ISSN: 1648-6897 / eISSN: 1822-4199

VILNIUS TECH
Vilnius Gediminas Technical University

JOURNAL of ENVIRONMENTAL ENGINEERING & LANDSCAPE MANAGEMENT

2025
Volume 33
Issue 2
Pages 226–241
https://doi.org/10.3846/jeelm.2025.23568

PREDICTION OF THE ANTI-CARBONATION PERFORMANCE OF CONCRETE BASED ON RANDOM FOREST – LEAST SQUARES SUPPORT VECTOR MACHINE MODEL

Sivaraja M.[1]✉, Swaminathen A. N.[2], Kuttimarks M. S.[3], Rajprasad J.[4], Sakthivel M.[5], Rex J.[6]

[1]*Department of Civil Engineering, Nehru Institute of Technology, Coimbatore, 641105 Tamilnadu, India*
[2]*Department of Civil Engineering, Adi Shankara Institute of Engineering and Technology, Kalady, Kerala*
[3]*Department of Civil Engineering, Shivajirao S. Jondhle College of Engineering & Technology, Thane, Maharastra*
[4]*Department of Civil Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, 603203 Kattakulathur, Tamilnadu, India*
[5]*Department of Civil Engineering, Kongunadu College of Engineering and Technology, 621215 Trichy, India*
[6]*Department of Civil Engineering, CK College of Engineering & Technology, 607003 Cuddalore, Tamilnadu, India*

Highlights:

- enhance the accuracy of predicting the anti-carbonation performance (ACP) of concrete structures;
- combine random forest (RF) regression and least squares support vector machine (LSSVM) for improved prediction accuracy;
- utilize RF regression to identify the most significant factors affecting ACP, optimizing the input features for the LSSVM model;
- validate the hybrid RF-LSSVM model against test data to demonstrate its robustness and reliability;
- exhibit superior performance in predicting concrete carbonation resistance compared to traditional methods;
- perform comprehensive error analysis to show the model's effectiveness in minimizing prediction errors;
- offer a reliable tool for assessing concrete durability, contributing to the longevity and safety of concrete structures;
- address the limitations of existing empirical and mathematical models, offering a more precise evaluation method for ACP;
- ensure the model can be widely applied in the construction industry for better maintenance and safety of concrete infrastructure;
- enhance the methodology for evaluating concrete ACP, promoting advancements in construction material science and structural engineering.

Article History:

- received 06 December 2023
- accepted 21 January 2025

Abstract. Concrete carbonation is a critical factor influencing the durability and longevity of concrete structures, particularly in urban environments. Traditional methods for predicting the anti-carbonation performance (ACP) of concrete often lack precision and fail to account for complex interactions between influencing factors. In this study, a novel hybrid model combining random forest (RF) regression with a least squares support vector machine (LSSVM) is proposed to enhance the accuracy of ACP predictions. The RF regression is utilized for feature selection, identifying the most significant factors affecting ACP and optimizing the input features for the LSSVM model. Our hybrid model is validated against a comprehensive dataset, demonstrating superior performance in predicting concrete carbonation resistance compared to conventional methods. Quantitative results show that the RF-LSSVM model achieves a root mean square error (RMSE) of $5e^{-5}$ and a high coefficient of determination (R-squared) of 0.999, indicating robust predictive capability and accuracy. The main novelty of this work lies in the integration of RF and LSSVM to create a robust, accurate, and practical tool for assessing the durability of concrete structures.

Keywords: carbonation resistance, least squares support vector machine (LSSVM) model, safety assessment, random forest (RF), concrete.

✉Corresponding author. E-mail: ***nitprincipal@nehrucolleges.com***

Notations

ACP – anti-carbonation performance;
RF – random forest;
LSSVM – least squares support vector machine;
RMSE – root mean square error;
RC – reinforced concrete;
C-Carb – carbonation of concrete;
XRDA – X-ray diffraction;
TGA – thermogravimetry;
FTIR – Fourier transform infrared spectroscopy;
cc – cement dosage;
fasc – fly ash content;
fagc – fine aggregate content;
cagc – coarse aggregate content;
wrc – water reducing agent content;

wc – water content;
RH – relative humidity;
TEMP – temperature;
C-CO_2 – carbon dioxide concentration;
csv – concrete specimen volume;
Tcarb – carbonation time;
CS – compressive strength;
OOB – out-of-bag;
OOBerr – OOB data error;
RFE – recursive feature backward elimination;
GKF – Gaussian kernel function;
GA – genetic algorithm;
SVM – support vector machine;
ANN – artificial neural network;
R^2 – Goodness of fit.

1. Introduction

Concrete has been a foundational material in construction for decades due to its versatility, cost-effectiveness, and performance in various environments. Its widespread use stems from its ease of manufacturing, availability of raw materials, low maintenance requirements, and adaptability for different structural applications. However, RC structures are prone to durability issues, particularly when exposed to harsh environmental conditions (Kar et al., 2020). These conditions often lead to a cascade of chemical and physical reactions that undermine the concrete's structural integrity over time (Yaseen et al., 2018; Zhang et al., 2022; Maryam et al., 2018).

1.1. The challenge of carbonation in concrete

One of the most significant durability challenges faced by concrete structures is carbonation. Carbonation refers to the gradual reduction in the alkalinity of the concrete due to the penetration of carbon dioxide (CO_2) from the atmosphere into the concrete matrix. When CO_2 reacts with calcium hydroxide ($Ca(OH)_2$) in the concrete, it forms calcium carbonate ($CaCO_3$). This reaction lowers the pH level of the concrete, compromising its alkaline environment, which is crucial for protecting the embedded steel reinforcements from corrosion. Once the concrete loses its protective alkaline properties, the steel reinforcements become vulnerable to corrosion, which can significantly reduce the lifespan of the structure (Li et al., 2022).

1.2. Protective measures against carbonation

To combat carbonation and its effects, several protective systems have been developed. Among these, surface coatings are one of the most commonly employed methods. These coatings act as barriers to CO_2 ingress, helping to preserve the concrete's alkalinity and, in turn, the durability of the steel reinforcements. Evaluating the effectiveness of such protective systems typically involves accelerated carbonation tests. These tests expose concrete samples to elevated levels of CO_2 to simulate the long-term effects of carbonation in a relatively short period (Verapathran et al., 2023; Heoa et al., 2016; Chaabene & Nehdi, 2020).

Despite the widespread use of protective systems and accelerated testing, a gap remains in the understanding of how these systems perform under real-world conditions. Many of the empirical and mathematical models developed to predict carbonation progression are based on unprotected systems and do not adequately address the complexities involved when protective coatings are applied. This limitation underscores the need for new and improved methods for predicting the ACP of concrete structures, particularly those with protective coatings.

1.3. Factors affecting carbonation

Several factors influence the carbonation process in concrete, including the properties of the cement, the composition and strength of the concrete, environmental exposure conditions, and the presence of protective coatings. Cement properties, such as type and quality, directly impact the permeability of concrete to CO_2 and other aggressive agents. Concrete strength and composition, particularly the water-cement ratio, also play key roles in determining how resistant a structure will be to carbonation (Junyoung & Yootaek, 2014). Structures with higher strength concrete, for instance, tend to have lower porosity and are therefore less permeable to CO_2 ingress (Chou et al., 2020; Al-Musawi et al., 2020).

Given the complexity of these interacting factors, researchers have sought to model and predict carbonation depth using mathematical and empirical approaches. However, solving the high-order differential equations involved in these models, which must account for multiple variables, is a labor-intensive and time-consuming process. Additionally, many of these models require extensive experimental data, making them costly and impractical for widespread application (España et al., 2017; Avci et al., 2021; Solhmirzaei et al., 2020).

To address these challenges, researchers are increasingly turning to machine learning (ML) and artificial intelligence (AI) techniques for predicting concrete carbonation. These approaches offer significant advantages over traditional empirical models by leveraging large datasets to uncover patterns and relationships between variables that may not be immediately apparent. ML algorithms can rapidly process historical data, enabling more efficient and accurate predictions of carbonation depth and concrete durability (Kamali et al., 2022; Barjouei et al., 2021; Anemangely et al., 2019; Ly et al., 2020).

Among the various intelligent algorithms, RF regression and LSSVM have shown considerable potential in predicting ACP. RF regression is particularly useful for feature selection, identifying the most important factors influencing ACP. By reducing the dimensionality of the input dataset, RF can optimize the performance of subsequent models, such as LSSVM, which is then used for precise regression analysis. The combination of RF and LSSVM provides a robust and accurate method for predicting ACP, outperforming many traditional approaches (Mehrad et al., 2020, 2022; Abad et al., 2022; Davoodi et al., 2023a, 2023b, 2023c, 2023d, 2023e).

1.4. The proposed predictive model: integrating RF and LSSVM

The objective of this study is to develop a hybrid predictive model for concrete's ACP by combining RF regression and LSSVM. The novelty of this approach lies in the synergy between the two models: RF regression is first employed to select the most relevant features, ensuring that only the most significant variables affecting ACP are considered. LSSVM is then applied to these optimized features to perform regression analysis and predict ACP with high accuracy.

This hybrid model will be trained and validated using a comprehensive dataset that encompasses a wide range of concrete properties, environmental conditions, and protective systems. By comparing the performance of the RF-LSSVM model to traditional methods, this study aims to demonstrate the superiority of the hybrid approach in accurately predicting concrete carbonation resistance (Sheykhinasab et al., 2023; Jafarizadeh et al., 2023; Matinkia et al., 2022a, 2022b; Talkhouncheh et al., 2023; Davoodi et al., 2024).

The proposed RF-LSSVM model represents a significant advancement in the field of concrete durability prediction. By leveraging the strengths of both RF and LSSVM, this approach offers a precise, efficient, and scalable solution for predicting the ACP of concrete, particularly in protected systems. This innovation not only enhances our understanding of concrete carbonation but also provides a practical tool for improving the long-term performance of RC structures in various environmental conditions.

1.5. Research significance

- Widespread Use of Concrete: Concrete's extensive use in various types of structures highlights the importance of ensuring its durability.
- Durability Concerns: Unfavorable exposure conditions can significantly reduce the durability of RC due to chemical and physical reactions.
- Protection Systems: Surface coatings are among the most used systems for protecting concrete against carbonation.
- Testing Methods: High concentration CO_2 tests are commonly used to evaluate the performance of concrete protection systems, providing accelerated results.
- Current Models: Several mathematical and empirical models exist to predict carbonation, but their applicability to protected systems is limited.
- Complexity of Calculations: Mathematical models often involve cumbersome calculations and high-order differential equations.
- Microscopic Studies: Early C-Carb studies at the microscopic level require extensive experimental data, leading to increased research costs.
- Data-Driven Prediction: Utilizing accumulated data and intelligent algorithms offers a more efficient approach to predicting early carbonation.
- Novel Hybrid Model: The proposed hybrid model combining RF and LSSVM offers improved accuracy in predicting ACP.
- Advancement in Research: This new approach addresses the deficiencies of existing models, contributing to advancements in construction material science and structural engineering.

2. Literature review

Most carbonation prediction models found in the literature (Jumaa'h et al., 2019; Nafees et al., 2021; Zhao et al., 2020) assume a variation in C-Carb depth proportional to the square root of the exposure time. This coefficient depends on many different factors, which makes it very difficult to correctly model the evolution of the carbonation depth. In (Zhao et al., 2020) even considers that it is doubtful that it will ever be possible to determine a formula for predicting the rate of carbonation with adequate accuracy, which takes into account all the parameters involved.

For the determination of the carbonation front, the most used method is through the sprinkling of an alcoholic solution of phenolphthalein. This indicator only reveals the separation of two distinct pH zones. Lilac colored zones with pH > 9 and therefore called non-carbonated, and zones without color change, called carbonated. This technique is easy to apply and the results are obtained quickly. In (Chang & Chen, 2016) found, through readings with phenolphthalein, which corrosion could start between 6 and 8 mm away from the carbonation front. However, this procedure is considered to provide a good estimate of carbonation depth and is undoubtedly the most widely used method.

According to a study carried out by Ibrahim et al. (2019), other more sophisticated methods such as XRDA, TGA and FTIR, when analyzing the microscopic structure of concrete in depth, present carbonation depth values higher than those measured with phenolphthalein. These methods require specialized equipment and personnel and are not very quick to apply, but allow the precise identification of the degree and C-Carb depth.

Another way of predicting the progress of the carbonation front is to use tests in accelerated carbonation chambers. In these tests, the chambers are prepared with an atmosphere enriched in and the depth of carbonation is measured at regular time intervals. It is possible to compare these results with results for the same concrete under real exposure conditions. Several authors have carried out studies in this respect (Khalaf et al., 2021; Sharafati et al., 2020; Sim & Park, 2011), but there is still no law that allows predicting the relationship between accelerated and real values, with based on concrete composition data and environmental characteristics. However, there are some clues that indicate that a direct extrapolation of the results obtained in these tests to a natural exposure scenario cannot be made by simply applying a scale factor associated with the concentration ratio (Nguyen-Sy et al., 2020).

Concrete protection systems against carbonation currently available on the market are surface film products (paints) and modified cementitious mortars. These products are characterized by having high coefficients of resistance to penetration of. Some studies carried out on this subject (Vasanthalin & Kavitha, 2021; Penido et al., 2022; Zeng et al., 2022), reveal that these products may indeed have a beneficial effect against carbonation. Several mathematical and empirical models have been proposed in order to predict the progress of carbonation (Gandomi et al., 2014; Doddy et al., 2020).

Machine learning was used to model mud loss rate in 305 wells at Marun Oilfield, with data preprocessed using the Savitzky-Golay method (Jafarizadeh et al., 2023). The NSGA-II algorithm identified the most significant features, showing accuracy improvements plateauing beyond nine features.

Four ML algorithms predicted CO_2 storage mass and nine reservoir variables (Talkhouncheh et al., 2024). The Mahalanobis distance technique excluded 520 and 439 outlier records, respectively, enhancing prediction reliability.

Predictive models for confined compressive strength were developed using mud-logging data from two wells in southwest Iran (Davoodi et al., 2024). Data were preprocessed with the Tukey method to remove outliers, ensuring high accuracy and generalizability. ML algorithms were applied and assessed using training and test data subsets.

3. Methodology

The methods employed in this study aim to predict the concrete carbonation depth (C-Carb depth) using a hybrid intelligent model combining RF and LSSVM. This approach addresses the challenges of multiple indices and noise interference, with RF handling feature selection and LSSVM constructing the predictive model. Here's a detailed description of the methods:

3.1. Influence factor index system and sample data acquisition

The study uses 14 influencing factors (variables) related to concrete carbonation, including:

- Water-cement ratio (wcr);
- Compressive strength (CS);
- Cement content (cc);
- Fine and coarse aggregate contents (fasc, fagc);
- Water-reducing content (wrc);
- Water content (wc);
- Relative humidity (RH);
- Temperature (TEMP);
- Carbon dioxide concentration (C-CO_2);
- Carbonation time (Tcarb).

These factors are selected as input variables, while the output variable is the concrete carbonation depth. A dataset of 96 samples is collected, where each sample includes the values of the input variables and the corresponding carbonation depth. The dataset is split into two sets: 76 samples are used as the training set, and 20 samples as the test set.

3.2. Random forest (RF) regression model for feature selection

RF is used to identify the most important variables that influence concrete carbonation. The RF model divides the dataset into training and test sets and selects a subset of important features based on their contributions to reducing residual errors and increasing node purity. The RF model parameters are set to:

- mtry = 5 (number of variables sampled at each split);
- ntree = 800 (number of trees in the forest).

The importance of each variable is visualized through metrics like the mean square error reduction (%IncMSE) and increase in node purity (IncNodePurity). Variables with higher importance, such as CS, Tcarb, RH, TEMP, C-CO_2, and wc, are selected for further analysis. Pearson correlation analysis is also conducted to evaluate the correlation between these variables and C-Carb depth.

3.3. Least squares support vector machine (LSSVM) modeling and evaluation

After feature selection, LSSVM is used to build the predictive model. The selected variables from RF (e.g., Tcarb, CS, C-CO_2, RH, TEMP, etc.) are input into the LSSVM model. The LSSVM differs from traditional SVM by transforming quadratic programming problems into linear equations, reducing computational complexity and improving prediction accuracy.

The model is optimized using a GKF, which provides better performance and anti-interference capabilities. Parameter tuning is done using 10-fold cross-validation to find the best values for gamma and cost parameters, ensuring the model generalizes well to unseen data.

3.4. Model evaluation and validation

The RF-LSSVM hybrid model is evaluated using metrics such as RMSE and R-squared (R^2). These metrics assess how well the model predicts the C-Carb depth. The performance of the RF-LSSVM model is compared to other models, including:

- SVM without feature selection;
- Artificial Neural Networks (ANNs).

The RF-LSSVM model achieves the lowest RMSE (indicating the highest prediction accuracy) and the highest R-squared (indicating better fit) compared to the other models, demonstrating its superior performance.

3.5. Cross-validation and recursive feature elimination (RFE)

The study employs recursive feature elimination (RFE) to refine the model by systematically eliminating less important variables based on their impact on the prediction error. The optimal set of variables is determined through cross-validation, with the RF-LSSVM model ultimately including the most critical factors influencing concrete carbonation.

3.6. Comparison with other machine learning approaches

The study benchmarks the performance of the RF-LSSVM model against SVM and ANN models. The RF-LSSVM model consistently outperforms the other methods, as shown through a comparative analysis of their RMSE and R-squared values. Additionally, the hybrid model is evaluated for its ability to handle feature selection, reduce computational complexity, and improve predictive performance.

By leveraging the strengths of both RF for feature selection and LSSVM for modeling, the proposed methodology offers an effective tool for predicting concrete carbonation depth with high accuracy and reduced computational cost.

In this study, a hybrid RF and LSSVM model is developed for predicting concrete carbonation depth (C-Carb depth). The combination of RF and LSSVM is chosen to leverage the strengths of both models–RF for feature selection and LSSVM for efficient regression analysis. This methodology addresses key challenges such as dimensionality reduction and computational complexity, leading to enhanced predictive accuracy and model stability.

The process begins with the selection of 14 influencing factors, including water-binder ratio (wcr), compressive strength (CS), carbon dioxide concentration (C-CO_2), relative humidity (RH), temperature (TEMP), and others. A total of 96 monitored data sets are used for analysis, with 76 for training and 20 for testing. Feature selection is performed using RF, with variable importance evaluated through %IncMSE and IncNodePurity metrics. Critical variables like CS, Tcarb (carbonation time), RH, TEMP, and C-CO_2 are identified as the most influential on C-Carb depth.

After feature selection, the LSSVM model is built using the selected variables, optimized through a 10-fold cross-validation approach. The performance is further enhanced by tuning key parameters, such as gamma and cost, using an optimization grid. The results demonstrate that the RF-LSSVM model achieves the lowest RMSE and the highest R-squared value, significantly outperforming both traditional SVM without feature selection and ANNs. The RF-LSSVM model provides a precise prediction with an RMSE of 5e-5 and an R^2 of 0.999, while the SVM and ANN models show higher RMSE values and lower accuracy.

The findings highlight that factors such as carbonation time (Tcarb), compressive strength (CS), carbon dioxide concentration (C-CO_2), and RH have the strongest positive correlations with C-Carb depth. In contrast, variables like concrete specimen volume (csv) and water content exhibit negative correlations. These relationships are validated through Pearson correlation analysis and visualized for better interpretability.

Finally, the study concludes that the RF-LSSVM model not only enhances prediction accuracy but also reduces computational complexity, making it a superior method for predicting concrete carbonation depth compared to existing machine learning approaches. The model's strong performance suggests its potential for practical applications in the construction industry for durability assessments and material optimization.

4. Implementation of the proposed work

As depicted in Figure 1, the RF-LSSVM model predicts concrete ACP in this study. The first step involves constructing an index system for concrete ACP based on raw materials and mixing factors that influence it. Sample data for this index system is collected, and an original sample set is established. Each sample undergoes denoising processing, after which the original sample set is divided into a training dataset and a testing dataset according to a specified ratio.

The concrete ACP index system comprises factors that affect ACP and evaluation criteria for ACP. Material factors include the water-binder ratio, CS, cement dosage (cc), fly ash content (fasc), fine aggregate content (fagc), coarse aggregate content (cagc), water reducing agent content (wrc), and water content (wc). Environmental factors consist of relative humidity (RH), temperature (TEMP), and carbon dioxide concentration (C-CO_2). Test factors include concrete specimen volume (csv), carbonation time (Tcarb), and CS. The evaluation index for ACP is the concrete carbonation (C-Carb) depth value.

The influencing factors and ACP evaluation index form the samples in both the training and testing datasets. In total, there are 14 factors monitored to observe the distribution of C-Carb depth under different conditions. The

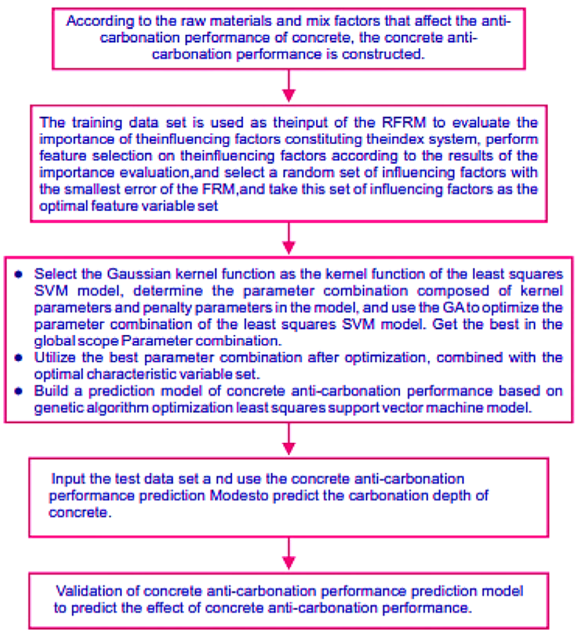

Figure 1. The flow chart of RF-LSSVM model prediction concrete ACP

training dataset, which constitutes 80% of the total sample data, is randomly selected for constructing the RF regression model and selecting the index set. The remaining 20% forms the test dataset used to evaluate the final model's predictive performance (Gandomi et al., 2014; Doddy et al., 2020).

As shown in Figure 1, RF-LSSVM model predicting concrete ACP that the embodiment of the present work provides, may further constructs the index system of concrete ACP as a first step according to the raw material that influences concrete ACP and mix factor, and collects the sample data of the index system of this concrete ACP, sets up original sample set, concentrates original sample Denoising processing is performed on each sample, and the original sample set after denoising processing is divided into a training data set and a testing data set according to a specified ratio. The index system of the ACP of the concrete includes factors affecting the ACP of the concrete and evaluation indexes of the ACP of the concrete. The material factors include: water-binder ratio, CS, cement dosage (cc), fly ash content (fasc), fine aggregate content (fagc), coarse aggregate content (cagc), water reducing agent content (wrc) and water content (wc), and the environmental factors include: relative humidity (RH), Temperature (TEMP) and carbon dioxide concentration (C-CO_2), the test factors include: concrete specimen volume (csv), carbonation time (Tcarb) and CS; the concrete carbonation (C-Carb) depth evaluation index is the C-Carb depth value. That is, based on the literature and engineering experience, select the main factors affecting the ACP of concrete, build an index system, collect corresponding sample data, and establish an original sample set. The above index system includes two parts: influencing factors and evaluation indexes of concrete ACP. Among them, the influencing factors include: influencing factors include material factors, environmental factors and test factors, including 14 factors in total, which are used to monitor the numerical distribution of C-Carb depth under different conditions; material factors include: wcr, CS, cc, fasc, fagc, cagc, wrc and wc; environmental factors include: RH, TEMP, C-CO_2; test factors include: csv, Tcarb and CS; the evaluation index of ACP is the C-Carb depth. Furthermore, each influencing factor of concrete ACP and concrete ACP evaluation index constitutes the samples in the training data set and the test data set, and the corresponding data is used as the sample data in the data set.

The ratio of the total number of samples in the training data set and the test data set is 2: 1~4: 1.

In this case, 80% of the total sample data is randomly used as the training data set for the construction of the RF regression model to select the index set; the remaining 20% is the test data set for the evaluation of the final model prediction performance (Mahmood & Mohammad, 2019; Najafgholipour et al., 2017).

With described training number set as the input of RF regression model, to carry out importance evaluation to the influence factor that constitutes described index system, carry out feature selection to influence factor according to the result of this importance evaluation, select the set of influencing factors with the smallest error of the RF regression model is obtained, and this set of influencing factors is used as the optimal feature variable set. Specifically, a RF regression model is constructed according to the number of features contained in the binary tree nodes in the RF regression model and the number of decision trees; the training data set is used as the input of the RF regression model, for each in the RF regression model A decision tree, using the OOB data corresponding to the decision tree to calculate its OOB data error OOBerr1; Randomly use a certain variable of all sample data in the OOB data corresponding to the decision tree as a feature X, and add noise to the feature X Interference, and then calculate the corresponding OOB data error OOBerr2 of the decision tree again; build an importance calculation model, and evaluate the variable importance of the above-mentioned feature X according to the importance calculation model; output the variables of all variables in the training data set Importance evaluation, and then draw a visual drawing of the variable importance evaluation, and arrange the variable importance evaluations of all variables in descending order, and perform a preliminary screening of the variable importance measurement according to the sorting results; for the variable set obtained after the preliminary screening. Use the RFE method to remove the variable of the specified proportion from the variable set one by one, and get one variable each time, compare the OOB error rate corresponding to the remaining variables after removing the variable, and use the variable set with the smallest error rate as the optimal feature variable set, and determine the number of optimal features in the optimal feature variable set. Among them, when calculating the OOB data error, the OOB data error obtained in Bootsrap sampling is calculated. When evaluating the importance of variables with variables in descending order, the importance of each variable is preliminarily measured by the reduction of the visual indicator mean square residual (%IncMSE) and the reduction of model accuracy (IncNodePurity). Sex is used as the importance evaluation of the corresponding variable, and the variable importance evaluations of all variables are arranged in descending order. In this second step mentioned, the determining method of optimal feature variable set comprises two input parameters of RF: the characteristic number mtry that binary tree node comprises and the tree number Ntree of decision tree, mtry = P/3 (regression model) under the default situation, P is variable number, Ntree = 500; Model and train the data set; secondly, by calculating the OOB data error obtained in Bootsrap sampling, to visualize the reduction of the indicator mean square residual (%IncMSE) and the reduction of model accuracy (IncNodePurity). Initially measure the importance of each variable and arrange them in descending order. Further, described original sample set is used as the input of RF model, carries out variable importance evaluation through RF model training, carries out feature selection to input variable by RFE method.

Selecting the optimal feature variable set with the smallest model error to realize RF dimensionality reduction includes variable importance evaluation and important variable screening (Wu et al., 2021; Jumaa'h et al., 2019; Nafees et al., 2021).

4.1. Variable importance evaluation

For each tree in RF, OOB is used to calculate its error, be OOBerr1; Noise interference is added to the feature X of all samples of OOB data randomly, and OOB data error again is calculated (OOBerr2); if there are Ntree trees in the RF, the the importance of feature X is calculated using the equation:

$$Importance = \sum \left(OOBerr2 - OOBerr1 \right) / Ntree, \tag{1}$$

where *Ntree* is the tree of decision tree in RF regression model.

4.2. Important variable screening

In the visualization of variable importance scores, the scores are arranged in descending order to determine the most influential factors in the feature set. The process utilizes RFE to iteratively remove less important variables from the set, comparing the corresponding OOB error rates for each reduced set. The set with the lowest error rate is selected as the optimal feature set, determining the optimal quantity of features.

For the third step, the GKF is chosen as the kernel function for the least squares SVM model. The parameters of the kernel function and the penalty parameter in the model are determined through a GA to find the optimal parameter combination globally. Initially, the SVM parameters are encoded and an initial population is randomly generated. The GA process includes setting the population size, termination criteria based on evolution iterations, crossover and mutation probabilities, and initializing the parameter combination within specified ranges.

Next, the least squares SVM model is established using the kernel width parameter and penalty parameter, trained with the training dataset, and evaluated using RMSE as the fitness function for each individual in the population. The GA proceeds by selecting individuals based on fitness using the roulette wheel selection rule, performing crossover to generate new individuals, and introducing mutation with a defined probability. This iterative process continues until convergence criteria are met, optimizing to find the individual with the highest fitness as the optimal solution or until the maximum number of iterations is reached.

In the process of weighing the decision function of the accuracy of the least squares SVM model with the fitness function of each individual of the population with the root mean square, by using the concrete raw material and the range of values of the mixing ratio as a constraint condition, it is guaranteed that each Reasonableness of the sample. The calculation model of the constraints is: (Chang & Chen, 2016).

$$\begin{cases} 56 \le X_4 \le 98 \\ 620 \le X_5 \le 860 \\ 1030 \le X_6 \le 1150 \\ 0.9 \le X_7 / (X_3 + X_4 + X_8) \le 1.6 \\ 18 \le X_8 \le 28 \\ 0.46 X_2 (X_1 + X_4 + X_8 / X_3 - 0.2 - f_{cu,k} - 8.225 \ge 0 , \end{cases} \tag{2}$$

where X_1 is wcr, X_3 is cement amount, X_4 is fly ash, X_5 is fine aggregate, X_6 is coarse aggregate, X_8 is silica fume consumption, $f_{cu,\,k}$ is concrete cube compressive strength (CS) standard value.

As another parameter optimization scheme of the present work, by determining kernel function, selection parameter, set up least squares SVM (LSSVM) model, the optimal feature set that step 2 obtains is used as the input variable of LSSVM, Train the sample data, output the prediction result of C-Carb depth, and use the test set to verify the prediction result of the model. Concretely: build the least squares SVM model, use the optimal feature variable set as the input variable of the least squares SVM model, and the corresponding C-Carb depth value as the output variable, to the LVSS. The machine model is trained, and then the test data set is used to verify the prediction result of the trained least squares SVM model on the ACP of concrete. That is, the GKF is selected to construct the least squares SVM model; the optimal feature variable set is used as the input variable of the least squares SVM model, and the corresponding C-Carb depth value is used as the output variable, and the ten-fold cross-validation is used to carry out Parameter optimization to determine the optimal parameter combination of the penalty parameter C of the least squares SVM model and the GKF kernel width parameter σ^2; the least squares SVM model using the optimal parameter combination based on the optimal feature variable set. Predict the test data set and output the prediction results to verify the carbonation resistance of concrete predicted by the trained least squares SVM model (Ibrahim et al., 2019).

More specifically, at first, the GKF (RBF) that selection mapping ability is strong, generalization performance is excellent, applicability is good establishes.

The least squares SVM (LSSVM) model, the GKF expression is as follows:

$$K\left(x_i, x_j\right) = \exp\left(-\frac{\left\| x - x_i \right\|^2}{2\sigma^2} \right), \tag{3}$$

where x is the variable of input, and x_i is the ith sample, and x_j is the jth sample, and σ^2 is the kernel width parameter.

Next, adopt ten-fold cross validation to carry out parameter optimization, determine the penalty parameter of model and the optimal parameter combination of GKF kernel width parameter σ^2.

Finally, based on the optimal feature set, the sample training set and the test set are trained and predicted, and the output prediction results are represented by the actual value and the predicted value fitting curve.

In fourth step of the current work, input test number set and utilize concrete ACP prediction model to predict C-Carb depth value, and verify the effect of concrete ACP prediction model of described concrete ACP prediction model. That is to analyze the error of the prediction results, select the SVM and the ANN without feature selection for modeling, and use the same error index for comparative analysis to verify the applicability and superiority of the model. That is to build the calculation model of the RMSE and R^2 of the model performance parameters, and use the SVM model without the feature selection of the influencing factors and the prediction results of the ANN prediction model and the result of the feature selection of the influencing factors. The prediction results of the concrete ACP prediction model are analyzed for errors, and the effect of the concrete ACP prediction model for predicting the C-Carb resistance is verified (Khalaf et al., 2021).

Concrete, at first, select model performance parameter *RMSE*, R^2 to evaluate the predictive accuracy of model, to the further verification of output fitting curve effect, and expression is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y^o - y^p)^2}{n}}; \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y^o - y^p)^2}{\sum_{i=1}^{n} (y^o - \bar{y}^o)^2}, \tag{5}$$

where y^o is sample data observation value, and y^p is model prediction value, and n is sample number.

Next, select the SVM that does not carry out feature selection and the result of ANN prediction model and do comparative analysis with it, affirmed the effectiveness and the correctness of RF feature screening again.

4.3. Control parameters for each of the algorithm

The control parameters for each algorithm separately as below:

- Index System and Data Collection: You've established an index system for concrete ACP (Alkali Carbonate Reaction Potential), incorporating factors like water-binder ratio, cement dosage, environmental factors (e.g., RH, temperature, CO_2 concentration), and test factors. Sample data was collected and processed through denoising, splitting into training and testing sets.
- RF Regression Model: The RF regression model was employed for feature selection. Features impacting concrete ACP depth were evaluated based on the RF model's importance metrics (%IncMSE, IncNodePurity). This step helped identify key variables influencing carbonation depth.
- Variable Importance Evaluation: Through iterative processes like RFE, you narrowed down to an optimal set of features (e.g., Tcarb, CS, C-CO_2, RH, TEMP, fasc, cement content, water-binder ratio) that minimized error rates, thereby enhancing model accuracy.
- LSSVM Model Construction: A GKF was selected for the Least Squares SVM (LSSVM) model, optimized using genetic algorithms to determine the best parameter combinations for predicting carbonation depth accurately.
- Model Evaluation: The trained RF-LSSVM model demonstrated superior performance compared to SVM without feature selection ANNs, as evidenced by lower RMSE (Root Mean Square Error) and higher R^2 (coefficient of determination) values. This validates the efficacy of your proposed approach in predicting concrete ACP reliably.

4.4. Peculiarity of their methods which used in this research

In the field of predicting the Accelerated Carbonation Performance (ACP) of concrete, the methods utilized in the research are characterized by several distinctive features:

- Integration of RF and LSSVM: The research combines RF for feature selection and LSSVM for modeling. This integration leverages RF's ability to handle complex datasets with LSSVM's robust regression capabilities, aiming to improve predictive accuracy while reducing computational complexity compared to traditional SVM approaches.
- Comprehensive Feature Selection: The study employs a detailed feature selection process using RF. RF is utilized to evaluate the importance of each feature (influencing factors such as water-binder ratio, cement dosage, environmental factors like relative humidity and temperature, etc.), helping to identify the most critical variables that affect concrete ACP. This approach ensures that only the most relevant factors are included in the predictive model, enhancing its efficiency and accuracy.
- Optimization Using Genetic Algorithms: Genetic algorithms are employed to optimize the parameters of the LSSVM model. This method systematically searches for the best combination of kernel parameters and penalty factors, ensuring that the SVM model is fine-tuned to provide the most accurate predictions of concrete carbonation depth. The use of genetic algorithms helps in finding global optima efficiently, which is crucial for enhancing model performance.
- Cross-validation and Error Analysis: The research applies rigorous cross-validation techniques (e.g., 10-fold cross-validation) to assess model performance objectively. This process helps in validating the predictive capability of the developed models across different datasets, ensuring their robustness and reliability. Additionally, error analysis metrics such as RMSE and coefficient of determination (R-squared) are used to quantitatively evaluate and compare the accuracy of different models, providing insights into their predictive efficacy.

The methods were validated internally through rigorous data preprocessing, model selection, feature selection, and evaluation using cross-validation techniques. This ensures that the models are robust and not overfitting to the training data.

5. Results and discussion

The methodology employed in this study utilizes RF to effectively handle sample data with multiple indices and noise interference, identifying pertinent feature variables to enhance the predictive accuracy of the LSSVM model. This approach results in more precise and reliable predictions of C-Carb depth. Moreover, the innovative method integrates RF with Least Squares SVM to establish a hybrid intelligent model, leveraging the strengths of different algorithms while mitigating shortcomings of existing methods. This hybrid model reduces model dimensionality, speeds up training, and resolves issues related to unstable prediction results, offering a rapid and effective tool for concrete ACP prediction.

The study specifically selects Least Squares SVM for constructing the predictive model, improving upon traditional SVM approaches. The RF-LSSVM model transforms complex quadratic programming problems into solutions of linear equations, thereby reducing computational complexity and minimizing error rates, leading to enhanced regression problem solving. The efficacy and correctness of RF feature screening are validated through error analysis. Additionally, GKF is chosen for its superior performance and simultaneous advantages of radial basis kernel functions, along with robust anti-interference capabilities.

The proposed methodology for predicting concrete ACP encompasses the acquisition of sample data for the influence factor index system, RF regression model feature selection, and the modeling and evaluation of the Least Squares SVM (Sharafati et al., 2020).

5.1. Influence factor index system sample data acquisition

Fourteen influencing factors–such as the water-binder ratio (wbr), compressive strength (CS), carbon dioxide concentration (C-CO_2), relative humidity (RH), temperature (TEMP), and others–are utilized as input variables. The output is the C-Carb depth. The dataset comprises 96 monitored samples, and a snapshot of the data is provided in Table 1, showing critical variables affecting concrete carbonation.

5.2. RF regression model feature selection

The dataset is split into a training set of 76 samples and a test set of 20. The RF regression model is trained with parameters such as *mtry* = 5 and *Ntree* = 800. The importance of each variable is visualized using metrics like %IncMSE and IncNodePurity, indicating that factors such as carbonation time (Tcarb), CS, C-CO_2, RH, and TEMP have a strong influence on the depth of carbonation. The Pearson correlation analysis confirms these findings, showing strong positive correlations for some variables (e.g., C-CO_2 and RH) and negative correlations for others (e.g., Tcarb).

The RF program package in R software was utilized to visualize the Importance function evaluation index, as depicted in Figure 2. The variables are arranged in descending order of importance, with a larger increase in node purity (IncNodePurity) indicating stronger variable importance. Similarly, a higher value of mean square error reduction (IncMSE) signifies greater variable importance.

From Figure 2, it is evident that factors such as CS, carbonation time (Tcarb), relative humidity (RH), temperature (TEMP), carbon dioxide concentration (C-CO_2), water content (wc), fly ash content (fasc), water-binder ratio (wbr), and concrete specimen volume (csv) exhibit relatively high importance. This indicates that these variables significantly influence the depth of concrete carbonation (Sim & Park, 2011).

Table 1. Sample data

csv m^3	wcr	CS MPa	cc kg/ m^3	fasc kg/ m^3	fagc kg/ m^3	cagc kg/ m^3	wrc (%)	wc kg/ m^3	RH (%)	TEMP (C)	C-CO_2	Tcarb d	CS MPa	C-Carb depth mm
0.0032	0.6	52.5	270	0	748	1220	2.7	160	70	20	20	28	46.8	6.6
0.0032	0.6	52.5	243	42.5	742	1212	2.84	160	70	20	20	28	43.5	10.4
0.0032	0.6	52.5	189	121.5	732	1194	3.11	160	70	20	20	28	33.4	18
0.004	0.45	52.5	280	120	714	1165	2.4	180	70	20	20	7	32	6.1
0.004	0.45	52.5	280	120	714	1165	2.4	180	70	20	20	28	38	7.9
0.004	0.45	52.5	200	200	714	1165	2.4	180	70	20	20	7	23	11.2
0.004	0.45	52.5	200	200	714	1165	2.4	180	70	20	20	28	50.5	12.5
0.004	0.6	42.5	295	0	729	1106	2.95	177	63	19	0.03	28	30.2	1.5
0.004	0.6	42.5	295	0	729	1106	2.95	177	63	19	0.03	60	32.9	2.1
0.004	0.6	42.5	295	0	729	1106	2.95	177	63	19	0.03	90	34.8	2.5
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
0.003	0.45	42.5	520	0	525	1020	0	234	70	20	20	28	58	1.7
0.003	0.45	42.5	607	0	464	901	0	273	70	20	20	28	46.5	8.7
0.003	0.45	42.5	693	0	403	782	0	312	70	20	20	28	43.5	9

The Pearson correlation function in R software was employed to analyze the correlation between influencing factors and C-Carb depth, validating the previously mentioned order of importance. Using the ggplot2 program package, Figure 3 illustrates the results. The Pearson correlation coefficient ranges from –1 to 1, where darker and larger circles indicate stronger absolute correlation coefficients between variables, and vice versa for weaker correlations.

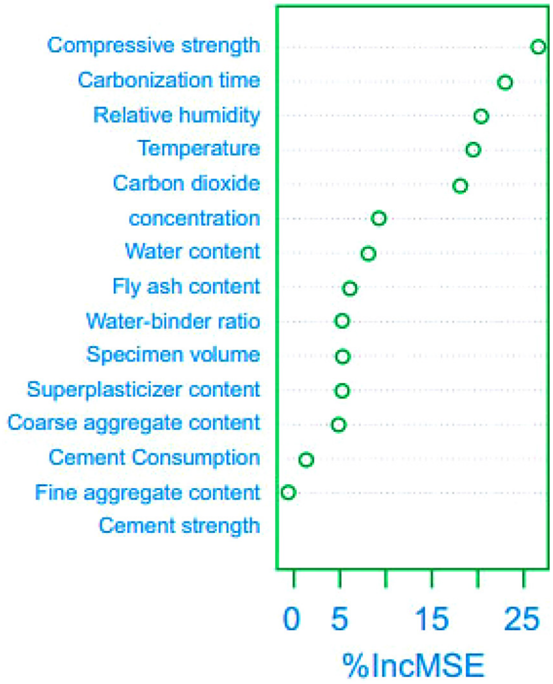

Figure 2. Schematic diagram of each variable with the visual index mean square residual error reduction (%IncMSE)

From Figure 3, it is evident that factors such as carbon dioxide concentration (C-CO_2), relative humidity (RH), and temperature (TEMP) exhibit a strong positive correlation with C-Carb depth. Conversely, variables like concrete specimen volume (csv), carbonation time (Tcarb), and water reducing agent (wrc) show a significant negative correlation with C-Carb depth, consistent with the importance ranking diagram.

Additionally, to validate the predictive performance, the RMSE and R^2 values for different variable combinations were obtained through 10 repetitions of 5-fold cross-validation, as summarized in Table 2 (Nguyen-Sy et al., 2020).

From the table, it is analysed that,

- The RMSE values range from 2.5902.5902.590 to 4.3324.3324.332. A lower RMSE suggests that the model predictions are closer to the actual observed values. Therefore, variables 9 to 14 generally have lower RMSE values, indicating better model performance in terms of prediction accuracy compared to variables 1 to 8.
- The R-squared values range from 0.1340.1340.134 to 0.6180.6180.618. A higher R-squared value indicates that a larger proportion of the variance in the

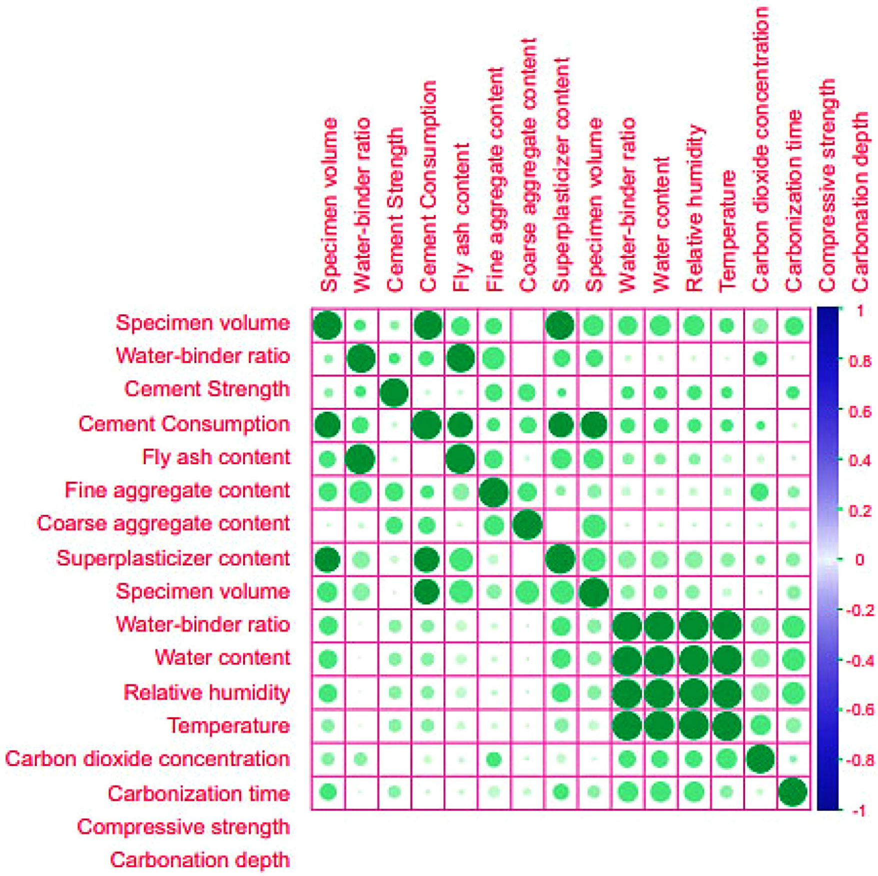

Figure 3. Influence factor correlation size figure

Table 2. RMSE and R^2 change table for different variable combinations

Variables	RMSE	R-squared
1	4.332	0.134
2	3.454	0.295
3	2.640	0.610
4	2.730	0.573
5	2.804	0.551
6	2.655	0.594
7	2.692	0.588
8	2.714	0.584
9	2.590	0.618
10	2.608	0.611
11	2.643	0.602
12	2.607	0.608
13	2.623	0.600
14	2.656	0.590

dependent variable is explained by the independent variables. Variables 3, 9, and 10 show higher R-squared values, suggesting these variables provide a better fit to the data and explain more variability in the dependent variable compared to others.

To visualize the change in RMSE according to Table 2, Figure 4 is drawn.

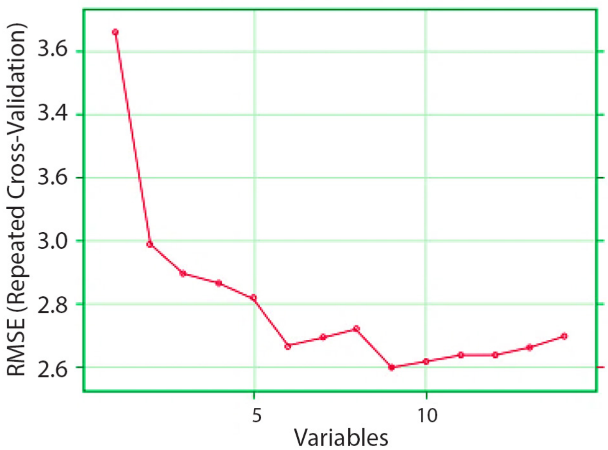

Figure 4. RMSE change trend figure when different variable combinations are involved

The RMSE values for the variables analyzed in this study (Figure 4) vary across a range of 2.6 to 3.7. Variable 1 had the highest RMSE at 3.7, indicating a larger discrepancy between predicted and actual values compared to other variables. Variables 6, 10, 11, and 12 exhibited the lowest RMSE values, ranging from 2.63 to 2.64, suggesting higher accuracy in prediction for these variables. Variables 3, 7, and 14 showed moderate RMSE values around 2.7, while variables 2, 4, 5, 8, and 13 had RMSE values between 2.65 and 2.81, indicating intermediate predictive performance. These RMSE values provide insights into the accuracy and variability of predictions across the 14 influencing factors considered in the analysis.

When the number of variables reaches 9, the RMSE value reaches its minimum, indicating the highest model accuracy at this point. Utilizing the RFE method based on importance ranking, the model is systematically built through cross-combination, progressively eliminating unimportant variables until all feature variables are evaluated. This process culminates in identifying the optimal set of variables.

The optimal index set includes variables such as carbonation time (Tcarb), compressive strength (CS), carbon dioxide concentration (C-CO_2), relative humidity (RH), temperature (TEMP), fine aggregate content (fasc), cement content (cc), water-binder ratio, and water content. These variables are selected to construct the least squares SVM model, ensuring optimal predictive performance (Nguyen-Sy et al., 2020).

5.3. Least squares SVM modeling and evaluation

Randomly selecting 76 data groups as the SVM training set and the remaining 20 groups as the test set, we utilize nine influencing factors–Tcarb, CS, C-CO_2, RH, TEMP, fly ash content, cement dosage, water-binder ratio, and water content–as input variables for the model. The output variable is C-Carb depth. We load the e1071 version 1.6-7 package in R language and optimize parameters using the 10-fold cross-validation (CV) method. The optimization results are then outputted as shown in Table 3.

Table 3. Output optimization results

S.No.	gamma	cost	error
1	1e-04	0.1	17.48027
2	1e-03	0.1	17.04788
3	1e-02	0.1	14.59542
4	1e-01	0.1	11.12708
5	1e-00	0.1	14.29068
6	1e-04	1	17.04084
7	1e-03	1	14.36732
8	1e-02	1	9.24183
9	1e-01	1	6.64675
10	1e-00	1	9.80437
11	1e-04	10	14.35399
12	1e-03	10	9.30037
13	1e-02	10	4.82437
14	1e-01	10	4.29084
15	1e-00	10	8.57455
16	1e-04	100	9.32133
17	1e-03	100	5.67290
18	1e-02	100	3.46779
19	1e-01	100	4.65478
20	1e-00	100	8.57457

The gamma values range from 10−410^{-4}10−4 to 111, across different scales (1, 10, 100), which suggests a logarithmic scale exploration (e.g., 10−410^{-4}10−4, 10−310^{-3}10−3, 10−210^{-2}10−2, 10−110^{-1}10−1, 111).

The cost values also vary across powers of ten (1, 10, 100), which is typical in SVM parameter tuning to explore a wide range of regularization strengths.

Error Values are the results obtained from applying each combination of gamma and cost parameters. Lower error values typically indicate better model performance, though the specific interpretation depends on the context of the problem (e.g., classification accuracy, mean squared error).

Observing the data, it is noted that,

- Higher values of gamma (e.g., 111 and 100100100) generally result in lower errors across different cost values, indicating that in this dataset, higher gamma values might be more suitable.

As cost increases (from 0.10.10.1 to 100100100), errors generally decrease, suggesting a preference for higher regularization strengths in this specific context.

The result of parameter optimization is: best c = 0.01, best g = 100, best performance = 3.467780 (minimum), input this parameter and set up the least squares SVM model, training fitting result to training sample as shown in Figure 5, the prediction results for the test samples (Figure 6). It can be seen (Figure 5) that the simulation value is very close to the actual value, and the simulation effect is better. Use the trained RF model to predict the test set (Figure 6), it can be seen that the predicted value curve on the test set of the RF model is closer to the real value (Vasanthalin & Kavitha, 2021; Penido et al., 2022).

The provided data (Figure 5) represents the concrete carbonation depth values along with the corresponding number of samples for each depth value. Concrete carbonation depth refers to the penetration depth of carbon dioxide into concrete, which can affect its durability and structural integrity over time.

The dataset appears to be structured with two columns:

- Actual Concrete carbonation depth value: This column lists the actual depth values in centimeters where carbonation has occurred within the concrete structure.
- No. of samples: This column indicates the number of instances or samples where the carbonation depth matches the corresponding value listed.

Interpretation and Analysis:

The dataset shows a range of carbonation depths from 0 cm to 73 cm. Here are some observations and insights:

- Distribution of Carbonation Depths: The data indicates that carbonation depths vary widely, with higher frequencies observed in the mid-range depths (around 4 cm to 14 cm) and some peaks at higher values (around 22 cm to 24 cm).

Implications for Concrete Durability: Deeper carbonation can weaken concrete structures over time, affecting

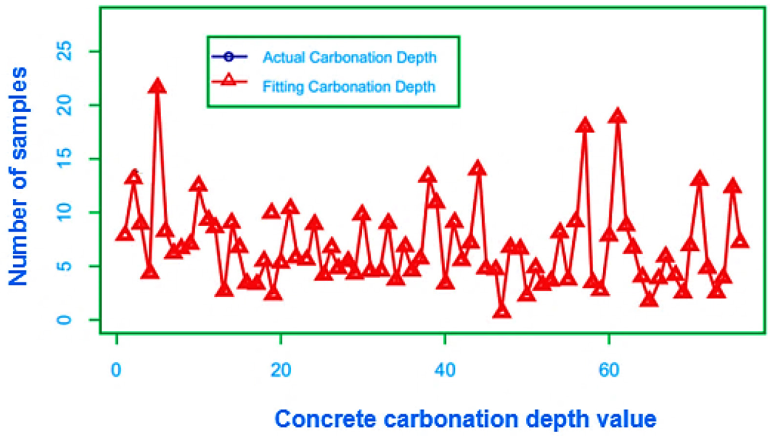

Figure 5. RF-LSSVM model training number set fitting result

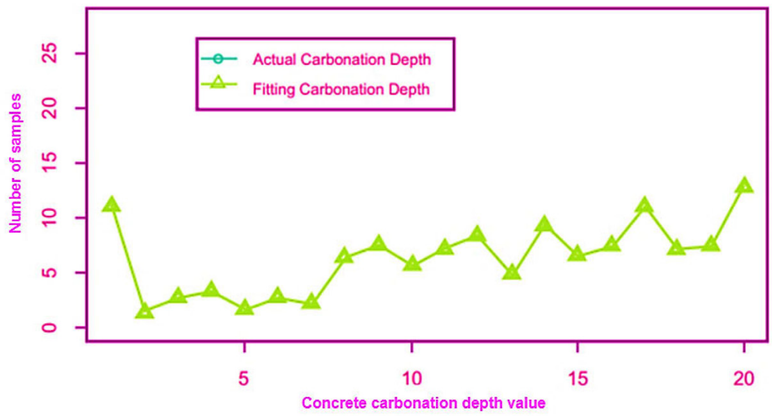

Figure 6. LSSVM model test number set prediction result

their ability to withstand environmental stresses and reducing their lifespan.

The data (Figure 6) presented details the relationship between the actual concrete carbonation depth and the fitting carbonation depth, along with the number of samples at each depth.

At a carbonation depth of 0 mm, there are 11 samples where the actual carbonation depth is also 0 mm. As the carbonation depth increases, the number of samples fluctuates. For example, at carbonation depth of 1 mm, there are 2 samples, and at 2 mm, there are 3 samples. The number of samples continues to increase with the carbonation depth up to 10 mm, where it reaches a peak of 10 samples.

Beyond 10 mm, the number of samples starts to show variability again. At 11 mm, there are 5 samples, while at 12 mm, there are 10 samples. The number of samples at 13 mm and 14 mm is 6 each, and at 15 mm, there are 11 samples. The trend continues with 5 samples at 16 mm, 6 samples at 17 mm, and 7 samples at 18 mm and 19 mm. Finally, at a carbonation depth of 20 mm, there are 13 samples.

This distribution suggests a varied data set with some depth intervals having significantly more samples than others, indicating possible areas of greater or lesser interest or data collection intensity in the study of concrete carbonation.

Using the optimized influencing factors, a Least Squares SVM (LSSVM) model is built, with parameter tuning performed using 10-fold cross-validation. Key parameters (gamma and cost) are varied across a wide range, and the optimal combination (best gamma = 100, best cost = 0.01) yields the lowest prediction error, indicating that the LSSVM model achieves a good fit with the data.

5.4. Prediction result evaluation

In order to check the superiority of RF fusion least squares SVM model (RF-LSSVM), select the SVM that does not carry out feature selection to model and do comparative analysis with ANN, select Equation (3) The RMSE and the certainty coefficient R^2 of the Equation (4) are used to measure the prediction accuracy of the model, and the error comparison results (Table 4) (Zeng et al., 2022; Gandomi et al., 2014).

Table 4. Comparison of error of the competent methods

Model	Performance	
	RMSE	R^2
RF-LSSVM	5e-5	0.999
SVMs (without feature selection)	0.984	0.896
ANNs	3.57	0.812

The results demonstrate that the RMSE of the RF-LSSVM model prediction is closest to 0, and the determination coefficient σ^2 is closest to 1. This indicates that the prediction accuracy of the RF-LSSVM model is superior to that of the simple SVM and artificial neural network models. Therefore, the RF-LSSVM model shows a promising application prospect in the research field of predicting the ACP of concrete based on materials and mix ratios. This structured comparison highlights the distinct features and relative advantages of the proposed RF-LSSVM approach compared to other ML algorithms in predicting concrete carbonation depth, as shown in Table 5.

Table 5. Comparison of machine learning approaches for predicting concrete carbonation depth

Aspect	Proposed work (RF-LSSVM)	Comparison with other ML algorithms
Dataset Characteristics	96 samples, 14 influencing factors (e.g., water-binder ratio, RH, TEMP)	Typically includes multiple influencing factors, exact details not specified
Machine Learning Models	RF for feature selection, LSSVM optimized with GKF	SVM, ANNs, Decision Trees, Random Forests, etc.
Feature Selection Approach	RF for identifying critical variables	PCA, Recursive Feature Elimination, Lasso Regression, etc.
Model Training and Optimization	Genetic algorithm for LSSVM optimization	Grid Search, Bayesian Optimization, Gradient Descent, etc.
Evaluation Metrics	RMSE, R-squared	MAE, Accuracy, Precision, Recall, F1-score, etc.
Cross-Validation Technique	10-fold cross-validation	Leave-One-Out, Stratified K-Fold, Nested Cross-Validation, etc.
Results	RF-LSSVM outperformed SVM and ANNs in terms of RMSE, R-squared	Comparative performance metrics across different algorithms
Visualization and Correlation Analysis	Visualized variable importance and correlations	Similar techniques used for interpretability and feature analysis
Validation (Internal)	Rigorous data preprocessing, model selection, cross-validation	Similar practices for ensuring model robustness and reliability

The study has limitations. Firstly, the dataset used consists of 96 samples, which may limit the generalizability of the findings to broader applications within the concrete industry. While RF was employed for feature selection, further exploration of its parameters and comparison with alternative methods could refine the variable selection process. Moreover, the model's performance across varying concrete compositions, environmental conditions, and geographical locations requires more extensive validation to ensure robustness and applicability. Although evaluation metrics such as RMSE and R-squared were utilized, additional metrics could provide a more comprehensive assessment of the model's predictive accuracy and reliability. Furthermore, enhancing the interpretability of the model's predictions and the underlying reasons for these predictions through explainable AI techniques could improve trust and usability in practical settings. Lastly, validating assumptions about input variables and their relationships with concrete carbonation depth through rigorous sensitivity analyses and validation studies would strengthen the study's conclusions and practical relevance.

The RF-LSSVM model's performance is superior compared to SVM without feature selection and ANNs. As shown in Table 4, the RF-LSSVM achieves an RMSE closest to 0 and an R^2 value closest to 1, suggesting that it offers the most accurate predictions of concrete carbonation depth among the models considered. The structured comparison in Table 5 also highlights the strengths of the RF-LSSVM approach, including its use of RF for feature selection, optimization via genetic algorithms, and high prediction accuracy across evaluation metrics like RMSE and R^2.

Hence the hybrid RF-LSSVM model proves to be a robust and reliable approach for predicting concrete carbonation depth, significantly outperforming traditional SVM and ANN models.

6. Conclusions

The RF-LSSVM model developed in this study represents a notable advancement in predicting concrete ACP, particularly in assessing carbonation depth. By combining the strengths of RF for feature selection with LSSVM for accurate modeling, this hybrid approach addresses the inherent complexities in concrete data analysis more effectively than traditional methods.

The quantitative results of the RF-LSSVM model are compelling, demonstrating a reduced RMSE of 5e-5 and a high R-squared value of 0.999. These metrics underscore the model's robustness and superior predictive accuracy. The RF-LSSVM model's performance outstrips that of conventional SVMs and ANNs, highlighting its potential for providing precise and reliable predictions crucial for concrete durability assessments.

Beyond enhancing prediction accuracy, the RF-LSSVM model significantly improves computational efficiency. The integration of RF for feature selection simplifies the model by reducing dimensions and optimizing training times. This efficiency is crucial in practical scenarios where timely and accurate assessment of concrete ACP is essential for informed decision-making in construction and infrastructure projects. The model's ability to provide rapid predictions can aid engineers and construction professionals in making better choices about materials and maintenance strategies, thus enhancing the overall durability and safety of concrete structures.

Additionally, the comprehensive feature analysis and correlation studies conducted as part of this research have provided valuable insights into the factors influencing concrete carbonation. By identifying and quantifying these factors, the study contributes to a deeper understanding of carbonation mechanisms and enriches the knowledge base in the field of construction material science. This detailed analysis not only aids in predicting ACP but also informs the development of more effective protective measures and treatments for concrete.

The success of the RF-LSSVM model in this study underscores its potential as a powerful tool for advancing concrete durability assessments. Future research could further refine this model and explore its application to a wider range of concrete types and environmental conditions. Overall, this innovative approach offers significant benefits for optimizing concrete durability assessments, contributing to the development of more resilient and long-lasting infrastructure.

References

Abad, A., R. B., Ghorbani, H., Mohamadian, N., Davoodi, S., Mehrad, M., Khezerloo-ye Aghdam, S., & Nasriani, H. R. (2022). Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields. *Fuel, 308*, Article 121872. https://doi.org/10.1016/j.fuel.2021.121872

Al-Musawi, A. A., Alwanas, A. A. H., Salih, S. Q., Ali, Z. H., Tran, M. T., & Yaseen, Z. M. (2020). Shear strength of SFRCB without stirrups simulation: Implementation of hybrid artificial intelligence model. *Engineering with Computers*, *36*(1), 1–11. https://doi.org/10.1007/s00366-018-0681-8

Anemangely, M., Ramezanzadeh, A., Amiri, H., & Hoseinpour, S.-A. (2019). Machine learning technique for the prediction of shear wave velocity using petrophysical logs. *Journal of Petroleum Science and Engineering, 174*, 306–327. https://doi.org/10.1016/j.petrol.2018.11.032

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., & Inman, D. J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications. *Mechanical Systems and Signal Processing*, *147*, Article 107077. https://doi.org/10.1016/j.ymssp.2020.107077

Barjouei, H. S., Ghorbani, H., & Mohamadian, N., Wood, D. A., Davoodi, S., Moghadasi, J., & Saberi, H. (2021). Prediction performance advantages of deep machine learning algorithms for two-phase flow rates through wellhead chokes. *Journal of Petroleum Exploration and Production Technology, 11*, 1233–1261. https://doi.org/10.1007/s13202-021-01087-4

Chaabene, W. B., & Nehdi, M. L. (2020). Novel soft computing hybrid model for predicting shear strength and failure mode of SFRC beams with superior accuracy. *Composites Part C: Open Access*, *3*, Article 100070. https://doi.org/10.1016/j.jcomc.2020.100070

Chang, C. F., & Chen, J. W. (2006). The experimental investigation of concrete carbonation depth. *Cement and Concrete Research*, *36*, 1760–1767. https://doi.org/10.1016/j.cemconres.2004.07.025

Chou, J. S., Pham, T. P. T., Nguyen, T. K., Pham, A. D., & Ngo, N. T. (2020). Shear strength prediction of reinforced concrete beams by baseline, ensemble, and hybrid machine learning models. *Soft Computing*, *24*(5), 3393–3411. https://doi.org/10.1007/s00500-019-04103-2

Davoodi, S., Mehrad, M., Wood, D. A., Ghorbani, H., & Rukavishnikov, V. S. (2023a). Hybridized machine-learning for prompt

prediction of rheology and filtration properties of water-based drilling fluids. *Engineering Applications of Artificial Intelligence, 123*, Article 106459.
https://doi.org/10.1016/j.engappai.2023.106459

Davoodi, S., Mehrad, M., Wood, D. A., Rukavishnikov, V. S., & Bajolvand, M. (2023b). Predicting uniaxial compressive strength from drilling variables aided by hybrid machine learning. *International Journal of Rock Mechanics and Mining Sciences, 170*, Article 105546. https://doi.org/10.1016/j.ijrmms.2023.105546

Davoodi, S., Vo Thanh, H., Wood, D. A., Mehrad, M., & Rukavishnikov, V. S. (2023c). Combined machine-learning and optimization models for predicting carbon dioxide trapping indexes in deep geological formations. *Applied Soft Computing, 143*, Article 110408. https://doi.org/10.1016/j.asoc.2023.110408

Davoodi, S., Vo Thanh, H., Wood, D. A., Mehrad, M., Al-Shargabi, M., & Rukavishnikov, V. S. (2023d). Machine-learning models to predict hydrogen uptake of porous carbon materials from influential variables. *Separation and Purification Technology, 316*, Article 123807.
https://doi.org/10.1016/j.seppur.2023.123807

Davoodi, S., Vo Thanh, H., Wood, D. A., Mehrad, M., Al-Shargabi, M., & Rukavishnikov, V. S. (2024). Machine learning insights to CO_2-EOR and storage simulations through a five-spot pattern – a theoretical study. *Expert Systems with Applications, 250*, Article 123944. https://doi.org/10.1016/j.eswa.2024.123944

Davoodi, S., Vo Thanh, H., Wood, D. A., Mehrad, M., Rukavishnikov, V. S., & Dai, Z. (2023e). Machine-learning predictions of solubility and residual trapping indexes of carbon dioxide from global geological storage sites. *Expert Systems with Applications, 222*, Article 119796.
https://doi.org/10.1016/j.eswa.2023.119796

Doddy, P., Cheng, M. Y., Wu, Y. W., & Duc-Hoc, T. (2020). Combining machine learning models via adaptive ensemble weighting for prediction of shear capacity of reinforced-concrete deep beams. *Engineering with Computers*, *36*(3), 1135–1153.

España, R. M., Hernández-Díaz, A. M., Cecilia, J. M., & García-Román, M. D. (2017). Evolutionary strategies as applied to shear strain effects in reinforced concrete beams. *Applied Soft Computing*, *57*, 164–176. https://doi.org/10.1016/j.asoc.2017.03.037

Gandomi, A. H., Mohammadzadeh, S., Pérez-Ordóñez, J. L., & Alavi, A. H. (2014). Linear genetic programming for shear strength prediction of reinforced concrete beams without stirrups. *Applied Soft Computing*, *19*, 112–120.
https://doi.org/10.1016/j.asoc.2014.02.007

Heoa, E. Y., Leeb, G., & Lee, G. (2016). Effect of elevated temperatures on chemical properties, microstructure and carbonation of cement paste. *Journal of Ceramic Processing Research*, *17*(6), 648–652.

Ibrahim, A. M., Salman, W. D., & Bahlol, F. M. (2019). Flexural behavior of concrete composite beams with new steel tube section and different shear connectors. *Tikrit Journal of Engineering Sciences*, *26*(1), 51–61. https://doi.org/10.25130/tjes.26.1.07

Jafarizadeh, F., Larki, B., Kazemi, B., Mehrad, M., Rashidi, S., Neycharan, J. G., Gandomgoun, M., & Gandomgoun, M. H. (2023). A new robust predictive model for lost circulation rate using convolutional neural network: A case study from Marun Oilfield. *Petroleum, 9*(3), 468–485.
https://doi.org/10.1016/j.petlm.2022.04.002

Jumaa'h, M., Kamil, B., & Baghabra, O. (2019). Mechanical and structural properties of a lightweight concrete with different types of recycling coarse aggregate. *Tikrit Journal of Engineering Sciences*, *26*(1), 33–40. https://doi.org/10.25130/tjes.26.1.05

Junyoung, P., & Yootaek, K. (2014). Property enhancement of supercritically carbonated specimen by particle-size separation of fly ash and cement. *Journal of Ceramic Processing Research*, *15*(4), 212–215.

Kamali, M. Z., Davoodi, S., Ghorbani, H., Wood, D. A., Mohamadian, N., Lajmorak, S., Rukavishnikov, V. S., Taherizade, F., & Band, S. S. (2022). Permeability prediction of heterogeneous carbonate gas condensate reservoirs applying group method of data handling. *Marine and Petroleum Geology, 139*, Article 105597. https://doi.org/10.1016/j.marpetgeo.2022.105597

Kar, S., Pandit, A. R., & Biswal, K. C. (2020). Prediction of FRP shear contribution for wrapped shear deficient RC beams using adaptive neuro-fuzzy inference system (ANFIS). *Structures*, *23*, 702–717. https://doi.org/10.1016/j.istruc.2019.10.022

Khalaf, J. A., Majeed, A. A., Aldlemy, M. S., Ali, Z. H., Zand, A. W. A., Adarsh, S., Bouaissi, A., Hameed, M. M., & Yaseen, Z. M. (2021). Hybridized deep learning model for perfobond rib shear strength connector prediction. *Complexity*, *2021*, Article 6611885. https://doi.org/10.1155/2021/6611885

Li, Y., Luo, Y., Du, H., Liu, W., Tang, L., & Xing, F. (2022). Evolution of microstructural characteristics of carbonated cement pastes subjected to high temperatures evaluated by MIP and SEM. *Materials*, *15*(17), Article 6037.
https://doi.org/10.3390/ma15176037

Ly, H. B., Le, T. T., Vu, H. L. T., Tran, V. Q., Le, L. M., & Pham, B. T. (2020). Computational hybrid machine learning based prediction of shear capacity for steel fiber reinforced concrete beams. *Sustainability*, *12*(7), 1–34.
https://doi.org/10.3390/su12072709

Mahmood, B. A., & Mohammad, K. I. (2019). Finite element analysis of RC deep beams under eccentric load. *Tikrit Journal of Engineering Sciences*, *26*(1), 41–50.
https://doi.org/10.25130/tjes.26.1.06

Maryam, S., Karzan, W., Majid, S., & Mahdi, S. (2018). Application of support vector machine with firefly algorithm for investigation of the factors affecting the shear strength of angle shear connectors. *Smart Materials and Structures*, *22*, 413–424.

Matinkia, M., Amraeiniya, A., Behboud, M. M., Mehrad, M., Bajolvand, M., Gandomgoun, M. H., & Gandomgoun, M. (2022a). A novel approach to pore pressure modeling based on conventional well logs using convolutional neural network. *Journal of Petroleum Science and Engineering, 211*, Article 110156.
https://doi.org/10.1016/j.petrol.2022.110156

Matinkia, M., Sheykhinasab, A., Shojaei, S., Vojdani Tazeh Kand, A., Elmi, A., Bajolvand, M., & Mehrad, M. (2022b). Developing a new model for drilling rate of penetration prediction using convolutional neural network. *Arabian Journal for Science and Engineering, 47*, 11953–11985.
https://doi.org/10.1007/s13369-022-06765-x

Mehrad, M., Bajolvand, M., & Ramezanzadeh, A., & Neycharan, J. G. (2020). Developing a new rigorous drilling rate prediction model using a machine learning technique. *Journal of Petroleum Science and Engineering, 192*, Article 107338.
https://doi.org/10.1016/j.petrol.2020.107338

Mehrad, M., Ramezanzadeh, A., Bajolvand, M., & Hajsaeedi, M. R. (2022). Estimating shear wave velocity in carbonate reservoirs from petrophysical logs using intelligent algorithms. *Journal of Petroleum Science and Engineering, 212*, Article 110254.
https://doi.org/10.1016/j.petrol.2022.110254

Nafees, A., Javed, M. F., Khan, S., Nazir, K., Farooq, F., Aslam, F., Musarat, M. A., & Vatin, N. I. (2021). Predictive modeling of mechanical properties of silica fume-based green concrete using artificial intelligence approaches: MLPNN, ANFIS, and GEP. *Materials*, *14*(24), Article 7531.
https://doi.org/10.3390/ma14247531

Najafgholipour, M. A., Dehghan, S. M., Dooshabi, A., & Niroomandi, A. (2017). Finite element analysis of reinforced concrete beam-column connections with governing joint shear failure mode. *Latin American Journal of Solids and Structures*, *14*(7), 1200–1225. https://doi.org/10.1590/1679-78253682

Nguyen-Sy, T., Wakim, J., To, Q. D., Vu, M. N., Nguyen, T. D., & Nguyen, T. T. (2020). Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. *Construction and Building Materials*, *260*, Article 119757. https://doi.org/10.1016/j.conbuildmat.2020.119757

Penido, R. E. K., da Paixão, R. C. F., Costa, L. C. B., Peixoto, R. A. F., Cury, A. A., & Mendes, J. C. (2022). Predicting the compressive strength of steelmaking slag concrete with machine learning – considerations on developing a mix design tool. *Construction and Building Materials*, *341*, Article 127896. https://doi.org/10.1016/j.conbuildmat.2022.127896

Sharafati, A., Haghbin, M., Aldlemy, M. S., Mussa, M. H., Al Zand, A. W., Ali, M., Bhagat, S. K., Al-Ansari, N., & Yaseen, Z. M. (2020). Development of advanced computer aid model for shear strength of concrete slender beam prediction. *Applied Sciences*, *10*(11), Article 3811. https://doi.org/10.3390/app10113811

Sheykhinasab, A., Mohseni, A. A., & Barahooie Bahari, A., Naruei, E., Davoodi, S., Aghaz, A., & Mehrad, M. (2023). Prediction of permeability of highly heterogeneous hydrocarbon reservoir from conventional petrophysical logs using optimized data-driven algorithms. *Journal of Petroleum Exploration and Production Technology, 13*, 661–689. https://doi.org/10.1007/s13202-022-01593-z

Sim, J., & Park, C. (2011). Compressive strength and resistance to chloride ion penetration and carbonation of recycled aggregate concrete with varying amount of fly ash and fine recycled aggregate. *Waste Management*, *31*, 2352–2360. https://doi.org/10.1016/j.wasman.2011.06.014

Solhmirzaei, R., Salehi, H., Kodur, V., & Naser, M. Z. (2020). Machine learning framework for predicting failure mode and shear capacity of ultra high performance concrete beams. *Engineering Structures*, *224*, Article 111221. https://doi.org/10.1016/j.engstruct.2020.111221

Talkhouncheh, M. Z., Davoodi, S., Larki, B., Mehrad, M., Rashidi, S., & Vasfi, M. (2023). A new approach to mechanical brittleness index modeling based on conventional well logs using hybrid algorithms. *Earth Science Informatics, 16*, 3387–3416. https://doi.org/10.1007/s12145-023-01098-1

Talkhouncheh, M. Z., Davoodi, S., Wood, D. A., Mehrad, M., Rukavishnikov, V. S., & Bakhshi, R. (2024). Robust machine learning predictive models for real-time determination of confined compressive strength of rock using mudlogging data. *Rock Mechanics and Rock Engineering, 57,* 6881–6907. https://doi.org/10.1007/s00603-024-03859-w

Vasanthalin, P. C., & Kavitha, N. C. (2021). Prediction of compressive strength of recycled aggregate concrete using artificial neural network and cuckoo search method. *Materials Today: Proceedings*, *46*, 8480–8488. https://doi.org/10.1016/j.matpr.2021.03.500

Verapathran, M., Vivek, S., Arunkumar, G. E., & Dhavashankaran, D. (2023). Flexural behaviour of HPC beams with steel slag aggregate. *Journal of Ceramic Processing Research*, *24*(1), 89–97.

Wu, J., Xu, J., Diao, B., & Panesar, D. K. (2021). Impacts of reinforcement ratio and fatigue load level on the chloride ingress and service life estimating of fatigue loaded reinforced concrete (RC) beams. *Construction and Building Materials*, *266*, Article 120999. https://doi.org/10.1016/j.conbuildmat.2020.120999

Yaseen, Z. M., Tran, M. T., Kim, S., Bakhshpoori, T., & Deo, R. C. (2018). Shear strength prediction of steel fiber reinforced concrete beam using hybrid intelligence models: A new approach. *Engineering Structures*, *177*, 244–255. https://doi.org/10.1016/j.engstruct.2018.09.074

Zeng, Z., Zhu, Z., Yao, W., Wang, Z., Wang, C., Wei, Y., Wei, Z., & Guan, X. (2022). Accurate prediction of concrete compressive strength based on explainable features using deep learning. *Construction and Building Materials*, *329*, Article 127082. https://doi.org/10.1016/j.conbuildmat.2022.127082

Zhang, G., Ali, Z. H., Aldlemy, M. S., Mussa, M. H., Salih, S. Q., Hameed, M. M., Al-Khafaji, Z. S., & Yaseen, Z. M. (2022). Reinforced concrete deep beam shear strength capacity modelling using an integrative bio-inspired algorithm with an artificial intelligence model. *Engineering with Computers*, *38*(Suppl. 1), 15–28. https://doi.org/10.1007/s00366-020-01137-1

Zhao, Z., Qu, X., & Li, J. (2020). Application of polymer modified cementitious coatings (PCCs) for impermeability enhancement of concrete. *Construction and Building Materials*, *249*, Article 118769. https://doi.org/10.1016/j.conbuildmat.2020.118769