Modeling the compressive strength of concrete containing waste glass using multi-objective automatic regression

Emadaldin Mohammadi Golafshani1 · Alireza Kashani1

Received: 1 May 2021 / Accepted: 27 April 2022 / Published online: 30 May 2022
© The Author(s) 2022

Abstract
Some grades of municipal and industrial waste glass (WG) discarded in landfills can cause environmental issues. One of the efficient solutions to overcome this issue is to use WG in concrete mixtures as aggregate or supplementary cementitious materials. Modeling the compressive strength (CS) of the produced concrete using machine learning methods can provide helpful insights into the effects of WG on concrete properties. In this study, a comprehensive database of concrete containing WG (CCWG) was gathered from 24 peer-reviewed papers. Two different scenarios were considered in the selection of input variables, and a novel machine learning method, called multi-objective multi-biogeography-based programming, was used to predict the CS of CCWG. This algorithm can automatically select the effective input variables, the structure of the equations, and its coefficients. Moreover, the proposed model optimizes the precision and complexity of the developed models simultaneously. The definition of complexity in the optimization problem can help achieve different mathematical equations with various accuracies and assist users in predicting the CS of CCWG even with a limited number of optimal input variables. The results show that the proposed algorithm can introduce several equations with different accuracies, complexities, and input variables to predict the CS of CCWG.

Keywords Concrete containing waste glass · Compressive strength · Automatic regression · Multi-objective biogeography-based programming · Machine learning

1 Introduction

The excessive consumption of natural resources in concrete production, the emission of carbon dioxide from Portland cement production, and disposing of waste materials in landfills are the main sustainability concerns in the concrete industry, which threaten the environment [1, 2]. There have been many endeavors to increase the environmental performance of the concrete industry by using waste materials as partially replaced natural aggregate and cement in the concrete mixture [3, 4]. Some of these waste materials, including recycled aggregate, have been used in concrete to reduce the consumption of natural materials and prevent their landfills without improving the mechanical properties and durability behavior of the produced concrete [5–7]. On the other hand, other waste materials, consisting of silica fume and blast furnace slag, participate in chemical reactions leading to stronger cement hydration products [8–10]. The use of all these waste materials in concrete can help move toward sustainability.

Glass is one of the high-demanding materials and consumes a substantial amount of natural resources, including sand, limestone, and soda ash [1, 11, 12]. By increasing industrialization, the consumption of glass in the industry and household has grown significantly, leading to increased waste glass (WG) production [11]. It is estimated that almost 200 million tons of WG are produced each year worldwide [13]. Although some WG is recycled into glass containers, a considerable amount of WG cannot be recycled because of the high recycling costs, color mixing, and high breaking potential [14, 15]. The discarded WG into the landfills causes a significant problem from the environmental viewpoint of inappropriate land use and
pollution of the soil and water. This non-biodegradable material can be used as coarse aggregate, fine aggregate, and powder in the concrete mix design. However, more endeavors have been paid to use WG as fine aggregate and powder in the concrete mixture. The reason is that using WG as coarse aggregate in concrete decreases the quality of concrete [16].

The compressive strength (CS) of concrete containing WG (CCWG) has been investigated by many researchers. Like other studies, the compressive and tensile strengths of concrete containing WG powder increases especially in the long term due to the formation of denser microstructure compared to concrete without WG powder [21–24]. In terms of WG fine aggregate (WGFA), the effect of fine aggregate on the CS of CCWG has been contradictory. Some researchers have found that the CS of concrete containing WGFA increases because of the better bonding of the surface texture of WGFA with the cement paste than the concrete without WGFA [20, 25]. On the other hand, several researchers have reported that the WGFA harms the CS. They have attributed the reduction of CS to the decreasing of concrete density, weaker bonding of the smooth surface of WGFA, the lower elastic modulus of WGFA compared to the natural aggregates, and the existence of lead in the WGFA [14, 26, 27].

Many attempts have been conducted to use several machine learning methods in modeling the CS of concrete containing waste materials [10, 28–42]. Different machine learning methods are available, including artificial neural networks, fuzzy systems, decision trees, and automatic regression methods. For predicting the different properties of various types of concrete containing WG, a limited number of studies have been conducted. Ray et al. [43] developed an artificial neural network model to predict the compressive and tensile strengths of concrete containing WR and tin can fiber based on their experiments. Ghorbani et al. [44] used an artificial neural network to model the dynamic characterization of recycled glass-recycled concrete according to their experiments. As can be observed, the numbers of data used for developing these models are small, and because of the existence of other types of components in the concrete mixtures of these studies, the applications of the developed models are limited. The automatic regression-based machine learning methods extend the metaheuristic optimization algorithms. They can automatically find the structure of the mathematical equation, its corresponding coefficients, and the effective input variables. Biogeography-based programming (BBP) is a relatively novel automatic regression method, which has been successfully employed in several practical applications, including the prediction of the CS of concrete containing silica fume [42], estimation of the climbing rate of slip formwork systems [45], prediction of the shear strength of concrete beams reinforced by fiber-reinforced polymer bars without transverse reinforcement [46], the formulation of the elastic modulus of recycled aggregate concrete [7], self-compacting concrete [47], and the dynamic modulus of asphalt mixture [48].

In this study, a comprehensive database consisting of 830 data records of CCWG was gathered from the literature. To model the CS of this type of concrete, various mathematical equations with different accuracies and complexities can be generated with the assumption of different control parameters for the BBP. To overcome this issue, the BBP was seen as a multi-objective optimization problem that was solved by a metaheuristic optimization algorithm called biogeography-based optimization. The proposed procedure can lead to developing several formulas for predicting the CS of CCWG with different precisions, complexities, and input variables.

2 Developed soft computing models

2.1 Biogeography-based optimization (BBO) algorithm

Biogeography-based optimization (BBO) algorithm is a metaheuristic optimization method inspired by the distribution of species among islands in an ecosystem [49]. Several factors, including topographic features, temperature, humidity, and climate, can affect the distribution of species in a nature biography so that more species live in high-quality islands, and they tend to migrate to low-populated islands. In the BBO, each possible solution (PS) and its related decision variables are simulated by an island and its characteristics, respectively. Each PS is evaluated by an objective function in which the lower objective value of a PS shows its better performance in a minimization problem. The quality of each PS is called habitat suitability index (HSI) in such a way that a high-quality PS has higher HSI. There are two mechanisms to look for the optimal solution in the BBO, including exploration and exploitation carried out by mutation and migration operators, respectively. For the migration operator, two emigration and immigration probabilities are assigned to each PS, which
are the probabilities of migration from and to the PS, respectively. In this regard, all existing PSs are ranked in each epoch from the worst to the best ones based on their HSI, and more emigration and fewer immigration probabilities are assigned to a PS with higher HSI. Figure 1 depicts the linear migration curve, including emigration and immigration curves in which the \( E \) and \( I \) are the maximum values of emigration and immigration probabilities, respectively. These values are assumed one in this study. The emigrated and immigrated PSs are selected according to their assigning immigration and emigration probabilities and the roulette wheel selection mechanism. Then, to generate the migrated PS (MPS\(_i\)), the migration is carried out from the emigrated PS (EPS\(_i\)) to the immigrated PS (IPS\(_i\)) using the following equation:

\[
\text{MPS}_i(\text{RDV}) = \alpha \text{IPS}_i(\text{RDV}) + (1 - \alpha)\text{EPS}_i(\text{RDV})
\]

where RDV is a random decision variable, and \( \alpha \) is a random number between zero and one. It means that the values of decision variables of MPS change between those of EPS and IPS.

The mutation operator is a probabilistic procedure that avoids trapping the BBO in the local optimum and enhances the diversity of PSs in the algorithm. Having a mutation probability, this operator gives this chance to PSs with low and high HSIs to be improved by modifying their decision variables randomly. Figure 2 summarizes the different steps of the BBO algorithm.

### 2.2 Multi-objective biogeography-based optimization (MOBBO) algorithm

From the viewpoint of objective function number (OFN), the optimization problems are divided into single and multi-objective optimization problems. In the single-objective problem, the quality of the two PSs can be easily assessed by comparing their objective function values. In the case of a multi-objective problem, the comparison of PSs is more complicated. The dominance concept is used to compare the performance of PSs in this type of problem.

The PS\(_i\) dominates the PS\(_j\), if only the following conditions are met.

\[
\forall k \in \{1, \ldots, \text{OFN}\} : z^k_i \leq z^k_j \quad \text{and} \quad \exists k \in \{1, \ldots, \text{OFN}\} : z^k_i < z^k_j
\]

where \( z^k_i \) and \( z^k_j \) are the \( k \)th object function values related to the \( i \)th and \( j \)th PSs, respectively. Indeed, there are several optimal solutions in the multi-objective optimization problem, which make an optimal Pareto front (PF). In the MOBBO, the emigration and immigration probabilities of each PS are specified based on their HSIs, and the ranking of PSs in each epoch is determined using two following strategies:

- All non-dominated PSs of each epoch are determined, which make the first PF, called optimal PF. Then, the members of this PF are disregarded, and the non-dominated PSs of the remained PSs are specified, which build the second PF. This process will continue until all possible PFs are obtained, and each PS has its own PF. The members of the first PF have the top priority compared to the PSs of other PFs, and the members of the last PF have the least priority.
- Motivated by the non-sorting genetic algorithm II (NSGA-II) [50], all members of each PF are ranked using the crowding distance (CD) measure, defined as follows:

\[
\text{CD}_j = \sum_{k=1}^{\text{OFN}} \frac{z^k_{j,\text{next}} - z^k_{j,\text{former}}}{z^k_{j,\max} - z^k_{j,\min}}
\]

where \( \text{CD}_j \) is the CD of the PS\(_j\) in the \( p \)th PF, \( z^k_{j,\max} \) and \( z^k_{j,\min} \) are the maximum and minimum objective function values of the \( k \)th function for PSs in the \( p \)th PF, respectively. Moreover, \( z^k_{j,\text{next}} \) and \( z^k_{j,\text{former}} \) are the closest next and former neighbors of the PS\(_j\) from the viewpoint of objective function values, respectively. To have more uniform PFs, the PSs with the higher CDs have more priority than PSs with lower CDs in each PF, and more emigration probabilities are assigned to them. However, the emigration probabilities of PSs in the high-ranked PF are higher than those in the low-ranked PF. The flowchart of the MOBBO algorithm is given in Fig. 3.

### 2.3 Multi-biogeography-based programming (MBBP) algorithm

Biogeography-based programming (BBP), as a branch of automatic regression, mimics the BBO algorithm in regression problems. This method looks for the best mathematical equation, which automatically maps the
Fig. 2 The flowchart of the BBO algorithm

Fig. 3 The flowchart of the MOBBO algorithm
input of a system on its output(s) [51]. In the BBP, an island is a mathematical equation with a tree structure. The first node located on the top of a tree structure is called the root node, while the nodes in the lowest positions of a tree are called leaf nodes. The maximum tree depth is defined as the longest path from the root node to the leaf nodes. The root and intermediate nodes can be mathematical functions defined by the user, and variables and constants can be assigned to the leaf nodes. Moreover, a root node can be a variable or a constant with a tree depth of zero. Figure 4 illustrates an island with a tree structure which represents the mathematical equation of \( y = 2.3 \ln(x_2) + \frac{x_1}{x_1} \). To evaluate an island, different statistical indicators can be used, defined later. Whatever the predicted values of an island are closer to the observed results, the prediction error of the related mathematical equation is lower, and the corresponding HSI would be higher.

In the multi-BBP (MBBP) algorithm, an archipelago is defined as several islands which combine as a linear combination. Figure 5 reveals an archipelago with two islands with the maximum tree depth of three. This archipelago represents the mathematical equation \( y = a_0 + a_1y_1 + a_2y_2 \) in which the regression coefficients \((a_0, a_1, a_2)\) are optimized in such a way that the prediction error of the corresponding mathematical equation is minimized. The root-mean-square deviation was used as the cost function in this study, defined later.

Based on the obtained HSIs of all archipelagos, their emigration and immigration probabilities are calculated. To migrate from an emigrated archipelago to an immigrated one, a random sub-tree (ST) of a random island of the immigrated archipelago is selected and is replaced by a random ST of a random island of the emigrated archipelago. In the case of mutation, an archipelago is chosen randomly from the existing archipelagos in each epoch based on the mutation probability. Then, a random ST is selected from a random island of the chosen archipelago, and it is replaced by a randomly generated ST. It is noteworthy that the maximum depth of the generated archipelagos should be less than or equal to the defined maximum tree depth in both migrated and mutated archipelagos. Figure 6 schematically demonstrates the migration process between archipelagos.

### 2.4 The proposed multi-objective MBBP (MOMBBP) algorithm

The MBBP has the ability to find the relationship between the input and output variables automatically. Unlike the conventional regression method in which the regression type should be set at first, the MBBP looks for the regression type using the predefined functions, as well as the required coefficients simultaneously. However, finding a suitable mathematical equation can be considered as a multi-objective optimization problem. The first objective is the accuracy of the developed mathematical equation, and the second one is the complexity of the developed archipelago. The statistical indicators determine the accuracy of a developed model. At the same time, the number of nodes (i.e., number of input variables, constants, and functions) in the related archipelago specifies the complexity of a model. For an archipelago with the maximum number of islands MNI, the maximum tree depth MTD and the maximum argument number of functions MANF, the maximum complexity is \( MNI \times MTD \times MANF \). In general, increasing the complexity of an archipelago can raise the training capacity of the MBPP, and it may reduce the generalization of the developed model. Moreover, fewer input variables participate in simple models, which shows their importance compared to other input variables. Additionally, more variables appear in the most complicated MBBP models, and the produced models can have more accuracy. In the proposed MOMBBP, different mathematical equations are developed with various architectures and input variables. In this regard, the MOBBO algorithm is used to find the optimal Pareto front. The pseudo-code of the MOMBBP is shown in Fig. 7, and more details are given below.

**Phase 1** The gathered dataset is divided randomly into training and testing datasets which the former is used to
train the MOMBBP, and the latter is served to verify the developed models.

**Phase 2** Prior to running the proposed algorithm, all control parameters related to the proposed MOMBBP algorithm are adjusted. These parameters include the training dataset ratio \( (n_{\text{train}}) \), the testing dataset ratio \( (n_{\text{test}}) \), the number of archipelagos \( (n_{\text{archipelagos}}) \), the maximum number of generation \( (n_{\text{generation}}) \), the maximum number of algorithm repetition \( (n_{\text{repetition}}) \), the maximum number of islands \( (n_{\text{islands}}) \), the maximum tree depth \( (d_{\text{tree}}) \), the used functions, the mutation probability \( (p_{\text{mutation}}) \), the range of constants \( (r_{\text{constants}}) \).

**Phase 3** The archipelagos are created randomly. To generate an archipelago, a random integer between 1 and \( n_{\text{islands}} \) is selected as the number of its islands, and the islands are produced using the ramped half-and-half method. This method is a compound of growing and full methods. In the grow method, a root node is randomly chosen as a function or terminal (variables and constants). If a terminal is chosen, it will be a leaf; otherwise, a randomly used function is assigned to this node, and the number of branches of the selected function is determined based on the number of its arguments. This procedure continues for the generated nodes of the lower levels of the tree. If a node reaches the depth of \( d_{\text{tree}} \), a random terminal is assigned to it. In the case of the full method, a random function is assigned to each node from the root to the nodes with the tree depth of less than \( d_{\text{tree}} \), and the random terminals are assigned to the nodes in the lowest positions. This method allows the tree to grow entirely as far as it can. In the ramped half-and-half method, half of the archipelagos are generated using the growing method, while the others are created using the full method.

**Phase 4** Using the training dataset, the regression coefficients of archipelagos are determined, and their HSIs are calculated using the statistical indicators. Moreover, the number of nodes used in the archipelagos is calculated as their complexities.

**Phase 5** Having the HSIs and the complexities of all archipelagos, the non-dominated archipelagos are specified. Then, the CDs of all archipelagos are determined, and all mathematical equations are sorted from the best to the worst based on their HSIs at first and then according to their CDs. After ranking the archipelagos, their emigration and immigration probabilities are calculated.

**Phase 6** In this phase, migration and mutation operators are used to generate the new archipelagos. In this regard, an archipelago is selected based on its immigration probability. Then, an emigrated archipelago is chosen according to its emigration probability and using the roulette wheel selection mechanism. Based on the predefined \( p_{\text{mutation}} \), the mutation is carried out to increase the variety of archipelagos.

**Phase 7** The HSIs and complexities of the generated archipelagos are calculated. Then, the old and new archipelagos are merged, their non-dominated solutions and SDs are determined and finally ranked from the best to the worst archipelagos. Next, the \( n \) best archipelagos are chosen as the main archipelagos, and the rest are discarded.

**Phase 8** The 7 and 8 phases are repeated iteratively until the termination conditions of the MOMBBP algorithm are satisfied. To do so, \( N_{\text{generation}} \) was used in this study. After
finishing the algorithm’s runs, the best archipelagos, which are the optimal Pareto front, are saved and used for more investigations.

3 Data collection

Collecting a comprehensive and credible database is necessary to model the CS of CCWG. In this regard, a total number of 830 data records from 24 peer-reviewed journals were used to develop the MOMBBP model [21–23, 25, 52–71]. In this study, two scenarios were considered to select input variables of the model. First, the amounts of Cement (C), Water (W), Blast Furnace Slag (BFS), Fly Ash (FA), Natural Coarse Aggregate (NCA), Natural Fine Aggregate (NFA), Superplasticizer (SP), and Waste Glass (WG) as well as the Age of Concrete (AC), the Passing Percentages of waste glass at the sieves of 75 μm (PP75) and 45 μm (PP45) are used as the input variables. Second, the W/B (where B is the usual binder consisting of cement, BFS and FA), BFS/C, FA/C, NCA/C, NFA/TA, SP/B (in percentage) ratios, the amount of Passing WG at the sieve of 75 μm to Cement (PWG75/C), the amount of Passing WG at the sieve of 45 μm to Cement (PWG45/C), the amount of Remaining WG on the sieve of 75 μm to Total Fine Aggregate (RWG75/TFA), and also the AC are the input variables. It is worth mentioning that the WG was used in the gathered mix designs of CCWG as sand and powder distinguished from each other by their sizes. The statistical characteristics of the gathered database are given in Table 1. Among 830 data records, 179 and 211 records include BFS and FA in the concrete

Begin
Prepare the dataset
Set the control parameters
Generate initial $n$ archipelagos
Calculate the HSIs and complexities of archipelagos
Determine the non-dominated archipelagos
Calculate the CDs of the archipelagos
Sort the archipelagos
Assign the emigration and immigration probabilities to the archipelagos

for $i = 1$ : Maximum generation number

for $i = 1$ : $n$

if rand() $\leq$ the immigration probability of the $i$th archipelago

Use the roulette wheel selection mechanism to choose an emigrated archipelago
Select a random ST of a random island of the $i$th archipelago
Select a random ST of a random island of the emigrated archipelago
Migrate from the emigrated archipelago to the $i$th archipelago

else if rand() $\leq$ mutation probability

Select a random ST of a random island of the $i$th archipelago
Replace the selected ST by a random ST

end

if there is a change in the $i$th archipelago

Calculate the HSI and complexity of the new archipelago

end

end

Combine the old and new archipelagos
Specify the non-dominated archipelagos of the combined archipelagos
Calculate the CDs of the combined archipelagos
Sort the combined archipelagos
Choose the best $n$ archipelagos of the combined archipelagos

end

Save the optimal archipelagos

Finish
mixture, respectively. Additionally, there are different types of standard specimens for measuring the CS of CCWG in the database, including cubic specimens with 100 mm and 150 mm and cylindrical specimens with dimensions of 100 mm × 200 mm and 150 mm × 300 mm. The CS of all specimens with different shapes and sizes was converted to cylindrical specimens with dimensions of 100 mm × 200 mm based on the conversion equations given in [72]. Meanwhile, the NFA was entirely replaced by the WG in the concrete mix design of 47 data records.

### 4 Results and discussion

To assess the accuracy of the developed models and compare their performance, several statistical indicators were served, including root-mean-square deviation (RMSD), relative RMSE (RRMDS), mean absolute deviation (MAD), coefficient of determination (R²), correlation coefficient (R), and OBJ, given in the following:

\[
RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (PV_i - ER_i)^2} \quad (4)
\]

\[
RRMDS = \frac{RMSD}{ER} \quad (5)
\]

\[
MAD = \frac{1}{n} \sum_{i=1}^{n} |PV_i - ER_i| \quad (6)
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (PV_i - ER_i)^2}{\sum_{i=1}^{n} (ER_i - \bar{ER})^2} \quad (7)
\]

\[
R = \frac{n \sum_{i=1}^{n} ER_i PV_i - (\sum_{i=1}^{n} ER_i)(\sum_{i=1}^{n} PV_i)}{(n \sum_{i=1}^{n} ER_i^2 - (\sum_{i=1}^{n} ER_i)^2)(n \sum_{i=1}^{n} PV_i^2 - (\sum_{i=1}^{n} PV_i)^2)} \quad (8)
\]

\[
OBJ = \frac{\frac{n_{\text{train}} - n_{\text{test}}}{} \frac{RMSD_{\text{train}} + MAD_{\text{train}}}{} + 1 + R_{\text{train}}}{\frac{n_{\text{train}} + n_{\text{test}}}{} + 2n_{\text{test}} \frac{RMSD_{\text{test}} + MAD_{\text{test}}}{} + 1 + R_{\text{test}}} \quad (9)
\]

where PVᵢ and ERᵢ are the predicted value and experimental results of the ith data record, respectively. ER is the average value of the experimental results. n, nᵢtrain, nᵢtest are the number of data records, the number of data records of the training dataset, and the number of data records of the testing dataset, respectively. The OBJ value is a combinational merit that simultaneously considers three statistical indicators for the training and testing datasets. For the NRMSD, there are different categories for determining the quality of the developed model: NRMSD ≤ 0.1 shows “Excellent efficiency,” 0.1 ≤ NRMSD ≤ 0.2 shows “Good efficiency,” 0.2 ≤ NRMSD ≤ 0.3 shows “Fair...

| Minimum | Maximum | Average | Standard deviation | Skewness | Kurtosis |
|---------|---------|--------|--------------------|----------|----------|
| C (kg/m³) | 100 | 578 | 328.61 | 117.56 | −0.33 | −0.50 |
| W (kg/m³) | 140 | 266.40 | 189.67 | 23.53 | 0.97 | 1.91 |
| BFS (kg/m³) | 0 | 292 | 22.75 | 59.86 | 3.41 | 11.42 |
| FA (kg/m³) | 0 | 146 | 15.35 | 38.34 | 2.47 | 4.39 |
| CA (kg/m³) | 315 | 1300 | 955.81 | 246.35 | −0.63 | −0.13 |
| NFA (kg/m³) | 0 | 1080 | 617.95 | 245.03 | −0.81 | 0.55 |
| SP (kg/m³) | 0 | 8.91 | 1.71 | 2.81 | 1.25 | −0.06 |
| PPₗₐ (%) | 0 | 100 | 22.26 | 39.14 | 1.39 | 0.01 |
| PPₗₘ (%) | 0 | 100 | 15.99 | 32.75 | 1.72 | 1.16 |
| WG (kg/m³) | 0 | 914 | 183.645 | 202.61 | 1.53 | 1.88 |
| W/B | 0.28 | 2.42 | 0.65 | 0.48 | 2.53 | 5.59 |
| BFS/C | 0 | 15.99 | 0.11 | 0.32 | 3.75 | 13.32 |
| FA/C | 0 | 0.43 | 0.05 | 0.10 | 2.22 | 3.94 |
| NCA/C | 1 | 6.14 | 3.30 | 1.30 | 0.44 | −0.61 |
| NFA/TA | 0.32 | 2.90 | 0.54 | 0.24 | 5.17 | 41.12 |
| SP/B × 100 | 0 | 2.96 | 0.43 | 0.71 | 1.40 | 0.69 |
| PWGₗₐ/C | 0 | 1.33 | 0.09 | 0.19 | 4.08 | 20.42 |
| PWGₗₘ/C | 0 | 1.23 | 0.05 | 0.15 | 5.33 | 33.92 |
| RWGₗₑ/TFA | 0 | 1 | 0.21 | 0.27 | 1.55 | 1.67 |
| AC (days) | 1 | 365 | 59.80 | 86.39 | 2.54 | 6.13 |
| CS (MPa) | 0.34 | 84.20 | 34.45 | 18.44 | 0.18 | −0.32 |
efficiency” and $0.3 \leq \text{NRMSD}$ shows “Poor efficiency” [35, 73].

Before running the proposed MOMBBP algorithm, the control parameters defined earlier should be set. Table 2 shows the control parameters adjusted for the MOMBBP algorithm in this study which were selected based on trial and error.

After running the proposed method ten times, the optimal PF with the best performance was achieved. Figure 8 depicts the best optimal PF for two scenarios in which the horizontal and vertical axes are the complexity and statistical indicator (i.e., RMSD and OBJ), respectively. As expected, the prediction error decreases when the complexity of models increases. The complexities of the optimal models in the first and second scenarios are in the range of [1, 108] and [1, 65], respectively. The numbers of optimal solutions of the first and second scenarios are 74 and 53, respectively. The best solutions, illustrated as solid-filled markers, have the $R$ equal or greater than 0.85, and their numbers are 65 and 46 for the first and second scenarios, respectively.

The simplest optimal models in both scenarios consist of only one node, which is the root node. These nodes are, respectively, $C$ and $W/B$ in the first and second scenarios, which shows their importance compared to other input variables. Moreover, there are no significant differences between the optimal models with the complexities of 56 and 108 in the first scenario and the optimal models with the complexities of 37 and 65 in the second scenario for unknown data. All the eleven input variables of the first

| Parameters          | Values       | Parameters          | Values       |
|---------------------|--------------|---------------------|--------------|
| $n_{\text{train}}$  | 0.75         | $n_{\text{test}}$  | 0.25         |
| $n_{\text{archipelagos}}$ | 100–150–200 | $n_{\text{generation}}$ | 1000–5000 |
| $n_{\text{repetition}}$ | 10          | $n_{\text{islands}}$ | 5           |
| $d_{\text{tree}}$  | 6–8          | $p_{\text{mutation}}$ | 0.1–0.4     |
| $r_{\text{constants}}$ | $[-10, 10]$ | Used functions     | +, −, ×, /, log, $\phi^2$, $\phi^3$, $\phi^4$, $\phi^5$, $\sqrt{\phi}$, $\sqrt{\phi}$, $\sqrt{\phi}$ |

Fig. 8 The optimal models obtained by the proposed method: (a) RMSD-complexity plot of the training dataset, (b) RMSD-complexity plot of the testing dataset, and (c) OBJ-complexity plot.
Fig. 9 The NRMSD-complexity plot of the best models for two scenarios

![NRMSD-complexity plot](image)

Table 3 Participated input variables in the best optimal models of the first scenario

| Complexity (Node number) | Participated input variables | Complexity (Node number) | Participated input variables |
|--------------------------|-----------------------------|--------------------------|-----------------------------|
|                          | C | W | BFS | FA | NCA | NFA | SP | PP75 | PP45 | WG | AC |
| 10                       |   |   |     |    |     |     |    |      |      |    |    |
| 11                       |   |   |     |    |     |     |    |      |      |    |    |
| 12                       |   |   |     |    |     |     |    |      |      |    |    |
| 13                       |   |   |     |    |     |     |    |      |      |    |    |
| 14                       |   |   |     |    |     |     |    |      |      |    |    |
| 15                       |   |   |     |    |     |     |    |      |      |    |    |
| 16                       |   |   |     |    |     |     |    |      |      |    |    |
| 17                       |   |   |     |    |     |     |    |      |      |    |    |
| 18                       |   |   |     |    |     |     |    |      |      |    |    |
| 19                       |   |   |     |    |     |     |    |      |      |    |    |
| 20                       |   |   |     |    |     |     |    |      |      |    |    |
| 21                       |   |   |     |    |     |     |    |      |      |    |    |
| 22                       |   |   |     |    |     |     |    |      |      |    |    |
| 23                       |   |   |     |    |     |     |    |      |      |    |    |
| 24                       |   |   |     |    |     |     |    |      |      |    |    |
| 25                       |   |   |     |    |     |     |    |      |      |    |    |
| 26                       |   |   |     |    |     |     |    |      |      |    |    |
| 27                       |   |   |     |    |     |     |    |      |      |    |    |
| 28                       |   |   |     |    |     |     |    |      |      |    |    |
| 29                       |   |   |     |    |     |     |    |      |      |    |    |
| 30                       |   |   |     |    |     |     |    |      |      |    |    |
| 31                       |   |   |     |    |     |     |    |      |      |    |    |
| 32                       |   |   |     |    |     |     |    |      |      |    |    |
| 33                       |   |   |     |    |     |     |    |      |      |    |    |
| 34                       |   |   |     |    |     |     |    |      |      |    |    |
| 35                       |   |   |     |    |     |     |    |      |      |    |    |
| 36                       |   |   |     |    |     |     |    |      |      |    |    |
| 38                       |   |   |     |    |     |     |    |      |      |    |    |
| 40                       |   |   |     |    |     |     |    |      |      |    |    |
| 42                       |   |   |     |    |     |     |    |      |      |    |    |
| 44                       |   |   |     |    |     |     |    |      |      |    |    |
| 45                       |   |   |     |    |     |     |    |      |      |    |    |
| 49                       |   |   |     |    |     |     |    |      |      |    |    |
| 54                       |   |   |     |    |     |     |    |      |      |    |    |
scenario participate in the optimal solution with the complexity of 108. In contrast, all input variables except for PP$_{75}$ are involved in the optimal solution with the complexity of 56. It means that knowing the value of PP$_{75}$ does not significantly improve the accuracy of the prediction. In the case of the second scenario, FA/C and RWG$_{75}$/TFA are the only input variables involved in the optimal solution with the complexity of 108. This indicates that these two variables are the most relevant for predicting the target variable in the second scenario.
TFA do not appear in the optimal models with the complexities of 37 and 65. Moreover, NFA/TA and PWG_{75/C} do not exist in the optimal model with the complexity of 37. It shows that it is possible to estimate the CS of CCWG with an acceptable accuracy given only six input variables.

Assuming similar complexities, the optimal solutions of the second scenario have better performance than those of the first scenario. Additionally, there are 11 and 8 input variables in the best optimal solutions of the first and second scenarios, respectively. With the smaller number of input variables and lower complexity, the best model of the second scenario outperforms the best model of the first scenario. It means that considering the effective ratios as the predictors of CS of CCWG can lead to more accurate results.

The NRMSDs of the best optimal models for both scenarios are plotted in Fig. 9. Most developed models in two scenarios have the NRMSDs between 0.2 and 0.3, which indicates their fair performance in predicting the CS of CCWG for both training and testing phases. However, the

Table 5 Selected developed models and the used variables

| Models    | Input variables                          |
|-----------|------------------------------------------|
| CSA,11    | C, BFS, FA, SP, PP_{45}, WG, AC          |
| CSA,24    | C, W, BFS, FA, SP, PP_{45}, WG, AC       |
| CSA,29    | C, W, BFS, FA, NFA, SP, PP_{45}, WG, AC  |
| CSA,36    | C, W, BFS, FA, NCA, SP, PP_{45}, WG, AC  |
| CSA,80    | C, W, BFS, FA, NCA, NFA, SP, PP_{45}, WG, AC |
| CSA,108   | C, W, BFS, FA, NCA, NFA, SP, PP_{75}, PP_{45}, WG, AC |
| CSB,9     | W/B, SP/B, PWG_{45/C}, AC                |
| CSB,10    | W/B, NCA/C, NFA/TA, SP/B, PWG_{45/C}, AC |
| CSB,12    | W/B, NCA/C, SP/B, PWG_{75/C}, AC        |
| CSB,14    | W/B, BFS/C, NCA/C, SP/B, PWG_{75/C}, AC |
| CSB,26    | W/B, NCA/C, SP/B, PWG_{45/C}, AC        |
| CSB,28    | W/B, FA/C, NCA/C, SP/B, PWG_{75/C}, AC  |
| CSB,54    | W/B, BFS/C, NCA/C, SP/B, PWG_{45/C}, AC |
| CSB,60    | W/B, BFS/C, FA/C, NCA/C, NFA/TA, SP/B, PWG_{75/C}, PWG_{45/C}, AC |
| CSB,65    | W/B, BFS/C, NCA/C, NFA/TA, SP/B, PWG_{75/C}, PWG_{45/C}, AC |
NRMSDs reduce by increasing the complexities of the developed models, and the NRMSDs of the most complicated models of the second scenario reach 0.2 for the training phase, which shows their good performance in mapping data from the input space into the output space.

Tables 3 and 4 show the participated input variables in the best-developed models for the first and second scenarios, respectively. Six effective input variables of (C, BFS, FA, SP, PP 45, WG, AC) and four effective input variables of (W/B, SP/B, PWG 45/TFA, AC) participate in the simplest models of the first and second scenarios, respectively. Moreover, all 11 input variables exist in the most complicated model of the first scenario. In the case of the second scenario, two input variables of FA/C and PWG75/TFA do not participate in the most complicated model.

Figure 10 illustrates the participation percentage of the input variables for the best-developed models of the first and second scenarios. In all 65 developed models of the first scenario, C, W, BFS, FA, SP, PP 45, WG, and AC have the main participation, while the PP 75 has the least participation. It means that PP 75 has the lowest effect on the CS of CCWG. In the second scenario, W/B, NCA/C, SP/B, PWG 45/C, and AC are the most effective variables on the CS of CCWG. Moreover, RWG75/TFA does not participate in any of the 46 developed models of the second scenario, which means that the waste glass content with the size of greater than 75 μm does not have any contribution on the CS of CCWG compared to other input variables. It is compatible with the results of Shao et al. [62, 74] who reported that the WG with the size of 150 μm does not have a pozzolanic reaction. Besides, FA/C, with a participation percentage of 8.70%, has the second order from the viewpoint of the least important variable. Meanwhile, PWG 45/C appeared about 2.7 times more than the PWG75/C in the 46 best-developed models of the second scenario. The reason can be related to the pozzolanic activity and filler effects of WG with a smaller size [74, 75].

The number of input variables that participated in the best models of the first and second scenarios is given in Fig. 11. The numbers of input variables of the simplest models of the first and second scenarios is 7 and 4, respectively. For the most complicated models, the numbers of input variables is 11 and 8 for the first and second scenarios, respectively. It means fewer input variables participate in the second scenario than the first scenario. However, the accuracy of the best-developed models of the second scenario is higher despite their lower complexities. From the best-developed models, the models with different variables and better precisions were chosen for more investigations. The numbers of 6 and 9 models were selected from the first and second scenarios, respectively. For the most complicated models, the numbers of input variables is 11 and 8 for the first and second scenarios, respectively. It means fewer input variables participate in the second scenario than the first scenario.

Table 6 Statistical indicators of the selected models for the training and testing phases

| Models | Training phase | | | Testing phase | | |
|--------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|        | RMSD (MPa)     | MAD (MPa)      | R              | RMSD (MPa)     | MAD (MPa)      | R              | RMSD (MPa)     | MAD (MPa)      | R              |
| CSA,11 | 8.9975         | 6.6977         | 0.8754         | 9.1911         | 6.7855         | 0.8569         | 8.4868         |                 |                |                |                |
| CSA,24 | 8.0278         | 5.6468         | 0.9022         | 8.5851         | 5.9980         | 0.8768         | 7.4804         |                 |                |                |                |
| CSA,29 | 7.8291         | 5.4780         | 0.9072         | 8.7694         | 6.1281         | 0.8711         | 7.4708         |                 |                |                |                |
| CSA,36 | 7.5436         | 5.2645         | 0.9142         | 8.1957         | 5.5404         | 0.8880         | 6.9840         |                 |                |                |                |
| CSA,80 | 7.2425         | 4.9235         | 0.9212         | 7.9108         | 5.2946         | 0.8971         | 6.6475         |                 |                |                |                |
| CSA,108| 7.1819         | 4.8245         | 0.9225         | 7.9312         | 5.2552         | 0.8964         | 6.6001         |                 |                |                |                |
| CSB,9  | 8.7192         | 6.3306         | 0.8834         | 9.0225         | 6.7717         | 0.8635         | 8.2336         |                 |                |                |                |
| CSB,10 | 8.6301         | 6.3846         | 0.8860         | 9.0164         | 6.6985         | 0.8641         | 8.1963         |                 |                |                |                |
| CSB,12 | 8.1333         | 6.0054         | 0.8994         | 8.3979         | 6.1405         | 0.8829         | 7.5829         |                 |                |                |                |
| CSB,14 | 7.9786         | 5.8634         | 0.9034         | 8.2847         | 6.0739         | 0.8861         | 7.4429         |                 |                |                |                |
| CSB,26 | 7.5000         | 5.2699         | 0.9152         | 7.7205         | 5.5051         | 0.9020         | 6.8109         |                 |                |                |                |
| CSB,28 | 7.4645         | 5.1843         | 0.9160         | 7.7079         | 5.3022         | 0.9023         | 6.7206         |                 |                |                |                |
| CSB,54 | 7.1287         | 4.8095         | 0.9237         | 7.1919         | 4.8130         | 0.9153         | 6.2369         |                 |                |                |                |
| CSB,60 | 7.0639         | 4.6509         | 0.9252         | 7.2934         | 4.8104         | 0.9131         | 6.2062         |                 |                |                |                |
| CSB,65 | 7.0192         | 4.6370         | 0.9261         | 7.1931         | 4.7376         | 0.9154         | 6.1405         |                 |                |                |                |
Table 6 gives the statistical indicators of the selected models for the training and testing phases. The CS$_{A,11}$ and CS$_{B,65}$ models with the RMSDs of 8.9975 MPa and 7.0192 MPa for the training dataset are the best and the worst selected models. Moreover, the RMSD of the CS$_{A,108}$ model, with the complexity of about ten times of the CS$_{A,11}$ model, is about 13.7% lower than the RMSD of the CS$_{A,11}$ model for the testing dataset. Comparing the RMSDs of the CS$_{B,9}$ and CS$_{B,65}$ models for the testing dataset shows that the latter model is almost 20.3% more accurate than the former model. In terms of OBJ, the simplest and complicated models of the second scenario are about 3% and 7% more precise than those of the first scenario, respectively. However, the simplest and complicated models of the second scenario are, respectively, almost 18% and 40% simpler than those of the first scenario.

To show the performance of the selected models, the predicted CSs by the 15 selected models versus the experimental CSs for the testing dataset are plotted in Fig. 12. For more complicated models, the majority of data scatter are around the baseline with the slope of one passed through the origin of coordinates. Moreover, with increasing the complexity of models, the $R^2$ values increase, and the slopes of models get closer to one. However, there are some data points with a considerable difference between the predicted and experimental CSs for
all selected models which can be related to the experimental errors.

Figure 13 depicts the error box plot of the selected models for the testing dataset. The error on the horizontal axis was defined as the logarithm of the predicted to experimental CSs. The interquartile range (IQR) of a model increases by increasing its complexity. All developed models overestimate the CS of CCWG. For both of the $\text{CS}_{A,108}$ and $\text{CS}_{B,65}$ models, as the best models of the first and second scenarios, the errors of the 50% of data records
happen in the range of $-0.04$ and $0.06$ with the average of zero. The $CS_{A,11}$ and $CS_{B,9}$ models, as the least accurate models of the first and second scenarios, have the IQRs of about 40% and 50% more than the most complicated models. Surprisingly, the number of outliers is almost the same for the best models of the first and second scenarios. More investigation shows that the most outliers happen in the same data records for both models indicating the good performance of the developed models in identifying the outliers.

To investigate the effect of $PP_{45}$, $PP_{75}$ on the CS of CCWG, the $CS_{A,108}$ model was used. In this regard, all other input variables were kept constant in their average ranges, and the desired variables were changed within their allowable ranges. The CSs of CCWG were predicted for the designed data using the best model, and the obtained surfaces were plotted for the AC of 7, 28, 56, and 90 days. As illustrated in Fig. 14, the CS of CCWG rises by increasing the $PP_{45}$ and $PP_{75}$. However, the growth of CS is higher in the case of $PP_{45}$, which is related to the higher pozzolanic activity of the smaller size of WG [24, 75–77]. The distance of surfaces indicates the growth of CS of CCWG over time. The distance between the surfaces corresponding to the AC of 56 and 90 days is recognizable, which indicates the secondary reaction of WG at older ages of concrete. This analysis shows that the developed models can correctly estimate the trend of CS of CCWG by changing the $PP_{45}$ and $PP_{75}$ as well as considering the age of concrete in their related models.

5 Conclusions

In this study, the multi-objective multi-biogeography-based programming (MOMBBBP) was used as a machine learning method to estimate the compressive strength of concrete containing waste glass. In this regard, two scenarios were considered. In the first scenario, the amounts of concrete constituents, the age of concrete, the passing percentages of waste glass at the sieves of 75 $\mu$m ($PP_{75}$) and 45 $\mu$m ($PP_{45}$) are used to predict the compressive strength. In the second scenario, the critical ratios of concrete constituents as well as the age of concrete, the ratios of passing WG at the sieve of 45 $\mu$m and 75 $\mu$m to cement (PWG$_{45}$/C and PWG$_{75}$/C) and the amount of remaining WG on the sieve of 75 $\mu$m to the total fine aggregate (RWG$_{75}$/TFA) were served to estimate the compressive strength of concrete containing WG. The obtained conclusions are listed below:

- The numbers of 65 and 46 mathematical models with R values of more than 0.85 were achieved for the first and second scenarios, respectively. The complexities of the developed models of the first and second scenarios are within [10, 108] and [8, 65], respectively.
- Despite the lower complexities of the developed models of the second scenario, their precisions are better than the models developed by the first scenario. It shows the significant effect of employing the critical ratios of mix design as the input variables in predicting the compressive strength of concrete containing WG.
- In the simplest model of the first scenario, the seven variables consisting of cement, blast furnace slag, fly ash, superplasticizer, $PP_{45}$, WG, and age of concrete are the factors for the prediction of the compressive strength. In contrast, only four variables including water to binder ratio, superplasticizer to binder ratio, PWG$_{45}$/C, and the age of concrete are the optimized predictors for the estimation of the compressive strength in the case of the less complicated model of the second scenario.
- In the complicated model of the first scenario, all eleven variables are used to predict the compressive strength of concrete containing WG. While only eight predictors are served to formulate the compressive strength, and the fly ash to binder ratio and RWG$_{75}$/TFA do not participate in this mathematical equation.
- In the first scenario, $PP_{75}$ cooperates only in one-third of the obtained 65 mathematical formulas in the optimization process, which shows its less influence on the compressive strength compared to other predictors. RWG$_{75}$/TFA does not appear in any of the 46 obtained mathematical formulas of the second scenario, which indicates that WG with the size of 150 $\mu$m to 4.75 mm does not have a significant effect on the compressive strength compared to other predictors.

The authors suggest using other feature selection and machine learning methods, as a hybrid model, for the determination of the effective factors influence on the compressive strength of concrete containing WG as well as increasing the precision of the developed models in this study. In the case of MOMBBP, the proposed technique can be served for any engineering regression problem, especially in the field of material science and structural engineering, in which different mathematical equations with various attributes are required.

Appendix: Mathematical equations obtained by different MOMBBP algorithm

$$CS_{1,11} = -30.1013 - 3.1593\sqrt{SP} + 8.2570\sqrt{AC} + 0.1211\cdot C + 0.0019PP_{45}\cdot WG + 0.1298(BFS + FA)$$
\[ \text{CS}_{1,24} = -60.4259 + 0.0014 \text{PP}_{45, \text{WG}} + 35.5896 \sqrt{\text{FA} + \text{AC}} \]

\[-0.0590 \left[ C^{\frac{1}{5}} \sqrt{171.0613 + \sqrt{\text{W.SP} - \text{BFS} - W}} \right] + 0.2235C + 0.1783\text{BFS} \]

\[ \text{CS}_{1,29} = -59.5140 + 0.0014 \text{PP}_{45, \text{WG}} + 34.9474 \sqrt{\text{FA} + \text{AC}} \]

\[-0.0629 \left[ C^{\frac{1}{5}} \sqrt{13.7427 \sqrt{\text{SP} - \sqrt{\text{NFA} - W + 179.6707}}} \right] + 0.2294C + 21.2264 \left( \frac{\text{BFS}}{C} \right) \]

\[ \text{CS}_{1,36} = -35.4258 + 0.0010 \text{PP}_{45, \text{WG}} + 5.9355 \sqrt{\text{NCA}(\text{FA} + \text{AC} - 1)} \]

\[-0.0618 \left[ C^{\frac{1}{5}} \sqrt{14.0253 \sqrt{\text{SP}^2 - \left( \frac{\text{NCA} + \text{AC}}{C} \right)^2} - W + 179.3528} \right] + 0.2132C + 0.1194\text{BFS} \]

\[ \text{CS}_{1,80} = -39.2617 + 1.39e^{-6} \left[ \text{NFA.PP}_{45, \text{WG}} - \frac{\text{NCA}^2 \sqrt{\text{AC}^6}}{C} + 2\text{C.NCA} - \text{C} \right] \]

\[ + 6.5588 \sqrt{\text{NCA}(\text{FA} + \text{AC} - 1)} + \text{FA} + \text{FA.WG} \]

\[-0.0686 \left[ C^{\frac{1}{5}} \sqrt{\text{NCA}^2 \sqrt{\text{SP}^2} - \left( \frac{\text{NCA} - \text{PP}_{45}}{C} \right)^2} - W + 178.1806 \right] + 0.2272C \]

\[ + 14.6525 \left[ C^{\frac{1}{4}} \sqrt{\frac{\text{BFS}}{W - \sqrt{\text{FA}^3 - \sqrt{\text{C} \sqrt{\text{AC}}}}} \right] \]

\[ \text{CS}_{1,108} = -37.8596 + 1.41e^{-6} \left[ \text{NFA.PP}_{45, \text{WG}} - \frac{\text{NCA}^2 \sqrt{\text{AC}^6}}{C} + 2\text{C.NCA} - \text{WG} \right] \]

\[ + 6.5421 \sqrt{\text{NCA}(\text{FA} + \text{AC} - 1)} + \text{FA} + \text{FA.WG} \]

\[-0.0690 \left[ C^{\frac{1}{4}} \sqrt{\text{NCA}^2 \sqrt{\text{SP}^2} - \left( \frac{\text{NCA} - \text{PP}_{45}}{C} \right)^2 + \text{AC}^2} - W + 178.1806 \right] \]

\[ + 0.2259 \left[ C^{\frac{1}{5}} \sqrt{\left( \text{BFS} - \text{WG} + 8.3376 \right)^2 - \text{C} \sqrt{\text{AC} + \text{W} - \text{WG} + \sqrt{\text{PP}_{45}}}} \right] \]

\[ + 14.5666 \left[ C^{\frac{1}{4}} \sqrt{\frac{\text{BFS}}{W - \sqrt{\text{FA}^3 - \sqrt{\text{C} \sqrt{\text{AC}}}}} \right] \]

\[ \text{CS}_{2,9} = -28.7466 + 4.5274 \sqrt{\text{AC}} \]

\[ + 28.1956 \sqrt{\left( \frac{\text{PW}_{45}}{B} \right)} - 3.7124 \left( \frac{\text{SP}}{\text{W}} \right) - 6.6784 \left( \frac{\text{SP}}{\text{W}} \right) \]

\[ \text{CS}_{2,10} = -13.4268 + 4.3418 \sqrt{\text{AC}} - 6.3202 \left( \frac{\text{NFA}}{\text{TA}} \right) + 35.2324 \left( \frac{\text{PW}_{45}}{B} \right) \]

\[ - 1.2992 \left( \frac{\text{NCA}}{C} \right) - 3.2489 \left( \frac{\text{SP}}{\text{W}} \right) - 6.6609 \left( \frac{\text{SP}}{\text{W}} \right) \]
\[ \begin{align*}
CS_{2,12} &= -13.0909 - 17.1859 \left( \frac{\text{NCA}}{C} \right) + 0.1530 \left( \frac{\text{SP}}{W} \right) - 6.6276 \right)^2 \\
+ 8.3411 \sqrt{\text{AC}} + 27.5978 \left( \frac{\text{PWG}_{75}}{B} \right) + 50.8306 \ln \left( \frac{\text{NCA}}{C} \right) \\
CS_{2,14} &= -4.6955 - 20.6538 \left( \frac{\text{NCA}}{C} \right) + 0.1735 \left( \frac{\text{SP}}{W} \right) - 5.8503 \right)^2 \\
+ 8.1548 \sqrt{\text{AC}} + 29.1911 \left( \frac{\text{PWG}_{75}}{B} \right) + 55.8578 \ln \left( \frac{\text{BFS}}{C} + \left( \frac{\text{NCA}}{C} \right) \right) \\
CS_{2,26} &= -56.9105 + 20.4994 \left( \frac{\text{NCA}}{C} \right) + 2.4035 \left( \frac{W}{B} + \frac{\text{NCA}}{C} \right) \\
+ 28.6491 \left[ \sqrt{\left( \frac{W}{B} \right)^5} + AC + 1.1976 \right] + \frac{\text{PWG}_{45}}{B} \right)^3 \\
- 28.7136 \ln \left( \frac{\text{NCA}}{C} \right)^3 - 0.1860 \left( \frac{\text{SP}}{W} \right) - 2.0061 \right)^3 \\
CS_{2,28} &= -873.6001 + 25.8653 \left( \frac{\text{PWG}_{75}}{B} \right)^3 - 13.4826 \left( \frac{\text{FA}}{C} + \frac{\text{NCA}}{C} \right) + \left( \frac{\text{SP}}{W} \right) \\
+ 964.4530 \left( \frac{\text{FA}}{C} + \ln \left( \frac{\text{NCA}}{C} \right) + 7.4239 \right) + \sqrt{\left( \frac{\text{AC}}{C} - \frac{\text{NCA}}{C} \right)} + 5.2742 \left( \frac{\text{AC}}{C} \right) \\
CS_{2,54} &= -55.9469 + 132.3788 \left( \frac{\text{PWG}_{45}}{B} \right) \left( \frac{\text{NCA}}{C} \right) + 239.9285 \ln \left( \frac{\text{BFS}}{C} + \frac{\text{NCA}}{C} \right) \\
+ 7.9998 \sqrt{\left( \frac{\text{SP}}{W} \right) + AC + \left( \frac{W}{B} \right)^3 + \ln \left( \frac{W}{B} \right)^4 + 21.7955 \left( \frac{\text{SP}}{W} \right) \\
- 29.1078 \left[ \frac{2.6927 \ln \left( 0.2217 \left( \frac{\text{NCA}}{C} - \frac{\text{BFS}}{C} \right) + \left( \frac{\text{SP}}{W} \right) \right) \left( \frac{\text{SP}}{W} \right) - \ln \left( \frac{\text{NCA}}{C} \right) \right]^2 \\
+ \left( \frac{\text{SP}}{W} \right)^2 \left( \frac{\text{SP}}{W} \right) + \frac{\text{NCA}}{C} \right)^2 \\
CS_{2,60} &= -69.4317 + 15.7714 \left[ \frac{\text{BFS}}{C} + \frac{\text{SP}}{W} + \left( \frac{\text{PWG}_{45}}{B} \right) \ln (AC) \right] + 128.2571 \left[ \frac{\ln \left( \frac{\text{NCA}}{C} \right)}{\ln \left( \frac{\text{NCA}}{C} \right) - \left( \frac{\text{PWG}_{45}}{B} \right)} \right] \\
+ 7.4912 \left[ \sqrt{\left( \frac{\text{AC}}{C} \right) + \ln \left( \frac{W}{B} \right)^4 + 18.9260 \left[ \frac{4.2842 \left( \frac{\text{NCA}}{C} \right) - \frac{\text{FA}}{C} - \left( \frac{\text{SP}}{W} \right) \right]}{\left( \frac{W}{B} \right)} \right] \\
- 3.78e^3 \left[ \frac{\ln \left( 0.1966 \left( \frac{\text{NCA}}{C} - \frac{\text{BFS}}{C} \right) + \left( \frac{\text{SP}}{W} \right) \right) \left( \frac{\text{SP}}{W} \right) - \ln \left( \frac{\text{NCA}}{C} \right) \right]^2}{64.3124 \left( \frac{\text{W}}{B} \right)^2 \left( \frac{\text{NCA}}{C} \right)^2} \right] \\
\end{align*} \]
CS_{2,05} = -59.2861 + 10.8611 \left( \frac{\text{BFS}}{C} \right) + \left( \frac{\text{SP}}{F} \right) + \frac{\text{PWGls}}{B} \ln \left( \frac{\text{NCA}}{C} \right) + \frac{\text{AC}}{C}^2 \right) \right] \\
+ 305.9236 \left[ \ln \left( \frac{\text{NCA}}{C} \right)^3 \right] + 7.2918 \left[ \sqrt{\frac{2}{W}} + \text{AC} + 4.3250 \ln \left( \frac{W}{B} \right) \right] \\
+ 18.6680 \left[ \sqrt{\frac{2}{W}} \left( \frac{\text{NCA}}{TA} \right) - \left( \frac{\text{SP}}{F} \right) \right] \\
- 37.3107 \left[ \left( \ln \left( \frac{\text{NCA}}{C} \right) - \left( \frac{W}{F} \right) \right)^2 \ln \left[ 0.1739 \left( \frac{\text{BFS}}{C} \right) - \left( \frac{\text{NCA}}{C} \right) + \left( \frac{\text{SP}}{F} \right) \right] \right] \\

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. He Z, Zhan P, Du S et al (2019) Creep behavior of concrete containing glass powder. Compos Part B Eng. https://doi.org/10.1016/j.compositesb.2018.11.133
2. Hajimohammadi A, Ngo T, Kashani A (2018) Sustainable one-part geopolymer foams with glass fines versus sand as aggregates. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2018.03.120
3. Almesfer N, Ingham J (2014) Effect of waste glass on the properties of concrete. J Mater Civ Eng. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001077
4. Hajimohammadi A, Ngo T, Kashani A (2018) Glass waste versus sand as aggregates: the characteristics of the evolving geopolymer binders. J Clean Prod. https://doi.org/10.1016/j.jclepro.2018.05.086
5. Guo H, Shi C, Guan X et al (2018) Durability of recycled aggregate concrete—a review. Cem Conc Res. https://doi.org/10.1007/j.cemconcomp.2018.03.008
6. Golafshani EM, Behnood A (2017) Application of soft computing methods for predicting the elastic modulus of recycled aggregate concrete. J Clean Prod. https://doi.org/10.1016/j.jclepro.2017.11.186
7. Golafshani EM, Behnood A (2018) Automatic regression methods for formulation of elastic modulus of recycled aggregate concrete. Appl Soft Comput J 64:377–400. https://doi.org/10.1016/j.asoc.2017.12.030
8. Duan P, Shui Z, Chen W, Shen C (2013) Effects of metakaolin, silica fume and slag on pore structure, interfacial transition zone and compressive strength of concrete. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2013.02.075
9. Behnood A, Golafshani EM (2018) Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves. J Clean Prod. https://doi.org/10.1016/j.jclepro.2018.08.065
10. Kandiri A, Mohammadi Golafshani E, Behnood A (2020) Estimation of the compressive strength of concretes containing ground granulated blast furnace slag using hybridized multi-objective ANN and salp swarm algorithm. Constr Build Mater 248:118676. https://doi.org/10.1016/j.conbuildmat.2020.118676
11. Jani Y, Hogland W (2014) Waste glass in the production of cement and concrete—a review. J Environ Chem Eng. https://doi.org/10.1016/j.jece.2014.03.016
12. Cassar J, Camilleri J (2012) Utilisation of imploded glass in structural concrete. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2011.10.005
13. Mirzahosseini M, Riding KA (2015) Influence of different particle sizes on reactivity of finely ground glass as supplementary cementitious material (SCM). Cem Conc Comp. https://doi.org/10.1016/j.cemconcomp.2014.10.004
14. Soliman NA, Tagnit-Hamou A (2017) Using glass sand as an alternative for quartz sand in UHPC. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2017.03.187
15. Shayan A, Xu A (2004) Value-added utilisation of waste glass in concrete. Cem Conc Res. https://doi.org/10.1016/S0008-8846(03)00251-5
16. Afshininia K, Rangaraju PR (2016) Impact of combined use of ground glass powder and crushed glass aggregate on selected
properties of Portland cement concrete. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2016.04.072
17. Chandra Paul S, Šavija B, Babafemi AJ (2018) A comprehensive review on mechanical and durability properties of cement-based materials containing waste recycled glass. J Clean Prod 198:891–906
18. De Castro S, De Brito J (2013) Evaluation of the durability of concrete made with crushed glass aggregates. J Clean Prod. https://doi.org/10.1016/j.jclepro.2012.09.021
19. Kou SC, Poon CS (2009) Properties of self-compacting concrete prepared with recycled glass aggregate. Cem Concr Compos 31:107–113
20. Ali EE, Al-Tersawy SH (2012) Recycled glass as a partial replacement for fine aggregate in self compacting concrete. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2012.04.117
21. Du H, Tan KH (2017) Properties of high volume glass powder concrete. Cem Concr Compos. https://doi.org/10.1016/j.cemconcomp.2016.10.010
22. Ismail ZZ, AL-Hashmi EA (2009) Recycling of waste glass as a partial replacement for fine aggregate in concrete. Waste Manag. https://doi.org/10.1016/j.wasman.2008.08.012
23. Kamali M, Ghahremaninezhad A (2015) Effect of glass powders on the mechanical and durability properties of cementitious materials. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2015.06.010
24. Nassar RUD, Soroushian P (2012) Strength and durability of recycled aggregate concrete containing milled glass as partial replacement for cement. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2011.10.061
25. Batayneh M, Marie I, Asi I (2007) Use of selected waste materials in concrete mixes. Waste Manag. https://doi.org/10.1016/j.wasman.2006.07.026
26. Patel HG, Dalal SP (2017) An experimental investigation on physical and mechanical properties of concrete with the replacement of fine aggregate by poly vinyl chloride and glass waste. Procedia Eng 173:1666–1671
27. Zhao H, Poon CS, Ling TC (2013) Utilizing recycled cathode ray tube funnel glass sand as river sand replacement in the high-density concrete. J Clean Prod. https://doi.org/10.1016/j.jclepro.2013.01.052
28. Naderpour H, Mirrashid M (2020) Estimating the compressive strength of eco-friendly concrete incorporating recycled coarse aggregate using neuro-fuzzy approach. J Clean Prod. https://doi.org/10.1016/j.jclepro.2020.121886
29. Adil M, Ullah R, Noor S, Gohar N (2020) Effect of number of hardened properties of self-compacting concrete using support vector regression approach. Neural Comput Appl. https://doi.org/10.1007/s00521-020-05305-8
30. Gholampour A, Mansouri I, Kisi O, Ozbakkaloglu T (2020) Evaluation of mechanical properties of concretes containing coarse recycled concrete aggregates using multivariate adaptive regression splines (MARS), M5 model tree (M5Tree), and least squares support vector regression (LSSVR) models. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3630-y
31. Gupta T, Sachdeva SN (2020) Recurrent neural network-based prediction of compressive and flexural strength of steel slag mixed concrete. Neural Comput Appl. https://doi.org/10.1007/s00521-020-05470-w
32. Saha P, Debnath P, Thomas P (2020) Prediction of fresh and hardened properties of self-compacting concrete using support vector regression approach. Neural Comput Appl. https://doi.org/10.1007/s00521-019-04267-w
33. Asteris PG, Kolovos KG (2019) Self-compacting concrete strength prediction using surrogate models. Neural Comput Appl. https://doi.org/10.1007/s00521-017-3007-7
34. Subaşi S, Beycioğlu A, Sancak E, Şahin I (2013) Rule-based Mamdani type fuzzy logic model for the prediction of compressive strength of silica fume included concrete using non-destructive test results. Neural Comput Appl. https://doi.org/10.1007/s00521-012-0879-4
35. Golafshani EM, Behnoood A, Arashpour M (2020) Predicting the compressive strength of normal and high-performance concretes using ANN and ANFIS hybridized with grey wolf optimizer. Constr Build Mater 232:117266. https://doi.org/10.1016/j.conbuildmat.2019.117266
36. Sadowski Ł, Piechówka-Mielnik M, Wdysziszowski T et al (2019) Hybrid ultrasonic-neural prediction of the compressive strength of environmentally friendly concrete screeds with high volume of waste quartz mineral dust. J Clean Prod. https://doi.org/10.1016/j.jclepro.2018.12.059
37. Nguyen KT, Nguyen QD, Le TA et al (2020) Analyzing the compressive strength of green fly ash based geopolymer concrete using experiment and machine learning approaches. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2020.118581
38. Hammoudi A, Moussaaceb K, Belebchouche C, Dahmoune F (2019) Comparison of artificial neural network (ANN) and response surface methodology (RSM) prediction in compressive strength of recycled concrete aggregates. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2019.03.119
39. Golafshani EM, Pazouki G (2018) Predicting the compressive strength of self-compacting concrete containing fly ash using a hybrid artificial intelligence method. Comput Concrr. https://doi.org/10.12989/cac.2018.22.4.419
40. Behnoood A, Golafshani EM (2018) Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves. J Clean Prod 202:54–64
41. Behnoood A, Golafshani EM (2020) Machine learning study of the mechanical properties of concretes containing waste foundry sand. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2020.118152
42. Golafshani EM, Behnoood A (2019) Estimating the optimal mix design of silica fume concrete using biogeography-based programming. Cem Concr Comps 96:95–105
43. Ray S, Haque M, Ahmed T, Nahin TT (2021) Comparison of artificial neural network (ANN) and response surface methodology (RSM) in predicting the compressive and splitting tensile strength of concrete prepared with glass waste and tin (Sn) can fiber. J King Saud Univ Eng Sci. https://doi.org/10.1016/j.jsues.2021.03.006
44. Ghorbani B, Arulrajah A, Narsilio G et al (2021) Dynamic characterization of recycled glass-recycled concrete blends using experimental analysis and artificial neural network modeling. Soil Dyn Earthq Eng. https://doi.org/10.1016/j.soildyn.2020.106544
45. Golafshani EM, Talahhari S (2018) Predicting the climbing rate of slip formwork systems using linear biogeography-based programming. Adv Soft Comput J 70:263–278. https://doi.org/10.1016/j.asoc.2018.05.036
46. Golafshani EM, Ashour A (2016) A feasibility study of BBP for predicting shear capacity of FRP reinforced concrete beams without stirrups. Adv Eng Softw. https://doi.org/10.1016/j.advengsoft.2016.02.007
47. Golafshani EM, Ashour A (2016) Prediction of self-compacting concrete elastic modulus using two symbolic regression techniques. Autom Constr 64:7–19. https://doi.org/10.1016/j.autcon.2015.12.026
48. Behnoood A, Mohammad Golafshani E (2021) Predicting the dynamic modulus of asphalt mixture using machine learning techniques: an application of multi biogeography-based programming. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2020.120983
49. Simon D (2008) Biogeography-based optimization. IEEE Trans Evol Comput 12:702–713
50. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput. https://doi.org/10.1109/4235.996017
51. Golafshani EM (2015) Introduction of biogeography-based programming as a new algorithm for solving problems. Appl Math Comput 270:1–12. https://doi.org/10.1016/j.amc.2015.08.026
52. Metwally IM (2007) Investigations on the performance of concrete made with blended finely milled waste glass. Adv Struct Eng. https://doi.org/10.12691/136943307780150823
53. Wang HY (2009) A study of the effects of LCD glass sand on the properties of concrete. Waste Manag. https://doi.org/10.1016/j.wasman.2008.03.005
54. Kim IS, Choi SY, Yang EJ (2018) Evaluation of durability of concrete substituted heavyweight waste glass as fine aggregate. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2018.06.221
55. Her-Yung W (2009) A study of the engineering properties of waste LCD glass applied to controlled low strength materials concrete. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2008.12.012
56. Aliabido AA, Abd Elmoaty AEM, Aboshama AY (2016) Utilization of waste glass powder in the production of cement and concrete. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2016.08.016
57. Islam GMS, Rahman MH, Kazi N (2017) Waste glass powder as partial replacement of cement for sustainable concrete practice. Int J Sustain Built Environ. https://doi.org/10.1016/j.ijisbe.2016.10.005
58. Wang CC, Wang HY (2017) Assessment of the compressive strength of recycled waste LCD glass concrete using the ultrasonic pulse velocity. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2017.01.017
59. Bostanci SC (2020) Use of waste marble dust and recycled glass for sustainable concrete production. J Clean Prod. https://doi.org/10.1016/j.jclepro.2019.119785
60. Ramakrishnan K, Pugazhmani G, Sripragadesh R et al (2017) Experimental study on the mechanical and durability properties of concrete with waste glass powder and ground granulated blast furnace slag as supplementary cementitious materials. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2017.08.183
61. Adaway M, Wang Y (2015) Recycled glass as a partial replacement for fine aggregate in structural concrete—effects on compressive strength. Electron J Struct Eng 14(1):116–122
62. Du H, Tan KH (2014) Waste glass powder as cement replacement in concrete. J Adv Conc Technol. https://doi.org/10.3151/jact.12.468
63. Elwell D, Fu G (1995) Compression testing of concrete: cylinders versus cubes. New York State Department of Transportation, New York
64. Li MF, Tang XP, Wu W, Bin LH (2013) General models for estimating daily global solar radiation for different solar radiation zones in mainland China. Energy Convers Manag. https://doi.org/10.1016/j.enconman.2013.03.004
65. Shao Y, Lefort T, Moras S, Rodriguez D (2000) Studies on concrete containing ground waste glass. Cem Concr Res. https://doi.org/10.1016/j.cemconres.2005.12.010
66. Limbachiya MC (2009) Bulk engineering and durability properties of washed glass sand concrete. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2008.05.022
67. De Oliveira LAP, Castro-Gomes JP, Santos P (2008) Mechanical and durability properties of concrete with ground waste glass sand. ISTANBUL, Turkey, 11–14 May 2008
68. Taha B, Nougu G (2008) Properties of concrete contains mixed colour waste recycled glass as sand and cement replacement. Constr Build Mater. https://doi.org/10.1016/j.conbuildmat.2007.01.019
69. Du H, Tan KH (2014) Waste glass powder as cement replacement in concrete. J Adv Conc Technol, https://doi.org/10.3151/jact.12.468
70. Elwell D, Fu G (1995) Compression testing of concrete: cylinders versus cubes. New York State Department of Transportation, New York
71. Li MF, Tang XP, Wu W, Bin LH (2013) General models for estimating daily global solar radiation for different solar radiation zones in mainland China. Energy Convers Manag. https://doi.org/10.1016/j.enconman.2013.03.004
72. Shao Y, Lefort T, Moras S, Rodriguez D (2000) Studies on concrete containing ground waste glass. Cem Concr Res. https://doi.org/10.1016/j.cemconres.2005.12.010
73. Li MF, Tang XP, Wu W, Bin LH (2013) General models for estimating daily global solar radiation for different solar radiation zones in mainland China. Energy Convers Manag. https://doi.org/10.1016/j.enconman.2013.03.004