Periodica Polytechnica Civil Engineering, 69(2), pp. 490–504, 2025

Developing Hybrid Algorithms with Fire Hawk Optimization on Concrete's Chloride Diffusion

Chunyang Zhang[1*]

[1] College of Civil Engineering, Baicheng Normal University, 57 Zhongxing Main Road, 137000 Baicheng, Jilin, China
[*] Corresponding author, e-mail: Lucy_zh@126.com

Received: 08 July 2024, Accepted: 30 November 2024, Published online: 31 January 2025

Abstract

Throughout their lifespan, reinforced concrete buildings may encounter many challenging circumstances, such as exposure to chloride ions. Exposure to the elements, particularly in coastal areas, may lead to a decrease in durability and degradation of concrete structures. Artificial intelligence (AI) may be utilized to create models that accurately predict the chloride diffusion coefficient (CD) of non-steady state concrete over a long duration by analyzing experimental field data. This approach has the potential to boost the evaluation of the durability of a certain building structure by highlighting the most significant factors. This work showcases the use of the support vector regression (SVR), multi-layered perceptron (MLP), and random forests (RF) for predicting the DC of concrete under different exposure conditions. The fire hawk optimization algorithm (FHOA) approach was employed to improve prediction models that were trained on a dataset consisting of 216 data points. The findings indicate that the RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} models have significant promise in properly forecasting the CD of concrete under different exposure situations while maintaining acceptable R^2 values. The results suggest that RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} may reliably predict specific CD values in various exposure situations. The RF_{FHOA} attained R^2 values of 0.9951 throughout training and 0.9971 throughout testing. In detail, MLP_{FHOA} had a R^2 value of 0.9659 throughout training and 0.9756 throughout testing. The R^2 value for SVR_{FHOA}'s test stage is 0.9835, whereas the training stage is 0.9659.

Keywords

concrete, chloride diffusion, fire hawk optimization, random forests regression

1 Introduction

Reinforced concrete (RC) is widely used in infrastructure because of its cost-effectiveness and durability [1]. Chloride-induced damage may significantly impact RC structures, especially those in coastal, marine, and offshore locations, resulting in costly repairs. Once chloride reaches a certain percentage in concrete, it de-passivates the reinforcing steel, causing corrosion imitation. Chloride infiltration might potentially cause rapid structural collapse [2]. Understanding the penetration level of chlorides in concrete is crucial for extending the endurance and service life of buildings and developing appropriate rehabilitation strategies. It is challenging to estimate the diffusion coefficient or chloride penetration in concrete for every project owing to time and budget constraints [3]. Several investigations have estimated chloride penetration utilizing Fick's second rule, as given in Eq. (1) [4]:

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2}. \tag{1}$$

Where $C(x, t)$ represents chloride concentration at depth x and time t, and D represents diffusion coefficient. Standard formulas may overlook important factors and provide false forecasts due to the complexity and time-dependent nature of chloride intrusion [5]. Utilizing mineral compounds, including fly ash (F), ground granulated blast furnace slag (GGBS), and silica fume (SF) may significantly lower the rate of diffusion and stop reinforcement corrosion, according to a number of empirical investigations [6, 7]. Using pozzolanic materials may prevent hostile species from breaching the steel-concrete contact by reducing porosity and limiting chloride ion motion [8, 9]. To avoid longitudinal empirical testing, an appropriate forecast model for D must be developed utilizing powerful computational tools and including the above phenomena.

Traditional approaches for determining the association among variables rely on statistical analysis using linear and non-linear regression formulas [10, 11]. Regression

Cite this article as: Zhang, C. "Developing Hybrid Algorithms with Fire Hawk Optimization on Concrete's Chloride Diffusion", Periodica Polytechnica Civil Engineering, 69(2), pp. 490–504, 2025. https://doi.org/10.3311/PPci.37856

models are not dependable because they are limited by pre-defined linear or non-linear equations binding the issue to the model [12, 13]. Additionally, these models' assumptions and extensive function range for curve fitting might lead to inaccurate forecasts, especially for complex datasets [14, 15]. Complicated engineering events are modeled using machine learning (ML) approaches according to natural tools to address these difficulties [16–19]. Precise and trustworthy ML methods are genetic programming, gene expression programming, and artificial neural networks (ANN). These methods primarily construct models by training on available data. These methods' ability to recognize patterns may lead to an easy way to comprehend engineering patterns [20]. ANN clearly has advantages due to:

1. It learns from examples, generating input-output correlations from data, and eliminates the need for scholars to choose the suitable regression formula.
2. It can simulate complicated procedures and integrate outliers, enabling wider applicability [21].

ANN is a great tool for tackling complicated engineering issues with many parameters because to its capabilities.

Recent research has focused on using ANN methods to forecast mechanical qualities [22] and design concrete blend proportions. Some researchers have modeled the endurance qualities of concrete owing to its heterogeneity and complexity. High-performance concrete (HPC) chloride ion permeability was simulated by Parichatprecha and Nimityongskul [23], Song and Kwon [24], and Hodhod and Ahmed [25] utilizing 86, 120, and 300 datasets, accordingly, and input factors pertaining to concrete blend proportions. Evaluation of model efficiency involves comparing findings to empirical values and using regression analysis. To enhance the performance of neural network models, it is advised to include a large and quantitative dataset, since the proposed models obtained excellent precision [24]. According to the empirical outcomes, an experimental model employing a neural network approach was developed to forecast the chloride permeability of concrete containing ground pozzolans as a function of 6 input variables (i.e., water to binder ratio, percent replacement, testing ages, pozzolans type, aggregate to cement ratio), with a correlation coefficient (R) as high as 0.97 [26]. In their work, Najimi et al. [3] used an ANN model and artificial bee colony algorithm (ABC) to forecast fast chloride ion permeability using blend variables. This research found a strong link and found that neural network models outperformed regression and genetic methods. Several investigations have associated blend characteristics with chloride diffusion coefficient in carbonated or non-carbonated concretes in steady or non-steady states [27]. Boğa et al. [28] modelled concrete chloride ion permeability using the ANN and adaptive neuro-fuzzy inference system (ANFIS). With the input variables of cure type, curing duration, GGBS, and corrosion inhibitor, the research attained a precision level of 73%. Whereas Hoang et al. [29] employed mortar age, depth of measured position, diffusion dimension, and presence of reinforcement as the input to model chloride penetration of concrete mortar employing ANN approach, Asghshahr et al. [30] took environmental conditions, penetration depth, water-to-cementitious material ratio, and SF mass into consideration. The three investigations examined empirical outcomes from 54, 162, and 132 datapoints, but did not account for potential factors like concrete blend proportion, age, or environment, leading to low precision and generalization capability [12].

The models from literature used two innovative types of ensembles artificial intelligence (AI) methods, genetic programming forest (GPF) and linear genetic programming forest (LGPF) methodologies, to mimic the chloride diffusion coefficient (CD) of concrete. As a result of the investigation, actual field data were gathered. As the control ensemble methodology, the random forests (RF) technique was used for comparison. The highest-performing LGPF model outperforms even the most sophisticated GPF and RF models. Findings show that the percentage of silica particles in the binder, exposure time, and circumstances most affect concrete endurance [31]. The ANN developed 4 metaheuristic optimization algorithms that rely on marine creatures: the whale optimization algorithm (WOA), the marine predator's algorithm (MPA), and the jellyfish search optimizer (JSO). The offered methods were used to mimic concrete's CD under atmospheric, tidal, impact, and submerged circumstances. The research included 216 field trial data points. Findings show that simpler synthesized approaches outperform the old strategy. The Wilcoxon rank-sum test shows that ANN-JSO outperforms other ANN methods. Additionally, the mean ANN-MPA, artificial neural networks salp swam algorithm (ANN-SSA), and ANN-WOA test results stay unaltered [32]. Research offered several ML techniques to estimate the chloride diffusion coefficient of concrete with supplementary cementitious materials (SCMs) like SF, ground granulated blast furnace slag, and fly ash. A database was established, consisting of nine input parameters. Eight ML models were assessed, including Support Vector Machine (SVM), Extreme Learning Machine (ELM), K-Nearest Neighbours (KNN), Light

Gradient Boosting (LGB), Extreme Gradient Boosting (EGB), RF, Gradient Boosting (GB), and AdaBoost (AB). Gradient Boosting predicted the chloride diffusion coefficient best. Choose the finest ML method Gradient Boosting helped establish a reliable soft computing method for concrete structure endurance design, such as mix design optimization and binder selection [33].

1.1 The study contribution and structure

Due to the intricate nature of forecasting the CD of concrete, which exhibits time-dependent behavior, standard models relying on short-term empirical studies suffer from several limitations. Several objections of the rapid chloride penetration test (RCPT) include the fact that it assesses flow prior to attaining a stable migration state, elevates the sample's temperature due to high voltage, and evaluates flow for all ions rather than only chloride ions. Conducting longitudinal field testing on the CD of concrete requires a significant investment of both time and financial resources. This topic emphasizes the significance of the findings derived from extensive and thorough field studies on the durability and crack resistance of concrete carried out over an extended period of time. To enhance the understanding of the CD of concrete in real-world scenarios and provide more dependable results, merging the outcomes of these significant studies and utilizing AI methods to evaluate the collected data as a comprehensive database is proposed. The present study compiled a comprehensive database of longitudinal investigations on the CD of concrete in maritime environments, employing data from prior research. The CD of concrete is then characterized using support vector regression (SVR), multi-layered perceptron (MLP), and RF models, which are constructed and compared with existing models to assess their effectiveness. In this work, the SVR, MLP, and RF models are improved utilizing a proven optimization approach called fire hawk optimization algorithm (FHOA). When estimating the CD of concrete, several significant factors are taken into account. These include the water-to-binder ratio, the ratio of coarse aggregate to total aggregate, the ratio of SF to binder, the ratio of superplasticizer to binder, the curing process, the exposure duration, and the exposure condition. This study makes a valuable contribution to the field of civil engineering by proposing a novel model for the coefficient of diffusion of concrete, denoted as CD. This tool facilitates the evaluation of the resilience of concrete buildings.

In the first section, the background information on the topic was addressed, the significance of the study was explained, and the objectives and scope of the research was scrutinized. Also, the existing literature and related studies were reviewed. Secondly, the steps related to data pre-processing were discussed. Next, the base description of the considered algorithms was accomplished such as training algorithms, and hyperparameter tuning. In the next step, the results of developed models discussed and compared with each other. Finally, the summary of key findings was discussed.

2 Methodology

2.1 Applied models and optimization algorithm

2.1.1 Fire hawk optimization algorithm (FHOA)

It has been presented FHO as one of the novel metaheuristic methods [34]. This method is inspired by the hawk's manner of pursuing a hunt by dispersing flames around the hunting area (Fig. 1). To build a tiny fire, the hawk will take a flaming stick and drop it somewhere else that has not burnt. The prey is scared off by this little fire and is forced to run away quickly and anxiously, making it easier for the hawk to grab them.

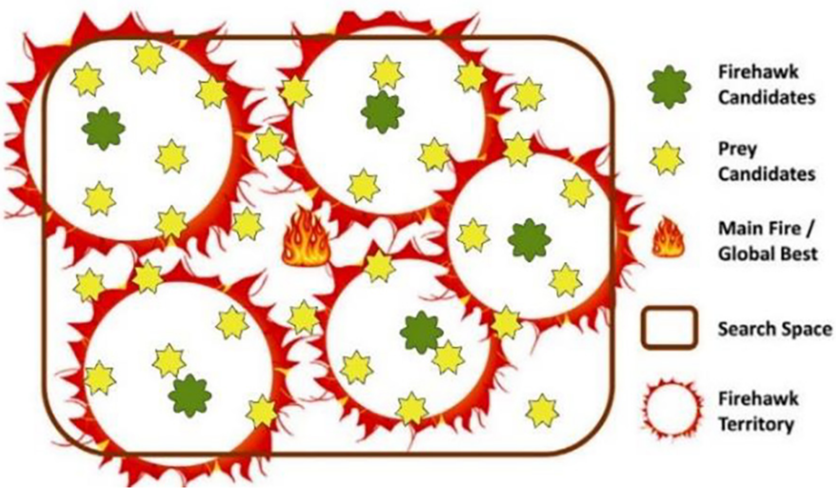


Fig. 1 Fire hawk model search space schematic

Initialization is the first step in the FHO process. The initial definition of a number of possible answers (***X***) is the location vectors of the fire hawk and its target. These vectors' starting locations in space are determined by an arbitrary initialization procedure. The location vectors are expressed as in Eqs. (2) and (3).

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\\ \vdots\\ \boldsymbol{X}_i\\ \vdots\\ \boldsymbol{X}_N\end{bmatrix}=\begin{bmatrix}\boldsymbol{X}_2^1 & \boldsymbol{X}_1^2 & \cdots & \boldsymbol{X}_j^i & \cdots & \boldsymbol{X}_1^d\\ \boldsymbol{X}_2^1 & \boldsymbol{X}_2^2 & \cdots & \boldsymbol{X}_2^j & \cdots & \boldsymbol{X}_2^d\\ & & \vdots & \ddots & \vdots & \\ \boldsymbol{X}_i^1 & \boldsymbol{X}_i^2 & \cdots & \boldsymbol{X}_i^j & \cdots & \boldsymbol{X}_i^d\\ & & \vdots & \ddots & \vdots & \\ \boldsymbol{X}_N^1 & \boldsymbol{X}_N^2 & \cdots & \boldsymbol{X}_N^j & \cdots & \boldsymbol{X}_N^d\end{bmatrix} \tag{2}$$

$$\boldsymbol{X}_i^j(0)=\boldsymbol{X}_{i,\min}^j+\mathrm{rand}\times\left(\boldsymbol{X}_{i,\max}^j-\boldsymbol{X}_{i,\min}^j\right),\begin{cases}i=1,2,\ldots,N\\ j=1,2,\ldots,d\end{cases} \tag{3}$$

The i^{th} answer candidate in the search area is represented by $\boldsymbol{X}_i$, where d is the dimension of the issue being considered. The total number of candidates for solutions in the search space is N; $\boldsymbol{X}_i^j(0)$ denotes the answer candidate's starting location. The i^{th} answer candidate's j^{th} decision parameter is denoted by $\boldsymbol{X}_i^j$. The i^{th} answer candidate's j^{th} decision parameter has minimum and maximum bounds denoted by $\boldsymbol{X}_{i,\min}^j$ and $\boldsymbol{X}_{i,\max}^j$, whereas rand is a uniformly distributed arbitrary integer in the interval [0,1].

The subsequent phase is to evaluate the target function for the candidate answer while taking into account the chosen optimization issue as expressed in Eqs. (4) and (5) in order to pinpoint the fire hawk's position in the region. where PR_k represents the k^{th} prey in the search area taking into account the total number of m prey, and FH_1 represents the 1st fire hawk taking into account the total number of n fire hawks in the search area.

$$PR=\begin{bmatrix}PR_1\\ PR_2\\ \vdots\\ PR_k\\ \vdots\\ PR_m\end{bmatrix},\ k=1,2,\ldots,m \tag{4}$$

$$FH=\begin{bmatrix}FH_1\\ FH_2\\ \vdots\\ FH_k\\ \vdots\\ FH_m\end{bmatrix},\ k=1,2,\ldots,n \tag{5}$$

The fire hawk's target's interval from it is computed in the subsequent stage. Equation (6) calculates the total interval that the fire hawk must travel to reach its target victim.

$$D_K^I\sqrt{\left(x_2-x_1\right)^2+\left(y_2-y_1\right)^2},\ \begin{cases}I=1,2,\ldots,n\\ k=1,2,..,m\end{cases} \tag{6}$$

Where the fire hawk and prey's locations in the search area are represented by (x_1, y_1) and (x_2, y_2). The variables m and n represent the total number of prey and fire hawks in the search area, respectively, and D_K^I denotes the entire interval between the first fire hawk and the k^{th} prey. Equation (7) illustrates the location modification process in the FHO primary search loop, which is the subsequent stage of the method.

$$\boldsymbol{FH}_I^{\mathrm{new}}=FH_I+\left(r_1\times GB-r_2FH_{\mathrm{near}}\right),\ I=1,2,\ldots,n \tag{7}$$

Where GB represents the worldwide greatest answer in the search area regarded as the primary fire, and $\boldsymbol{FH}_I^{\mathrm{new}}$ is the updated location vector of the 1st fire hawk (FH_I). The other fire hawk in the search area is called , and the uniformly distributed arbitrary integers r_1 and r_2 in the range of (0,1) are used to calculate the fire hawk's movement in the direction of the primary fire and the other fire hawk territories.

The movement of the hunt inside every fire hawk area is taken into consideration as a crucial component of animal manner for the location update procedure in the subsequent phase of the method. Equation (8) may be used to take this action into account throughout the location modification procedure.

$$\boldsymbol{PR}_q^{\mathrm{new}}=PR_q+\left(r_3\times FH_I-r_4\times SP_I\right),\ \begin{cases}I=1,2,\ldots,n\\ q=1,2,\ldots,r\end{cases} \tag{8}$$

To ascertain the prey's progress in the direction of the fire hawk and the safe spot, r_3 and r_4 are uniformly distributed arbitrary integers in the range of (0,1). SP_I represents the safe spot under the fire hawk realm. Whereas $\boldsymbol{PR}_q^{\mathrm{new}}$ is the 1st fire hawk's (FH_I) new position vector around the q^{th} prey (PR_q). In the area of search that is thought to be the primary fire, GB provides the finest option.

Furthermore, the prey may attempt to flee to a safer area out of the fire hawk zone where they are caught, or they may migrate toward another fire hawk zone. It is also conceivable for the hunt to approach the fire hawk nearer to the ambush. Equation (9) may be used to take these changes into account throughout the location modification procedure (Fig. 2).

$$\boldsymbol{PR}_q^{\mathrm{new}}=PR_q+\left(r_5\times FH_{\mathrm{Alter}}-r_6\times SP\right),\ \begin{cases}I=1,2,\ldots,n\\ q=1,2,\ldots,r\end{cases} \tag{9}$$

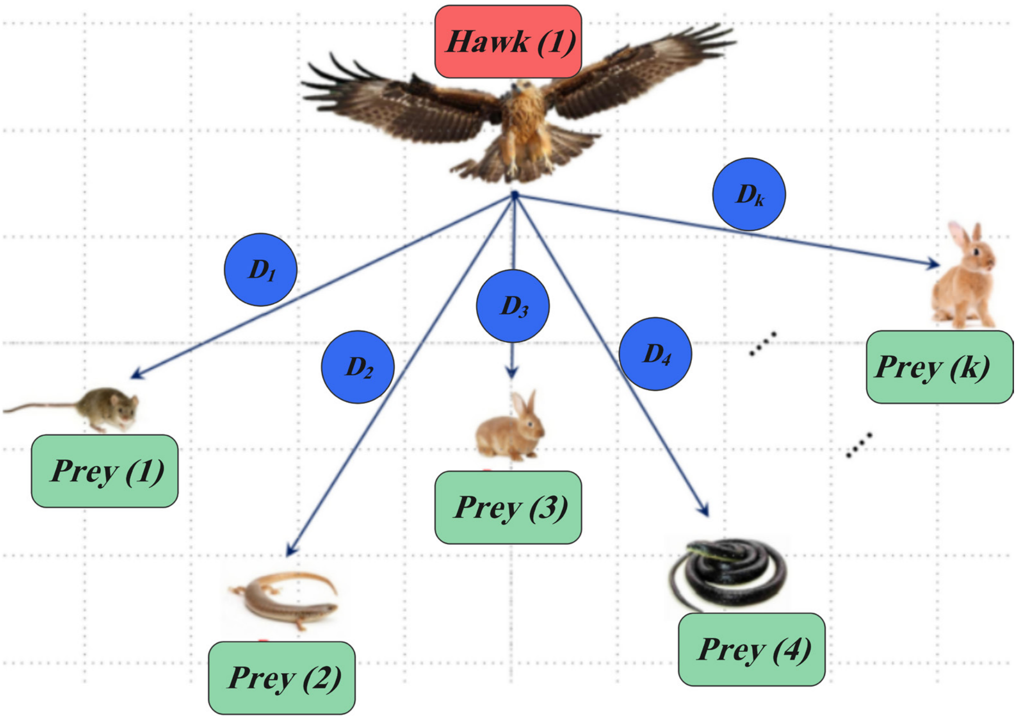


Fig. 2 Fire hawk position updating

In the search area, where FH_{Alter} is one of the other fire hawks. The 1st fire hawk (FH_l) is encircling the new location vector of the q^{th} prey (PR_q), which is $\boldsymbol{PR}_q^{\text{new}}$. SP is a secure location out of the l^{th} fire hawk's domain. To ascertain the motion of hunt towards other fire hawks and secure locations beyond the region, r_5 and r_6 are uniformly distributed arbitrary values in an interval of (0,1).

Equations (10) and (11) illustrate SP_l and SP mathematically. This is according to the observation that most animals gather in secure locations when they are in threat in order to keep their health and safety.

$$SP_l = \frac{\sum_{q=1}^{r} PR_q}{r}, \begin{cases} q = 1,2,\ldots,r \\ l = 1,2,\ldots,n \end{cases} \tag{10}$$

$$SP = \frac{\sum_{k=1}^{m} PR_K}{m}, \ k = 1,2,\ldots,m \tag{11}$$

PR_K represents the k^{th} prey in the search area, while PR_q represents the q^{th} prey encircled by the l^{th} fire hawk (FH_l).

2.1.2 Random forests (RF) regression

In ensemble learning, RF is a bagging method. Breiman [35] integrated decision trees into an RF, which is a large number of decision trees produced by randomly selecting parameters (columns) and data (rows). Subsequently, the decision tree outcomes were combined, leading to a significant improvement in the RF's forecast precision while maintaining the same computation expense. The Bagging method and the decision tree make up the RF. The following is the bagging method procedure: By using the bootstrapping technique, k training samples are chosen at random from the initial sample collection of n samples. Novel training collections are then created by sampling k times without replacing any samples (k training collections are independent of one another, and components may be reused). The following describes the percentage of the novel training set that comprises samples from the original sample set:

$$x = 1 - \left(1 - \frac{1}{k}\right)^n \tag{12}$$

when $k \to \infty$

$$\lim_{k\to\infty} x = 1 - \lim_{k\to\infty}\left(1 - \frac{1}{k}\right)^k = 1 - e^{-1} \approx 0.632 = 63.2\% . \tag{13}$$

Out-of-bag forecasters (OOB) are samples from the original sample set that are absent from the novel training set. They are used to assess how well the decision trees produced by the novel training set function. The creation of decision trees is assessed using Eq. (14).

$$M = \arg\min \sum_{u\in U} \sum_{v\in V} \left(h_u - h_v\right)^2 \tag{14}$$

Where h_v represents the anticipated value of the decision tree's terminal leaf node, h_u represents the result value of the u^{th} sample in the data set, and M represents the total of the decision tree's squared errors.

K decision trees are created for each of the k training sets, and the outcome T is merged by taking the average of each of the k decision trees' outcomes (T_1, T_2, …, T_k); each decision tree has the same weight.

$$T = \frac{1}{k}\sum\left(T_1 + T_2 + \ldots + T_k\right) \tag{15}$$

2.1.3 Support vector regression (SVR)

SVM is a type of particular method that is able to be utilized to handle regression and classification issues. A hyperplane with the largest margin in the feature area serves as their fundamental model. The research goal of SVR, taking into account the provided train dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, is to discover a function indicating the connection between x and y, and the function may obtain the matching predicted value when a novel x is provided. Equation (16) represents this function.

$$f(x) = \sum_{i=1}^{n} w\varphi(x) + b \tag{16}$$

Where the SVR's ultimate research objectives are w and b, which determine a linear hyperplane that is able to match the training dataset. When the connection between x and y is non-linear, the non-linear mapping $\varphi(x)$ transfers x to a novel area. The link between $\varphi(x)$ and y in the novel area is linear.

In Eq. (17), where L_ε is referred to as the ε-insensitive loss function given by Drucker et al. [36], the anticipated risk may be characterized as the objective of SVR, which is to reduce it. In Eq. (18), L_ε is determined.

$$R_{emp} = \frac{1}{n}\sum_{i=1}^{n} L_\varepsilon\left(y_i, f(x_i)\right) \tag{17}$$

$$L_\varepsilon\left(y, f(x)\right) = \begin{cases} 0, & \text{if } \left|y - f(x)\right| \le \varepsilon \\ \left|y - f(x)\right| - \varepsilon, & \text{otherwise} \end{cases} \tag{18}$$

In order to lessen the predicted risk utilizing an ε-insensitive loss, SVR executes linear regression in the feature area. At the same time, it attempts to simplify the model by reducing $\|w^2\|$. Equation (19), in which $\xi_i, \xi_i^* \left(i = 1, \ldots, n\right)$ represent the non-negative slack parameters, which may be used to actualize this. These parameters indicate the difference between the training dataset's function $f(x)$ and the true value.

$$\begin{aligned} &\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w^2\| + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \\ &\text{subjected to} \quad \begin{aligned} & w\varphi(x_i) + b - y_i \le \varepsilon + \xi_i, \\ & y_i - w\varphi(x_i) - b \le \varepsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, n, \end{aligned} \end{aligned} \tag{19}$$

This optimization issue may be converted into a dual issue, and the dual issue's answer is provided by Eq. (20), where a_i^*, a_i represent the Lagrange multipliers, which are able to be obtained by addressing the dual issue and $K(x_i, x_j)$ represents the kernel function, which is equivalent to the inner product of $\varphi(x_i)$ and $\varphi(x_j)$. As the kernel function, every function that fulfills Mercer's requirement [37] is acceptable.

$$\begin{aligned} & f(x) = \sum_{i=1}^{n}\left(a_i^* - a_i\right)K\left(x_i, x\right) + b \\ & \text{subjected to} \quad 0 \le a_i^* \le C, \quad 0 \le a_i \le C \end{aligned} \tag{20}$$

The sigmoid kernel function, radial basis kernel function, and polynomial kernel function are the three most often utilized kernel functions. The radial basis kernel function, which is denoted by Eq. (21), is used in this study.

$$K(x, z) = \exp\left(\frac{\|x - z\|^2}{2\gamma^2}\right) \tag{21}$$

Where γ represents a manually adjustable variable, analogous to ε and C in Eq. (20), all of that has a significant impact on the SVR's predicting precision.

2.1.4 Multi-layered perceptron (MLP)

One of the most often used neural network models is back-propagation, which is a multilayer feed-forward in ANN. The back-propagation makes use of gradient reduction and average square fault to modify the link weight of the minimal fault total of squares. As a training sample for the network in this approach, certain calculated values are provided. The link weights' starting values are then supplied [38]. The difference between calculated values and predicted values is back-propagated across the network to update weights. After the supervised learning technique, the difference between the predicted and calculated values will be reduced. The network of the non-linear model is organized into three levels: back-propagation and feed-forward. Input level, neurons' concealed level with non-linear transfer functions, and neurons' outcome level with linear transfer functions make up the layout of this network. The input parameters are presented by x_j(j = 1, 2, …, n), the neurons' outcome in the concealed levels indicated by

$z_i (i = 1, 2, \ldots, m)$, and the neural network's result is presented by $y_t (t = 1, 2, \ldots, l)$.

By providing enough input data, neural networks may generate any kind of pattern. In order to match the inputs and objectives, the network will be trained using an appropriate approach, such as Levenberg-Marquardt back-propagation. Two crucial stages may be implemented throughout the training procedure to update the values of the weights. The concealed level is the initial stage and the following function in Eqs. (22) and (23) illustrate how to calculate the concealed level for entire neuronal results. neti represents the i^{th} node's activation value, z_i includes the concealed level's result, and f_H shows the activation function, which, in this instance, is a sigmoid function.

$$\text{net}_i = \sum_{j=0}^{n} w_{ji} x_j v_i \quad i, j = 1, 2, \ldots, m \tag{22}$$

$$z_i = f_H\left(\text{net}_i\right) \quad i = 1, 2, \ldots, m \tag{23}$$

$$f_H\left(x\right) = \frac{1}{1 + \exp\left(-x\right)} \tag{24}$$

The result, or second stage, in which the whole neurons in the outcome level's result is shown using the function in Eq. (25).

$$y_t = f_t\left(\sum_{i=0}^{n} w_{it} z_i\right) \quad t = 1, 2, \ldots, l \tag{25}$$

Here, a line function is indicated by $f_t (t = 1, 2, \ldots, l)$. Using learning samples, the delta rule minimizes the weights that are established using predicted values.

2.2 Collected data pre-processing

Preparing raw data for further analysis or modeling is the goal of pre-processing in data analysis and ML. In order to ensure that algorithms are able to make good use of the data, pre-processing entails a series of procedures that involve cleaning, transforming, and organizing the data. The procedures for dealing with missing values, noisy data, and outliers were carried out in the first stage. After that, the dataset was standardized by selectively shortening the aforementioned data. After this, the sensitivity analysis was carried out to select features based on the literature, which included selecting the characteristics that were the most relevant to the problem at hand. In addition, the dataset obtained from various sources was meticulously merged and divided into two distinct stages for learning and evaluation. The distribution of the data in both phases was randomized, ensuring that all ranges of each characteristic were included. The dataset used for estimating the CD of concrete in the models consisted of 216 rows of data, representing different exposure situations [39–49]. The dataset was split into 10 equal-sized subsets (or folds). 10-fold was typically chosen when the dataset is moderately sized, providing a good balance between bias and variance. In this study, the observations were split into two stages: the training stage, which comprised 70% of the data (150 recordings), and the testing stage, which comprised 30% of the data (66 recordings). These proportions are based on the literature and would be considered for data dividing, along with 70/30, 75/25, 80/20, and 90/10. For 3 to 60 months, the concrete examples were exposed to a variety of air, splash, tidal, and submerged environments. The water-to-binder ratio (W/B), coarse aggregate to total aggregate (CAG/TAG), silica fume-to-binder ratio (SF/B), superplasticizer to binder ratio (SP/B), curing mechanism (CM), exposure time (ET), and exposure condition (EC) are the non-dimension form of the dataset that was introduced for development. CM = $X + 0.01Z$ is the formula for CM, where X and Z stand for the curing method and period (in days), respectively. For curing circumstances of air-curing, humid-curing, water-curing, and 95% humidity-curing, X takes on values of 1, 2, 3, and 4, correspondingly. For both the training and testing sets of characteristics, Table 1 shows their statistical qualities. Fig. 3 shows the scatter and box plots of input traits against the target. The scatter plot will allow the reader to visually assess how strongly the input traits relate to the target (e.g., whether linear or non-linear patterns exist). Also, the box plot will offer insights into the distribution of these traits and whether any outliers may affect the relationship between input traits and the target.

2.3 Performance evaluators

To assess the effectiveness of the RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} models and facilitate comparison, eight efficiency factors were taken into consideration. The coefficient of determination (R^2), the root mean square error (RMSE), the mean absolute error (MAE), the relative absolute error (RAE), root relative square error (RRSE), normalized mean square error (NMSE), Theil inequality coefficient (TIC), agreement of forecasting results (IA) are the indices that are being taken into consideration.

Table 1 Characteristics of the chosen input variables

Subset	Index						
	Minimum	Maximum	Standard deviation	Variance	Average	Skewness	Kurtosis
Input 1: W/B							
Train phase	0.3	0.5	0.0681	0.00464	0.431	−0.397	−1.206
Test phase	0.3	0.5	0.0631	0.00398	0.428	−0.201	−1.1964
Input 2: CAG/TAG							
Train phase	0.51	0.65	0.0316	0.001	0.554	1.8188	3.739
Test phase	0.51	0.65	0.0255	0.00065	0.552	2.43	7.887
Input 3: SF/B							
Train phase	0	0.143	0.046	0.0021	0.0459	0.355	−1.283
Test phase	0	0.143	0.0481	0.0023	0.0445	0.41	−1.4043
Input 4: SP/B							
Train phase	0	2.4	0.692	0.4791	0.609	1.306	0.899
Test phase	0	2.4	0.5486	0.301	0.4301	2.3038	5.9377
Input 5: CM							
Train phase	2.07	3.28	0.444	0.1975	2.968	−1.382	0.2836
Test phase	2.07	3.28	0.4129	0.1705	3.0101	−1.648	1.352
Input 6: ET (Days)							
Train phase	3	60	18.6314	347.129	19.81	0.8713	−0.533
Test phase	3	84	20.02	400.74	19.67	1.185	0.5944
Input 7: Exposure type (Count)							
EC	Tidal (T)		Splash (SP)	Atmosphere (A)		Submerged (SU)	
Train phase	84		32	21		11	
Test phase	30		22	9		5	
Target: CD							
Train phase	0.21	27.55	4.552	20.722	3.667	2.373	6.8196
Test phase	0.21	21.79	4.2584	18.1346	3.4839	2.484	7.339

$$R^2 = \left(\frac{\sum_{d=1}^{D} (m_d - \bar{m})(z_d - \bar{z})}{\sqrt{\left[\sum_{d=1}^{D} (m_d - m)^2\right]\left[\sum_{d=1}^{D} (z_d - \bar{z})^2\right]}} \right)^2 \tag{26}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{D}\sum_{d=1}^{D} (z_d - m_d)^2} \tag{27}$$

$$\mathrm{MAE} = \frac{1}{D}\sum_{d=1}^{D} |z_d - m_d| \tag{28}$$

$$\mathrm{RAE} = \frac{\sum_{d=1}^{D} |m_d - z_d|}{\sum_{d=1}^{D} |m_d - \bar{m}|} \tag{29}$$

$$\mathrm{RRSE} = \sqrt{\frac{\sum_{d=1}^{D} (m_d - z_d)^2}{\sum_{d=1}^{D} (m_d - \bar{m})^2}} \tag{30}$$

$$\mathrm{NMSE} = \mathrm{MSE} / Var(m) \tag{31}$$

Where MSE is the mean squared error and *Var* means variance.

$$\mathrm{TIC} = \frac{\sqrt{\frac{1}{D}\sum_{d=1}^{D} (z_d - m_d)^2}}{\left(\sqrt{\frac{1}{D}\sum_{d=1}^{D} {z_d}^2} + \sqrt{\frac{1}{D}\sum_{d=1}^{D} {m_d}^2}\right)} \tag{32}$$

$$\mathrm{IA} = 1 - \frac{\sum_{d=1}^{D} (m_d - z_d)^2}{\sum_{d=1}^{D} \left(|m_d - \bar{m}| + |z_d - \bar{m}|\right)^2} \tag{33}$$

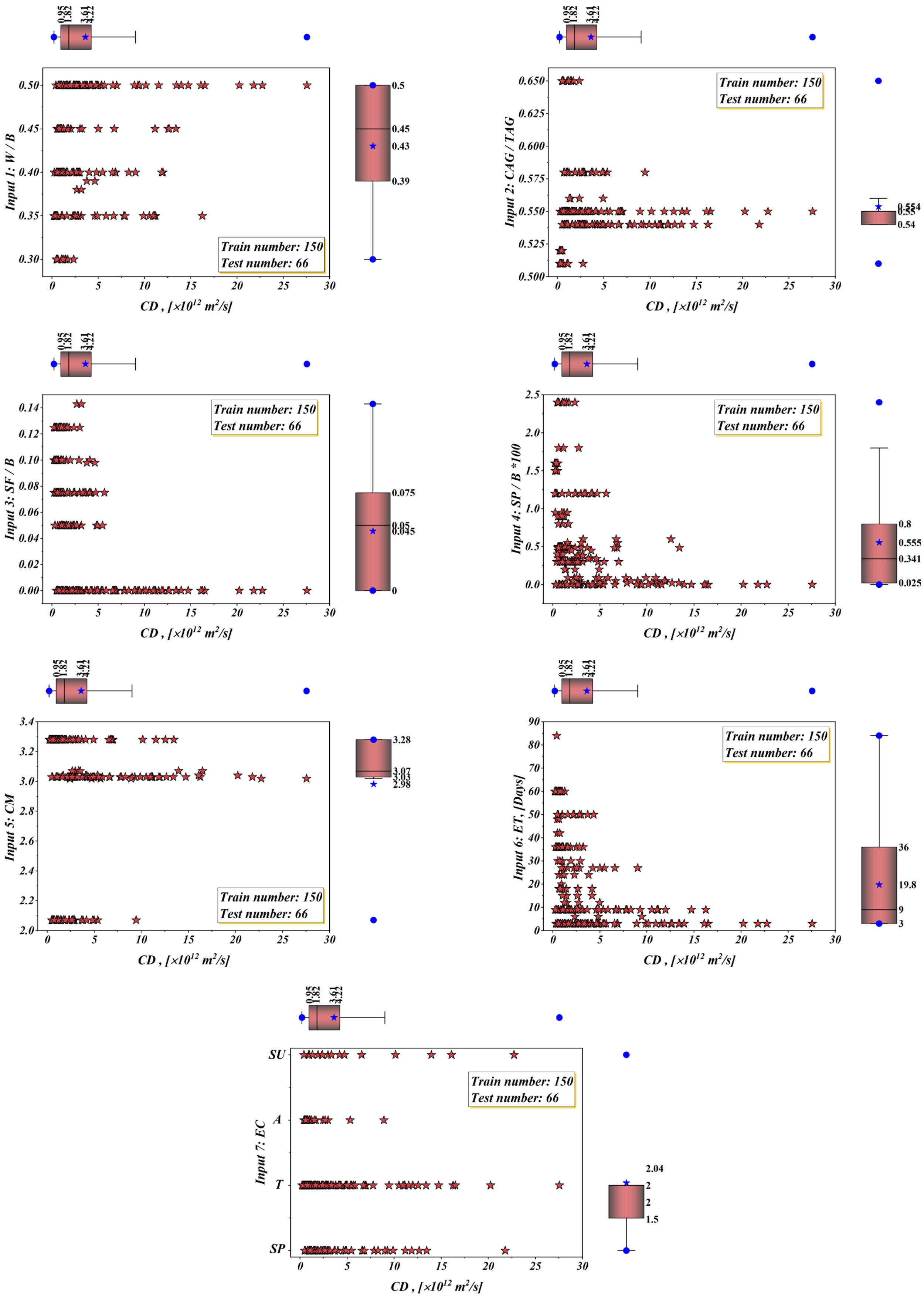


Fig. 3 The chosen input variables vs. target

The formulas include the parameters m_d, $\bar{m}$, z_d, and $\bar{z}$, which represent the observed values, the mean of the observed values, the simulated values, and the mean of the simulated values, respectively. Furthermore, D represents the overall quantity of datasets.

3 Results and discussions

3.1 The procedure of FHOA-based models

Data preparation, simulated training, and hyperparameter tweaking are essential steps in constructing RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} models. The FHOA method may successfully identify the optimal hyperparameters to improve simulation efficiency. Table 2 shows the parameters related to the procedure of FHOA-based models.

Cleaning and preparation of the incoming dataset involved encoding category features and fixing missing values. In order to assess the procedure's efficacy, the data set was split up into many classes for training and testing. It was found that a few hyperparameters required adjusting. The conventional hyperparameters for SVR are c, ε, and γ, for RF are n_e, $\max_d$, and $\max_f$, and for MLP are neurons in the first, second, and third hidden layers. Next, one target metric that could be used as a productivity indicator was the RMSE function, which could be optimized or minimized based on a particular set of hyperparameters. Moreover, the FHOA was used to determine the ideal configuration of the hyperparameters. Utilizing the entire training dataset and the identified ideal hyperparameters, the RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} models were built. Finally, the accuracy and reliability were assessed using the test process.

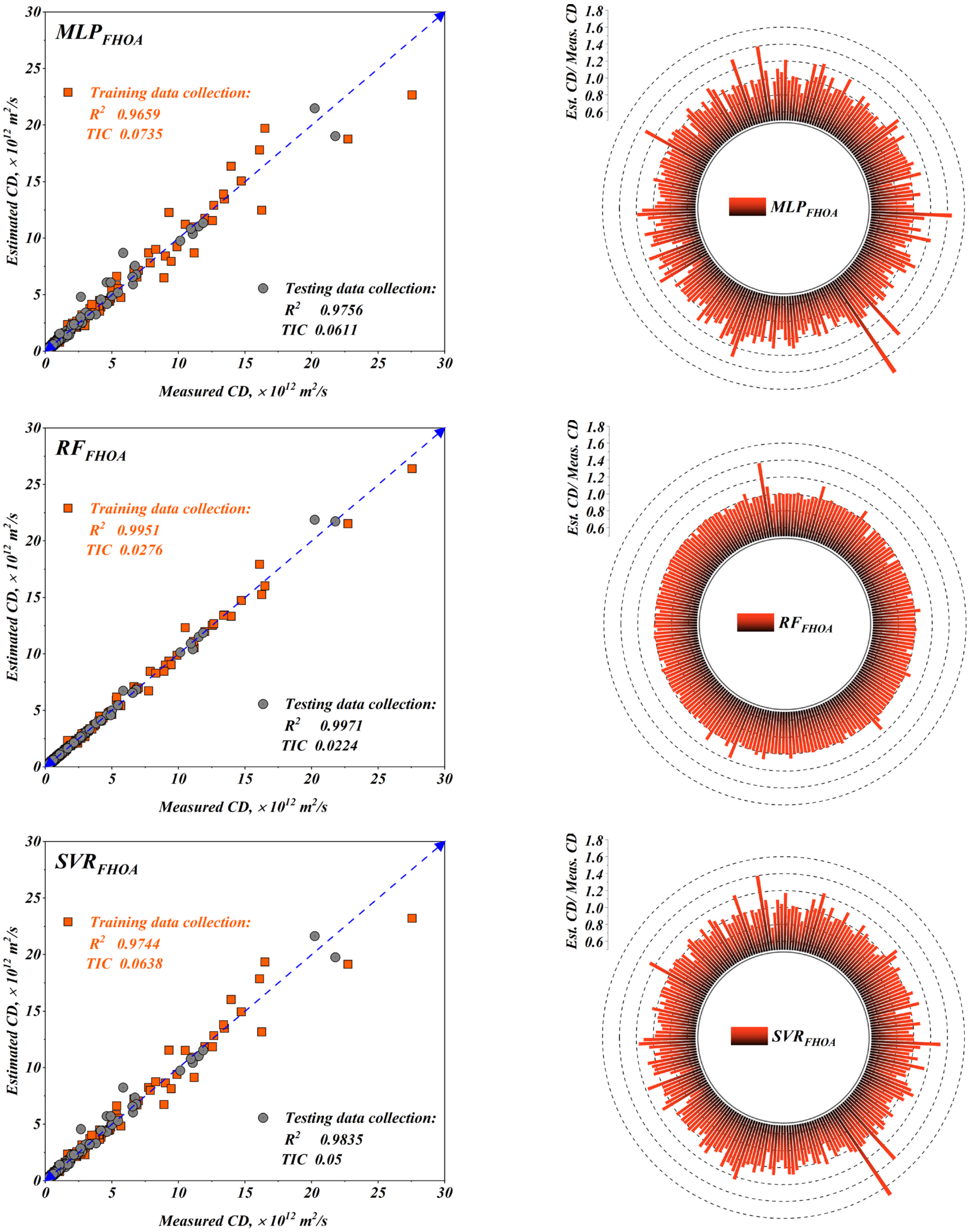
3.2 Discussion

This study aims to evaluate the effectiveness of the RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} techniques in determining the CD of concrete. Fig. 4 displays the observed and expected amounts of concrete CD under various exposure conditions via the testing and training stages of the recommended RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} procedures. Moreover, the error ratio is also presented for the training and test stages with the aim of residual presentation. The precision of RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} in predicting CD was evaluated using the metrics R^2, RMSE, MAE, RAE, RRSE, NMSE, IA, and TIC, as shown in Table 3. In addition, this research evaluated the outcomes of the developed models with the most related research to evaluate the dependability and effectiveness of the models [50, 51].

The results suggest that RF_{FHOA}, MLP_{FHOA} and SVR_{FHOA} have significant promise in properly forecasting the CD of concrete under different exposure situations. During the training stage, RF_{FHOA} achieved R^2 value of 0.9951, and during the testing stage, it achieved a value of 0.9971. To be more specific, MLP_{FHOA} received a R^2 value of 0.9659 during the training phase and 0.9756 during the testing phase. SVR_{FHOA}'s testing phase has a R^2 value of 0.9835, whereas the training phase's value was 0.9659. It is necessary to thoroughly evaluate the effectiveness of auxiliary measures such as NMSE, RAE, RRSE, MAE, TIC, and IA for this particular purpose. The lowest values were shown by the RF_{FHOA} model for the error-based metrics RMSE, RRSE, and RAE. These numbers were more than 50% lower than those of the MLP_{FHOA} and SVR_{FHOA}. For instance, RF_{FHOA}

Table 2 The parameters related to the procedure of FHOA-based models

Optimization	Initialization	Value	Coupled models	Parameter	Optimal value
FHOA	Parameter free		RF_{FHOA}	n_e	186
	Iterations	200		$\max_d$	126
	Runs	10		$\max_f$	102
	Populations	50	MLP_{FHOA}	Function	Back-propagation
				Neurons in the input layer	7
				Hidden layers	3
				Neurons in the first hidden layer	20
				Neurons in the second hidden layer	20
				Neurons in the third hidden layer	10
				Neurons in the output layer	1
			SVR_{FHOA}	C	193
				ε	2.21
				σ	8.32
				Function	RBF



(a) (b)

Fig. 4 The FHOA-based models' findings: (a) scatter plots, and (b) ratio plots

had the lowest value throughout the training stage (0.113) based on the MAE index, which was lower than the scores of 0.354 for MLP$_{FHOA}$ and 0.3028 for SVR$_{FHOA}$. Furthermore, RF$_{FHOA}$ displays the lowest value at 0.0661 throughout the testing stage according to MAE values, which was smaller than MLP$_{FHOA}$ value at 0.3373 and SVR$_{FHOA}$ value at 0.2736.

Fig. 4 displays the performance of the RF$_{FHOA}$, MLP$_{FHOA}$ and SVR$_{FHOA}$ networks by presenting the error ratio among anticipated and observed values. The analysis encompasses both the training and evaluation phases. A greater frequency

Table 3 The FHOA-based models' findings

FHOA-based models	Performance evaluators							
	R^2	RMSE	MAE	RAE	RRSE	NMSE	IA	TIC
Training data collection								
MLP_{FHOA}	0.9659	0.847	0.354	0.112	0.186	0.0107	0.9909	0.0735
RF_{FHOA}	0.9951	0.321	0.113	0.036	0.071	0.0020	0.9987	0.0276
SVR_{FHOA}	0.9744	0.7366	0.3028	0.0958	0.1618	0.0079	0.9932	0.0638
ANN-JSO [50]	0.9033							
Testing data collection								
MLP_{FHOA}	0.9756	0.6702	0.3373	0.1153	0.1574	0.0185	0.9936	0.0611
RF_{FHOA}	0.9971	0.2478	0.0661	0.0226	0.0582	0.0008	0.9992	0.0224
SVR_{FHOA}	0.9835	0.5502	0.2736	0.0936	0.1292	0.0131	0.9958	0.05
ANN-JSO [50]	0.8815							
All data collection								
RF_{FHOA}	0.9685	0.7972						
MLP_{FHOA}	0.9955	0.3008						
SVR_{FHOA}	0.9768	0.685						
LGPF [49]	0.9495	1.033						

of errors near the 1 line and narrower distribution plots indicates greater precision and more pleasing outcomes. The RF_{FHOA} framework displays a substantially focused error distribution with confined upper and lower limits, with the majority of cases clustered around the 1 line.

As previously mentioned, [50, 51], the results of the superior model (RF_{FHOA}) in the present study are compared with the research literature. To facilitate comparative analysis, the R^2 and RMSE measures were utilized. It is clear that the R^2 values increased in the learning and evaluation sections, going from 0.9033 [51] and 0.8815 [51] to 0.9951 and 0.9971, accordingly. This clearly demonstrates the expansion. The recent findings from a study [50], conducted throughout the All-data phase show a significant improvement in performance. The R^2 value increased from 0.9495 to 0.9685 in RF_{FHOA}, from 0.9495 to 0.9955 in MLP_{FHOA}, and from 0.9495 to 0.9768 in SVR_{FHOA}. Additionally, the RMSE value decreased from 1.033 to 0.7972 in RF_{FHOA}, from 1.033 to 0.3008 in MLP_{FHOA}, and from 1.033 to 0.685 in SVR_{FHOA}.

Hybrid RF model is generally easier to interpret. Feature importance can be easily extracted, allowing insights into which variables contribute most to predictions. Neural networks, especially deep ones, are often seen as black-box models, making it difficult to interpret individual feature contributions. Next, due to its ensemble nature, RF is less prone to overfitting, as it averages the results across many trees, reducing variance. While ANNs can overfit if not properly regularized, especially when the network is deep or when the dataset is small. Moreover, RF is more robust to outliers and noise in the data. Since it builds multiple trees, it can ignore outliers or noise in some trees, reducing the impact on the overall model. In contrast, ANNs can be sensitive to noise and outliers, especially if they are not properly handled in pre-processing.

3.3 Parameters' importance analysis

During this investigation, sensitivity analysis is employed to assess the influence of input variables or factors on performance. Model sensitivity assessments give a methodical approach to comprehending the impact of input variables and factors on the model's performance. The objective of this endeavor is to enhance the process of making decisions, comprehension, and the optimization of models. In this work, each of the built models was enhanced with a distinct set of input variables, thereby creating the augmented model (RF_{FHOA}). In order to analyze the impact of various inputs, three metrics were developed and compared with RF_{FHOA}: R^2, IA, and TIC. The results of this comparison are shown in Table 4. The disparity in metrics will escalate in direct correlation to the extent that the absence of components impacts the outcome. The findings suggest that the bulk of the input elements have little impact on the outcome when compared to the RF_{FHOA}. It is important to mention that there is a significant rise in the TIC, and a significant fall in the R^2 and IA value when parameters associated with ET and EC are eliminated from the input set. After deleting the EC, the R^2 and IA

Table 4 The parameters' importance analysis using RF_{FHOA}

Removed attribute (All variables: W/B, CAG/TAG, SF/B, SP/B, CM, ET, EC)	Training data collection			Testing data collection		
	R^2	IA	TIC	R^2	IA	TIC
–	0.995	0.998	0.027	0.997	0.999	0.022
W/B	0.979	0.995	0.057	0.971	0.992	0.066
CAG/TAG	0.995	0.999	0.028	0.997	0.999	0.022
SF/B	0.968	0.992	0.071	0.959	0.989	0.078
SP/B	0.978	0.994	0.059	0.976	0.993	0.063
CM	0.950	0.985	0.093	0.932	0.979	0.109
ET	0.861	0.962	0.148	0.776	0.932	0.189
EC	0.8533	0.959	0.152	0.8547	0.9585	0.153

values declined from 0.9959 to 0.8558 and 0.9987 to 0.959, respectively. Additionally, the TIC value rose from 0.0276 to 0.152 during the learning phase. In addition, when the ET was eliminated, the R^2 and IA values decreased from 0.9971 to 0.776 and 0.9992 to 0.9329, respectively. Conversely, the TIC value grew from 0.0224 to 0.1898 throughout the evaluation stage.

4 Conclusions

The researchers developed a technique that integrates interconnected SVR, MLP, and RF regression to create models that can accurately forecast the diffusion coefficient (CD) of concrete under different exposure situations. The present study utilized the FHOA techniques to discover crucial variables in the MLP, SVR, and RF approaches that might be improved. The study investigates and contrasts the statistical metrics utilized to evaluate the precision and dependability of every model. Moreover, the distinctive sensitivity analysis technique is applied to assess the impact of eliminating every factor on the objective. Furthermore, the present study evaluated the dependability and efficacy of the developed models by comparing them to the most related studies.

Results indicate that RF_{FHOA}, MLP_{FHOA}, and SVR_{FHOA} significantly predict concrete CD under various exposure conditions. The RF_{FHOA} attained R^2 values of 0.9951 during training and 0.9971 during testing. In detail, MLP_{FHOA} had a R^2 value of 0.9659 during training and 0.9756 during testing. The R^2 value for SVR_{FHOA}'s testing phase is 0.9835, whereas the training phase is 0.9659.

The lowest values were found in the RF_{FHOA} model for RMSE, RRSE, and RAE error measures. The results were about 50% lower than MLP_{FHOA} and SVR_{FHOA}. For example, RF_{FHOA} had the lowest MAE index (0.113) during training, compared to 0.354 for MLP_{FHOA} and 0.3028 for SVR_{FHOA}. During testing, RF_{FHOA} had the lowest MAE value of 0.0661, followed by MLP_{FHOA} at 0.3373 and SVR_{FHOA} at 0.2736.

The R^2 and RMSE metrics were used for comparison analysis with literature. In the learning and assessment phases, R^2 values rose from 0.9033 and 0.8815 related to literature to 0.9951 and 0.9971 related to this article, respectively. The obtained results depicted the improved accuracy of developed models with respect to publications.

When ET and EC variables are removed from the input set, the TIC increases, and the R^2 and IA values decrease significantly. After removing the EC, TIC increased from 0.0276 to 0.152 via learning. Eliminating the ET resulted in a drop in R^2 and IA values from 0.9971 to 0.776 and 0.9992 to 0.9329, respectively.

From parameter importance analysis, after deleting the EC, the R^2 and IA values declined from 0.9959 to 0.8558 and 0.9987 to 0.959, respectively. Additionally, the TIC value rose from 0.0276 to 0.152 during the learning phase. In addition, when the ET was eliminated, the R^2 and IA values decreased from 0.9971 to 0.776 and 0.9992 to 0.9329, respectively. Conversely, the TIC value grew from 0.0224 to 0.1898 throughout the evaluation stage.

References

[1] Benemaran, R. S., Esmaeili-Falak, M., Kordlar, M. S. "Improvement of recycled aggregate concrete using glass fiber and silica fume", Multiscale and Multidisciplinary Modeling, Experiments and Design, 7(3), pp. 1895–1914, 2024. https://doi.org/10.1007/s41939-023-00313-2

[2] Mao, L., Hu, Z., Xia, J., Feng, G., Azim, I., Yang, J., Liu, Q. "Multiphase modelling of electrochemical rehabilitation for ASR and chloride affected concrete composites", Composite Structures, 207, pp. 176–189, 2019. https://doi.org/10.1016/j.compstruct.2018.09.063

[3] Najimi, M., Ghafoori, N., Nikoo, M. "Modeling chloride penetration in self-consolidating concrete using artificial neural network combined with artificial bee colony algorithm", Journal of Building Engineering, 22, pp. 216–226, 2019. https://doi.org/10.1016/j.jobe.2018.12.013

[4] Patnana, N., Pattnaik, S., Varshney, T., Singh, V. P. "Self-learning salp swarm optimization based PID design of Doha RO plant", Algorithms, 13(11), 287, 2020. https://doi.org/10.3390/a13110287

[5] van Noort, R., Hunger, M., Spiesz, P. "Long-term chloride migration coefficient in slag cement-based concrete and resistivity as an alternative test method", Construction and Building Materials, 115, pp. 746–759, 2016.
https://doi.org/10.1016/j.conbuildmat.2016.04.054

[6] Du, H., Gao, H. J., Pang, S. D. "Improvement in concrete resistance against water and chloride ingress by adding graphene nanoplatelet", Cement and Concrete Research, 83, pp. 114–123, 2016.
https://doi.org/10.1016/j.cemconres.2016.02.005

[7] Ouldkhaoua, Y., Benabed, B., Abousnina, R., Kadri, E.-H., Khatib, J. "Effect of using metakaolin as supplementary cementitious material and recycled CRT funnel glass as fine aggregate on the durability of green self-compacting concrete", Construction and Building Materials, 235, 117802, 2020.
https://doi.org/10.1016/j.conbuildmat.2019.117802

[8] Sakai, Y. "Relationship between pore structure and chloride diffusion in cementitious materials", Construction and Building Materials, 229, 116868, 2019.
https://doi.org/10.1016/j.conbuildmat.2019.116868

[9] Du, H., Pang, S. D. "High performance cement composites with colloidal nano-silica", Construction and Building Materials, 224, pp. 317–325, 2019.
https://doi.org/10.1016/j.conbuildmat.2019.07.045

[10] Liu, Q., Hu, Z., Lu, X., Yang, J., Azim, I., Sun, W. "Prediction of chloride distribution for offshore concrete based on statistical analysis", Materials, 13(1), 174, 2020.
https://doi.org/10.3390/ma13010174

[11] Esmaeili-Falak, M., Benemaran, R. S. "Ensemble extreme gradient boosting based models to predict the bearing capacity of micropile group", Applied Ocean Research, 151, 104149, 2024.
https://doi.org/10.1016/j.apor.2024.104149

[12] Iqbal, M. F., Liu, Q., Azim, I., Zhu, X., Yang, J., Javed, M. F., Rauf, M. "Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming", Journal of Hazardous Materials, 384, 121322, 2020.
https://doi.org/10.1016/j.jhazmat.2019.121322

[13] Esmaeili-Falak, M., Sarkhani Benemaran, R. "Application of optimization-based regression analysis for evaluation of frost durability of recycled aggregate concrete", Structural Concrete, 25(1), pp. 716–737, 2024.
https://doi.org/10.1002/suco.202300566

[14] Azim, I., Yang, J., Javed, M. F., Iqbal, M. F., Mahmood, Z., Wang, F., Liu, Q. "Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming", Structures, 25, pp. 212–228, 2020.
https://doi.org/10.1016/j.istruc.2020.02.028

[15] Benemaran, S. R., Esmaeili-Falak, M. "Predicting the Young's modulus of frozen sand using machine learning approaches: State-of-the-art review", Geomechanics and Engineering, 34(5), pp. 507–527, 2023.
https://doi.org/10.12989/GAE.2023.34.5.507

[16] Yaseen, Z. M., Deo, R. C., Hilal, A., Abd, A. M., Bueno, L. C., Salcedo-Sanz, S., Nehdi, M. L. "Predicting compressive strength of lightweight foamed concrete using extreme learning machine model", Advances in Engineering Software, 115, pp. 112–125, 2018.
https://doi.org/10.1016/j.advengsoft.2017.09.004

[17] DeRousseau, M. A., Kasprzyk, J. R., Srubar III, W. V. "Computational design optimization of concrete mixtures: A review", Cement and Concreter Research, 109, pp. 42–53, 2018.
https://doi.org/10.1016/j.cemconres.2018.04.007

[18] Aslay, S. E., Dede, T., Kaveh, A. "Integrated Design Optimization Process for Building Projects", Periodica Polytechnica Civil Engineering, 68(4), pp. 1175–1183, 2024.
https://doi.org/10.3311/PPci.37113

[19] Aslay, S. E., Dede, T. "Reduce the construction cost of a 7-story RC public building with metaheuristic algorithms", Architectural Engineering and Design Management, 20(2), pp. 214–229, 2024.
https://doi.org/10.1080/17452007.2023.2195612

[20] Sarkhani Benemaran, R. "Application of extreme gradient boosting method for evaluating the properties of episodic failure of borehole breakout", Geoenergy Science and Engineering, 226, 211837, 2023.
https://doi.org/10.1016/j.geoen.2023.211837

[21] Topçu, İ. B., Sarıdemir, M. "Prediction of properties of waste AAC aggregate concrete using artificial neural network", Computational Materials Science, 41(1), pp. 117–125, 2007.
https://doi.org/10.1016/j.commatsci.2007.03.010

[22] Premkumar, R., Hariharan, P., Rajesh, S. "Effect of silica fume and recycled concrete aggregate on the mechanical properties of GGBS based geopolymer concrete", Materials Today Proceedings, 60, pp. 211–215, 2022.
https://doi.org/10.1016/j.matpr.2021.12.442

[23] Parichatprecha, R., Nimityongskul, P. "Analysis of durability of high performance concrete using artificial neural networks", Construction and Building Materials, 23(2), pp. 910–917, 2009.
https://doi.org/10.1016/j.conbuildmat.2008.04.015

[24] Song, H.-W., Kwon, S.-J. "Evaluation of chloride penetration in high performance concrete using neural network algorithm and micro pore structure", Cement and Concrete Research, 39(9), pp. 814–824, 2009.
https://doi.org/10.1016/j.cemconres.2009.05.013

[25] Hodhod, O. A., Ahmed, H. I. "Developing an artificial neural network model to evaluate chloride diffusivity in high performance concrete", HBRC Journal, 9(1), pp. 15–21, 2013.
https://doi.org/10.1016/j.hbrcj.2013.04.001

[26] Inthata, S., Kowtanapanich, W., Cheerarot, R. "Prediction of chloride permeability of concretes containing ground pozzolans by artificial neural networks", Materials and Structures, 46(10), pp. 1707–1721, 2013.
https://doi.org/10.1617/s11527-012-0009-x

[27] Delnavaz, A., Ramezanianpour, A. A. "The assessment of carbonation effect on chloride diffusion in concrete based on artificial neural network model", Magazine of Concrete Research, 64(10), pp. 877–884, 2012.
https://doi.org/10.1680/macr.11.00059

[28] Boğa, A. R., Öztürk, M., Topçu, İ. B. "Using ANN and ANFIS to predict the mechanical and chloride permeability properties of concrete containing GGBFS and CNI", Composites Part B: Engineering, 45(1), pp. 688–696, 2013.
https://doi.org/10.1016/j.compositesb.2012.05.054

[29] Hoang, N.-D., Chen, C.-T., Liao, K.-W. "Prediction of chloride diffusion in cement mortar using Multi-Gene Genetic Programming and Multivariate Adaptive Regression Splines", Measurement, 112, pp. 141–149, 2017.
https://doi.org/10.1016/j.measurement.2017.08.031

[30] Asghshahr, M. S., Rahai, A., Ashrafi, H. "Prediction of chloride content in concrete using ANN and CART", Magazine of Concrete Research, 68(21), pp. 1085–1098, 2016.
https://doi.org/10.1680/jmacr.15.00261

[31] Golafshani, E. M., Kashani, A., Arashpour, M. "Chloride diffusion modeling of concrete using tree-based forest models", Structural Concrete, 24(4), pp. 5614–5634, 2023.
https://doi.org/10.1002/suco.202300245

[32] Mohammadi Golafshani, E., Kashani, A., Kim, T., Arashpour, M. "Concrete chloride diffusion modelling using marine creatures-based metaheuristic artificial intelligence", Journal of Cleaner Production, 374, 134021, 2022.
https://doi.org/10.1016/j.jclepro.2022.134021

[33] Quan Tran, V. "Machine learning approach for investigating chloride diffusion coefficient of concrete containing supplementary cementitious materials", Construction and Building Materials, 328, 127103, 2022.
https://doi.org/10.1016/j.conbuildmat.2022.127103

[34] Azizi, M., Talatahari, S., Gandomi, A. H. "Fire Hawk Optimizer: a novel metaheuristic algorithm", Artificial Intelligence Review, 56(1), pp. 287–363, 2023.
https://doi.org/10.1007/s10462-022-10173-w

[35] Breiman, L. "Random forests", Machine Learning, 45(1), pp. 5–32, 2001.
https://doi.org/10.1023/A:1010933404324

[36] Drucker, H., Burges, C. J., Kaufman, L., Smola, A., Vapnik, V. "Support Vector Regression Machines", In: Mozer, M. C., Jordan, M., Petsche, T. (eds.) Advances in Neural Information Processing Systems 9 (NIPS 1996), MIT Press, Cambridge, MA, USA, 1996, pp. 155–161.

[37] Smith, B. L., Williams, B. M., Keith Oswald, R. "Comparison of parametric and nonparametric models for traffic flow forecasting", Transportation Research Part C: Emerging Technologies, 10(4), pp. 303–321, 2002.
https://doi.org/10.1016/S0968-090X(02)00009-8

[38] Sagiroglu, S., Colak, I., Bayindir, R. "Power factor correction technique based on artificial neural networks", Energy Conversion and Management, 47(18–19), pp. 3204–3215, 2006.
https://doi.org/10.1016/j.enconman.2006.02.018

[39] Alizadeh, R., Ghods, P., Chini, M., Hoseini, M., Ghalibafian, M., Shekarchi, M. "Effect of curing conditions on the service life design of RC structures in the Persian Gulf region", Journal of Materials in Civil Engineering, 20(1), pp. 2–8, 2008.
https://doi.org/10.1061/(ASCE)0899-1561(2008)20:1(2)

[40] Shekarchi, M., Rafiee, A., Layssi, H. "Long-term chloride diffusion in silica fume concrete in harsh marine climates", Cement and Concrete Composites, 31(10), pp. 769–775, 2009.
https://doi.org/10.1016/j.cemconcomp.2009.08.005

[41] Ghods, P., Chini, M., Alizadeh, R., Hoseini, M., Shekarchi, M., Ramezanianpour, A. A. "The effect of different exposure conditions on the chloride diffusion into concrete in the Persian Gulf region", In: 3rd International Conference on Construction Materials: Performance, Innovations and Structural Implications, Vancouver, Canada, 2005, 272. ISBN 0888658109

[42] Tadayon, M. H., Shekarchi, M., Tadayon, M. "Long-term field study of chloride ingress in concretes containing pozzolans exposed to severe marine tidal zone", Construction and Building Materials, 123, pp. 611–616, 2016.
https://doi.org/10.1016/j.conbuildmat.2016.07.074

[43] Khanzadeh Moradllo, M., Sadati, S., Shekarchi, M. "Quantifying maximum phenomenon in chloride ion profiles and its influence on service-life prediction of concrete structures exposed to seawater tidal zone – A field oriented study", Construction and Building Materials, 180, pp. 109–116, 2018.
https://doi.org/10.1016/j.conbuildmat.2018.05.284

[44] Farahani, A., Taghaddos, H., Shekarchi, M. "Prediction of long-term chloride diffusion in silica fume concrete in a marine environment", Cement and Concrete Composites, 59, pp. 10–17, 2015.
https://doi.org/10.1016/j.cemconcomp.2015.03.006

[45] Costa, A., Appleton, J. "Chloride penetration into concrete in marine environment—Part I: Main parameters affecting chloride penetration", Materials and Structures, 32(4), pp. 252–259, 1999.
https://doi.org/10.1007/BF02479594

[46] Farahani, A., Taghaddos, H., Shekarchi, M. "Chloride diffusion modeling in pozzolanic concrete in marine site", ACI Materials Journal, 115(4), pp. 509–517, 2018.
https://doi.org/10.14359/51702185

[47] Valipour, M., Shekarchi, M., Arezoumandi, M. "Chlorine diffusion resistivity of sustainable green concrete in harsh marine environments", Journal of Cleaner Production, 142, pp. 4092–4100, 2017.

[48] Valipour, M., Pargar, F., Shekarchi, M., Khani, S., Moradian, M. "In situ study of chloride ingress in concretes containing natural zeolite, metakaolin and silica fume exposed to various exposure conditions in a harsh marine environment", Construction and Building Materials, 46, pp. 63–70, 2013.
https://doi.org/10.1016/j.conbuildmat.2013.03.026

[49] Khaghanpour, R., Dousti, A., Shekarchi, M. "Prediction of cover thickness based on long-term chloride penetration in a marine environment", Journal of Performance of Constructed Facilities, 31(1), 04016070, 2017.
https://doi.org/10.1061/(ASCE)CF.1943-5509.0000931

[50] Golafshani, E. M., Kashani, A., Arashpour, M. "Chloride diffusion modeling of concrete using tree-based forest models", Structural Concrete, 24(4), pp. 5614–5634, 2023.
https://doi.org/10.1002/suco.202300245

[51] Mohammadi Golafshani, E., Kashani, A., Kim, T., Arashpour, M. "Concrete chloride diffusion modelling using marine creatures-based metaheuristic artificial intelligence", Journal of Cleaner Production, 374, 134021, 2022.
https://doi.org/10.1016/j.jclepro.2022.134021