scientific reports

OPEN

Physics-assisted machine learning methods for predicting the splitting tensile strength of recycled aggregate concrete

Jianguo Liu, Xiangyu Han, Yin Pan, Kai Cui & Qinghua Xiao✉

Recycled aggregate concrete (RAC) has become a popular building material due to its eco-friendly features, but the difficulty in predicting the crack resistance of RAC is increasingly impeding its application. In this study, splitting tensile strength is adopted to describe the crack resistance ability of RAC, and physics-assisted machine learning (ML) methods are used to construct the predictive models for the splitting tensile strength of RAC. The results show that the AdaBoost model has excellent predictive performance with the help of the Firefly algorithm, and physical assistance plays a remarkable role in selecting features and verifying the ML models. Due to the limit in data size and the generalizability of the model, the dataset should be supplemented with more representative data, and an algorithm for small sample sizes could be studied in the future.

According to reports from the National Development and Reform Commission of China, the total production of commercial concrete exceeded 32 billion cubic meters in 2021[1]. Meanwhile, with the demolition of old and abandoned buildings, more than 2 billion tons of concrete waste has been produced. Hence, recycled aggregate concrete (RAC), which uses recycled concrete waste to replace some coarse aggregates, has become a new trend in concrete production. Normally, the cost of using recycled aggregate in concrete is generally lower than that of using new materials. RAC also contributes to reducing the amount of waste that goes to landfills as well as the amounts of energy and carbon emissions needed to produce new materials.

During the recycled aggregate manufacturing process, the obtained recycled aggregate always consists of natural aggregate and hardened mortar. Hence, compared with that of natural aggregate, the recycled aggregate is weaker due to the weakness of hardened mortar and the interfaces between the mortar and natural aggregate. Furthermore, a large number of cracks are generated inside the recycled aggregate during the crushing process[2,3]. As shown in Fig. 1, various weak regions are introduced into recycled aggregate concrete, and such material is peculiarly prone to cracking when exposed to external loads, which severely restricts its application. Hence, compared with other mechanical properties, the cracking properties of RAC deserve more attention.

The presence of various fibers improves the ability of fiber-reinforced RAC to resist cracking[4–6]. Akca et al.[7] applied polypropylene fibers to reinforce RAC and proved that both flexural tensile strength and splitting tensile strength increase with increasing fiber content. Ali et al.[8] compared the mechanical properties of glass fiber-reinforced RAC and plain RAC, and an obvious increase was witnessed in the splitting tensile strength of fiber-reinforced recycled aggregate concrete. Gao et al.[9] studied the performance of steel fiber-reinforced RAC, and the flexural strength was improved significantly with increasing steel fiber volume fraction. It can be concluded that the fibers could efficiently improve the cracking resistance of RAC, but the mechanism of reinforcement is affected by various factors.

For plain concrete, various methods have been proposed to predict its cracking performance, such as fracture tests, numerical methods and mechanical models[10–13], and different factors are studied to guarantee the prediction reliability. Even so, it is still difficult to precisely describe the cracking characteristics of concrete. Considering the influence of recycled aggregate, the prediction of the cracking behavior of RAC, especially of RAC with fibers, is more complex. Machine learning methods are novel approaches for prediction issues[14–16]. Pan and Amin attempted to use machine learning (ML) to predict the cracking characteristics of RAC[17,18] and established several ML models with superior performance. However, during the construction of ML models, the physical meaning of the model is neglected, and the influence of fibers is not considered.

School of Civil Engineering, Southwest Jiaotong University, Chengdu 610031, People's Republic of China. ✉email: xqhbp@swjtu.edu.cn
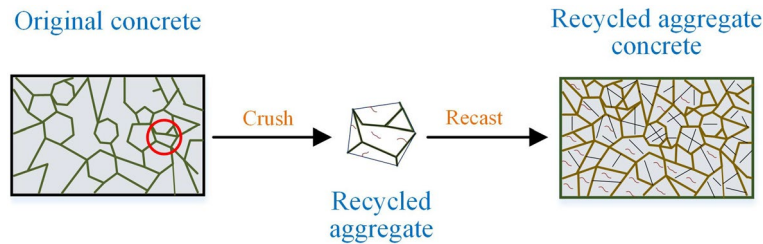
Figure 1. Various weak regions of recycled aggregate concrete.

In this study, machine learning approaches are combined with existing mechanical models and physical experiments, and a precise prediction of the splitting tensile strength of RAC is attempted. This paper is arranged as shown in Fig. 2 and includes (1) the construction of a database with the help of physical theories, (2) the illustration of several machine learning algorithms used in the current study, (3) the presentation of the construction process of predictive models, (4) the testing of the performance of predictive models with mathematical and physical methods and (5) the presentation of conclusions.

Methods

Physics-assisted construction of the database. For concrete materials, a splitting test is always adopted to observe their cracking behavior, and the splitting tensile strength is applied to evaluate their ability to resist cracking. Hence, the splitting tensile strength of RAC is selected as the predicted target (output variables) to represent the cracking performance, and various influencing factors are regarded as the input variables. These two types of variables compromise machine learning databases.

As the first step of machine learning, the selection of variables and collection of data are significantly important, which affects the prediction accuracy and generalizability of ML models. For conventional ML models, the physical meaning of variables is always neglected, leading to blind ML models that cannot handle cases outside of the constructed database. Therefore, physical experiments and existing mechanical models are used to assist in the construction of the database.

Since there are no well-established mechanical models for RAC, plain concrete cracking models are referenced here. According to well-known mechanical models, such as the fictitious-crack model, size effect model, and boundary effect model, the cracking/fracture properties of concrete are irrelevant to the specimen geometries but determined by mixture design[13,19,20]. Hence, the contents of water, cement and coarse aggregate and the size of aggregate are chosen as the input variables. Moreover, based on published studies[4,6,21–23], the splitting tensile strength is also affected by other factors, such as the characteristics of the recycled aggregate (content, density, water absorption, size), the fiber types and the volume fractions.

Based on the established guidance for selecting features, a total of 257 data points were collected from the published literature[2,4–6,8,9,21–44]. It should be noted that in these collected experiments, the testing process follows the requirements of relevant standards, such as the Standard Test Method for Splitting Tensile Strength of Cylindrical Concrete Specimens (ASTM C496/C496M-2017) and the Standard for Test Method of Mechanical Properties on Ordinary Concrete (GB/T 20081-2002), and the obtained splitting tensile strength can be regarded as the material property. In most experiments on RAC, to eliminate the entropy factor in the results, natural and
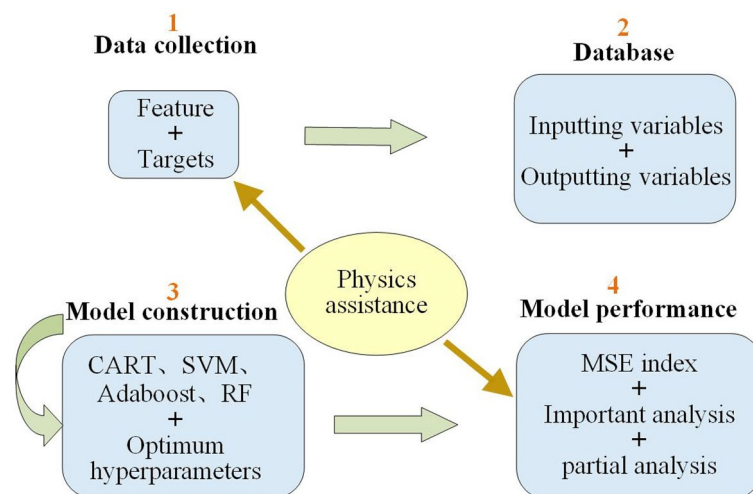
Figure 2. Flowchart of the physics-assisted machine learning process in this study.

recycled aggregates are always sieved together and have the same size distribution, and the sizes of natural and recycled aggregates are incorporated into the index of the maximum size of coarse aggregates. Consequently, the mathematical characteristics of 10 numeric variables and one nonnumeric variable are listed in Table 1. These include water, cement, NCA (natural concrete aggregate) content, RCA (recycled concrete aggregate) content, SP (superplasticizer), D_{max}_RCA (maximum aggregate size of RCA), ρ_{RCA} (density of RCA), W_{RCA} (water absorption of RCA), fibers, STS (splitting tensile strength) and fiber type. A scattered distribution can be found in these numeric variables, and their values are quite different. Hence, to guarantee that the values of each feature can be reasonably used during the training process, preprocessing work is implemented for these data. First, the missing values are treated with K-nearest neighbor (KNN) methods. Then, to ensure the equal status of different features, min–max normalization is applied to these variables, as shown in Eq. (1), where x_{min} and x_{max} are the minimum and maximum values of feature x, respectively.

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} (i = 1, 2, 3 \dots n)$$ (1)

Next, since the fiber type value is nonnumeric and cannot be directly used in the training or testing process, these values are converted into numeric values. Finally, the correlation between 9 numeric input variables is analyzed with a correlation matrix, as shown in Fig. 3. There is a strong relationship between the NCA and RCA content. From the physics perspective, to ensure the workability and mechanical properties of recycled aggregate concrete, the total contents of coarse aggregate are determined using a mixture design, and the NCA and RCA contents are directly related. Hence, the features of NCA are dropped here. Consequently, 8 numeric features and 1 nonnumeric feature are identified as the input variables, and the splitting tensile strength is regarded as the output variable.

Machine learning algorithms.

1. Classification and regression tree (CART) As the most frequently used supervised machine learning method, the classification and regression tree (CART) algorithm can be easily applied in classification and regression problems[45]. The difference between a classification tree and a regression tree is the type of target values. When the target belongs to a discrete variable, the CART is a classification tree. When the target belongs to a consecutive variable, the CART is a regression tree. In this study, the CART is a regression tree.

A conventional CART is composed of a root node, a decision node and a leaf node, as shown in Fig. 4. The root node that contains all the data is split into two subsets following the recursive binary splitting criterion. During the splitting procedure, the data in each subset should be kept as homogeneous as possible. Then, the decision node is split into two sets following the same principle, and the complexity of the variance in each subset is reduced, but the model becomes increasingly complicated. Such a partitioning process will stop when one of the following conditions is met: (1) when the data in each leaf node share the same characteristics or (2) when the depth of the tree reaches its maximum value. Consequently, a fully grown tree is generated.

For a fully grown tree, a predictive model often has excellent performance during the training process but cannot precisely predict the target values with the testing dataset. This overfitting phenomenon is caused by the complex structure of CART. Hence, another essential procedure of the CART algorithm is pruning work. First, redundant branches are removed from the bottom of the fully grown tree, forming a sequence that consists of different subtrees. Then, the cross-validation method is used to test the performance of subtrees and select the optimum subtree as the ultimate predictive model. During the training of each CART model, the following parameters should be determined: max_depth, min_samples_split, and min_samples_leaf.
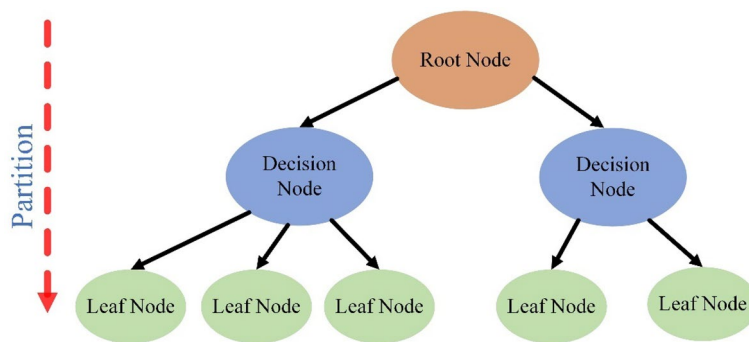
2. Support vector regression (SVR).

Item	Mean	Minimum	Maximum	Standard deviation
Water (kg/m³)	185.6	98.3	343.5	37.4
Cement (kg/m³)	369.9	158.0	600.0	62.0
NCA (kg/m³)	344.8	0	1143.0	363.6
RCA (kg/m³)	732.8	57.0	1474.0	387.4
SP (kg/m³)	1.0	0	7.8	1.6
D_{max}_RCA (mm)	18.1	10.0	25.0	4.3 k
ρ_{RCA} (kg/m³)	2431.0	2010.0	2661.0	156.7
W_{RCA} (%)	5.4	1.9	10.9	1.8
Fibers (vol. %)	0.3	0	2.0	0.4
STS (MPa)	3.1	1.4	7.6	1.1
Fiber type	Steel fiber, Carbon fiber, Polypropylene fiber, Basalt fiber, Glass fiber, Woolen fiber			

Table 1. The characteristics of the collected data.
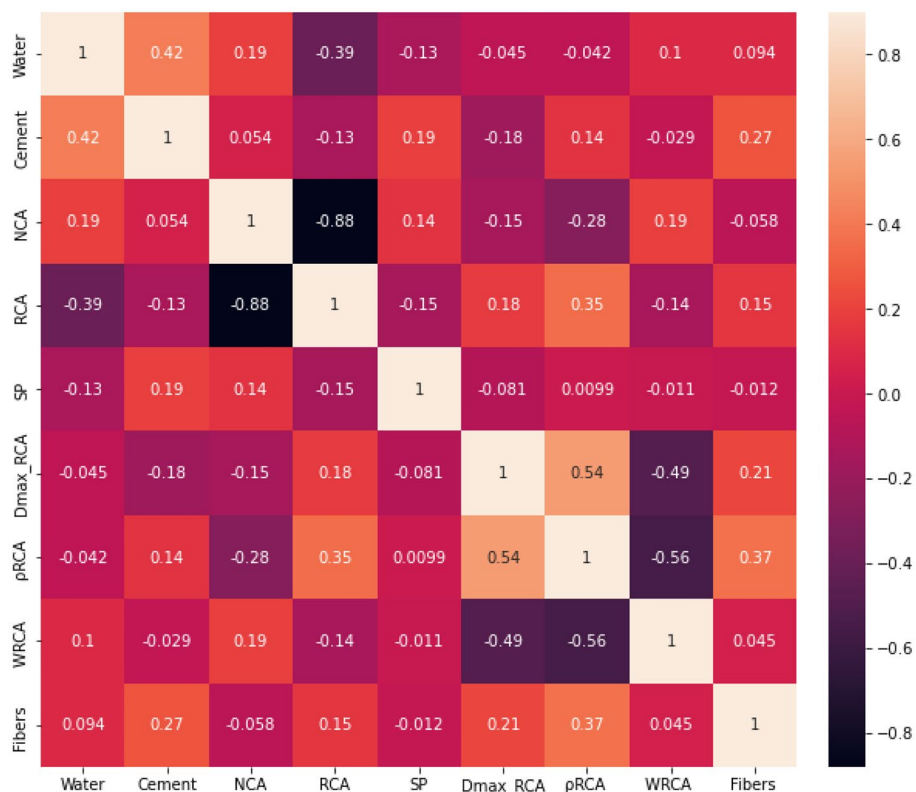
**Figure 3.** Correlation analysis of 9 numeric input variables.

Figure 4. The schematic of CART.

Support vector regression (SVR) is a kind of support vector machine algorithm that addresses regression problems[46]. Due to its high prediction accuracy and lower computational power requirements, SVR has been widely used in various fields.

The aim of SVR is to search a hyperplane that distinctly classifies the data points in the training data, and the dimensions of the hyperplane are determined by the number of features. There are various candidate hyperplanes, and the optimal plane often has the maximum distance between data points of both classes so that the data points can be partitioned with more confidence, as shown in Fig. 5. The data points that are close to the hyperplane are referred to as support vectors, and the position and orientation of the hyperplane are determined by these vectors. The two lines that are drawn around the hyperplane at a distance of ε are referred to as boundary lines, and they are used to create a margin between the data points. The value of ε reflects the tolerance of error in the training process, and a higher ε indicates a higher generalizability. Moreover, another essential part of SVR is the kernel, which consists of a set of mathematical functions. With the assistance of different kernels, the data can be transformed into the required form so that the hyperplane can be found in a higher dimensional space.

From the description of the SVR algorithm, it can be found that SVR is easy to implement and robust to outliers. However, the model will have poor performance when the dataset has too much noise. For one given SVR model, the C_penalty, kernel and tolerance should be determined during training.
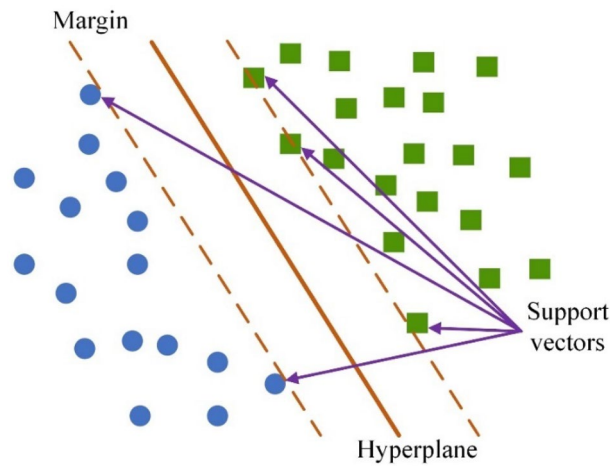
Figure 5. Schematic of the SVR algorithm.

3. Adaptive boosting (AdaBoost).

During the construction of machine learning models, it is essential to balance the prediction accuracy and generalizability of ML models. Prediction models with higher accuracy always have poor generalizability, and vice versa. Hence, ensemble learning algorithms are proposed to solve such problems.

The AdaBoost algorithm is a kind of ensemble learning algorithm that integrates various base learners[47]. As shown in Fig. 6, the procedure of Adaboost can be summarized as follows: (1) evenly assign the weights of all data points; (2) use the data to train the base learner and calculate its error; (3) adjust the weights of the data point by reducing the weights of accurately predicted data and increasing the weights of incorrectly predicted data; (4) repeat Step (2) and Step (3); and (5) integrate these base learners into the ensembled learner. The base learners with low error rates will have larger weights, and the base learners with high error rates will have smaller weights. Consequently, prediction models with good robustness are established. During construction of Adaboost models, the max_depth, min_samples_split, min_samples_leaf, n_estimators and loss are needed.
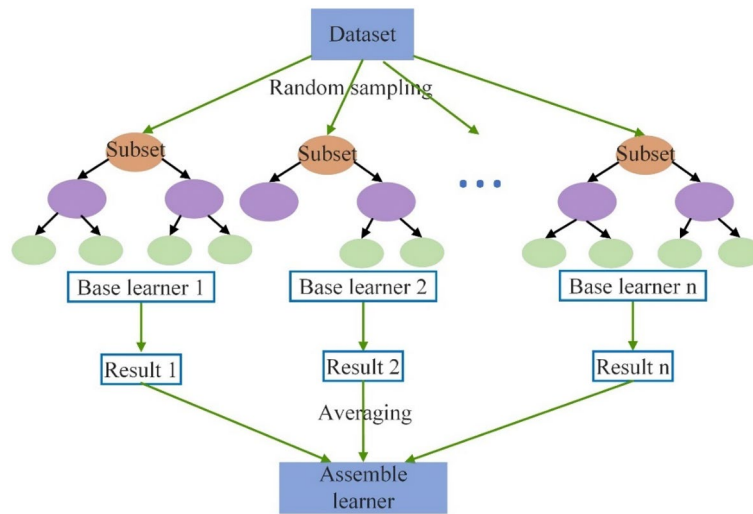
4. Random forest (RF).

The random forest (RF) algorithm is another important ensemble algorithm that was proposed by Breiman[48]. To improve the prediction accuracy and control overfitting, the bagging technique is applied to average the parallel base learners in RF.

As shown in Fig. 7, the random forest algorithm consists of sampling, model training, predicting and averaging procedures. First, the data points are randomly selected to construct subsets with replacements, and overlaps exist between different subsets. Next, during the growth of trees, the attributes to split the nodes are randomly selected. Thus, the pruning process is not necessary here. Then, these isolated base learners are used to implement prediction work. Finally, for regression problems, the base learners are integrated by averaging the predictive results in each iteration. It should be noted that the randomness of RF is introduced by data sampling and node splitting, and all the base learners are independent. Moreover, the following parameters could be set for better RF performance, such as max_depth, min_samples_split, min_samples_leaf, and n_estimators.
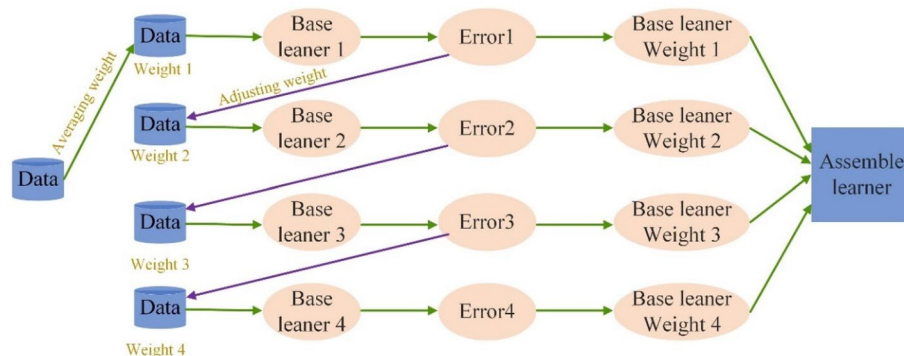
Figure 6. The procedure of the Adaboost algorithm.

Figure 7. Flowchart of the random forest algorithm.

Construction of predictive models. As illustrated in the above section, a total of 257 data points are prepared in the current study. To train and test the predictive models, the dataset, which consists of input variables and output variables, should be split into two subsets, a training set and a testing set, and the ratio between the training set and testing set is set as 0.7:0.3. Namely, 179 cases are assigned to construct the predictive models, and the remaining 78 cases are used to evaluate the model performance. It should be noted that to guarantee the generalizability of the predictive models, all the data should be randomly shuffled before dataset splitting.

During the training process, the predictive models are always constructed in a complex form, leading to different performance during training and testing. Hence, K-fold cross-validation is proposed to solve such problems by introducing an extra validation procedure during training[49]. First, the training subset is evenly partitioned into K parts. Next, in the first iteration, the former K−1 parts are applied to search the parameters of the predictive models, and the remaining part is used to validate the constructed models and calculate the prediction accuracy. Then, the above procedure is repeated K times to guarantee that each part can be used as a training set K-1 times and a validation set one time. Finally, the average value of the accuracy score in each iteration is regarded as the performance index of the constructed models. Therefore, the predictive models can be more adaptive to the data out of the range of the training set, and the generalizability is also enhanced. Moreover, K-fold cross-validation is helpful for tasks with limited data. In this study, the value of K is set as 10.

There is a large difference between the performance of AI models with different hyperparameters. Hence, another essential task during the training process is to search for the optimum hyperparameters. In this study, the firefly algorithm (FA) is applied to tune the hyperparameters of the prediction models. As a heuristic algorithm, the FA is inspired by the flashing behavior of fireflies[47]. The following conditions are assumed for FA: (1) fireflies are neutral with respect to sex, and these fireflies will be attracted to each other; (2) the attractiveness between fireflies is proportionate to their brightness, which is determined by the distance between them; and (3) if there is no brighter firefly than a given firefly, it will move randomly. After several iterations, the brightest firefly will be found. For the current predictive models, the mean squared error (MSE) is regarded as the objective function, as shown in Eq. (2), where y_i and $\hat{y}_i$ are experimental and predicted values, respectively. When the MSE reaches the lowest value during tenfold cross-validation, the optimum hyperparameters are obtained.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 (i = 1, 2, 3 \ldots n)$$
(2)

Therefore, ten-fold cross-validation and the FA algorithm are used to tune the hyperparameters of the 4 algorithms in section "Physics-assisted verification of prediction models". As shown in Fig. 8, the MSE values of all the models reach a stable state within 6 iterations. That is, the FA algorithm is efficient for tuning the hyperparameters of the prediction models. Moreover, the RF and Adaboost models have better performance than that of the CART and SVR models. The optimum hyperparameters of the 4 models are listed in Table 2.

After the AI models are established, the following step is to test the generalizability of these models. As illustrated in above sections, the RF predictive model obtained the optimum performance during training. However, a qualified AI model should not only perform well in the training process but also have excellent performance in predicting unseen data.

Therefore, the 78 cases in the testing set are used to test the above 4 predictive models. First, the input variables are used to feed the AI models. Then, the predicted targets are obtained for each case. Then, the predicted targets are compared with the output variables. For regression problems, the difference between the predicted targets and real targets can evaluated with the MSE score.
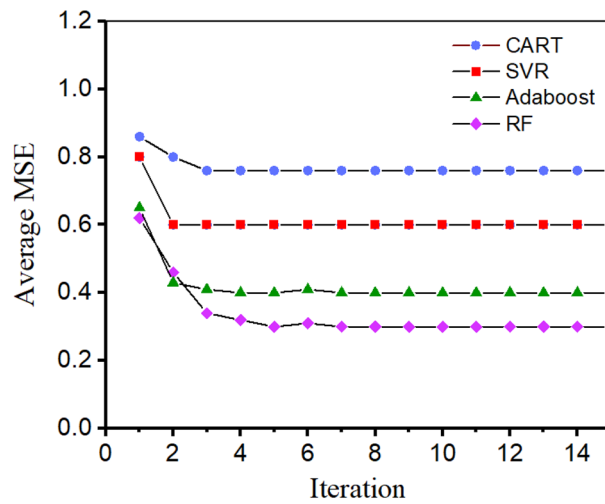
Figure 8. The variation in the average MSE with tuning iterations.

AI models	Hyperparameters	Optimum value
CART	max_depth	11
	min_samples_split	7
	min_samples_leaf	6
SVR	C_penelty	7.9
	kernel	linear
	tolerance	0.0001
Adaboost	max_depth	12
	min_samples_split	8
	min_samples_leaf	4
	n_estimators	323
	loss	exponential
RF	max_depth	19
	min_samples_split	3
	min_samples_leaf	2
	n_estimators	126

Table 2. The optimum hyperparameters of 4 AI models.
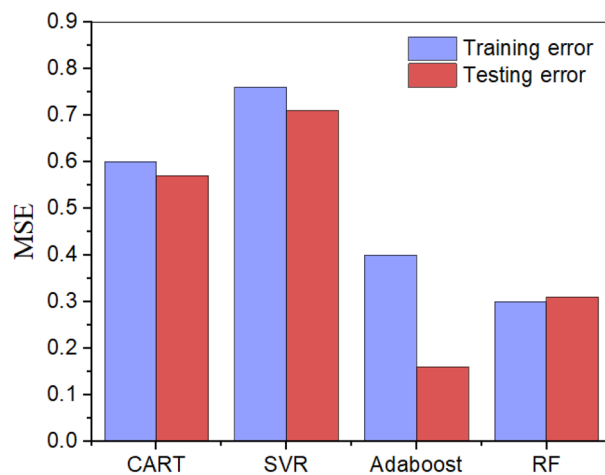
**Figure 9.** Comparison of the prediction performance between four AI models.

As shown in Fig. 9, the performance of the four AI models is compared. The ensemble models have better performance than that of the CART and SVM models in both the training and testing procedures. By integrating several different base learners, the prediction accuracy and generalizability of ensembled models are enhanced. Although the MSE of the AdaBoost model is slightly higher than that of the RF model in the training procedure, the AdaBoost model has better performance in the testing process. Then, the predicted values using AdaBoost and the real values are compared, as shown in Fig. 10, and the high consistency of the results reflects the excellent prediction ability of the AdaBoost model.

Discussion

Physics-assisted verification of prediction models. Normally, AI models are regarded as a black box. These models pay much more attention to the reliability of the data than the physical meaning of the parameters. Consequently, the constitution of AI models is always too complex to interpret. Hence, the importance of features is proposed as an index to evaluate the contribution of various features to the AI model.

In this section, the AdaBoost model is selected to analyze the importance of 9 features due to its excellent prediction performance. For each feature, its contribution to the impurity of the base learner of the AdaBoost model can be calculated. Then, by averaging the calculated values and implementing normalization, the importance score of each of the 9 features is obtained. As shown in Fig. 11, the importance of these features is ranked. The recycled aggregate content has the highest importance score of 0.23. The size of the recycled aggregate and water content also have considerable influence on the AI models. The fiber type is regarded as a negligible feature for the predictive models of the splitting tensile strength of RAC, which obtains the lowest importance score of 0.02.

After that, the predicted model of splitting tensile strength is explained from the perspective of physics. As illustrated above, the recycled aggregate content, recycled aggregate size and water content are regarded as the three most influential factors. For a given RAC mixture, an increase in the recycled aggregate content will lead to a reduction in the natural aggregate content. Due to the inner voids and cracks, the RAC always fractures
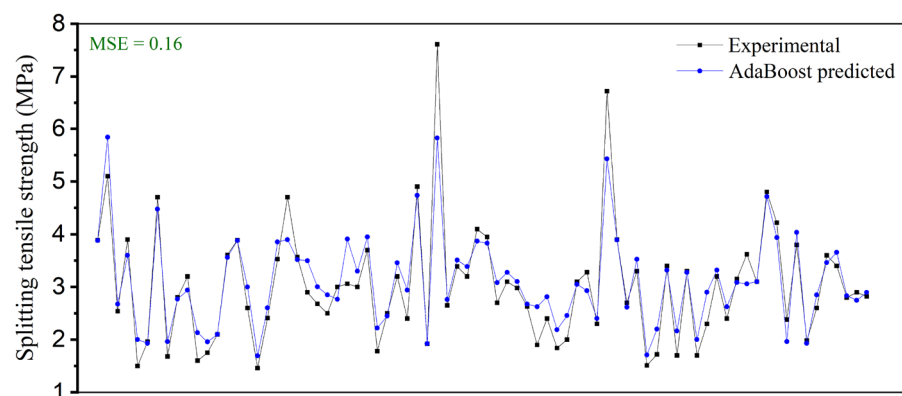
Figure 10. Comparison between the experimental and AdaBoost-predicted splitting tensile strength values.
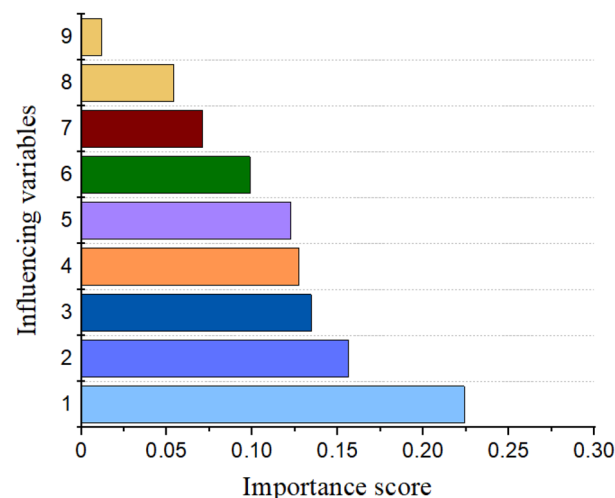
Figure 11. The importance of various influencing variables.

around these weak regions, which will further impair the splitting tensile strength of RAC. Moreover, when incorporating recycled aggregates, the contents of ITZs (interfacial transition zones) are enhanced, which will introduce a weak RAC location. Hence, the recycled aggregate content has a significant impact on the splitting tensile strength of RAC, and this conclusion is consistent with that of other AI models[18]. Next, the influence of the maximum aggregate size is analyzed. In the collected literature, the maximum size of recycled aggregate and natural aggregate is always consistent, and the maximum aggregate size index indicates the recycled aggregate and natural aggregate. In the splitting process of RAC, a crack will propagate along the surface of the natural aggregate or cross the recycled aggregate. Hence, the aggregate size will determine the cracking path and further affect the splitting strength. Then, the water and cement contents, which determine the w/c ratios, have been proven to be relevant to the tensile strength of various concretes. Finally, the fiber volume and fiber types are also regarded as influential factors for the splitting tensile strength of RAC but are not fully reflected in the AI models. This phenomenon may be caused by the distribution of collected data.

Then, the AI models are verified by physical experiments. From the importance analysis of different features, the recycled aggregate content is proven to be the most influential feature for the AI predictive models. Then, the dependence of AI models on these features is analyzed using partial dependence analysis.

The partial dependence analysis can be implemented as follows: (1) select one feature as the research object, (2) change the object in a reasonable range, keeping other features unchanged, (3) feed the AI models with different input variables, obtaining the predictive targets, (4) analyze the dependence of the AI models on the selected feature, and (5) compare the variation tendency with the experimental results.

Taking the experiments in Reference[18] as an example, the recycled aggregate content varies from 100 to 1000 kg/m³, and the other 8 features are kept consistent with the mixture design in the experiments. These features are set as the input variables, and the constructed AdaBoost models in section "Discussion" are used to predict the splitting tensile strength of RAC. Figure 12 illustrates the variation tendency of the AI-predicted splitting tensile strength with the change in recycled aggregate content, and the experimental results are also plotted in the same coordinates. The prediction accuracy of the splitting tensile strength is acceptable. An obvious reduction occurs in the splitting tensile strength when the recycled aggregate content increased from 100 to 1000 kg/m³, which was proven by physical experiments. Despite the visible deviation, the prediction of AI models is convincing.

Compared with that of previous research[17,18], the construction of predictive models in the current study is convincing, and the established models are reliable with the assistance of physics. First, the influencing features are effectively selected from a physical view. Then, the feature importance analysis of predictive models is deeply discussed with mechanical theories, making the models more explainable. After that, partial dependence analysis of AI models is implemented with physical experiments, making the models more reliable.

Conclusions

Due to various defects in recycled aggregates, the splitting tensile strength of RAC is weakened and interferes with the application of RAC. Fiber reinforcement methods have been proposed to solve such problems, but the cracking performance of these RACs is more complicated. In this study, four artificial intelligence methods are used to predict the tensile behavior of RAC, and physical assistance is used in the construction and verification of predictive models. The conclusions can be drawn as follows:

1. With the help of the FA, the optimum parameters of the models are efficiently searched. The mean square error (MSE) is selected to evaluate the performance of the model, and the MSE values for the CART, SVR,
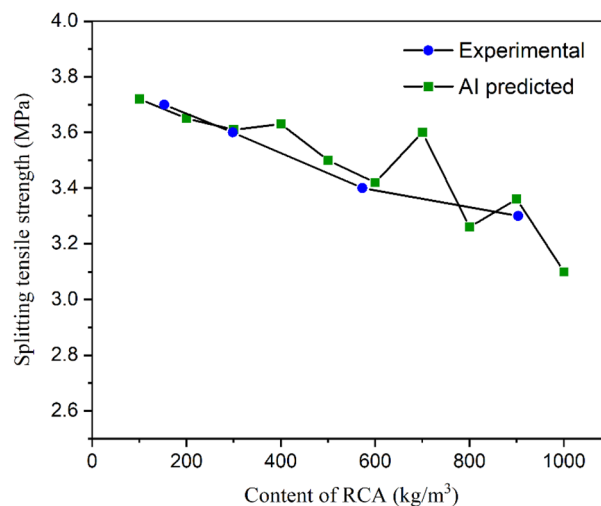
Figure 12. Variation in splitting tensile strength with the change in RCA content.

AdaBoost and RF models are 0.572, 0.713, 0.164 and 0.311, respectively. The superiority of the AdaBoost models can be attributed to the ensemble of various base learners.

2. Feature importance is determined, and the following input importance with an increasing pattern was observed for the AdaBoost models: recycled concrete aggregate (0.224) > maximum aggregate size (0.156) > water content (0.134) > cement content (0.127) > fiber superplasticizer (0.099) > density of RCA (0.071) > water absorption of RCA (0.054) > fiber type (0.012).

3. Physics is used to assist the machine learning models as follows: improving the efficiency of the construction of AI models by selecting highly related features, analyzing the feature importance from the fracture mechanism of RAC, and verifying the reliability of AI models with physical experiments.

4. The application of established predictive models can be described as follows: by collecting 9 features of one RAC and inputting these features into the models, the splitting tensile strength can be obtained. It should be noted that the physical verification of predictive results must be implemented to guarantee its reliability.

Data availability

All data generated or analysed during this study are included in the supplementary information files.

References

1. The consumption of construction material of China in 2021, National Development and Reform Commission (2022; accessed 20 September 2022). https://www.ndrc.gov.cn/fgsj/tjsj/jjyx/mdyqy/202201/t20220130_1314182.html?code=&state=123.
2. Younis, K. H. & Pilakoutas, K. Strength prediction model and methods for improving recycled aggregate concrete. *Constr. Build. Mater.* **49**, 688–701 (2013).
3. Fonseca, N., De Brito, J. & Evangelista, L. The influence of curing conditions on the mechanical performance of concrete made with recycled concrete waste. *Cement Concr. Compos.* **33**(6), 637–643 (2011).
4. Chakradhara Rao, M., Bhattacharyya, S. K. & Barai, S. V. Influence of field recycled coarse aggregate on properties of concrete. *Mater. Struct.* **44**(1), 205–220 (2011).
5. Katkhuda, H. & Shatarat, N. Improving the mechanical properties of recycled concrete aggregate using chopped basalt fibers and acid treatment. *Constr. Build. Mater.* **140**, 328–335 (2017).
6. Meesala, C. R. Influence of different types of fiber on the properties of recycled aggregate concrete. *Struct. Concr.* **20**(5), 1656–1669 (2019).
7. Akça, K. R., Çakır, Ö. & İpek, M. Properties of polypropylene fiber reinforced concrete using recycled aggregates. *Constr. Build. Mater.* **98**, 620–630 (2015).
8. Ali, B. & Qureshi, L. A. Influence of glass fibers on mechanical and durability performance of concrete with recycled aggregates. *Constr. Build. Mater.* **228**, 116783 (2019).
9. Gao, D. & Zhang, L. Flexural performance and evaluation method of steel fiber reinforced recycled coarse aggregate concrete. *Constr. Build. Mater.* **159**, 126–136 (2018).
10. Yuan, F. *et al.* Full-field measurement and fracture and fatigue characterizations of asphalt concrete based on the SCB test and stereo-DIC. *Eng. Fract. Mech.* **235**, 107127 (2020).
11. Sun, X. *et al.* Fracture performance and numerical simulation of basalt fiber concrete using three-point bending test on notched beam. *Constr. Build. Mater.* **225**, 788–800 (2019).
12. Chen, Y., Hu, Y. & Hu, X. Quasi-brittle fracture analysis of large and small wedge splitting concrete specimens with size from 150 mm to 2 m and aggregates from 10 to 100 mm. *Theor. Appl. Fract. Mech.* **121**, 103474 (2022).
13. Hu, X. *et al.* Modelling fracture process zone width and length for quasi-brittle fracture of rock, concrete and ceramics. *Eng. Fract. Mech.* **259**, 108158 (2022).
14. Shah, M. I. *et al.* Machine learning modeling integrating experimental analysis for predicting the properties of sugarcane bagasse ash concrete. *Constr. Build. Mater.* **314**, 125634 (2022).
15. Javed, M. F. *et al.* New prediction model for the ultimate axial capacity of concrete-filled steel tubes: An evolutionary approach. *Crystals* **10**(9), 741 (2020).
16. Khan, S. *et al.* Predicting the ultimate axial capacity of uniaxially loaded cfst columns using multiphysics artificial intelligence. *Materials* **15**(1), 39 (2021).
17. Pan, X. *et al.* Use of artificial intelligence methods for predicting the strength of recycled aggregate concrete and the influence of raw ingredients. *Materials* **15**(12), 4194 (2022).
18. Amin, M. N. *et al.* Split tensile strength prediction of recycled aggregate-based sustainable concrete using artificial intelligence methods. *Materials* **15**(12), 4296 (2022).
19. Hillerborg, A., Modéer, M. & Petersson, P. E. Analysis of crack formation and crack growth in concrete by means of fracture mechanics and finite elements. *Cem. Concr. Res.* **6**(6), 773–781 (1976).
20. Bažant, Z. P. & Planas, J. *Fracture and Size Effect in Concrete and Other Quasibrittle Materials* (Routledge, 2019).
21. Gómez-Soberón, J. M. V. Porosity of recycled concrete with substitution of recycled concrete aggregate: An experimental study. *Cem. Concr. Res.* **32**(8), 1301–1311 (2002).
22. Yang, K. H., Chung, H. S. & Ashour, A. F. Influence of type and replacement level of recycled aggregates on concrete properties. *ACI Mater. J.* **105**(3), 289–296 (2008).
23. Das, C. S. *et al.* Performance evaluation of polypropylene fibre reinforced recycled aggregate concrete. *Constr. Build. Mater.* **189**, 649–659 (2018).
24. Butler, L., West, J. S. & Tighe, S. L. Effect of recycled concrete coarse aggregate from multiple sources on the hardened properties of concrete with equivalent compressive strength. *Constr. Build. Mater.* **47**, 1292–1301 (2013).
25. Thomas, C. *et al.* Durability of recycled aggregate concrete. *Constr. Build. Mater.* **40**, 1054–1065 (2013).
26. Andreu, G. & Miren, E. Experimental analysis of properties of high performance recycled aggregate concrete. *Constr. Build. Mater.* **52**, 227–235 (2014).
27. Pereira, P., Evangelista, L. & De Brito, J. The effect of superplasticizers on the mechanical performance of concrete made with fine recycled concrete aggregates. *Cem. Concr. Compos.* **34**(9), 1044–1052 (2012).
28. Duan, Z. H. & Poon, C. S. Properties of recycled aggregate concrete made with recycled aggregates with different amounts of old adhered mortars. *Mater. Des.* **58**, 19–29 (2014).

29. Folino, P. & Xargay, H. Recycled aggregate concrete–Mechanical behavior under uniaxial and triaxial compression. *Constr. Build. Mater.* **56**, 21–31 (2014).
30. Pedro, D., De Brito, J. & Evangelista, L. Performance of concrete made with aggregates recycled from precasting industry waste: Influence of the crushing process. *Mater. Struct.* **48**(12), 3965–3978 (2015).
31. Thomas, C. *et al.* Evaluation of the fatigue behavior of recycled aggregate concrete. *J. Clean. Prod.* **65**, 397–405 (2014).
32. Etxeberria, M., Marí, A. R. & Vázquez, E. Recycled aggregate concrete as structural material. *Mater. Struct.* **40**(5), 529–541 (2007).
33. Dong, J. F., Wang, Q. Y. & Guan, Z. W. Material properties of basalt fibre reinforced concrete made with recycled earthquake waste. *Constr. Build. Mater.* **130**, 241–251 (2017).
34. Fang, S. E., Hong, H. S. & Zhang, P. H. Mechanical property tests and strength formulas of basalt fiber reinforced recycled aggregate concrete. *Materials* **11**(10), 1851 (2018).
35. Ibrahm, H. A. & Abbas, B. J. Mechanical behavior of recycled self-compacting concrete reinforced with polypropylene fibres. *J. Architect. Eng. Technol.* **6**(2), 1–7 (2017).
36. Fathifazl, G. *et al.* Creep and drying shrinkage characteristics of concrete produced with coarse recycled concrete aggregate. *Cem. Concr. Compos.* **33**(10), 1026–1037 (2011).
37. Liu, H. *et al.* Basic mechanical properties of basalt fiber reinforced recycled aggregate concrete. *Open Civ. Eng. J.* **11**, 1 (2017).
38. Kou, S. C., Poon, C. S. & Chan, D. Influence of fly ash as cement replacement on the properties of recycled aggregate concrete. *J. Mater. Civ. Eng.* **19**(9), 709–717 (2007).
39. Wang, Y. *et al.* A new method to improve the properties of recycled aggregate concrete: Composite addition of basalt fiber and nano-silica. *J. Clean. Prod.* **236**, 117602 (2019).
40. Zhang, X. *et al.* Orthogonal experimental study on strength of steel fiber reinforced fly ash recycled concrete. *J. Build. Mater.* **17**, 677–694 (2014) (**(In Chinese)**).
41. Kou, S. C., Poon, C. S. & Chan, D. Influence of fly ash as a cement addition on the hardened properties of recycled aggregate concrete. *Mater. Struct.* **41**(7), 1191–1201 (2008).
42. Zega, C. J. & Di Maio, A. A. Recycled concretes made with waste ready-mix concrete as coarse aggregate. *J. Mater. Civ. Eng.* **23**(3), 281–286 (2011).
43. Chen, S. *et al.* The evaluation of influence of different fibers on the properties of recycled aggregate pervious concrete. *J. Basic Sci. Eng.* **30**(1), 208–218 (2022) (**(In Chinese)**).
44. Guo, L. *et al.* Study on mechanical properties, water permeability and wear resistance of fiber modified recycled aggregate pervious concrete. *Trans. Chin. Soc. Agric. Eng.* **35**(02), 153–160 (2019) (**(In Chinese)**).
45. Lewis, R. J. An introduction to classification and regression tree (CART) analysis. In *Annual Meeting of the Society For Academic Emergency Medicine in San Francisco, California. Citeseer* (2000).
46. Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **14**(3), 199–222 (2004).
47. Solomatine, D. P., Shrestha, D. L. AdaBoost. RT: A boosting algorithm for regression problems. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541). IEEE, vol.* 2 1163–1168 (2004).
48. Breiman, L. Random forests. *Mach. Learn.* **45**(1), 5–32 (2001).
49. Bengio, Y. & Grandvalet, Y. No unbiased estimator of the variance of k-fold cross-validation. *Adv. Neural Inf. Process. Syst.* **2013**, 16 (2003).

Author contributions
Methodology, software, validation, writing, review and editing, J.G.L.; writing, methodology, funding acquisition, X.Y.H.; review, editing, validation, Y.P.; writing, review and editing, K.C.; conceptualization, validation, Q.H.X. All authors reviewed the paper and contributed to the discussions.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-36303-0.

Correspondence and requests for materials should be addressed to Q.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.