

Article

Optimization and Predictive Modeling of Reinforced Concrete Circular Columns

Gebrail Bekdaş^{1,*}, Celal Cakiroglu², Sanghun Kim³ and Zong Woo Geem^{4,*}¹ Department of Civil Engineering, Istanbul University-Cerrahpasa, 34320 Istanbul, Turkey² Department of Civil Engineering, Turkish-German University, 34820 Istanbul, Turkey³ Department of Civil and Environmental Engineering, Temple University, Philadelphia, PA 19122, USA⁴ Department of Smart City & Energy, Gachon University, Seongnam 13120, Korea

* Correspondence: bekdas@iuc.edu.tr (G.B.); geem@gachon.ac.kr (Z.W.G.)

Abstract: Metaheuristic optimization techniques are widely applied in the optimal design of structural members. This paper presents the application of the harmony search algorithm to the optimal dimensioning of reinforced concrete circular columns. For the objective of optimization, the total cost of steel and concrete associated with the construction process were selected. The selected variables of optimization include the diameter of the column, the total cross-sectional area of steel, the unit costs of steel and concrete used in the construction, the total length of the column, and applied axial force and the bending moment acting on the column. By using the minimum allowable dimensions as the constraints of optimization, 3125 different data samples were generated where each data sample is an optimal design configuration. Based on the generated dataset, the SHapley Additive exPlanations (SHAP) algorithm was applied in combination with ensemble learning predictive models to determine the impact of each design variable on the model predictions. The relationships between the design variables and the objective function were visualized using the design of experiments methodology. Applying state-of-the-art statistical accuracy measures such as the coefficient of determination, the predictive models were demonstrated to be highly accurate. The current study demonstrates a novel technique for generating large datasets for the development of data-driven machine learning models. This new methodology can enhance the availability of large datasets, thereby facilitating the application of high-performance machine learning predictive models for optimal structural design.

Keywords: predictive modeling; optimization; structural design

Citation: Bekdaş, G.; Cakiroglu, C.; Kim, S.; Geem, Z.W. Optimization and Predictive Modeling of Reinforced Concrete Circular Columns. *Materials* **2022**, *15*, 6624. <https://doi.org/10.3390/ma15196624>

Academic Editor: Ayman El-Zohairy

Received: 7 August 2022

Accepted: 19 September 2022

Published: 23 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Structural optimization aims at designing structures with the best possible dimensions that minimize cost without any impact on structural performance. In recent years metaheuristic optimization techniques have been increasingly applied to the optimization of different structures such as cylindrical reinforced concrete walls [1,2], retaining walls [3–5], plate girders [6], laminated composite plates [7–9], concrete-filled steel tubes [10,11], truss systems [12], timber structures [13], and liquid mass dampers [14–16]. These algorithms can be divided into evolutionary, physics-based, swarm-based, and population-based algorithms [17–19]. A detailed classification of the state-of-the-art metaheuristic optimization techniques can be found in Figure 1 [20–35]. Among the metaheuristic optimization algorithms applied to structural optimization, the harmony search algorithm stands out as one of the most widely used techniques. Besides structural optimization, the harmony search technique has been applied to various areas of science and engineering such as transportation engineering [36–39], environmental engineering [40,41], healthcare systems [42], bioinformatics [43–45], and cloud computing [46].

Reinforced concrete (RC) circular columns have been used in broad applications in structural engineering. The total amount of longitudinal reinforcement determines their load-carrying capacity of them. Therefore, the accurate determination of the right amount

of reinforcement in these structural members under axial forces and bending moments bears the utmost importance. Figure 2 shows a general description of an RC circular column where the lateral and longitudinal reinforcements can be seen. The outer diameter D and the total length L describe the geometry of the column in addition to the longitudinal reinforcement area A_s .

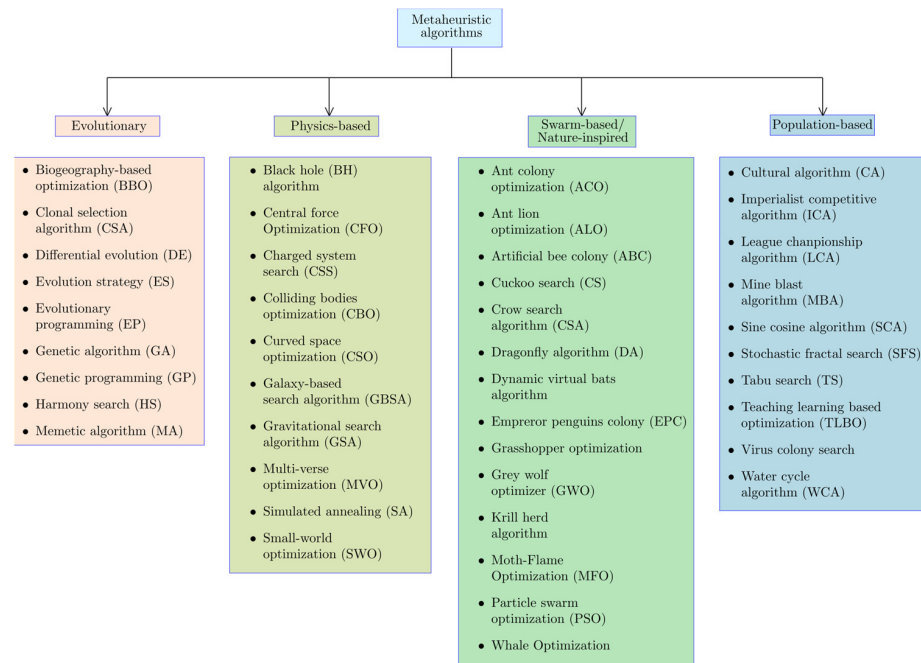


Figure 1. Classification of metaheuristic optimization algorithms.

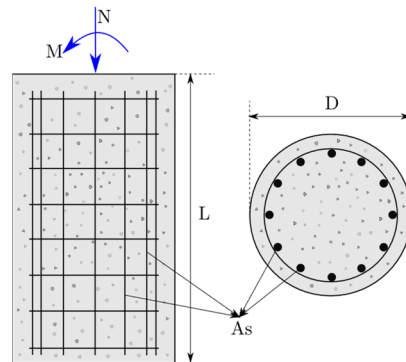


Figure 2. Cross-section and longitudinal section of a circular RC column.

This paper presents a novel data-driven technique for the prediction of the area of the longitudinal reinforcement (A_s) in RC circular columns. To this end, four different ensemble learning algorithms have been utilized to obtain predictive models. The performances of these algorithms in terms of predicting A_s accurately have been compared using the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) as the metrics of accuracy. The datasets needed to train these predictive models have been generated using the harmony search algorithm such that each data sample corresponds to an optimal design configuration that satisfies certain load-carrying requirements defined by the design codes. A combination of axial force and the bending moment was applied in each data sample. A dataset of 3125 samples was generated where each sample consists of six input variables and the corresponding output variable. The input variables in this dataset consist of the outer diameter of the column (D), the unit

cost of concrete used in the construction (C_c), the unit cost of steel (C_s), the total length of the column (L), the bending moment acting on the column (M), and the axial force acting on the column (N). The corresponding output variable is the optimal longitudinal reinforcement area (A_s). These input variables were selected in order to have a description of the geometry, material properties, and external loading for each data sample. In this regard, the unit costs of concrete and steel quantify the material properties whereas the column length and outer diameter describe the geometry corresponding to a data sample. To clarify the impact of each input variable on the output of the predictive models and to show the dependencies between different variables, the SHAP algorithm has been utilized. Furthermore, a four-level factorial analysis has been carried out to visualize the variation of the output variable for the different levels of each input variable [47]. Based on the dependencies of the input variables, a predictive equation has been proposed for the reinforcement area. The equation has been developed using the harmony search algorithm to minimize the difference between the predicted and true optimal reinforcement areas. An R^2 score of 0.998 could be achieved by the developed equation.

The current paper investigates the optimal design of circular RC columns under combined loading, which is an area of structural engineering that has not been investigated using data-driven machine learning techniques to the best of the authors' knowledge. The work related to failure mode classification and capacity prediction of RC columns using an ensemble machine learning algorithm AdaBoost by Feng et al. [48] can be counted among the recent machine learning-related research works in the field of conventional RC columns. The ensemble learning algorithm was developed based on a data set consisting of 254 data samples collected from cyclic loading tests. Also, Dogan et al. [49] investigated the damage levels of RC columns under cyclic lateral loading conditions using machine learning methods for classification such as support vector machines, K-nearest neighbors and discriminant analysis. The machine learning models were trained on a set of 390 damage images. However, the research activity in the area of RC columns using machine learning methods has been limited compared to other areas such as concrete columns with fiber-reinforced polymer wrappings [50,51], or crack propagation and corrosion in RC structures [52–54]. Nasrollahzadeh and Nouhi [50] proposed fuzzy inference system models to predict the strength and strain capacity of square concrete columns wrapped with fiber-reinforced polymer. Experimental data sets consisting of 261 and 112 test samples were used for the prediction of compressive strength and ultimate strain respectively. Naderpour et al. [51] utilized artificial neural network and gene expression programming techniques to predict the compressive strength of columns confined with fiber reinforced polymers based on a data set consisting of 95 data samples. However, despite the benefits of using composite materials as reinforcement or confinement for columns reported in the literature, the overwhelming majority of new constructions and existing infrastructure rely on conventional reinforced concrete.

One of the reasons for the lack of machine learning related research in the field of RC columns is the difficulty of obtaining large datasets. Machine learning-based predictive models need to be trained using large and comprehensive data sets in order to be relevant in general-purpose structural design. On the other hand, experimental research in the field of RC members is generally costly and experimental programs usually deliver a limited number of data points. Therefore, alternative techniques need to be devised for the training of machine learning models in order to use these powerful techniques in the field of RC design.

An important current issue in previous literature pertaining to machine learning applications in structural engineering is the size of the data sets used during the model training process. Evidently, most studies in this field depend on data sets with less than a thousand data samples. However, the reliability of a machine learning model heavily depends on the size and quality of the data used in its training. The aim of the current study is to present a methodology for the generation of large datasets related to the optimal design of RC columns. The novelty of this current work shows the applicability of newly

developed techniques of artificial intelligence to the design process of RC columns. The availability of quality data is a major requirement in this process. However, large datasets are needed for the development of accurate predictive machine learning models, which is one of the limiting factors in the process of machine learning model development. The current paper proposes a novel technique for the generation of large datasets. This technique is highly valuable since it can remove one of the major bottleneck points in the application of machine learning techniques to structural design.

2. Dataset Generation and Analysis

A large dataset consisting of 3125 unique configurations has been generated with the help of the harmony search optimization algorithm. Each sample in this dataset corresponds to a design configuration that minimizes the total cost associated with the construction process while keeping the load-carrying capacity above a certain level. To this end, the axial load and moment capacities were kept above the applied load which can be described in Equation (1):

$$\phi P_n \geq P_u, \quad \phi M_n \geq M_u \tag{1}$$

In Equation (1), P_u and M_u are the applied loads, P_n and M_n are the nominal strengths of the column cross-section, and ϕ is the strength reduction factor. This process starts with the generation of a randomly populated harmony memory matrix (HM) as shown in Equation (2) where f denotes the cost function, HMS is the size of the solution candidate population, V_c and W_s are the total volume of concrete and the total weight of steel respectively and x^i is a solution candidate vector containing the variables $D^i, C_c^i, C_s^i, L^i, M^i, N^i, A_s^i$. The cost function f determines the performance of each population member and the solution candidates can be ranked accordingly.

$$HM = \begin{bmatrix} D^1 & C_c^1 & C_s^1 & L^1 & M^1 & N^1 & A_s^1 & f(x^1) \\ D^2 & C_c^2 & C_s^2 & L^2 & M^2 & N^2 & A_s^2 & f(x^2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ D^{HMS} & C_c^{HMS} & C_s^{HMS} & L^{HMS} & M^{HMS} & N^{HMS} & A_s^{HMS} & f(x^{HMS}) \end{bmatrix} \tag{2}$$

$$f(x^i) = (C_c)^i V_c + (C_s)^i W_s$$

In Equation (2), $C_c^i, C_s^i, L^i, M^i, N^i$, and A_s^i stand for the concrete cost per unit volume, steel cost per unit weight, length of the column, bending moment, axial force and the total area of the longitudinal reinforcement in the i -th solution candidate respectively. The harmony search technique obtains an optimum solution that minimizes the total cost by an evolutionary process in which the solution candidates improve incrementally and eventually converge to an optimum solution. The evolutionary process of incremental improvement of the harmony memory matrix can be described in Equations (3)–(6).

$$k = \text{int}(\text{rand} \cdot \text{HMS}), \text{ rand} \in (0, 1) \tag{3}$$

$$x_{i,\text{new}} = x_{i, \text{min}} + \text{rand} \cdot (x_{i, \text{max}} - x_{i, \text{min}}), \text{ if HMCR} + \text{rand} \tag{4}$$

$$x_{i,\text{new}} = x_{i,k} + \text{rand} \cdot \text{PAR} \cdot (x_{i, \text{max}} - x_{i, \text{min}}), \text{ if HMCR} + \text{rand} \tag{5}$$

$$\text{HMCR} = 0.5 \left(1 - \frac{i}{\text{max}(i)} \right), \text{ PAR} = 0.05 \left(1 - \frac{i}{\text{max}(i)} \right) \tag{6}$$

In Equations (3)–(6), HMCR, PAR, $x_{i, \text{min}}$ and $x_{i, \text{max}}$ stand for the harmony memory consideration rate, the pitch adjustment rate, the minimum and the maximum values of the i -th input variable in the population respectively. After each harmony search iteration, the updated solution candidates replace the old ones if they have superior performance and satisfy the structural design code requirements. For a more detailed review of the harmony search technique, the reader is referred to [55].

The variable ranges of the dataset generated using the harmony search algorithm can be seen in Figure 3 and Table 1. In Figure 3, the variable ranges have been divided into

four partitions and the ranges of these partitions have been shown on the horizontal bars whereas the number of samples in each partition is shown inside the horizontal bars. The length of each partition in Figure 3 is in proportion to the number of samples belonging to that partition. Figure 3 shows that the largest partition for the outer diameter D consists of 1213 samples ranging between 0.574 m and 0.661 m. The second largest partition for this variable with 1000 samples ranges between 0.661 m and 0.747 m. The remaining two partitions ranging between the lower bound of the outer diameter of 0.4 m and 0.574 m constitute 29% of the entire dataset. The partitions for the variables C_c, C_s, L, M, and N are evenly distributed. The largest partition for the column length L consists of 1250 samples ranging between 3 m and 4 m, which is 40% of the entire dataset. The horizontal bars for C_c and C_s in Figure 3 show the unit prices of concrete and steel in USD/m³ and USD/ton respectively.



Figure 3. Design variable ranges in the dataset.

Table 1. Variable ranges and statistical properties.

Variable	Min		Max		Average		Standard Deviation		Variance	
	Actual	Normalized	Actual	Normalized	Actual	Normalized	Actual	Normalized	Actual	Normalized
D [m]	0.4	0.647	0.747	1.209	0.618	1	0.087	0.141	0.0076	0.02
C _c [USD/m ³]	50	0.5	150	1.5	100	1	35.4	0.354	1250	0.125
C _s [USD/ton]	750	0.6	1750	1.4	1250	1	354	0.283	125,000	0.08
L [m]	3	0.6	7	1.4	5	1	1.41	0.283	2	0.08
M [kNm]	100	0.333	500	1.667	300	1	141	0.471	20,000	0.222
N [kN]	1000	0.333	5000	1.667	3000	1	1414	0.471	2,000,000	0.222
A _s [mm ²]	1385	0.447	4524	1.459	3101	1	799	0.258	639,080	0.067

Table 1 includes the upper and lower bounds as well as statistical properties of the design variables represented in the dataset. These statistical properties are the average value, standard deviation, and variance of each variable inside the dataset. Furthermore, for each variable, the corresponding boundaries and statistical properties have been listed after normalizing the variables by their average values. These normalized values are used at a later stage for the development of a predictive equation. Also, the partitions presented in Figure 3 are used as the basis of a four-level factorial analysis to determine the variation of A_s for each design variable. In addition to the partition plot in Figure 3, also a correlation plot has been generated for the dataset (Figure 4). For each input variable and the output variable A_s, Figure 4 shows the Pearson correlation coefficient between any two of these variables in the upper right portion of the diagram. Pearson correlation values close to 1 indicate a high correlation between two variables. The highest correlation coefficient in Figure 4 can be observed between A_s and D which indicates that the outer diameter has a significant impact on the reinforcement area. The second highest correlation can be observed between A_s and N with a Pearson correlation value of 0.79. Another relatively

high correlation is observed between D and N with a correlation value of 0.77. Finally, the column length L is correlated to A_s and D with correlation coefficients of 0.53 and 0.54 respectively. Greater correlation between variables is represented by larger font size and stars inside the tiles of the matrix. In Figure 4, each variable occupies one of the diagonal tiles and the scale of this variable is shown in one of the horizontal axes and one of the vertical axes. Furthermore, each diagonal tile contains a histogram showing the distribution of the variable in it. The lower left portion of the correlation matrix contains bivariate scatter plots with regression lines. The equation for the computation of the Pearson correlation coefficient as well as the other metrics of accuracy used in this paper can be found in Appendix A.

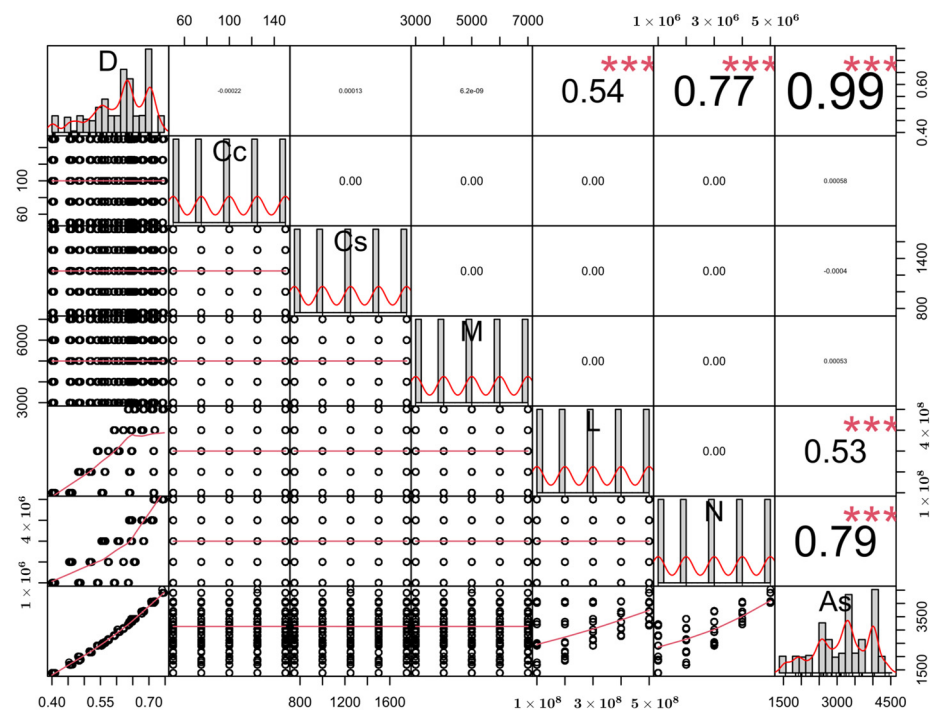


Figure 4. Correlation matrix of the dataset.

Using the dataset whose properties have been shown in Figure 3, Table 1, and Figure 4, four different data-driven predictive models have been trained using the ensemble learning algorithms XGBoost, LightGBM, Random Forest, and CatBoost. The results of them and their interpretations using the SHapley Additive exPlanations (SHAP) technique have been presented in the next section. The theoretical background of ensemble learning and SHAP algorithms can be found in [3,56–62]. In addition to the SHAP analysis also, a four-level factorial analysis has been carried out to further investigate the sensitivity of A_s to the variations in different design variables. Afterward, a closed-form equation has been proposed for the prediction of A_s as a function of the six design variables shown in Figures 3 and 4. The overall process of dataset generation, training of the machine learning models, and the development of a predictive closed-form equation have been summarized in a flow chart in Figure 5.

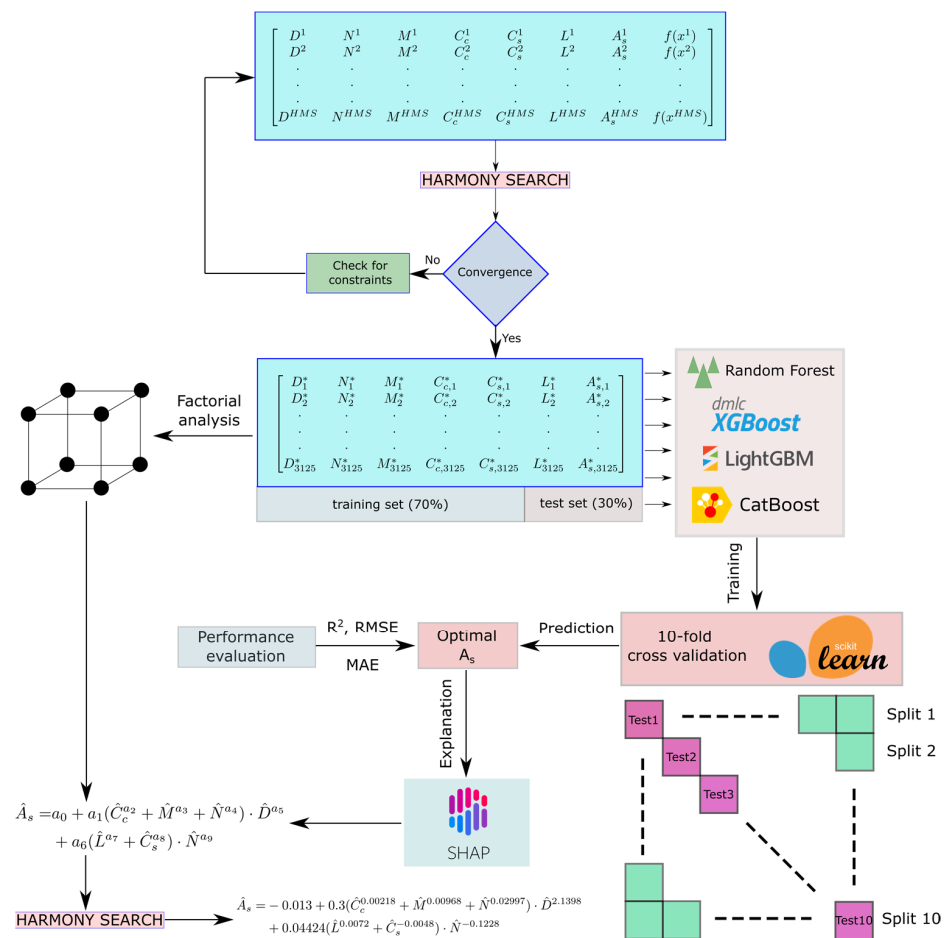


Figure 5. Machine learning and equation development process.

3. Results

This section presented the comparison of the optimal reinforcement areas predicted by the ensemble learning algorithms with the true optimal values obtained through the harmony search algorithm. The performances of each predictive model have been measured by the metrics of the coefficient of determination, mean absolute error and root mean squared error. The outcome of the ensemble learning models has been analyzed with the SHAP algorithm to determine the variables with the highest impact on the model predictions. Furthermore, a four-level factorial analysis has been performed to visualize the variation of the optimal reinforcement cross-section for each input variable. Based on the outcome of the SHAP and factorial analyses, a predictive equation format has been proposed. This predictive equation has been developed using the harmony search algorithm, and the accuracy of the obtained equation has been demonstrated by the same accuracy metrics applied to the ensemble learning models.

3.1. Ensemble Learning Model Predictions

The ensemble learning models have been trained by splitting the entire dataset into a training set and a test set in 70% to 30% proportions. This division was made based on past machine learning studies in the area of structural engineering. Particularly, the study of Nguyen et al. [63] demonstrated that among the 10/90, 20/80, 30/70, 40/60, 50/50, 60/40, 70/30, 80/20, and 90/10 training set to test set ratios, the 70/30 ratio delivered the best performance. The models have been trained on the training set using ten-fold cross-validation. After the completion of the model training, the test set was used to measure the model performances. The performances of the ensemble learning models have been visualized by plotting the true optimal A_s values of the test set against the

A_s values predicted by the ensemble learning models. Figure 6 shows the comparison of the predicted and actual optimal A_s values for each of the four predictive models. In Figure 6, the diagonal solid lines represent the case when the actual and predicted values are equal, whereas the dotted lines represent $\pm 10\%$ deviation from a perfect prediction. The performances of these predictive models are compared to each other in Table 2. According to Table 2, the Random Forest algorithm demonstrated the best performance in terms of both prediction accuracy on the test set and the speed of execution (3.71 s).

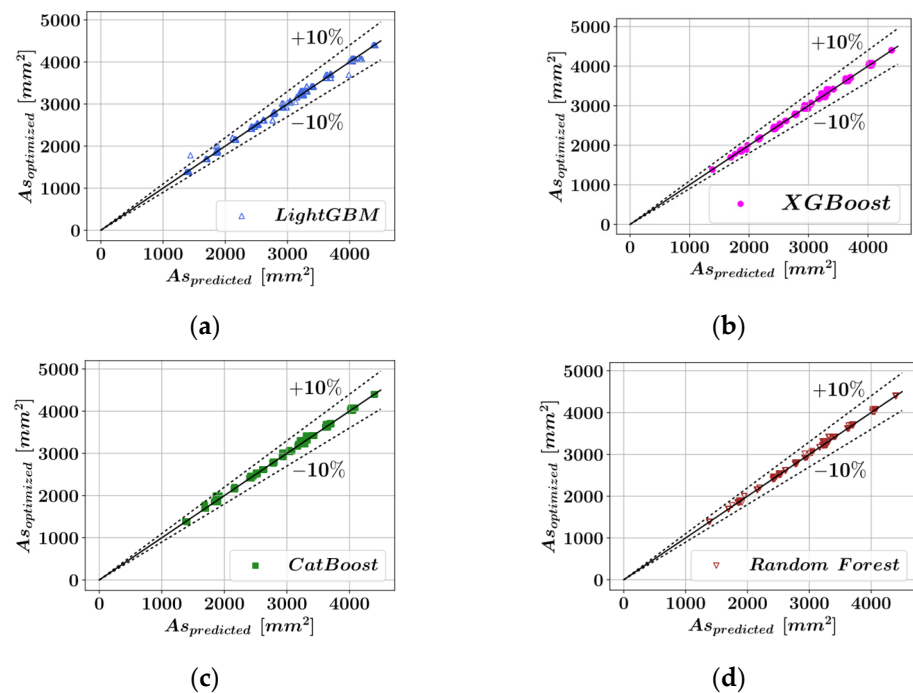


Figure 6. Comparison of the predicted and optimized dimensions. (a) LightGBM. (b) XGBoost. (c) CatBoost. (d) Random Forest.

Table 2. Prediction accuracy of the machine learning models.

Algorithm	Variable	R^2		MAE		RMSE		Duration [s]
		Train	Test	Train	Test	Train	Test	
XGBoost	A_s	0.9999	0.9995	1.998	7.523	3.839	17.072	5.14
Random Forest	A_s	0.9999	0.9996	2.593	7.111	6.095	15.929	3.71
LightGBM	A_s	0.9994	0.9988	9.962	12.767	19.673	27.157	6.07
CatBoost	A_s	0.9998	0.9994	7.579	10.788	12.440	18.940	28.23

3.2. SHAP Analysis

The SHAP analysis visually describes the contribution of each design variable to the prediction of a machine learning model. The SHAP summary plot and feature dependence plots in this section are based on the Random Forest algorithm selected due to its superior performance on this dataset. The SHAP summary plot in Figure 7 ranks the six input variables according to their impact on the predictive model output. In Figure 7, each data sample is represented by a dot and positive SHAP values indicate an increasing effect of a variable on the model output whereas negative SHAP values indicate a decreasing effect on the model output. The impact of a variable on model output for a particular data sample is a function of the position of a dot along the horizontal axis. On the other hand, the numerical values of the input parameters are represented with color so that high parameter values are shown with shades of red and the low parameter values are shown with shades of blue. According to Figure 7, the outer diameter D has the greatest impact on the model

output. It can be deduced that an increase in the value of D also leads to an increase in the model prediction. Conversely, decreasing the D value leads to lower model predictions. On the other hand, the impacts of the remaining parameters on the model output are an order of magnitude smaller than the impact of D according to Figure 7.

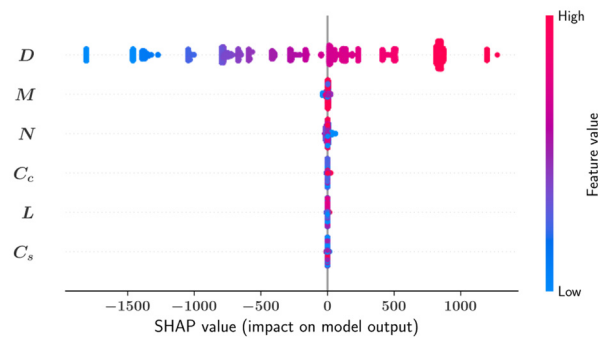


Figure 7. SHAP summary plot for A_s (Random Forest).

The feature dependence plots in Figure 8 can help better understand the interdependencies between different variables and their effects on the model output. Figure 8a clearly shows that as the D value increases, so does the SHAP value for this variable. This confirms the inference from Figure 7 that increased D values lead to increased reinforcement area. The colors of the dots in Figure 8a indicate the numerical values of C_c which is the variable most dependent on D . On the other hand, the feature dependence plots of C_c , C_s and L show that the SHAP values of these variables have a horizontal trend as the variable values increase. For these three variables, most of the SHAP values stay in the range of -5 to 5 for the entire dataset. For a significant portion of the dataset, the SHAP values are clustered around zero. Also, for all three of these variables, N is the variable most dependent on them. The feature dependence plots of M and N exhibit certain patterns for different levels of these variables. Figure 8e,f show that the feature dependence plots for M and N are fragmented and can be investigated separately for different levels of these variables. For each value of M and N , the SHAP values of these variables are concentrated around different levels depending on the value of D which is the parameter most dependent on M and N . It can be observed that M has the greatest impact on the model output when $M = 3 \times 10^8 \text{ Nmm}$ and D have large values shown in red and the least impact when $M = 10^8 \text{ Nmm}$ and D have small values shown in blue. Similarly, N has the greatest impact when $N = 10^6 \text{ N}$ and D have small values shown in blue and the least impact when $N = 3 \times 10^6 \text{ N}$ and D have large values shown in red.

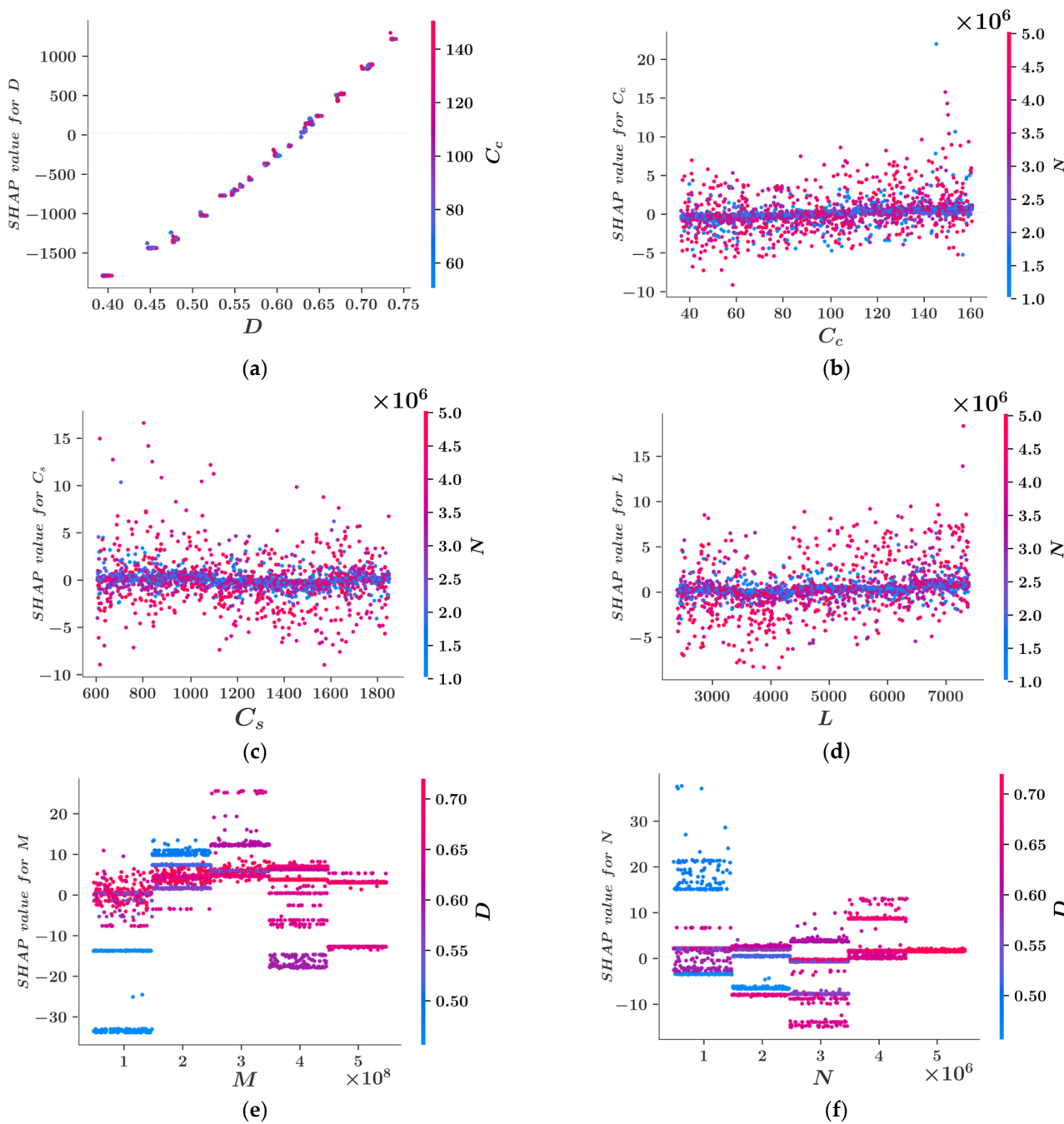


Figure 8. Feature dependence plots for the variables (a) D , (b) C_c , (c) C_s , (d) L , (e) M , (f) N .

3.3. Four-Level Factorial Analysis

The factorial analysis technique is widely used for gaining insights into the response and sensitivity of a system that depends on multiple variables to the variations in a single variable. The factorial analysis technique is particularly useful when the variables of a system can be broken down into different levels. In this paper, the cross-section of reinforcement is predicted as a function of six variables. Each of these variables has been broken down into four levels as shown in Figure 3. The four-level factorial analysis enables the visualization of the nonlinear variations in the target variables or curvatures in the system response [64]. Afterward, for each level of each variable, the average value of the area of reinforcement has been calculated. These average values are plotted for different levels of each variable in Figure 9. A significant variation of A_s in Figure 9 with respect to changes in a certain variable indicates the high sensitivity of A_s to the changes in this

variable. According to Figure 9, the greatest change in the value of A_s is observed when D increases from its lowest level (level 0) to its highest level (level 3). A total increase from 1641 mm^2 to 3993 mm^2 can be observed which corresponds to a 143% increase. The second largest increase in the average value of A_s can be observed when N increases from its lowest level to its highest level. In this case, the increase of the area from 2459 mm^2 to 4119 mm^2 can be observed which corresponds to a 68% increase. The third largest percentage-wise increase in the average A_s value can be observed when M increases from its lowest level to its highest level. An increase from an average area of 2661 mm^2 to an average area of 3750 mm^2 can be observed which corresponds to a 41% increase. For the remaining three variables C_c , C_s and L , the changes in the average value of A_s was negligible in comparison to D , N , and M .

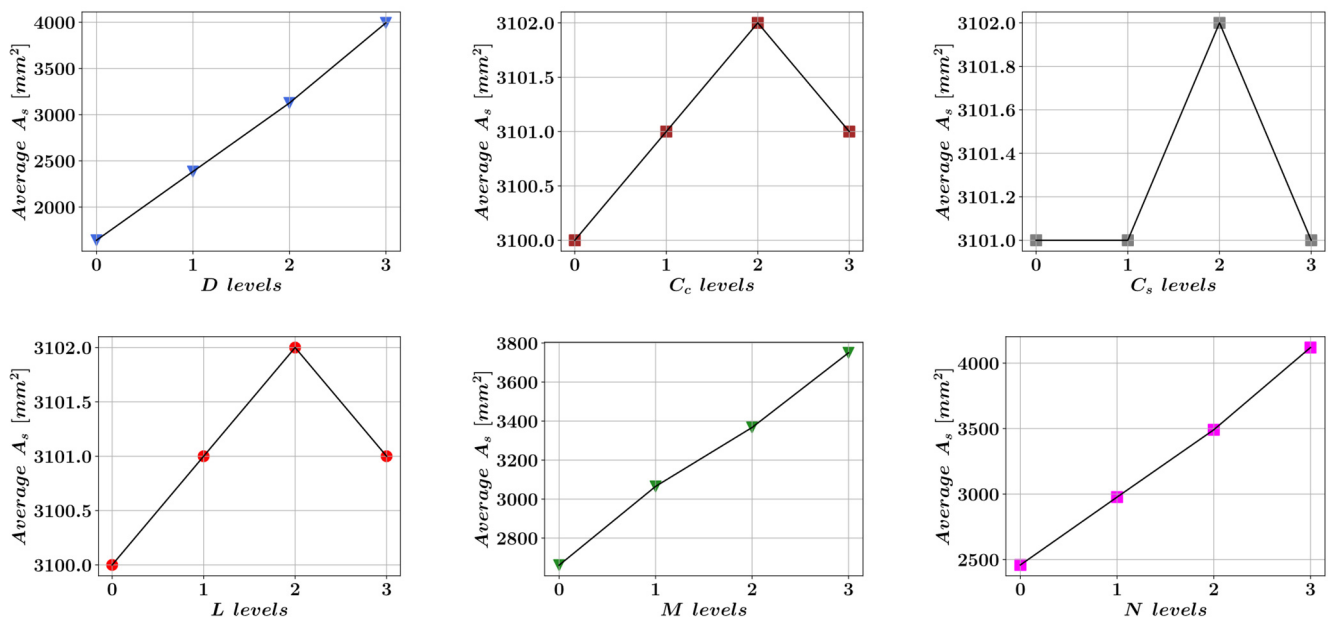


Figure 9. Four-level factorial analysis.

3.4. Development of an Equation for the Prediction of the Optimum Reinforcement Area

In light of the results presented in the previous sections, the formula in Equation (7) has been proposed for the prediction of the reinforcement area in an optimal design. In Equation (7) the variables $\hat{A}_s, \hat{C}_c, \hat{M}, \hat{N}, \hat{D}, \hat{L}, \hat{C}_s$ are normalized by the average value of each variable in the dataset consisting of 3125 samples.

$$\hat{A}_s = a_0 + a_1 \left(\hat{C}_c^{a_2} + \hat{M}^{a_3} + \hat{N}^{a_4} \right) \cdot \hat{D}^{a_5} + a_6 \left(\hat{L}^{a_7} + \hat{C}_s^{a_8} \right) \cdot \hat{N}^{a_9} \quad (7)$$

The coefficients a_0 to a_9 in Equation (7) have been adjusted using harmony search iterations. This process necessitates the declaration of a new harmony memory matrix that contains the coefficients of Equation (7) as shown in Equation (8) where the population consists of 30 different solution candidates.

$$HM = \begin{bmatrix} a_0^1 & a_1^1 & a_2^1 & a_3^1 & a_4^1 & a_5^1 & a_6^1 & a_7^1 & a_8^1 & a_9^1 \\ a_0^2 & a_1^2 & a_2^2 & a_3^2 & a_4^2 & a_5^2 & a_6^2 & a_7^2 & a_8^2 & a_9^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_0^{30} & a_1^{30} & a_2^{30} & a_3^{30} & a_4^{30} & a_5^{30} & a_6^{30} & a_7^{30} & a_8^{30} & a_9^{30} \end{bmatrix} \quad (8)$$

After every harmony search iteration, the performances of the solution candidates are measured by comparing their predictions of \hat{A}_s with the actual \hat{A}_s values for the entire dataset. The prediction error is represented by the Euclidean norm of the vector containing the differences between the actual and predicted optimal \hat{A}_s values. The development of

these vector norms for the best- and worst-performing members of the harmony memory population is presented in Figure 10.

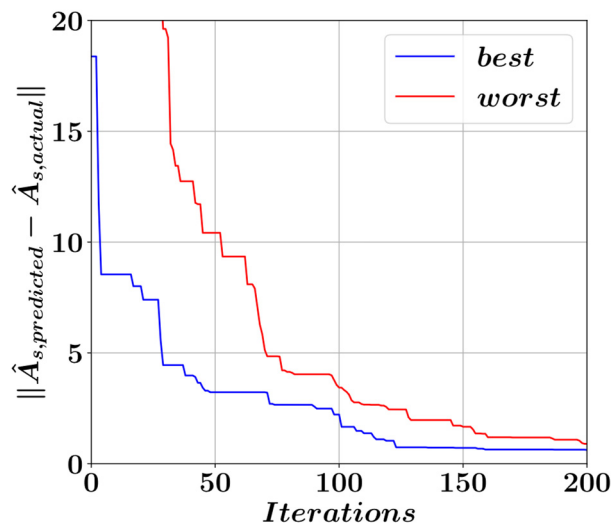


Figure 10. Equation performance throughout the harmony search iterations.

A total of 5000 harmony search iterations was carried out to obtain the best possible equation coefficients with the smallest possible error norm. Figure 10 shows the development of the best and worst solution candidates in the initial 200 iterations. It should be noted that the largest improvements in the solution candidates take place during the initial phases of the harmony search iterations. Figure 11 shows the process of obtaining the coefficients a_0 to a_9 that minimize the difference between the optimal \hat{A}_s values predicted by Equation (7) and the actual optimal \hat{A}_s values.

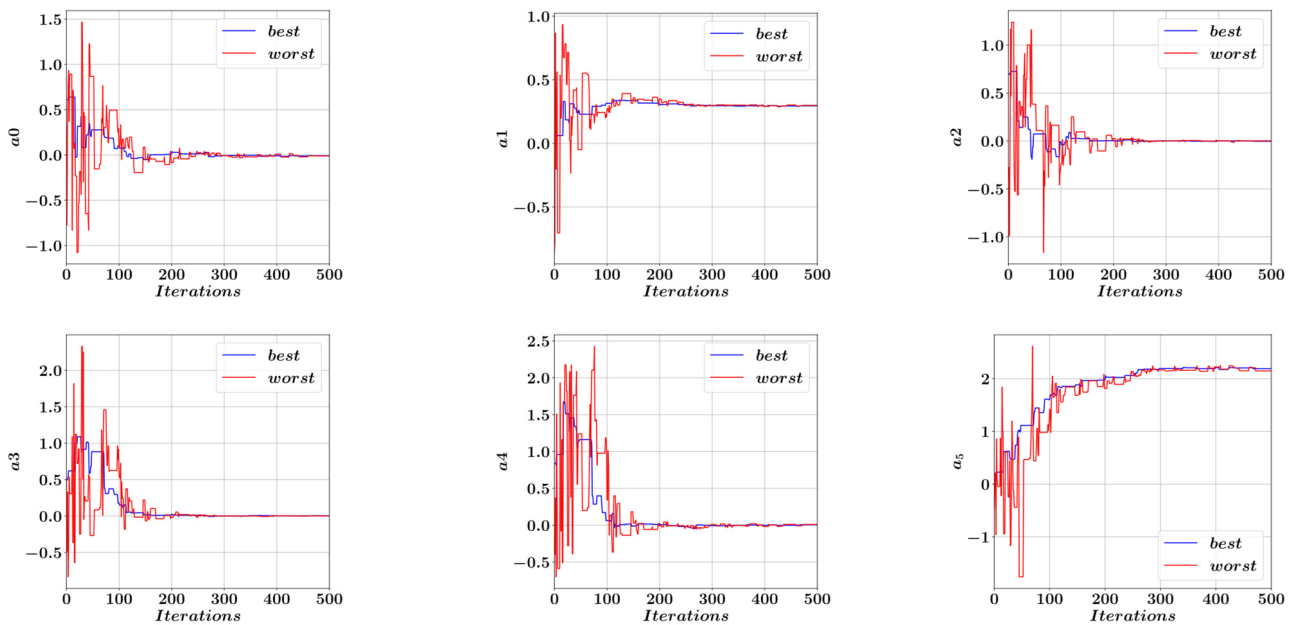


Figure 11. Cont.

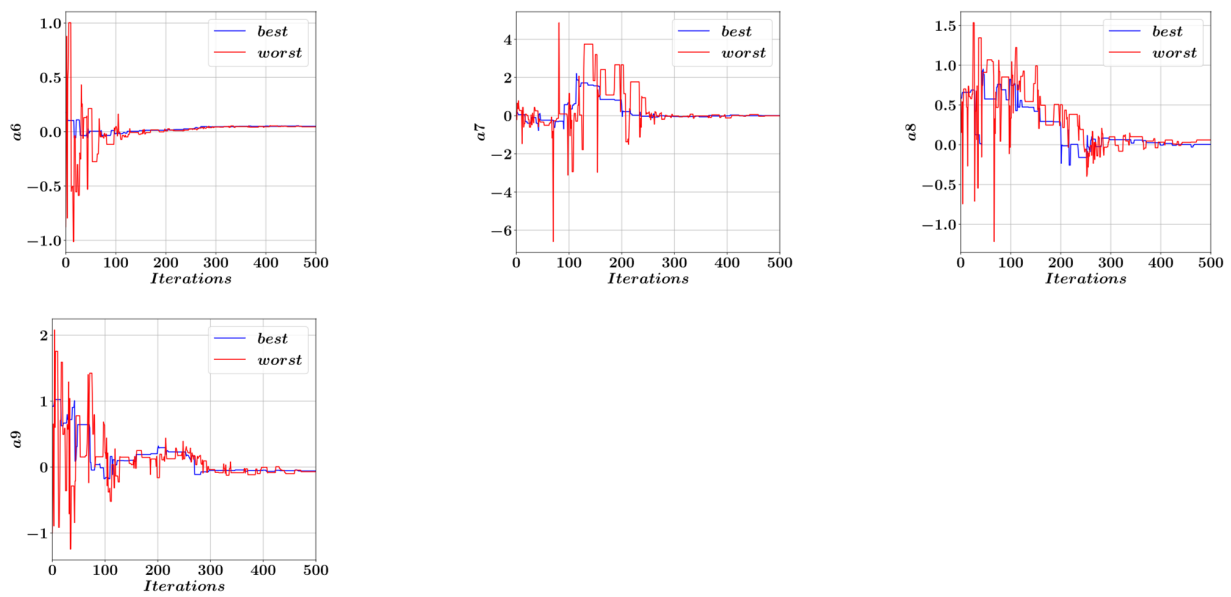


Figure 11. Development of the equation coefficients in Equation (7).

Figure 11 shows the values of the coefficients a_0 to a_9 in the first 500 harmony search iterations. The coefficient values corresponding to the best- and worst-performing members of the harmony memory population are shown in blue and red respectively. It can be observed that after the initial fluctuations, these coefficients tend to converge to their optimal limit values. Inserting these limit values of the coefficients a_0 to a_9 in Equation (7), we obtain Equation (9) for the prediction of the normalized reinforcement area \hat{A}_s from which the actual reinforcement area A_s can be obtained after multiplication with the average value of A_s from Table 1.

$$\hat{A}_s = -0.013 + 0.3 \left(\hat{C}_c^{0.00218} + \hat{M}^{0.00968} + \hat{N}^{0.02997} \right) \cdot \hat{D}^{2.1398} + 0.04424 \left(\hat{L}^{0.0072} + \hat{C}_s^{-0.0048} \right) \cdot \hat{N}^{-0.1228} \tag{9}$$

A coefficient of determination of 0.9985, mean absolute error of 0.0076, and root mean square error of 0.0099 could be achieved using Equation (9). The \hat{A}_s values predicted by Equation (9) are plotted against the actual \hat{A}_s values in Figure 12 where the dotted lines indicate the $\pm 10\%$ deviation from a perfect match between the predicted and actual values.

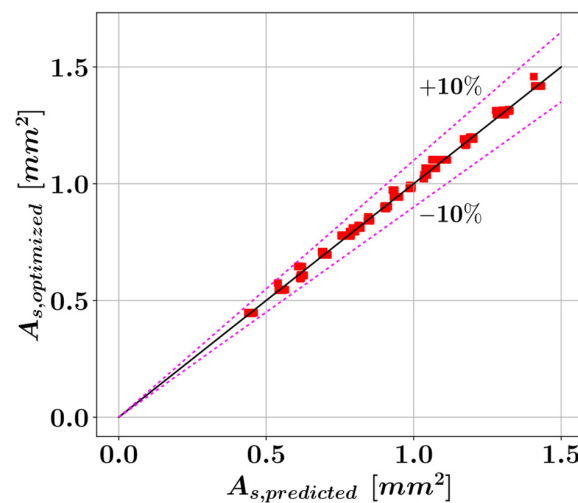


Figure 12. Comparison of the predicted and actual optimal reinforcement areas.

4. Discussion

Circular RC columns are ubiquitous in the structural design of buildings, bridges, and ports. Therefore, the optimal design of these structural members can have enormous economic and environmental benefits. This paper demonstrates the applicability of machine learning models to the structural optimization of RC circular columns. The total cross-sectional area of the longitudinal reinforcement was selected as the decisive parameter in the design of these structures. A novel technique has been applied in the training phase of these machine learning models. The harmony search algorithm was used to generate a large dataset consisting of 3125 samples where each sample represents an optimum configuration of column geometry and external loading. XGBoost, LightGBM, Random Forest, and CatBoost algorithms are used to generate four different ensemble learning predictive models. The performances of these models have been compared using the coefficient of determination (R^2), mean absolute error (MAE) and root mean squared error (RMSE) as the metrics of accuracy. It was found that the Random Forest algorithm performed better than the other three ensemble learning algorithms in terms of both the accuracy of predicted optimal cross-sectional areas and the computational speed. However, it should be noted that all ensemble learning algorithms demonstrated high accuracy with an R^2 score greater than 0.99. On the other hand, the execution speed of the CatBoost algorithm was significantly slower than the remaining three algorithms. The output of the Random Forest algorithm has been further analyzed using the SHAP algorithm to better understand the impact of each input variable on the model prediction and the interdependencies between the input variables. Furthermore, a four-level factorial analysis has been carried out to visualize the sensitivity of the reinforcement cross-sectional area (A_s) to various input parameters.

The results of the four-level factorial analysis showed that the column diameter and applied external loads largely determine the cross-sectional area of steel reinforcement necessary for a safe design. The outer diameter of the column was also singled out by the SHAP summary plot as the most impactful design variable that determines the output of the random forest model. The outcome of the SHAP analysis and the factorial analysis can be interpreted as the external loading and column size being an order of magnitude more significant than the unit costs of steel and concrete in terms of their effect on the necessary amount of reinforcement.

In the design of a predictive equation, both the factorial analysis and the SHAP feature dependence plots have been decisive. In addition to a bias term, the predictive equation consists of two other terms. The first term after the bias term is chosen to be a product of the normalized diameter with the sum of the three parameters most dependent on D according to the feature dependence plots. The second term is chosen to be a product of the axial load with the sum of the normalized values of the two design variables most dependent on the axial load. In order to capture the nonlinear variations in the optimum reinforcement area, each term in the equation was raised to a power and multiplied by a coefficient. Afterward, the optimal coefficient values were determined through harmony search optimization. A coefficient of determination greater than 0.99 could be obtained by the predictive equation thus developed. However, it should be noted that more comprehensive studies including a larger number of design variables and greater ranges for the variable values should be performed to enhance the reliability of the developed equation.

Since machine learning techniques are data-driven, the applicability of these techniques to structural design depends on the availability of quality datasets. Furthermore, the accuracy of the obtained predictive models depends largely on the size of the dataset. However, most of the recent machine learning related research in the field of structural engineering is based on data sets not large enough to be statistically significant. This issue is being addressed in this study by proposing a harmony search-based novel technique for generating large data sets. The application of the harmony search methodology to the problem of data generation can solve the problem of limited data availability in structural engineering. Using optimization techniques such as harmony search, large datasets can

be generated where the structural performances of the generated optimal samples are controlled according to the requirements in the existing design codes.

5. Conclusions

The performance of any predictive model depends mostly on the size and quality of the dataset used in its training. The current paper demonstrated a novel technique for the generation of large datasets using the harmony search optimization methodology. The generated dataset was used in the prediction of the optimal amount of longitudinal reinforcement for circular RC columns. Four different ensemble learning models were demonstrated to perform well in the prediction process. Furthermore, a closed-form equation was proposed that predicts the optimal amount of reinforcement that minimizes the cost associated with the construction process without violating the design code requirements. The most important results of this research work can be listed as follows:

- Four different machine learning models were developed using the XGBoost, LightGBM, Random Forest, and CatBoost algorithms. All of these algorithms performed well on the dataset with an R^2 score greater than 0.99. Among these models, the Random Forest algorithm performed best in terms of both accuracy and computational speed whereas the CatBoost algorithm was nearly an order of magnitude slower than the rest of the algorithms.
- The results of the SHAP analysis showed that the outer diameter of the circular column has the greatest impact on the machine learning model predictions. The impacts of the applied axial loading (N) and bending moment (M) were found to be dependent on the value of D . At smaller values of D , N was shown to have a larger impact on the model output.
- After dividing the dataset into four segments for each variable the four-level factorial analysis showed that a 59% increase in the outer diameter can lead to a 143% increase in the optimal value of A_s . A_s was also found to be highly sensitive to variations in N and M . Doubling the magnitude of N was observed to cause a 68% increase in the optimal value of A_s whereas doubling the magnitude of M led to a 41% increase in the optimal value of A_s .
- A closed-form equation with an R^2 score of 0.9985 was proposed which predicts the optimal value for A_s as a function of column outer diameter, axial loading, bending moment, column length, and the unit prices of concrete and steel.

The availability of closed-form equations that deliver optimal dimensions for structural design can greatly facilitate and accelerate the design process for practicing engineers. With the help of these equations, the most favorable design combinations can be obtained without the need for complex optimization methodologies. However, it should be noted that the proposed equation and machine learning models in this paper are limited by the range of variables that constitute the dataset. Therefore, further research needs to be carried out for the development of more comprehensive predictive models. Furthermore, the scope of the variables included in the dataset could be enhanced to include variables such as the number and diameter of longitudinal reinforcing bars. In its current form, the output of the predictive equation could be evenly distributed to determine the area and a total number of the longitudinal reinforcing bars. Also, the spacing and dimensions of the lateral reinforcements can be included in the database. Future research towards the design of RC columns using machine learning methodologies can include composite materials such as carbon fiber or glass fiber reinforced polymers as the material of reinforcement.

Author Contributions: Methodology, G.B.; formal analysis (coding), C.C.; writing—original draft preparation, C.C. and G.B.; writing—review & editing, S.K. and Z.W.G.; visualization, C.C.; supervision, G.B., S.K. and Z.W.G.; funding acquisition, Z.W.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Energy Cloud R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (2019M3F2A1073164).

Data Availability Statement: Data will be made available upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In Table A1, x_i and y_i denote the values of two different data series, \tilde{y}_i denotes the predicted values of the data series y_i , and n is the total number of data points in the series.

Table A1. Metrics of Model Accuracy.

Root mean square error (RMSE)	$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{n}}$
Coefficient of determination (R^2):	$R^2 = \left(\frac{n \sum_{i=1}^n y_i \tilde{y}_i - \sum_{i=1}^n y_i \sum_{i=1}^n \tilde{y}_i}{\sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2} \sqrt{n \sum_{i=1}^n \tilde{y}_i^2 - (\sum_{i=1}^n \tilde{y}_i)^2}} \right)^2$
Mean absolute error (MAE):	$\text{MAE} = \frac{\sum_{i=1}^n y_i - \tilde{y}_i }{n}$
Pearson correlation coefficient:	$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}$

References

- Bekdaş, G.; Cakiroglu, C.; Islam, K.; Kim, S.; Geem, Z.W. Optimum Design of Cylindrical Walls Using Ensemble Learning Methods. *Appl. Sci.* **2022**, *12*, 2165. [\[CrossRef\]](#)
- Bekdaş, G. Harmony Search Algorithm Approach for Optimum Design of Post-Tensioned Axially Symmetric Cylindrical Reinforced Concrete Walls. *J. Optim. Theory Appl.* **2015**, *164*, 342–358. [\[CrossRef\]](#)
- Bekdaş, G.; Cakiroglu, C.; Kim, S.; Geem, Z.W. Optimal Dimensioning of Retaining Walls Using Explainable Ensemble Learning Algorithms. *Materials* **2022**, *15*, 4993. [\[CrossRef\]](#)
- Arama, Z.A.; Kayabekir, A.E.; Bekdaş, G.; Kim, S.; Geem, Z.W. The Usage of the Harmony Search Algorithm for the Optimal Design Problem of Reinforced Concrete Retaining Walls. *Appl. Sci.* **2021**, *11*, 1343. [\[CrossRef\]](#)
- Kayabekir, A.E.; Arama, Z.A.; Bekdaş, G.; Nigdeli, S.M.; Geem, Z.W. Eco-Friendly Design of Reinforced Concrete Retaining Walls: Multi-objective Optimization with Harmony Search Applications. *Sustainability* **2020**, *12*, 6087. [\[CrossRef\]](#)
- Cakiroglu, C.; Bekdaş, G.; Kim, S.; Geem, Z.W. Optimisation of Shear and Lateral-Torsional Buckling of Steel Plate Girders Using Meta-Heuristic Algorithms. *Appl. Sci.* **2020**, *10*, 3639. [\[CrossRef\]](#)
- Cakiroglu, C.; Bekdaş, G. Buckling Analysis and Stacking Sequence Optimization of Symmetric Laminated Composite Plates. In *Advances in Structural Engineering—Optimization. Studies in Systems, Decision and Control*; Nigdeli, S.M., Bekdaş, G., Kayabekir, A.E., Yucel, M., Eds.; Springer: Cham, Switzerland, 2021; Volume 326. [\[CrossRef\]](#)
- Cakiroglu, C.; Bekdaş, G.; Geem, Z.W. Harmony Search Optimisation of Dispersed Laminated Composite Plates. *Materials* **2020**, *13*, 2862. [\[CrossRef\]](#)
- Cakiroglu, C.; Islam, K.; Bekdaş, G.; Kim, S.; Geem, Z.W. Metaheuristic Optimization of Laminated Composite Plates with Cut-Outs. *Coatings* **2021**, *11*, 1235. [\[CrossRef\]](#)
- Cakiroglu, C.; Islam, K.; Bekdaş, G.; Billah, M. CO₂ emission and cost optimization of concrete-filled steel tubular (CFST) columns using metaheuristic algorithms. *Sustainability* **2021**, *13*, 8092. [\[CrossRef\]](#)
- Cakiroglu, C.; Islam, K.; Bekdaş, G.; Kim, S.; Geem, Z.W. CO₂ Emission Optimization of Concrete-Filled Steel Tubular Rectangular Stub Columns Using Metaheuristic Algorithms. *Sustainability* **2021**, *13*, 10981. [\[CrossRef\]](#)
- Toklu, Y.C.; Temur, R.; Bekdaş, G. Teaching learning based optimization algorithm for analyses of trusses considering elastoplastic behavior. In *Fluids, Heat and Mass Transfer, Mechanical and Civil Engineering*; Recent Advances in Mechanical Engineering Series 17; WSEAS Press: Athens, Greece, 2015; ISBN 978-1-61804-358-0.
- Ulusoy, S. Optimum design of timber structures under fire using metaheuristic algorithm. *Grđevinar* **2022**, *74*, 115–124. [\[CrossRef\]](#)
- Ocak, A.; Nigdeli, S.M.; Bekdaş, G.; Kim, S.; Geem, Z.W. Adaptive Harmony Search for Tuned Liquid Damper Optimization under Seismic Excitation. *Appl. Sci.* **2022**, *12*, 2645. [\[CrossRef\]](#)
- Ocak, A.; Bekdaş, G.; Nigdeli, S.M.; Kim, S.; Geem, Z.W. Optimization of Tuned Liquid Damper Including Different Liquids for Lateral Displacement Control of Single and Multi-Story Structures. *Buildings* **2022**, *12*, 377. [\[CrossRef\]](#)
- Ulusoy, S.; Nigdeli, S.M.; Bekdaş, G. Introduction and Review on Active Structural Control. In *Optimization of Tuned Mass Dampers. Studies in Systems, Decision and Control*; Bekdaş, G., Nigdeli, S.M., Eds.; Springer: Cham, Switzerland, 2022; Volume 432. [\[CrossRef\]](#)
- Khalilpourazari, S.; Khalilpourazary, S. An efficient hybrid algorithm based on Water Cycle and Moth-Flame Optimization algorithms for solving numerical and constrained engineering optimization problems. *Soft Comput.* **2019**, *23*, 1699–1722. [\[CrossRef\]](#)

18. Mirjalili, S. SCA: A sine cosine algorithm for solving optimization problems. *Knowl. -Based Syst.* **2016**, *96*, 120–133. [[CrossRef](#)]
19. Harifi, S.; Mohammadzadeh, J.; Khalilian, M.; Ebrahimnejad, S. Giza Pyramids Construction: An ancient-inspired metaheuristic algorithm for optimization. *Evol. Intel.* **2021**, *14*, 1743–1761. [[CrossRef](#)]
20. Holland, H.; Reitman, J.S. Cognitive systems based on adaptive algorithms. *ACM SIGART Bull.* **1977**, *63*, 49. [[CrossRef](#)]
21. Storn, R.; Price, K. Differential Evolution—A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *J. Glob. Optim.* **1997**, *11*, 341–359. [[CrossRef](#)]
22. Fogel, D.B. Artificial Intelligence through Simulated Evolution. In *Evolutionary Computation: The Fossil Record*; IEEE: Piscataway, NJ, USA, 1998; pp. 227–296. [[CrossRef](#)]
23. Koza, J.R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*; MIT Press: Cambridge, MA, USA, 1992; ISBN 0-262-11170-5.
24. Simon, D. Biogeography-based optimization. *IEEE Trans. Evol. Comput.* **2008**, *12*, 702–713. [[CrossRef](#)]
25. Rashedi, E.; Nezamabadi-Pour, H.; Saryazdi, S. GSA: A gravitational search algorithm. *Inf. Sci.* **2009**, *179*, 2232–2248. [[CrossRef](#)]
26. Kaveh, A.; Mahdavi, V.R. Colliding bodies optimization method for optimum discrete design of truss structures. *Comput. Struct.* **2014**, *139*, 43–53. [[CrossRef](#)]
27. Hatamlou, A. Black hole: A new heuristic optimization approach for data clustering. *Inf. Sci.* **2013**, *222*, 175–184. [[CrossRef](#)]
28. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [[CrossRef](#)]
29. Mirjalili, S.; Mirjalili, S.M.; Hatamlou, A. Multi-verse optimizer: A nature-inspired algorithm for global optimization. *Neural Comput. Appl.* **2016**, *27*, 495–513. [[CrossRef](#)]
30. Du, H.; Wu, X.; Zhuang, J. Small-World Optimization. In *Advances in Natural Computation. ICNC 2006: Lecture Notes in Computer Science*; Jiao, L., Wang, L., Gao, X., Liu, J., Wu, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4222. [[CrossRef](#)]
31. Mirjalili, S. The ant lion optimizer. *Adv. Eng. Soft.* **2015**, *83*, 80–98. [[CrossRef](#)]
32. Harifi, S.; Khalilian, M.; Mohammadzadeh, J. Emperor Penguins Colony: A new metaheuristic algorithm for optimization. *Evol. Intel.* **2019**, *12*, 211–226. [[CrossRef](#)]
33. Li, M.D.; Zhao, H.; Weng, X.W.; Han, T. A novel nature-inspired algorithm for optimization: Virus colony search. *Adv. Eng. Soft.* **2016**, *92*, 65–88. [[CrossRef](#)]
34. Salimi, H. Stochastic fractal search: A powerful metaheuristic algorithm. *Knowl. -Based Syst.* **2015**, *75*, 1–18. [[CrossRef](#)]
35. Kashan, A.H. League Championship Algorithm (LCA): An algorithm for global optimization inspired by sport championships. *Appl. Soft Comput.* **2014**, *16*, 171–200. [[CrossRef](#)]
36. Geem, Z.W.; Lee, K.S.; Park, Y. Application of harmony search to vehicle routing. *Am. J. Appl. Sci.* **2005**, *2*, 1552–1557. [[CrossRef](#)]
37. Zhang, Y.; Su, R.; Zhang, Y.; Gammuna Guruge, N.S. A Multi-Bus Dispatching Strategy Based on Boarding Control. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 5029–5043. [[CrossRef](#)]
38. Villalobos, O.A.R.; Rojas, E.J.R.; Velandia, J.B.; Cuchango, H.E.E. Application of the Harmonic Search Algorithm for Identification of Model Parameters of Traffic Lights for a High Way of Bogotá. *Int. J. Eng. Res. Technol.* **2020**, *13*, 240–246. [[CrossRef](#)]
39. Ganeshkumar, N.; Kumar, S. QoS Aware Modified Harmony Search Optimization for Route Selection in VANETs. *Indian J. Comput. Sci. Eng. (IJCSE)* **2022**, *13*, 2. [[CrossRef](#)]
40. You, C.-H.; Suh, S.-H.; Jung, W.; Kim, H.-J.; Lee, D.-I. Dual-Polarization Radar-Based Quantitative Precipitation Estimation of Mountain Terrain Using Multi-Disdrometer Data. *Remote Sens.* **2022**, *14*, 2290. [[CrossRef](#)]
41. Loy-Benitez, J.; Li, Q.; Nam, K.; Nguyen, H.T.; Kim, M.; Park, D.; Yoo, C. Multi-objective optimization of a time-delay compensated ventilation control system in a subway facility—A harmony search strategy. *Build. Environ.* **2021**, *190*, 107543. [[CrossRef](#)]
42. Abdulkhaleq, M.T.; Rashid, T.A.; Alsadoon, A.; Hassan, B.A.; Mohammadi, M.; Abdullah, J.M.; Vimal, S. Harmony search: Current studies and uses on healthcare systems. *Artif. Intell. Med.* **2022**, *131*, 102348. [[CrossRef](#)]
43. Taghipour, S.; Zarrineh, P.; Ganjtabesh, M. Improving protein complex prediction by reconstructing a high-confidence protein-protein interaction network of Escherichia coli from different physical interaction data sources. *BMC Bioinform.* **2017**, *18*, 10. [[CrossRef](#)]
44. Li, X.; Wang, B.; Lv, H.; Yin, Q.; Zhang, Q.; Wei, X. Constraining DNA Sequences With a Triplet-Bases Unpaired. *IEEE Trans. NanoBiosci.* **2020**, *19*, 299–307. [[CrossRef](#)]
45. Mohsen, A.M.; Khader, A.T.; Ramachandram, D. HSRNAFold: A harmony search algorithm for RNA secondary structure prediction based on minimum free energy. In Proceedings of the 2008 International Conference on Innovations in Information Technology, Al Ain, United Arab Emirates, 16–18 December 2008; pp. 11–15. [[CrossRef](#)]
46. Bibartiu, O.; Dürr, F.; Rothmel, K. Optimal Refinement for Component-based Architectures. In Proceedings of the 2021 IEEE 25th International Enterprise Distributed Object Computing Conference (EDOC), Gold Coast, Australia, 25–29 October 2021; pp. 142–151. [[CrossRef](#)]
47. Schofield, J.S.; Evans, K.R.; Hebert, J.S.; Marasco, P.D.; Carey, J.P. The effect of biomechanical variables on force sensitive resistor error: Implications for calibration and improved accuracy. *J. Biomech.* **2016**, *49*, 786–792. [[CrossRef](#)]
48. Feng, D.C.; Liu, Z.T.; Wang, X.D.; Jiang, Z.M.; Liang, S.X. Failure mode classification and bearing capacity prediction for reinforced concrete columns based on ensemble machine learning algorithm. *Adv. Eng. Inform.* **2020**, *45*, 101126. [[CrossRef](#)]
49. Dogan, G.; Arslan, M.H.; Baykan, O.K. Determination of damage levels of RC columns with a smart system oriented method. *Bull. Earthquake Eng.* **2020**, *18*, 3223–3245. [[CrossRef](#)]

50. Nasrollahzadeh, K.; Nouhi, E. Fuzzy inference system to formulate compressive strength and ultimate strain of square concrete columns wrapped with fiber-reinforced polymer. *Neural Comput. Appl.* **2018**, *30*, 69–86. [[CrossRef](#)]
51. Naderpour, H.; Nagai, K.; Fakharian, P.; Haji, M. Innovative models for prediction of compressive strength of FRP-confined circular reinforced concrete columns using soft computing methods. *Compos. Struct.* **2019**, *215*, 69–84. [[CrossRef](#)]
52. Bossio, A.; Lignola, G.P.; Fabbrocino, F.; Monetta, T.; Prota, A.; Bellucci, F.; Manfredi, G. Nondestructive assessment of corrosion of reinforcing bars through surface concrete cracks. *Struct. Concr.* **2017**, *18*, 104–117. [[CrossRef](#)]
53. Bossio, A.; Fabbrocino, F.; Monetta, T.; Lignola, G.P.; Prota, A.; Manfredi, G.; Bellucci, F. Corrosion effects on seismic capacity of reinforced concrete structures. *Corros. Rev.* **2019**, *37*, 45–56. [[CrossRef](#)]
54. Fabbrocino, F.; Funari, M.F.; Greco, F.; Lonetti, P.; Luciano, R.; Penna, R. Dynamic crack growth based on moving mesh method. *Compos. Part B Eng.* **2019**, *174*, 107053. [[CrossRef](#)]
55. Kayabekir, A.E.; Bekdaş, G.; Yücel, M.; Nigdeli, S.M.; Geem, Z.W. Harmony Search Algorithm for Structural Engineering Problems. In *Nature-Inspired Metaheuristic Algorithms for Engineering Optimization Applications*; Springer Tracts in Nature-Inspired Computing, Carbas, S., Toktas, A., Ustun, D., Eds.; Springer: Singapore, 2021.
56. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
57. Feng, D.C.; Wang, W.J.; Mangalathu, S.; Taciroglu, E. Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. *J. Struct. Eng.* **2021**, *147*, 04021173. [[CrossRef](#)]
58. Rahman, J.; Ahmed, K.S.; Khan, N.I.; Islam, K.; Mangalathu, S. Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach. *Eng. Struct.* **2021**, *233*, 111743. [[CrossRef](#)]
59. Degtyarev, V.V.; Naser, M.Z. Boosting machines for predicting shear strength of CFS channels with staggered web perforations. *Structures* **2021**, *34*, 3391–3403. [[CrossRef](#)]
60. Dorogush, A.V.; Ershov, V.; Gulin, A. Catboost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363.
61. Lee, S.; Vo, T.P.; Thai, H.T.; Lee, J.; Patel, V. Strength prediction of concrete-filled steel tubular columns using Categorical Gradient Boosting algorithm. *Eng. Struct.* **2021**, *238*, 112109. [[CrossRef](#)]
62. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
63. Nguyen, Q.H.; Ly, H.B.; Ho, L.S.; Al-Ansari, N.; Le, H.V.; Tran, V.Q.; Prakash, I.; Pham, B.T. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Math. Probl. Eng.* **2021**, *2021*, 4832864. [[CrossRef](#)]
64. Montgomery, D.C. *Design and Analysis of Experiments*; John Wiley & Sons: Hoboken, NJ, USA, 2017.