

INTELLIGENT PREDICTION OF THE FROST RESISTANCE OF HIGH-PERFORMANCE CONCRETE: A MACHINE LEARNING METHOD

Jian ZHANG¹, Yuan CAO², Linyu XIA³, Desen ZHANG¹, Wen XU^{2*}, Yang LIU^{4#}

¹*School of Economics and Management, Wuhan University, Wuhan 430072, PR China*

²*School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*

³*Qilu Institute of Technology, Jinan, Shandong 250200, China*

⁴*ZhongNan Hospital of Wuhan University, Wuhan University, Wuhan 430071, PR China*

Received 9 June 2022; accepted 19 March 2023

Abstract. Frost resistance in very cold areas is an important engineering issue for the durability of concrete, and the efficient and accurate prediction of the frost resistance of concrete is a crucial basis for determining reasonable design mix proportions. For a quick and accurate prediction of the frost resistance of concrete, a Bayesian optimization (BO)-random forest (RF) approach was used to establish a frost resistance prediction model that consists of three phases. A case study of a key national engineering project results show that (1) the RF can be used to effectively screen the factors that influence concrete frost resistance. (2) R^2 of BO-RF for the training set and the test set are 0.967 and 0.959, respectively, which are better than those of the other algorithms. (3) Using the test data from the first section of the project for prediction, good results are obtained for the second section. The proposed BO-RF hybrid algorithm can accurately and quickly predict the frost resistance of concrete, and provide a reference basis for intelligent prediction of concrete durability.

Keywords: frost resistance, durability of concrete, random forest, Bayesian optimization, mix proportion.

Introduction

With continuous economic development, new infrastructure has become an important aspect of policy and local high-quality development in China. Due to its advantages, high-performance concrete is increasingly used in many major infrastructure construction projects, such as the “Belt and Road Initiative”, the Northeast revitalization project, the development of the Guangdong-Hong Kong-Macao Greater Bay Area, and the national three-dimensional comprehensive transportation planning and construction initiative (Amran et al., 2022). Due to the continuous construction of major projects, the number of special and complex projects, such as projects in cold environments and underground environments, which include the Zhonghai Heshan Grand View Project, the Taishan Station Snow Engineering Construction Project, and the Sichuan Tibet Railway, is increasing. Due to the particularity and complexity of the environments of these projects, the frost resistance of the concrete used must be higher than average. Frost resistance is an important in-

dicator of the durability of concrete. Improving the frost resistance of concrete is important for ensuring structural safety and reducing structural damage. Additionally, using the appropriate concrete can reduce the consumption of resources and energy and the environmental impact during reconstruction or repair.

Concrete has cost and durability advantages compared with other construction materials; it is one of the most widely used building materials in China (Boukhatem et al., 2011). Durability is one of the most important properties of concrete during its use. However, with the broad application of concrete in engineering, the deterioration of concrete structures and the damage caused by insufficient durability have become increasingly prominent (Wu et al., 2022b). Given the complex, severe and cold working conditions and the effects of salt erosion, freezing, and melting, the frost resistance of concrete structures changes with the ratios of cement, water, aggregates, admixtures and other raw materials used. Such conditions greatly

*Corresponding author. E-mail: xuwen0227@hust.edu.cn

#Corresponding author. E-mail: dabailiu@whu.edu.cn

impact the safety and service life of concrete structures (Yazıcı, 2008). Therefore, accurately and rapidly forecasting the frost resistance of concrete has important application value for civil engineering (DeRousseau et al., 2018).

The frost resistance of concrete is related to many factors, such as the internal pore structure, bubble content, water saturation degree, freezing age, and concrete strength. Yang et al. (2011) studied the permeability and frost resistance of aerated concrete for the Qingdao Bay Bridge. The results showed that proper air entrainment can greatly improve the frost resistance and impermeability of concrete. Chen et al. (2022) studied the influence of different curing conditions and construction methods on the frost resistance of concrete surfaces. The results showed that the frost resistance of concrete can be improved by curing before the initial stages of freezing and salt application to ensure that the concrete is not in a dry shrinkage state and to avoid the loss of air through the concrete surface caused by surface disruption (Chen et al., 2022). In addition, internal factors, such as the water binder ratio, cement dosage, admixture, mineral admixture, and aggregate content, and external factors also influence the frost resistance of concrete. Environmental factors indirectly influence frost resistance, while internal factors are directly related to frost resistance. However, there are few studies that use relevant mathematical models to comprehensively evaluate the relationship between frost resistance and various constituent materials. Scholars have studied the relationships among various constituent materials through experiments, but those studies had certain shortcomings, including long study periods and heavy workloads (Gao et al., 2021). Emerging machine learning technology provides an opportunity to accurately predict the performance of concrete. In this paper, a prediction model of frost resistance is presented based on the random forest (RF) method.

RFs provide good prediction performance and can rank the importance of influential factors (Wu et al., 2022a). Aulia et al. (2019) used an RF to investigate automatic production history matching in reservoir engineering and ranked the importance of the input parameters. Liu et al. (2021b) used an RF model to rank the importance of relevant factors in engineering specialty selection. Ding et al. (2021) used an RF to rank the importance of the characteristic variables of artificial terraces. Marcos-Pasero et al. (2021) established an RF model to evaluate and rank the importance of variables that affect childhood obesity. Wei et al. (2021) used an RF to rank the importance of the input features of shale gas production and thereby improved the interpretability of the modeling results. The importance of factors can be evaluated with RFs to control the main influential factors and effectively reveal the underlying mechanisms (Zhou et al., 2016).

To obtain more accurate prediction results, a prediction model for concrete frost resistance is proposed based on the RF algorithm. The main research questions are as follows: (1) How can a complex relationship be established

between the input and output variables based on an RF model? (2) How can the importance of and correlations among the factors affecting the mix proportion of the raw materials in concrete be evaluated? The RF method is used to screen the important factors that affect the frost resistance of concrete, and the dynamic modulus of elasticity is used as an index to accurately predict frost resistance. The main contributions of this research are as follows: (a) An effective and accurate RF prediction model is established, providing a solid basis to design the concrete mix proportion and achieve frost resistance; (b) The importance of the variables that influence frost resistance are revealed, and the accuracy of the proposed method is verified through correlation analysis; and (c) The prediction performance of the RF is compared with that of a support vector machine (SVM) model, a back propagation (BP) model, and a gradient boost decision tree (GBDT).

Overall, the structure of the paper is as follows. In the third section, basic theoretical knowledge is presented. In the fourth section, the execution process for the RF prediction model is introduced. In the fifth section, specific cases are analyzed. In the sixth section, a discussion and a comparison with the other methods are provided to highlight the superiority of the proposed method. The seventh section summarizes the paper, and future research directions are identified.

1. Literature review

1.1. Study of the frost resistance of concrete with traditional methods

Researchers mainly use experimental methods, formula-based analysis approaches and statistical models to explore the frost resistance of concrete. Luan et al. (2020) and Wu et al. (2022b) found that the addition of mixed recycled aggregate and recycled cement enabled recycled concrete to resist an increased number of freeze–thaw cycles, thereby improving its frost resistance. Yuan et al. (2019) experimentally simulated the effects of freeze–thaw cycles and a sulfate environment on the durability of recycled concrete under the coupled effect of different aggregate replacement rates. However, the test method requires considerable human and material resources, and it may negatively influence the environment. Moreover, an insufficient number of samples and selection errors can affect the test results. Dvorkin (2019) obtained a formula to predict frost resistance based on a theoretical analysis of the structural parameters of concrete. Smith et al. (2018) used a developed limit state function to demonstrate how to quantitatively select design variables that limit frost resistance. However, this method requires tedious calculations and a heavy workload and is characterized by low efficiency. Ashraf et al. (2018) developed a probability model to analyze the freeze–thaw performance of concrete. Keleştemur et al. (2014) used statistical methods to study the impact of marble dust and glass fiber on cement mortar for different numbers of freeze–thaw cycles. In summary, most

scholars have used traditional methods to study the changes in and effects of frost resistance. However, due to systematic errors, the randomness of measurement data and other factors, the level of discreteness of the experimental observation data is high, and traditional methods often fail to provide reliable results (Koya et al., 2022). In addition, previous probabilistic evaluations and predictions of the frost resistance of concrete mixtures have often only considered a limited amount of experimental data when establishing the “best fit” function. Therefore, the use of traditional methods to forecast the frost resistance of concrete has some limitations.

1.2. Studies of concrete using artificial intelligence

In recent decades, many machine learning algorithms have been applied to predict the compressive strength, durability, and service life of concrete, and these methods include artificial neural networks (ANNs) (Kewalramani & Gupta, 2006) and SVMs (Lakshmanprabu et al., 2019; Liu et al., 2020). Belalia Douma et al. (2016) predicted the performance of fly ash self-compacting concrete using an ANN method. Özcan et al. (2009) used an ANN and fuzzy logic to predict the long-term compressive strength of silica fume concrete. Nguyen et al. (2021) studied the relationships between different input variables and the compressive strength of ordinary concrete and high-performance concrete based on an ANN prediction model. Azimi-Pour et al. (2020) proposed appropriate linear and nonlinear SVM models with different cores to predict the compressive strength of self-compacting concrete with a high fly ash content. Sonebi et al. (2016) studied the feasibility of using an SVM to predict the freshness of self-compacting concrete. Cheng and Hoang (2016) proposed an adaptive fuzzy least-squares support vector machine inference model to predict the compressive strength of rubber concrete. These studies showed that models based on machine learning algorithms can provide better prediction results than traditional models and that if the base predictor variables are correctly selected, the models based on integrated algorithms yield the highest accuracy (Cai et al., 2020).

The machine learning algorithms currently being used are mainly classical methods, such as ANNs and SVMs, that can obtain better prediction results than traditional methods, although they still have several limitations. For example, the structure of ANNs must be determined based on experience, and ANNs heavily rely on samples; moreover, SVMs struggle to support large-scale sample training, rely heavily on typical samples and are inefficient (Liu & Zhang, 2020). Therefore, superior methods are needed to predict the frost resistance of concrete. As an emerging machine learning ensemble approach, the RF algorithm can solve complex nonparametric and nonlinear classification problems and reduce the complexity of calculations under the premise of improving accuracy. This approach requires fewer parameters than traditional methods and provides strong generalization ability, strong antioverfit-

ting ability and other advantages (Zhang & Min, 2016). Therefore, in recent years, the use of RFs for forecasting and ranking the importance of influential factors has become common in many industries (Lundström & Verikas, 2013). Li et al. (2020) used an RF model for the crime analysis. Chun et al. (2020) used an RF to evaluate the internal loss of reinforced concrete. On the basis of existing concrete data, Nilsen et al. (2019) obtained high-accuracy predictions of the coefficient of thermal expansion (CTE) and the relative elastic modulus of concrete with an RF. Benedet et al. (2021) used an RF to quickly predict soil fertility. Lee et al. (2021) developed an RF model to predict pediatric mortality within 72 hours of ICU admission. Niu et al. (2020) applied an RF to short-term photovoltaic power generation forecasting. RFs are superior to many other prediction methods, so applying an RF to forecast the frost resistance of concrete is worthwhile.

Based on the above literature studies, it can be found that concrete durability is a very important indicator. Currently, experiments and numerical simulation are used in more researches, while machine learn-based researches are few. Among the prediction models adopted in existing research, RF has better prediction performance. Therefore, an RF frost resistance prediction model is developed in this paper considering the ratio of raw materials used. According to this model, the frost resistance of concrete is predicted and analyzed, and the corresponding influential factors are ranked based on importance. The prediction results are compared with those of three other prediction models to verify the reliability of the proposed RF model.

2. Preliminary information

2.1. Bayesian optimization

If appropriate hyperparameters are set, the predefined loss function can be simplified, thus improving the prediction or classification accuracy for given independent data (Chen et al., 2023b). The search for the best combination of hyperparameters requires experienced insight, which can be difficult to obtain. Two-hyperparameter optimization in a prediction model is used as an example. The parameter search processes, such as grid searches, random searches and Bayesian optimization, are shown in Figure 1. The black dots represent unsearched points, the blue dots represent searched points, and the red arrows represent the search directions. Suppose hyperparameter 1 is associated with n selectable items and hyperparameter 2 is associated with m selectable items. Figure 1a shows the grid search process. A grid search should traverse all nodes in the grid plane, which may require a long training time and consume excessive resources. Figure 1b shows the random search process. In the random search method, a general search range is first determined; then, the points within this range are randomly compared, and an upper limit is set regarding the number of iterations. If the optimal value is found within the set number of iterations, the search process is terminated; otherwise, the number of iterations

is increased, and the iteration process restarts (Lin & Liu, 2006). However, this approach can easily be influenced by local optimality. Figure 1c shows the search process of Bayesian optimization. The Bayesian optimization algorithm searches the superparameter candidates according to historical observations to determine the next evaluation location to obtain the global optimal solution at the fastest speed possible (Jones et al., 1998).

In Bayesian optimization, Bayes’ theorem is used in the optimization process (Puga et al., 2015):

$$p(f|D_{1:t}) = \frac{p(D_{1:t}|f)p(f)}{p(D_{1:t})}, \tag{1}$$

where f represents either the unknown objective function (or the parameters in the parametric model), $D_{1:t} = \{(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t)\}$ represents the observed set, x_t is the decision vector, $y_t = f(x_t) + \varepsilon_t$ is the observed value, ε_t is the observed error, $P(D_{1:t} | f)$ is the likelihood distribution of y considering the observation error (this parameter is also known as the noise), $p(f)$ represents the prior probability distribution of f , that is, the assumption regarding the state of the unknown objective function, and $P(D_{1:t})$ is the marginal likelihood distribution, or “evidence”, of marginal f . Since the marginal likelihood is determined from the product and integral of probability density functions, it is often difficult to obtain an explicit analytical expression. This marginal likelihood is mainly used to optimize hyperparameters in Bayesian optimization. $P(f | D_{1:t})$ represents the posterior probability distribution of f .

The Bayesian optimization framework consists of two primary parts: the probability in the agent model and the acquisition function. The probabilistic proxy model is used to replace the unknown objective function. The model begins with the assumed prior function and then modifies it by iteratively increasing the amount of information to provide the best possible representation of the unknown objective function in the proxy model. The sampling function is constructed according to the posterior probability distribution, the next-most-likely evaluation point is selected by maximizing the acquisition function, and the optimal sample point is finally selected to minimize the value of the objective function (Nguyen et al., 2018).

2.2. Regression and prediction with the random forest model

RFs use a randomly split feature set to construct decision trees based on the bagging algorithm (Q. Wu et al., 2014; X. Wu et al., 2023), which is used to randomly replace the original dataset and form multiple sample sets. First, based on the bagging algorithm, training samples are randomly selected, and the original dataset is randomly replaced to generate multiple training data subsets. Small sample sizes in all categories are used to generate training sets in each step, and data are randomly extracted from large samples in categories and combined with small samples to form training sets. In this way, many training sets and training models can be obtained after repeated iterations. Therefore, the RF algorithm can solve problems related to an unbalanced data distribution. The data distribution determines the accuracy of the model, and the ability of the RF model to solve the data imbalance problem ensures relatively accurate predictions. The probability of each sample not being extracted is $p = (1 - \frac{1}{n})^n$ and unused samples are called out-of-bag (OOB) data, which can later be applied to calculate the model generalization error (Liu et al., 2021a). The RF generalization error can be expressed as

$$E^* = P_{X,Y} (M(X,Y) < 0). \tag{2}$$

The subscripts X and Y indicate that the probability of inclusion in the X and Y spaces, respectively. Thirty-six percent of the samples in the original dataset are not included in each new dataset. Then, a decision tree ($h_i(x, \theta_k), k = 1, 2, 3, \dots, K$) is established for every sample set. As the model is generated, the inner nodes of each decision tree are stochastically divided based on certain characteristics. Multiple split decision trees form an RF in the classification and regression tree (CART) algorithm. The trees then output the results. For regression algorithms, the final forecasting result is the average value of the output results of all decision trees. The final regression result is given in Eqn (3) (Rayal et al., 2017):

$$\bar{H}(x) = \frac{1}{k} \sum_{i=1}^k (h_i(x= Y)), \tag{3}$$

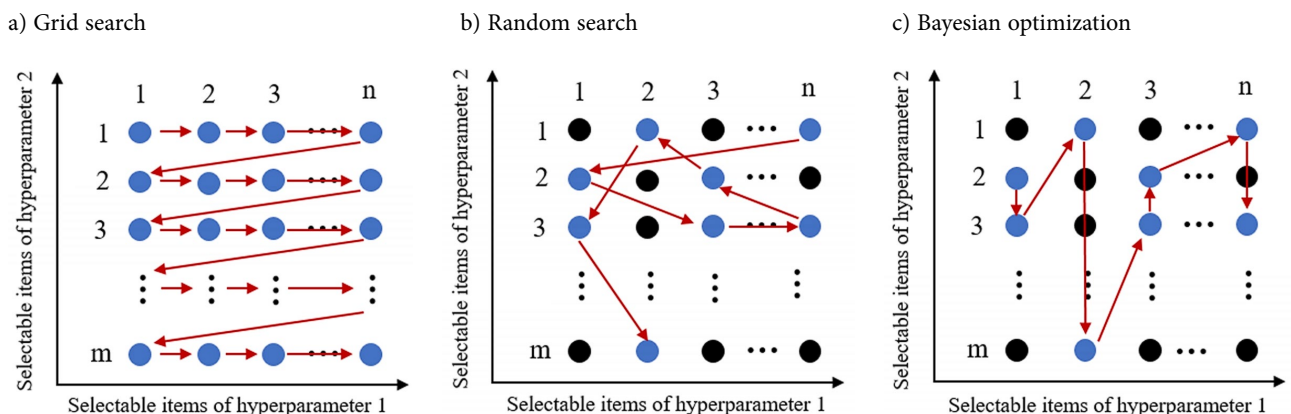


Figure 1. Hyperparameter searches based on three methods

where $\bar{H}(x)$ represents the prediction result, k is the number of decision trees in the RF model, h_i is a single decision tree, and Y is the output variable (target variable).

2.3. Importance evaluation of the random forest

The characteristics of indices can be ranked by using the RF model (Wu et al., 2022b). By randomly changing the order of the input variables of each decision tree, that is, by adding noise or interference, the effect of each input variable is evaluated by calculating the OOB mean square error. The average error of all decision trees represents the importance of the input variables. Compared with other methods, this method not only separately considers the degree of influence of each variable but also estimates the relationships among the other variables. The importance of the input variable X is expressed by Eqn (4):

$$VIM_i(MSE) = \frac{\left(\frac{1}{b} \sum_{j=1}^b (MSE_j - MSE_{ij}) \right)}{S_E}, \quad (1 \leq i \leq p), \quad (4)$$

where MSE_j represents the mean squared residual for the j th sample and S_E represents the standard deviation. The larger the value of VIM_i is, the greater the influence of the characteristic variable on the output. If the relationship between the characteristic variable and the prediction target is strong, the forecasting precision of the decision tree will decrease after random arrangement. Additionally, the stronger the relevance of the variable is, the greater the decrease observed.

Once the importance of each feature has been obtained, a reasonable ranking is established; that is, an important analysis based on the RF is realized. The selection of the target seeds is mainly based on the literature, rel-

evant specifications and engineering experience. Based on standard concrete structures (GB/T 50476-2019) (People's Republic of China, 2019), a large number of studies and practical project experience, 6 factors that influence the relative dynamic elastic modulus of concrete are selected.

3. Methodology

The flow chart of our proposed model for RF-based concrete frost resistance prediction is depicted in Figure 2. The model includes three main steps: establishing a sample dataset, establishing a prediction model, and evaluating the model.

3.1. Dataset acquisition

Many factors, such as the type of cement, additives used, and water-binder ratio, affect the frost resistance of concrete. The relationships among these factors and the RDEM are nonlinear and greatly impact the performance of concrete (Yaseen et al., 2018). In this paper, the commonly used parameters that influence the frost resistance of concrete are selected (Łaźniewska-Piekarczyk, 2013). The output index is the RDEM, which is used to construct an initial index system. Then, related experiments are conducted, data are collected, and these data are used to establish an original dataset.

To increase the reliability of the prediction results, a primitive sample is generally divided into two parts: a training set used to train the model and a test set used to test the prediction effect. The primitive sample is randomly and equally assigned to K , and $K - 1$ is the training sample set of the RF model; the other sample set is the test sample set of the RF model used to test the prediction effect of the model.

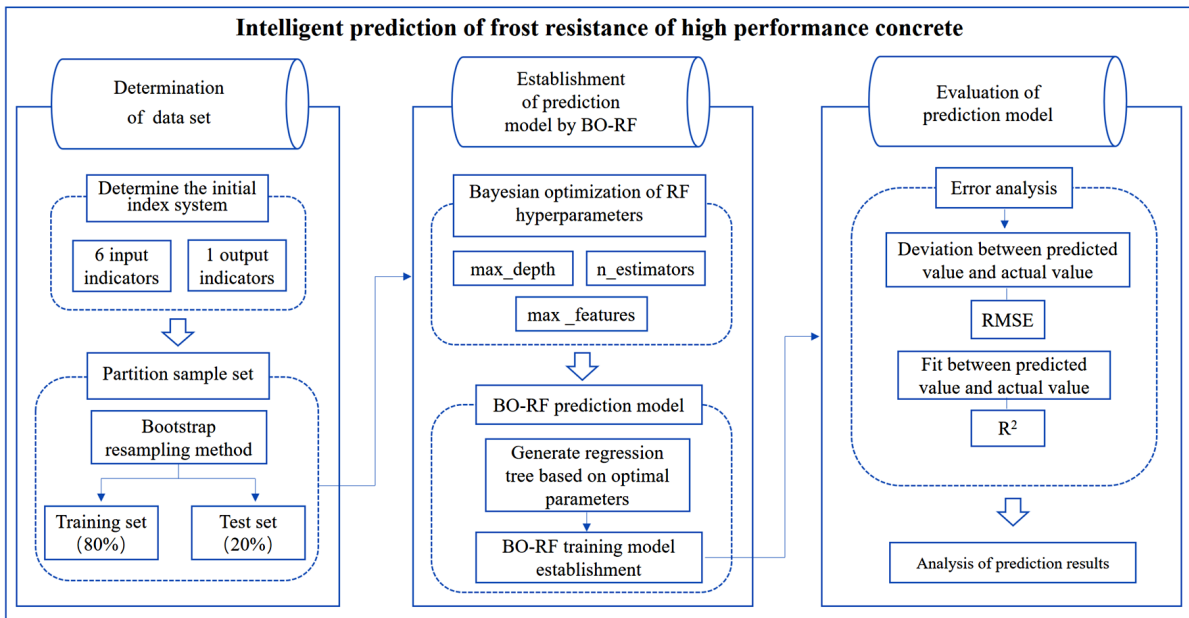


Figure 2. Flowchart of the frost resistance regression model based on an RF

3.2. Selection of RF hyperparameters

The setting of the RF parameters directly affects the regression fitting performance of the model. Therefore, when using the RF regression algorithm to train samples, the important parameters of the model must be set (Chen et al., 2023a). The parameters that must be set mainly include the parameters of the bagging framework and the parameters of the CART framework (Ala'raj & Abbod, 2016). The most important parameter in the RF bagging framework is the maximum number of iterations $n_estimators$. The value of this parameter has a direct impact on the prediction performance of the RF model. If this number is too small, the fitting performance will be poor, and a number that is too large will increase the computing cost. The important parameters of the CART algorithm include the maximum number of features $max_features$ and the maximum depth max_depth , which affect the establishment of the decision tree model and thereby affect the regression fitting performance of the model. The larger the $max_features$ value is, the more information the model can learn, but the algorithm speed will also decrease. The max_depth value is related to the complexity of the decision tree. After determining the parameters of the RF model, training and testing sample sets for the RF regression model can be established according to the RF algorithm.

3.3. Bayesian optimization of hyperparameters

The values of hyperparameters in the modeling process directly affect the prediction effect of the model. It is necessary to optimize these hyperparameters. In extreme cases, the model relearns all hyperparameters during each iteration based on the present data. Although this method can ensure modeling accuracy, the learning of hyperparameters requires a high computational load, and it is often inefficient. Thus, Bayesian optimization methods, such as the Bayes optimization of hyperparameters (Martinez-Cantin, 2014), usually involve relearning hyperparameters after multiple iterations, such as 100 iterations. The RF hyperparameters optimized with the Bayesian method are shown in Table 1.

3.4. Validation of the prediction model

The final forecasting outcome is the average value of the output results of all decision trees according to Eqn (1). To evaluate the prediction performance of the RF regres-

sion model, the prediction accuracy of the RF model is comprehensively evaluated in terms of its precision and stability, and the commonly used average is selected (Liu et al., 2022; Qian et al., 2021). The root mean square error ($RMSE$) and goodness of fit (R^2) are the two parameters used to assess the accuracy of the prediction results. The $RMSE$ is used to describe the bias between the forecasted value and the measured value, thereby reflecting the rate of divergence of the sample. As the $RMSE$ approaches 0, the model deviation decreases, and the precision increases. R^2 reflects the fitting degree between variables; if the value is close to 1, the fit is good, and the interpretability of the model is high. R^2 reflects the grade of agreement between the forecasted value and the real value. These two indicators should be comprehensively considered when evaluating RF models. Additionally, the prediction results of the RF model are compared with the prediction results of an SVM model, a BP neural network and a GBDT, and the results verify the superiority of the RF model. Equations (5) and (6) show the expressions for the $RMSE$ and R^2 :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y^{obs} - y^{pred})^2}{n}}; \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y^{obs} - y^{pred})^2}{\sum_{i=1}^n (y^{obs} - \bar{y}^{obs})^2}, \quad (6)$$

where y^{obs} represents the measured value and y^{pred} represents the forecasted value for sample i .

4. Case study

4.1. Project context

As one of the important provincial highways in Jilin, the Yulin-Changchun Highway is an important national construction project. The main line is 208 kilometers in length and passes through eight bridges and six tunnels. The project is located in an alpine and high-salt-alkali area in Northeast China. The area has long winters, receives little snowfall, and is characterized by cold, dry, and relatively harsh geographic and climatic conditions. The project environment is shown in Figure 3a. To mitigate the issues caused by freeze–thaw cycles and salt-alkali corrosion, concrete must provide high frost resistance, and many projects now require high-durability concrete. The concrete freeze resistance grade is now greater than 300 (based on the number of freeze-thaw cycles) in most

Table 1. Search ranges of RF hyperparameters

ML model	Hyperparameter	Description	Search range
RF	$n_estimators$	Number of tree models	[1, 800]
	max_depth	Maximum depth of tree models	[1, 15]
	$min_samples_split$	Minimum number of samples required to split a node	[2, 20]
	$min_samples_leaf$	Minimum number of samples required in a leaf	[1, 20]
	$min_impurity_decrease$	Criterion for node splitting, which results in a decrease in impurity	[0, 10]

projects, and the slump limit is 10–30 mm. To design the concrete mix proportion in the early stage of construction, a raw material mix proportion test for expressway concrete was performed to obtain test data and RDEM data samples (where the RDEM is the 300 freeze–thaw cycle RDEM) according to the Durability Design Standard for Concrete Structures, GB/T 50476-2019 (People’s Republic of China, 2019). For C50 concrete and with the proposed machine learning algorithm for forecasting concrete frost resistance, durability tests of concrete were performed, as shown in Figure 3.

a) Expressway project



b) Test block of concrete



c) Dynamic elastic modulus test of concrete



Figure 3. Concrete durability test

4.2. Establishment of a sample data set

Based on a large number of previous studies and a summary of practical engineering experience (Tumidajski, 2005), the freeze–thaw mechanism of concrete was analyzed (Ben Chaabene et al., 2020), and the water-binder ratio, cement content, fly ash content, fine aggregate content, coarse aggregate content and six factors influencing the frost resistance of the water reducer were used as input variables. A primary indicator system for frost resistance was constructed by selecting the RDEM as the independent variable. Based on an actual project, 100 sets of orthogonal test data were obtained, among which 80 sets of sample data were used as the training sample set and 20 sets of data were used as the inspection sample set. The training set was used to select the RF parameters and build the RF model, and the test set was used to evaluate and validate the prediction performance of the model. Table 2 shows the detailed sample data. Due to space limitations, details for all the datasets are not provided. The full datasets are available upon reasonable request from the corresponding authors.

Table 2. Concrete frost resistance data

x_1	x_2	x_3	x_4	x_5	x_6	y
380	48	1106	706	1.1	0.34	95.66
358	45	1190	670	1.3	0.32	94.48
372	47	1174	661	1.3	0.32	94.81
...
408	72	1160	652	1.1	0.33	99.94
435	24	1165	656	1.1	0.31	96.28
406	24	1160	652	1.2	0.33	93.84
377	22	1237	637	1.3	0.3	91.50

4.3. Bayesian optimization of hyperparameters

Based on a theoretical study, the generalizability of the RF regression model is directly affected by the RF parameters, so important parameters such as max_features, n_estimators, and max_depth, which affect the establishment of the decision tree model and the regression fitting ability of the model, must be adjusted. Because the number of selected training sample characteristics is low, only 6 max_features could be automatically set by default, and the goodness of fit was used as the performance evaluation index. The range of n_estimators was set to 1~100, and the step size was 10. Max_depth was set to 5~8 according to the sample size, the step distance was 1, and the two parameters were cross-combined and modeled. The parameters of RF model were optimized by BO method and 5-fold cross-validation method. The optimization result is shown in Figure 4. After calculations, the values of n_estimators and max_depth that minimized the error of the frost resistance prediction model were max_depth = 5 and n_estimators = 54, with $R^2 = 0.9674$.

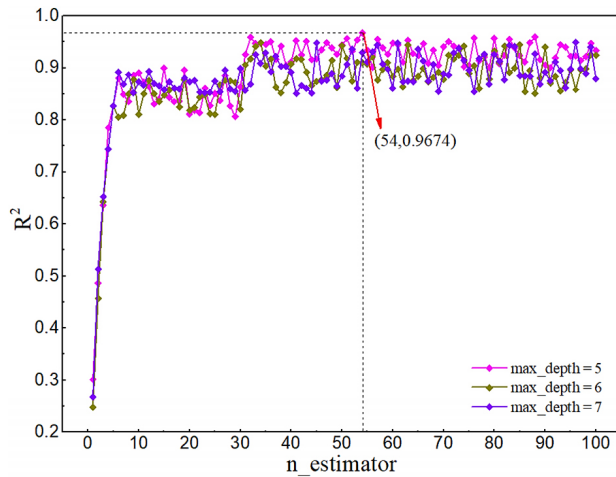


Figure 4. Selection of the number of decision trees

4.4. Evaluation of the regression prediction results

Based on the Bayesian optimization of hyperparameters, the regression test results were obtained by modeling the training samples and test samples. Figure 5a and Figure 5b show the regression fitting curve of the training sample set and the prediction result of regression fitting with the test sample set, respectively. The following results were obtained.

- (1) The discrepancy between the forecasted and actual values of the RDEM predicted with the BO-RF model is small. The RMSE between the real and predicted values for the RDEM for the training set is 0.04504, and the RMSE for the test set is 0.09578. The closer the RMSE is to 0, the higher the prediction precision of the model is.
- (2) The BO-RF prediction model displays a good fitting effect. The goodness of fit R^2 of the actual and predicted values of the RDEM for the training set is 0.9674, and the goodness of fit R^2 for the test set is 0.9592. The closer R^2 is to 1, the better the fitting effect of the model is.
- (3) The predictive performance of the BO-RF model is good. Similar to the conclusions of this study, Abou El Assad et al. (2020) used an RF model to accurately predict the shear

strength of reinforced concrete beams. Mai et al. (2021) used a BO-RF model to determine the best compressive strength of concrete containing ground granulated blast furnace slag (GGBFS). Studies have shown that the BO-RF output is an excellent predictor that can improve the prediction accuracy of subsequent models.

5. Discussion

5.1. Importance evaluation and correlation analysis

(1) Importance evaluation

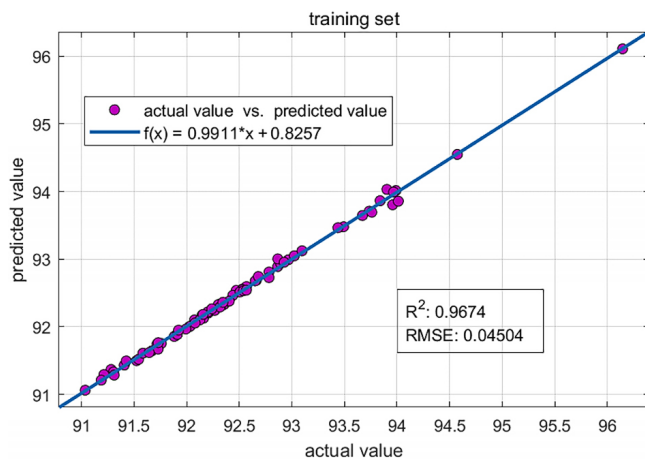
The significance of influential factors on the RDEM can be confirmed by importance evaluation. The importance score of each influential factor in the training set was calculated, the RF package in R software was used for the computation, and a significance ranking of every variable in the training model was obtained, as shown in Figure 6.

The water-binder ratio and the amount of cement have the greatest influence on the frost resistance. If the importance score is high, the effect of the variable on the evaluation indicator is correspondingly high, as is the importance of the variable. As shown in Figure 6, the importance ranking is as follows: water-binder ratio, cement content, fine aggregate content, coarse aggregate content, water reducer content and fly ash dosage. Similar to this conclusion, Ke and Duan (2021) found that the water-binder ratio has an important influence on the performance of high-performance concrete. Yang et al. (2012) showed that the water-binder ratio is a key factor affecting the compressive strength of concrete. Therefore, in actual projects, to ensure frost resistance, attention should be given to controlling the amounts of the most important raw materials mixed into concrete.

(2) Correlation analysis

The Pearson correlation coefficient (PCC) was used to analyze the linear relationships between various mix proportion factors and frost resistance. The PCC is the

a) Fitting results for training samples



b) Fitting results for test samples

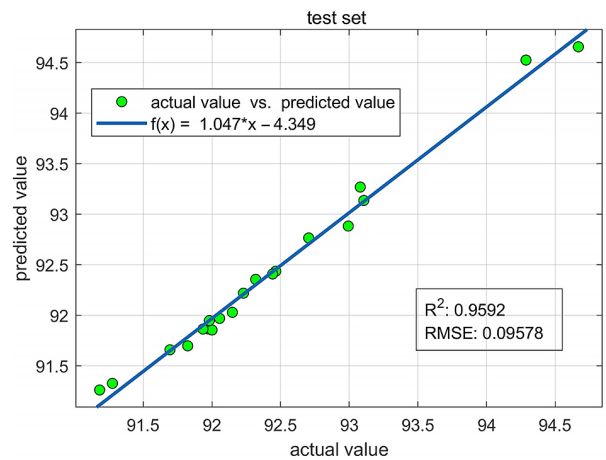


Figure 5. Prediction results

product of the mean difference and the mean difference sum of squares between two variables, so it is also called the correlation coefficient of product differences. The formula for the overall PCC is expressed as follows (Nápoles et al., 2020):

$$p = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E(X - \mu_X)(Y - \mu_Y)}{\sigma_X \sigma_Y}, \quad (7)$$

where $\text{cov}(X, Y)$ represents the covariance of quantities X and Y , μ_X and μ_Y represent the average values of variables X and Y , and σ_X and σ_Y are the deviations of variables X and Y , respectively.

The Pearson function can be used to analyze the correlations between influential factors and frost resistance. The stronger the factor correlation is, the greater the degree of influence. Figure 7 shows the PCCs between the calculated characteristic variables and frost resistance and a correlation graph of the results obtained with R software. The range of PCCs is from -1 to 1 . Blue represents a positive correlation between two variables, and red represents a negative correlation. The darker the color of the square is and the larger the size is, the greater the absolute value of the PCCs between the two variables is, and the higher the correlation. The opposite is true for weaker correlations.

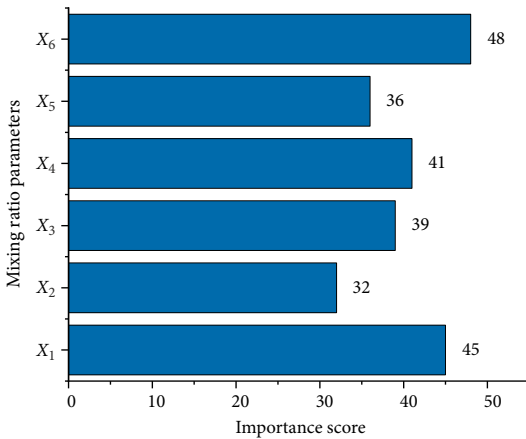


Figure 6. Importance scores for influential factors

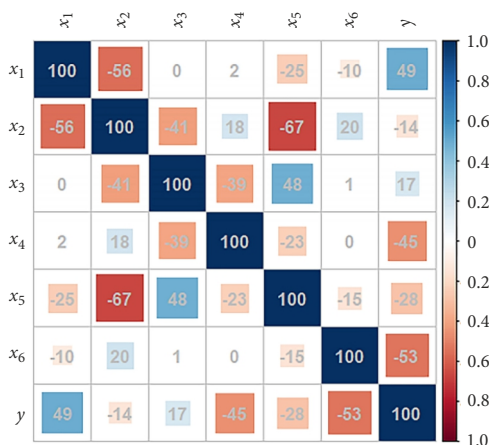


Figure 7. Correlation results

Figure 7 shows that (1) the results of the correlation analysis are roughly the same as the importance ranking results obtained with the RF algorithm. The water-binder ratio is the most important factor, and its correlation with the relative elastic modulus is significantly higher than that of the other factors. This verifies the accuracy of the RF algorithm. (2) Reducing the water-binder ratio and increasing the amount of cement can improve the frost resistance of concrete. Figure 7 shows that the correlation coefficients between the water-binder ratio and the cement dosage and the relative elastic modulus are -53% and 49% , respectively, indicating that the relative elastic modulus of concrete is negatively correlated with the water-binder ratio and positively correlated with the cement dosage. Therefore, in actual projects, priority can be given to reducing the water-binder ratio and increasing the amount of cement to improve the frost resistance of concrete. (3) The correlations between the input parameters are weak, and no coupling phenomenon exists. The correlation coefficient among the six parameters in Figure 7 is low, indicating no obvious coupling phenomenon among the parameters that would affect the reliability of the prediction results.

5.2. Prediction accuracy analysis

To further verify the credibility of the RF, BO-RF, SVM, BO-SVM, BP, BO-BP, GBDT and BO-GBDT models were used to forecast the frost resistance of concrete. Overall, the forecasting results were similar to those obtained with the RF model. For comparative analysis, the RMSE and certainty coefficient were chosen to assess the prediction abilities of the models. Comparisons of the prediction performance of the various models are shown in Table 3 and Figure 8.

Table 3 and Figure 8 show that (1) the RMSE of the BO-RF prediction model is the lowest and the closest to 0 among those of all models. For the training set, the RMSE of the BO-RF model is 0.045. For the test set, the RMSE of the BO-RF model is only 0.096, which is significantly lower than the values obtained for the other models. The results of the BO-RF prediction model are the closest to the actual values, and the prediction accuracy is the highest. (2) The R^2 value of the BO-RF model is the largest and closest to 1. For the training set and test set, the R^2 values of the BO-RF prediction model are 0.967 and 0.959, which are higher than those of the other three models, indicating that the BO-RF model yields the highest degree of fit to the data and the best prediction effect. (3) The BO-RF algorithm displays stronger adaptability than other models and is superior in forecasting concrete frost resistance. This conclusion has been verified by other scholars. For instance, by comparing the prediction performance of five machine learning algorithms, Yan and Shen (2022) found that the prediction precision of the BO-RF model was better than that of other algorithms.

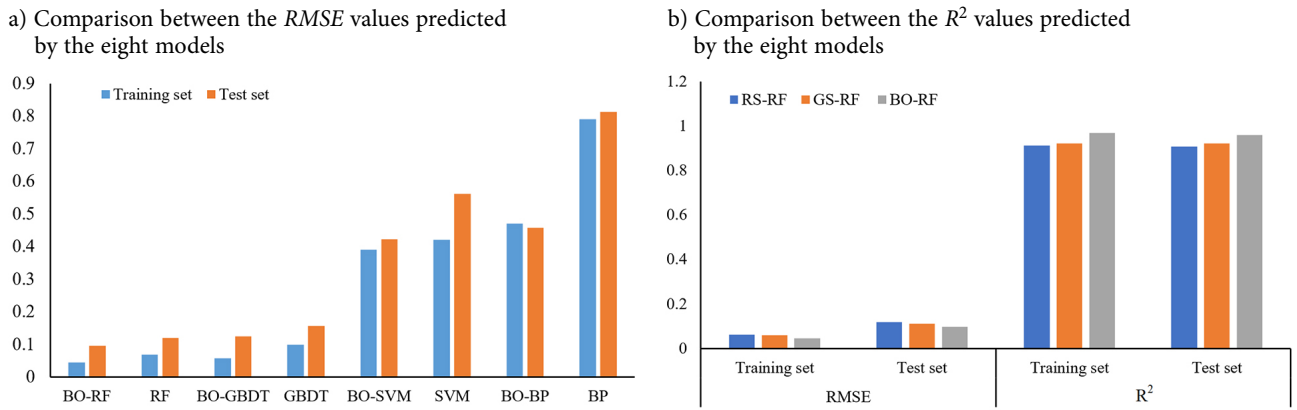


Figure 8. Comparison of the prediction performance of the eight algorithms

Table 3. Comparison of the prediction performance of the eight algorithms

Model	RMSE		R ²	
	Training set	Test set	Training set	Test set
BO-RF	0.045	0.096	0.967	0.959
RF	0.069	0.120	0.901	0.896
BO-GBDT	0.057	0.125	0.932	0.910
GBDT	0.098	0.156	0.891	0.821
BO-SVM	0.390	0.422	0.880	0.842
SVM	0.421	0.562	0.821	0.819
BO-BP	0.470	0.457	0.851	0.806
BP	0.791	0.812	0.801	0.801

Table 4. Comparison of the prediction performance of the three optimization algorithms

Model	RMSE		R ²	
	Training set	Test set	Training set	Test set
RS-RF	0.061	0.118	0.911	0.906
GS-RF	0.059	0.110	0.921	0.921
BO-RF	0.045	0.096	0.967	0.959

To further verify the advantages of BO optimization, the GS, RS and BO methods were selected to combine with the RF algorithm in this study. The corresponding prediction performance results obtained for the test set are shown in Table 4 and Figure 9.

Three hyperparameter optimization methods, RS, GR and BO, are used with the RF prediction model, and the BO-RF model displays the best prediction performance for the training set and the test set. The goodness-of-fit values are 0.967 and 0.959, respectively, which are better than those of the other algorithms. Additionally, the RMSE and MAE of the proposed algorithm are lower than those of other algorithms for concrete performance and cost. Thus, in terms of model hyperparameter optimization, the BO-RF algorithm yields the best prediction effect, indicating that BO provides excellent hyperparameter optimization ability.

The results are consistent with the conclusion of Liang (2019), who found that a hyperparameter optimization

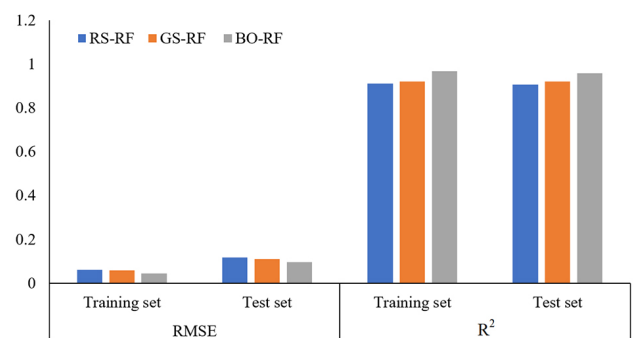


Figure 9. Comparison of the prediction performance of the three optimization algorithms

method based on Bayesian optimization performed well (Liang, 2019; Yang & Shami, 2020). Bayesian optimization was also noted by Y. Xia to be an excellent hyperparameter optimization approach that can be applied in the engineering field. In specific performance scenarios,

the prediction accuracy of BO models is higher than that of other hyperparameter optimization methods (Xia et al., 2017). Therefore, BO is the best choice for optimizing the hyperparameters of the model in this paper.

The set of mix proportions that yielded the largest RDEM (99.94%) in Table 2 was applied to the second section of the project, and experimental verification was performed. After 300 dynamic melting cycles, the RDEM was 99.84%, the deviation was only 0.1%, and the concrete design and application were effective.

The results of the above study indicate that the established BO-RF prediction model yields the best prediction effect among all studied models.

Conclusions

In complex and extreme environments, the frost resistance requirements of concrete are very high. To study the importance of relevant factors and accurately and efficiently predict the frost resistance of concrete, an intelligent prediction model based on the RF algorithm is proposed, six factors that have the greatest impact on frost resistance are selected, and the dynamic elastic modulus is used as the evaluation index. With a key national project as an example, the effectiveness of the method is verified. The main conclusions of this study are as follows.

(1) The proposed RF model can be used to effectively screen the factors that influence concrete frost resistance. The results show that the most important variable is the water-binder ratio, followed by the amounts of cement, fine aggregate, coarse aggregate, water reducing agent and fly ash. This conclusion is consistent with the specification requirements in actual applications.

(2) BO provides excellent model hyperparameter optimization ability. Notably, BO is used to optimize the hyperparameters of the RF prediction model, and the R^2 values of the BO-RF model for the training set and the test set are 0.967 and 0.959, respectively, which are better than those of the other algorithms.

(3) The proposed BO-RF hybrid algorithm can accurately and quickly predict the frost resistance of concrete. Using the test data from the first section of the project for prediction, the R^2 values are between 0.959 and 0.967, and the MAEs are between 0.045 and 0.096. Additionally, good results are obtained in an application involving the second section. Thus, the proposed approach can reduce the requirements of engineering tests in similar cases and save time and effort.

The frost resistance of concrete in the project is good, and the algorithm displays good potential application value for engineering projects. In addition, the model can be applied to a wide range of concrete research projects (such as those involving concrete strength, the concrete mix proportion and other factors). However, the study has some limitations. The common principles affecting the frost resistance of concrete are selected based on the concrete material mix proportion. In fact, concrete curing measures, climate and environmental all have a certain

impact on the frost resistance of concrete. Additionally, in this study, the RF algorithm was adopted to establish the prediction model of concrete frost resistance, improve the durability of concrete, optimize the concrete mix design, and preliminarily explore the cross-integration of computer and material disciplines. However, other performance indicators that influence the application of concrete, such as strength and resistance to chloride ion permeability, must be further analyzed.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 72031009), the Social Science Foundation of Hubei Province (Grant No. 22ZD004), the Philosophy and Social Science research Project in the Department of Education of Hubei Province (Grant No. 21G001), and the Zhongnan Hospital of Wuhan University Science, Technology and Innovation Seed Fund, Project CXPY2020013.

Author contributions

Jian Zhang and Yuan Cao contributed equally to the work. Jian Zhang: Conceptualization, Methodology, Software, Writing – original draft, Data curation. Yuan Cao: Conceptualization, Methodology, Software, Writing – original draft, Data curation. Linyu Xia: Methodology, Writing – review & editing. Desen Zhang: Writing – review & editing, Software. Wen Xu: Writing – review & editing. Yang Liu: Writing – review & editing, Funding acquisition, Resources.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Abou Elasad, E. Z., Mousannif, H., & Al Moatassime, H. (2020). A proactive decision support system for predicting traffic crash events: A critical analysis of imbalanced class distribution. *Knowledge-Based Systems*, 205, 106314. <https://doi.org/10.1016/j.knosys.2020.106314>
- Ala'raj, M., & Abbod, M. F. (2016). Classifiers consensus system approach for credit scoring. *Knowledge-Based Systems*, 104, 89–105. <https://doi.org/10.1016/j.knosys.2016.04.013>
- Amran, M., Huang, S.-S., Onaizi, A. M., Makul, N., Abdelgader, H. S., & Ozbakkaloglu, T. (2022). Recent trends in ultra-high performance concrete (UHPC): Current status, challenges, and future prospects. *Construction and Building Materials*, 352, 129029. <https://doi.org/10.1016/j.conbuildmat.2022.129029>
- Ashraf, W., Glinicki, M. A., & Olek, J. (2018). Statistical analysis and probabilistic design approach for freeze–thaw performance of ordinary Portland cement concrete. *Journal of Materials in Civil Engineering*, 30(11), 04018294. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002494](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002494)

- Aulia, A., Jeong, D., Saaid, I. M., Kania, D., Shuker, M. T., & El-Khatib, N. A. (2019). A Random Forests-based sensitivity analysis framework for assisted history matching. *Journal of Petroleum Science and Engineering*, 181, 106237. <https://doi.org/10.1016/j.petrol.2019.106237>
- Azimi-Pour, M., Eskandari-Naddaf, H., & Pakzad, A. (2020). Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete. *Construction and Building Materials*, 230, 117021. <https://doi.org/10.1016/j.conbuildmat.2019.117021>
- Belalia Douma, O., Boukhatem, B., Ghrici, M., & Tagnit-Hamou, A. (2016). Prediction of properties of self-compacting concrete containing fly ash using artificial neural network. *Neural Computing and Applications*, 28(S1), 707–718. <https://doi.org/10.1007/s00521-016-2368-7>
- Ben Chaabene, W., Flah, M., & Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. *Construction and Building Materials*, 260, 119889. <https://doi.org/10.1016/j.conbuildmat.2020.119889>
- Benedet, L., Acuña-Guzman, S. F., Faria, W. M., Silva, S. H. G., Mancini, M., Teixeira, A. F. d. S., Pereira Pierangeli, L. M., Acerbi Júnior, F. W., Gomide, L. R., Pádua Júnior, A. L., de Souza, I. A., de Menezes, M. D., Marques, J. J., Guilherme, L. R. G., & Curi, N. (2021). Rapid soil fertility prediction using X-ray fluorescence data and machine learning algorithms. *CATENA*, 197, 105003. <https://doi.org/10.1016/j.catena.2020.105003>
- Boukhatem, B., Kenai, S., Tagnit-Hamou, A., & Ghrici, M. (2011). Application of new information technology on concrete: An overview. *Journal of Civil Engineering and Management*, 17(2), 248–258. <https://doi.org/10.3846/13923730.2011.574343>
- Cai, R., Han, T., Liao, W., Huang, J., Li, D., Kumar, A., & Ma, H. (2020). Prediction of surface chloride concentration of marine concrete using ensemble machine learning. *Cement and Concrete Research*, 136, 106164. <https://doi.org/10.1016/j.cemconres.2020.106164>
- Chen, H., Deng, T., Du, T., Chen, B., Skibniewski, M. J., & Zhang, L. (2022). An RF and LSSVM-NSGA-II method for the multi-objective optimization of high-performance concrete durability. *Cement and Concrete Composites*, 129, 104446. <https://doi.org/10.1016/j.cemconcomp.2022.104446>
- Chen, H., Cao, Y., Liu, Y., Qin, Y., & Xia, L. (2023a). Enhancing the durability of concrete in severely cold regions: Mix proportion optimization based on machine learning. *Construction and Building Materials*, 371, 130644. <https://doi.org/10.1016/j.conbuildmat.2023.130644>
- Chen, H., Li, X., Feng, Z., Wang, L., Qin, Y., Skibniewski, M. J., Chen, Z.-S., & Liu, Y. (2023b). Shield attitude prediction based on Bayesian-LGBM machine learning. *Information Sciences*, 632, 105–129. <https://doi.org/10.1016/j.ins.2023.03.004>
- Cheng, M.-Y., & Hoang, N.-D. (2016). A self-adaptive fuzzy inference model based on least squares SVM for estimating compressive strength of rubberized concrete. *International Journal of Information Technology & Decision Making*, 15(3), 603–619. <https://doi.org/10.1142/s0219622016500140>
- Chun, P.-j., Ujike, I., Mishima, K., Kusumoto, M., & Okazaki, S. (2020). Random forest-based evaluation technique for internal damage in reinforced concrete featuring multiple non-destructive testing results. *Construction and Building Materials*, 253, 119238. <https://doi.org/10.1016/j.conbuildmat.2020.119238>
- DeRousseau, M. A., Kasprzyk, J. R., & Srubar, W. V. (2018). Computational design optimization of concrete mixtures: A review. *Cement and Concrete Research*, 109, 42–53. <https://doi.org/10.1016/j.cemconres.2018.04.007>
- Ding, H., Na, J., Jiang, S., Zhu, J., Liu, K., Fu, Y., & Li, F. (2021). Evaluation of three different machine learning methods for object-based artificial terrace mapping – A case study of the Loess Plateau, China. *Remote Sensing*, 13(5), 1021. <https://doi.org/10.3390/rs13051021>
- Dvorkin, L. (2019). Design estimation of concrete frost resistance. *Construction and Building Materials*, 211, 779–784. <https://doi.org/10.1016/j.conbuildmat.2019.03.108>
- Gao, Y., Jing, H. W., Zhao, Z. L., Shi, X. S., & Li, L. (2021). Influence of ultrasonication energy on reinforcing-roles of CNTs to strengthen ITZ and corresponding anti-permeability properties of concrete. *Construction and Building Materials*, 303, 124451. <https://doi.org/10.1016/j.conbuildmat.2021.124451>
- Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4), 455–492. <https://doi.org/10.1023/A:1008306431147>
- Ke, X., & Duan, Y. (2021). A Bayesian machine learning approach for inverse prediction of high-performance concrete ingredients with targeted performance. *Construction and Building Materials*, 270, 121424. <https://doi.org/10.1016/j.conbuildmat.2020.121424>
- Keleştemur, O., Yildiz, S., Gökçer, B., & Arici, E. (2014). Statistical analysis for freeze–thaw resistance of cement mortars containing marble dust and glass fiber. *Materials & Design*, 60, 548–555. <https://doi.org/10.1016/j.matdes.2014.04.013>
- Kewalramani, M. A., & Gupta, R. (2006). Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks. *Automation in Construction*, 15(3), 374–379. <https://doi.org/10.1016/j.autcon.2005.07.003>
- Koya, B. P., Aneja, S., Gupta, R., & Valeo, C. (2022). Comparative analysis of different machine learning algorithms to predict mechanical properties of concrete. *Mechanics of Advanced Materials and Structures*, 29(25), 4032–4043. <https://doi.org/10.1080/15376494.2021.1917021>
- Lakshmanaprabu, S. K., Shankar, K., Ilayaraja, M., Nasir, A. W., Vijayakumar, V., & Chilamkurti, N. (2019). Random forest for big data classification in the internet of things using optimal features. *International Journal of Machine Learning and Cybernetics*, 10(10), 2609–2618. <https://doi.org/10.1007/s13042-018-00916-z>
- Łażniewska-Piekarczyk, B. (2013). The type of air-entraining and viscosity modifying admixtures and porosity and frost durability of high performance self-compacting concrete. *Construction and Building Materials*, 40, 659–671. <https://doi.org/10.1016/j.conbuildmat.2012.11.032>
- Lee, B., Kim, K., Hwang, H., Kim, Y. S., Chung, E. H., Yoon, J.-S., Cho, H. J., & Park, J. D. (2021). Development of a machine learning model for predicting pediatric mortality in the early stages of intensive care unit admission. *Scientific Reports*, 11, 1263. <https://doi.org/10.1038/s41598-020-80474-z>
- Li, Y.-S., Chi, H., Shao, X.-Y., Qi, M.-L., & Xu, B.-G. (2020). A novel random forest approach for imbalance problem in crime linkage. *Knowledge-Based Systems*, 195, 105738. <https://doi.org/10.1016/j.knsys.2020.105738>
- Liang, X. (2019). Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with

- Bayesian optimization. *Computer-Aided Civil and Infrastructure Engineering*, 34(5), 415–430.
<https://doi.org/10.1111/mice.12425>
- Lin, J.-P., & Liu, J.-h. (2006). Support vector machine and optimized method for spectral analysis. *Spectroscopy and Spectral Analysis*, 26(12), 2232–2235.
- Liu, J.-C., & Zhang, Z. (2020). A machine learning approach to predict explosive spalling of heated concrete. *Archives of Civil and Mechanical Engineering*, 20(4), 134.
<https://doi.org/10.1007/s43452-020-00135-w>
- Liu, Y., Chen, H. Y., Zhang, L. M., & Wu, X. (2020). Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China. *Journal of Cleaner Production*, 272, 129232.
<https://doi.org/10.1016/j.jclepro.2020.122542>
- Liu, Y., Chen, H., Zhang, L., & Feng, Z. (2021a). Enhancing building energy efficiency using a random forest model: A hybrid prediction approach. *Energy Reports*, 7, 5003–5012.
<https://doi.org/10.1016/j.egy.2021.07.135>
- Liu, Y., Chen, H., Zhang, L., & Wang, X. (2021b). Risk prediction and diagnosis of water seepage in operational shield tunnels based on random forest. *Journal of Civil Engineering and Management*, 27(7), 539–552.
<https://doi.org/10.3846/jcem.2021.14901>
- Liu, Y., Cao, Y., Wang, L., Chen, Z.-S., & Qin, Y. (2022). Prediction of the durability of high-performance concrete using an integrated RF-LSSVM model. *Construction and Building Materials*, 356, 129232.
<https://doi.org/10.1016/j.conbuildmat.2022.129232>
- Luan, H., Wu, J., & Pan, J. (2020). Saline water absorption behavior and critical saturation degree of recycled aggregate concrete during freeze-thaw cycles. *Construction and Building Materials*, 258, 119640.
<https://doi.org/10.1016/j.conbuildmat.2020.119640>
- Lundström, J., & Verikas, A. (2013). Assessing print quality by machine in offset colour printing. *Knowledge-Based Systems*, 37, 70–79. <https://doi.org/10.1016/j.knsys.2012.07.022>
- Mai, H.-V. T., Nguyen, T.-A., Ly, H.-B., Tran, V. Q., & Zhang, J. (2021). Prediction compressive strength of concrete containing GGBFS using random forest model. *Advances in Civil Engineering*, 2021, 6671448. <https://doi.org/10.1155/2021/6671448>
- Marcos-Pasero, H., Colmenarejo, G., Aguilar-Aguilar, E., de Molina, A. R., Reglero, G., & Loria-Kohen, V. (2021). Ranking of a wide multidomain set of predictor variables of children obesity by machine learning variable importance techniques. *Scientific Reports*, 11, 1910.
<https://doi.org/10.1038/s41598-021-81205-8>
- Martinez-Cantin, R. (2014). BayesOpt: A Bayesian Optimization Library for nonlinear optimization, experimental design and bandits. *Journal of Machine Learning Research*, 15, 3735–3739.
- Nápoles, G., Grau, I., & Salgueiro, Y. (2020). Recommender system using Long-term Cognitive Networks. *Knowledge-Based Systems*, 206, 106372.
<https://doi.org/10.1016/j.knsys.2020.106372>
- Nguyen, T. D., Gupta, S., Rana, S., & Venkatesh, S. (2018). Stable Bayesian optimization. *International Journal of Data Science and Analytics*, 6(4), 327–339.
<https://doi.org/10.1007/s41060-018-0119-9>
- Nguyen, T. T., Ngoc, L. T., Vu, H. H., & Thanh, T. P. (2021). Machine learning-based model for predicting concrete compressive strength. *International Journal of GEOMATE*, 20, 197–204. <https://doi.org/10.21660/2020.77.j2019>
- Nilsen, V., Pham, L. T., Hibbard, M., Klager, A., Cramer, S. M., & Morgan, D. (2019). Prediction of concrete coefficient of thermal expansion and other properties using machine learning. *Construction and Building Materials*, 220, 587–595.
<https://doi.org/10.1016/j.conbuildmat.2019.05.006>
- Niu, D., Wang, K., Sun, L., Wu, J., & Xu, X. (2020). Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study. *Applied Soft Computing*, 93, 106389.
<https://doi.org/10.1016/j.asoc.2020.106389>
- Özcan, F., Atiş, C. D., Karahan, O., Uncuoğlu, E., & Tanyildizi, H. (2009). Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete. *Advances in Engineering Software*, 40(9), 856–863.
<https://doi.org/10.1016/j.advengsoft.2009.01.005>
- People's Republic of China. (2019). *Standard for design of concrete structure durability* (GB/T 50476-2019). National standard.
- Puga, J. L., Krzywinski, M., & Altman, N. (2015). Bayes' theorem. *Nature Methods*, 12(4), 277–278.
<https://doi.org/10.1038/nmeth.3335>
- Qian, Y. W., Zeng, G., Pan, Y., Liu, Y., Zhang, L. M., & Li, K. (2021). A prediction model for high risk of positive RT-PCR test results in COVID-19 patients discharged from Wuhan Leishenshan Hospital, China. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.778539>
- Royal, R., Khanna, D., Sandhu, J. K., Hooda, N., & Rana, P. S. (2017). N-semble: Neural network based ensemble approach. *International Journal of Machine Learning and Cybernetics*, 10(2), 337–345. <https://doi.org/10.1007/s13042-017-0718-0>
- Smith, S. H., Kurtis, K. E., & Tien, I. (2018). Probabilistic evaluation of concrete freeze-thaw design guidance. *Materials and Structures*, 51, 124.
<https://doi.org/10.1617/s11527-018-1259-z>
- Sonebi, M., Cevik, A., Grünwald, S., & Walraven, J. (2016). Modelling the fresh properties of self-compacting concrete using support vector machine approach. *Construction and Building Materials*, 106, 55–64.
<https://doi.org/10.1016/j.conbuildmat.2015.12.035>
- Tumidajski, P. J. (2005). Relationship between resistivity, diffusivity and microstructural descriptors for mortars with silica fume. *Cement and Concrete Research*, 35(7), 1262–1268.
<https://doi.org/10.1016/j.cemconres.2004.10.007>
- Wei, Y., Han, A., & Xue, X. (2021). A data-driven study for evaluating the compressive strength of high-strength concrete. *International Journal of Machine Learning and Cybernetics*, 12(12), 3585–3595.
<https://doi.org/10.1007/s13042-021-01407-4>
- Wu, Q., Ye, Y., Zhang, H., Ng, M. K., & Ho, S.-S. (2014). Forest-Texter: An efficient random forest algorithm for imbalanced text categorization. *Knowledge-Based Systems*, 67, 105–116.
<https://doi.org/10.1016/j.knsys.2014.06.004>
- Wu, X., Wang, L., Chen, B., Feng, Z., Qin, Y., Liu, Q., & Liu, Y. (2022a). Multi-objective optimization of shield construction parameters based on random forests and NSGA-II. *Advanced Engineering Informatics*, 54, 101751.
<https://doi.org/10.1016/j.aei.2022.101751>
- Wu, X., Zheng, S., Feng, Z., Chen, B., Qin, Y., Xu, W., & Liu, Y. (2022b). Prediction of the frost resistance of high-performance concrete based on RF-REF: A hybrid prediction approach. *Construction and Building Materials*, 333, 127132.
<https://doi.org/10.1016/j.conbuildmat.2022.127132>

- Wu, X., Li, X., Qin, Y., Xu, W., & Liu, Y. (2023). Intelligent multi-objective optimization design for NZEBs in China: Four climatic regions. *Applied Energy*, 339, 120934. <https://doi.org/10.1016/j.apenergy.2023.120934>
- Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78, 225–241. <https://doi.org/10.1016/j.eswa.2017.02.017>
- Yan, M., & Shen, Y. (2022). Traffic accident severity prediction based on random forest. *Sustainability*, 14(3), 1729. <https://doi.org/10.3390/su14031729>
- Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neuro-computing*, 415, 295–316. <https://doi.org/10.1016/j.neucom.2020.07.061>
- Yang, Q. R., Guo, B. L., Yang, Q. B., & Li, Q. Q. (2011). Experimental study on durability of air-entrained concrete of Qingdao Bay Bridge. In *Applied Mechanics and Materials* (Vols. 99–100, pp. 142–145). Trans Tech Publications, Ltd. <https://doi.org/10.4028/www.scientific.net/AMM.99-100.142>
- Yang, K.-H., Cho, A.-R., & Song, J.-K. (2012). Effect of water–binder ratio on the mechanical properties of calcium hydroxide-based alkali-activated slag concrete. *Construction and Building Materials*, 29, 504–511. <https://doi.org/10.1016/j.conbuildmat.2011.10.062>
- Yaseen, Z. M., Deo, R. C., Hilal, A., Abd, A. M., Bueno, L. C., Salcedo-Sanz, S., & Nehdi, M. L. (2018). Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Advances in Engineering Software*, 115, 112–125. <https://doi.org/10.1016/j.advengsoft.2017.09.004>
- YAZICI, H. (2008). The effect of silica fume and high-volume Class C fly ash on mechanical properties, chloride penetration and freeze–thaw resistance of self-compacting concrete. *Construction and Building Materials*, 22(4), 456–462. <https://doi.org/10.1016/j.conbuildmat.2007.01.002>
- Yuan, J., Du, Z., Wu, Y., & Xiao, F. (2019). Salt-frost resistance performance of airfield concrete based on meso-structural parameters. *Journal of Materials in Civil Engineering*, 31(9). [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002789](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002789)
- Zhang, H.-R., & Min, F. (2016). Three-way recommender systems based on random forests. *Knowledge-Based Systems*, 91, 275–286. <https://doi.org/10.1016/j.knosys.2015.06.019>
- Zhou, Q., Zhou, H., & Li, T. (2016). Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features. *Knowledge-Based Systems*, 95, 1–11. <https://doi.org/10.1016/j.knosys.2015.11.010>

APPENDIX

Abbreviation	
RF	Random forest
RDEM	Relative dynamic elastic modulus
SVM	Support vector machine
BP	Back propagation
GBDT	Gradient boosting decision tree
ANN	Artificial neural network
CTE	Coefficient of thermal expansion
OBB	Out-of-bag data
CART	Classification and regression tree
GGBFS	Ground granulated blast furnace slag
PCCs	Pearson correlation coefficient
Nomenclature	
p	Out-of-bag data
n	The number of samples
E^*	The generalization error of the RF model
$P_{X,Y}$	The probability for the X and Y spaces
$p = (1 - \frac{1}{n})^n$	The prediction result
H_i	A single decision tree
Y	The output variable
K	The number of decision trees in the RF model
MSE_j	The mean square residual of the j -th sample
S_E	The standard error
VIM_i	The variable importance of input variables
K	The quantity of original data
$RMSE$	The root mean square error
R^2	The goodness of fit
y^{obs}	The observed value of sample i
y^{pred}	The predicted value of sample i
$Cov(X, Y)$	The covariance of the quantity X and the quantity Y
μ_X	The average value of the variable X
μ_Y	The average value of the variable Y
σ_X	The deviations of the variable X
σ_Y	The deviation of the variable Y
x_1	The cement content
x_2	The fly ash dosage
x_3	Coarse aggregate
x_4	Fine aggregate
x_5	Water reducer
x_6	The water-binder ratio
Y	The relative dynamic elastic modulus