

## *Retraction*

# **Retracted: Prediction of Concrete Compressive Strength Based on the BP Neural Network Optimized by Random Forest and ISSA**

### **Journal of Function Spaces**

Received 12 December 2023; Accepted 12 December 2023; Published 13 December 2023

Copyright © 2023 Journal of Function Spaces. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] G. Chen, D. Zhu, X. Wang, C. Zhou, and X. Chen, "Prediction of Concrete Compressive Strength Based on the BP Neural Network Optimized by Random Forest and ISSA," *Journal of Function Spaces*, vol. 2022, Article ID 8799429, 20 pages, 2022.

## Research Article

# Prediction of Concrete Compressive Strength Based on the BP Neural Network Optimized by Random Forest and ISSA

Gang Chen,<sup>1</sup> Donglin Zhu ,<sup>2</sup> Xiao Wang ,<sup>3</sup> Changjun Zhou ,<sup>4</sup> and Xiangyu Chen<sup>1</sup>

<sup>1</sup>College of Physics and Information Engineering, Fuzhou University, Fuzhou, Fujian 350108, China

<sup>2</sup>School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

<sup>3</sup>Xingzhi College, Zhejiang Normal University, Jinhua, Zhejiang 321000, China

<sup>4</sup>College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua, Zhejiang 321004, China

Correspondence should be addressed to Xiao Wang; [tianzhu213@zjnu.cn](mailto:tianzhu213@zjnu.cn)

Received 16 June 2022; Revised 2 August 2022; Accepted 8 August 2022; Published 27 August 2022

Academic Editor: Miaochao Chen

Copyright © 2022 Gang Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In modern engineering construction, the compressive strength of concrete determines the safety of engineering structure. BP neural network (BPNN) tends to converge to different local minimum points, and the prediction accuracy is not high in the prediction of the compressive strength of concrete. Therefore, a prediction model based on the BPNN optimized by improved sparrow search algorithm (ISSA) and random forest (RF) is proposed to enhance the generalization ability and prediction accuracy of BPNN for compressive strength of concrete. In terms of algorithm improvement, three improvements are proposed for SSA: Latin hypercube sampling is introduced to initialize the location of sparrows and increase the diversity of sparrows; the somersault foraging strategy is used to enrich the optimal position of producers; and combining with the cyclone foraging mechanism, the position updating process of the scroungers is optimized to obtain a better foraging position. In terms of performance evaluation of the algorithm, the ablation experiment verifies that the three improved strategies have improved effects in SSA, and the performance of ISSA on the CEC2017 benchmark function is better than other peers. In terms of predictive index screening, the important features are selected as the input variables of the model by random forest. The prediction results show that compared with the RF-BPNN model and models optimized by other algorithms, RF-ISSA-BPNN model has the lowest prediction error, and the expected value fits the real value better.

## 1. Introduction

Machine learning can mine the inherent relationships from a large number of historical data for classification or prediction. However, in addition to deep learning, random forest, support vector machine, and other methods in machine learning, BP neural network is also gradually applied to the prediction of various engineering fields for its great effect. Liang W et al. [1] study and analyze coal ash deformation temperature with linear regression method and FactSage calculation and introduce BP neural network to obtain accurate prediction results. Xu B et al. [2] use static and dynamic methods to simulate in BP neural network, respectively, which ensure the accuracy of road temperature prediction

in different stages. Liu Y et al. [3] prove that BP neural network is practical and feasible in predicting thermal error of five-axis machining center. Dai S et al. [4] propose a prediction model combining multiple regression and BP neural network, which showed good prediction performance in WFFZ height prediction. Similarly, it can also be applied to predict the output pressure of the sensor [5], performance evaluation of manufacturing collaborative logistics [6], depth of concrete carbonation, and amount of steel corrosion [7].

Because BP neural network has the advantages of self-learning, generalization ability, and fault tolerance, it has been widely used by scholars in many fields. However, as the scope of application becomes wider and wider, many

shortcomings are exposed. The traditional BP neural network is a local search optimization method, and the weight of the network is adjusted along the direction of local optimization, which tends to converge to different local minima, leading to the failure of network training. In addition, it is also susceptible to the influence of initial weights and thresholds, and different weights and thresholds result in different training results. Therefore, many scholars use intelligent optimization algorithm to improve its generalization performance and learning ability by optimizing weights and thresholds on the original basis. For example, Dou K et al. [8], Wang H et al. [9], and Supraja P et al. [10] both adopt genetic algorithm (GA) for optimization, while Yuan H et al. [11], Jiang G et al. [12], and Wang W et al. [13] optimize key parameters of BP neural network through particle swarm optimization (PSO) and mind evolutionary algorithm (MEA) with better optimization performance, and this indicates that the performance of the model with BP neural network as the core and intelligent optimization algorithm as the auxiliary is more outstanding than the original BP neural network. In order to seek a breakthrough in the performance of the predictive model, some scholars have done a lot of work on the improvement of the algorithm. For instance, Wu Y et al. [14] adopt the adaptive learning rate to improve the prediction model, and optimize the model with a new improved algorithm integrating GA and SA. Zhang W et al. [15] improve the convergence factor and position update formula in the standard gray wolf algorithm, and used the improved gray wolf algorithm to find the two optimal values in the prediction model, so that the model can meet the requirements of accuracy and real time of short-term traffic flow prediction. Tian H et al. [16] add nonlinear decline factor to the inertia weight of particle swarm optimization, and the IPSO-BPNN model can effectively predict the yield of winter wheat. Wu L et al. [17] introduce the crossover and mutation operation in GA into the improved fruit fly optimization algorithm (FOA) to establish a corresponding GAIFOA-BP model, and the prediction of fatigue life and fatigue consumption by the model can be closer to the actual results.

The above researches only optimize the prediction model by improving the algorithm, but ignore the importance of feature selection that affects the prediction accuracy [18]. In a sense, the contribution of screening good predictive indicators to the improvement of prediction accuracy may be greater than algorithm optimization and combination of models [19]. RF is selected to measure the importance of each characteristic variable [20–22], and a certain threshold is set for screening to find several characteristic variables highly correlated with the dependent variable and eliminate the characteristic variables with low importance, thus reducing the complexity of the prediction model. Similarly, the improvement of the algorithm has also become an important breakthrough to improve the prediction model. Sparrow search algorithm (SSA) has attracted the attention of scholars in recent years, which divides the search population into three roles: producer, scrounger, and scout. The three roles cooperate with each other to find the optimal

value by their position updating mechanism. SSA has more advantages than GWO, PSO, and GSA in terms of search accuracy, convergence speed and stability [23]. However, there are three problems as follows: (1) Randomly generated initial positions may cause the sparrow population to be unevenly distributed throughout the search solution space; (2) in the stage of updating position of producers, the overall trend decreases as the iteration goes on; and (3) in the stage of updating position of scroungers, the value tends to 0 when the population size is relatively large or the sparrow population converges. In terms of algorithm optimization, Latin hypercube sampling firstly is used to replace the formation mode of the original population to enhance the quality of the initial individuals. Secondly, somersault foraging strategy is introduced in the stage of updating the location of producers to enrich the optimal position of producers and expand its search space. Finally, cyclone foraging mechanism is introduced in the stage of updating the location of scroungers to obtain a better foraging position and enhance the escape ability of the local optimal solution. In this paper, the CEC2017 benchmark function is selected for simulation experiments, the feasibility and rationality of the three strategies for improving the algorithm are verified by ablation experiments, and ISSA is compared with 4 classic optimization algorithms and 2 other improved algorithms. The comprehensive ranking of the simulation results shows that the optimization ability of ISSA is superior to other six algorithms. In the prediction of compressive strength of concrete, several features with high importance are screened out as input variables by RF first, and then, the optimal weight and threshold value are found in BPNN by optimization ability of ISSA, and the RF-ISSA-BPNN prediction model is established; finally, three algorithms (PSO, SSA, and chaotic sparrow search algorithm [24](CSSA)) with excellent performance on the CEC2017 benchmark function are selected to establish the corresponding prediction model (RF-BPNN, RF-PSO-BPNN, RF-SSA-BPNN, RF-CSSA-BPNN, and ISSA-BPNN), and they are compared horizontally and vertically. From the predicted data of the final concrete compressive strength, it can be found that the MAE and RMSE values of RF-ISSA-BPNN are smaller than those of the other five models, the predicted data fit the actual data better, and the improvement of SSA and the feature selection based on random forest can improve the prediction model, which can provide a reliable theoretical reference for the safe construction of the project.

In summary, the main contributions of this paper are the following aspects:

- (i) In terms of algorithm optimization, Latin hypercube sampling is used to initialize the population, somersault foraging strategy is proposed to enrich the optimal position of producers, and cyclone foraging mechanism is proposed to obtain a better foraging position
- (ii) Ablation experiments are conducted to verify the feasibility and rationality of the three strategies, and the performance of ISSA algorithm is verified

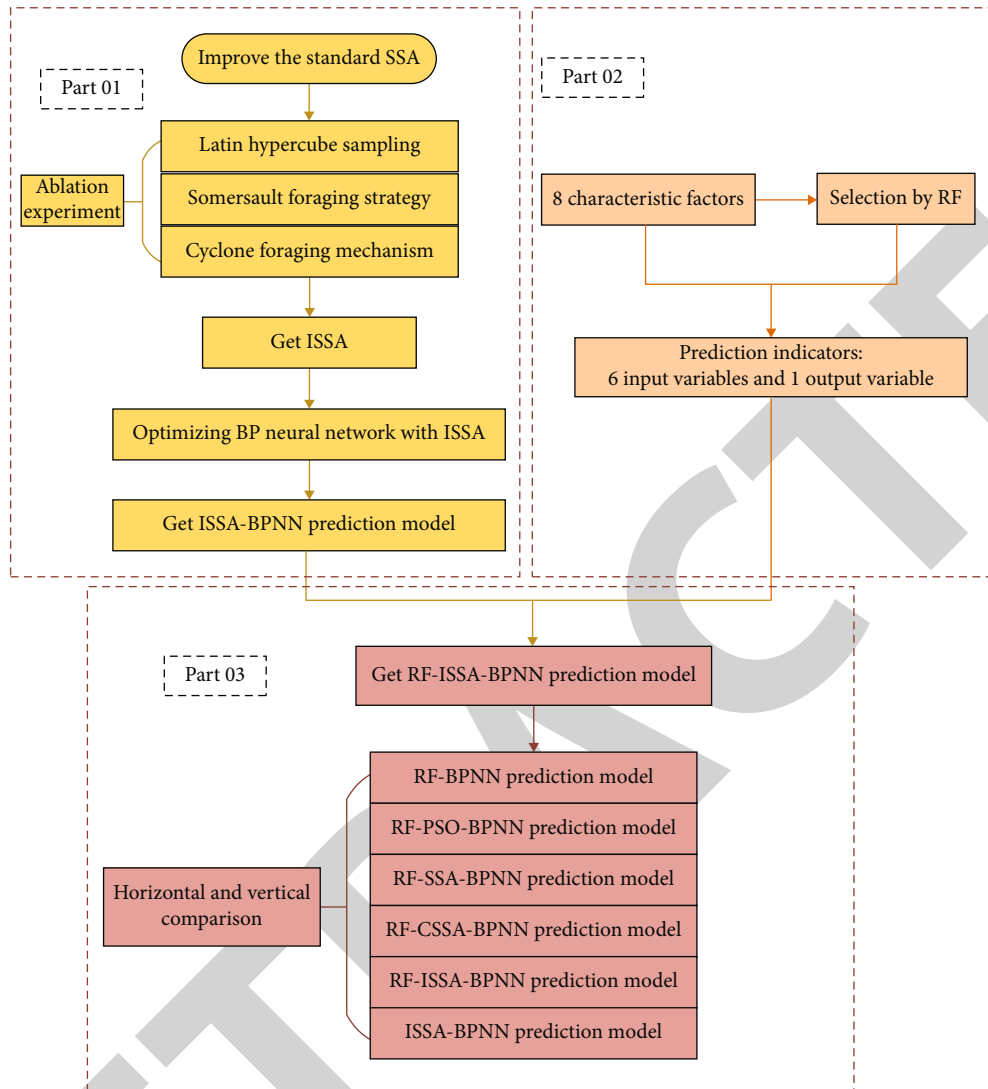


FIGURE 1: Brief flow chart of this research work.

by comparing it with some classical algorithms and two improved sparrow algorithms

- (iii) RF is used to measure the importance of each feature, and 6 features with VIM greater than 0.09 are screened as input variables
- (iv) The RF-ISSA-BPNN prediction model of concrete compressive strength is established and compared with the prediction model without RF and four kinds of prediction models with RF in both horizontal and vertical aspects. In addition, two evaluation indexes (MAE and RMSE) are selected as the measurement standards of the prediction accuracy

The rest of this paper is summarized below. Section 2 is an introduction to the related theoretical background and three improvement strategies of the sparrow search algorithm and gives the ISSA algorithm flow chart. Section 3 is divided into three parts: the ablation experiment, the comparison experiment with other algorithms, and the analysis

of the time complexity of the algorithm. Section 4 mainly includes the introduction of experimental data, the establishment of prediction model, and the comparison of prediction simulation experiment results. Finally, Section 6 is the summary of the whole paper. Figure 1 is a brief flow chart of this research work.

## 2. Theory

### 2.1. Related Theoretical Background

**2.1.1. BP Neural Network.** Figure 2 shows the basic structural framework of BPNN. BPNN is a multilayer feedforward network trained according to error back propagation algorithm [25]. It can learn and realize any complex nonlinear mapping between input and output through training of a large number of data samples [26] and then adjust weights and thresholds continuously through back propagation to minimize the error of output signals. However, it has shortcomings such as sensitive weight setting during fitting. Once

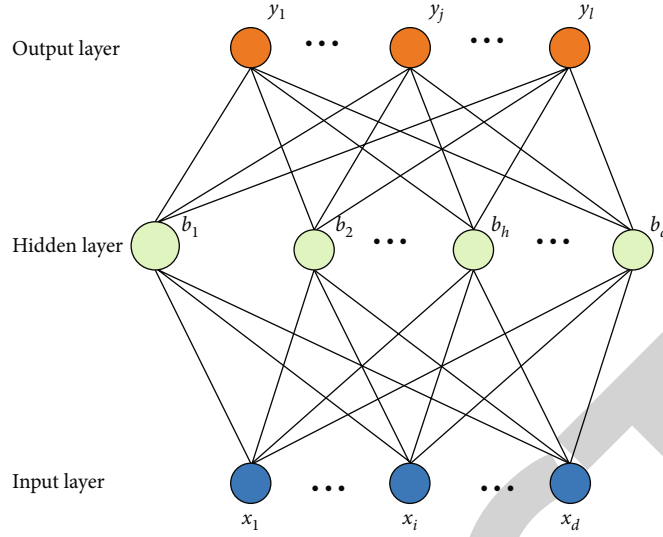


FIGURE 2: Schematic diagram of basic structure of BPNN.

the threshold and weight are set incorrectly, the performance of the model may be greatly reduced.

**2.1.2. Random Forest for Feature Selection.** In practical applications, there are often dozens of attributes of data set, and even the curse of dimensionality may occur. Therefore, data preprocessing for dimensionality reduction is a critical step in machine learning tasks. Feature selection relies on the feature selection function of the machine learning model itself, and selects the more important features from the input feature variables to achieve dimensionality reduction and simplify the complexity of the model to a certain extent. Generally speaking, there are two purposes for feature selection [27]. The first is to find highly correlated important variables to achieve the purpose of explanation, and the second is to find a small number of feature variables that can make good predictions.

Random forest [28] is not only widely used in prediction, but also in feature selection. It has better robustness and faster learning speed to noise and missing data, and its feature importance can be used as a feature selection tool for high-dimensional data [29]. The measurement indicators of characteristics in random forest mainly include Gini index [30] and out-of-bag data error rate [31, 32]. In this study, RF mainly uses the Gini index to calculate the average impurity and is used as an evaluation index to measure the contribution of each characteristic variable in the compressive strength of concrete. The higher the Gini index, the higher the average impurity, which shows that the importance of characteristic variables is more significant. The Gini index is represented by GI, and the variable importance score is represented by VIM.

**2.1.3. Sparrow Search Algorithm.** According to the foraging behavior of sparrows, the sparrow population is divided into producers and scroungers. Producers usually store high energy, which determines the direction of the whole population during the foraging process, while the scroungers

update their positions according to the foraging information provided by the producers. As long as sparrows can find a better source of food or have a higher energy reserve, they can become producers, but it is worth noting that the ratio of the producers and the scroungers remains constant in the entire population. The position of the sparrow represents a set of effective solutions in the search space, which is defined as  $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})$ ,  $i = 1, 2, \dots, n$ , where  $d$  is the dimension and  $n$  is population size. The energy reserve of the sparrow represents fitness value, which is defined as  $f(x_i)$ . The location of the producers is updated as

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \cdot \exp\left(\frac{-i}{n \cdot \text{iter}_{\max}}\right), R_2 < ST \\ x_{i,j}^t + Q \cdot L, R_2 \geq ST \end{cases}, \quad (1)$$

where  $x_{i,j}^t$  is the position of the  $i$ -th sparrow in the  $j$ -dimension ( $j = 1, 2, \dots, d$ ) at the  $t$ -th iteration.  $ST$  ( $ST \in [0.5, 1.0]$ ) is the safety threshold.  $R_2$  ( $R_2 \in [0, 1]$ ) is the alarm value.  $\alpha$  ( $\alpha \in (0, 1)$ ) is a random number.  $\text{iter}_{\max}$  is the maximum number of iterations.  $L$  is a  $d$ -dimensional row vector with all elements of 1.  $Q$  is a random number that obeys normal distribution. When  $R_2 < ST$ , there is no threat from predators in the environment, and producers can search for food in a wide range. When  $R_2 \geq ST$ , some sparrows are aware of the presence of predators, and all individuals should move away from their current position to avoid predators.

Scroungers monitor the producers over time, and they often move around producers in the best position and compete with them for resources. The location of the scroungers is updated as

$$x_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{x_{\text{worst}}^t - x_{i,j}^t}{t^2}\right), i > n/2 \\ x_p^{t+1} + |x_{i,j}^t - x_p^{t+1}| \cdot A^+ \cdot L, \text{otherwise} \end{cases}, \quad (2)$$

where  $x_{\text{worst}}^t$  is the current global worst position.  $x_p^{t+1}$  is the best position currently found by the producer.  $A$  is a row



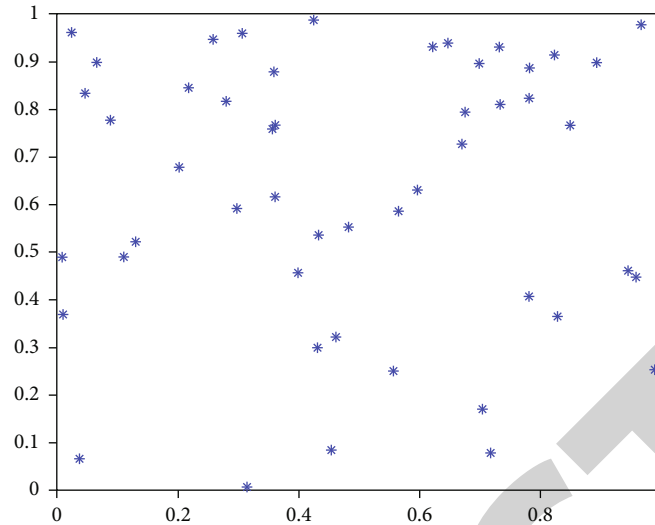


FIGURE 3: Rand function.

vector whose element can only be 1 or  $-1$ , and  $A^+ = A^T (AA^T)^{-1}$ . When  $i \leq n/2$ , scroungers search for food around the best location found by the producers. The remaining sparrows are starving and can only fly to other locations to find food.

In SSA, some sparrows will be selected to adopt the reconnaissance and early warning mechanism, so sparrows in different positions of the population will choose different coping methods in the face of incoming danger. When a sparrow is aware of danger, it will actively approach its partners in or around the safety circle to increase its own safety factor. The position of the scouts is updated as in

$$x_{i,j}^{t+1} = \begin{cases} x_{\text{best}}^t + \beta \cdot |x_{i,j}^t - x_{\text{best}}^t| \cdot f_i > f_g \\ x_{i,j}^t + K \cdot \left( \frac{|x_{i,j}^t - x_{\text{worst}}^t|}{(f_i - f_w) + \varepsilon} \right) \cdot f_i = f_g \end{cases}, \quad (3)$$

where  $x_{\text{best}}^t$  is the current global optimal position.  $K$  ( $K \in [-1, 1]$ ) is a random number.  $\varepsilon$  is the smallest constant to prevent the occurrence of 0 in the denominator.  $f_w$  is the current global worst fitness value,  $f_i$  is the fitness value of the scouts, and  $f_g$  is the current global best fitness value.  $\beta$  represents a step size control parameter that obeys the standard normal distribution. When  $f_i > f_g$ , the surroundings of the current global optimal position are safe, and the sparrows at the edge of the population realize the appearance of predators and quickly move around  $x_{\text{best}}^t$ . When  $f_i = f_g$ , sparrows in the center of the population should change their search strategy in time and seek protection from nearby partners to reduce the risk of being predation.

## 2.2. Improved Sparrow Search Algorithm

**2.2.1. Latin Hypercube Sampling.** Population initialization is an indispensable part of swarm intelligence optimization algorithm, and the convergence of swarm intelligence algorithm is easily affected by the distribution of the initial population [33]. The initial population of the traditional

sparrow search algorithm is generated based on random function in the feasible region. It can be found that the randomly generated population is not evenly distributed in Figure 3, which greatly reduces the efficiency of optimization. Latin hypercube sampling (LHS) is adopted to initialize the population to ensure the randomness and uniform distribution of sample points and improve the efficiency of optimization. As shown in Figure 4, random sample points are evenly distributed within the feasible region, which enrich the diversity of the primary population. Among them, population size is 50, dimension is 2, and the interval is  $[0, 1]$ .

Taking  $m$  samples in  $n$ -dimensional vector space as an example, the specific steps of LHS [34] are as follows:

- (1) The sample number and dimension of vector space are determined
- (2) Nonoverlapping  $m$  equal parts with equal probability is generated in each dimension
- (3) A random number is generated in each cell, so the sampling matrix  $[m, n]$  is formed
- (4) A number is randomly selected in each column of the sampling matrix to form a vector

Latin hypercube sampling is used to initialize the population in SSA, so the number of populations is  $m$  in the sampling matrix, and the multidimensional decision variables correspond to the  $n$ -dimensional vector space. LHS can be used to generate a sampling matrix  $[m, n]$ , in which each number in each column is generated by different cells and arranged in disorder. Therefore, a population with a wider and more uniform distribution range is formed, and the probability of obtaining a solution with good diversity and convergence is higher.

**2.2.2. Somersault Foraging Strategy.** In the stage of updating the location of producers, the position of sparrows shows an overall decreasing trend with the progress of iteration when

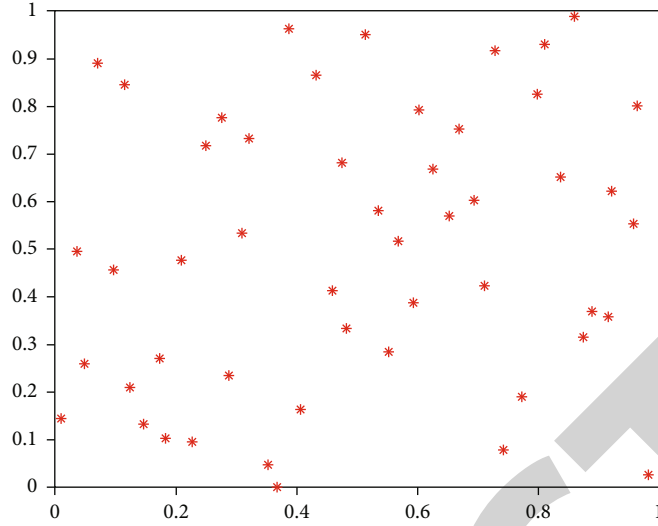
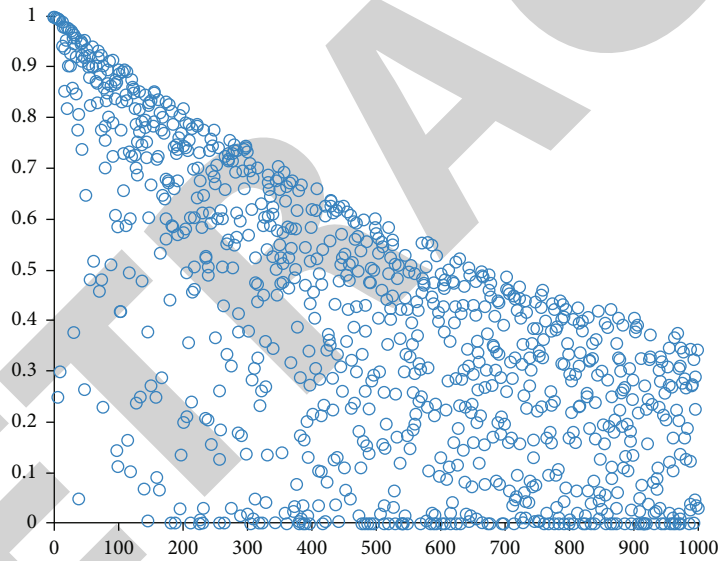


FIGURE 4: LHS.

FIGURE 5:  $f(x) = \exp(-x/1000 \times \alpha)$ .

$R_2 < ST$ . It can be found that the value range of  $f(x)$  has changed from  $[0,1]$  to  $[0,0.4]$  in Figure 5, which means that the diversity of the population may gradually decrease in the later iterations, and the probability of obtaining excellent solutions will also decrease. Therefore, the somersault foraging strategy [35] is introduced to enrich the optimal position of producers, which opened new foraging horizons for the whole population. The somersault foraging strategy is to take the position of the food (the optimal position) as the center point, and the individual always update its position by somersaults around the optimal position, and the expression is

$$x_{i,j}^{t+1} = x_{i,j}^t + S \cdot (r_2 \cdot x_{\text{best}}^t - r_3 \cdot x_{i,j}^t), \quad (4)$$

where  $x_{\text{best}}^t$  is the optimal position;  $S$  is the somersault factor, and the general value is 2; and  $r_2$  and  $r_3$  are two random numbers in  $[0,1]$ .

It can be seen from Equation (4) that the search area of producers is between the current position and the symmetrical position around the optimal position it currently finds, and it uses the global optimal position as the fulcrum to update its position. But with the iteration of the population, the range of sparrows foraging for somersaults is also shrinking, and all sparrows will gradually approach the optimal position. We use somersault foraging strategy into the stage of updating the location of producers to get an opposite position with the optimal position as the center point, and the optimal position is selected between the current sparrow and the opposite position. If the fitness value at

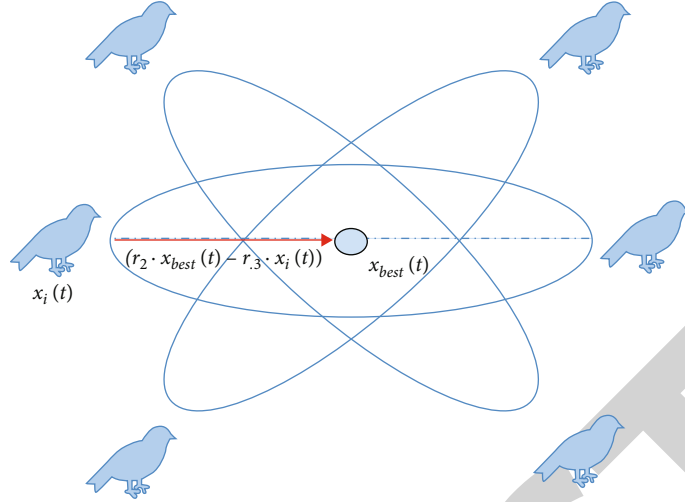


FIGURE 6: Producer's somersault foraging behavior.

the original position is inferior to that at the opposite position during each iteration, the original position is replaced. Otherwise, the original position is retained for the next generation. Different from the reverse learning strategy, the somersault strategy revolves around the optimal solution when updating the position, which makes the algorithm more convergent [36]. Figure 6 is a schematic diagram of the producer somersaulting.

**2.2.3. Cyclone Foraging Mechanism.** In the stage of updating the location of scrungers, the  $i$ -th scrunger is to get rid of the worst position in the current foraging process by virtue of the property of  $\exp()$  function, so as to obtain a better foraging position when  $i > n/2$ . However, its value will gradually tend to 0 when  $n$  is relatively large or the sparrow population converges, which means that the scrungers cannot fly to other positions and the population diversity is also lost. Theoretically, all starving sparrows should have the chance to search randomly using food as a reference location. The cyclone foraging mechanism [35] can create such an opportunity for hungry individuals to randomly designate a reference position in the entire search space, which can make the hungry individuals far away from the optimal position to find a new location and improve the global search capability of algorithm. Figure 7 shows the cyclone foraging behavior of sparrows in a two-dimensional space. It can be seen that the sparrows follow the preceding sparrows along the spiral path towards the food. Mathematical expressions are shown in

$$x_{\text{rand}}^d = Lb^d + r \cdot (Ub^d - Lb^d), \quad (5)$$

$$x_i^d(t+1) = x_{\text{rand}}^d + r \cdot (x_{i-1}^d(t) - x_i^d(t)) + \beta \cdot (x_{\text{rand}}^d - x_i^d(t)), \quad (6)$$

$$\beta = 2e^{r_1 \cdot (\text{iter}_{\text{max}} - t + 1 / \text{iter}_{\text{max}})} \cdot \sin(2\pi r_1), \quad (7)$$

where  $x_{\text{rand}}^d$  is a randomly generated position in the search

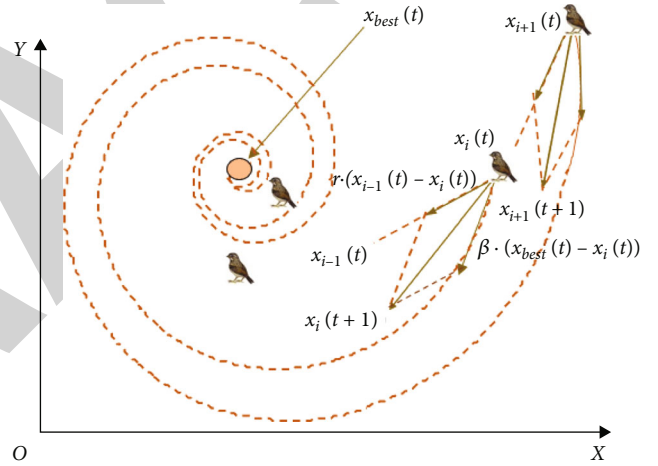


FIGURE 7: Cyclone foraging mechanism of individual sparrows.

space.  $Ub^d$  and  $Lb^d$  are defined upper and lower limits, respectively.  $r$  and  $r_1$  are two random numbers in  $[0,1]$ . Therefore, the position of scrungers in ISSA is updated as

$$x_{i,j}^{t+1} = \begin{cases} x_{\text{rand}}^d + r \cdot (x_{i-1,j}^t - x_{i,j}^t) + \beta \cdot (x_{\text{rand}}^d - x_{i,j}^t), & i > n/2 \\ x_p^{t+1} + |x_{i,j}^t - x_p^{t+1}| \cdot A^+ \cdot L, & \text{otherwise} \end{cases} \quad (8)$$

**2.3. ISSA Algorithm Flow Chart.** In summary, the overall flow chart of ISSA implementation is shown in Figure 8.

### 3. Algorithm Performance Test

**3.1. Simulation Environment and Test Function.** All simulation tests are performed on a computer with memory: 16 GB DDR4, CPU: AMD Ryzen 5 4600H with Radeon Graphics and operating system: Windows10; the compilation and operation of the program are all carried out in the Matlab2018a environment. Each optimization algorithm is tested to look for the optimal solution on the CEC2017 benchmark function solving problem in this paper, and they are run independently for 30 times to eliminate the



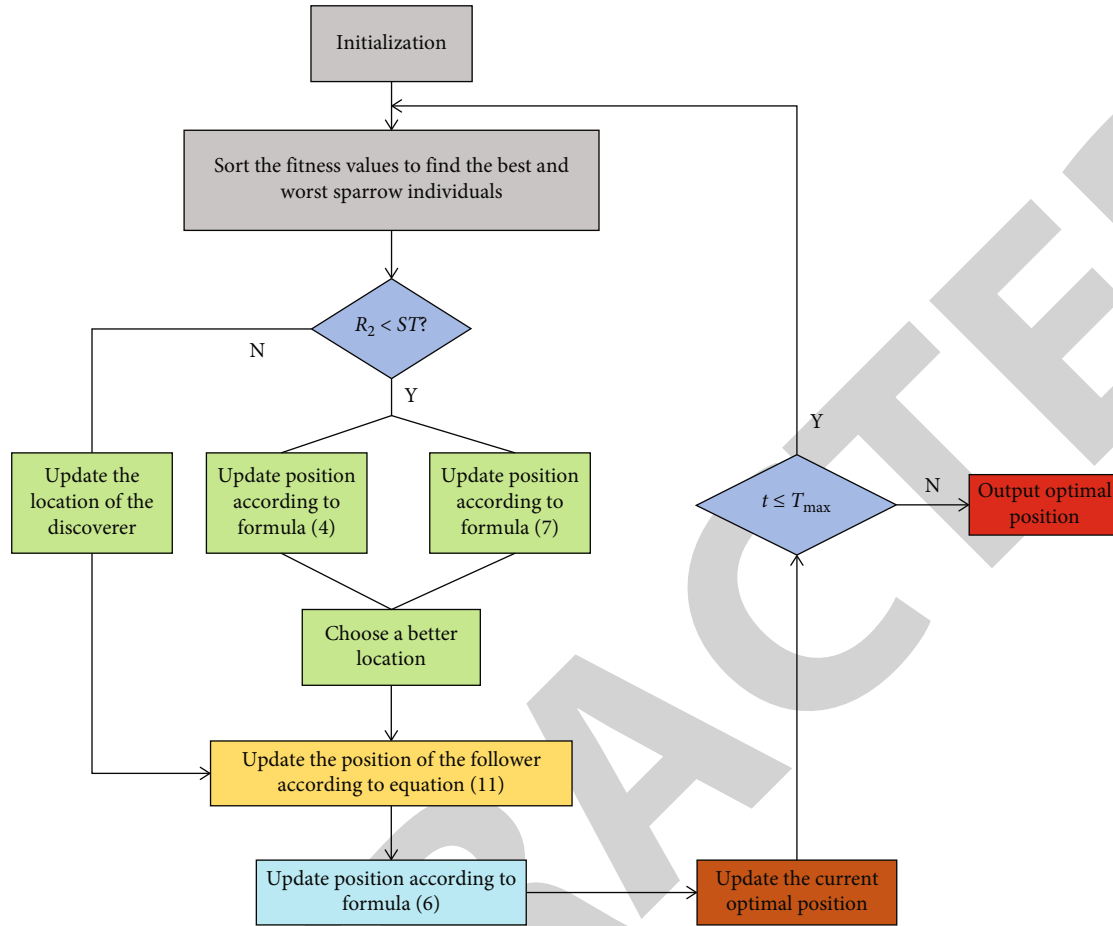


FIGURE 8: ISSA algorithm flow chart.

interference of accidental factors, and the results of mean, standard deviation (Std), and Friedman's ranking test [37] are recorded to obtain the average ranking (Ar) and the final ranking (Fr) on the entire CEC2017 benchmark function. The population size is set as 100, the  $iter_{max}$  is set as 200, and the dimension in the benchmark function is set to 30.

**3.2. Ablation Experiment.** It is worth checking whether the three introduced strategies play a role in the improvement of SSA separately or at the same time, and the ablation experiment [38] is used to investigate the improvement effects of each strategy. SSA1 (SSA with the introduction of LHS), SSA2 (SSA with the introduction of somersault foraging strategy), SSA3 (SSA with the introduction of cyclone foraging mechanism), ISSA (SSA with three strategies introduced at the same time), and SSA are simulated and compared, as shown in Table 1. The algorithm parameters are uniformly set to  $PD = 30$ ,  $ST = 0.6$ , and  $SD = 70$ .

From the final overall performance ranking, it can be found that the introduction of each strategy significantly improves the algorithm, and the performance of the SSA with the three strategies added at the same time is better than the SSA with the three strategies introduced separately, which means that LHS, somersault foraging, and cyclone foraging mechanism all play an improved role in SSA.

**3.3. Simulation Comparison with Other Algorithms.** The CEC2017 test function is often used to test the performance of intelligent optimization algorithms. In order to further evaluate the optimization effect of ISSA, it is compared with SSA, 3 classic swarm intelligence optimization algorithms, and 2 other improved sparrow algorithms; they are as follows: whale optimization algorithm (WOA) [39], artificial bee colony algorithm (ABC) [40], particle swarm optimization algorithm (PSO), chaos sparrow search algorithm (CSSA), and new chaos sparrow search algorithm (NCSSA) [41]. The general parameters are set the same to reflect the objectivity of each algorithm in the process of optimization. Table 2 lists other parameter settings of the seven algorithms.

As can be seen from the final results of Friedman's ranking test in Table 3, the overall ranking of ISSA is the first, followed by CSSA. Therefore, the optimization performance of ISSA on the CEC2017 benchmark function is the best. Compared with the classic algorithms ABC, WAO, and PSO, ISSA is better than WOA on 29 benchmark functions and is better than ABC and PSO on most benchmark functions. However, ABC is better than ISSA in optimization performance only on function F27, and the optimization performance of PSO on functions F4 and F11 is stronger than ISSA. On the whole, it can be seen that optimization

TABLE 1: Ablation experiments.

function	index	SSA	SSA1	SSA2	SSA3	ISSA
F1	Mean	$1.3 \times 10^5$	$1.3 \times 10^5$	$6.4 \times 10^4$	$1.4 \times 10^5$	$9.4 \times 10^4$
	Std	$5.7 \times 10^4$	$8.0 \times 10^4$	$4.3 \times 10^4$	$7.6 \times 10^4$	$5.6 \times 10^4$
	Rank	4	4	2	1	3
F3	Mean	$5.1 \times 10^4$	$5.1 \times 10^4$	$4.7 \times 10^4$	$6.5 \times 10^4$	$4.8 \times 10^4$
	Std	6784.0	$1.2 \times 10^4$	5714.2	$1.5 \times 10^4$	9803.3
	Rank	3	3	1	4	2
F4	Mean	511.2	505.8	521.4	501.5	523.5
	Std	16.6	24.7	21.5	27.3	62.3
	Rank	3	2	4	1	5
F5	Mean	676.0	677.5	665.8	683.2	670.8
	Std	45.1	48.8	59.7	31.1	33.9
	Rank	3	4	1	5	2
F6	Mean	626.9	634.7	636.7	629.9	623.5
	Std	7.4	8.8	12.2	9.9	9.7
	Rank	2	4	5	3	1
F7	Mean	1053.8	1017.1	1038.8	1008.2	981.8
	Std	98.7	111.4	86.2	41.1	69.5
	Rank	5	3	4	2	1
F8	Mean	947.3	943.2	942.1	941.3	934.5
	Std	38.8	32.1	31.0	26.9	22.8
	Rank	5	4	3	2	1
F9	Mean	4969.7	4729.5	4301.5	3650.6	3805.2
	Std	940.7	984.3	830.1	1238.4	746.5
	Rank	5	4	3	1	2
F10	Mean	5076.4	4983.1	5465.1	4740.0	4504.6
	Std	683.9	744.3	804.9	661.6	696.8
	Rank	4	3	5	2	1
F11	Mean	1325.6	1314.1	1305.3	1300.9	1302.1
	Std	53.5	70.0	76.0	77.0	63.6
	Rank	5	4	3	1	2
F12	Mean	$2.3 \times 10^6$	$2.0 \times 10^6$	$1.3 \times 10^6$	$1.9 \times 10^6$	$2.4 \times 10^6$
	Std	$1.4 \times 10^6$	$9.7 \times 10^5$	$6.6 \times 10^5$	$1.0 \times 10^6$	$1.5 \times 10^6$
	Rank	5	4	1	2	3
F13	Mean	$3.8 \times 10^4$	$2.6 \times 10^4$	$1.6 \times 10^4$	$4.5 \times 10^4$	$2.6 \times 10^4$
	Std	$4.7 \times 10^4$	$2.4 \times 10^4$	$2.2 \times 10^4$	$4.1 \times 10^4$	$1.3 \times 10^4$
	Rank	3	2	1	4	2
F14	Mean	$9.5 \times 10^4$	$9.2 \times 10^4$	$3.3 \times 10^4$	$9.3 \times 10^4$	$3.5 \times 10^4$
	Std	$7.4 \times 10^4$	$5.0 \times 10^4$	$2.1 \times 10^4$	$7.2 \times 10^4$	$2.8 \times 10^4$
	Rank	5	3	1	4	2
F15	Mean	$1.1 \times 10^4$	$1.7 \times 10^4$	$1.1 \times 10^4$	$3.7 \times 10^4$	$1.3 \times 10^4$
	Std	$1.0 \times 10^4$	$1.3 \times 10^4$	$1.2 \times 10^4$	$2.4 \times 10^4$	$1.3 \times 10^4$
	Rank	1	3	1	4	2
F16	Mean	2778.1	2728.7	2864.8	2670.1	2610.3
	Std	145.3	259.6	256.1	387.0	293.1
	Rank	4	3	5	2	1
F17	Mean	2300.0	2207.1	2325.4	2245.1	2263.6

TABLE 1: Continued.

function	index	SSA	SSA1	SSA2	SSA3	ISSA
F18	Std	189.4	306.2	290.7	344.7	245.0
	Rank	4	1	5	2	3
	Mean	$3.5 \times 10^5$	$1.0 \times 10^6$	$9.9 \times 10^5$	$4.4 \times 10^5$	$5.0 \times 10^5$
	Std	$4.7 \times 10^5$	$1.0 \times 10^6$	$9.9 \times 10^5$	$3.2 \times 10^5$	$8.5 \times 10^5$
F19	Rank	1	5	4	2	3
	Mean	$2.2 \times 10^4$	$1.4 \times 10^4$	8133.5	$1.2 \times 10^4$	$1.3 \times 10^4$
	Std	$1.6 \times 10^4$	$1.3 \times 10^4$	5704.1	$1.5 \times 10^4$	$1.2 \times 10^4$
	Rank	5	4	1	2	3
F20	Mean	2654.8	2358.4	2769.2	2521.9	2519.2
	Std	160.4	248.2	169.5	128.1	149.4
	Rank	4	1	5	3	2
	Mean	2471.8	2460.5	2477.4	2445.1	2440.7
F21	Std	57.3	41.4	47.0	49.9	30.7
	Rank	4	3	5	2	1
	Mean	5352.7	3642.1	2545.9	2938.4	5087.8
	Std	1895.5	1992.0	927.5	1457.3	2009.7
F22	Rank	5	3	2	1	4
	Mean	2817.6	2831.5	2789.5	2858.7	2801.6
	Std	60.5	48.0	35.0	63.8	26.2
	Rank	5	3	1	4	2
F23	Mean	2975.0	2968.0	2991.7	2952.2	2992.5
	Std	63.4	33.1	46.7	37.7	30.3
	Rank	3	2	4	1	5
	Mean	2905.6	2894.1	2911.0	2900.2	2897.3
F24	Std	20.4	13.8	21.5	16.0	14.0
	Rank	4	1	5	3	2
	Mean	4675.2	5580.6	5229.6	5384.8	4684.3
	Std	856.7	712.2	971.1	701.3	1332.2
F25	Rank	1	5	3	4	2
	Mean	3235.7	3244.3	3202.5	3237.2	3227.7
	Std	15.1	25.3	7.7	12.5	38.1
	Rank	3	5	1	4	2
F26	Mean	3233.7	3247.9	3246.0	3247.4	3254.9
	Std	22.0	31.0	18.0	20.8	35.6
	Rank	1	4	2	3	5
	Mean	3972.0	3981.6	4078.6	4039.0	3912.8
F27	Std	192.2	191.8	169.7	269.7	193.8
	Rank	2	3	5	4	1
	Mean	$4.1 \times 10^4$	$2.6 \times 10^4$	$3.5 \times 10^4$	$5.2 \times 10^4$	$1.9 \times 10^4$
	Std	$2.5 \times 10^4$	$2.7 \times 10^4$	$5.9 \times 10^4$	$2.6 \times 10^4$	$1.6 \times 10^4$
F28	Rank	4	2	3	5	1
	Ar	3.552	3.172	2.966	2.621	2.276
	Fr	5	4	3	2	1

performance of ISSA is only slightly worse than the classical algorithm in a few benchmark functions. Compared with SSA, ISSA has better optimization performance on 22 benchmark functions. Compared with the improved algo-

rithms CSSA and NCSSA, ISSA outperforms them on 16 benchmark functions and 26 benchmark functions, respectively, and is slightly worse than NCSSA only on functions F19, F22, and F24. In summary, the global search and local

TABLE 2: Parameter settings.

Algorithm	Parameter settings
ABC	$L = \text{round}(0.6 * \text{dim} * N), a = 1$
WOA	$b = 1, a = 2 - (2t/T_{\max}), P^* = 0.5$
PSO	$c_1 = 2, c_2 = 2, w_{\max} = 0.9, w_{\min} = 0.2$
SSA	PD = 30, ST = 0.6, SD = 70
CSSA	PD = 30, ST = 0.6, SD = 70
NCSSA	PD = 30, ST = 0.6, SD = 70-round((70-50) × (t/200))
ISSA	PD = 30, ST = 0.6, SD = 70

development capabilities of ISSA are better than the other six algorithms, which fully explain the successful introduction of LHS, somersault foraging strategy, and cyclone foraging mechanism into sparrow search algorithm.

**3.4. Time Complexity.** The time complexity of an algorithm is used to measure the operating efficiency of the algorithm, and it reflects the pros and cons of an algorithm to a large extent. Therefore, ISSA is compared with SSA and other improved algorithms (CSSA and NCSSA) in terms of time complexity as shown in Table 4, where the dimension is denoted by  $d$ , and the time to solve the fitness function is denoted by  $f(d)$ . According to the introduction of the principle of the standard sparrow algorithm in Section 2.1.3, the algorithm is mainly composed of five phases: population initialization, location of producers updated, location of scroungers updated, location of scouts updated, and update of optimal location.

We discuss and analyze the time complexity of five phases of the algorithm, respectively, and it can be obtained through analysis in Table 4 that the time complexity of SSA, CSSA, NCSSA, and ISSA is equal, so the three improved algorithms do not increase the algorithm complexity in exchange for the improvement of performance. Combined with the experimental results on the CEC2017 benchmark function in Section 3.3, it can be seen that ISSA has better optimization performance than CSSA and NCSSA under the same time complexity of algorithm.

## 4. Prediction of Concrete Compressive Strength Based on RF-ISSA-BPNN Model

**4.1. Experimental Data.** With the rapid development of cement and concrete production technology, concrete has now become the largest amount of man-made building materials in the world. High-performance concrete (HPC) is a new type of concrete; using modern concrete technology, the use of high-quality raw materials, in addition to cement, aggregates, and water, must be used in a suitable water-cement ratio, mixed with sufficient high-quality mineral admixtures and efficient admixtures. As the academic community generally agreed that strength is the most important indicator of the performance of concrete, so many studies have long been done around how to improve the strength. Concrete as a very common material on modern construction projects, its compressive strength also determines the

quality of construction. Therefore, predicting the compressive strength from the available data is a challenging task. The prediction of compressive strength of concrete is a very complex nonlinear curve, and many factors directly or indirectly affect the compressive strength of concrete. The experimental data for the compressive strength of concrete used in this study consisted of 1,030 sets of data as well as nine properties. Among them, the compressive strength of concrete is affected by age, fine aggregate, cement, superplasticizer, blast furnace slag, water, coarse aggregate, and fly ash. The units of the last 7 factors are  $kg/m^3$ , the unit of age is calculated by days, and the unit of compressive strength of concrete is MPa. The compressive strength of concrete is highly nonlinear with age and ingredients. It can be analyzed from Figures 9 and 10 that the compressive strength of concrete is highly nonlinearly correlated with age and composition. Detailed information about the input properties is mentioned in Table 5.

**4.2. Establishment of RF-ISSA-BPNN Model.** ISSA shows strong optimization ability in the CEC2017 test function, so ISSA can directly participate in the process of network parameter optimization. The so-called network parameters optimization is to find good weights and thresholds to minimize the global error in network [42], and then, train the optimized model to obtain the final prediction result. The dimensionality is  $q \times (l + d + 1) + l$  because the dimension of individual sparrows is decided by the weights and thresholds obtained together.

The specific steps for establishing the RF-ISSA-BPNN model are as follows:

- (1) The sample data are imported, and RF is used to conduct feature selection on the sample data. The number of neurons in each layer, the transfer function, and the number of training times are determined so as to start building the network
- (2) Parameter and sparrow population are initialized, and the sum of the absolute values of the prediction errors obtained by training is set as the objective function, as shown in

$$f(x) = \sum_{i=1}^n |\tilde{x}_i - x_i|. \quad (9)$$

- (3) The optimal solution corresponding to the minimum fitness function value is found by ISSA, the optimal solution obtained by optimization is assigned to weights and thresholds, and then, the training begins. When the training accuracy requirements are met, the final prediction results are output

### 4.3. Application of RF-ISSA-BPNN Prediction Model

**4.3.1. Evaluation of the Importance of Features.** In the Python 3.7 operating environment, the importance score  $V$

TABLE 3: Experimental results of 7 algorithms applied to CEC2017 test set.

function	index	ABC	WOA	PSO	SSA	CSSA	NCSSA	ISSA
F1	Mean	$4.8 \times 10^9$	$4.0 \times 10^9$	$2.4 \times 10^7$	$1.3 \times 10^5$	$1.5 \times 10^5$	$5.9 \times 10^5$	$9.4 \times 10^4$
	Std	$8.0 \times 10^8$	$1.3 \times 10^9$	$6.2 \times 10^6$	$5.7 \times 10^4$	$1.1 \times 10^5$	$2.6 \times 10^5$	$5.6 \times 10^4$
	Rank	7	6	5	2	3	4	1
F3	Mean	$3.0 \times 10^5$	$2.5 \times 10^5$	$7.5 \times 10^4$	$5.1 \times 10^4$	$4.6 \times 10^4$	$5.9 \times 10^4$	$4.8 \times 10^4$
	Std	$8.8 \times 10^4$	$8.8 \times 10^4$	$1.4 \times 10^4$	6784.0	7117.0	3913.0	9803.3
	Rank	7	6	5	3	1	4	2
F4	Mean	1900.0	1408.6	485.4	511.2	524.2	528.1	523.5
	Std	259.0	411.5	25.6	16.6	23.3	21.6	62.3
	Rank	7	6	1	2	4	5	3
F5	Mean	785.1	841.0	709.0	676.0	649.9	676.0	670.8
	Std	22.1	54.2	39.6	45.1	43.1	25.3	33.9
	Rank	5	6	4	3	1	3	2
F6	Mean	646.2	674.9	652.0	626.9	624.2	624.2	623.5
	Std	3.4	5.2	5.5	7.4	12.7	4.9	9.7
	Rank	4	6	5	3	2	2	1
F7	Mean	1060.4	1255.8	1098.6	1053.8	945.4	1038.7	981.8
	Std	20.3	117.8	40.2	98.7	90.3	123.7	69.5
	Rank	5	7	6	4	1	3	2
F8	Mean	1084.3	1046.1	966.5	947.3	946.3	944.4	934.5
	Std	17.3	22.5	28.1	38.8	23.7	31.2	22.8
	Rank	7	6	5	4	3	2	1
F9	Mean	$1.1 \times 10^4$	9696.4	7369.8	4969.7	4221.6	4871.8	3805.2
	Std	1471.7	3438.2	1098.3	940.7	1153.1	789.6	746.5
	Rank	7	6	5	4	2	3	1
F10	Mean	9299.1	7029.2	5620.8	5076.4	4858.0	4767.8	4504.6
	Std	328.5	714.8	437.3	683.9	396.0	432.5	696.8
	Rank	7	6	5	4	3	2	1
F11	Mean	$1.5 \times 10^4$	7897.1	1271.3	1325.6	1283.9	1324.6	1302.1
	Std	2741.5	3154.0	33.2	53.5	68.0	89.8	63.6
	Rank	7	6	1	5	3	4	2
F12	Mean	$8.4 \times 10^8$	$3.8 \times 10^8$	$1.1 \times 10^7$	$2.3 \times 10^6$	$1.6 \times 10^6$	$2.5 \times 10^6$	$2.4 \times 10^6$
	Std	$2.3 \times 10^8$	$3.0 \times 10^8$	$5.7 \times 10^6$	$1.4 \times 10^6$	$1.2 \times 10^6$	$2.6 \times 10^6$	$1.5 \times 10^6$
	Rank	7	6	5	2	1	4	3
F13	Mean	$4.5 \times 10^7$	$1.1 \times 10^7$	$1.1 \times 10^5$	$3.8 \times 10^4$	$4.4 \times 10^4$	$7.7 \times 10^4$	$2.6 \times 10^4$
	Std	$7.9 \times 10^6$	$9.7 \times 10^6$	$5.6 \times 10^4$	$4.7 \times 10^4$	$6.5 \times 10^4$	$6.6 \times 10^4$	$1.3 \times 10^4$
	Rank	7	6	5	2	3	4	1
F14	Mean	$2.7 \times 10^5$	$1.9 \times 10^6$	$1.7 \times 10^5$	$9.5 \times 10^4$	$2.0 \times 10^4$	$8.0 \times 10^4$	$3.5 \times 10^4$
	Std	$1.8 \times 10^5$	$1.0 \times 10^6$	$9.6 \times 10^4$	$7.4 \times 10^4$	$1.8 \times 10^4$	$7.1 \times 10^4$	$2.8 \times 10^4$
	Rank	6	7	5	4	1	3	2
F15	Mean	$3.8 \times 10^6$	$1.8 \times 10^6$	$2.1 \times 10^4$	$1.1 \times 10^4$	8246.8	$1.6 \times 10^4$	$1.3 \times 10^4$
	Std	$9.5 \times 10^5$	$1.7 \times 10^6$	8248.9	$1.0 \times 10^4$	6448.7	$1.4 \times 10^4$	$1.3 \times 10^4$
	Rank	7	6	5	2	1	4	3
F16	Mean	4233.3	4227.7	3348.2	2778.1	2921.8	2741.1	2610.3
	Std	102.8	384.3	435.1	145.3	506.3	213.8	293.1
	Rank	7	6	5	3	4	2	1
F17	Mean	3077.4	3061.4	2321.6	2300.0	2304.7	2383.4	2263.6



TABLE 3: Continued.

function	index	ABC	WOA	PSO	SSA	CSSA	NCSSA	ISSA
F18	Std	179.8	542.8	269.5	189.4	287.1	217.1	245.0
	Rank	7	6	4	2	3	5	1
	Mean	$2.2 \times 10^7$	$1.2 \times 10^7$	$6.7 \times 10^5$	$3.5 \times 10^5$	$6.6 \times 10^5$	$1.1 \times 10^6$	$5.0 \times 10^5$
F19	Std	$9.5 \times 10^6$	$9.4 \times 10^6$	$4.2 \times 10^5$	$4.7 \times 10^5$	$8.4 \times 10^5$	$1.6 \times 10^6$	$8.5 \times 10^5$
	Rank	7	6	4	1	3	5	2
	Mean	$1.4 \times 10^6$	$1.7 \times 10^7$	$4.2 \times 10^4$	$2.2 \times 10^4$	8868.6	$1.1 \times 10^4$	$1.3 \times 10^4$
F20	Std	$4.8 \times 10^5$	$1.3 \times 10^7$	$4.9 \times 10^4$	$1.6 \times 10^4$	$1.0 \times 10^4$	7202.9	$1.2 \times 10^4$
	Rank	6	7	5	4	1	2	3
	Mean	3179.1	2790.0	2725.0	2654.8	2515.0	2570.3	2519.2
F21	Std	137.9	174.2	217.2	160.4	203.7	173.0	149.4
	Rank	7	6	5	4	1	3	2
	Mean	2583.3	2637.1	2536.4	2471.8	2466.6	2449.2	2440.7
F22	Std	13.6	83.5	32.5	57.3	56.7	29.0	30.7
	Rank	6	7	5	4	3	2	1
	Mean	$1.0 \times 10^4$	8467.4	7041.4	5352.7	4247.9	4209.8	5087.8
F23	Std	177.8	734.1	1912.7	1895.5	2115.3	2113.3	2009.7
	Rank	7	6	5	4	2	1	3
	Mean	2977.9	3063.9	3456.9	2817.6	2810.7	2802.0	2801.6
F24	Std	13.0	90.7	221.6	60.5	45.5	51.7	26.2
	Rank	5	6	7	4	3	2	1
	Mean	3130.6	3270.3	3360.5	2975.0	2987.6	2986.9	2992.5
F25	Std	14.5	77.5	59.5	63.4	60.4	41.3	30.3
	Rank	5	6	7	1	3	2	4
	Mean	3342.6	3184.2	2928.3	2905.6	2899.1	2904.8	2897.3
F26	Std	60.6	46.2	18.9	20.4	14.9	17.5	14.0
	Rank	7	6	5	4	2	3	1
	Mean	6507.6	8639.1	6774.7	4675.2	5305.7	5067.9	4684.3
F27	Std	254.1	548.0	2309.6	856.7	1086.6	756.5	1332.2
	Rank	5	7	6	1	4	3	2
	Mean	3200.0	3494.4	3861.9	3235.7	3249.1	3253.5	3227.7
F28	Std	$5.4 \times 10^{-5}$	123.5	395.0	15.1	17.8	20.1	38.1
	Rank	1	6	7	3	4	5	2
	Mean	3300.0	3748.8	3271.5	3233.7	3261.4	3267.4	3254.9
F29	Std	0.03	256.0	21.4	22.0	50.7	26.6	35.6
	Rank	6	7	5	1	3	4	2
	Mean	5184.5	5206.8	4550.6	3972.0	3983.7	3969.8	3912.8
F30	Std	113.7	640.5	264.7	192.2	175.7	196.1	193.8
	Rank	6	7	5	3	4	2	1
	Mean	$1.6 \times 10^6$	$4.5 \times 10^7$	$2.1 \times 10^6$	$4.1 \times 10^4$	$2.4 \times 10^4$	$6.0 \times 10^4$	$1.9 \times 10^4$
	Std	$8.6 \times 10^5$	$2.9 \times 10^7$	$9.5 \times 10^5$	$2.5 \times 10^4$	$1.7 \times 10^4$	$4.5 \times 10^4$	$1.6 \times 10^4$
	Rank	5	7	6	3	2	4	1
	Ar	6.103	6.276	4.931	2.966	2.379	3.172	1.724
	Fr	6	7	5	3	2	4	1

IM of each feature is calculated through RF and sorted. The higher the ranking, the higher the importance of the feature, which also means the more correlated it is with the compressive strength of concrete.

According to the ranking results of V IM in Figure 11, it can be seen that age has the greatest influence on the compressive strength of concrete, while the importance of fly ash is the lowest. This paper uses 0.09 as the threshold to

TABLE 4: Time complexity comparison of algorithm.

Phase	Algorithm			
	SSA	CSSA	NCSSA	ISSA
Population initialization	$O(d + f(d))$	$O(d + f(d))$	$O(d + f(d))$	$O(d + f(d))$
Location of producers updated	$O(d)$	$O(d)$	$O(d)$	$O(d)$
Location of scroungers updated	$O(d)$	$O(d)$	$O(d)$	$O(d)$
Location of scout updated	$O(d)$	$O(d)$	$O(d)$	$O(d)$
Update of optimal location	$O(d)$	$O(d)$	$O(d)$	$O(d)$

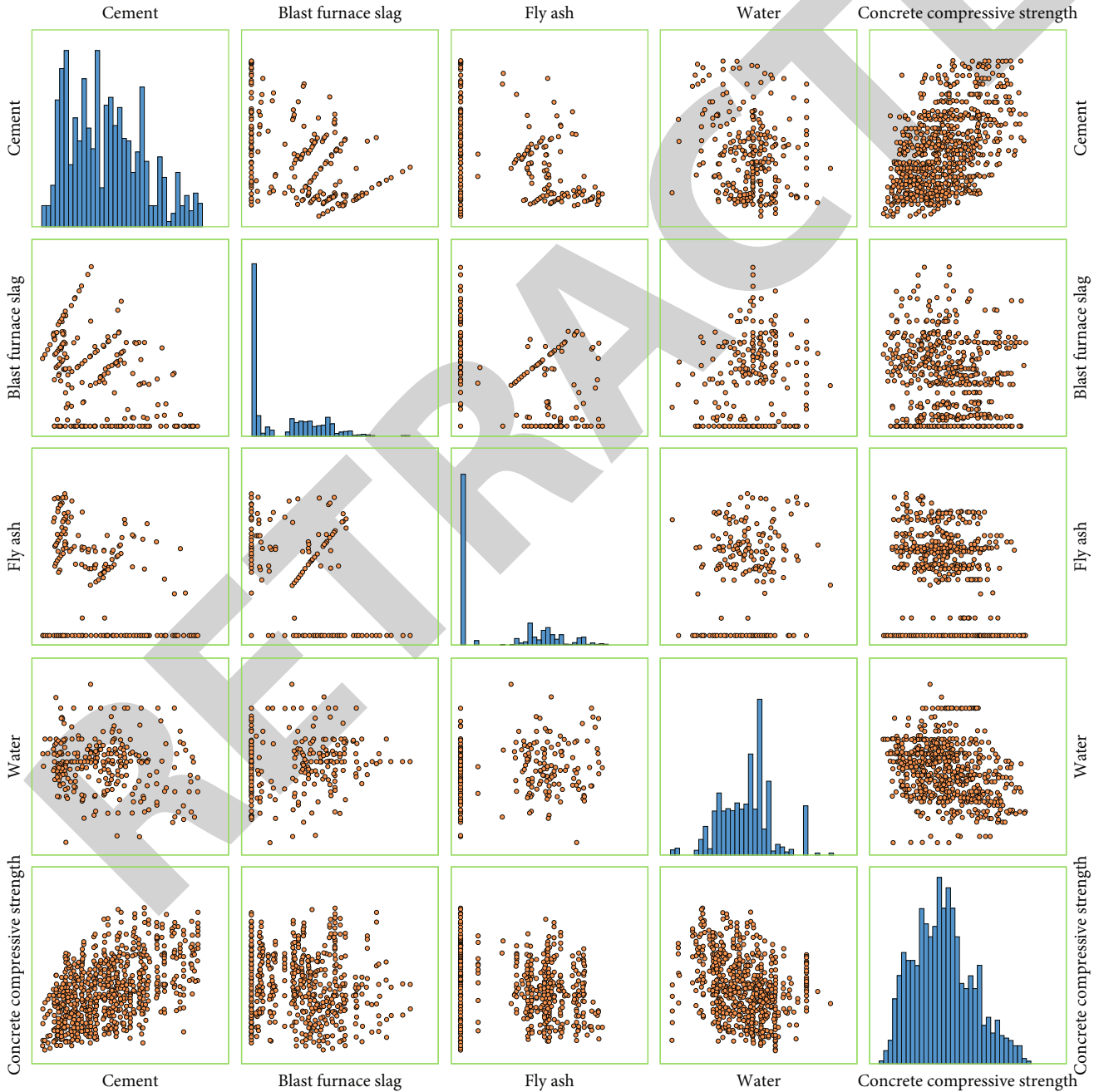


FIGURE 9: Plot of the relationship between output variables and the first 4 input feature variables (cement, blast furnace slag, fly ash, and water).

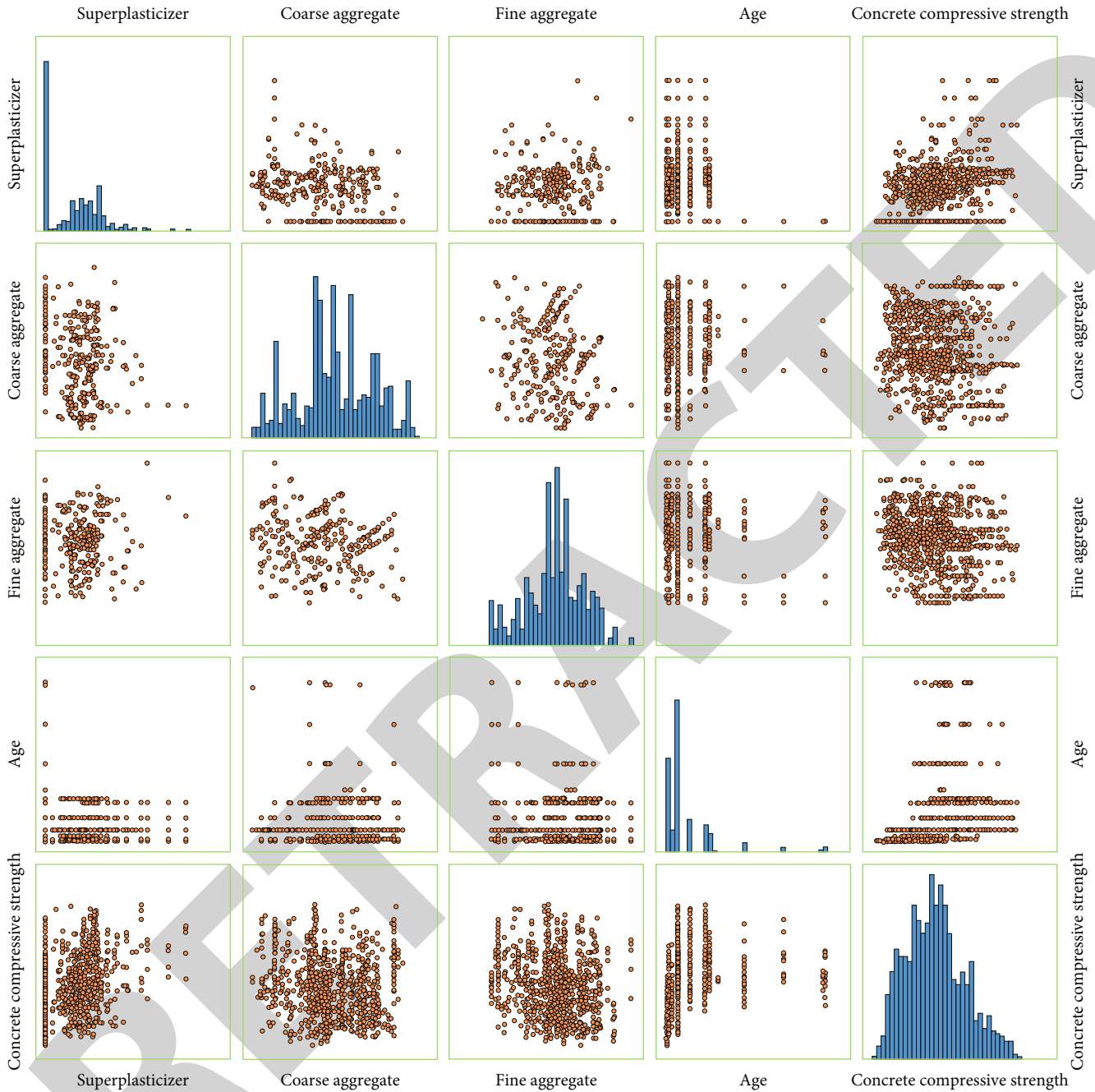


FIGURE 10: Plot of the relationship between output variables and the last 4 input feature variables (superplasticizer, coarse aggregate, fine aggregate, and age).

filter out 6 characteristic factors (age, cement, fine aggregate, coarse aggregate, water, and blast furnace slag) whose  $V_{IM}$  is greater than 0.09, as input variables of the prediction model, the two characteristic variables (superplasticizer and fly ash) that have little influence on the compressive strength of concrete are eliminated, so that the model is simplified without losing important features and the calculation efficiency of the model is improved.

4.3.2. *Select an Appropriate Number of Nodes.* The 6 features selected by RF are set as the 6 nodes of the input layer of the BPNN, and the compressive strength of concrete is the only

node of output, but improper setting of node number may lead to poor prediction results in the hidden layer. Therefore, how to select the appropriate number of nodes in the model is a key problem, so as to achieve the best prediction performance. Kolmogorov theorem [43] proposes that the optimal number of points of the hidden layer is generally  $2n + 1$ , where  $n$  is the number of nodes in the input layer. When  $n = 6$ , the optimal number of points is 13. Figure 12 is the basic structure diagram of BPNN.

4.3.3. *Prediction of Concrete Compressive Strength.* The trained RF-ISSA-BPNN model is used to predict the

TABLE 5: Statistical features of concrete data sets.

Concrete components	Maximum	Minimum	Mean	Std deviation
Cement	540.0	102.0	281.2	104.50
Blast furnace slag	359.4	0.0	73.9	86.27
Fly ash	200.1	0.0	54.2	63.99
Water	247.0	121.8	181.6	21.35
Superplasticizer	32.2	0.0	6.2	5.97
Coarse aggregate	1145.0	801.0	972.9	77.75
Fine aggregate	992.6	594.0	773.6	80.17
Age	365.0	1.0	45.7	63.16
Concrete compressive strength	82.6	2.3	35.8	16.70

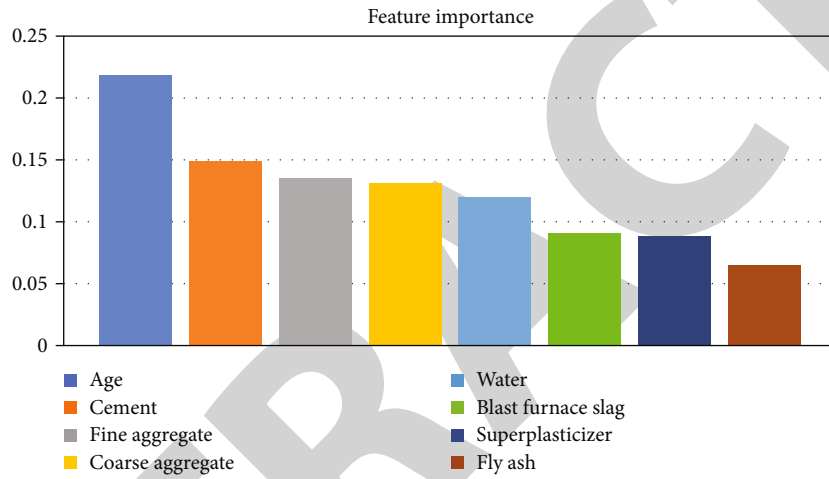


FIGURE 11: Feature importance ranking.

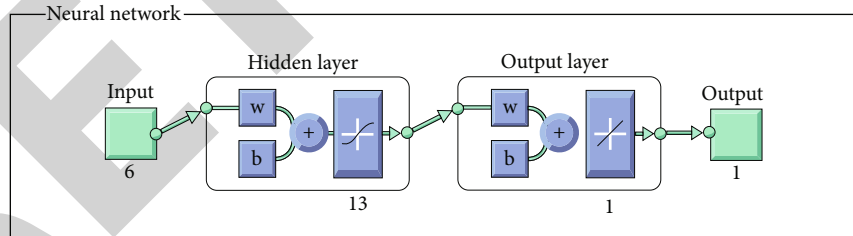


FIGURE 12: The basic structure of BP neural network.

compressive strength of concrete, and the feasibility and superiority of the model are proved. It is longitudinally compared with the prediction model ISSA-BPNN that has not been processed by the RF feature selection process. In addition, the algorithms (PSO, SSA, and CSSA) with excellent performance on the CEC2017 benchmark function are selected to participate in the process of network parameter optimization, and the prediction models of RF-BPNN, RF-PSO-BPNN, RF-SSA-BPNN, and RF-CSSA-BPNN are established for horizontal comparison. In this paper, 1,000 sets of data are used as the training set, and 30 sets of data are used as the test set for simulation and prediction. The

model parameters of BP neural network are set uniformly: The number of training is 100, the accuracy is 0.001, and the learning rate is 0.01. The parameters of the selected optimization algorithm are consistent with those in Table 2, and the simulation experiments are performed in the Matlab2018a compilation environment. The prediction results of concrete compressive strength of each model are shown in Figure 13.

Cross-validation is an important parameter for the evaluation of machine learning algorithms how well prediction capability is with generalization for an independent data set and can be used to assess the model capability to predict

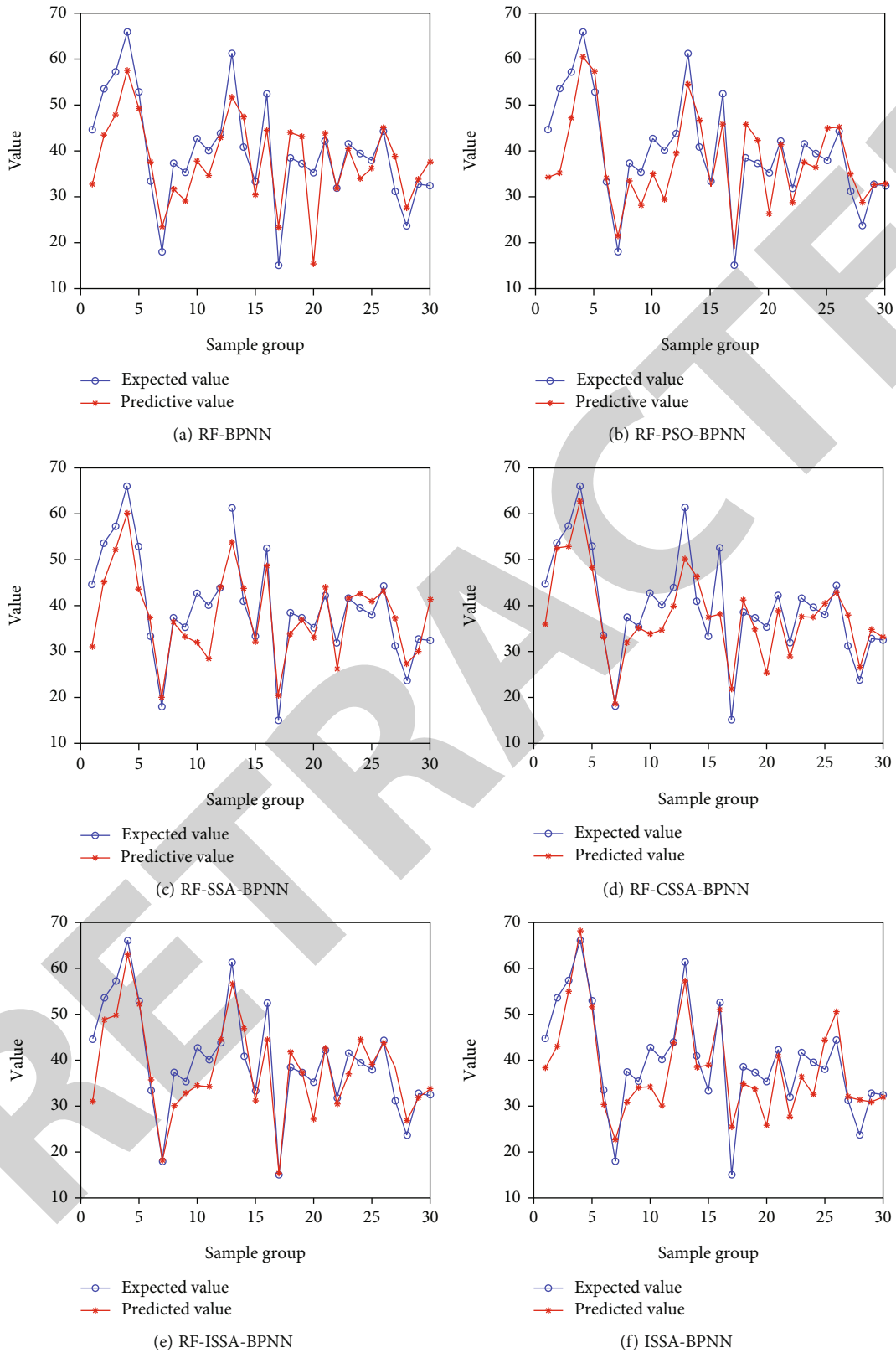


FIGURE 13: The prediction results of the concrete compressive strength of each model.

the new data and to obtain insight about the model capability for prediction of the independent data set [44]. Tenfold cross-validation [45] is a commonly used technique for

cross-validation, and this section discusses the accuracy of the RF-ISSA-BPNN model from the perspective of tenfold cross-validation. The error evaluation indexes can fully



TABLE 6: Predicted results of 6 models.

Index	Model					
	RF-BPNN	RF-PSO-BPNN	RF-SSA-BPNN	RF-CSSA-BPNN	RF-ISSA-BPNN	ISSA-BPNN
MAE	5.6620	4.9291	5.4083	5.1226	4.3798	4.9887
RMSE	7.5048	6.7247	6.8324	6.4402	5.6687	6.5417

reflect the real gap between predicted results, and the actual results, two commonly used error evaluation indexes (MAE and RMSE) are selected as a measure of each model performance, as shown in Equations (10) and (11). The closer the final result of the three is to 0, the better the performance of the model, as shown in Table 6.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \left| \tilde{y}_i - y_i \right|, \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\tilde{y}_i - y_i)^2}, \quad (11)$$

where  $N$  is the predicted array length,  $y_i$  is the actual concrete compressive strength, and  $\tilde{y}_i$  is the predicted concrete compressive strength.

## 5. Results and Discussion

As can be seen from Figure 13, whether horizontal or vertical comparison, the predicted data of RF-ISSA-BPNN model has the best fitting degree with the actual data, and the prediction performance is better than other models. The validity of the prediction model is verified by the tenfold cross-validation method, as can be seen from the data of the predictive indicators MAE and RMSE in Table 6:

- (1) The RF-ISSA-BPNN model has the lowest values of MAE and RMSE, which means that the error between the actual and predicted values is the smallest, so the RF-ISSA-BPNN model has the highest accuracy in predicting the compressive strength of concrete
- (2) The prediction performance of BPNN model is significantly enhanced after the optimized weights and thresholds, and the prediction results (MAE and RMSE) of RF-PSO-BPNN, RF-SSA-BPNN, RF-CSSA-BPNN, and RF-ISSABPNN models are better than RF-BPNN models
- (3) The significant difference between MAE and RMSE values of RF-ISSA-BPNN and ISSA-BPNN indicates the importance of RF feature selection. MAE is reduced by 0.6089, and RMSE is reduced by 0.873
- (4) The performance of the improved prediction model for SSA has also been further improved. MAE and RMSE of RF-CSSA-BPNN and RF-ISSA-BPNN models are all smaller than those of RF-SSA-BPNN model. It is fully verified that the improvement of the algorithm can also seek a breakthrough in the

performance of the prediction model, which also indirectly confirmed that the three improvement strategies proposed provide key help to the improvement of the original SSA

## 6. Conclusion

In this study, the standard sparrow search algorithm is improved by integrating LHS, somersault foraging strategy, and cyclone foraging mechanism to solve the problems of premature convergence and insufficient global search ability. In the prediction application of concrete compressive strength, the RF-ISSA-BPNN prediction model is established by finding the optimal threshold and weight by ISSA and using the feature selection function of RF. The main conclusions are as follows:

- (1) The ablation experiment on the CEC2017 benchmark function verifies that the three strategies introduced separately and simultaneously can significantly improve the optimization effect of SSA, which fully demonstrates the effectiveness and feasibility of the three strategies introduced
- (2) Compared with the classic algorithms (PSO, ABC, WOA, and SSA), ISSA has better optimization ability. In the case of the same time complexity of algorithm, the performance of ISSA is more outstanding than other improved algorithms (CSSA and NCSSA).
- (3) The importance of 8 features is sorted by RF, and 6 features with  $V_{IM} > 0.09$  are selected as input variables, so that the complexity of the model is reduced. It can be seen that there is a great correlation between age and compressive strength of concrete, which indirectly indicates that the effect of age on the compressive strength of concrete should be emphasized in practical application. The results of the longitudinal comparison between RF-ISSA-BPNN and ISSA-BPNN show that the feature selection of RF can make the model have more efficient running efficiency and higher prediction accuracy
- (4) The RF-ISSA-BPNN prediction model is horizontally compared with other models (RF-BPNN, RF-PSO-BPNN, RF-SSA-BPNN, and RF-CSSA-BPNN), the prediction results show that the MAE and RMSE values of RF-ISSA-BPNN are lower than those of other models, and the predicted data has the highest fitting degree with the actual data. It not only improves the shortcomings that BPNN is low prediction accuracy and easy to converge to different local

minima, but also enhances its accuracy and stability in prediction

Although RF-ISSA-BPNN can greatly reduce the risk of falling into local optimal solutions, the situation of falling into local minima still occurs in the later stage of the search. How to fundamentally eliminate the occurrence of this phenomenon and make the prediction model more characteristic and practical value will be the focus of our future work. In addition, due to the small number of samples, training the neural network model will be affected to some extent. In order to build a more accurate prediction model, a complete and higher quality data is needed. In the next research, the main work is to extend the database and introduce more factors related to the compressive strength of concrete to further improve the generalization ability of the model.

## Appendix

Data source: <http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

## Acknowledgments

This research was funded by the National Youth Science Foundation of China No. 62002046 and No. 61802040, the project supported by the Zhejiang Provincial Natural Science Foundation of China (No. LQ21F020005), and the Basic Public Welfare Research Program of Zhejiang Province (No. LGG18E050011).

## References

- [1] W. Liang, G. Wang, X. Ning et al., "Application of BP neural network to the prediction of coal ash melting characteristic temperature," *Fuel*, vol. 260, p. 116324, 2020.
- [2] B. Xu, H. C. Dan, and L. Li, "Temperature prediction model of asphalt pavement in cold regions based on an improved BP neural network," *Applied Thermal Engineering*, vol. 120, pp. 568–580, 2017.
- [3] Y. Liu, X. Wang, X. Zhu, and Y. Zhai, "Thermal error prediction of motorized spindle for five-axis machining center based on analytical modeling and BP neural network," *Journal of Mechanical Science and Technology*, vol. 35, no. 1, pp. 281–292, 2021.
- [4] S. Dai, B. Han, S. Liu, N. Li, F. Geng, and X. Hou, "Neural network-based prediction methods for height of water-flowing fractured zone caused by underground coal mining," *Arabian Journal of Geosciences*, vol. 13, no. 12, pp. 1–11, 2020.
- [5] Y. Li, J. Li, J. Huang, and H. Zhou, "Fitting analysis and research of measured data of SAW micro-pressure sensor based on BP neural network," *Measurement*, vol. 155, p. 107533, 2020.
- [6] J. Gao, "Performance evaluation of manufacturing collaborative logistics based on BP neural network and rough set," *Neural Computing and Applications*, vol. 33, no. 2, pp. 739–754, 2021.
- [7] Q. Liu, P. Sun, X. Fu et al., "Comparative analysis of BP neural network and RBF neural network in seismic performance evaluation of pier columns," *Mechanical Systems and Signal Processing*, vol. 141, p. 106707, 2020.
- [8] K. Dou and X. Sun, "Long-term weather prediction based on GA-BP neural network," *IOP conference series: earth and environmental science*, vol. 668, no. 1, article 012015, 2021.
- [9] H. Wang, Z. Zhang, and L. Liu, "Prediction and fitting of weld morphology of Al alloy-CFRP welding-rivet hybrid bonding joint based on GA-BP neural network," *Journal of Manufacturing Processes*, vol. 63, pp. 109–120, 2021.
- [10] P. Supraja, V. M. Gayathri, and R. Pitchai, "Optimized neural network for spectrum prediction using genetic algorithm in cognitive radio networks," *Cluster Computing*, vol. 22, no. S1, pp. 157–163, 2019.
- [11] H. Yuan, X. Wang, X. Sun, and Z. Ju, "Compressive sensing-based feature extraction for bearing fault diagnosis using a heuristic neural network," *Measurement Science and Technology*, vol. 28, no. 6, article 065018, 2017.
- [12] G. Jiang, M. Luo, K. Bai, and S. Chen, "A precise positioning method for a puncture robot based on a PSO-optimized BP neural network algorithm," *Applied Sciences*, vol. 7, no. 10, p. 969, 2017.
- [13] W. Wang, R. Tang, C. Li, P. Liu, and L. Luo, "A BP neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights," *Ocean Engineering*, vol. 162, pp. 98–107, 2018.
- [14] Y. Wu, R. Gao, and J. Yang, "Prediction of coal and gas outburst: a method based on the BP neural network optimized by GASA," *Process Safety and Environmental Protection*, vol. 133, pp. 64–72, 2020.
- [15] Z. Wensheng and J. Z. Haozqi, "BP neural network model for short-time traffic flow forecasting based on transformed grey wolf optimizer algorithm," *Journal of Transportation Systems Engineering and Information Technology*, vol. 20, no. 2, pp. 196–203, 2020.
- [16] H. Tian, P. Wang, K. Tansey, S. Zhang, J. Zhang, and H. Li, "An IPSO-BP neural network for estimating wheat yield using two remotely sensed variables in the Guanzhong Plain, PR China," *Computers and Electronics in Agriculture*, vol. 169, p. 105180, 2020.
- [17] L. Wu, Y. Yang, and M. Maheshwari, "Strain prediction for critical positions of FPSO under different loading of stored oil using GAIFOA-BP neural network," *Marine Structures*, vol. 72, p. 102762, 2020.
- [18] D. Lianjie, C. Degang, W. Ningling, and L. Zhanhui, "Key energy-consumption feature selection of thermal power systems based on robust attribute reduction with rough sets," *Information Sciences*, vol. 532, pp. 61–71, 2020.
- [19] L. Wen and X. Yuan, "Forecasting CO<sub>2</sub> emissions in Chinas commercial department, through BP neural network based on random forest and PSO," *Science of the Total Environment*, vol. 718, p. 137194, 2020.
- [20] J. K. Jaiswal and R. Samikannu, "Application of random forest algorithm on feature subset selection and classification and

- regression,” in *2017 world congress on computing and communication technologies (WCCCT)*, pp. 65–68, Tiruchirappalli, India, 2017.
- [21] X. K. Li, W. Chen, Q. Zhang, and L. Wu, “Building auto-encoder intrusion detection system based on random forest feature selection,” *Computers & Security*, vol. 95, p. 101851, 2020.
- [22] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, “Benchmark for filter methods for feature selection in high-dimensional classification data,” *Computational Statistics & Data Analysis*, vol. 143, p. 106839, 2020.
- [23] J. Xue and B. Shen, “A novel swarm intelligence optimization approach: sparrow search algorithm,” *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.
- [24] X. Lv, X. D. Mu, J. Zhang, and Z. Wang, “Chaos sparrow search optimization algorithm,” *Journal of Beijing University of Aeronautics and Astronautics*, vol. 47, pp. 1–10, 2020.
- [25] J. Li, J. Cheng, J. Shi, and F. Huang, “Brief introduction of back propagation (BP) neural network algorithm and its improvement,” in *Advances in computer science and information engineering*, pp. 553–558, Springer, Berlin, Heidelberg, 2012.
- [26] T. Y. Liu, *Improvement Research and Application of BP Neural Network*, Dongbei Agricultural University master degree thesis, 2011.
- [27] R. Genuer, J. M. Poggi, and C. Tuleau-Malot, “Variable selection using random forests,” *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2225–2236, 2010.
- [28] P. M. Chinta and M. N. Murty, “Discriminative feature analysis and selection for document classification,” in *International Conference on Neural Information Processing*, pp. 366–374, Berlin, Heidelberg, 2012.
- [29] D. Niu, K. Wang, L. Sun, J. Wu, and X. Xu, “Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: a case study,” *Applied Soft Computing*, vol. 93, p. 106389, 2020.
- [30] B. H. Menze, B. M. Kelm, R. Masuch et al., “A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data,” *BMC Bioinformatics*, vol. 10, no. 1, pp. 1–16, 2009.
- [31] P. Zhang, Z. Y. Yin, Y. F. Jin, and T. H. T. Chan, “A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest,” *Engineering Geology*, vol. 265, p. 105328, 2020.
- [32] W. Huo, W. Li, Z. Zhang, C. Sun, F. Zhou, and G. Gong, “Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and random forest feature selection,” *Energy Conversion and Management*, vol. 243, p. 114367, 2021.
- [33] A. M. Zalzal, P. J. Fleming, and P. Fleming, *Genetic Algorithms in Engineering Systems*, Iet, 1997.
- [34] M. D. McKay, R. J. Beckman, and W. J. Conover, “A comparison of three methods for selecting values of input variables in the analysis of output from a computer code,” *Technometrics*, vol. 42, no. 1, pp. 55–61, 2000.
- [35] W. Zhao, Z. Zhang, and L. Wang, “Manta ray foraging optimization: an effective bio-inspired optimizer for engineering applications,” *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103300, 2020.
- [36] W. Zhengtong, C. Fengqin, and L. S. Juventus, “Grey wolf optimization algorithm based on somersault foraging strategy,” *Computer Application Research*, vol. 38, no. 5, pp. 1434–1437, 2021.
- [37] W. Long, J. Jiao, X. Liang, and M. Tang, “An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization,” *Engineering Applications of Artificial Intelligence*, vol. 68, pp. 63–80, 2018.
- [38] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, Columbus, Ohio, USA, 2014.
- [39] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.
- [40] H. Wang, W. Wang, S. Xiao, Z. Cui, M. Xu, and X. Zhou, “Improving artificial Bee colony algorithm using a new neighborhood selection mechanism,” *Information Sciences*, vol. 527, pp. 227–240, 2020.
- [41] C. Zhang and S. Ding, “A stochastic configuration network based on chaotic sparrow search algorithm,” *Knowledge-Based Systems*, vol. 220, p. 106924, 2021.
- [42] Y. Wang, D. Niu, and L. Ji, “Short-term power load forecasting based on IVL-BP neural network technology,” *Systems Engineering Procedia*, vol. 4, pp. 168–174, 2012.
- [43] R. Hecht-Nielsen, “Kolmogorov’s mapping neural network existence theorem,” in *Proceedings of the international conference on neural networks*, vol. 3, pp. 11–14, New York, NY, USA, 1987.
- [44] V. Vakharia and R. Gujar, “Prediction of compressive strength and portland cement composition using cross-validation and feature ranking techniques,” *Construction and Building Materials*, vol. 225, pp. 292–301, 2019.
- [45] S. C. Hsieh, “Prediction of compressive strength of concrete and rock using an elementary instance-based learning algorithm,” *Advances in Civil Engineering*, vol. 2021, 10 pages, 2021.