



# An Improved Multivariate Adaptive Regression Splines (MARS) Method for Prediction of Compressive Strength of High-Strength (HS) Concrete

Limei Fu<sup>1</sup> · Zhihao Peng<sup>1</sup>

Received: 28 November 2021 / Accepted: 17 April 2022 / Published online: 19 July 2022  
© King Fahd University of Petroleum & Minerals 2022

## Abstract

In today's engineering problems, an adaptive computer-aided system is recruited to relieve the computational cost of design evaluations. The capability of machine learning techniques can be regarded in learning the complex interrelations between the design variables and the response specifically in the concrete mix design. High-Strength (HS) concrete is a complex material, which makes modeling its behavior very challenging. In the present study, possible applicability of Multivariate Adaptive Regression Splines (MARS) to predict the compressive strength of HS concrete is proposed using a limited number of input variables. In order to overcome the conventional drawback of the machine learning approach (e.g., local minima), a novel meta-heuristic technique, namely Water Cycle Algorithm (WCA), is employed to modify MARS model. In addition, a trial and error process is recruited to optimize the complexity of the proposed models. It was also deduced that the WCA outperforms two benchmark optimizers of Crow Search Algorithm (CSA) and Cat Swarm Optimization (CSO) in terms of modeling accuracy in the prediction of the compressive strength of HS concrete. Experimental results using several statistical metrics show that MARS-WCA model ( $R = 0.994$ ,  $NSE = 0.981$ ,  $RMSE = 0.991$  MPa and  $LMI = 0.906$  (training phase) and  $R = 0.991$ ,  $NSE = 0.981$ ,  $RMSE = 1.336$  MPa and  $LMI = 0.880$  (testing phase)) outperformed MARS-CSA, MARS-CSO and standalone MARS to formulation of compressive strength of HS concrete, respectively. In addition, Monte Carlo uncertainty, external validation and sensitivity of variables importance analysis were carried out to verify the results.

**Keywords** High-strength concrete · Evolutionary machine learning models · Compressive strength · Monte Carlo uncertainty

## 1 Introduction

Nowadays, concrete has an undeniably special position in the building industry and is one of the most widely used building materials in the world with its ever-increasing widespread application. The reason for such a special position is the fulfilment of the technical, economic and environmental requirements of human societies [1]. The implementation of concrete structures as the cheapest and most durable structures has driven the researchers to use concrete with higher strength, which has brought great successes in this area [2]. The most important factor that determines the strength of High-Strength (HS) concrete is the porosity in three concrete

phases (aggregates, cement paste and interfacial transition zone (ITZ)) [3, 4]. To produce HS concrete, the contrasting effect of water and cement content on the consistency and strength of concrete cannot be counteracted without the use of water-reducing additives. For this reason, in the past decade, the use of super-plasticizers has played an important role in the production of HS concrete. Basically, the ITZ is the weakest part of the concrete with an ordinary weight made of resistant aggregates with the maximum size of 12–20 mm and the water to cement ratio of 0.4–0.7. The concrete strength can be greatly increased by reducing the maximum size of coarse aggregates for a given water-to-cement ratio, as this will improve the ITZ strength [5–8]. Therefore, in the mix design of HS concrete, the maximum aggregate size is usually considered 19 mm or less. The advantage of using HS concrete is the relatively low hydration heat per unit strength, which reduces the possibility of thermal cracking and minimizes the column dimensions and thus, ensures the beauty of the structure [9–11]. HS concrete columns can hold more weight and therefore, be made slimmer than OPC columns,

✉ Zhihao Peng  
pengzhihao@neusoft.edu.cn

Limei Fu  
fulimei@neusoft.edu.cn

<sup>1</sup> School of Software, Neusoft University of Information, Dalian, China



which allows for more usable space, especially in the lower floors of buildings [10]. The use of pozzolans in the production of composite cement has reduced the total cost of concrete, but the unavailability of various types of pozzolans for the construction of structures and their contribution to the mix design of HS concrete has made it challenging to use the pozzolans [11, 12]. Therefore, in this study, a mix design containing four main components of concrete (water, cement, coarse and fine aggregates) and super-plasticizer was used to present computational models and estimate the compressive strength of concrete.

Up to date, various applied machine learning models have been recruited to simulate the mechanical properties of concretes including Gene Expression Programming (GEP) [13], Random Forest (RF) [14], Model Tree (MT) [15], Artificial Neural Network (ANN) [16], Adaptive Neuro-Fuzzy Inference Systems (ANFIS) [17], Support Vector Machines (SVM) [13, 18] and Multivariate Adaptive Regression Splines (MARS) [19, 20]. It is worth noting that ML techniques have also been used to model several other properties of concrete [21–25]. For example, Ashrafi et al. presented an integrative predictive formula-based model to indicate relationship of mixture and hardened strength of foamed concrete [26]. Compressive strength prediction of plastic concrete using four computational intelligence and statistical approaches was studied by Amlashi et al. [27]. Sun et al. developed an enhanced SVM method to evaluate CS and permeability coefficient [28]. Shahmansouri et al. performed GEP technique to estimate properties of eco-environmentally concrete containing zeolite [29]. Feng et al. applied machine learning method named an adaptive boosting to predict the hardened strength [30]. Iqbal et al. implemented collected data records to propose goodness of fit model for formulation of the mechanical properties of eco-friendly concrete inspired by GEP method [31]. Asteris et al. utilized artificial computing systems to predict CS of self-compacting concrete as an alternative predictive models in terms of MARS predictive strategy [32]. Behnood et al. used decision tree approach for mechanical properties of ordinary concretes [33]. In case of hybrid evolutionary MARS application in civil engineering, Al Sudani et al. [33] developed MARS Integrated with differential evolution algorithm to forecast streamflow pattern in semi-arid region. Rezaie-Balf et al. [34] predicting daily solar radiation using MARS model trained by crow search algorithm. Tien Bui et al. [35] presented new intelligent model based on GIS-based MARS and particle swarm optimization to estimate flash flood susceptible areas.

Applied regression-based soft computing methods do not consider any clear knowledge about the relationship between the inputs and corresponding target variables. It is a weakness of all machine learning models that the independent instances of the estimative database should fall within the range of predictors of calibration dataset. It seems that machine learning

techniques could be optimistically applied but due to their dependence on the initial control parameters their use is limited [35]. Aiming at improving the performance capacity of soft computing models in terms of accuracy and error, the optimization algorithms can be employed. To this end, soft computing models should be hybridized with evolutionary algorithms to create intelligent solutions called integrated evolutionary models. Based on this concept, the MARS user-defined parameters are optimized using Water Cycle Algorithm (WCA), Crow Search Algorithm (CSA) and Cat Swarm Optimization (CSO). The main parameters for tuning of MARS are maximum interaction between variables ( $I_{max}$ ), the maximum number of basis functions ( $M_{max}$ ) and smooth parameters ( $c$ ). The main characteristics of MARS model as they are the controlling elements of the complexity and generalization in the modeling process as these parameters reported by Andalib and Atry [36]. Thus, predictability of MARS model highly depends on the selected parameters quality. Friedman has selected these parameters among widespread ranges which their actual values are dependent upon the database [37]. Selection of optimum values of the parameters is a challenge in SC methods, and it is similar to an optimization effort.

In this paper, the application of WCA, CSA and CSO search engines for optimization of MARS parameters was investigated. In terms of WCA, to prove the concept of evolutionary optimization, the proposed algorithm selected to integration process based on high convergency and stability of the final solutions. The performance capability of WCA to fully explore the search space more accurately was implemented using the process of water cycle in nature. This may be utilized to search the local solution space, while the computational procedure of flowing can be utilized for the global search for enhancement of the convergence speed. Besides, population-based optimization algorithm including CSA and CSO which was applied to control the diversity due to application of limited and simple adjustable parameters. Also, aforementioned evolutionary algorithms have developed optimal performances for designing reliable solution for the mathematical test functions and operations of the systems [38–40]. Many researchers have proven superiority of WCA, CSA and CSO algorithms over other available algorithms [36–39]. It is an advantage for proposed meta-heuristic algorithms that they can be combined with MARS to detect hyper-parameters such as  $M_{max}$ ,  $I_{max}$  and  $c$ .

As mentioned previously, classical machine learning models are trustworthy means for prediction the concrete properties but they could not generate a relationship between the main concrete properties and the mixture design codes. The main goal of this research is to develop a new evolutionary formula-based models for predicting the Compressive Strength (CS) of HS concrete through adoption of a self-adaptive approach which is based on MARS integrated WCA

(MARS-WCA), MARS integrated CSA (MARS-CSA) and MARS integrated CSO (MARS-CSO). Besides, another motivation of this research is to investigate the effect of uncertainties of proposed SC models for HS concrete mixtures. For this aim, by considering the influential input parameters as random variables, the effects of uncertain parameters on the reliability analysis results are investigated for different values of target CS using the Monte Carlo Simulation (MCS) method and external validations to satisfy model's verification.

This research is presented in four sections. Section one introduces the topic of this research. Section two deals with the applied MARS and evolutionary algorithms, theoretical framework and HS concrete database details. Section 3 describes the standalone MARS and the proposed evolutionary models together with the evaluation process. Section 4 summarizes this the key-findings of this study.

## 2 Materials and Methods

### 2.1 Theoretical Framework and Data Description

In Construction of buildings sector, to optimize the size of columns in the lower floors HS concrete can be utilized specially in in high-rise buildings. One of the most advantages of HS concrete is for deflection reduction due to a higher elastic modulus [41]. In addition, it provides a higher tensile strength, greater durability, reduced creep than traditional concrete. By using HS concrete, it can be concluded that the cost of construction of buildings can be decreased with the currently prevailing prices of concrete and steel and make the members slender for more flexible structure [4, 11, 41]. This type of concrete can also be used to reduce slab depths and, therefore, a building's overall height. This can result in significant cost savings while slimmer high-strength concrete columns can increase the overall net area.

In this study, 324 experimental mix design aiming to determine the compressive strength of HS concrete based on the influencing factors was gathered from the Al-Shamiri et al. [42] for constructing accurate models. In the mentioned research, the extreme learning machine (ELM) and ANN approaches was developed to estimate the CS of HS concrete. The developed ELM model requires more hidden neurons than ANN model, but its training speed is extremely fast as no iteration is required during the learning phase. The results shown that the developed model was reliable for predicting the CS of HS concrete, but they do not have capabilities to extract equation from their predicted values for future studies. Recently, Hameed et al. [43] used high-order response surface model as the prediction model to accurately predict the compressive strength of HS concrete based on dataset

presented by Al-Shamiri et al. [44]. They also used dimensional variables to develop proposed model for predicting of compressive strength of HS concrete.

The descriptive measures including mean, minimum, maximum, standard deviation, Skewness and Kurtosis are summarize in Table 1 to introduce the characteristics of explained dataset. This data set can represent the whole population or a sample of it. Statistical measures can be classified as measures of central tendency and measures of spread. Skewness assesses the extent to which a variable's distribution is symmetrical. If the distribution of responses for a variable stretches toward the right or left tail of the distribution, then the distribution is referred to as skewed. Kurtosis is a measure of whether the distribution is too peaked. Based on the statistical analysis presented in Table 1, it can be deduced that CS and cement content ranges [37.5 – 73.6MPa] and [284 – 600kg/m<sup>3</sup>], respectively.

Principally, some of admixture materials which is considered as the input parameters for the development of AI models could have a high positive and negative dependency to each other, which can result in poor efficiency and accuracy of the modeling. In addition, extracting the most influential independent variables on the compressive strength could be difficult to explain for modeling. In this regard, the correlation matrix which depicts the correlation between all the possible independent variables is proposed in this study (Table 2). According to Table 2, no significant correlations are observed between the independent input variables for the prediction of CS of HS concrete.

For developing of experimental records, three scenarios including 70–30, 75–25, and 80–20 percentages of the training and testing groups, respectively, were considered. To do so, linear regression method was performed to evaluate proposed scenarios. Table 3 presented results of proposed scenarios that investigated in this study. According to Table 3, optimal scenario for development of AI methods was 75–25 splitting percentage in terms of statistical metrics.

One of the important steps in machine learning models is randomization in which dataset should randomly mixed to prevent overfitting, and then, processed data are categorized into two groups, training (to build up the predictive model) and testing (to evaluate the constructed model). In this study, about 75% of the experimental data (243 data samples) is for development (training phase) and the remainder (81 data samples) should be kept unseen for evolutionary MARS models. For experimental details on material specimens, mixture proportion and curing conditions of these specimens, the readers are referred to [44]. The descriptive measures and marginal plot of the input parameters for modeling of the CS of HS concrete are presented in Table 2 and Fig. 1. The marginal plot allows studying the relationship between 2 numeric variables and the distribution of x and y axes variables using a histogram. This helps to visualize the

**Table 1** Statistical analysis of input and output variables

Variables	Mean	Standard deviation	Median	Kurtosis	Skewness	Minimum	Maximum
W (Kg/m <sup>3</sup> )	170	8.18	170	− 1.5	0	160	180
C (Kg/m <sup>3</sup> )	417.81	77.22	411	− 0.53	0.4	284	600
FA (Kg/m <sup>3</sup> )	767.71	85.45	769.5	− 0.44	− 0.19	552	951
CA (Kg/m <sup>3</sup> )	898.51	42.93	898	− 1.5	− 0.01	845	989
SP (%)	0.95	0.5495	1	− 0.72	0.02	0	2
CS (MPa)	51.93	9.446	48.9	− 1.02	0.44	37.5	73.6

**Table 2** Correlation matrix of the input variables against target variable

Parameter	CS	W	C	FA	CA	SP
CS	1					
W	0.28	1				
C	0.76	0.18	1			
FA	0.46	0.16	0.32	1		
CA	0.19	0	0.01	0.22	1	
SP	0.58	0.019	0.34	0.105	0.013	1

**Table 3** Data splitting evaluation

Model	Data splitting percentage					
	70–30%		75–25%		80–20%	
	R	RMSE	R	RMSE	R	RMSE
Training phase	0.887	1.239	0.911	0.891	0.862	1.785
Testing phase	0.856	1.386	0.905	0.886	0.829	2.214

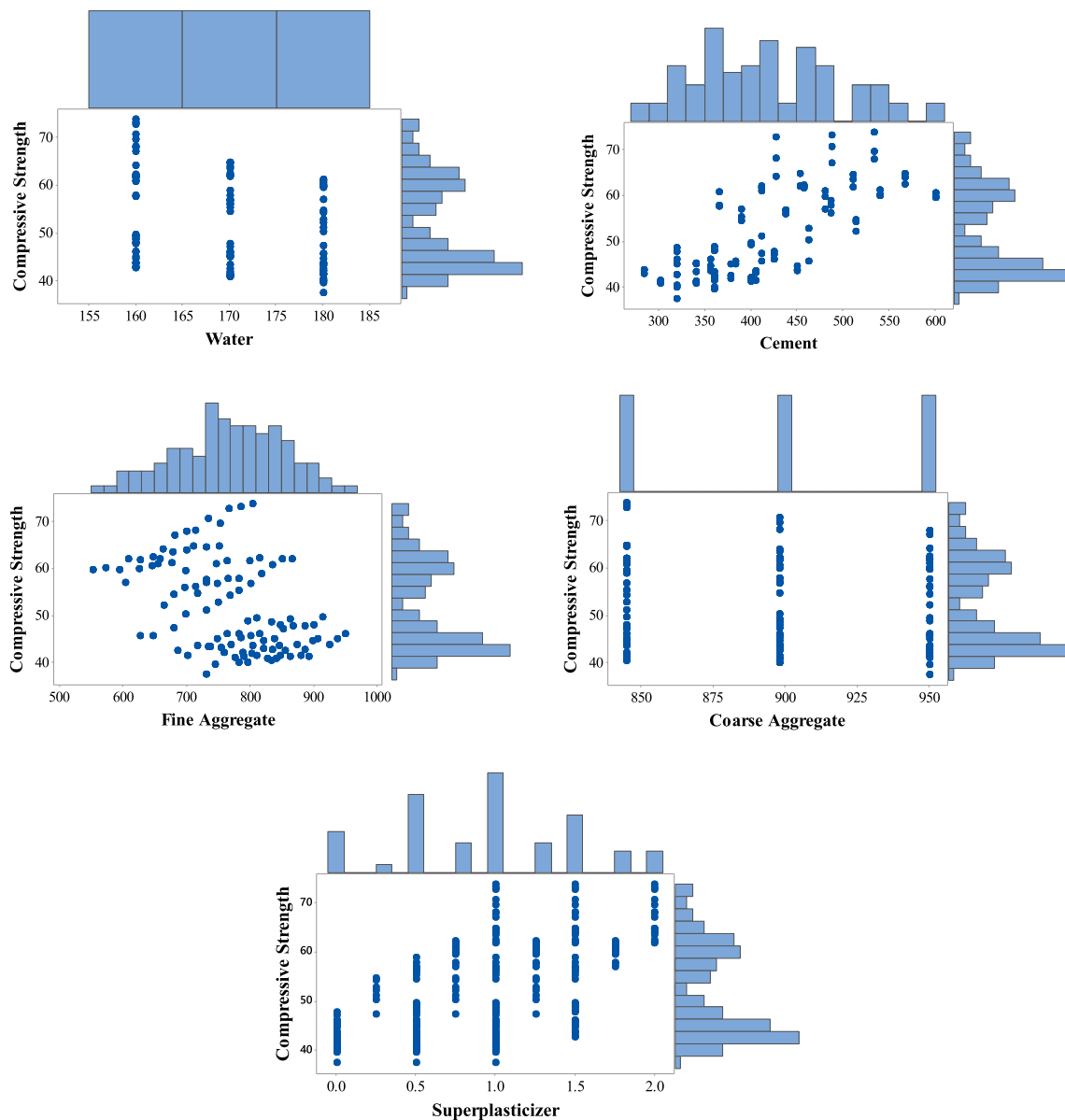
distribution intensity at different values of variables along both axes.

### 2.2 Multivariate Adaptive Regression Splines (MARS)

MARS is a nonparametric paradigm consisting splines for nonlinear modeling between the independent and target variables of a knowledge system. For this purpose, it utilizes a regression-based intelligence algorithm [37]. The Basis Function (BF) based on slope of regression line which extracted in final model (one spline to more) has changed. MARS being a flexible method in assessing nonlinear procedure in predictor and target variables, in comparison with other common method applied in regression-based data intelligent approaches [43, 45]. MARS could search probable situation between in all degrees and finding solution for variable’s interactions due to extract complicated data structures from multi-dimensional datasets [37]. The general equation for MARS function is as follows:

$$f(x) = \beta_0 + \sum_{m=1}^M \beta_m \lambda_m(x) \tag{1}$$

Here,  $f(x)$  denotes the predicted response corresponding to the predictor variable  $x$ .  $\beta_0$  and  $\beta_m$  are the predicted constant coefficients in order to attain the best data fit and  $\lambda_m$  is a spline function.  $M$  is the number of basis functions. In MARS, the knots locations are selected through an adaptive regression algorithm and BFs are created using a stepwise search [46]. Two steps are adopted in MARS for optimization namely the forward stage and backward stage. To reduced data over-fitting, a large number of BFs are utilized in the forward stage. MARS adopted the generalized cross-validation (GCV) algorithm to remove pruned BFs, which as presented in Eq. 4, helps with preventing over-fitting in the backward stage [47]. The generalized cross-validation formula is as



**Fig. 1** Marginal plot of each input variable vs. compressive strength along with distribution of each variables (unite of compressive strength, super-plasticizer, and other variables are MPa, percentage of cement, and Kg/m<sup>3</sup>, respectively)

follows:

$$GCV = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - f(x_i)]^2}{\left[1 - \frac{C(B)}{N}\right]^2} \tag{2}$$

In the above equation, N is the number of training data records, C (B) indicated the effective number of parameters in BFs, C denotes the penalty factor corresponding to each basis function and B is the number of BF. C (B) could be presented as:

$$C(B) = (B + 1) + dB \tag{3}$$

where B is the number of spline function and d is the penalizing parameter. After obtaining the MARS model, overall relationship is acquired using combination of all BFs including intercept BF and pairwise BF with interaction.

### 2.3 Water Cycle Algorithm (WCA)

WCA has been a nature-inspired optimizing solution proposed based on the observations from nature phenomena such as river, sea, mountain cloud and rain. Within the hydrologic cycle or water cycle, the clouds are formed by evaporation, also they are generated through photosynthesis by plants and

trees and transfer to atmosphere. A precipitation when happen that the atmosphere gets colder and clouds get condense [38]. Thus, underground water is filled with melted snow or rainfalls. Then, the water moves under the ground as the surface water which flows over the ground. Therefore, the cycle continues by evaporation from the rivers and streams [39].

Similar to other meta-heuristic algorithms, each raindrop is taken as an initial population in this method. Initially, we assume we have rain or precipitation. So, the best individual or rainfalls are chosen as a sea and some other good rainfalls are assumed as a river. Thus, in a  $N_{var}$  dimensional optimization procedure a raindrop would be an array of  $1 \times N_{var}$  which is calculated by the follow [48]:

$$\text{Raindrop} = [x_1, x_2, x_3, \dots, x_N] \tag{4}$$

To compute the evolutionary process, alternative procedure which is raindrop’s matrix for the size of  $N_{pop} \times N_{var}$  is created which is the raindrop’s population:

$$\text{Population of raindrops} = \begin{bmatrix} \text{Raindrops}_1 \\ \text{Raindrops}_2 \\ \text{Raindrops}_3 \\ \vdots \\ \text{Raindrops}_{N_{pop}} \end{bmatrix} \tag{5}$$

The remained raindrops are considered as streams which flow into the rivers and sea. Here, we assume a parameter named  $N_{sr}$ , which is the sum of rivers plus a single sea (Eq. 6). The remained population is computed using Eq. 7:

$$N_{sr} = \text{Number of Rivers} + 1(\text{sea}) \tag{6}$$

$$N_{\text{Raindrops}} = N_{\text{pop}} - N_{sr} \tag{7}$$

Now, each river adsorbs the stream water in proportion with the flow intensity. Therefore, the amount of water entering the river from each stream is different from the other streams. Eq. 8 is used to designate/assign raindrops to rivers and sea, on the basis of flow intensity:

$$NS_n = \text{round} \left\{ \left| \frac{\text{cost}_n}{\sum_{i=1}^{N_{sr}} \text{cost}_i} \right| \times N_{\text{Raindrops}} \right\}, \tag{8}$$

$n = 1, 2, \dots, N_{sr}$

In the above equation,  $NS_n$  denotes the number of streams flowing to a certain sea or rivers. The flowchart of the water cycle optimization algorithm is shown in Fig. 2.

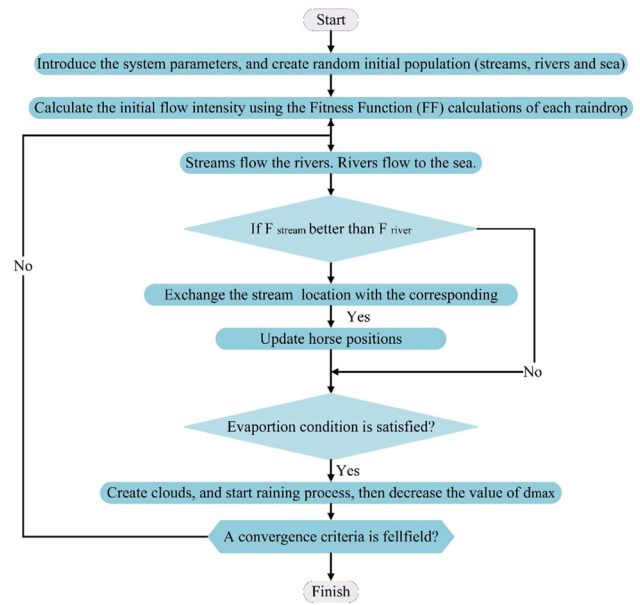


Fig. 2 Flowchart of water cycle optimization algorithm

### 2.4 Crow Search Algorithm (CSA)

Crows have several characteristics: one of them is the memorization of faces, the second is communication via advanced ways and third is their capability in hiding and recovering the food when needed. These features of crows help them in discovering what others have hidden and steal them after leaving of the owner. Askar Zadeh [39], considering the abovementioned characteristics of crows, proposed an evolutionary algorithm which himself called it CSA for solving of the complicated engineering challenge.

As other meta-heuristic algorithms, in CSA we assume a dimensional environment with a number of crows. The population size which is the crow number here is denoted by  $N$ . The crow position  $i$  at time (iteration) within the search environment is defined by a vector  $x^{i, iter} = [x_1^{i, iter}, x_2^{i, iter}, \dots, x_d^{i, iter}]$  where  $i = (1, 2, \dots, N)$  and  $iter = (1, 2, \dots, iter_{max})$ . Here,  $iter_{max}$  represents the maximum number of iterations.

The parameter of awareness probability (AP) in CSA helps with balancing between the intensification and diversification [33]. By decrease in the AP value, CSA tends to apply the local region searching, so a small AP value increases intensification. Also, by increase in the AP value, the searching probability of the neighborhood of current ideal solutions reduces, leading to increase in diversification. The steps for implementing CSA are as follows [49]:

- Step 1: Definition of the optimization problem, decision variables and constrains and valuation of CSA user-defined

parameters including the flock size, maximum number of iterations ( $iter_{max}$ ), Flight Length (FL) and AP.

- Step 2: Initializing the memory of crows in a d-dimensional search space using Eqs. 9 and 10.

$$\text{Position} = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_d^1 \\ x_1^2 & x_2^2 & \dots & x_d^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^N & x_2^N & \dots & x_d^N \end{bmatrix} \tag{9}$$

$$\text{Memory} = \begin{bmatrix} m_1^1 & m_2^1 & \dots & m_d^1 \\ m_1^2 & m_2^2 & \dots & m_d^2 \\ \vdots & \vdots & \ddots & \vdots \\ m_1^N & m_2^N & \dots & m_d^N \end{bmatrix} \tag{10}$$

- Step 3: Evaluation of the fitness function so that per each crow, the quality of its position is determined based on decision variable and objective function values.
- Step 4: Generating new positions in this way that if crow  $i$  wants to generate a new position, it should select randomized of the flock crows (for example  $j$ ) and attempt to find the position of foods that are hidden by this crow ( $m^j$ )
- Step 5: Checking capability of the new location for each crow.
- Step 6: Assessing the objective function of new position per each crow
- Step 7: crows' memory will be updated using Eq. 11.

$$m^{i, iter+1} = \begin{cases} x^{i, iter+1} & f(x^{i, iter+1}) \text{ is better than } f(m^{i, iter}) \\ m^{i, iter} & \text{otherwise} \end{cases} \tag{11}$$

In the above equation,  $f(\cdot)$  represents the value of objective function.  $M^{i, iter}$  denote the memory of crow  $i$  in iteration  $iter$  and  $x^{i, iter}$  is the position of crow  $i$  in iteration ( $iter$ ). In the final step, the termination criterion is checked by repeating steps 4–7 to reach  $iter_{max}$ . Figure 3 presents flowchart of the CSA optimization method.

### 2.5 Cat swarm Optimization (CSO)

Although resting behavior of cats is generalized, they are highly alert about their environment and any moving object. This inherent behavior helps cats with finding their preys. The cats spend very little time for chasing prey with respect to their resting time. Chu and Tsai proposed their CSO method which was inspired by hunting pattern of the cats in two modes namely the "seeking mode" (for the resting time of cats) and "tracing mode" (for the time of chasing preys) [41].

In CSO, the number of populations comprised of cats is implemented and prepared randomized distribution in the N-dimensional solution space. This population is divided into

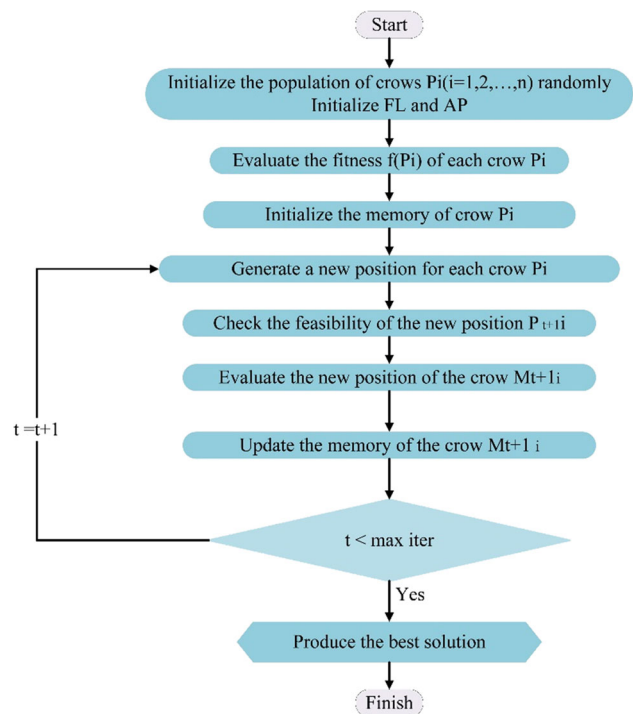


Fig. 3 Flowchart of crow search algorithm

two sub-groups of "seeking mode" where cats are resting and are looking around their surroundings. And the "tracing mode" where they begin to move around and chase their preys. The mixture of these two modes helps with global solution in the N-dimensional space. As the cats spend too little time in the tracing mode, their number in this subset is small. For this purpose, the Mixture Ratio (MR) is utilized which has a low value. After grouping the cats into these two modes, new positions and objective functions could be calculated, and the cat with the optimistic searching criteria is keep in the memory. The above stages are repeated till the stopping criteria are met [50]. Flowchart of cat swarm optimization algorithm is depicted in Fig. 4. For more information about the computational procedures of CSO [41] study have considered.

### 2.6 Description of Proposed Evolutionary-Based MARS Trained by Optimization Algorithm

In present study, an evolutionary-based predictive methods were implemented and proposed. One of the main motivations of this study was to achieve the optimum values for three hyper-parameters of MARS model in the prediction of CS of HS concrete using evolutionary-based solutions. Three hyper-parameters of MARS that can affect the model's predictivity in terms of accuracy are  $M_{max}$ ,  $c$ ,  $I_{max}$ . Among various evolutionary optimization techniques such as random search, Nelder–Mead search, genetic algorithms, grid search,

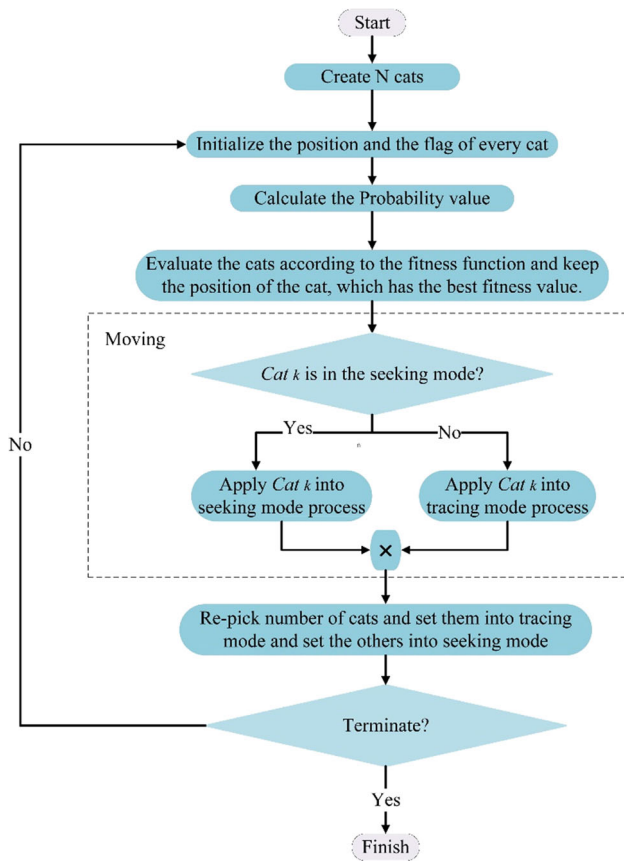


Fig. 4 Flowchart of cat swarm optimization algorithm

pattern search, heuristic search, this study recruited two bio-inspired algorithms based on the strategic behavior of cats and crows, namely CSO and CSA, and a natural-inspired optimization algorithm (e.g., WCA)-based water cycle production to calculate the optimum hyper-parameters. The four main steps of the proposed adaptive machine learning techniques are described as following:

- Step 1: The accuracy or performance of the hybrid MARS models was compared with evolutionary algorithm's greedy selector to assess the fitness function. In the present research, the MSE metric (objective function) was used as the stopping criteria. Hence, Eq. 12 was proposed as follows to calculate the summation of the training and testing RMSE for controlling model fitness function.

$$F = E_{\text{Calibrate}} + E_{\text{Validate}} \quad (12)$$

Where in the above equation,  $E_{\text{calibrate}}$  and  $E_{\text{validate}}$  denote the calibration and validation procedure's error, respectively. In Eq. 12, mean square error is incorporated as the index for prediction error. In Eq. 12, the fitness function represents the balance achieved between the model complexity and generalization.

- Step 2: The initial values of three hyper-parameters (i.e.,  $M_{\text{max}}$ ,  $c$ ,  $I_{\text{max}}$ ) of MARS model for prediction of CS of HS concrete were randomly generated based on a uniform distribution, as given in Eq. 13. The initial values of the abovementioned hyper-parameters are expressed in the following formula.

$$X = (Ub - Lb) \times \text{rand} + Lb \quad (13)$$

- Step 3: The evolutionary algorithm's population during the optimization process included a large number of the variables which were the MARS's hyper-parameters. At each generation, MSE as the candidate objective function of each set of hyper-parameters was calculated. The three algorithmic approaches, CSO, CSA, and WCA, discarded inferior set of hyper-parameters and gradually guided the population toward regions featuring good set of hyper-parameters. According to this mechanism, better solutions of  $bfs$  and  $c$  can be employed for modeling CS of HS concrete in the training phase of MARS model.
- Step 4: The optimization process would stop when the considered  $rmse$  as the stopping criteria was achieved. Finally, the optimal prediction of CS with fittest parameter values was determined when the stop criterion was reached. The pseudocode of hybrid mars-based three evolutionary algorithms in predicting CS of HS concrete is proposed in Fig. 5.

## 2.7 Statistical Measures

In this study, the effectiveness of the WCA, CSO, and CSA algorithms to optimize MARS model is evaluated several performance metrics (Eqs. 14–19) as follows,

- (1) Pearson correlation coefficient (R):

$$R = \frac{\sum_{i=1}^M (CS_{\text{Obs}} - \overline{CS_{\text{Obs}}}) \cdot (CS_{\text{Pre}} - \overline{CS_{\text{Pre}}})}{\sqrt{\sum_{i=1}^M (CS_{\text{Obs}} - \overline{CS_{\text{Obs}}})^2 \sum_{i=1}^M (CS_{\text{Pre}} - \overline{CS_{\text{Pre}}})^2}},$$

Range = (0, 1); Ideal value = 1 (14)

- (2) Nash–Sutcliffe Efficiency (NSE):

$$NSE = 1 - \frac{\sum_{i=1}^M (CS_{\text{Pre}} - CS_{\text{Obs}})^2}{\sum_{i=1}^M (CS_{\text{Obs}} - \overline{CS_{\text{Obs}}})^2},$$

Range =  $(-\infty, 1)$ ; Ideal value = 1 (15)



**Fig. 5** Pseudo code of proposed evolutionary algorithms for MARS model optimization

```

Hybridized MARS model using WCA, CSO, and CSA algorithms

Set the input parameters.

Define the considered MSE error function

Initialize population size randomly for selected algorithms individually.

While (MSEtr+MSEte)>defined MSE do

    Create the population size (particles)

    Calculate the fitness function
    {

    For each particle in the population do

        Update raindrop, cat, and crow according to "updating position" for WCA,
        CSO, and CSA algorithms, respectively.

    End for

    }

    End while

Return best raindrop, cat, and crow position.

Train the optimized MARS parameters based WCA, CSO, and CSA algorithms

Calculate MSEtr+MSEte for evolutionary-based MARS model

Optimized MARS model is obtained to simulate CS of HSC.
    
```

(3) Mean Absolute Error (MAE):

$$MAE = \frac{\sum_{i=1}^M |CS_{Pre} - CS_{Obs}|}{N},$$

Range =  $(-\infty, +\infty)$ ; Ideal value = 0 (16)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^M (CS_{Pre} - CS_{Obs})^2},$$

Range =  $(-\infty, +\infty)$ ; Ideal value = 0 (18)

(4) Wilmot’s Index of agreement (WI):

$$WI = 1 - \frac{\sum_{i=1}^M (CS_{Obs} - CS_{Pre})^2}{\sum_{i=1}^M (|CS_{Pre} - \overline{CS_{Obs}}| + |CS_{Obs} - \overline{CS_{Obs}}|)^2},$$

Range = (0, 1); Ideal value = 1 (17)

(6) Legates–McCabe’s Index (LMI):

$$LMI = 1 - \left[ \frac{\sum_{i=1}^M |CS_{Pre} - \overline{CS_{Pre}}|}{\sum_{i=1}^M |CS_{Obs} - \overline{CS_{Obs}}|} \right],$$

Range =  $(-\infty, 1)$ ; Ideal value = 1 (19)

(5) Root Mean Square Error (RMSE):

In the above equations,  $CS_{Obs}$  and  $CS_{Pre}$  denote the experimental and prediction of CS values, respectively.  $\overline{CS_{Obs}}$  and  $\overline{CS_{Pre}}$  are the average values of experimental and prediction, respectively. Also, M represent the total number of the samples.

**Table 4** Basis functions of developed MARS method

BF	Equation
BF1	$\max(0, 533 - C)$
BF2	$\max(0, W - 170)$
BF3	$\max(0, 170 - W)$
BF4	$\max(0, FA - 786)$
BF5	$\max(0, 786 - FA)$
BF6	$\max(0, CA - 898)$
BF7	$\max(0, 898 - CA)$

**Table 5** Analysis of variance for input variables of MARS

Function	GCV	STD	No. of BFs	Variable(s)
1	42.570	7.367	2	W
2	12.603	4.144	1	C
3	19.275	7.494	2	FA
4	15.299	4.641	2	CA

### 3 Application Results and Discussion

#### 3.1 MARS Formulation for CS Prediction

In this study, Jekabsons open source code which presented [47] was employed for development of piecewise-linear MARS. The best performance of MARS model for CS of HSNc possessed 8 basis functions, while the maximum interaction level was set in second-order interaction. In the present research, smooth parameters investigated in default value (1–8) recommended by Jekabsons and best value was defined  $c_{best} = 1$  based on minimum mean square error. Total number of effective parameters was 18.5, and GCV was 4.102. The details of BFs are reported in Table 4. Moreover, in this study, cross-validation method was utilized for prevention of model capability assessing bias. Table 5 presented Analysis of Variance (ANOVA) that is known for assessing the interactions of predictors and recognizing the most influential ones using GCV evaluation metric. The interpretable nonparametric formula-based model was defined as following:

$$\begin{aligned}
 CS_{MARS} = & 57.768 - 0.058 \times BF1 - 0.852 \times BF2 \\
 & + 0.967 \times BF3 - 0.107 \times BF4 + 0.073 \\
 & \times BF5 - 0.102 \times BF6 + 0.111 \times BF7 \quad (20)
 \end{aligned}$$

#### 3.2 Evolutionary-MARS Formulation for CS Prediction

In computer sciences, accurate selection of parameters corresponding to the soft computing methods is the very important procedure in achieving high performance. For example, in ANN model weights, neurons, layers and biases are among the crucial control parameters that should be determined. Even if in addressing a problem, the model provides good results but an incorrect selection of parameters could result into undesirable outcome. One technique to find the appropriate parameters is combination of the previous experiences with a restricted heuristic search of optimistic solutions. But this may take a lot of time and not yield a result.

MARS model is dependent on the parameters included in it such as the smooth parameter  $c$ , maximum basis function  $M_{max}$ , and maximum interaction  $I_{max}$ . On the other hand, finding optimal parameters simultaneously is very difficult due to the many available choices, but finding appropriate parameters can significantly improve MARS capability in predicting the procedure although suitable values may fall outside the suggested ranges. Therefore, in this study, evolutionary MARS model is incorporated to help users to face this challenge.

At the first stage, MARS is incorporated to handle the underlying function. Then, a new MARS method is built per each set of parameters that are provided by meta-heuristic algorithm (e. g., WCA, CSA and CSO). The greedy selector in aforementioned algorithms will compare the capability of proposed formula-based model with respect to the objective function assessment. After the end of train step, MARS model is incorporated to verify the data rows. It is obvious that due to good training of data we may face over-fitting. Therefore, combination of calibration error and validation error can result into an optimally balanced model with minimum errors. At the second stage, meta-heuristic algorithms applied to search for the best parameter setting value such as  $M_{max}$ ,  $I_{max}$ , and  $c$ . When the stopping condition is met, the optimization would be terminated. In this study, the generation number is depicted to present the convergence rate of each optimization algorithm up to 100 generations. Ultimately, when the stopping criterion is satisfied, the optimal predictive model with the fittest parameter setting is obtained. So, evolutionary MARS has accomplished the calibrating process and is now prepared to simulate new input patterns. The related equations of BFs of evolutionary MARS are indicated in Table 6 for formulation of CS. The interpretable evolutionary MARS predictive formulas were developed as following:

$$\begin{aligned}
 CS_{MARS-WCA} = & 52.721 - 0.070 \times BF1 - 1.549 \\
 & \times BF2 - 0.056 \times BF3 - 0.087 \\
 & \times BF4 + 0.015 \times BF5 - 0.005
 \end{aligned}$$

**Table 6** Basis functions of developed evolutionary MARS methods

BF	Equation		
	MARS-WCA	MARS- CSA	MARS- CSO
BF1	$\max (0,533 -C)$	$\max (0,533 -C)$	$\max (0,533 -C)$
BF2	$\max (0, W -170)$	$\max (0, W -170)$	$\max (0, W -170)$
BF3	$\max (0, FA -786)$	$\max (0,170 -W)$	$\max (0,170 -W)$
BF4	$\max (0, CA -898)$	$\max (0, FA -786)$	$\max (0, FA -786)$
BF5	$BF2 \times \max (0, C -450)$	$\max (0,786 -FA)$	$\max (0,786 -FA)$
BF6	$BF3 \times \max (0,170 -W)$	$\max (0, CA -898)$	$\max (0, CA -898)$
BF7	$\max (0,170 -W) \times \max (0, CA -898)$	$BF2 \times \max (0, C -450)$	$\max (0,898 -CA)$
BF8	$\max (0,170 -W) \times \max (0,898 -CA)$	$BF2 \times \max (0,450 -C)$	$BF2 \times \max (0, C -450)$
BF9	$BF1 \times \max (0,790 -FA)$	$BF4 \times \max (0,170 -W)$	$BF2 \times \max (0,450 -C)$
BF10	$BF2 \times \max (0, FA -628)$	$BF3 \times \max (0, CA -898)$	$BF4 \times \max (0, W -170)$
BF11	$BF2 \times \max (0,628 -FA)$	$BF3 \times \max (0,898 -CA)$	$BF4 \times \max (0,170 -W)$
BF12	$\max (0,898 -CA) \times \max (0, C -480)$	$BF1 \times \max (0,790 -FA)$	$BF3 \times \max (0, CA -898)$
BF13	$\max (0,898 -CA) \times \max (0,480 -C)$	$BF2 \times \max (0, FA -628)$	$BF3 \times \max (0,898 -CA)$
BF14	$BF1 \times \max (0,1.5 -SP)$	$BF2 \times \max (0,628 -FA)$	$BF1 \times \max (0, FA -790)$
BF15	$BF1 \times \max (0,1.5 -SP)$	$\max (0,898 -CA) \times \max (0, C -480)$	$BF1 \times \max (0,790 -FA)$
BF16	$BF3 \times \max (0, C -425)$	$\max (0,898 -CA) \times \max (0,480 -C)$	
BF17	$BF3 \times \max (0,425 -C)$	$BF1 \times \max (0,1.5 -SP)$	
BF18	$\max (0,898 -CA) \times \max (0, C -378)$	$\max (0,898 -CA) \times \max (0,769 -FA)$	
BF19	$\max (0,898 -CA) \times \max (0,378 -C)$	$BF4 \times \max (0, C -425)$	
BF20	$\max (0,170 -W) \times \max (0, C -411)$	$BF4 \times \max (0,425 -C)$	
BF21	$\max (0,170 -W) \times \max (0,411 -C)$	$\max (0,898 -CA) \times \max (0, C -378)$	
BF22	$BF2 \times \max (0,898 -CA)$	$\max (0,898 -CA) \times \max (0,378 -C)$	
BF23	$BF2 \times \max (0, C -480)$	$BF3 \times \max (0, C -411)$	
BF24	$BF2 \times \max (0,480 -C)$	$BF3 \times \max (0,411 -C)$	
BF25	$\max (0,170 -W) \times \max (0, C -400)$	$BF2 \times \max (0,898 -CA)$	
BF26	$\max (0,170 -W) \times \max (0, C -427)$		
BF27	$\max (0,170 -W) \times \max (0,427 -C)$		
BF28	$\max (0,899 -FA)$		

$$\begin{aligned}
 & \times BF6 - 0.003 \times BF7 + 0.004 \\
 & \times BF8 + 0.0002 \times BF9 + 0.003 \\
 & \times BF10 + 0.005 \times BF11 - 0.001 \\
 & \times BF12 + 0.001 \times BF13 - 0.003 \\
 & \times BF14 + 0.0007 \times BF15 + 0.001 \\
 & \times BF16 + 0.0006 \times BF17 + 0.0007 \\
 & \times BF18 - 0.002BF19 + 0.033 \\
 & \times \times BF20 - 0.071BF21 - 0.004 \\
 & \times BF22 - 0.012 \times BF23 + 0.003 \\
 & \times BF24 + 0.022 \times BF25 - 0.057 \\
 & \times BF26 + 0.070 \times BF27 + 0.055 \\
 & \times BF28
 \end{aligned}$$

(21)

$$\begin{aligned}
 CS_{MARS-CSA} = & 59.491 - 0.073 \times BF1 - 1.339 \\
 & \times BF2 + 1.173 \times BF3 - 0.114 \\
 & \times BF4 + 0.053 \times BF5 - 0.088 \\
 & \times BF6 + 0.004 \times BF7 + 0.003 \\
 & \times BF8 - 0.005 \times BF9 - 0.002 \\
 & \times BF10 + 0.006 \times BF11 + 0.0002 \\
 & \times BF12 + 0.003 \times BF13 + 0.006 \\
 & \times BF14 - 0.001 \times BF15 + 0.001 \\
 & \times BF16 - 0.004 \times BF17 + 0.0008 \\
 & \times BF18 + 0.001 \times BF19 + 0.0007 \\
 & \times BF20 + 0.0007 \times BF21 - 0.002 \\
 & \times BF22 - 0.001 \times BF23 - 0.002
 \end{aligned}$$

$$\times \text{BF24} - 0.003 \times \text{BF25} \quad (22)$$

$$\begin{aligned} \text{CS}_{\text{MARS-CSO}} = & 58.646 - 0.067 \times \text{BF1} - 1.029 \\ & \times \text{BF2} + 1.091 \times \text{BF3} - 0.111 \\ & \times \text{BF4} + 0.066 \times \text{BF5} - 0.095 \\ & \times \text{BF6} + 0.084 \times \text{BF7} + 0.002 \\ & \times \text{BF8} + 0.003 \times \text{BF9} + 0.003 \\ & \times \text{BF10} - 0.004 \times \text{BF11} - 0.003 \\ & \times \text{BF12} + 0.005 \times \text{BF13} + 0.0002 \\ & \times \text{BF14} + 0.0001 \times \text{BF15} \quad (23) \end{aligned}$$

Overall, the proposed approach of this paper is presented new insight into the investigating of the compressive strength of the HS concrete using the experimental mixture variables. To do this, two-step procedure is implemented; (1) compute the BFs of the problem as per Table 6. The interpretable MARS method generates many of BFs for experimental variables. A model is then learned from the output of each of BFs with the target variable. This means that the output of each BFs is weighted by a coefficient. A prediction is made by summing the weighted output of all of the BFs in the model. (2) Compute the CS applying Eqs. 20–23 depending on the MARS, MARS-WCA, MARS-CSA and MARS-CSO developed methods.

### 3.3 Results Comparison of Evolutionary and Benchmark Methods

The predictability and suitability of evolutionary nonparametric multivariate paradigm named MARS-WCA, MARS-CSA, MARS-CSO and standalone MARS was evaluated for compressive strength of HS concrete formulation. The six-performance metrics (R, NSE, MAE, WI, RMSE, and LMI) of the developed machine learning models on the formulation of CS is represented in Table 6. Before evaluating the results extracted by standalone and hybrid models, the overfitting issue for every model should be checked and controlled. One of the simple approaches that can assess this issue is the difference of models' performance in calibration and validation stages. When overfitting occurs, the performance metrics of the predictive models are satisfactory in the calibration stage; however, the significant weakness can be seen in the validation stage. For instance, when the difference in the indices R, NSE, and WI is less than 5%, the probability of overfitting occurrence is quite low. Therefore, overfitting of the proposed models based on the values represented in Table 7 has been avoided.

Based on the results, MARS-WCA model (R = 0.994 and WI = 0.997, NSE = 0.988) and (R = 0.991 and WI = 0.995, NSE = 0.981), outperformed other machine learning

methods in both calibration and validation stages, respectively. MARS-WCA has the lowest RMSE (0.991 MPa) and the highest LMI (0.906); it enhanced the precision of the validation stage to formulate CS in terms of LMI of the MARS-CSA, MARS-CSO and standalone MARS by 1.1%, 4.9% and 7.1%, respectively. Moreover, for further evaluation of the capacity of proposed evolutionary and standalone MARS, objective function (OBJ) index was computed. The optimistic predictive formula-based models are chosen based on OBJ through the best model's predictability. An OBJ (with range of (0, +∞) and optimum value of zero) is measured as follows:

$$\begin{aligned} \text{OBJ} = & \left( \frac{\text{No. training} - \text{No. testing}}{\text{NO. all}} \right) \frac{\text{MAE}_{\text{training}}}{R_{\text{training}}^2} \\ & + \frac{2\text{No. testing}}{\text{NO. all}} \times \frac{\text{MAE}_{\text{testing}}}{R_{\text{testing}}^2} \quad (24) \end{aligned}$$

The comparison between the simulated and observation CS values has indicated that the developed MARS-WCA model (with OBJ = 0.932), presented more efficiency beside another models (OBJ<sub>MARS</sub> = 1.734, OBJ<sub>MARS-CSA</sub> = 1.038 and OBJ<sub>MARS-CSO</sub> = 1.379).

Figures 6 and 7 illustrated the comparison of the predicted CS values that are modeled with standalone MARS model and three hybrid MARS models coupling with WCA, CSA and CSO algorithms. According to these figures, the MARS-WCA model has better visual agreement between actual and predicted values and the highest accuracy in the formulation of CS during the train and test steps based on  $y = x$  (black and dotted) lines. Figure 8 shows that hybrid MARS model coupled with WCA could predict the local maximum and minimum slightly higher than other standalone and hybrid MARS models. It can trustworthy express that the compressive strength that predicted by evolutionary models were in robustness coherence with the actual data points.

The estimation errors of CS models of HSNC in Fig. 9 presented significantly improvement of error production in model-based WCA integrated with MARS compared with other proposed models. Distribution plot of the relative errors of the standalone and evolutionary MARS models of the CS of HSNC to propagate the magnitude is presented in Fig. 10. The histogram values relative error of MARS-WCA, MARS-CSA, MARS-CSO and standalone MARS models in the interval of [-3%, 3%] account for 87.9%, 84.50%, 83.00% and 58.30% of the total samples. The number of relative error values estimated by the proposed models in the range of [-6%, 6%] further reduce. Through error interval analytical process, it is found that the estimation errors of MARS-WCA (mean = -0.004 and Standard Deviation (StDev) = 2.224) lower than other evolutionary and standalone models.

**Table 7** Performance capability of proposed soft computing models

Subset	Models	Statistical metrics					
		R	NSE	RMSE	MAE	WI	LMI
Calibration	MARS	0.979	0.958	1.871	1.615	0.989	0.803
	MARS-WCA	<i>0.994</i>	<i>0.988</i>	<i>0.991</i>	<i>0.763</i>	<i>0.997</i>	<i>0.906</i>
	MARS-CSA	0.993	0.986	1.085	0.871	0.996	0.893
	MARS-CSO	0.988	0.976	1.415	1.174	0.994	0.856
Validation	MARS	0.978	0.957	2.043	1.705	0.988	0.809
	MARS-WCA	<i>0.991</i>	<i>0.981</i>	<i>1.336</i>	<i>1.073</i>	<i>0.995</i>	<i>0.880</i>
	MARS-CSA	0.989	0.978	1.457	1.170	0.994	0.869
	MARS-CSO	0.983	0.966	1.819	1.504	0.991	0.831

Italic text refers to the best performance

**Fig. 6** Scatter plots of experimental and formulated CS for training performance

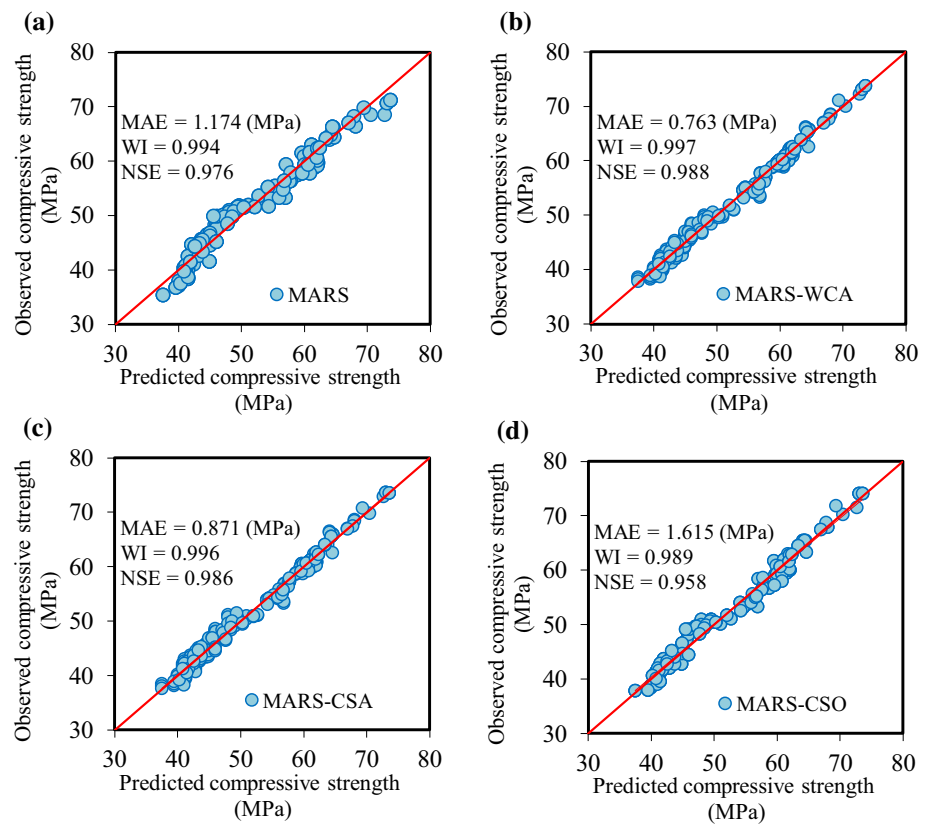
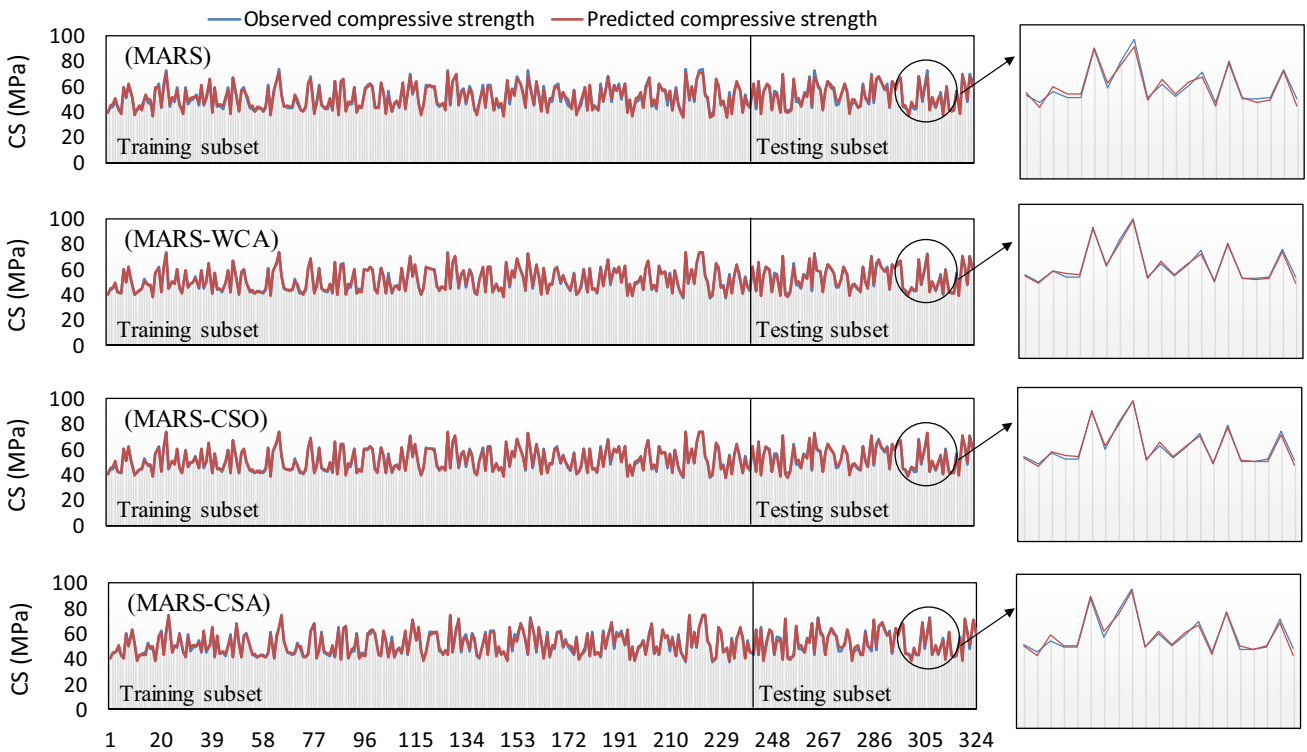
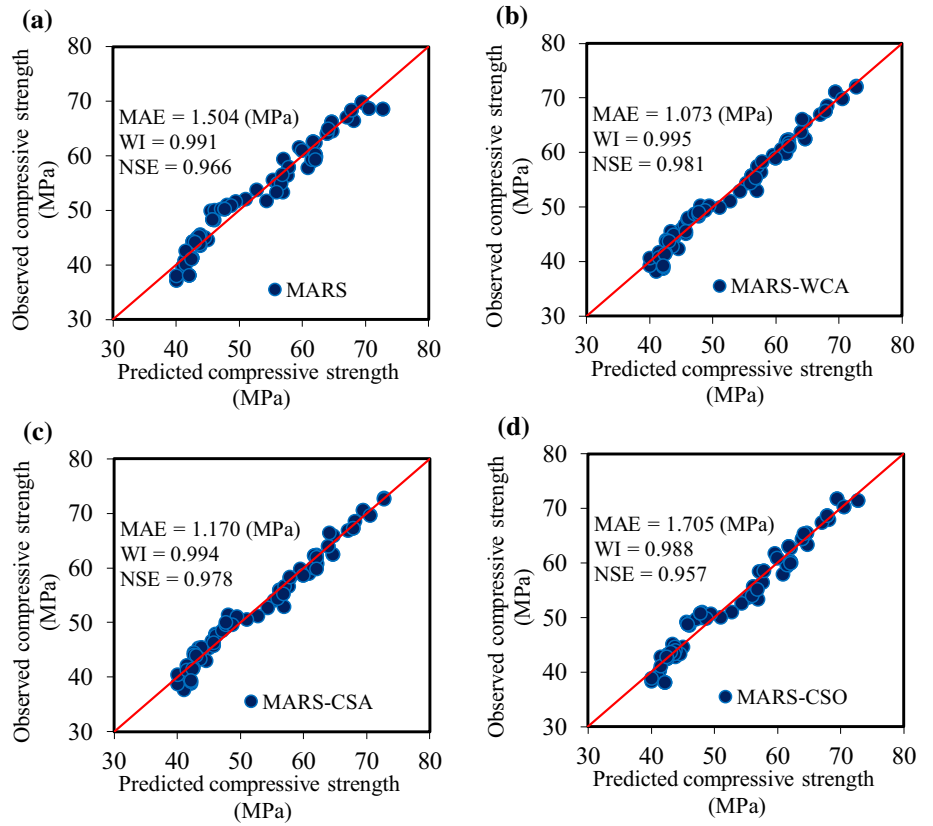


Figure 11 revealed the mean square error values as an objective function against the iteration cycles for the MARS integrated with meta-heuristic algorithms. Through Fig. 11, as the algorithm searching and optimizing the MARS parameters, the obtained results satisfied the constraints, while the value of constraint violation decreased. In this part, the CS of HS concrete estimated by the MARS-WCA, MARS, CSA and MARS-CSO models. As those evolutionary application for the hybrid models, the WCA, CSA and CSO were utilized to optimize the  $M_{max}$ ,  $I_{max}$  and  $c$ . The parameters of WCA (e.g., Number of rivers + sea (Nsr) and Evaporation condition

constant ( $d_{max}$ )), CSA (e. g., Flight Length (FL) and Awareness Probability (AP)) and CSO (e. g., Seeking Memory Pool (SMP) and Seeking Range of the selected Dimension (SRD)) are also needed to be tune up before optimization procedure of MARS method.

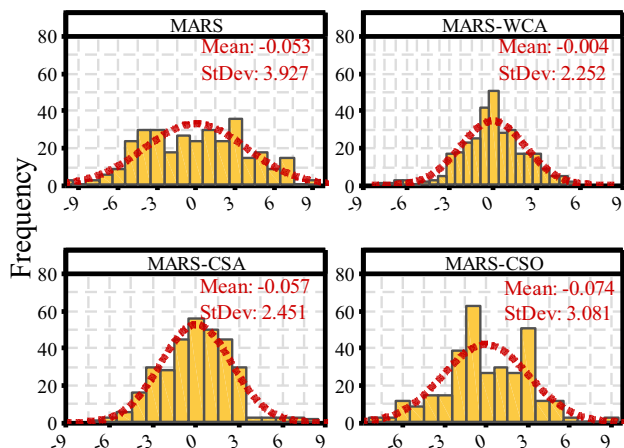
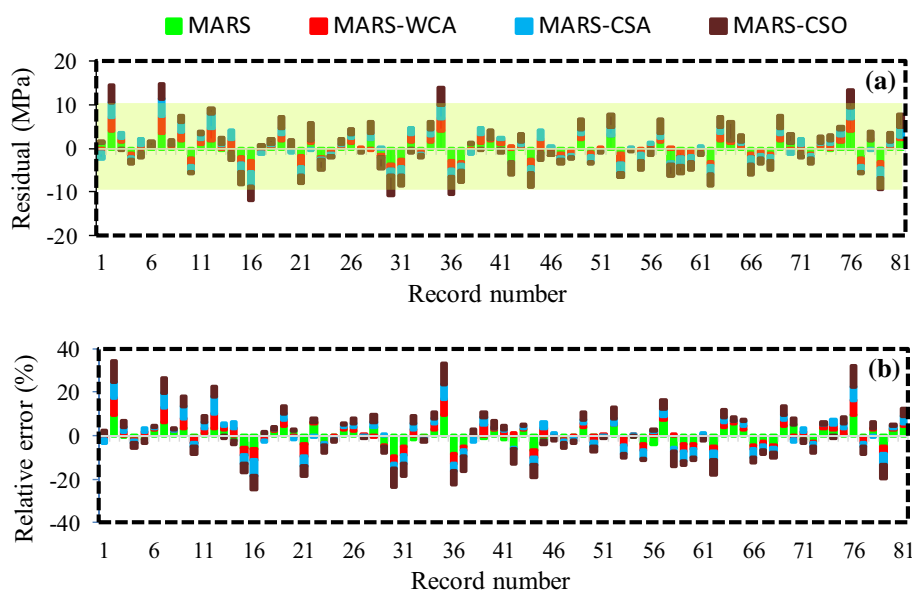
Based on Fig. 11, the speed of convergence of WCA after 100 generations is very high in probability of searching the optimum solution (decision variables) compared to CSA and CSO algorithms in same populations [46]. It is obvious that the integration of MARS with WCA also provides efficient

**Fig. 7** Scatter plots of experimental and formulated CS for testing performance

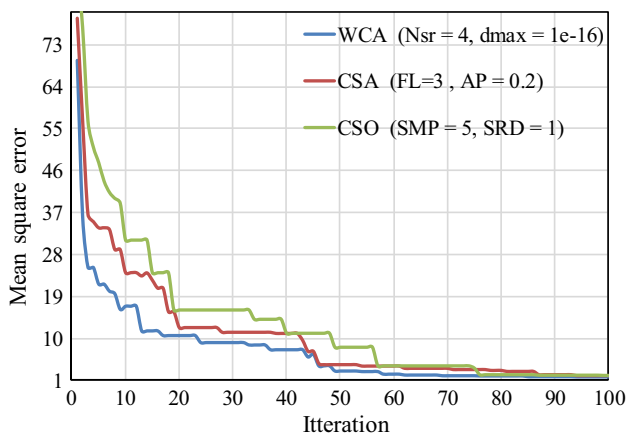


**Fig. 8** Curve fitting plots of experimental and formulated CS of HS Concrete

**Fig. 9** Graphical comparisons of error in model’s development; **a**: Residual, **b**: Relative error for validation stage



**Fig. 10** Relative error histograms of proposed models



**Fig. 11** Meta-heuristic algorithms performance to optimize MARS parameters

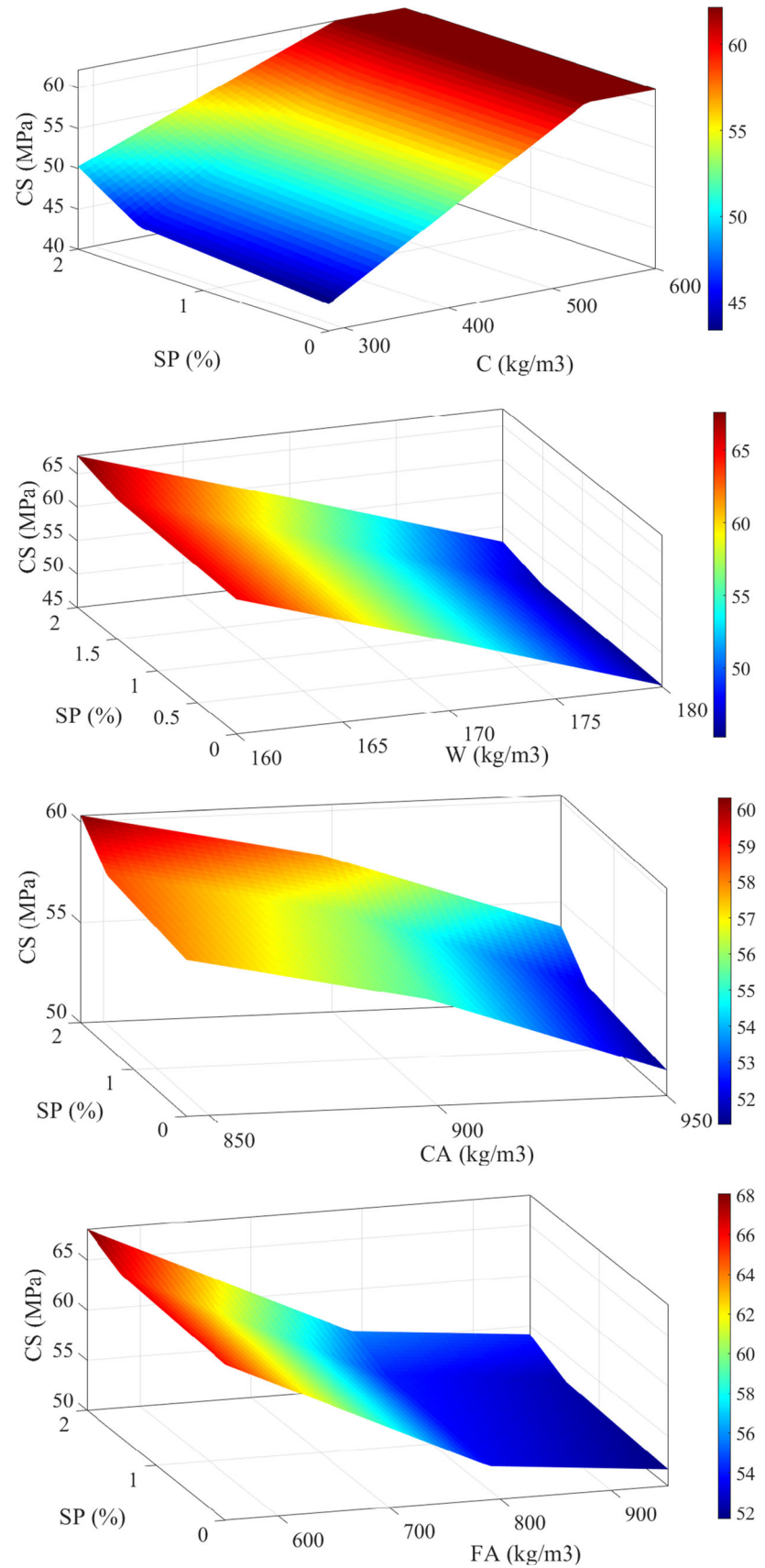
and reliable prediction for CS of HS concrete, and the well-known optimizer in terms of computational effort ( $Nsr = 4$  and  $d_{max} = 1e-16$ ) that are challenging for regularizing the ability of the WCA algorithm [51]. As can be seen from Fig. 8, the selected hyper-tuning parameters is reported  $FL = 3$  and  $AP = 0.2$  (for CSA), and  $SMP = 5$  and  $SRD = 1$  (for CSO). Finally, for further evaluation of influence of input variables on the CS of HS concrete proposed model, 3D surface diagrams for MARS-WCA (optimistic model) are presented in Fig. 12.

To compare the results of the proposed MARS-WCA model with other AI models for CS prediction of HS concrete, the best performance model of Al-Shamiri’s study [44], Back-Propagation Network (BPN) was selected. Due to the proximity of the modeling results in terms of R, RMSE, NSE and MAE, a composite metric (i.e., Performance Index (PI)) that involves two R and RMSE criteria was considered to combine and generate a single result for better comparison. The PI statistical metric changes from zero to one, with smaller values, presented a better prediction and is calculated as following:

$$PI = \frac{1}{|CS_{obs}|} \frac{RMSE}{(R + 1)} \tag{25}$$

Comparing the results of integrated MARS-WCA with BPN, it was found that the performance of both models is similar in terms of PI where the values for MARS-WCA and BPN are 0.109 and 0.104, respectively, for the prediction of CS of HS concrete. Moreover, the evaluation metrics indicate that applying optimization algorithms like WCA would improve the models’ performances remarkably. The

**Fig. 12** 3D surface diagrams of MARS-WCA model for evaluation of inputs





most important point in comparison of MARS-WCA and Al-Shamiri’s method was the applicability of the BPN model which cannot be used by researchers in the specific field of study. In other words, in this study, an explicit formulation was presented by integrated MARS-WCA which can be apply in order to predict and determine the compressive strength of high-strength concrete.

### 3.4 Validation of Formula-Based Models

#### 3.4.1 External Analysis of the Predictive Models

Validation is a crucial aspect of any quantitative structure–activity relationship modeling and a special emphasis on statistical significance and predictive ability of those models as their most crucial characteristics. Checking the robustness of the predictive models just with coefficient determination ( $R^2 > 0.5$ ) could not provide the ultimate proof of the model’s ability or capability for prediction of phenomena, and it needs an approach to establish the model robustness by randomization of response (i.e., compressive strength) [52]. External analysis as a reliable analytical method is used for comparison between the results of estimated and observation event data. Golbraikh and Tropsha [53] have adopted the new external validation criteria for evaluation of the estimation capacity of the models corresponding to the performance of testing subset based on coefficient of determination. External validation means assessing the model performance with independent samples [52]. In this method, minimum one of the coefficient of determination regression line gradients that passes through the source for estimated values against experimental CS or vice versa should be close to unity.

$$K = \sum_{i=1}^n \frac{CS_{obs} \times CS_{pre}}{CS_{pre}^2} \tag{26}$$

$$K' = \frac{CS_{obs} \times CS_{pre}}{CS_{obs}^2} \tag{27}$$

$CS_{obs}$  and  $CS_{pre}$  represent the experimental and formulated CS values, respectively. The determination coefficients passing through the source between the predicted and experimental values ( $R_0^2$ ) and conversely ( $R_0'^2$ ) are derived using the following equations:

$$R_0^2 = 1 - \sum_{i=1}^n \frac{CS_{pre}^2 (1 - k)^2}{\sum_{i=1}^n CS_{pre} - \overline{CS_{pre}})^2} \tag{28}$$

$$R_0'^2 = 1 - \sum_{i=1}^n \frac{CS_{obs}^2 (1 - k')^2}{\sum_{i=1}^n (CS_{obs} - \overline{CS_{obs}})^2} \tag{29}$$

$$m = (R^2 - R_0^2) / R^2 \tag{30}$$

$$n = (R^2 - R_0'^2) / R^2 \tag{31}$$

As the result of verification metric through the analysis based on slope of the regression ( $R_m$ ) should be greater than 0.5.

$$R_m = R^2 \times \left( 1 - \sqrt{|R^2 - R_0^2|} \right) > 0.5 \tag{32}$$

The validation measures and the related performance of CS formulation obtained by various models are reported in Table 8. Based on the results, the MARS-WCA models for compressive strength which yielded  $R_m = 0.852$  satisfied the conditions with best validation with respect to other used evolutionary and standalone SC methods such as MARS-CSA, MARS-CSO and standalone MARS models. Therefore, it is found that MARS-WCA has the highest validity for predicting CS of HS concrete and the computed correlations had not been accidentally.

#### 3.4.2 Monte Carlo Uncertainty Analysis

In this sub-section, the randomness of model’s uncertainty is performed by Monte Carlo Simulation (MCS). This procedure was first applied in military projects for simulation of the probabilistic events [51]. In this content, the target variable of the presented study (i.e., compressive strength) comprises several uncertainties like parameters uncertainty of the models, uncertainty of input variables, etc. Therefore, quantitative uncertainty of the MARS-WCA, MARS-CSA, MARS-CSO and standalone MARS models that associated with estimated CS is investigated. Based on this technique, the individual error of the predictions is calculated for all datasets (Eq. 31). The mean ( $\bar{e}$ ) and standard deviation ( $S_e$ ) of the formulation error is calculated by Eqs. 32 and 33, respectively [48]:

$$e_i = \log_{10}(CS_{pre}^i) - \log_{10}(CS_{obs}^i) \tag{33}$$

$$\bar{e} = \sum_{i=1}^n e_i \tag{34}$$

$$S_e = \sqrt{\sum_{i=1}^n \left( \frac{(e_i - \bar{e})^2}{n - 1} \right)} \tag{35}$$

Also,  $\pm 1.96 S_e$  yields 95% confidence level around predicted  $P_i$  as follows:

$$\left\{ P_i \times 10^{-\bar{e} - 1.96 S_e}, P_i \times 10^{-\bar{e} + 1.96 S_e} \right\} \tag{36}$$

The uncertainty percentage and Mean Absolute Deviation (MAD) as the two main factors in assessing the uncertainty

**Table 8** External analysis measures of predictive models

Model	$K$	$K'$	$m$	$N$	$R_m$
MARS	1.001	0.997	− 0.044	− 0.044	0.759
MARS-WCA	1.002	0.996	− 0.017	− 0.017	0.852
MARS-CSA	1.001	0.997	− 0.022	− 0.021	0.834
MARS-CSO	0.998	1.000	− 0.034	− 0.034	0.790

**Table 9** Monte Carlo result for uncertainty of models

Model	$\bar{e}$	$S_e$	Median	MAD	Uncertainty (%)
MARS	− 0.043	1.089	49.569	8.049	16.238
MARS-WCA	− 0.020	1.918	51.230	7.819	15.264
MARS-CSA	0.009	1.528	50.268	7.880	15.676
MARS-CSO	− 0.024	1.191	49.790	8.009	16.085

of the proposed models are reported in Table 9. Based on the result, the positive mean prediction error indicates that the predicted CS calculated by all these methods is higher than the observations. In addition, MARS-WCA and MARS-CSA models for CS yielded the minimum (15.26% and 15.67%) bandwidth uncertainties, respectively. Moreover, in other developed models, MARS-WCA satisfied bandwidth criteria and had lowest uncertainty.

### 3.5 Variable importance

In this study, to determine the difference of predictor’s value that will affected on target, analysis of variables importance (AVI) technique is recruited. The AVI% is calculated for each independent variable as follows [54]:

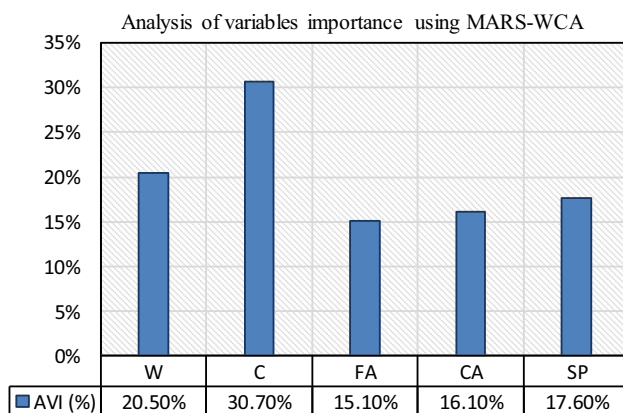
$$L_i = CS_{\max}(x_i) - CS_{\min}(x_i) \tag{37}$$

$$AVI_i = \frac{L_i}{\sum_{j=1}^M L_j} \times 100 \tag{38}$$

where  $CS_{\max}$  and  $CS_{\min}$  = maximum and minimum of the predicted compressive strength over the  $i_{th}$  input domain, where other independent variable values are equal to their average values. Figure 13 indicated the variable importance results for the simulation of CS of HS concrete using MARS-WCA model that was selected as the best model among other hybrid and standalone models. The bar plot presented that the most effective variable in CS of HS concrete is the cement content (30.7%) and water content (20.5%).

### 4 Conclusion

Achieving a high-accurate model to predict compressive strength is very important especially, in the field of concrete



**Fig. 13** Results of analysis of variables importance for corresponding compressive strength

technology. This paper evaluates the ability of the evolutionary algorithms, WCA, CSA and CSO, for optimizing hyperparameter of MARS model in order to extract new accurate formulas for the compressive strength of HS concrete. Through the results of this research, the following conclusion can be presented:

- New closed-form evolutionary nonparametric multivariate regression-based formula is proposed for the compressive strength of HS concrete. The overall nonlinear formulations given by non-parametric paradigm are presented with 29, 26, 16 and 8 BFs (Eqs. 20–22 and 18) for MARS-WCA, MARS-CSA, MARS-CSO and standalone MARS, respectively.
- Proposed CS developed models manifested that the MARS-WCA model presented more accurate prediction in comprising with the other three evolutionary and standalone techniques, with respect to R, NSE, RMSE, MAE, WI, and LMI indicators for calibrating and validating subsets. As the results of the evolutionary and original models

were similar, OBJ as the composite metric is recruited to illustrate the superiority of the model. The integrated MARS-WCA model with OBJ = 0.932 presented more efficiency compared to other models (OBJ of MARS = 1.734, OBJ of MARS-CSA = 1.038 and OBJ of MARS-CSO = 1.379).

- The proposed evolutionary models as novel machine learning tools confirmed all of the required criterion of the statistical regression line gradients for external validation. Results indicated that MARS-WCA which yielded K and k' near 1 and  $R_m$  equal to 0.852 satisfied the conditions with best validation with respect to MARS-CSA, MARS-CSO and standalone MARS models.
- The Monte Carlo uncertainty procedure for implemented evolutionary models was employed. MARS-WCA with 15.26% and MARS-CSA with 15.67% bandwidth uncertainties yielded the minimum uncertainty for predicting CS of HS concrete. The robustness of the proposed evolutionary machine learning techniques was verified in this study. Moreover, sensitivity analysis of variable importance considered the highest importance variables influenced on the compressive strength of HS concrete to be the cement content and water content with 30.7 and 20.5%.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. NoParast, M.; Hematian, M.; Ashrafi, A.; Amiri, M.J.T.; AzariJafari, H.: Development of a non-dominated sorting genetic algorithm for implementing circular economy strategies in the concrete industry. *Sustain. Prod. Consum.* **27**, 933–946 (2021)
2. Le, H.T.N.; Poh, L.H.; Wang, S.; Zhang, M.H.: Critical parameters for the compressive strength of high-strength concrete. *Cement Concrete Compos.* **82**, 202–216 (2017)
3. Xiao, H.; Zhang, F.; Liu, R.; Zhang, R.; Liu, Z.; Liu, H.: Effects of pozzolanic and non-pozzolanic nanomaterials on cement-based materials. *Constr. Build. Mater.* **213**, 1–9 (2019)
4. Dhanalakshmi, A.; Hameed, M.S.: Review study on high strength self compacting concrete. *IJSTE-Int. J. Sci. Technol. Eng.* **4**(12), 451 (2018)
5. Xu, J.; Wang, B.; Zuo, J.: Modification effects of nanosilica on the interfacial transition zone in concrete: a multiscale approach. *Cement Concr. Compos.* **81**, 1–10 (2017)
6. Golewski, G.L.: An assessment of microcracks in the interfacial transition zone of durable concrete composites with fly ash additives. *Compos. Struct.* **200**, 515–520 (2018)
7. Medina, C.; Zhu, W.; Howind, T.; De Rojas, M.I.S.; Frías, M.: Influence of interfacial transition zone on engineering properties of the concrete manufactured with recycled ceramic aggregate. *J. Civil Eng. Manage.* **21**(1), 83–93 (2015)
8. Liu, J.; Farzadnia, N.; Shi, C.: Effects of superabsorbent polymer on interfacial transition zone and mechanical properties of ultra-high performance concrete. *Constr. Build. Mater.* **231**, 117142 (2020)
9. Carrasquillo, R.L.; Nilson, A.H.; Slate, F.O.: Properties of high strength concretesubject to short-term loads. *ACI J.* **78**, 171–178 (1981)
10. Singh, J.; Siddique, R. G.: Effect of specimen shape and size on the strength properties of high strength concrete (Doctoral dissertation) (2015)
11. Isaia, G.C.; Gastaldini, A.L.G.; Moraes, R.: Physical and pozzolanic action of mineral additions on the mechanical strength of high-performance concrete. *Cement Concr. Compos.* **25**(1), 69–76 (2003)
12. Bonen, D.; Shah, S.: The effects of formulation on the properties of self-consolidating concrete. *Concr. Sci. Eng.: Tribute Arnon Bentur* **89**, 43–56 (2004)
13. Zhang, J.; Shi, R.; Shi, S.; Alzoubi, A.K.; Roco-Videla, A.; Hussein, M.; Khan, A.: Numerical assessment of rectangular tunnels configurations using support vector machine (SVM) and gene expression programming (GEP). *Eng. Computers* **51**, 1–17 (2021)
14. Ashrafi, A.; Taheri Amiri, M.J.; Masoumi, P.; Asadi-shiadeh, M.; Yaghoubi-chenari, M.; Mosavi, A.; Nabipour, N.: Classification-based regression models for prediction of the mechanical properties of roller-compacted concrete pavement. *Appl. Sci.* **10**(11), 3707 (2020)
15. Ashrafi, A.; Taheri Amiri, M.J.; Haghighi, F.: Modeling the slump flow of self-compacting concrete incorporating Metakaolin using soft computing techniques. *J. Struct. Constr. Eng.* **6**(2), 5–20 (2019)
16. Shariati, M.; Mafipour, M.S.; Ghahremani, B.; Azarhomayun, F.; Ahmadi, M.; Trung, N.T.; Shariati, A.: A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement. *Eng. Computers* **25**, 1–23 (2020)
17. Naderpour, H.; Mirrashid, M.; Nagai, K.: An innovative approach for bond strength modeling in FRP strip-to-concrete joints using adaptive neuro–fuzzy inference system. *Eng. Computers* **36**(3), 1083–1100 (2020)
18. Bui, H.B.; Nguyen, H.; Choi, Y.; Bui, X.N.; Nguyen-Thoi, T.; Zandi, Y.: A novel artificial intelligence technique to estimate the gross calorific value of coal based on meta-heuristic and support vector regression algorithms. *Appl. Sci.* **9**(22), 4868 (2019)
19. Ashrafi, A.; Amiri, M.J.T.; Rezaie-Balf, M.; Ozbakkaloglu, T.; Lotfi-Omran, O.: Prediction of compressive strength and ultrasonic pulse velocity of fiber reinforced concrete incorporating nano silica using heuristic regression methods. *Constr. Build. Mater.* **190**, 479–494 (2018)
20. Taheri Amiri, M.J.; Ashrafi, A.; Haghighi, F.R.; Javaheri Barforooshi, M.: Prediction of the compressive strength of self-compacting concrete containing rice husk ash using data driven models. *Modares Civil Eng. J.* **19**(1), 209–221 (2019)
21. Amlashi, A.T.; Abdollahi, S.M.; Goodarzi, S.; Ghanizadeh, A.R.: Soft computing based formulations for slump, compressive strength, and elastic modulus of bentonite plastic concrete. *J. Cleaner Prod.* **230**, 1197–1216 (2019)
22. Tavana Amlashi, A.; Alidoust, P.; Pazhouhi, M.; Pourrostami Niavol, K.; Khabiri, S.; Ghanizadeh, A.R.: AI-based formulation for mechanical and workability properties of eco-friendly concrete made by waste foundry sand. *J. Mater. Civil Eng.* **33**(4), 04021038 (2021)
23. Ghanizadeh, A.R.; Rahrovan, M.: Modeling of unconfined compressive strength of soil-RAP blend stabilized with Portland cement using multivariate adaptive regression spline. *Front. Struct. Civil Eng.* **13**(4), 787–799 (2019)



24. Ghanizadeh, A.R.; Fakhri, M.: Prediction of frequency for simulation of asphalt mix fatigue tests using MARS and ANN. *Scientif. World J.* **2014**, 1–16 (2014)
25. Ghanizadeh, A.R.; Safi Jahanshahi, F.; Khalifeh, V.; Jalali, F.: Predicting flow number of asphalt mixtures based on the marshall mix design parameters using multivariate adaptive regression spline (MARS). *Int. J. Transp. Eng.* **7**(4), 433–448 (2020)
26. Ashrafiyan, A.; Shokri, F.; Amiri, M.J.T.; Yaseen, Z.M.; Rezaie-Balf, M.: Compressive strength of Foamed Cellular Lightweight Concrete simulation: New development of hybrid artificial intelligence model. *Constr. Build. Mater.* **230**, 117048 (2020)
27. Amlashi, A.T.; Alidoust, P.; Ghanizadeh, A.R.; Khabiri, S.; Pazhouhi, M.; Monabati, M.S.: Application of computational intelligence and statistical approaches for auto-estimating the compressive strength of plastic concrete. *Eur. J. Environ. Civil Eng.* **65**, 1–32 (2020)
28. Sun, J.; Zhang, J.; Gu, Y.; Huang, Y.; Sun, Y.; Ma, G.: Prediction of permeability and unconfined compressive strength of pervious concrete using evolved support vector regression. *Constr. Build. Mater.* **207**, 440–449 (2019)
29. Shahmansouri, A.A.; Bengar, H.A.; Jahani, E.: Predicting compressive strength and electrical resistivity of eco-friendly concrete containing natural zeolite via GEP algorithm. *Constr. Build. Mater.* **229**, 116883 (2019)
30. Feng, D.C.; Liu, Z.T.; Wang, X.D.; Chen, Y.; Chang, J.Q.; Wei, D.F.; Jiang, Z.M.: Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Constr. Build. Mater.* **230**, 117000 (2020)
31. Iqbal, M.F.; Liu, Q.F.; 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. *J. Hazardous Mater.* **384**, 121322 (2020)
32. Asteris, P.G.; Ashrafiyan, A.; Rezaie-Balf, M.: Prediction of the compressive strength of self-compacting concrete using surrogate models. *Comput. Concr.* **24**, 137–150 (2019)
33. Behnood, A.; Behnood, V.; Gharehveran, M.M.; Alyamac, K.E.: Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Constr. Build. Mater.* **142**, 199–207 (2017)
34. Al-Sudani, Z.A.; Salih, S.Q.; Yaseen, Z.M.: Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. *J. Hydrol.* **573**, 1–12 (2019)
35. Rezaie-Balf, M.; Maleki, N.; Kim, S.; Ashrafiyan, A.; Babaie-Miri, F.; Kim, N.W.; Alaghmand, S.: Forecasting daily solar radiation using CEEMDAN decomposition-based MARS model trained by crow search algorithm. *Energies* **12**(8), 1416 (2019)
36. Bui, D.T.; Hoang, N.D.; Pham, T.D.; Ngo, P.T.T.; Hoa, P.V.; Minh, N.Q.; Samui, P.: A new intelligence approach based on GIS-based Multivariate Adaptive Regression Splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. *J. Hydrol.* **575**, 314–326 (2019)
37. Andalib, A.; Atry, F.: Multi-step ahead forecasts for electricity prices using NARX: a new approach, a critical analysis of one-step ahead forecasts. *Energy Conver. Manage.* **50**, 739–747 (2009). <https://doi.org/10.1016/j.enconman.2008.09.040>
38. Friedman, J.H.: Multivariate adaptive regression splines. *Annals Stat.* **19**, 1–67 (1991). <https://doi.org/10.1214/aos/1176347963>
39. Eskandar, H.; Sadollah, A.; Bahreininejad, A.; Hamdi, M.: Water cycle algorithm - a novel metaheuristic optimization method for solving constrained engineering optimization problems. *Computers Struct.*, 45:151–166 (2012). <https://doi.org/10.1016/j.compstruc.2012.07.010>
40. Haddad, O.B.; Moravej, M.; Loáiciga, H.A.: Application of the water cycle algorithm to the optimal operation of reservoir systems. *J. Irrig. Drain Eng.* **141**, 451 (2015). [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000832](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000832)
41. Askarzadeh, A.: A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. *Computers Struct.* **169**, 1–12 (2016)
42. Chu, S.C.; Tsai, P.W.; Pan, J.S.: Cat swarm optimization In Pacific Rim international conference on artificial intelligence. Springer, Heidelberg (2006)
43. El-Dieb, A.S.: Performance of reinforced concrete columns using ultra-high-strength fiber-reinforced self-compacted concrete (UHS-FR-SCC). *MOJ Civil Eng.* **1**(2), 00010 (2016)
44. Al-Shamiri, A.K.; Kim, J.H.; Yuan, T.F.; Yoon, Y.S.: Modeling the compressive strength of high-strength concrete: an extreme learning approach. *Constr. Build. Mater.* **208**, 204–219 (2019)
45. Hameed, M.M.; AlOmar, M.K.; Baniya, W.J.; AlSaadi, M.A.: Prediction of high-strength concrete: high-order response surface methodology modeling approach. *Eng. Computers* **54**, 1–14 (2021)
46. Nguyen, H.; Bui, X.N.; Tran, Q.H.; Nguyen, H.A.; Nguyen, D.A.; Le, Q.T.: Prediction of ground vibration intensity in mine blasting using the novel hybrid MARS–PSO–MLP model. *Eng. Computers* **15**, 1–19 (2021)
47. Hastie, T.; Tibshirani, R.; Friedman, J.: Overview of supervised learning, in: *The Elements of Statistical Learning*, 2008, pp. 1–33, doi: [https://doi.org/10.1007/b94608\\_2](https://doi.org/10.1007/b94608_2).
48. Jekabsons, G.; VariReg.: A software tool for regression modeling using various modeling methods, Riga Technical University, 2010.
49. Sadollah, A.; Yoo, D.G.; Yazdi, J.; Kim, J.H.; Choi, Y.: Application of water cycle algorithm for optimal cost design of water distribution systems. *Eng. Optim.* **47**, 1602 (2014)
50. Gupta, D.; Sundaram, S.; Khanna, A.; Hassanien, A.E.; De Albuquerque, V.H.C.: Improved diagnosis of Parkinson's disease using optimized crow search algorithm. *Computers Electr. Eng.* **68**, 412–424 (2018)
51. Saha, S.K.; Ghoshal, S.P.; Kar, R.; Mandal, D.: Cat swarm optimization algorithm for optimal linear phase FIR filter design. *ISA transactions* **52**(6), 781–794 (2013)
52. Benigni, R.; Bossa, C.: Predictivity of QSAR. *J. Chem. Inf. Model.* **48**(5), 971–980 (2008)
53. Roy, P.P.; Roy, K.: On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* **27**, 302–313 (2008). <https://doi.org/10.1002/qsar.200710043>
54. Golbraikh, A.; Tropsha, A.: Beware of q<sup>2</sup>! *J. Mol. Gr. Modell.* **20**, 269–276 (2002). [https://doi.org/10.1016/s1093-3263\(01\)00123-1](https://doi.org/10.1016/s1093-3263(01)00123-1)
55. Binder, K.; Ceperley, D.M.; Hansen, J.-P.; Kalos, M.H.; Landau, D.P.; Levesque, D.; Mueller-Krumbhaar, H.; Stauffer, D.; Weis, J.-J.: Monte Carlo methods in statistical physics. Springer Science & Business Media, Berlin (2012)
56. Newcombe, R.G.: Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat. Med.* **17**, 857–872 (1998)

