregate ratio; (e) water-to-binder ratio; (f) water absorption of coarse aggregate; (g) recycling cycle; and (h) replacement ratio.

The SHAP dependence plots provide valuable insights for further research. For example, as the number of R-Cycle increases, the CA-WA rises, which in turn leads to a reduction in CS. Therefore, employing various methods such as removing attached mortar or coating aggregates to reduce the CA-WA is expected to contribute to improving the CS, regardless of the number of R-Cycle. This finding is particularly significant given the practical challenges in accurately tracking and managing the number of recycling cycles in the construction industry.

As previously emphasized, the performance of the model is heavily influenced by the quantity of data used during training. A larger volume of training data allows the model to better capture underlying patterns, leading to improved accuracy and reduced prediction variance [49,50]. Furthermore, the dispersion observed in the dependence plots of certain features, which might currently appear scattered or unclear, is expected to become more distinct and consistent as the size of the data samples grows.

7. Conclusions

This study developed and analyzed ML models to predict the strength of MRAC. The following conclusions were drawn:

(1) ML models with high accuracy ($R^2 > 0.9$) could be developed using only basic ML algorithms, without the need for complex modifications or the combination of multiple algorithms.

(2) When SCMs, which influence concrete strength in different ways, are included in the dataset with a small sample size, the ML models may fail to capture these complex interactions, potentially leading to greater discrepancies between predicted and actual values.

(3) The SHAP analysis provided valuable insights based on experimental data. Through this analysis, the effects of recycling cycles on aggregate properties and concrete strength were systematically understood, offering directions for future research to optimize the efficient use of MRAC. This approach can contribute to enhancing experimental efficiency.

(4) This study demonstrated that ML can be a promising solution to overcome the physical time and resource limitations associated with modeling the strength of MRAC for achieving sustainable practices.

8. Limitations

This study demonstrates novelty by developing models to predict the strength of MRAC and discussing the reasons behind underperforming data points during training and prediction. Despite these contributions, a notable limitation is the relatively small dataset, which is due to the emerging nature of the topic. Moreover, in line with the "no free lunch" theorem, it is important to recognize that there is no universal optimization algorithm that outperforms all others across all possible problems. Algorithm performance is problem-specific, emphasizing the necessity of selecting the most suitable algorithm for a given task. From this perspective, the strong performance of the ML models developed in this study is limited to the conditions defined within the scope of this research. Thus, future studies should focus on expanding the dataset and exploring diverse ML environments, including hyperparameter optimization, the application of various algorithms.

Acknowledgment

This research was funded in whole by the National Science Centre, Poland (Grant number 2023/51/B/ST8/00621). For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

References

[1] M.A. DeRousseau, J.R. Kasprzyk, W.V. Srubar, Computational design optimization of concrete mixtures: A review, Cem Concr Res 109 (2018) 42–53. https://doi.org/10.1016/j.cemconres.2018.04.007.

[2] U. Reuter, A. Sultan, D.S. Reischl, A comparative study of machine learning approaches for modeling concrete failure surfaces, Advances in Engineering Software 116 (2018) 67–79. https://doi.org/10.1016/j.advengsoft.2017.11.006.

[3] F. Ceia, J. Raposo, M. Guerra, E. Júlio, J. De Brito, Shear strength of recycled aggregate concrete to natural aggregate concrete interfaces, Constr Build Mater 109 (2016) 139–145. https://doi.org/10.1016/j.conbuildmat.2016.02.002.

[4] A.B. Ajdukiewicz, A.T. Kliszczewicz, Comparative Tests of Beams and Columns Made of Recycled Aggregate Concrete and Natural Aggregate Concrete, Journal of Advanced Concrete Technology 5 (2007). https://doi.org/10.3151/jact.5.259.

[5] R. V. Silva, J. de Brito, R.K. Dhir, Comparative analysis of existing prediction models on the creep behaviour of recycled aggregate concrete, Eng Struct 100 (2015) 31–42. https://doi.org/10.1016/J.ENGSTRUCT.2015.06.004.

[6] B. Li, S. Dai, Y. Zhan, J. Xu, X. Guo, Y. Yang, Y. Chen, Strength criterion of recycled aggregate concrete under triaxial Compression: Model calibration, Constr Build Mater 320 (2022) 126201. https://doi.org/10.1016/j.conbuildmat.2021.126201.

[7] S. Wang, P. Xia, F. Gong, Q. Zeng, K. Chen, Y. Zhao, Multi objective optimization of recycled aggregate concrete based on explainable machine learning, J Clean Prod 445 (2024) 141045. https://doi.org/10.1016/j.jclepro.2024.141045.

[8] M. Fan, Y. Li, J. Shen, K. Jin, J. Shi, Multi-objective optimization design of recycled aggregate concrete mixture proportions based on machine learning and NSGA-II algorithm, Advances in Engineering Software 192 (2024) 103631. https://doi.org/10.1016/j.advengsoft.2024.103631.

[9] J. Kim, Sustainable Construction Exploration: A Review of Multi-Recycling of Concrete Waste, Int J Environ Res 18 (2024) 103. https://doi.org/10.1007/s41742-024-00652-z.

[10] J. Kim, N. Nciri, A. Sicakova, N. Kim, Characteristics of waste concrete powders from multi-recycled coarse aggregate concrete and their effects as cement replacements, Constr Build Mater 398 (2023) 132525. https://doi.org/10.1016/j.conbuildmat.2023.132525.

[11] C. Thomas, J. de Brito, A. Cimentada, J.A. Sainz-Aja, Macro- and micro- properties of multi-recycled aggregate concrete, J Clean Prod 245 (2020). https://doi.org/10.1016/j.jclepro.2019.118843.

[12] A. Behnood, E.M. Golafshani, Machine learning study of the mechanical properties of concretes containing waste foundry sand, Constr Build Mater 243 (2020) 118152. https://doi.org/10.1016/j.conbuildmat.2020.118152.

[13] S.T. Yi, E.I. Yang, J.C. Choi, Effect of specimen sizes, specimen shapes, and placement directions on compressive strength of concrete, Nuclear Engineering and Design 236 (2006) 115–127. https://doi.org/10.1016/J.NUCENGDES.2005.08.004.

[14] J.H. Kim, Multicollinearity and misleading statistical results, Korean J Anesthesiol 72 (2019) 558–569. https://doi.org/10.4097/kja.19087.

[15] C.G. Thompson, R.S. Kim, A.M. Aloe, B.J. Becker, Extracting the Variance Inflation Factor and Other Multicollinearity Diagnostics from Typical Regression Results, Basic Appl Soc Psych 39 (2017) 81–90. https://doi.org/10.1080/01973533.2016.1277529.

[16] Z. Li, J. Yoon, R. Zhang, F. Rajabipour, W. V. Srubar III, I. Dabo, A. Radlińska, Machine learning in concrete science: applications, challenges, and best practices, NPJ Comput Mater 8 (2022) 127. https://doi.org/10.1038/s41524-022-00810-x.

[17] K.-H. Liu, J.-K. Zheng, F. Pacheco-Torgal, X.-Y. Zhao, Innovative modeling framework of chloride resistance of recycled aggregate concrete using ensemble-machine-learning methods, Constr Build Mater 337 (2022) 127613. https://doi.org/10.1016/j.conbuildmat.2022.127613.

[18] G.A.F. Seber, A.J. Lee, Linear Regression Analysis, Wiley, 2003. https://doi.org/10.1002/9780471722199.

[19] C. Cortes, V. Vapnik, Support-vector networks, Mach Learn 20 (1995) 273–297. https://doi.org/10.1007/BF00994018.

[20] N.S. Altman, An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, Am Stat 46 (1992) 175–185. https://doi.org/10.1080/00031305.1992.10475879.

[21] A. Krogh, What are artificial neural networks?, Nat Biotechnol 26 (2008) 195–197. https://doi.org/10.1038/nbt1386.

[22] L. Breiman, Random Forests, Mach Learn 45 (2001) 5–32. https://doi.org/10.1023/A:1010933404324.

[23] 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, Association for Computing Machinery, New York, NY, USA, 2016: pp. 785–794. https://doi.org/10.1145/2939672.2939785.

[24] G. Ke, Q. Meng, T. Finely, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Advances in Neural Information Processing Systems 30 (NIP 2017), 2017. https://www.microsoft.com/en-us/research/publication/lightgbm-a-highly-efficient-gradient-boosting-decision-tree/.

[25] D. Castelvecchi, Can we open the black box of AI?, Nature 538 (2016) 20–23. https://doi.org/10.1038/538020a.

[26] J.S. Chou, C.F. Tsai, A.D. Pham, Y.H. Lu, Machine learning in concrete strength simulations: Multi-nation data analytics, Constr Build Mater 73 (2014) 771–780. https://doi.org/10.1016/J.CONBUILDMAT.2014.09.054.

[27] K. Liu, L. Zhang, W. Wang, G. Zhang, L. Xu, D. Fan, R. Yu, Development of compressive strength prediction platform for concrete materials based on machine learning techniques, Journal of Building Engineering 80 (2023) 107977. https://doi.org/10.1016/j.jobe.2023.107977.

[28] J. de-Prado-Gil, C. Palencia, N. Silva-Monteiro, R. Martínez-García, To predict the compressive strength of self compacting concrete with recycled aggregates utilizing ensemble machine learning models, Case Studies in Construction Materials 16 (2022) e01046. https://doi.org/10.1016/j.cscm.2022.e01046.

[29] H.-V.T. Mai, M.H. Nguyen, H.-B. Ly, Development of machine learning methods to predict the

compressive strength of fiber-reinforced self-compacting concrete and sensitivity analysis, Constr Build Mater 367 (2023) 130339. https://doi.org/10.1016/j.conbuildmat.2023.130339.

[30] J. Kim, D. Lee, A. Ubysz, Comparative analysis of cement grade and cement strength as input features for machine learning-based concrete strength prediction, Case Studies in Construction Materials 21 (2024) e03557. https://doi.org/10.1016/j.cscm.2024.e03557.

[31] Y.C. Ho, D.L. Pepyne, Simple Explanation of the No-Free-Lunch Theorem and Its Implications, J Optim Theory Appl 115 (2002) 549–570. https://doi.org/10.1023/A:1021251113462.

[32] G.A. Rao, Development of strength with age of mortars containing silica fume, Cem Concr Res 31 (2001) 1141–1146. https://doi.org/10.1016/S0008-8846(01)00540-3.

[33] B. Persson, Seven-Year Study on the Effect of Silica Fume in Concrete, Advanced Cement Based Materials 7 (1998) 139–155. https://doi.org/10.1016/S1065-7355(98)00003-0.

[34] W. Bai, X. Lu, J. Guan, C. Yuan, Experimental study on uniaxial compression mechanical properties of recycled concrete with silica fume considering the effect of curing age, Constr Build Mater 350 (2022) 128758. https://doi.org/10.1016/j.conbuildmat.2022.128758.

[35] Z. Liu, K. Takasu, H. Koyamada, H. Suyama, A study on engineering properties and environmental impact of sustainable concrete with fly ash or GGBS, Constr Build Mater 316 (2022) 125776. https://doi.org/10.1016/j.conbuildmat.2021.125776.

[36] F. Yang, Y. Yao, X. Wang, J. Wei, Z. Feng, Preparation of Recycled and Multi-Recycled Coarse Aggregates Concrete with the Vibration Mixing Process, Buildings 12 (2022) 1369. https://doi.org/10.3390/buildings12091369.

[37] J. Kim, A.M. Grabiec, A. Ubysz, S. Yang, N. Kim, Influence of Mix Design on Physical, Mechanical and Durability Properties of Multi-Recycled Aggregate Concrete, Materials 16 (2023) 2744. https://doi.org/10.3390/ma16072744.

[38] S.B. Huda, M.S. Alam, Mechanical behavior of three generations of 100% repeated recycled coarse aggregate concrete, Constr Build Mater 65 (2014) 574–582. https://doi.org/10.1016/j.conbuildmat.2014.05.010.

[39] P. Zhu, X. Zhang, J. Wu, X. Wang, Performance degradation of the repeated recycled aggregate concrete with 70% replacement of three-generation recycled coarse aggregate, Journal of Wuhan University of Technology-Mater. Sci. Ed. 31 (2016) 989–995. https://doi.org/10.1007/s11595-016-1480-y.

[40] O. Belabbas, F. Bouziadi, B. Boulekbache, M. Hamrat, A. Haddi, S. Amziane, Mechanical properties of multi-recycled coarse aggregate concrete, with particular emphasis on experimental and numerical assessment of shrinkage at different curing temperatures, Journal of Building Engineering 89 (2024) 109333. https://doi.org/10.1016/j.jobe.2024.109333.

[41] P. Zhu, Y. Hao, H. Liu, D. Wei, S. Liu, L. Gu, Durability evaluation of three generations of 100% repeatedly recycled coarse aggregate concrete, Constr Build Mater 210 (2019) 442–450. https://doi.org/10.1016/j.conbuildmat.2019.03.203.

[42] C. Chen, R. Liu, P. Zhu, H. Liu, X. Wang, Carbonization Durability of Two Generations of Recycled Coarse Aggregate Concrete with Effect of Chloride Ion Corrosion, Sustainability 12 (2020) 10544. https://doi.org/10.3390/su122410544.

[43] W.-T. Lin, Effects of sand/aggregate ratio on strength, durability, and microstructure of self-compacting concrete, Constr Build Mater 242 (2020) 118046. https://doi.org/10.1016/j.conbuildmat.2020.118046.

[44] T. Dittmer, H. Beushausen, The effect of coarse aggregate content and size on the age at cracking of

bonded concrete overlays subjected to restrained deformation, Constr Build Mater 69 (2014) 73–82. https://doi.org/10.1016/j.conbuildmat.2014.06.056.

[45] M.S. Meddah, S. Zitouni, S. Belâabes, Effect of content and particle size distribution of coarse aggregate on the compressive strength of concrete, Constr Build Mater 24 (2010) 505–512. https://doi.org/10.1016/J.CONBUILDMAT.2009.10.009.

[46] S. Silva, L. Evangelista, J. de Brito, Durability and shrinkage performance of concrete made with coarse multi-recycled concrete aggregates, Constr Build Mater 272 (2021) 121645. https://doi.org/10.1016/J.CONBUILDMAT.2020.121645.

[47] H. Liu, M. Hua, P. Zhu, C. Chen, X. Wang, Z. Qian, Y. Dong, Effect of Freeze–Thaw Cycles on Carbonation Behavior of Three Generations of Repeatedly Recycled Aggregate Concrete, Applied Sciences 11 (2021) 2643. https://doi.org/10.3390/app11062643.

[48] L. Dadd, T. Xie, B. Bennett, P. Visintin, Exploring the physical and mechanical characteristics of multi-generation recycled aggregate concrete at equivalent compressive strengths, J Clean Prod 451 (2024) 141965. https://doi.org/10.1016/j.jclepro.2024.141965.

[49] D.C. Feng, Z.T. Liu, X.D. Wang, Y. Chen, J.Q. Chang, D.F. Wei, Z.M. Jiang, Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach, Constr Build Mater 230 (2020) 117000. https://doi.org/10.1016/J.CONBUILDMAT.2019.117000.

[50] Z. Cui, G. Gong, The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features, Neuroimage 178 (2018) 622–637. https://doi.org/10.1016/j.neuroimage.2018.06.001.