

Article

Absorption and Utilization of Pollutants in Water: A Novel Model for Predicting the Carrying Capacity and Sustainability of Buildings

Enyang Mei ^{1,2,*}  and Kunyang Yu ³¹ School of Bioengineering, Huainan Normal University, Huainan 232038, China² Key Laboratory of Bioresource and Environmental Biotechnology of Anhui Higher Education Institutes, Huainan Normal University, Huainan 232038, China³ School of Civil Engineering, Harbin Institute of Technology, Harbin 150090, China; yukunyang@hit.edu.cn

* Correspondence: meienyang17@mails.ucas.ac.cn

Abstract: The combination of water management and urban planning can promote the sustainable development of cities, which can be achieved through buildings' absorption and utilization of pollutants in water. Sulfate ions are one of the important pollutants in water, and concrete is an important building material. The absorption of sulfate ions by concrete can change buildings' bearing capacity and sustainability. Nevertheless, given the complex and heterogeneous nature of concrete and a series of chemical and physical reactions, there is currently no efficient and accurate method for predicting mechanical performance. This work presents a deep learning model for establishing the relationship between a water environment and concrete performance. The model is constructed using an experimental database consisting of 1328 records gathered from the literature. The utmost essential parameters influencing the compressive strength of concrete under a sulfate attack such as the water-to-binder ratio, the sulfate concentration and type, the admixture type and percentage, and the service age are contemplated as input factors in the modeling process. The results of using several loss functions all approach 0, and the error between the actual value and the predicted value is small. Moreover, the results also demonstrate that the method performed better for predicting the performance of concrete under water pollutant attacks compared to seven basic machine learning algorithms. The method can serve as a reference for the integration of urban building planning and water management.



Citation: Mei, E.; Yu, K. Absorption and Utilization of Pollutants in Water: A Novel Model for Predicting the Carrying Capacity and Sustainability of Buildings. *Water* **2023**, *15*, 3152. <https://doi.org/10.3390/w15173152>

Academic Editors: Laura Bulgariu and Shunwen Bai

Received: 2 August 2023

Revised: 24 August 2023

Accepted: 1 September 2023

Published: 3 September 2023



Copyright: © 2023 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/>).

Keywords: water management; sustainable development; deep learning; mechanical performance; concrete

1. Introduction

The uneven distribution of urban water resources is not coordinated with land resources, and there are many areas where groundwater levels drop and rivers dry up, which may further lead to a shortage of national water resources and have extremely adverse effects on social development [1]. Therefore, there is an urgent need to address the issues of water resource waste and pollution at this stage, and to strengthen water resource protection. The protection of water resources involves both source control and the absorption and treatment of water pollutants. There are many ways to treat pollutants in water, and building absorption and utilization methods is one of the solutions. However, there are many studies on the removal of pollutants from water through plant absorption and other methods [2], with few studies on building absorption. Buildings are important components of a city, and if they can effectively absorb pollutants from water, they are of great significance for the sustainable development of the city.

A concrete structure is the most used form of infrastructure and civil buildings. It is beneficial for its durability, very high strength, low cost, and easy availability [3–10]. Concrete can absorb wastewater, and the absorption of various ions in wastewater has a certain

impact on the mechanical properties and durability of materials. The degree of influence is determined by factors such as the type and total amount of ions [11]. Concrete absorbs important pollutants in water, such as sulfate ions, which can cause complex physical and chemical reactions, thereby affecting bearing capacity and sustainability. Sulfate ions form soluble gypsum through chemical and physical interactions with hardened cement paste, and then generate ettringite [7]. These corrosion products will fill the pores of the concrete, thereby enhancing the bearing capacity and sustainability of concrete in the early stages. Due to the lengthy process of making concrete that has lasted for decades, the gradually generated expansion stress in the later stage of the concrete making process will also cause a decrease in strength, leading to building problems. Therefore, the performance of this process is predicted to understand whether the absorption of water pollutants has a positive or negative impact on buildings [12,13]. A successful prediction can avoid absorbing pollutants in water while reducing the bearing capacity in architectural design, thereby achieving the absorption and reuse of water pollutants. The final change rate of compressive strength (CSR) of concrete material is one of the important points of the carrying capacity and sustainability of buildings [14–16]. Traditional methods usually study concrete strength data through a large number of laboratory tests, and some laboratory data and engineering detection data have been accumulated [17–22]. At present, the current research mostly used the mathematical method of statistical regression to obtain the performance change curve by fitting the experimental data [8,23–25]. The functional relationships used in the fitting were different, and usually only one was obtained. Nevertheless, the achievements of these regression methods are chiefly the size, quality, and reliability of the used experimental data, as well as the chosen function [26]. Milad et al. [27] summarized the processes and drawbacks of the basic mathematical regression formula and expert experience method. These methods are universal and can be used in various disciplines. Nevertheless, it is obvious that the complex nonlinear problem predicting the CSR of concrete under sulfate corrosion does not apply to these methods, because it cannot be expressed by one or several formulas due to many factors affecting the CSR. Because there are some defects in empirical methods, it seems that more advanced tools for analyzing complex nonlinear problems should be considered, which can reduce the number of laboratory tests and overcome the defects of traditional methods. This new method can better predict CSR at the same time.

To these aims, the use of progressive methods is needed to establish evaluation relationships. Recently, researchers used machine learning methods to investigate the compressive strength of concrete, e.g., artificial neural networks (ANNs) [28], ensemble learning [29], adaptive neuro-fuzzy inference systems and gene expression programming [30], and gray correlation analysis [31]. There is a common feature in these studies, which is the use of multiple machine learning models integrated together so as to prevent the limitations of evaluation results, such as low accuracy and generalization. Nevertheless, the method is only effective for specific problems and is not conducive to engineering popularization. Obviously, the CSR of a concrete structure under sulfate corrosion in a real service environment is worth studying. Sahoo et al. [32] studied a double-hidden-layer back propagation ANN to estimate the compressive strength of concrete against sulfate corrosion. The input parameters included sulfate exposure periods, water curing days, and different ratios of fly ash replacement to binder. And 162 cubes were applied to train the ANN algorithm employing EXCEL of MS OFFICE. They only studied fly ash concrete, the corresponding scope of application was narrow, and the data and software were limited. Admixtures are indeed helpful to improve the sulfate resistance of concrete, but this is only one of the factors affecting the CSR. The factors should be studied together and should not be separated. Therefore, deep learning (DL) methods that can solve this problem appear. The basis of DL is a series of new structures and methods evolved in order to make an ANN with more layers that can be trained and run. DL has a priori knowledge about the physical world that cannot be obtained from the rigor of mathematics compared to ordinary ANN [33]. It has more advantages in practical engineering problems, because like biological learning, it is used to find the ‘physical relationship’ that increases entropy. The

learning goal is not simply to reduce the loss function. In addition, a shallow ANN can simulate any function, but the cost of the data volume is unacceptable. The amount of data in the field of concrete engineering is not as much as commercial data. Fortunately, DL solves this problem and can learn better fitting with less data compared with shallow ANN. Moreover, the TensorFlow (TF) method has been used for DL and has achieved significant performance and power-saving improvements.

With the consideration the above points, this paper implements a DL algorithm based on the TF framework to estimate the CSR of concrete under sulfate corrosion, which is necessary for the carrying capacity and sustainability of buildings. The aim of this paper is to construct a detailed configured DL model that can evaluate the CSR of concrete and ensure an ample applicability. The capability of this model is investigated, and the predicted results are visualized. Moreover, to verify the performance of the model, seven typical machine learning approaches will be applied to simulate the experiment, and the results will be compared with the DL model. The rest of this paper is formulated as follows. Section 1 introduces the comprehensive structure of the proposed DL model and introduces the detailed approach. Section 2 gives all the dataset generation and the correlation analysis processes. Section 3 shows the results of the testing and training, discusses the development, performance, and interpretation of the model, and compares the proposed method with seven basic machine learning algorithms. Finally, the conclusions and future works are discussed in Section 4.

2. Data and Methods

2.1. Methodologies

TensorFlow (TF) is an open-source and commutative program software library for machine learning (ML), proposed by the Google Brain Team, and it can be applied to compute information to evaluate the accuracy output. Version 2.0 was released in March 2019, and it was used in this study. As the name TF implies, operations are performed by DL on multidimensional data arrays. TF offers various popularly applied and powerful DL algorithms. TF applies dataflow graphs to represent the shared state, computation, and the operations that mutate that state [34].

ANN is the basis of DL and is scalable, powerful, and versatile. DL is a subfield of ML that is a series of algorithms. At present, most ANNs are shallow structure algorithms, and the representation ability of complex functions is limited in the case of limited samples and calculation units. Therefore, for complex problems, the generalization ability is restricted to some extent. DL can realize complex function approximation by learning a deep nonlinear network structure, it can characterize the distributed representation of input data, and shows a strong ability to learn data and essential characteristics from a small number of samples. The advantage of a multilayer is that complex functions can be represented with fewer parameters. The essence of DL is to learn more useful features by building ML models with many hidden layers and massive training data so as to finally improve the accuracy of prediction. Therefore, a depth model is the means and feature learning is the purpose. Different from traditional shallow learning, DL is characterized by the following: (1) it emphasizes the depth of the model structure, which usually exceeds the hidden layer of two layers, and (2) it clearly highlights the importance of feature learning. Through layer-by-layer feature transformation, the feature representation of the sample in the original space is transformed into a new feature space, which makes the prediction easier. Compared with the method of constructing features using artificial rules, using big data to learn features can better describe the rich internal information of the data. The connection between neurons is shown by the straight line. DL is very similar to the ANN. In short, an ANN with two or more hidden layers is called a DL model. Combined with the TF framework, the model is mainly composed of the following parts:

Layers: They contain input, hidden, and output layers. The full connection layer was used in this paper. Each node was connected with all nodes of the previous layer, which was used to synthesize the previously extracted features.

Loss function: It is used to define the error between a single training sample and the real value. In the process of training, the weight needs to be constantly adjusted. Its purpose is to obtain a set of final weights to make the output of the input characteristic data reach the expected value.

Optimizer: The principal function of the optimizer is to guide the parameters of the loss function to update the appropriate size in the correct direction in the process of back propagation, so that the updated value continues to approach the global minimum.

Activation function: In the model, the input features act on the activation function after a series of weighted summation. Similar to the neuron-based model in the human brain, the activation function ultimately determines whether to transmit the signal and what to transmit to the next neuron. However, it is not the function with excellent characteristics that is the most suitable. Each function needs to be tested continuously in the model so that the most suitable function is selected. The swish function proved to be the most suitable activation function for the model after many experiments, which has several characteristics: (1) it is unbounded above (avoids saturation), (2) bounded below (strong regularization, especially for large negative numbers), (3) and smooth (accessible to learn and less sensitive to it). The swish function is represented as

$$f(x) = x \frac{1}{1 + e^{-x}} \quad (1)$$

Overfitting: The model performs too well in the training set, resulting in poor performance in the validation data and testing data, that is, the generalization error is relatively large. From the perspective of variance and deviation, overfitting means high variance and low deviation on the training set. Overfitting is usually caused by an amount of data that is too small, inconsistent data distribution between the training set and validation set, model complexity that is too large, poor data quality, overtraining, and so on. In order to reduce the complexity of the model, the regularization method was used in this paper. There are two regularization methods, $L1$ and $L2$:

$$L1 = \frac{\lambda}{n} \sum_w |w| \quad (2)$$

$$L2 = \frac{\lambda}{2n} \sum_w w^2 \quad (3)$$

where λ represents the penalty number (regularization rate), and w presents weight value. $L2$ has some advantages: a unique solution, more convenient calculation, better effect of preventing overfitting, and making the optimizer solver more stable and faster. Therefore, $L2$ was chosen for this model.

If there are too few samples, the iterative calculation can be carried out in the form of full-batch learning, which has two advantages: (1) the direction decided by the complete data can represent the overall samples well; (2) it is difficult to choose a global learning rate due to the various gradient values of various weights. However, the above two advantages become disadvantages for a slightly larger dataset. It becomes infeasible to load all the data at one time, and the gradient correction values will offset each other because of sampling divergences between batches, and they cannot be corrected when iterated by the Rprop algorithm. Therefore, this model selected the mini-batch learning technique to promote the parallelization efficiency of matrix multiplication and decrease the loss value of the model. In addition, this model selected the exponential decay method to update the learning rate for the convergence speed because this update method is fast, and it selected the exponential moving average method to update the weight because it can make the model become more robust.

$$decayed_learning_rate = learning_rate \times decay_rate^{\frac{global_step}{decay_steps}} \quad (4)$$

$$shadow_variable = decay \times shadow_variable + (1 - decay)variable \quad (5)$$

where the shadow variable is maintained for each variable. The initial value of the shadow variable is the initial value of the corresponding variable, and decay represents the exponential moving average decay rate.

2.2. Data and Preparation

The preparation of data is very important for the establishment of the model. The quality/quantity and division of dataset have a great impact on the results. The data of this study were from 148 sources in the literature, with a total of 1328 groups of data (<https://www.zotero.org/groups/5162364/water-2552672-data/library>). Obviously, the data sources were relatively extensive. Even if individual data results were inappropriate, they would not affect the final prediction results. This was of great benefit to the generalization of the model because the ultimate purpose of this study is to use the model in engineering practice, and the data in actual engineering cannot be as accurate as that in the laboratory. According to the composition of data and expert experience, six input parameters were selected: (1) water-to-binder ratio (W/B), (2) sulfate concentration (SC), (3) sulfate type (ST), (4) admixture type (AT), (5) admixture percentage in relation to the total binder (A/B), and (6) service age (SA). The output training vector's output parameter has a dimension of 1×1 and consists of the final change rate of compressive strength (CSR). Table 1 shows the range, mean and standard deviation (STD) values, and the regression coefficient of the parameters.

Table 1. Statistical summary of datasets.

Input/Output Variables	Symbol	Unit	Category	Statistics			
				Min	Max	Mean	STD
Water-to-binder ratio	W/B	/	Input	0.21	0.95	0.46	0.08
Sulfate concentration	SC	mg/L	Input	0.00	30.00	6.58	3.97
Sulfate type	ST	/	Input	1.00	5.00	1.74	1.03
Admixture type	AT	/	Input	0.00	30.00	3.16	3.89
Admixture percentage in relation to the total binder	A/B	%	Input	0.00	100.00	15.83	20.53
Service age	SA	month	Input	1.00	120.00	12.71	16.99
Final change rate of compressive strength	CSR	%	Output	−100.00	144.08	−6.18	32.18

Among the inputs selected by the model, there may be variables that are redundant to each other. Therefore, Pearson correlation coefficients between input and output parameters were obtained and are displayed in Figure 1. The technique is a measure of the strength of the linear relationship between two parameters. The correlation coefficient cannot precisely and appropriately represent the relationship strength between those two parameters if the relationship between the parameters is not linear. Statistically, the range is $[-1, 1]$, where '1' represents a flawless positive linear relationship between variables, '−1' pinpoints a perfect negative, and '0' indicates no linear relationship. The efficiency of the model will become low when the correlation coefficient between the input variables is a negative of a high positive. This is because they are not independent of each other. Fortunately, the situation mentioned above does not appear in this model. It is shown that the correlation between CSR and the input parameter of the SA is still positive.

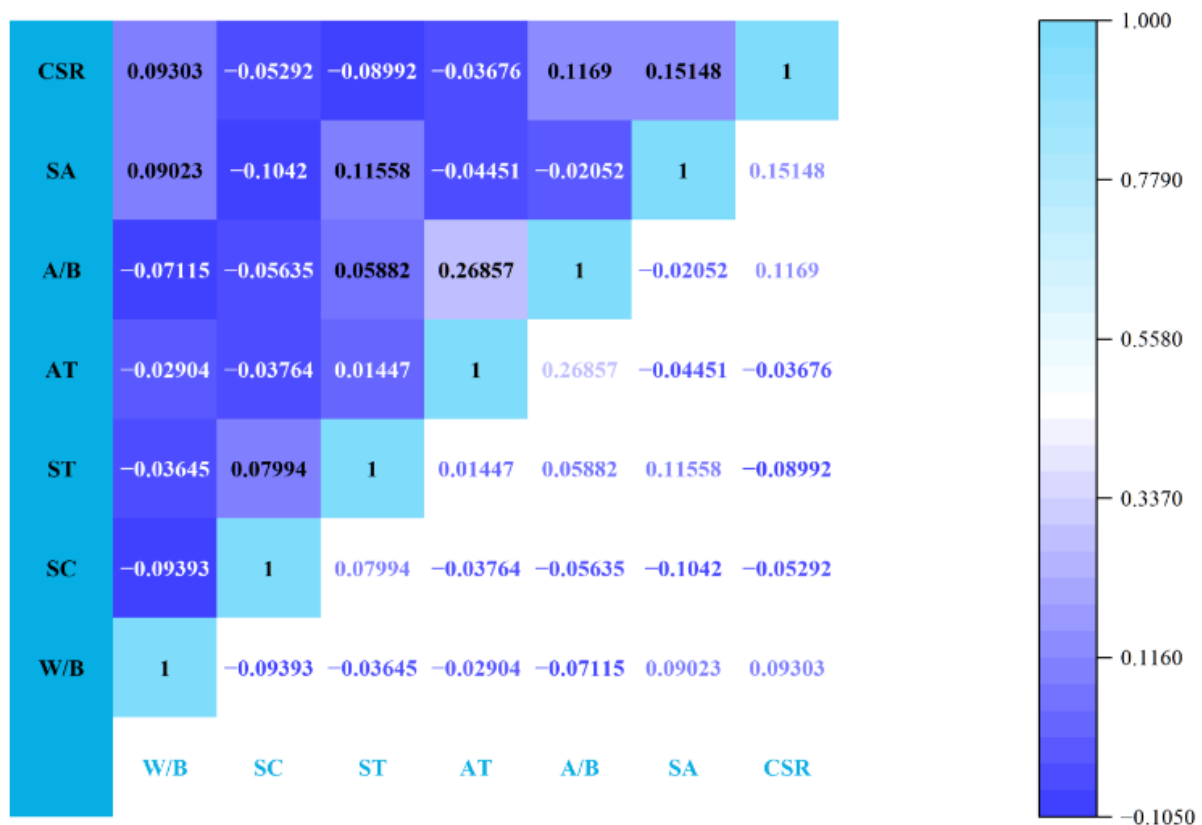
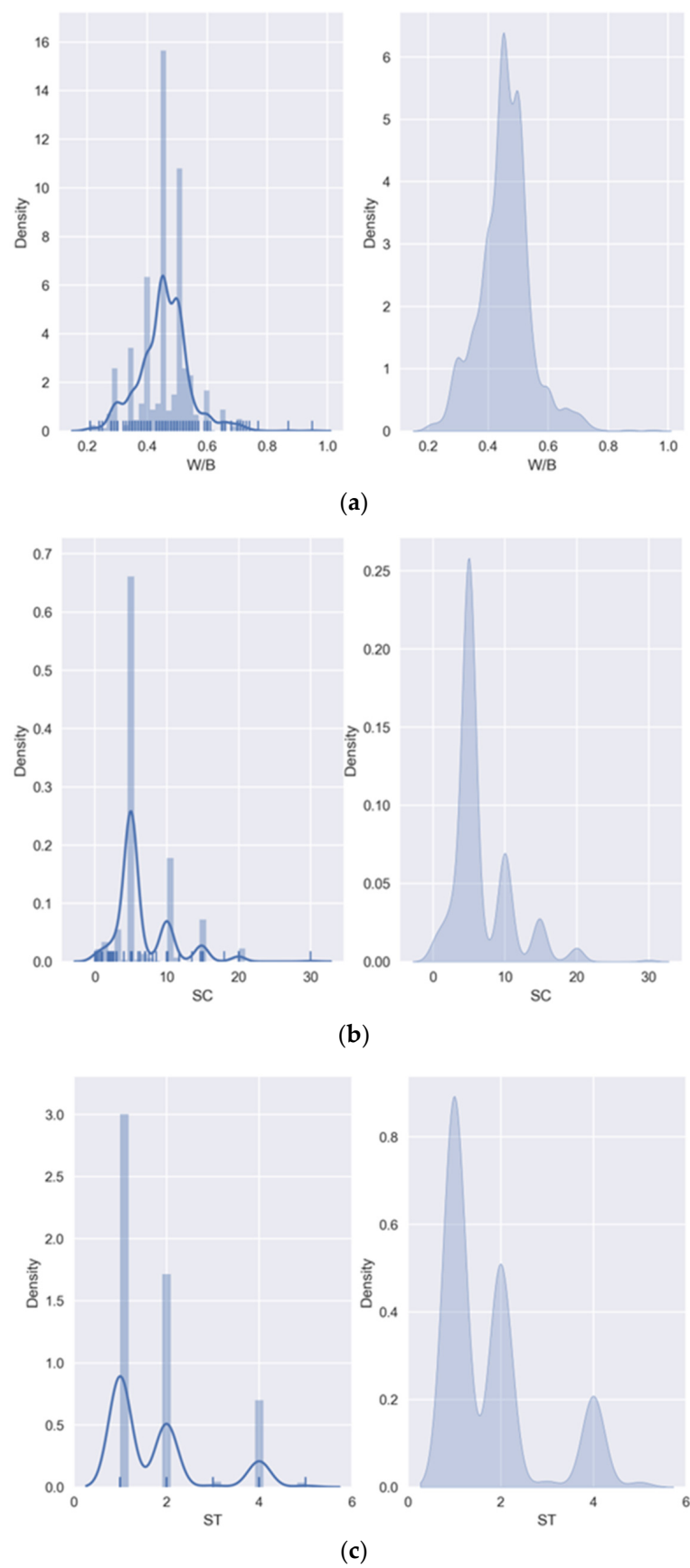


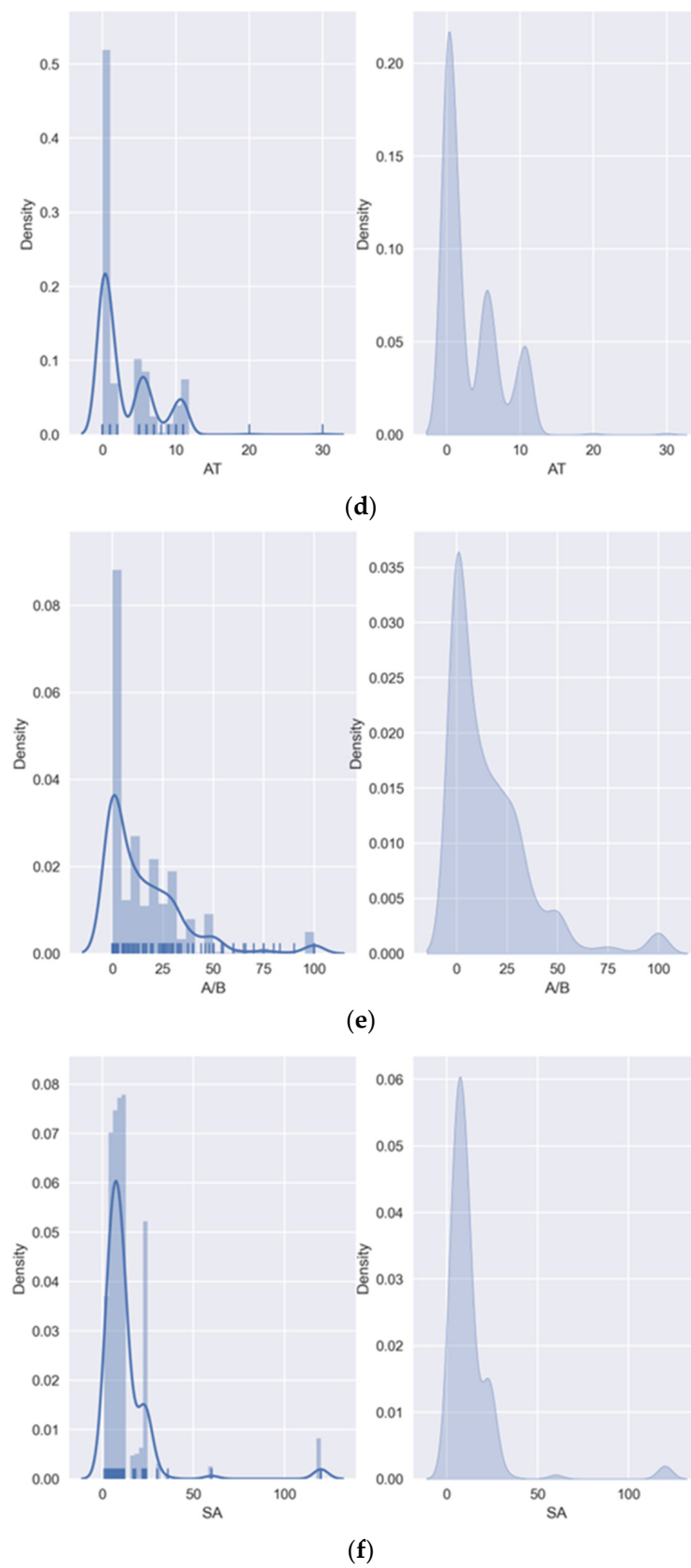
Figure 1. Heat map of the correlation matrix of the input variables.

In addition, the distribution of the variable values should be noted. Solving the distribution density function of random variables with a given sample set is one of the basic problems of probability statistics. The solutions include parametric estimation and nonparametric estimation. Parameter estimation requires basic assumptions, and there is often a large gap between it and the actual physical model. Therefore, the nonparametric estimation method is a good alternative; kernel density estimation (KDE) is one of the better methods. It does not use the prior knowledge of data distribution and does not attach any assumptions. It is a method used to study the characteristics of data distribution from the data sample itself and calculates the distance of data points by weighting, and the results are shown in Figure 2. The theory of weighting the distances of observations (x_i) from a particular point x is as follows:

$$f^{Kernel}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (6)$$

where $K(\cdot)$ is the chosen Kernel (weight function) and h is bandwidth. These figures are exceedingly advantageous, because through them, researchers can understand the distribution scope of the data and whether the amount of data is sufficient. It can be seen from Figure 2 that the amount of data used in this study is sufficient and the distribution is reasonable.

**Figure 2.** *Cont.*

**Figure 2.** *Cont.*

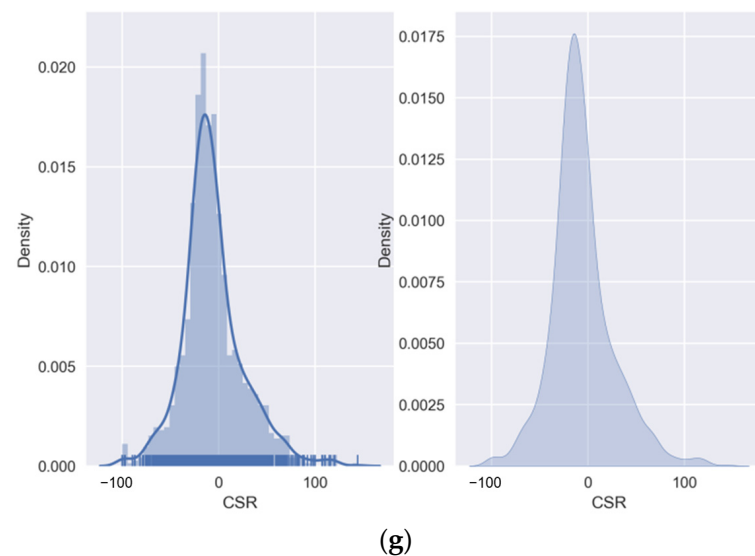


Figure 2. Distribution of several parameters constructed using KDE modeling technique (the right figure of each is Gaussian KDE): (a) W/B; (b) SC; (c) ST; (d) AT; (e) A/B; (f) SA; and (g) CSR.

3. Results and Discussions

The goal of the calculation investigation is to evaluate the CSR by applying 1328 sets of data. To this aim, 70% of all the data (i.e., training dataset of 930 samples) were employed to train the models, while the remaining data (398 samples) were utilized for testing. This indicated that the training effect of the model was satisfactory if the error of the training set was similar to that of the testing set. On the contrary, it indicated that there was an overfitting problem in the model if the error of the testing set was much greater than that of the training set. To evaluate the strength of fit and the robustness of the DL model, five performance parameters (loss functions) are used to assess the predicting achievement. Then, this section explores the performance of the model. Finally, a thorough comparison between the performance of existing analytical ML models and the proposed DL model is investigated. Figure 3 presents, in detail, the flow chart of the DL model based on the TF framework used in this paper. In what follows, the model performance indicators, development, and results are shown and discussed comprehensively.

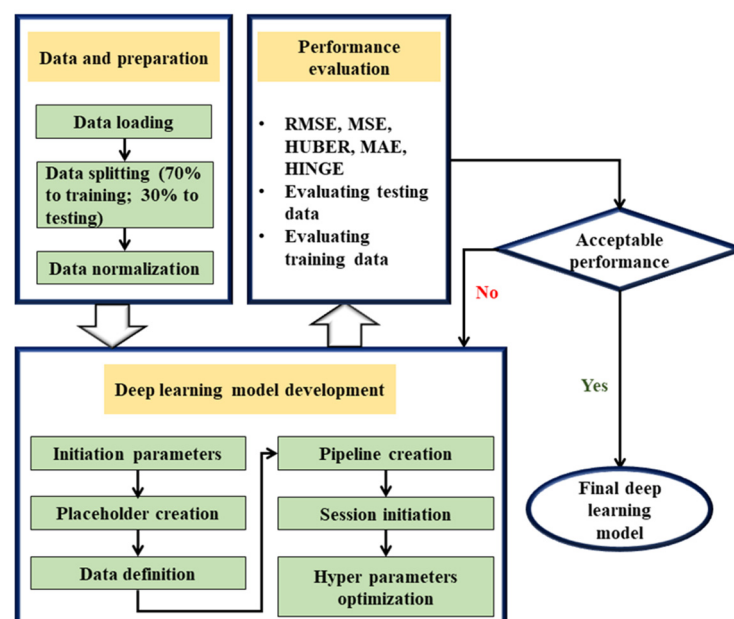


Figure 3. Flowchart of the deep learning model.

3.1. Model Performance Indicators

Metric is applied to assess the performance of the model. It mainly includes the root mean square error (RMSE), mean absolute error (MAE), mean square error (MSE), HUBER, and HINGE. The following are the principle, the scope of application, and the limitations of each loss function.

- The MAE measures the average error amplitude, the distance between the predicted value \hat{y}_i , and the real value y_i , and the action range is 0 to positive infinity. It is more robust to outliers on the line. However, the derivative at point 0 is discontinuous, which makes the solution efficiency lower and the convergence speed slower. For smaller loss values, the gradient is as large as that of other interval loss values.
- The MSE measures the square sum of the distance between the predicted value and the real value, and the scope of action is consistent with the MAE. There is something about the fast convergence speed. It can give an appropriate penalty weight to the gradient instead of the same, so that the direction of the gradient update can be more accurate. The disadvantage is that it is very sensitive to outliers, and the direction of the gradient update is easily dominated by outliers, so it is not robust.
- The HUBER function combines the MAE and MSE and takes their advantages. The principle is to use the MSE when the error is close to 0 and to use the MAE when the error is large.
- The HINGE function cannot be optimized using the gradient descent method, but needs a sub-gradient descent method. It is a proxy function based on a 0–1 loss function.
- The RMSE is the square root of the ratio of the square sum of the deviation between the observed value and the real value to the number of observations, which is more sensitive to outliers.
- R^2 is defined as using the mean as the error benchmark to evaluate whether the prediction error is greater than or less than the mean benchmark error. Generally, it is more effective in a linear model. It cannot fully reflect the prediction ability of the model, and cannot represent the good generalization performance of the model. The specific equations are as follows:

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \quad (7)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (8)$$

$$HUBER = \begin{cases} \frac{1}{2}(\hat{y}_i - y_i)^2 & \text{for } |\hat{y}_i - y_i| \leq \delta \\ \delta|\hat{y}_i - y_i| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases} \quad (9)$$

$$HINGE = \max(0, 1 - t \times \hat{y}) \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (11)$$

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \quad (12)$$

where δ is a constant, n represents the observed value, and \bar{y} is the average output of the data.

3.2. Model Developments

The performance of the DL model largely depends on the number of layers, the neurons in each layer, the activation function, and the hyper parameters. The optimal selection of these parameters was obtained through countless experiments. In this process,

the control variable method was used for the initial parameters, and one of the parameters was modified when the other parameters remained unchanged. Because DL is usually applicable to big data, the amount of data in this study was not enough to support the excessive number of hidden layers. Therefore, it was finally determined that four hidden layers were the most appropriate number combined with the tests. The number of hidden layers and neurons in each layer can be obtained simultaneously through tests. On this basis, different activation function effects were tested. The same method was used to determine the hyper parameters. The batch size is usually set as a multiple of two, and this model was tried according to this law. The learning rate base was combined with its decay rate, and was tested from 0.001, and gradually enlarged. The regularization rate and exponential moving average decay rate are usually set 0.001 and 0.99, respectively. And the model had excellent results after the tests, so they were not modified. The number of training steps depended on the specific loss function attenuation and the use of a GPU. Figure 4 and Table 2 presents the activation functions and neuron number of every layer of the DL model. Table 3 shows the optimal hyper parameters of the DL model in this study.

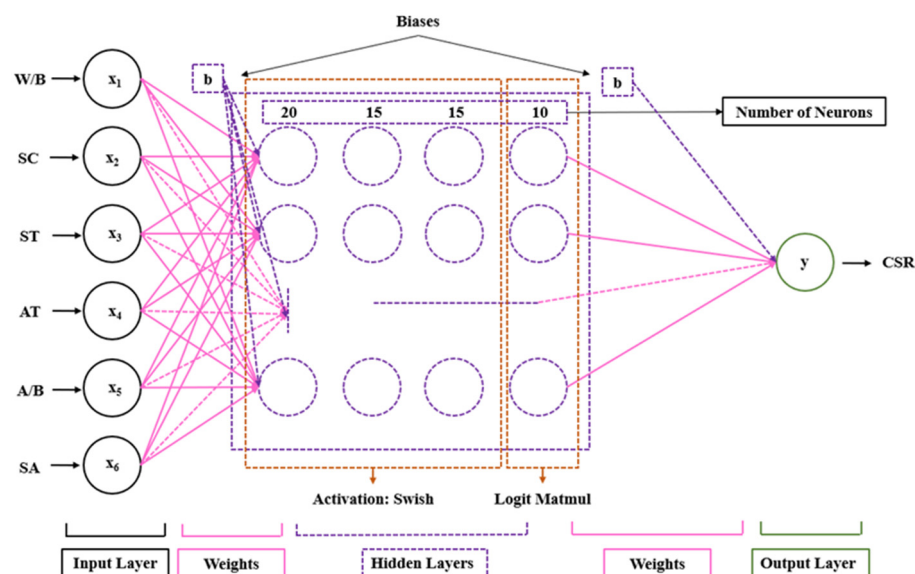


Figure 4. Schematic structure of deep learning model.

Table 2. Activation functions and neuron number of every layer of deep learning model.

Layer	Neuron Number	Activation Function
Input	6	/
Hidden 1	20	Swish
Hidden 2	15	Swish
Hidden 3	15	Swish
Hidden 4	10	/
Output	1	/

Table 3. Hyper parameters of deep learning model.

Hyper Parameters	Batch Size	Learning Rate Base	Learning Rate Decay Rate	Regularization Rate	Training Steps	Exponential Moving Average Decay Rate
Value	16	0.01	0.99	0.001	300,000	0.99

The choice of the optimizer was determined by the decline in the loss value. Figure 5 contrasts the performance of various optimizers, including Adam, AdaGrad, Ftrl, and ProximalAdaGrad. It can be seen that the Adam method performs best in the loss descent

curve, both in speed and convergence accuracy. This is because it is computationally efficient, requires less memory, and the update of parameters is not affected by the scaling transformation of the gradient. And the hyper parameters are well interpretable, and usually do not need to be adjusted or only require a little fine-tuning, which is very important for engineering applications. For the problem of CSR prediction, it is suitable for a sparse gradient and a large noise, so the effect is so significant.

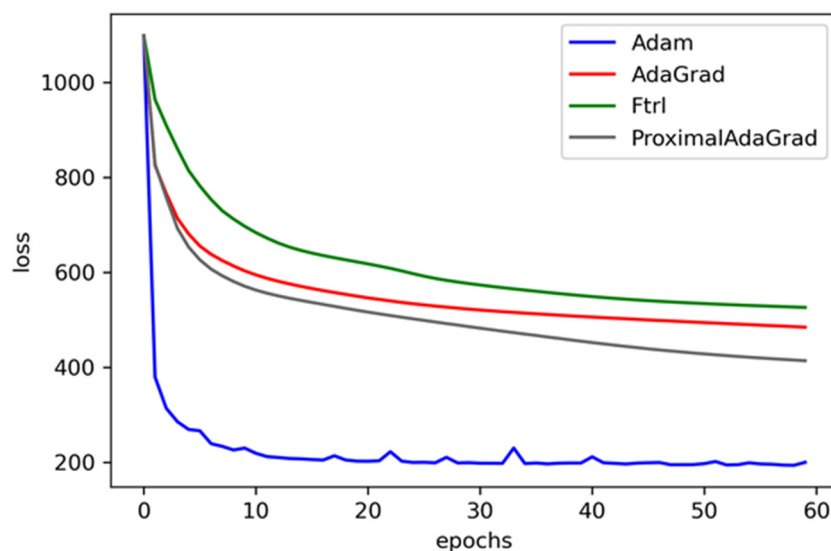


Figure 5. Performance of different optimization methods.

3.3. Evaluation of Model Performance

The performance of the DL model was assessed based on the MSE, MAE, HUBER, HINGE, and RMSE while considering the testing phase and training phase. Evaluating the performance of the DL model from various perspectives is the reason for choosing these indexes. For comprehensive discussion, the evaluation results of the DL model are shown in Figure 6. Among these figures, Figure 6a shows the performance of five loss functions on the training phase. It can be seen that MSE can converge to the minimum value the fastest; the minimum results of the MAE and HUBER are similar, but the decline process fluctuates and is unstable; the result of the RMSE is relatively stable, while the value of HINGE is the largest and the most unstable. These functions jointly show the performance of the model. The result of the minimum value and whether it fluctuates in the decline process are affected by its calculation process. It is inappropriate to specify a loss function alone, which depends on the trend of several combinations. Among them, the value of the HINGE function (Figure 6f) rises first and then fluctuates, and there is no obvious downward trend. The reason for this abnormal phenomenon is that it is more suitable for the classification prediction model, and it fails in this regression model. Among the other loss graphs, only the testing phase of the MSE (Figure 6b) has different coordinate axes. This is because the MSE is more sensitive to outliers. Therefore, the loss value is large in the process of initial learning iteration, but the error of the final training phase gradually tends to 0, and the decline speed is very fast. It is necessary to compare the results of the testing and training phases because the overfitting problem may still occur even if the regularization method is used. On the whole, the prediction results of the training and testing phases are very good, and it proves that the model has not over-fitted because the loss value of the training phase does not far exceed that of the testing phase. On the contrary, in Figure 6c–e, the loss value of the training phase is even greater than that of the test set, which is not a problem with the model, but shows that the amount of data in the training phase is large, including some abnormal points, resulting in the error being greater than that of the testing phase. This is a normal phenomenon in data science prediction.

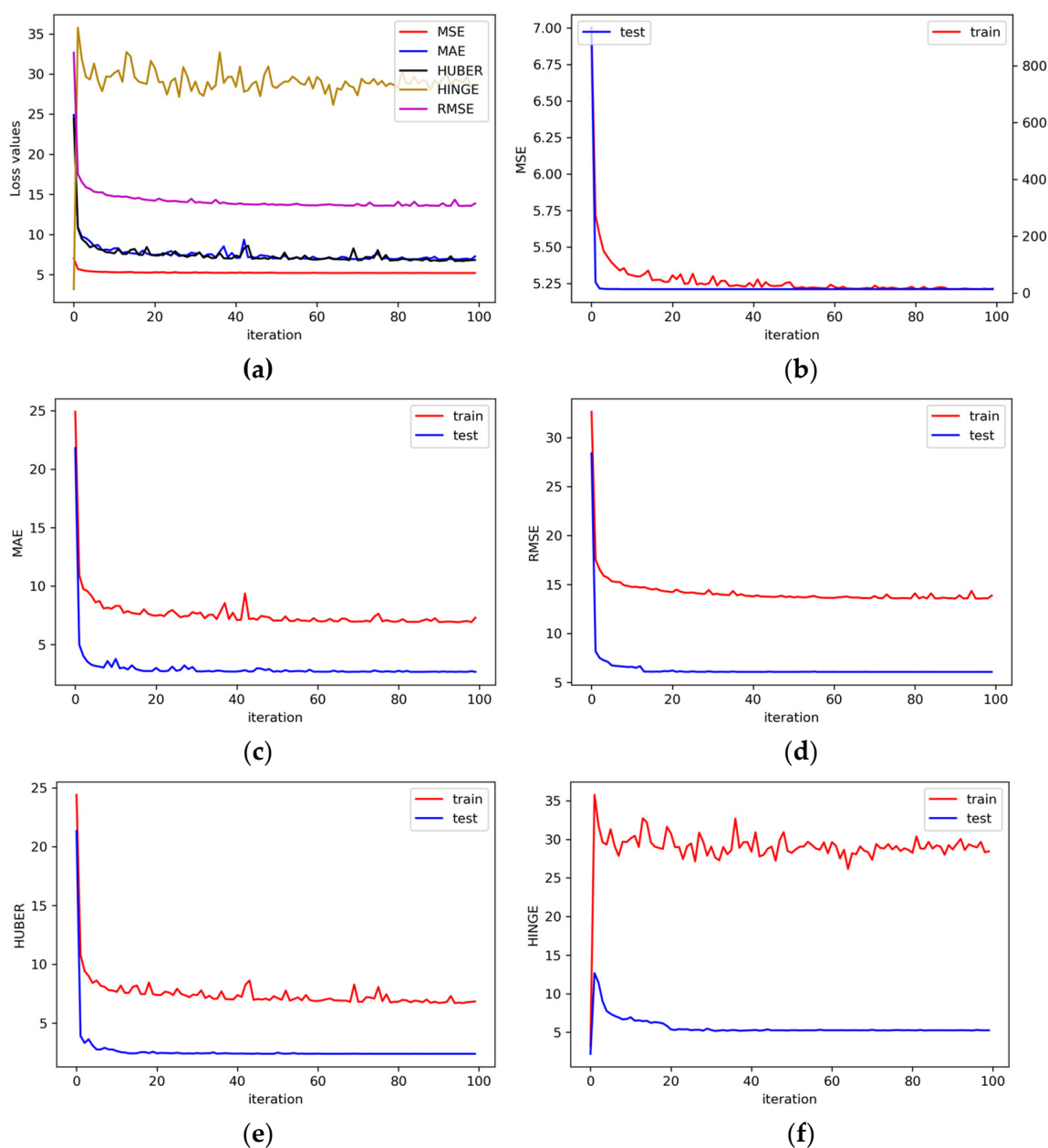


Figure 6. Loss curves for the dataset: (a) for five functions of training phase; (b) for MSE; (c) for MAE; (d) for RMSE; (e) for HUBER; and (f) for HINGE training and testing phases, respectively.

3.4. Visual Interpretation of Results

Table 4 presents the validation set that consists of 48 groups of randomly selected data from the total dataset. Intuitively, it can be seen from the table that most CSR values are negative, and the prediction in the negative range is relatively closer to the actual value. Individual results have large deviations, such as Numbers 19 and 38; the error comes from the fact that the value of W/B is too small for the Number 38, and the correlation between W/B and CSR is large. Most of the values of W/B in the data are concentrated between 0.4 and 0.6; the error of Number 38 comes from the fact that the other input values are similar, but the CSR is negative. Such a large error result is also the reason why the loss value does not fall to zero and it is a normal situation. It is impossible for any model to obtain accurate results for each set of data. Figure 7 shows the relationship (dispersion (a) and value (b)) between the predicted and actual values through the validation set. The intuitive comparison cannot illustrate all aspects of the results, and the error band is

introduced to explain the results. It can be seen from the DL model that the CSR produces a pretty close estimation with low dispersion and small measurement errors.

Table 4. Validation set and prediction results of CSR.

Numbers/Variables	W/B	SC	ST	AT	A/B	SA	CSR	Prediction
1	0.6	5.00	4	0	0.00	5	−72.34	−27.60
2	0.6	5.00	4	0	0.00	5	−62.16	−27.60
3	0.50	5.00	1	9	29.00	12	−44.23	−20.49
4	0.54	5.00	1	1	20.00	5	−38.11	−19.86
5	0.4	5.00	2	5	10.00	23	−28.00	−24.70
6	0.54	5.00	1	1	20.00	5	−25.70	−19.86
7	0.50	3.00	2	8	0.05	9	−22.24	−21.09
8	0.50	5.00	4	6	10.00	24	−22.00	−27.52
9	0.50	0.50	1	0	0.00	4	−21.00	−12.45
10	0.54	5.00	1	1	20.00	5	−20.95	−19.86
11	0.45	10.00	1	0	0.00	120	−18.87	−22.19
12	0.46	5.00	1	6	40.00	4	−17.50	−24.05
13	0.45	3.00	2	2	100.00	3	−16.56	−14.83
14	0.54	5.00	1	1	20.00	5	−16.34	−19.86
15	0.55	10.00	1	1	20.00	12	−16.00	−15.50
16	0.46	5.00	1	6	40.00	10	−15.00	−24.77
17	0.50	5.00	4	6	10.00	24	−14.00	−27.52
18	0.45	10.00	1	1	20.00	12	−13.00	−13.39
19	0.28	6.00	1	0	0.00	30	−10.30	20.33
20	0.45	3.00	2	2	50.00	3	−9.88	−14.88
21	0.45	10.00	1	1	20.00	12	−9.70	−13.39
22	0.45	5.04	1	0	0.00	120	−9.18	−12.59
23	0.50	0.05	1	0	0.00	4	−7.50	−4.47
24	0.40	5.00	1	0	0.00	2	−7.29	−20.26
25	0.40	5.00	1	0	0.00	2	−7.17	−20.26
26	0.45	1.00	1	1	20.00	12	−7.00	−6.94
27	0.45	5.00	4	0	0.00	12	−6.52	4.51
28	0.50	3.00	2	1	20.00	6	−4.41	−3.66
29	0.50	3.00	2	1	20.00	6	−4.37	−3.66
30	0.50	3.00	2	1	20.00	6	−3.45	−3.66
31	0.50	3.00	2	1	20.00	6	−2.95	−3.66
32	0.45	5.00	4	0	0.00	12	−2.17	4.51
33	0.60	5.00	1	0	0.00	6	−1.27	0.95
34	0.30	5.00	1	0	0.00	6	0.01	2.78
35	0.30	5.00	1	7	8.00	6	1.23	1.26
36	0.50	11.70	4	5	40.00	10	2.00	5.48
37	0.30	5.00	1	5	8.00	6	2.44	0.20
38	0.54	5.00	1	1	20.00	5	6.86	−19.86
39	0.52	10.00	1	1	27.00	5	7.53	18.85
40	0.52	10.00	3	1	27.00	5	7.88	18.14
41	0.54	5.00	1	1	20.00	7	9.76	20.09
42	0.45	5.00	1	7	8.00	6	11.67	2.20
43	0.56	14.70	2	0	0.00	7	12.80	18.49
44	0.40	5.00	4	0	0.00	12	15.25	7.13
45	0.5	5.00	1	7	10.00	18	20.00	11.45
46	0.54	5.00	4	1	30.00	120	20.29	33.19
47	0.40	5.00	4	1	10.00	12	24.59	22.49
48	0.40	5.00	4	1	30.00	12	34.04	26.99

Note: ST: 1—sodium sulfate; 2—magnesium sulfate, 3—ammonium sulfate, 4—mixed solution; 5—sulfuric acid. AT: 0—none; 1—fly ash; 2—granulated blast furnace slag; 3—steel slag powder; 4—phosphorus slag powder; 5—silica fume; 6—multicomponent admixture; 7—metakaolin; 8—fiber; 9—volcanic ash; 10—limestone powder; 11—others, e.g., coal gangue powder, bagasse ash, rice husk ash, and waste glass powder.

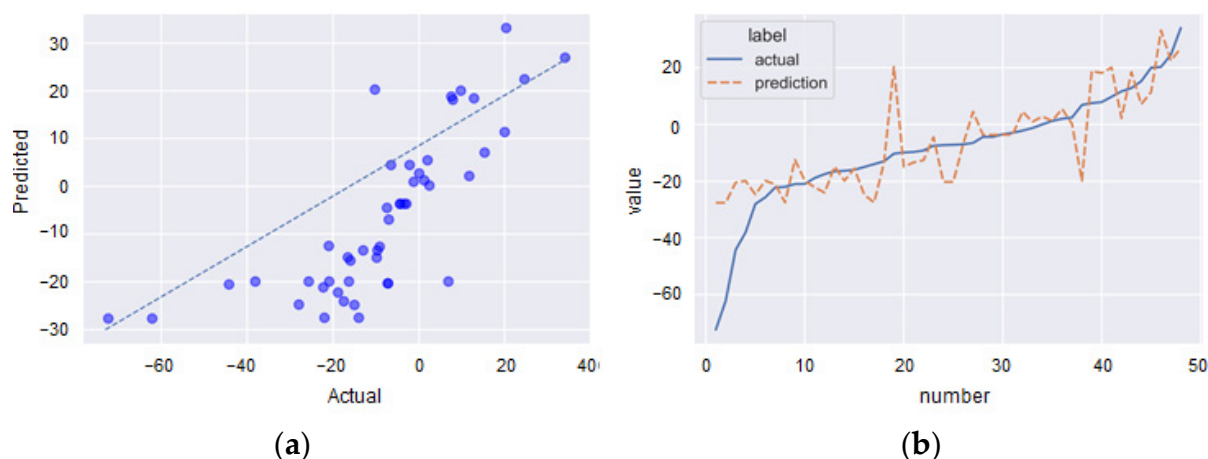


Figure 7. Prediction results of the deep learning model between actual and predicted results of 48 randomly selected tests: (a) dispersion; (b) value.

3.5. Comparison with Other Machine Learning Methods

3.5.1. Regression

In regression analysis, if there are two or more independent variables, it becomes multiple linear regressions (MLPR). The purpose of MLPR is to construct a regression equation and estimate the dependent variable by using multiple independent variables so as to explain and predict the value of the dependent variable. In addition to the most basic MLPR, there are various kinds of regression algorithms. This paper also used classical k-nearest neighbor regression (KNNR). The advantage of the KNNR model is that it can obtain a relatively accurate performance without too many variables, and the speed of constructing the model is usually very fast. However, the effect will become worse for datasets with more 0 values. Therefore, although the KNNR algorithm is easy to understand, it is not suitable for this study because of its slow prediction speed and because it cannot deal with multi-feature datasets.

3.5.2. Ensemble Learning

ML is chiefly divided into supervised learning and unsupervised learning. The learning model is stable in all aspects. Ensemble learning combines multiple weak supervised learning models to obtain a better and more comprehensive strong supervised learning model. Among them, the random forest (RF) algorithm establishes a forest in a random way. The forest is composed of many decision trees (DT), and there is no correlation between each DT. Each DT predicts the input, and the average result is the output of the RF. Each DT belongs to a weak learner and may become strong when combined. Bootstrap aggregation (Bagging) combines logistic regression and DT for calculation. In fact, the RF algorithm is an evolutionary algorithm of bagging. The improvement is mainly aimed at the establishment of DT. DT in bagging selects an optimal feature of all the data features n on the node to divide the right and left sub-trees of the DT. However, RF randomly selects some sample features n_{sub} on the nodes so as to improve the generalization ability. The calculation process of the extremely randomized trees (ET) algorithm is very similar to RF, with only one significant difference. RF obtains the best splitting attribute in a random subset, while ET is completely random, so ET has better consequences than the RF algorithm in some problems.

In addition to bagging methods, boosting is also an important strategy in ensemble learning. Contrary to the parallel idea of the bagging method, boosting trains a weak learner with an initial weight from the training set, and then updates the weight of the sample according to its error rate. This process makes the weight of samples with a high error rate higher, so it is more valued in the second learner. The second learner is trained based on the adjusted weight. This belongs to tandem thought. Extreme gradient boosting

(XGBT), adaptive boosting (AB), and gradient boosting (GB) are the most widely used in the boosting method. GB also combines DT models. Note that GB uses the gradient boosting method for training, while gradient descent is often used as a training method in logistic regression and ANN. The comparison of the two methods is shown in Table 5. It can be noted that the information of the loss function relative to the negative gradient direction of the model can be used to update the current model in each round of iteration. Nevertheless, the model is expressed in parametric form in the gradient descent. Hence, the update of the model is equivalent to the update of the parameters. In gradient boosting, the model does not need parametric representation, but is directly defined in the function space, which greatly expands the types of models that can be used. The idea of the XGBT algorithm is the same as that of GB, but some optimizations are made. The second derivative is used to make the loss function more accurate, and the regular term avoids tree overfitting. The AB algorithm is different from the other two algorithms. AB emphasizes adapting. It continuously modifies the sample weight and adds weak classifiers to the boost. The other two are designed to continuously reduce the residual and establish a new model in the direction of a negative gradient by adding new trees.

Table 5. Comparison of gradient descent and gradient boosting methods.

Algorithm	Training Method	Model Definition Space	Optimization Rule	Optimization Rule
Artificial neural networks, logistic regression	Gradient descent	Parameter space	$\theta_t = \theta_{t-1} + \Delta\theta_t$	$L = \sum_t l(y_t, f(\theta_t))$
Gradient boosting	Gradient boosting	Function space	$f_t(x) = f_{t-1}(x) + \Delta f_t(x)$	$L = \sum_t l(y_t, F(x_t))$

Note: θ_t represents the variable at time t ; $L(\bullet)$ is the loss function; and $F(\bullet)$ and $f(\bullet)$ represent the function of variable, respectively.

3.5.3. Score Analysis

All of the above-mentioned algorithms are essential basic algorithms in supervised learning, and are calculated based on the essence of regression prediction. Therefore, R^2 evaluation is enough to compare its applicability and accuracy in this problem. Table 6 and Figure 8 present the results of the R^2 analysis, from which it can be seen that in terms of the total data, except for the MLPR model, the values were between 0.60 and 0.80. The amount of total data was large, so the prediction result was relatively good. In terms of the training phase and testing phase, only the results of XGBT and RF were greater than or equal to 0.50. The results of the two regression algorithms were not very good, which were not applicable to this problem. From the side, it also reflected that the traditional regression method was not suitable for complex prediction problems. Among ensemble learning, the effect of the basic DT model was the worst, and the effect of bagging was also general. Only RF and XGBT had better effects and were relatively stable. The predicted values for the training phase and testing phase were similar, which indicated that the DL method had a certain generalization capability. However, the overall effect of the above models was not as good as the DL model, even if the DL model did not use the R^2 evaluation method.

Table 6. Details of R^2 analysis.

Phase/Model	KNNR	MLPR	DT	ET	XGBT	RF	AB	GB	Bagging
Total	0.73	0.31	0.61	0.75	0.72	0.70	0.73	0.77	0.69
Train	0.23	0.12	0.35	0.29	0.52	0.50	0.18	0.36	0.42
Test	0.22	0.11	0.32	0.36	0.52	0.51	0.16	0.35	0.45

Because R^2 depends on the variance of output variables, it is a measure of correlation, not accuracy. For these basic regression and ensemble learning algorithms in machine learning, R^2 can be used to compare, but it is not suitable for complex networks such as

the DL model. Moreover, the problem predicted in this paper was more complex. There is not a direct formula to calculate the change of concrete compressive strength caused by sulfate corrosion, and the internal reaction mechanism is not completely clear. And the direct and indirect relationship between the generated products and compressive strength is only qualitative, but its interaction is not clear. In addition, the amount of data (1328) is relatively large compared with the general concrete performance prediction problems, i.e., 144 and 120 samples for flexural and compressive strength [30], 254 samples for failure modes [35], 241 samples for ultimate compressive strength [36], 214 samples for shear capacity [37], and 209 samples for compressive strength [38]. The quality of the data is naturally uneven, and the data with large separation are not eliminated. Because the actual project may also have such large dispersion data, they cannot be deleted due to the reduction in prediction accuracy. On the contrary, when the data quality is not high, it can also predict the change in the relative compressive strength, which is more favorable for practical engineering. In general, the prediction results of the DL model can be adopted by engineering practice. Predicting the compressive strength in advance in the whole life cycle of a concrete structure will be of great benefit to structural design, urban and water management, and sustainability.

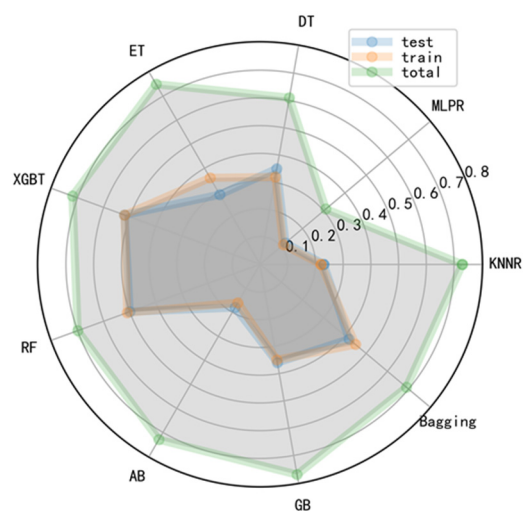


Figure 8. Illustration of score analysis showing in the form of radar diagram.

4. Conclusions

It is pertinent to mention that a reliable and accurate evaluation of the CSR can provide ways to evaluate concrete buildings' absorption of sulfide pollutants in water and can further make urban sustainability and building capacity predictable. The essential purpose of this method was to apply a soft computing method to evaluate the CSR under a sulfate attack. The model adopted optimization algorithms including mini-batch size, exponential moving average decay, and exponential decay of learning rate. It was proven that the proposed DL model can well predict the CSR of concrete caused by sulfate corrosion by analyzing the decline process of five loss functions in the testing phase and the training phase and the error comparison results between the real and the predicted value. It eliminated the need for expensive tests and saved the time cost. In addition, several traditional machine learning models, two regression algorithms, and seven ensemble learning algorithms were not suitable for this problem. The prediction accuracy of these traditional machine learning models in the training phase was not high, and the results of the testing phase were even worse, indicating that the generalization ability of these models was not good as the DL model.

In summary, the DL model is suitable for predicting the degradation performance of building materials after absorbing pollutants. Although this paper only predicts the performance after sulfate ion corrosion, the deep learning model can be extrapolated to other water pollutants when there are sufficient data.

Author Contributions: E.M. performed all of the experiments and wrote the article. K.Y. analyzed the data. All authors read and approved of its content. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors gratefully thank Chengrun Wang from the School of Bioengineering, Huainan Normal University for providing guidance.

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

References

- Bai, S.; Tu, Y.; Sun, H.; Zhang, H.; Yang, S.; Ren, N.-Q. Optimization of wastewater treatment strategies using life cycle assessment from a watershed perspective. *J. Clean. Prod.* **2021**, *312*, 127784. [\[CrossRef\]](#)
- Bai, S.; Chen, J.; Guo, M.; Ren, N.; Zhao, X. Vertical-scale spatial influence of radial oxygen loss on rhizosphere microbial community in constructed wetland. *Environ. Int.* **2023**, *171*, 107690. [\[CrossRef\]](#)
- Gil-Martín, L.M.; Hdz-Gil, L.; Molero, E.; Hernández-Montes, E. The Relationship between Concrete Strength and Classes of Resistance against Corrosion Induced by Carbonation: A Proposal for the Design of Extremely Durable Structures in Accordance with Eurocode 2. *Sustainability* **2023**, *15*, 7976. [\[CrossRef\]](#)
- He, S.; Jiang, Z.; Chen, H.; Chen, Z.; Ding, J.; Deng, H.; Mosallam, A.S. Mechanical Properties, Durability, and Structural Applications of Rubber Concrete: A State-of-the-Art-Review. *Sustainability* **2023**, *15*, 8541. [\[CrossRef\]](#)
- Soltaninejad, S.; Marandi, S.M.; Bp, N. Performance Evaluation of Clay Plastic Concrete of Cement and Epoxy Resin Composite as a Sustainable Construction Material in the Durability Process. *Sustainability* **2023**, *15*, 8987. [\[CrossRef\]](#)
- Robayo-Salazar, R.; Martínez, F.; Vargas, A.; Mejía de Gutiérrez, R. 3D Printing of Hybrid Cements Based on High Contents of Powders from Concrete, Ceramic and Brick Waste Chemically Activated with Sodium Sulphate (Na₂SO₄). *Sustainability* **2023**, *15*, 9900. [\[CrossRef\]](#)
- Wang, H.; Pang, J. Mechanical Properties and Microstructure of Rubber Concrete under Coupling Action of Sulfate Attack and Dry–Wet Cycle. *Sustainability* **2023**, *15*, 9569. [\[CrossRef\]](#)
- Jiang, L.; Niu, D. Study of deterioration of concrete exposed to different types of sulfate solutions under drying-wetting cycles. *Constr. Build. Mater.* **2016**, *117*, 88–98. [\[CrossRef\]](#)
- Yu, K.; Jia, M.; Yang, Y.; Liu, Y. A clean strategy of concrete curing in cold climate: Solar thermal energy storage based on phase change material. *Appl. Energy* **2023**, *331*, 120375. [\[CrossRef\]](#)
- Yu, K.; Liu, Y.; Jia, M.; Wang, C.; Yang, Y. Thermal energy storage cement mortar containing encapsulated hydrated salt/fly ash cenosphere phase change material: Thermo-mechanical properties and energy saving analysis. *J. Energy Storage* **2022**, *51*, 104388. [\[CrossRef\]](#)
- Raza, A.; Shah, S.A.R.; Kazmi, S.N.H.; Ali, R.Q.; Akhtar, H.; Fakhar, S.; Khan, F.N.; Mahmood, A. Performance evaluation of concrete developed using various types of wastewater: A step towards sustainability. *Constr. Build. Mater.* **2020**, *262*, 120608. [\[CrossRef\]](#)
- Arooj, M.F.; Haseeb, F.; Butt, A.I.; Irfan-Ul-Hassan, D.M.; Batool, H.; Kibria, S.; Javed, Z.; Nawaz, H.; Asif, S. A sustainable approach to reuse of treated domestic wastewater in construction incorporating admixtures. *J. Build. Eng.* **2021**, *33*, 101616. [\[CrossRef\]](#)
- Li, S.; Zhang, J.; Li, Z.; Liu, C.; Chen, J. Feasibility study on grouting material prepared from red mud and metallurgical wastewater based on synergistic theory. *J. Hazard. Mater.* **2021**, *407*, 124358. [\[CrossRef\]](#)
- Salami, B.A.; Maslehuddin, M.; Mohammed, I. Mechanical properties and durability characteristics of SCC incorporating crushed limestone powder. *J. Sustain. Cem. -Based Mater.* **2014**, *4*, 176–193. [\[CrossRef\]](#)
- Zhou, Y.; Tian, H.; Sui, L.; Xing, F.; Han, N. Strength Deterioration of Concrete in Sulfate Environment: An Experimental Study and Theoretical Modeling. *Adv. Mater. Sci. Eng.* **2015**, *2015*, 951209. [\[CrossRef\]](#)
- Liu, P.; Chen, Y.; Wang, W.; Yu, Z. Effect of physical and chemical sulfate attack on performance degradation of concrete under different conditions. *Chem. Phys. Lett.* **2020**, *745*, 137254. [\[CrossRef\]](#)
- Akiiz, F.; Tiirker, F.; Koral, S.; Yiizer, N. Effects of Sodium Sulfate Concentration on the Sulfate. *Cem. Concr. Res.* **1995**, *25*, 1360–1368. [\[CrossRef\]](#)
- Kumar, S.; Rao, C.V.S.K. Sulfate attack on concrete in simulated cast-in-situ and precast situations. *Cem. Concr. Res.* **1995**, *25*, 1–8. [\[CrossRef\]](#)
- Kumar, S.; Rao, C.V.S.K. Strength loss in concrete due to varying sulfate exposures. *Cem. Concr. Res.* **1995**, *25*, 57–62. [\[CrossRef\]](#)
- Torii, K.; Taniguchi, K.; Kawamura, M. Sulfate resistance of high fly ash content concrete. *Cem. Concr. Res.* **1995**, *25*, 759–768. [\[CrossRef\]](#)

21. Freidin, C. Stableness of new concrete on the quartz bond in water and sulphate environments. *Cem. Concr. Res.* **1996**, *26*, 1683–1687. [\[CrossRef\]](#)
22. Irassar, E.F.; Maio, A.D.; Batic, O.R. Sulfate attack on concrete with mineral admixtures. *Cem. Concr. Res.* **1996**, *26*, 113–123. [\[CrossRef\]](#)
23. Wei, L.; Xiao-Guang, J.; Zhong-Ya, Z. Triaxial test on concrete material containing accelerators under physical sulphate attack. *Constr. Build. Mater.* **2019**, *206*, 641–654. [\[CrossRef\]](#)
24. Kazmi, S.M.S.; Munir, M.J.; Wu, Y.-F.; Patnaikuni, I.; Zhou, Y.; Xing, F. Effect of different aggregate treatment techniques on the freeze-thaw and sulfate resistance of recycled aggregate concrete. *Cold Reg. Sci. Technol.* **2020**, *178*, 103126. [\[CrossRef\]](#)
25. Cheng, H.; Liu, T.; Zou, D.; Zhou, A. Compressive strength assessment of sulfate-attacked concrete by using sulfate ions distributions. *Constr. Build. Mater.* **2021**, *293*, 123550. [\[CrossRef\]](#)
26. Sadrossadat, E.; Basarir, H.; Karrech, A.; Elchalakani, M. Multi-objective mixture design and optimisation of steel fiber reinforced UHPC using machine learning algorithms and metaheuristics. *Eng. Comput.* **2021**, *38*, 2569–2582. [\[CrossRef\]](#)
27. Milad, A.; Hussein, S.H.; Khekan, A.R.; Rashid, M.; Al-Msari, H.; Tran, T.H. Development of ensemble machine learning approaches for designing fiber-reinforced polymer composite strain prediction model. *Eng. Comput.* **2021**, *38*, 3625–3637. [\[CrossRef\]](#)
28. Işık, M.F.; Avcil, F.; Harirchian, E.; Bülbül, M.A.; Hadzima-Nyarko, M.; Işık, E.; İzol, R.; Radu, D. A Hybrid Artificial Neural Network—Particle Swarm Optimization Algorithm Model for the Determination of Target Displacements in Mid-Rise Regular Reinforced-Concrete Buildings. *Sustainability* **2023**, *15*, 9715. [\[CrossRef\]](#)
29. Liu, Y.; Wang, Y.; Zhou, M.; Huang, J. Improvement of Computational Efficiency and Accuracy by Firefly Algorithm and Random Forest for Compressive Strength Modeling of Recycled Concrete. *Sustainability* **2023**, *15*, 9170. [\[CrossRef\]](#)
30. Shishegaran, A.; Saeedi, M.; Mirvalad, S.; Korayem, A.H. Computational predictions for estimating the performance of flexural and compressive strength of epoxy resin-based artificial stones. *Eng. Comput.* **2022**, *39*, 347–372. [\[CrossRef\]](#)
31. Gong, S.; Bai, L.; Tan, Z.; Xu, L.; Bai, X.; Huang, Z. Mechanical Properties of Polypropylene Fiber Recycled Brick Aggregate Concrete and Its Influencing Factors by Gray Correlation Analysis. *Sustainability* **2023**, *15*, 1135. [\[CrossRef\]](#)
32. Sahoo, S.; Mahapatra, T.R. ANN Modeling to study strength loss of Fly Ash Concrete against Long term Sulphate Attack. *Mater. Today Proc.* **2018**, *5*, 24595–24604. [\[CrossRef\]](#)
33. Orejarena, L.; Fall, M. The use of artificial neural networks to predict the effect of sulphate attack on the strength of cemented paste backfill. *Bull. Eng. Geol. Environ.* **2010**, *69*, 659–670. [\[CrossRef\]](#)
34. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 582–598.
35. Thai, D.-K.; Tu, T.M.; Bui, T.Q.; Bui, T.T. Gradient tree boosting machine learning on predicting the failure modes of the RC panels under impact loads. *Eng. Comput.* **2019**, *37*, 597–608. [\[CrossRef\]](#)
36. Nguyen, M.-S.T.; Trinh, M.-C.; Kim, S.-E. Uncertainty quantification of ultimate compressive strength of CCFST columns using hybrid machine learning model. *Eng. Comput.* **2021**, *38*, 2719–2738. [\[CrossRef\]](#)
37. Prayogo, D.; Cheng, M.-Y.; Wu, Y.-W.; Tran, D.-H. Combining machine learning models via adaptive ensemble weighting for prediction of shear capacity of reinforced-concrete deep beams. *Eng. Comput.* **2019**, *36*, 1135–1153. [\[CrossRef\]](#)
38. Duan, J.; Asteris, P.G.; Nguyen, H.; Bui, X.-N.; Moayed, H. A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model. *Eng. Comput.* **2020**, *37*, 3329–3346. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.