scientific reports

OPEN

Splitting tensile strength prediction of Metakaolin concrete using machine learning techniques

Qiang Li, Guoqi Ren, Haoran Wang, Qikeng Xu, Jinquan Zhao, Huifen Wang & Yonggang Ding ✉

Splitting tensile strength (STS) is an important mechanical property of concrete. Modeling and predicting the STS of concrete containing Metakaolin is an important method for analyzing the mechanical properties. In this paper, four machine learning models, namely, Artificial Neural Network (ANN), support vector regression (SVR), random forest (RF), and Gradient Boosting Decision Tree (GBDT) were employed to predict the STS. The comprehensive comparison of predictive performance was conducted using evaluation metrics. The results indicate that, compared to other models, the GBDT model exhibits the best test performance with an R^2 of 0.967, surpassing the values for ANN at 0.949, SVR at 0.963, and RF at 0.947. The other four error metrics are also the smallest among the models, with MSE = 0.041, RMSE = 0.204, MAE = 0.146, and MAPE = 4.856%. This model can serve as a prediction tool for STS in concrete containing Metakaolin, assisting or partially replacing laboratory compression tests, thereby saving costs and time. Moreover, the feature importance of input variables was investigated.

Concrete is the second most consumed material in the world after water. Portland cement is the primary binder used in most concrete applications[1–4]. The production of cement consumes a significant amount of energy and releases approximately 7% of global carbon dioxide emissions into the atmosphere[5–8]. However, the demand for cement continues to rise, and it is projected that annual cement consumption will reach 6 billion metric tons by 2060. One of the methods to reduce cement consumption is to use industrial by-products or more environmentally friendly materials that require less energy during manufacturing, such as Metakaolin. Studies have found that using Metakaolin as a partial replacement for cement can reduce carbon dioxide emissions by up to 170 kg per ton of cement produced[9,10]. Metakaolin is a highly reactive pozzolan that reacts with calcium hydroxide to form C-S–H and aluminate phases[11]. Incorporating Metakaolin as a partial replacement for cement in concrete helps to reduce pore size distribution, improve pore structure, and enhance various mechanical properties[12–14].

Previous studies have demonstrated that the addition of Metakaolin can effectively improve the performance of concrete, including enhancing compressive strength and durability, providing better resistance to freezing, weathering, chemical erosion, and permeability, as well as improving early-age concrete properties[15,16]. Therefore, estimating the mechanical properties of concrete, such as splitting tensile strength (STS), based on concrete mix proportions can help save time and costs, facilitate activities such as formwork removal, and promote the application of Metakaolin in the concrete industry. Previous research has mainly focused on finding the optimal Metakaolin content required to ensure the desired mechanical properties of concrete through experimental approaches[17,18]. However, the experimental process is time-consuming and labor-intensive. It would be highly useful to establish an intelligent model based on previous experimental data to predict the mechanical properties under given input mix proportions, which will significantly save experimental time and testing costs.

With the development of artificial intelligence, an increasing number of algorithms and models have provided new perspectives for addressing these problems[19–34]. The implementation of intelligent models for predicting the mechanical properties of concrete has also gained increasing attention[35–46]. Wu et al.[47] achieved accurate prediction of high-performance concrete tensile strength by utilizing a combination of support vector regression (SVR) and artificial neural network (ANN) models with optimization algorithms. Sourav et al.[48] employed support vector machine (SVM) and gradient boosting machine (GBM) models to predict the tensile strength of concrete, and the results indicated that GBM outperformed SVM in terms of prediction performance. Hammad et al.[49] utilized four models, namely gene expression programming (GEP), ANN, M5P model tree algorithm, and random forest (RF), to predict the flexural strengths of concrete with metakaolin, and the results demonstrated

College of Civil Engineering, Henan University of Technology, Zhengzhou 450001, China. ✉email: dyg9606@126.com

that random forest achieved the best predictive performance. Nozar et al. [50] studied the compressive strength of concrete containing metakaolin using the Multi-Layer Perceptron (MLP) model, and the results showed that the MLP network had reliable accuracy in predicting the compressive strength of concrete with metakaolin. Furthermore, user-friendly software was developed to facilitate the use of the proposed MLP network based on machine learning methods. Huang et al. [51] proposed a hybrid machine learning model combining RF and firefly algorithm (FA) to accurately predict the compressive strength of cementitious materials containing expansive clays based on a database of 361 samples. Abdulrahman et al. [16] compared the predictive performance of multiple individual models and ensemble models in predicting the compressive strength of cementitious materials containing expansive clays, and it was found that the DT AdaBoost model and the improved bagging model achieved the best predictive performance in predicting the STS of Metakaolin concrete.

However, there is relatively limited research and analysis on using machine learning models to predict the STS of concrete containing Metakaolin. Further research is needed in this area. Therefore, this paper aims to model and compare the STS of concrete containing Metakaolin using individual and ensemble models based on variables such as cement, Metakaolin, water-to-binder ratio (w/b), fine aggregate (FA), coarse aggregate (CA), superplasticizer (SP), age, height (H), and diameter (D) of concrete column specimen. The framework of this study is illustrated in Fig. 1.

## Machine learning models
Artificial neural network
ANN is a useful machine learning technique based on biological neural networks, designed to simulate complex relationships between inputs and outputs. The simplest processing element in a neural network is a neuron. Each neuron i may have multiple inputs, $x_1, x_2, …, x_d$, which are combined with corresponding weights, $w_{i1}, w_{i2}, …$, to produce a single output. More specifically, the propagation function combines these inputs with their weights and then applies an activation function to the resulting sum to generate the corresponding output [52]. The structure of an ANN is depicted in Fig. 2.

Support vector regression
SVM was originally proposed for studying linear problems. The basic idea behind SVM for pattern recognition is to transform the input space into a high-dimensional space through a non-linear transformation[53]. In this new space, the algorithm solves a convex quadratic programming problem to find the optimal linear classification hyperplane. However, when used for regression prediction, the fundamental idea is not to find an optimal classification plane that separates the samples but to find an optimal hyperplane that minimizes the distance
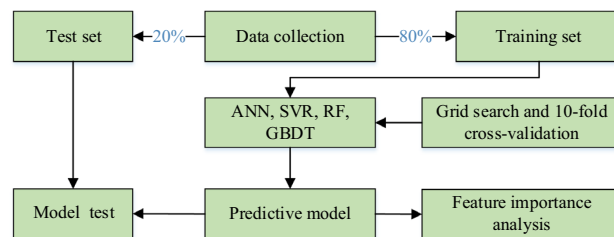
Figure 1. Paper framework workflow.
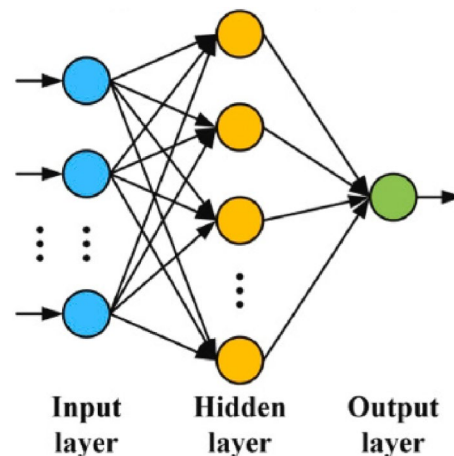
Figure 2. ANN network structure.

between the hyperplane and all training samples. This hyperplane can be considered a well-fitted curve, and the approach of using SVM for function approximation is known as SVR. SVR can be summarized as using a non-linear mapping function to map the input samples to a high-dimensional feature space and learning a linear regression quantity in the feature space to obtain the estimation function. The steps for implementing SVR for regression prediction are illustrated in Fig. 3.

Random forest

RF is an integrated learning model consisting of multiple decision trees. Its core idea is to improve prediction accuracy and stability by constructing multiple decision trees. As shown in Fig. 4, each decision tree is constructed based on random samples and random features, and this randomness makes Random Forest able to avoid overfitting and has good robustness. Advantages include: (1) Since random forests can utilize multiple
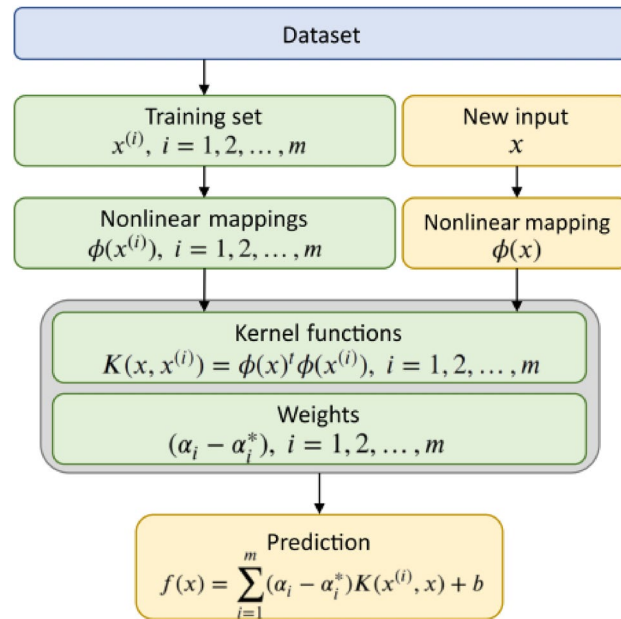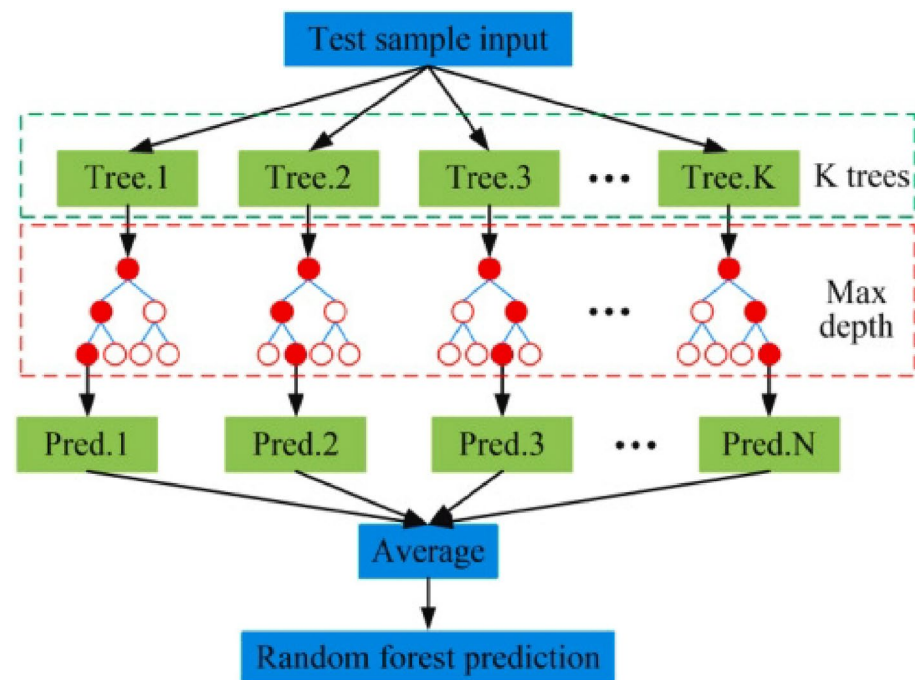

Figure 3. SVR algorithm flow[54].

Figure 4. Flowchart of random forest model[55].

decision trees for prediction, their prediction accuracy is higher than that of a single decision tree. (2) Random forests can handle a large number of input features, so they can be used for classification and regression problems with high-dimensional data. (3) Random forests are constructed using random samples and random features, and this randomness avoids the problem of overfitting.

Gradient boosting decision tree

Gradient Boosting Decision Trees (GBDT) is established based on the Boosting method in ensemble learning. It requires multiple iterations and the construction of multiple decision trees to form an ensemble model. During each iteration, the decision tree learners reduce the residuals along the direction of the steepest gradient descent. The algorithm is widely applied due to its strong interpretability, fast prediction speed, and the ability to freely combine multiple influential factors. When constructing the model, there is a strong correlation between each decision tree. Each subsequent decision tree adjusts its own weights based on the training results of the previous decision tree, and this process iterates until the desired residual or the maximum number of iterations is reached. The predictive model of GBDT can be represented as:

$$F(x) = \sum_{k=1}^{K} \omega_k g(x, \varphi_k) \tag{1}$$

where $F(x)$ is the response value of the input variable x; ω_k and φ_k are the weights and parameters of the k-th decision tree, respectively; and $g(x, \varphi_k)$ is the predicted value of the k-th decision tree.

Dataset collection

For machine learning models, a representative dataset is necessary and important. Therefore, this study collected a total of 204 samples from the literature[17,18,56–70]. The descriptive statistics and histogram distributions of the variables in these samples are shown in Table 1 and Fig. 5, respectively, where the input variables include component ratio, curing age, and specimen size. It can be observed that the content of "Metakaolin" ranges from 0 to 256 across the entire dataset, indicating a high degree of data variability. Furthermore, Fig. 6 presents the Pearson correlation coefficients between the variables. It can be seen that among the 9 input features listed, the linear correlation between "cement", "w/b" and "STS" is the strongest, with correlation coefficients of 0.3776 and −0.4362 respectively. However, this correlation is still weak, indicating that relying on multiple linear regression for predicting STS is unreliable due to the existence of complex nonlinear relationships between these variables and the output. This is why this study adopts machine learning models to achieve accurate predictions of STS. Moreover, the linear correlation between the input variables is weak, which is also an important prerequisite for machine learning applications.

Results and analysis
Model building

A total of 163 samples (80%) were randomly selected as the training set, and the remaining 41 samples (20%) were used as the test set for the trained model. After splitting the data, normalize the features to [0,1] to avoid scale effects. Referring to the literature[71], tenfold cross-validation and grid search methods were used to obtain the optimal hyper-parameters. The parameter value was determined in Table 2.

Performance comparison

Figure 7 illustrates the deviations between the predicted results and the actual results of each sample for different models. The training and testing results of different models are shown in Fig. 8. From the perspective of the coefficient of determination (R^2), all four models achieve good predictive performance. Among them, the GBDT model achieved the highest correlation coefficient of 0.967, followed by 0.963 for SVR, 0.949 for ANN,

Variable	Cement	Metakaolin	w/b	FA	CA	SP	Age	D	H	STS
Type	Input	Input	Input	Input	Input	Input	Input	Input	Input	Output
Size	204	204	204	204	204	204	204	204	204	204
Unit	kg/m³	kg/m³	/	kg/m³	kg/m³	kg/m³	Day	mm	mm	MPa
Max	570.00	256.00	0.75	989.00	1264.97	12.40	120.00	150.00	300.00	5.88
medium	400.00	41.00	0.43	818.10	848.00	4.12	28.00	150.00	300.00	3.30
Min	266.00	0.00	0.21	272.50	175.10	0.00	1.00	100.00	150.00	1.20
Std	65.63	38.86	0.11	180.72	266.44	3.34	31.66	24.30	56.21	1.05
Average	400.42	43.83	0.44	756.64	865.01	4.23	34.66	130.88	248.53	3.58
Cov	0.16	0.89	0.25	0.24	0.31	0.79	0.91	0.19	0.23	0.29
Skewness	0.13	1.33	0.40	−0.58	−1.11	0.41	1.22	−0.49	−0.32	0.44
Kurtosis	−0.36	4.29	−0.02	−0.39	1.60	−0.69	0.35	−1.78	−1.59	−0.51

Table 1. Characteristics of the variables.
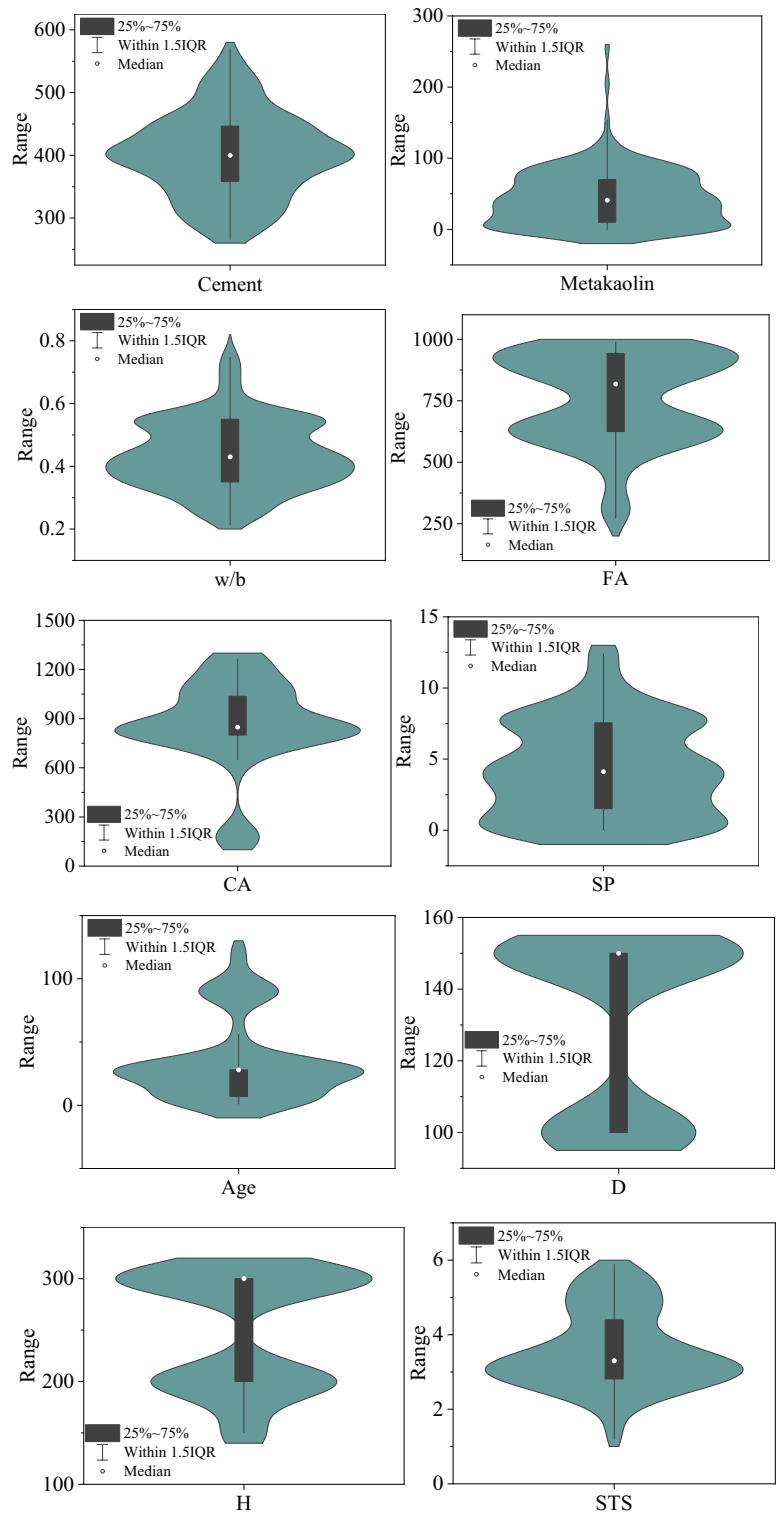
Figure 5. Distribution histogram of variables.

and 0.947 for RF. In general, relying on a single metric for evaluation may be unreliable. Therefore, four error metrics for each model's predictions were calculated, as shown in Table 3. It can be observed that compared to the other three models, the GBDT model achieves smaller error metrics. Specifically, the MSE, RMSE, MAE, and MAPE for the GBDT model are 0.041, 0.204, 0.146, and 4.856%, respectively.

For a more intuitive comparison, Fig. 9 presents the histograms of different model evaluation metrics. It can be concluded that overall, the GBDT model exhibits the best predictive performance among the machine learning models. Figure 10 shows a violin plot of the relative error percentages for different models. It can be
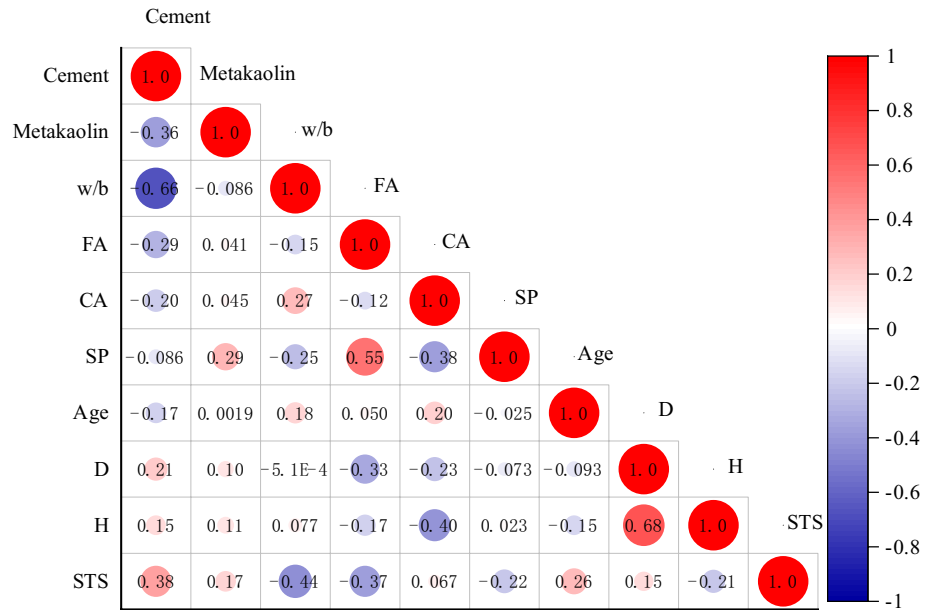
Figure 6. Pearson correlation coefficients between variables.

Model	Parameters	Value
ANN	Hidden layer	2
	Hidden neuron	55
SVR	kernel	RBF
	C	3
	gamma	0.25
RF	Estimators	48
	Maximum depth of the tree	10
	Minimum samples for split	2
	Minimum samples of leaf node	1
GBDT	Estimators	80
	Learning rate	0.25
	Minimum samples for split	6
	Maximum depth of the tree	3
	Minimum samples of leaf node	5

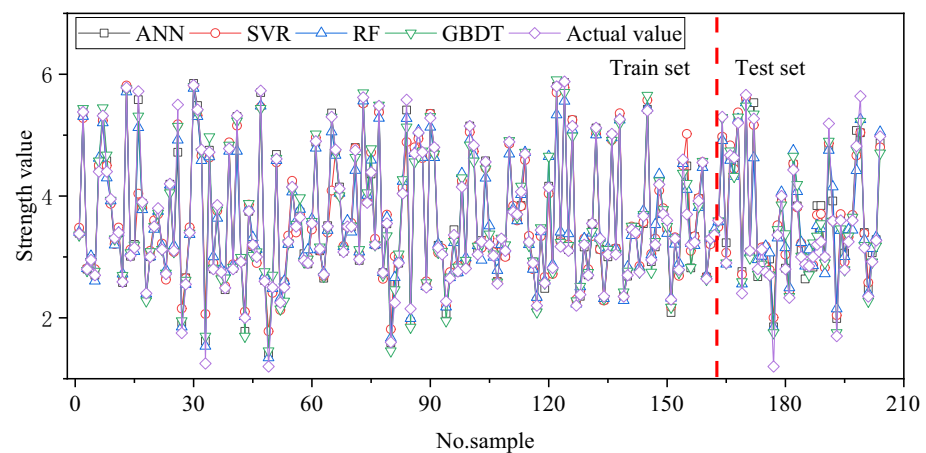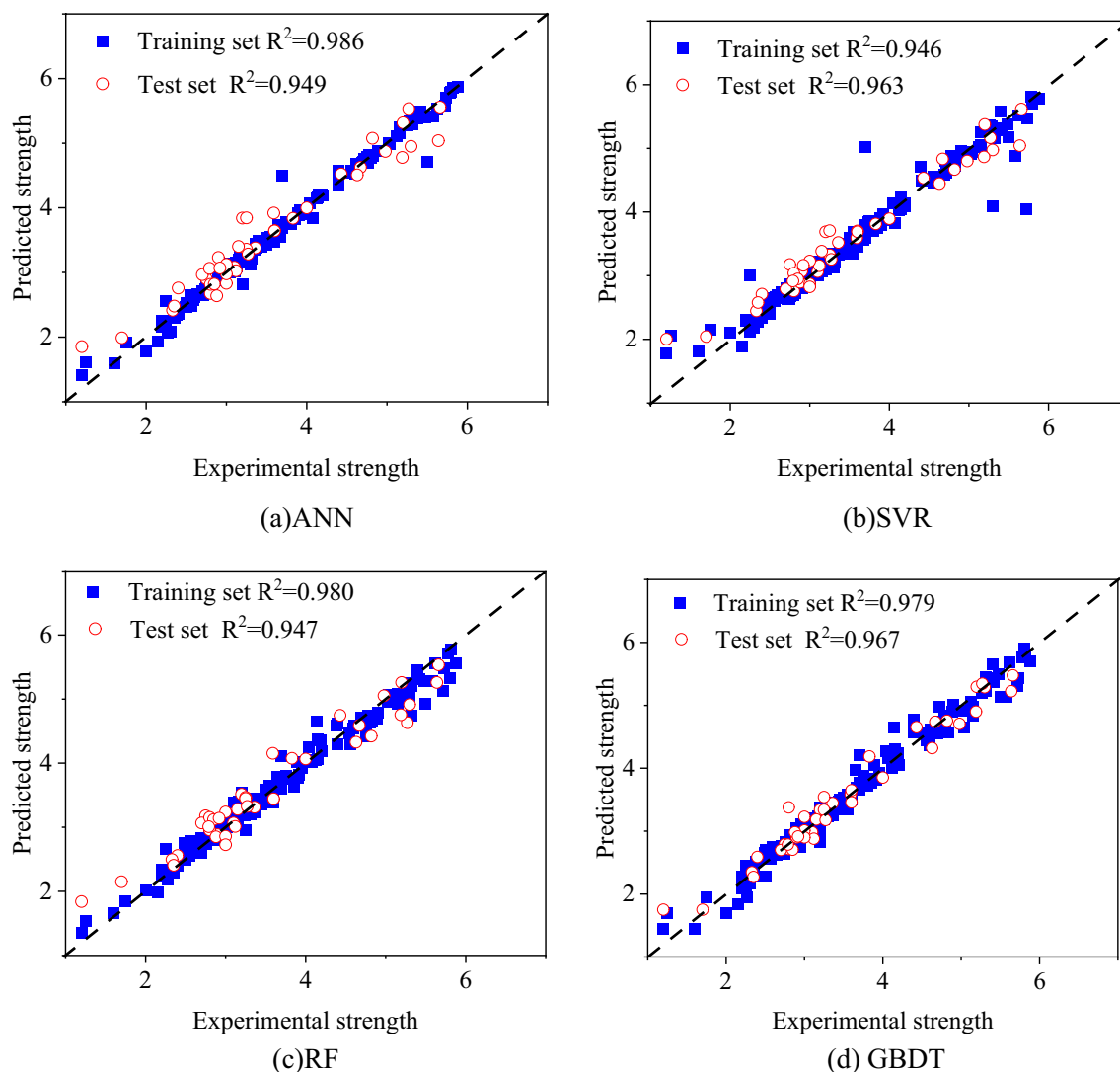**Table 2.** Parameters setting for different models.

Figure 7. Comparison between actual and predicted values of each sample for different models.

Figure 8. Correlation between predicted and actual values for different models.

| Evaluation index	ANN	SVR	RF	GBDT
R^2	0.949	0.963	0.947	0.967
MSE	0.071	0.066	0.083	0.041
RMSE	0.266	0.256	0.287	0.204
MAE	0.196	0.194	0.236	0.146
MAPE/%	6.716	6.948	7.841	4.856

Table 3. Evaluation index calculation of different model test results.

observed that, compared to other models, the GBDT model exhibits a more concentrated and closer-to-zero relative prediction error in the test dataset. The statistical analysis of the errors further underscores the positive predictive performance.
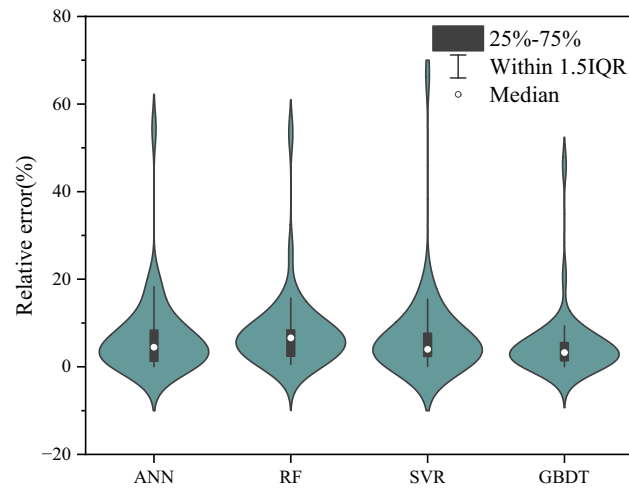
Feature importance

Feature importance analysis is the most commonly used method for interpreting model outputs. This analysis directly indicates the degree of influence of each feature on the final predictions. The greater the impact of a feature on the model's predictions, the more significant it is. Figure 11 presents the relative importance results of various features in predicting STS output using the GBDT model. Age is the most important feature for STS, which is as expected, as different ages exhibit significant differences in mechanical performance. Normalizing the relative importance of Age to 100%, the subsequent importance rankings are Cement and Metakaolin, with their
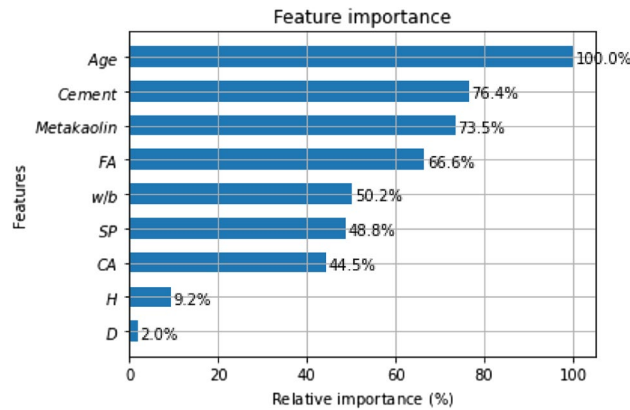
MSE	0.071	0.066	0.083	0.041
RMSE	0.266	0.256	0.287	0.204
MAE	0.196	0.194	0.236	0.146
MAPE/%	6.716	6.948	7.841	4.856
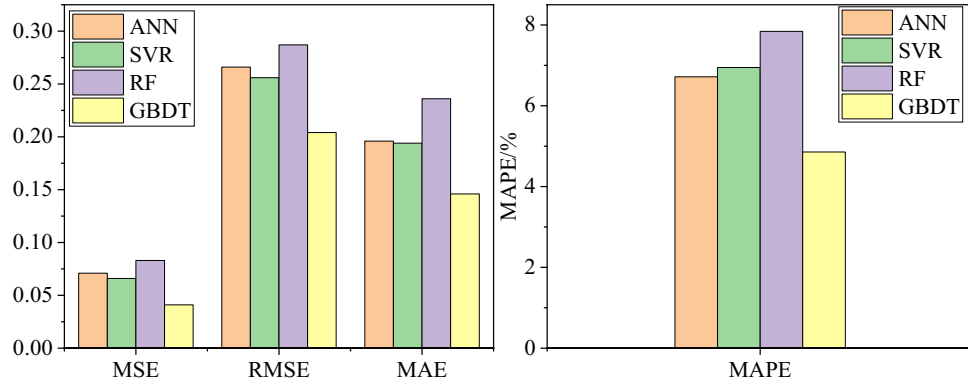
Figure 9. Histogram of evaluation indicator for test set.

Figure 10. The violin diagram for relative error percentage of different models.

Figure 11. Feature importance analysis.

8

importance being approximately three-quarters of the Age. Following that, in descending order, are FA, w/b, SP, and CA. The dimensions of the specimens are less important features, accounting for 9.2% for H and 2% for D.

Conclusions

This study proposes an STS prediction method involving concrete containing Metakaolin using individual and ensemble learning models. These machine learning models demonstrate good performance in reflecting the complex nonlinear relationships between input and output parameters in the prediction of STS for concrete containing Metakaolin. Based on the correlation coefficient between the predicted results and actual values, and considering other error metrics, the GBDT ensemble model exhibits the best prediction performance and is recommended as an intelligent method for STS prediction.

In the current dataset, the feature importance analysis based on the GBDT model shows that the most influential feature affecting STS is Age, followed by Cement, Metakaolin, FA, w/b, SP, and CA. The specimen dimensions have a relatively minor impact on STS. Feature importance analysis can provide guidance for obtaining the expected STS of Metakaolin concrete.

Although the machine learning methods developed in this study have achieved good prediction results, it should be noted that the research is conducted on a specific dataset. In the future, it is necessary to expand the dataset with more samples and search for samples that encompass a wider range of input parameters. Moreover, using Shapley Additive explanations analysis to further investigate the impact of these features on the output is also a focal point of future research.

Data availability

The datasets used during the current study available from the corresponding author on reasonable request.

References

1. Zhou, F. *et al.* Early shrinkage modeling of complex internally confined concrete based on capillary tension theory. *Buildings* **13**, 2201 (2023).
2. Zhou, F. *et al.* Moisture diffusion coefficient of concrete under different conditions. *Buildings* **13**, 2421 (2023).
3. Singh, A. *et al.* Utilization of antimony tailings in fiber-reinforced 3D printed concrete: A sustainable approach for construction materials. *Constr. Build. Mater.* **408**, 133689. https://doi.org/10.1016/j.conbuildmat.2023.133689 (2023).
4. Chen, L. *et al.* Recent developments on natural fiber concrete: A review of properties, sustainability, applications, barriers, and opportunities. *Dev. Built Environ.* **16**, 100255. https://doi.org/10.1016/j.dibe.2023.100255 (2023).
5. Farooq, F. *et al.* A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC). *Appl. Sci.* **10**, 7330 (2020).
6. Aslam, F. *et al.* Applications of gene expression programming for estimating compressive strength of high-strength concrete. *Adv. Civ. Eng.* **2020**, 8850535. https://doi.org/10.1155/2020/8850535 (2020).
7. Tang, H., Yang, Y., Li, H., Xiao, L. & Ge, Y. Effects of chloride salt erosion and freeze–thaw cycle on interface shear behavior between ordinary concrete and self-compacting concrete. *Structures* **56**, 104990. https://doi.org/10.1016/j.istruc.2023.104990 (2023).
8. He, H. *et al.* Employing novel N-doped graphene quantum dots to improve chloride binding of cement. *Constr. Build. Mater.* **401**, 132944. https://doi.org/10.1016/j.conbuildmat.2023.132944 (2023).
9. Jindal, B. B., Alomayri, T., Hasan, A. & Kaze, C. R. Geopolymer concrete with metakaolin for sustainability: A comprehensive review on raw material's properties, synthesis, performance, and potential application. *Environ. Sci. Pollut. Res.* **30**, 25299–25324. https://doi.org/10.1007/s11356-021-17849-w (2023).
10. Lenka, S. & Panda, K. C. Effect of metakaolin on the properties of conventional and self compacting concrete. *Adv. Concr. Constr.* **5**, 31–48 (2017).
11. Jin, M. *et al.* Multi-scale investigation on composition-structure of C-(A)-S-H with different Al/Si ratios under attack of decalcification action. *Cement Concr. Res.* **172**, 107251. https://doi.org/10.1016/j.cemconres.2023.107251 (2023).
12. Siddique, R. & Klaus, J. Influence of metakaolin on the properties of mortar and concrete: A review. *Appl. Clay Sci.* **43**, 392–400. https://doi.org/10.1016/j.clay.2008.11.007 (2009).
13. Xu, F., Wang, S., Li, T. & Li, Z. Effect of metakaolin on the mechanical properties and pore characteristics of fiber-reinforced tailing recycled aggregate concrete. *Structures* **35**, 15–25. https://doi.org/10.1016/j.istruc.2021.10.071 (2022).
14. Mohanraj, A. & Senthilkumar, V. Effect of metakaolin on the durability property of superabsorbent polymer blended self-compacting concrete. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **46**, 2099–2110. https://doi.org/10.1007/s40996-021-00660-5 (2022).
15. Lahoti, M., Narang, P., Tan, K. H. & Yang, E.-H. Mix design factors and strength prediction of metakaolin-based geopolymer. *Ceramics International* **43**, 11433–11441. https://doi.org/10.1016/j.ceramint.2017.06.006 (2017).
16. Bulbul, A. M. *et al.* In-depth analysis of cement-based material incorporating metakaolin using individual and ensemble machine learning approaches. *Materials* **15**, 7764 (2022).
17. Madandoust, R. & Mousavi, S. Y. Fresh and hardened properties of self-compacting concrete containing metakaolin. *Constr. Build. Mater.* **35**, 752–760. https://doi.org/10.1016/j.conbuildmat.2012.04.109 (2012).
18. Dinakar, P., Sahoo, P. K. & Sriram, G. Effect of metakaolin content on the properties of high strength concrete. *Int. J. Concr. Struct. Mater.* **7**, 215–223. https://doi.org/10.1007/s40069-013-0045-0 (2013).
19. Nafees, A. *et al.* Predictive modeling of mechanical properties of silica fume-based green concrete using artificial intelligence Approaches: MLPNN, ANFIS, and GEP. *Materials* **14**, 7531 (2021).
20. Song, H. *et al.* Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. *Constr. Build. Mater.* **308**, 125021. https://doi.org/10.1016/j.conbuildmat.2021.125021 (2021).
21. Wu, Y. & Zhou, Y. Prediction and feature analysis of punching shear strength of two-way reinforced concrete slabs using optimized machine learning algorithm and Shapley additive explanations. *Mech. Adv. Mater. Struct.* **30**, 3086–3096. https://doi.org/10.1080/15376494.2022.2068209 (2023).
22. Farooq, F. *et al.* A comparative study for the prediction of the compressive strength of self-compacting concrete modified with fly ash. *Materials* **14**, 4934 (2021).
23. Ahmad, A. *et al.* Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature. *Materials* **14**, 4222 (2021).

24. Wang, M., Yang, X. & Wang, W. Establishing a 3D aggregates database from X-ray CT scans of bulk concrete. *Constr. Build. Mater.* **315**, 125740. https://doi.org/10.1016/j.conbuildmat.2021.125740 (2022).

25. Ahmad, A. *et al.* Compressive strength prediction via gene expression programming (GEP) and artificial neural network (ANN) for concrete containing RCA. *Buildings* **11**, 324 (2021).

26. Wu, Y., Wang, Y., Li, D. & Zhang, J. Two-step detection of concrete internal condition using array ultrasound and deep learning. *Ndt&E Int.* **139**, 102945. https://doi.org/10.1016/j.ndteint.2023.102945 (2023).

27. Farooq, F., Ahmed, W., Akbar, A., Aslam, F. & Alyousef, R. Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *J. Clean. Prod.* **292**, 126032. https://doi.org/10.1016/j.jclepro.2021.126032 (2021).

28. Wang, Z., Wang, Q., Jia, C. & Bai, J. Thermal evolution of chemical structure and mechanism of oil sands bitumen. *Energy* **244**, 123190. https://doi.org/10.1016/j.energy.2022.123190 (2022).

29. Yin, L. *et al.* Study on the thermospheric density distribution pattern during geomagnetic activity. *Appl. Sci.* **13**, 5564 (2023).

30. Huang, H., Yuan, Y., Zhang, W. & Li, M. Seismic behavior of a replaceable artificial controllable plastic hinge for precast concrete beam-column joint. *Eng. Struct.* **245**, 112848. https://doi.org/10.1016/j.engstruct.2021.112848 (2021).

31. Huang, H., Yuan, Y., Zhang, W. & Zhu, L. Property assessment of high-performance concrete containing three types of fibers. *Int. J. Concr. Struct. Mater.* **15**, 39. https://doi.org/10.1186/s40069-021-00476-7 (2021).

32. Wu, Y., Zhang, J., Gao, C. & Xu, J. Internal defect detection quantification and three-dimensional localization based on impact echo and classification learning model. *Measurement* **218**, 113153. https://doi.org/10.1016/j.measurement.2023.113153 (2023).

33. Liu, C. *et al.* Development of crack and damage in shield tunnel lining under seismic loading: Refined 3D finite element modeling and analyses. *Thin-Walled Struct.* **185**, 110647. https://doi.org/10.1016/j.tws.2023.110647 (2023).

34. Yao, W. *et al.* Experimental and theoretical investigation of coupled damage of rock under combined disturbance. *Int. J. Rock Mech. Min. Sci.* **164**, 105355. https://doi.org/10.1016/j.ijrmms.2023.105355 (2023).

35. Rajender, A. & Samanta, A. K. Compressive strength prediction of metakaolin based high-performance concrete with machine learning. *Mater. Today: Proc.* https://doi.org/10.1016/j.matpr.2023.03.522 (2023).

36. Biswal, U. S., Mishra, M., Singh, M. K. & Pasla, D. Experimental investigation and comparative machine learning prediction of the compressive strength of recycled aggregate concrete incorporated with fly ash, GGBS, and metakaolin. *Innov. Infrastruct. Solut.* **7**, 242. https://doi.org/10.1007/s41062-022-00844-6 (2022).

37. Asteris, P. G. *et al.* Revealing the nature of metakaolin-based concrete materials using artificial intelligence techniques. *Constr. Build. Mater.* **322**, 126500. https://doi.org/10.1016/j.conbuildmat.2022.126500 (2022).

38. Abdulalim Alabdullah, A. *et al.* Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Constr. Build. Mater.* **345**, 128296. https://doi.org/10.1016/j.conbuildmat.2022.128296 (2022).

39. Hu, Y. *et al.* Strength evaluation sustainable concrete with waste ingredients at elevated temperature by employing interpretable algorithms: Optimization and hyper tuning. *Mater. Today Commun.* **36**, 106467. https://doi.org/10.1016/j.mtcomm.2023.106467 (2023).

40. Xu, G. *et al.* Evaluation of properties of bio-composite with interpretable machine learning approaches: Optimization and hyper tuning. *J. Mater. Res. Technol.* **25**, 1421–1446. https://doi.org/10.1016/j.jmrt.2023.06.007 (2023).

41. Amin, M. N. *et al.* Prediction of sustainable concrete utilizing rice husk ash (RHA) as supplementary cementitious material (SCM): Optimization and hyper-tuning. *J. Mater. Res. Technol.* **25**, 1495–1536. https://doi.org/10.1016/j.jmrt.2023.06.006 (2023).

42. Jiao, H. *et al.* A novel approach in forecasting compressive strength of concrete with carbon nanotubes as nanomaterials. *Mater. Today Commun.* **35**, 106335. https://doi.org/10.1016/j.mtcomm.2023.106335 (2023).

43. Wu, Y. & Zhou, Y. Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. *Constr. Build. Mater.* **330**, 127298. https://doi.org/10.1016/j.conbuildmat.2022.127298 (2022).

44. Zheng, W. *et al.* Sustainable predictive model of concrete utilizing waste ingredient: Individual alogrithms with optimized ensemble approaches. *Mater. Today Commun.* **35**, 105901. https://doi.org/10.1016/j.mtcomm.2023.105901 (2023).

45. Ullah, H. S., Khushnood, R. A., Ahmad, J. & Farooq, F. Predictive modelling of sustainable lightweight foamed concrete using machine learning novel approach. *J. Build. Eng.* **56**, 104746. https://doi.org/10.1016/j.jobe.2022.104746 (2022).

46. Ullah, H. S. *et al.* Prediction of compressive strength of sustainable foam concrete using individual and ensemble machine learning approaches. *Materials* **15**, 3166 (2022).

47. Wu, Y. & Zhou, Y. Splitting tensile strength prediction of sustainable high-performance concrete using machine learning techniques. *Environ. Sci. Pollut. Res.* https://doi.org/10.1007/s11356-022-22048-2 (2022).

48. Ray, S., Rahman, M. M., Haque, M., Hasan, M. W. & Alam, M. M. Performance evaluation of SVM and GBM in predicting compressive and splitting tensile strength of concrete prepared with ceramic waste and nylon fiber. *J. King Saud Univ. Eng. Sci.* **35**, 92–100. https://doi.org/10.1016/j.jksues.2021.02.009 (2023).

49. Shah, H. A. *et al.* Application of machine learning techniques for predicting compressive, splitting tensile, and flexural strengths of concrete with metakaolin. *Materials* **15**, 5435 (2022).

50. Moradi, N. *et al.* Predicting the compressive strength of concrete containing binary supplementary cementitious material using machine learning approach. *Materials* **15**, 5336 (2022).

51. Huang, J., Zhou, M., Yuan, H., Sabri, M. M. & Li, X. Prediction of the compressive strength for cement-based materials with metakaolin based on the hybrid machine learning method. *Materials* **15**, 3500 (2022).

52. Lee, S.-C. Prediction of concrete strength using artificial neural networks. *Eng. Struct.* **25**, 849–857. https://doi.org/10.1016/S0141-0296(03)00004-X (2003).

53. Wu, Y. & Li, S. Damage degree evaluation of masonry using optimized SVM-based acoustic emission monitoring and rate process theory. *Measurement* **190**, 110729. https://doi.org/10.1016/j.measurement.2022.110729 (2022).

54. Migallón, V., Penadés, H., Penadés, J. & Tenza-Abril, A. J. A machine learning approach to prediction of the compressive strength of segregated lightweight aggregate concretes using ultrasonic pulse velocity. *Appl. Sci.* **13**, 1953 (2023).

55. Zhang, K., Zhang, K. & Bao, R. Machine learning models to predict the residual tensile strength of glass fiber reinforced polymer bars in strong alkaline environments: A comparative study. *J. Build. Eng.* **73**, 106817. https://doi.org/10.1016/j.jobe.2023.106817 (2023).

56. Shafiq, N., Nuruddin, M. F., Khan, S. U. & Ayub, T. Calcined kaolin as cement replacing material and its use in high strength concrete. *Constr. Build. Mater.* **81**, 313–323. https://doi.org/10.1016/j.conbuildmat.2015.02.050 (2015).

57. Muduli, R. & Mukharjee, B. B. Effect of incorporation of metakaolin and recycled coarse aggregate on properties of concrete. *J. Clean. Prod.* **209**, 398–414. https://doi.org/10.1016/j.jclepro.2018.10.221 (2019).

58. Lenkaa, S. & Panda, K. C. Effect of metakaolin on the properties of conventional and self compacting concrete. *Adv. Concr. Constr.* **5**, 31–48 (2017).

59. Kannan, V. & Ganesan, K. Evaluation of mechanical and permeability related properties of self compacting concrete containing metakaolin. *Sci. Res. Essays* **7**, 4081–4091 (2012).

60. Güneyisi, E., Gesoğlu, M., Qays, M. A., Mermerdaş, K. & İpek, S. Fracture properties of high strength metakaolin and silica fume concretes. *In Proceedings of the 3rd International Conference on Chemical, Civil and Environmental Engineering (CCEE-2016), Antalya, Turkey, 20–21 April 2016.* (2016).

61. Al-Oran, A. A. A., Safiee, N. A. & Nasir, N. A. M. Fresh and hardened properties of self-compacting concrete using metakaolin and GGBS as cement replacement. *Eur. J. Environ. Civ. Eng.* **26**, 379–392. https://doi.org/10.1080/19648189.2019.1663268 (2022).
62. Kavitha, O. R., Shanthi, V. M., Prince Arulraj, G. & Sivakumar, P. Fresh, micro- and macrolevel studies of metakaolin blended self-compacting concrete. *Appl. Clay Sci.* **114**, 370–374. https://doi.org/10.1016/j.clay.2015.06.024 (2015).
63. Guneyisi, E., Gesoglu, M. & Mermerdas, K. Improving strength, drying shrinkage, and pore structure of concrete using metakaolin. *Mater. Struct.* **41**, 937–949. https://doi.org/10.1617/s11527-007-9296-z (2008).
64. Sharbatdar, M., Abadi, M. A. R. & Fakharian, P. F. Improving the properties of self-compacted concrete with using combined silica fume and metakaolin. *Period. Polytech. Civ. Eng.* **64**, 535–544 (2020).
65. Shehab El-Din, H. K., Eisa, A. S., Abdel Aziz, B. H. & Ibrahim, A. Mechanical performance of high strength concrete made from high volume of Metakaolin and hybrid fibers. *Constr. Build. Mater.* **140**, 203–209. https://doi.org/10.1016/j.conbuildmat.2017.02.118 (2017).
66. Rashad, A. M. A preliminary study on the effect of fine aggregate replacement with metakaolin on strength and abrasion resistance of concrete. *Constr. Build. Mater.* **44**, 487–495. https://doi.org/10.1016/j.conbuildmat.2013.03.038 (2013).
67. Kannan, V. Strength and durability performance of self compacting concrete containing self-combusted rice husk ash and metakaolin. *Constr. Build. Mater.* **160**, 169–179. https://doi.org/10.1016/j.conbuildmat.2017.11.043 (2018).
68. Zoe, Y., Hanif, I. M., Adzmier, H. M., Eyzati, H. H. & Noor Syuhaili, M. R. Strength of self-compacting concrete containing metakaolin and nylon fiber. *IOP Conf. Series: Earth Environ. Sci.* **498**, 012047. https://doi.org/10.1088/1755-1315/498/1/012047 (2020).
69. John, N. Strength properties of metakaolin admixed concrete. *Int. J. Sci. Res. Publ.* **3**, 2250–3153 (2013).
70. Guneyisi, E., Gesoglu, M., Karaoglu, S. & Mermerdas, K. Strength, permeability and shrinkage cracking of silica fume and metakaolin concretes. *Constr Build Mater* **34**, 120–130. https://doi.org/10.1016/j.conbuildmat.2012.02.017 (2012).
71. Feng, D.-C., Wang, W.-J., Mangalathu, S., Hu, G. & Wu, T. Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements. *Eng Struct* **235**, 111979. https://doi.org/10.1016/j.engstruct.2021.111979 (2021).

Author contributions

Q.L.: methodology, writing-original draft preparation; G.R.: software, H.W.: formal analysis; Q.X.: formal analysis; J.Z.: validation; H.W.: resources, Y.D.: methodology, supervision. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.D.

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.