

## Research Article

# SSA-BP Neural Network Model for Predicting Rice-Fish Production in China

Junlei Wang <sup>1</sup>, Guorui Zeng <sup>1</sup>, Maosen Xu <sup>1</sup>, Xuanchen Wan <sup>1</sup>, Keke Wang <sup>1</sup>,  
Jiegang Mou <sup>1</sup>, Chenchen Hua <sup>1</sup>, Chuanhao Fan <sup>1</sup> and Pengfei Han <sup>2</sup>

<sup>1</sup>College of Metrology Measurement and Instrument, China Jiliang University, Hangzhou 310018, China

<sup>2</sup>Rundian Energy Science and Technology Co., Ltd, Henan 450003, China

Correspondence should be addressed to Maosen Xu; [msxu@cjlu.edu.cn](mailto:msxu@cjlu.edu.cn)

Received 25 January 2024; Revised 17 April 2024; Accepted 22 May 2024

Academic Editor: Pronob Das

Copyright © 2024 Junlei Wang 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.

The rice-fish system has gained significant interest in recent years because of its effective usage of limited land and freshwater resources. To scientifically guide the improvement of rice field fishery production, the data in this study were selected from the latest China Fishery Statistical Yearbook, and therefore the development of paddy aquaculture was investigated. In order to more precisely predict the production of rice-fish in China, this paper introduces an artificial neural network with the SSA-BP model, which solves the drawbacks of the BP neural network such as easy to fall into local optimum and slow convergence speed when it is used for prediction. Firstly, the SSA-BP model incorporates the aquaculture area (split by water area), the national freshwater fish seedling output, the national end-of-year ownership of inland fishing vessels, the number of freshwater fisheries practitioners as input variables, and the production of rice-fish as an output variable; secondly, the SSA optimization algorithm was used to find the optimal initial thresholds and weights for the BP neural network, and finally the SSA-BP prediction model was constructed. The results revealed that the overall expansion of the rice field fishery was swift in the last five years, and the output of cultivated fish in China's rice fields rose by nearly 20% yearly in the past five years. Compared with the BP neural network and GA-BP models, the accuracy of the SSA-BP prediction model was enhanced by 61.01% and 16.15%, respectively, which was more suited for predicting the production of rice-fish.

## 1. Introduction

As a particular farming mode, fish culture in the paddy field makes maximum use of the limited water and soil resources by employing the freshwater resources in the paddy field and artificial help, finally providing the double benefits of rice and fish [1, 2]. China accounts for about 25% of the world's population, but it only has a quarter of the world's renewable water resources and a third of the world's per capita cultivated land area. At the same time, the shortage of arable land and freshwater resources impacts both freshwater aquaculture, which provides one-third of China's aquatic products, and rice, which makes up half of the country's staple food. As a result, the pairing of rice growing and aquaculture has garnered major attention in Chinese history [3]. The rice-fish paradigm has grown quickly in the last few

decades, and it is now used by almost 30 countries on 6 continents around the world [4]. China has a long history of fish growing on rice fields and has also brought in new improvements in recent years. Traditional fish farming in rice fields typically confronts hurdles such as solitary farming, restricted scale, and difficulty in raising productivity. The breeding mode is currently expanding in the direction of size, specialization, mechanization, and breeding diversification due to ongoing technological advancements. In establishing the rice-fish system, it is a problem worth paying attention to how to scientifically improve the productivity of fish culture in paddy fields [5]. The essential factors affecting the target yield usually come from production factors, while the yield of cultivated fish in paddy fields in the cross field of agriculture and fishery is generally affected by factors such as the breeding area, the

number of fish fry, the mechanized equipment, and the number of relevant employees. By mastering the production parameters and attaining a more accurate prediction of cultured fish yield in the paddy field, more scientific and theoretical guidance will be supplied to help people invest and allocate production resources. However, the selected components of production typically have complicated interconnections. So, finding such a complicated link is a hard task to be tackled at present.

The neural network is a convenient and effective way to establish the association between several elements [6]. Among various neural network models, the BP model is one of the typical representatives [7]. Zai et al. employed the BP neural network to estimate agricultural yield. The results showed that the model has potential utility in grain yield prediction [8]. However, Li et al. deployed a BP neural network to estimate the grain yield of Henan Province, found its limits, and suggested that it might be improved by employing the optimization technique [9]. This improved model will have better superiority and flexibility. The common BP neural network optimization approaches include PSO and GA. These algorithms need to be improved in terms of convergence speed, prediction accuracy, and flexibility even though they have shown some positive results [10, 11]. Jia et al. employed the sparrow search algorithm to optimize the BP neural network [12]. After multiple sets of trials, they showed that the SSA algorithm is more effective in increasing the performance of BP neural networks compared to other common algorithms.

This study will quickly evaluate the growth of rice field aquaculture in China in recent years. Then the effort seeks to employ the BP neural network to construct a prediction model of cultivated fish yield in the paddy field. To further boost model prediction accuracy, this research proposes to employ the sparrow search algorithm to optimize the prediction model and construct an SSA-BP model to forecast yield and compare the new prediction results with the prediction results of the BP neural network and GA-BP model.

## 2. Materials and Methods

**2.1. Data Selection.** The China Fishery Statistical Yearbook is a crucial resource for conducting research related to China's fisheries. It offers comprehensive statistics on the production, variety, dimensions, and costs of different fishery products. Through the examination of pertinent data and indicators in the yearbook, it is feasible to evaluate the production level, exploitation of resources, potential for growth, and trend of development in different fisheries in China. Hence, to guarantee the dependability and precision of the data, this study selected input and output parameters for the prediction model from the China Fishery Statistical Yearbook 2009–2022. Table 1 displays specific information. Subsequently, we conducted the following tasks based on fishery production factors. In predicting rice-fish production in China, four key factors were selected: the aquaculture area (split by water area), the national output of freshwater fish seedlings, the national ownership of inland fishing vessels at

the end of the year, and the number of practitioners in freshwater fisheries. The aquaculture area directly determines the available space and potential scale for fish farming, while the production of fish fry is essential for initiating and sustaining aquaculture activities. The number of inland fishing vessels indicates the capacity for resource management and environmental stability. Additionally, the workforce size reflects the scale and vitality of the local fishery community. Collectively, these factors provide a comprehensive understanding of the physical, technical, economic, and social dimensions influencing rice-fish production, thereby enhancing the accuracy and relevance of the predictive model. The four key production factors mentioned above show correlation with the variables in the model. However, the predicted target is rice-fish production. After 2009, the China Fishery Statistical Yearbook underwent changes in its statistical techniques and content, resulting in differences from prior editions. Each year's edition now documents the relevant state of China's fishery industry in the preceding year, due to the time-specific nature of the data. Some of the data in this research were obtained indirectly in order to address this challenge. For example, in some years:  $P_{fw} = P_f - P_{fm}$  ( $P_{fw}$ : the number of freshwater fishing employees,  $P_f$ : fishery employees, and  $P_{fm}$ : marine fishery employees).

Four production characteristics were selected for the linear regression fitting of the target output, as shown in Figure 1. The rice-fish aquaculture region with the best correlation was utilized to represent the aquaculture area. The determination coefficient  $R^2$  in Table 2 demonstrates that the linear regression model is unable to predict the production of farmed fish in rice fields. As can be seen from Table 2, the accuracy of fitting using the national freshwater fish seedling yield production factor was the lowest, while the accuracy of fitting the aquaculture area of rice-fish production factor was the highest, but fitting individual production factors by linear regression model alone could not accurately predict the rice-field yield of rice-fish culture. Therefore, in order to more properly anticipate the production of rice-fish in China, it is needed to take into account a more intelligent algorithm [13] and to analyze different production factors.

**2.2. BP Neural Network.** The BP neural network includes both the error backward propagation and forward propagation processes. It is a model that has been extensively investigated and applied in the field of evaluation and prediction [14]. The input, hidden, and output layers of the usual three-layer BP neural network structure are connected by network nodes, as seen in Figure 2 [15]. The connection weights ( $W_i$  and  $W_j$ ) between the layers, where  $W_i$  denotes the weight of the input layer to the implicit layer and  $W_j$  denotes the weight of the implicit layer to the output layer, are the most important parameters in the error back-propagation process. After the BP neural network gets the learning samples, the neuron activation values are transmitted from the input layer to the output layer through each hidden layer, and the optimal mapping relationship between

TABLE 1: Statistical data of fishery-related production factors in China.

Year	Aquaculture area (by water area) (hectare)					Rice-fish	National freshwater fish seedling yield (100 million tails)	National end-of-year inland fishing vessels (vessel)	Number of freshwater fishery practitioners (person)	Rice-fish farming production (ton)
	Pond	Lake	Reservoir	Brook	Other					
2008	2,144,715	961,335	1,549,612	202,183	113,178	1,477,501	6,873	343,636	11,190,876	1,169,968
2009	2,331,900	998,232	1,726,407	249,674	117,612	1,339,714	9,816	376,786	10,459,096	1,162,667
2010	2,377,001	1,007,103	1,795,579	264,126	120,534	1,326,113	3,607	377,436	10,546,071	1,242,732
2011	2,449,911	1,023,009	1,851,877	272,684	131,087	1,207,914	11,197	405,620	10,895,311	1,199,862
2012	2,566,859	1,024,785	1,911,468	274,817	129,547	1,294,919	11,181	415,105	10,601,100	1,331,378
2013	2,623,176	1,022,692	1,957,966	275,809	126,487	1,520,685	19,143	411,584	10,523,702	1,450,459
2014	2,661,901	1,015,327	1,994,819	274,965	133,876	1,489,501	12,746	409,313	10,437,477	1,456,719
2015	2,701,222	1,022,350	2,012,411	277,102	134,156	1,501,629	12,665	402,451	10,343,352	1,558,187
2016	2,762,604	990,816	2,010,928	267,694	147,577	1,516,093	13,005	392,966	10,058,153	1,632,263
2017	2,527,781	886,492	1,615,407	213,735	121,543	1,682,689	13,189	354,619	9,817,542	1,947,507
2018	2,666,835	746,155	1,441,670	179,414	112,381	2,028,262	13,110	323,171	9,514,123	2,333,269
2019	2,644,726	770,093	1,416,569	155,390	129,542	2,317,488	12,517	247,951	9,272,890	2,913,330
2020	2,625,404	720,648	1,420,871	147,447	126,186	2,562,686	13,097	160,740	8,901,561	3,249,109
2021	2,604,629	663,395	1,439,298	147,512	129,031	2,644,077	13,351	147,146	8,494,333	3,556,854

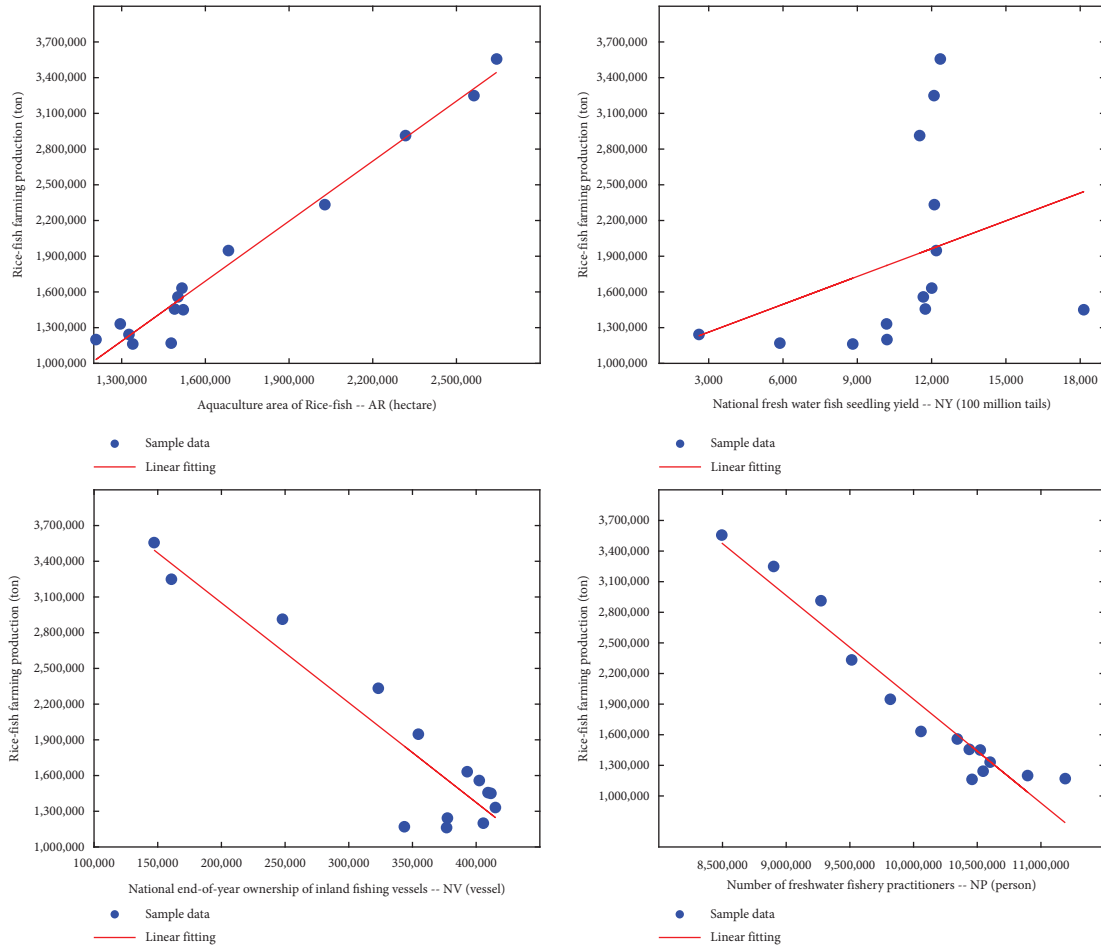


FIGURE 1: Linear fitting of production factors (AR, NY, NV, and NP).

TABLE 2:  $R^2$  of different production factors.

Production factors	AR	NY	NV	NP
$R^2$	0.97500	0.11280	0.87169	0.93709

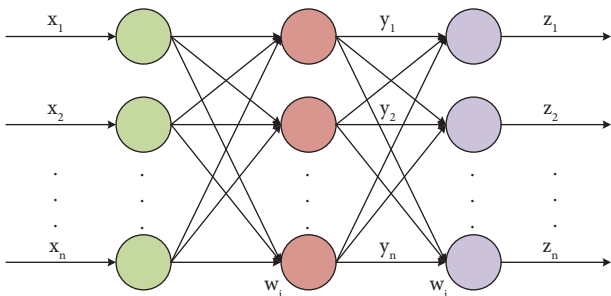


FIGURE 2: Structure of the three-layer BP neural network model.

the input and output variables is continuously established by incremental training. The neuron activation value is reversed from the output layer through the intermediate layer and back to the input layer based on the error between the target output and the actual value. This allows the weights and thresholds to be gradually adjusted until

a predetermined error accuracy or number of training times is reached [16].

In this study, a neural network comprising 9 input neurons, 5 hidden neurons, and 1 output neuron has been established, and 14 years' worth of data were separated into 9 groups for the training set and 5 groups for the test set at random. Other parameter choices included setting the learning rate to 0.1, the training goal residual to  $1 \times 10^{-5}$ , and the number of training sessions at 2000. Data preparation, which normalizes the training data using formula (1), is the initial stage in developing the model. The prediction model's accuracy is highly dependent on the number of hidden neurons chosen [17]. Neural network theory's Kolmogorov theorem also says that empirical formula (2) may usually be used to determine the number of hidden neurons  $h$  [18]. The model's hidden neurons can range from 4 to 14, according to the formula. To explore the training effects of many hidden neurons, the loop approach was applied [19]. By comparing the learning performance, the ideal number of hidden neurons is determined, and the results indicate that  $h = 5$  can improve the model's training. After calculating the required number of hidden neurons, the four main production factors in the training dataset were used as input variables for the training model, except for the rice-fish production which was used as the specified output variable. The results

demonstrate that the first training achieves optimal performance after 1762 steps, once the number of hidden neurons and input and output parameters have been established. The model's ideal performance essentially matches the training objective residuals and satisfies the predetermined parameters.

$$Y = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (1)$$

where  $Y$  is the normalized value of the initial sample data;  $X$  is the value of the initial data;  $X_{\max}$  is the maximum value in the initial dataset; and  $X_{\min}$  is the minimum value in the initial dataset.

$$h = \sqrt{m + n} + a, \quad (2)$$

where  $h$  is the number of hidden neurons;  $m$  is the number of input neurons;  $n$  is the number of output neurons; and  $a$  is an integer,  $a \in [1, 10]$ .

**2.3. GA Algorithm.** Genetic algorithm [20] is an optimization search strategy based on natural selection and genetics. The fundamental principle involves emulating the biological concept of "survival of the fittest" and evolutionary laws, wherein the problem is treated as a simulation of biological evolution. Through iterative updates, solutions to the problem are continuously refined, aiming to find an optimal or near-optimal solution.

The GA algorithm first initializes the population that may exist in the problem to be solved utilizing genetic coding and then, according to the fitness size of each individual in the population, retains the individuals with good fitness values, eliminates the individuals with poor fitness values, and combines crossover and mutation of the individuals with the help of the genetic operator in natural genetics to produce the population that represents the new solution set. The basic parameters of the genetic algorithm are established as indicated in Table 3, the maximum number of iterations is 60, the population size is 30, the position dimension is 56, the crossover probability is 0.8, and the mutation probability is 0.05. The three fundamental phases of the GA algorithm are as follows.

**2.3.1. Selection Stage.** The best members of the existing population are chosen for the selection stage with the intention of increasing the possibility that they will become parents, generate better offspring, and improve the likelihood that they will search for the global optimum.

**2.3.2. Crossover Stage.** By matching population members two to one and transferring a portion of their chromosomes based on a specific crossover probability  $P_c$  for each pair, the crossover stage aims to produce a new generation of individuals.

**2.3.3. Mutation Stage.** The goal of the mutation stage is to change the gene values on one or more loci for each individual in the population with a given mutation probability  $P_m$ , which gives an opportunity to generate new people and

improves the probability of the algorithm jumping out of the local optimum.

**2.4. SSA Optimization Algorithm.** In 2020, SSA was put forth. The foraging and anti-predation behaviors of sparrows were the source of inspiration for this algorithm [21]. There are two typical sparrow predation strategies in nature, the finder's strategy and the joiner's strategy. Individuals in the population will monitor the behavior of other individuals. At the same time, participants in the population compete for food with individuals in the population with high food intake. When sparrows feel danger, they will also have related anti-predator behavior [22]. After that, the monitor concept will be explained. There will be three positions in the population: discoverer, joiner, and monitor. The monitor, who makes up 10% to 20% of the population, is often picked at random and shares the monitor's responsibilities with the discoverer and joiner. Although the identities of the discoverer and joiner are dynamic and variable, their relative proportions never alter [23].

Specifically, the position of the sparrow can be expressed by the matrix of formula

$$x = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,d} \end{bmatrix}, \quad (3)$$

where  $d$  is the dimension of the variable that needs to be optimized and  $n$  is the number of sparrows in the population. Simultaneously, formula (4) expresses the fitness factors of every sparrow.

$$F_x = \begin{bmatrix} f[x_{1,1} & x_{1,2} & \cdots & x_{1,d}] \\ f[x_{2,1} & x_{2,2} & \cdots & x_{2,d}] \\ \vdots & \vdots & \ddots & \vdots \\ f[x_{n,1} & x_{n,2} & \cdots & x_{n,d}] \end{bmatrix}. \quad (4)$$

Within the context of  $F_x$ , every row vector represents the fitness of an individual. During the foraging process of sparrows, the individual with the superior fitness value will be the first to obtain food. The discoverer has a greater capacity to locate food over a larger territory compared to the joiner, as they are tasked with locating prey for the entire group and guiding their collective movement. The location of the discoverer will be changed after each repetition, and this update process is illustrated by

$$x_{i,j}^{k+1} = \begin{cases} x_{i,j}^k \cdot e^{(-i/\alpha \cdot \text{iter}_{\max})}, & \text{if } R < ST, \\ x_{i,j}^k + Q \cdot L, & \text{if } R \geq ST, \end{cases} \quad (5)$$

where  $k$  denotes the current iteration,  $x_{i,j}^{k+1}$  denotes the value of the  $j$ th dimension of the  $i$ th sparrow after  $k$  iterations, and  $\text{iter}_{\max}$  represents the maximum number of iterations. The variable  $\alpha$  is a randomly generated number that falls within the range of values between 0 and 1. The sparrow's alarm value and safety value are represented by the values  $R \in [0, 1]$  and  $ST \in [0.5, 1.0]$ , respectively.  $L$  is a matrix of

TABLE 3: Parameters of the BP neural network, the GA-BP model, and the SSA-BP model.

Algorithm	Parameter	Definition	Value
BP	m	The number of input neurons	9
	h	The number of hidden neurons	5
	n	The number of output neurons	1
	$\mu$	Learning rate	0.1
	t	The number of iterations	2000
	$\delta$	Training target residual	$1 \times 10^{-5}$
GA	T	The maximum number of iterations	60
	N	The population size	30
	D	The location dimension	56
	$P_c$	The crossover probability	0.8
	$P_m$	The mutation probability	0.05
SSA	T	The maximum number of iterations	60
	N	The population size	30
	D	The location dimension	56
	P_percent	The number of monitors as a proportion of population size	0.2

dimensions  $1 \times d$ , where each element has a value of 1. Q is a random number that is distributed according to the normal distribution. When  $R < ST$ , which signals that there are no predators around the sparrow population, the discoverer changes to a broad search mode.  $R \geq ST$  signals that some sparrows have identified predators, and all sparrows must promptly flee to other safe sites.

The discoverer is often watched over by participants. Participants will immediately leave their existing locations to compete for food once the discoverer finds better food, and the winner of the competition will receive the explorer's food. The update strategy algorithm is presented in the following formula:

$$x_{i,j}^{k+1} = \begin{cases} Q \cdot e^{\left(\frac{x_{\text{worst}}^k - x_{i,j}^k}{\alpha \cdot \text{iter}_{\text{max}}}\right)}, & \text{if } i > \frac{n}{2}, \\ x_p^{k+1} + |x_{i,j}^k - x_p^{k+1}| \cdot A^+ \cdot L, & \text{otherwise,} \end{cases} \quad (6)$$

where  $x_p$  is the best position occupied by all current explorations.  $x_{\text{worst}}$  denotes the current global worst position. A is a  $1 \times d$  dimensional matrix, in which each element is randomly assigned 1 or -1 and meets  $A^+ = A^T (AA^T)^{-1}$ . If  $i > n/2$ , it signifies that the  $i$ th follower in the population, who has a low moderate value, is starving and must move to another location to acquire food.

The final one affects the monitor. It is generated at random in its initial location. The location update strategy is shown in the following formula:

$$x_{i,j}^{k+1} = \begin{cases} x_{\text{best}}^k + \beta \cdot |x_{i,j}^k - x_{\text{best}}^k|, & \text{if } f_i > f_b, \\ x_{i,j}^k + K \cdot \left( \frac{|x_{i,j}^k - x_{\text{worst}}^k|}{(f_i - f_w) + \varepsilon} \right), & \text{if } f_i = f_b, \end{cases} \quad (7)$$

where  $x_{\text{best}}$  is the current global optimal location.  $\beta$  is a random number with a mean of 0 and a variance of 1, satisfying the random distribution. A random number of

$K \in [-1, 1]$ . Here  $f_i$  is the fitness value of the current sparrow;  $f_b$  is the current best fitness value; and  $f_w$  represents the current worst fitness value.  $\varepsilon$  is the smallest constant to avoid causing the denominator to be zero. For simplicity, when  $f_i > f_b$ , it indicates that sparrows are on the edge of the group. And  $x_{\text{best}}$  represents the location of the sparrow population center, surrounded by safe areas. When  $f_i = f_b$ , it indicates that the sparrow in the middle of the population is aware of the danger, and the sparrow in this position needs to update its position to get close to other sparrows. K is the direction of sparrow movement and is also the step control coefficient.

As shown in Table 3, several of the settings in the GA-BP model and SSA-BP model are compatible with the non-optimized BP neural network, notably including establishing 5 hidden neurons and a learning rate of 0.1 and limiting the residual error of the training goal to  $1 \times 10^{-5}$ . Furthermore, the location dimension was set to 56, the maximum evolutionary algebra to 60, and the SSA population size to 30, of which 20% were monitors. Based on the aforementioned setup parameters, the optimization program flow of the sparrow search algorithm for the BP neural network is illustrated in Figure 3. Firstly, the SSA optimization algorithm initializes the population. After receiving the weights and thresholds of the BP neural network, the fitness size of the population is determined. Then, the position of the discoverer, participant, and monitor is updated using formulas (5)–(7). After completing one iteration, the fitness of different individuals is recalculated, and the best individual's position information is selected based on the fitness size. Finally, after the iteration is completed, the optimal result, weight, and threshold are output.

### 3. Results

**3.1. Analysis of the Development of the Rice-Fish Industry in China.** Fishery statistics can be utilized to construct Figure 4. The most crucial one was that, between 2016 and 2021, the breeding area of the rice field fish industry expanded quickly, and the output of cultured fish in rice fields

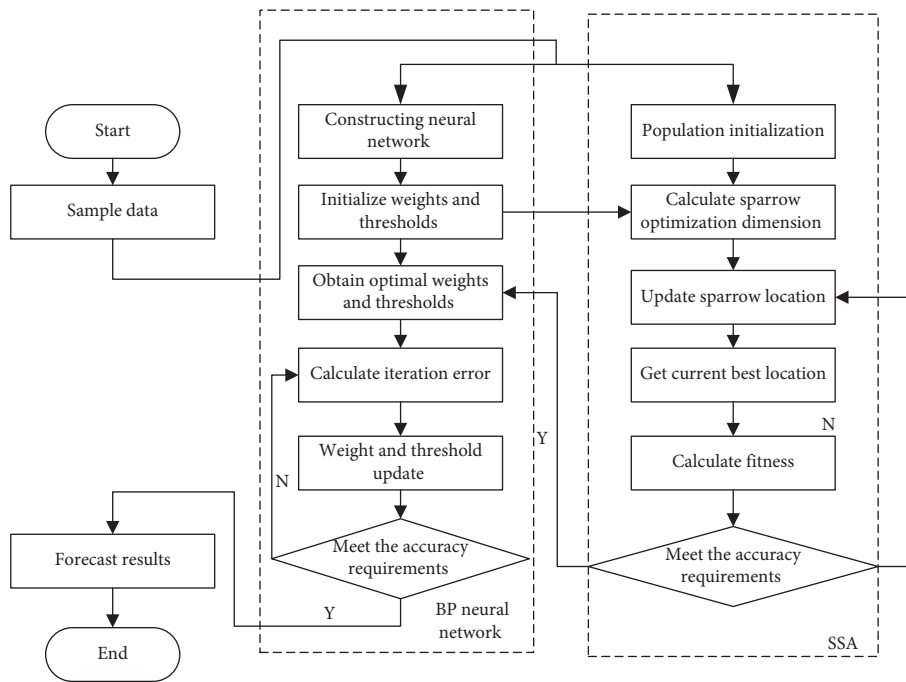


FIGURE 3: SSA optimizes the BP neural network process.

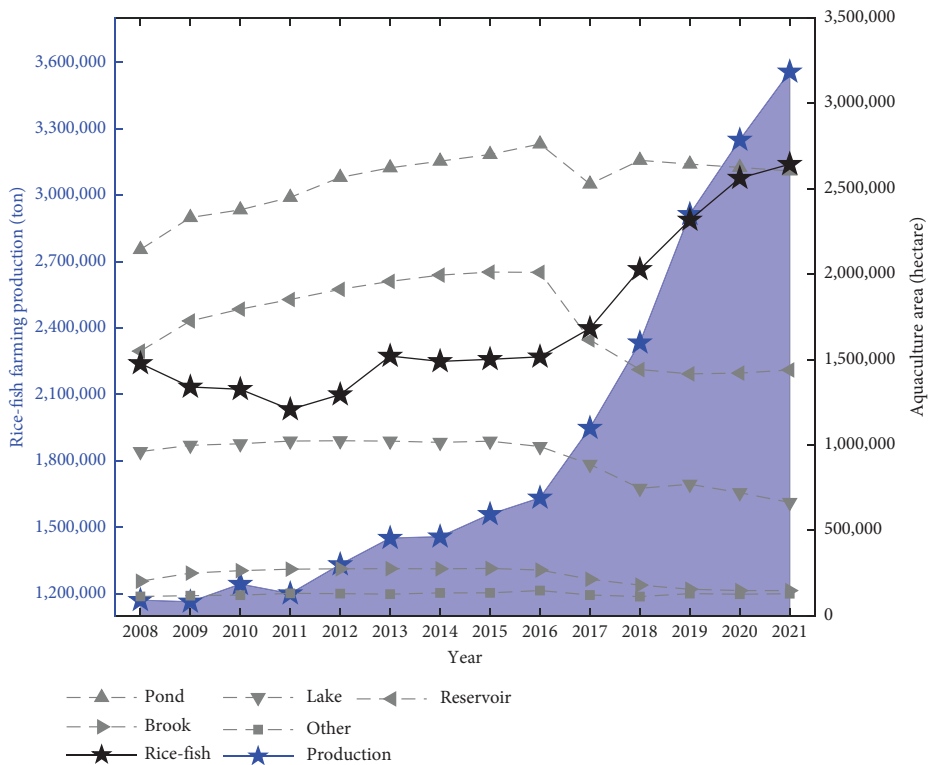


FIGURE 4: Trend chart of rice-fish aquaculture production and aquaculture area.

followed the pace. The data supported the determination coefficient  $R^2$  in Table 2 and revealed a distinct correlation between the rice field's cultivated area and yield. However, the paddy field aquaculture area increased annually from 2016 to 2021 at rates of 11%, 21%, 14%, 11%, and 3%, respectively, whereas the similar growth rate for rice-fish

output during the same time was 19%, 20%, 25%, 12%, and 9%. From the results, even though the modification in the breeding area had a large impact on the yield, its speed trend was not related. As a result, it was difficult to measure the yield accurately by depending simply on a single breeding area, and extra factor analysis was required.

This paper chose influential variables from four aspects based on the standpoint of production factors. Initially, a comprehensive examination of the cultural region that is essential for cultivating fish in rice fields was undertaken. The scarcity of geographic resources restricts the overall breeding area on a large scale, and the positioning of several breeding zones also mutually influences one another. Thus, one of the factors that had an impact was the diversity of breeding habitats. The surface area of ponds, lakes, reservoirs, brooks, and other conventional aquaculture systems decreased by 6%, 33%, 28%, 45%, and 13% accordingly, from 2016 to 2021, as depicted in Figure 4. Nevertheless, the rice-fish aquaculture sector has had significant growth throughout the same timeframe. The findings indicate the presence of an intricate translational connection between regions. Consequently, changes in the different aquaculture regions should be considered when making projections about productivity.

Then, other production factors were incorporated as the fundamental influence quantity, such as the output of freshwater fish fry (because fish fry was the fundamental influence quantity corresponding to the output and the vast majority of China's rice field systems were freshwater systems), the ownership of inland fishing boats (in actuality, the utilization rate of fishing boats in the rice field system was low, but this element represents the function of technical instruments in output production), and the freshwater fishery practitioners (it can demonstrate the impact of personnel on the industry, and this factor indirectly influences output) [24, 25]. Figure 5 represents the changes in other factors of production and rice-fish aquaculture production since 2008, from which it can be seen that from 2016 to 2022, national freshwater fish seedling yield increased by 3% and national end-of-year ownership of inland fishing vessels and number of freshwater fishery practitioners decreased by 63% and 16%, respectively, while rice-fish production increased rapidly during this period, indicating that the other basic factors of production are also not directly related to rice-fish production, and further use of SSA-BP neural network is needed to fit their relationship.

**3.2. Prediction Effect of BP Neural Network.** First, the tested data were utilized to verify the validity of the model. In Figure 6(a), the  $x$ -axis coordinate sample points represent nine randomly selected years, from left to right, 2021, 2013, 2014, 2010, 2020, 2012, 2016, 2015, and 2017, and in Figure 6(b), the  $x$ -axis, from left to right, is 2019, 2008, 2018, 2011, and 2009, and the  $y$ -axis coordinates represent the corresponding years of rice-fish yield. The real value, predicted value, and absolute error of the training samples of cultivated fish production in rice fields can be seen in Figure 6(a), while the calculated values of the relative errors were 1.78%, 0.74%, 3.26%, 0.77%, 1.38%, 0.19%, 6.87%, 0.64%, and 0.11%, respectively. The average relative error was only 1.75%. The results show that the BP model had good reduction performance for data, and the model was

accurate. As shown in Figure 6(b), the actual value and the predicted value were compared. The results show that the BP neural network can predict the output, and the calculated relative errors were 2.57%, 11.61%, 17.93%, 5.80%, and 10.70%. It can be seen from the results that the BP model may have a large difference in the prediction of individual samples of cultivated fish in rice fields, and the average relative error was as high as 9.72%.

**3.3. GA-Optimized BP Neural Network.** Figure 7 represents the training and prediction results of the GA-BP model, where the  $x$ -axis coordinate sample points are consistent with Figure 6. Figure 7(a) represents the relative error value corresponding to each year after the GA-BP model is trained, and its average value is 1.35%. Figure 7(b) shows the relative error values corresponding to each year after the GA-BP model is predicted, which are 3.69%, 4.83%, 8.14%, 3.25%, and 2.68%, with an average error of 4.52%. The results showed that the training and prediction performance of the GA-BP model for rice-fish production was improved compared to the BP neural network by 22.86% and 53.29%, respectively.

**3.4. SSA-Optimized BP Neural Network.** To visually compare the optimization effect of SSA, this section defines the training and validation data to obtain the effect map under the same data sample. First, it can be seen from the population iteration curve that the optimization was effective, as shown in Figure 8. When the population evolved to 18 generations, the best fitness was achieved, and in fact, the establishment of the optimized neural network model was completed with only 7 iterations of training.

Figure 9 displays the SSA-BP model's prediction effect, where the  $x$ -axis coordinate sample points are consistent with Figure 6. This model also had some influence on the forecast for farmed fish in rice fields. The average relative error of the model training, as displayed in Figure 9(a), was 0.66%. The outcome demonstrated the model's accuracy but also highlighted the BP model's superior reducibility for the training set of data. The prediction impact in Figure 9(b) shows that the average error was 3.79% and that the individual prediction errors were 2.82%, 2.25%, 7.41%, 2.29%, and 4.16%. According to relative error, the SSA-BP model performed noticeably better in training and forecasting the yield of farmed fish in rice fields than the BP neural network and GA-BP model, as shown in Figure 10. From the selected samples, it can be seen that the average training performance is relatively improved by 62.29% and 51.11% and the prediction performance is relatively improved by 61.01% and 16.15%. Figures 11(a), 11(b), and 11(c) represent, respectively, the regression curves of the BP neural networks, GA-BP model, and SSA-BP fitted to the rice-fish production from 2008 to 2021, where the  $x$ -axis represents

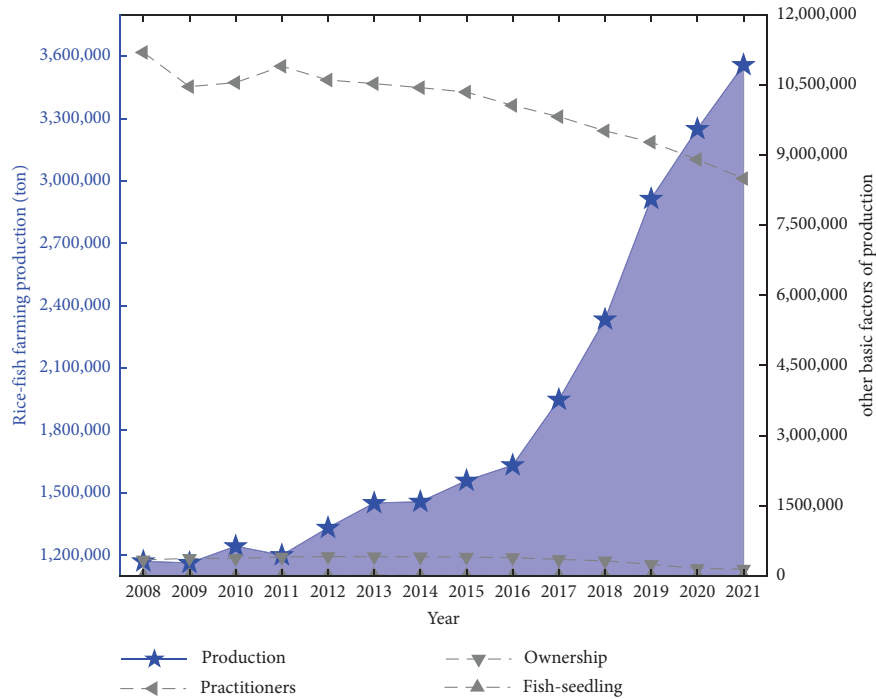


FIGURE 5: Trend chart of rice-fish aquaculture production and other basic factors of production.

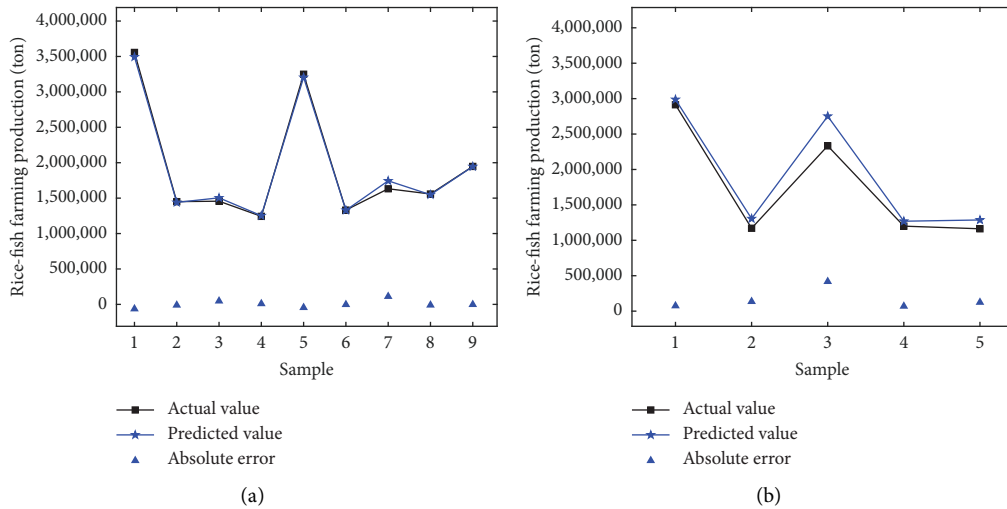


FIGURE 6: BP model training and prediction effect. (a) Training. (b) Prediction.

the target value and the  $y$ -axis represents the output value, both normalized. In addition, the correlation coefficient  $R$  of these three models can be seen in the figure as 0.96448, 0.98343, and 0.99788, respectively, and the closer the value is to 1 indicates a superior fitting effect. The results showed that the SSA-BP model provided a better fit than the BP neural network and the GA-BP model, confirming the feasibility of SSA-BP for rice-fish yield prediction.

The mean absolute error (MAE), mean squared error (MSE), and root mean square error (RMSE) are the next three error assessment criteria that examine each model's viability and efficacy. The following are the formulas:

$$MAE = \frac{1}{N} \sum_{t=1}^N |y(t) - \hat{y}(t)|, \tag{8}$$

$$MSE = \frac{1}{N} \sum_{t=1}^N (\hat{y}(t) - y(t))^2, \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{y}(t) - y(t))^2}, \tag{10}$$

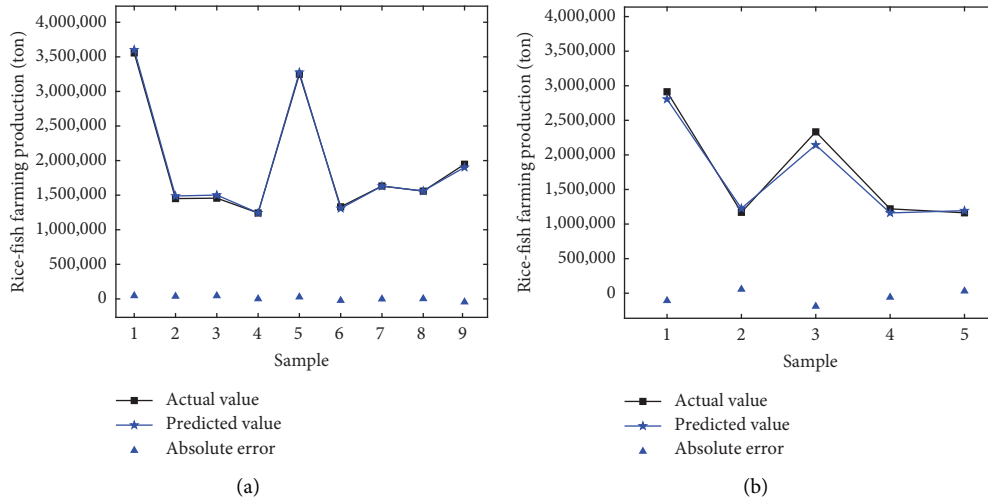


FIGURE 7: GA-BP model training and prediction effect. (a) Training. (b) Prediction.

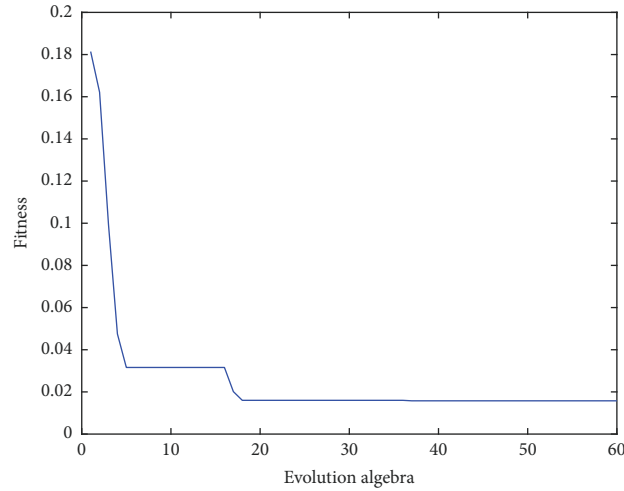


FIGURE 8: Population evolution curve.

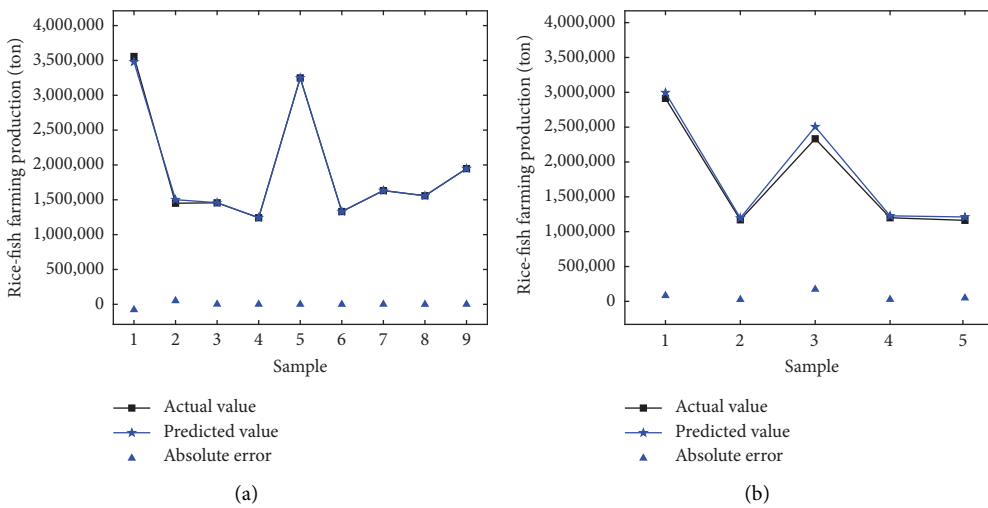


FIGURE 9: SSA-BP model training and prediction effect. (a) Training. (b) Prediction.

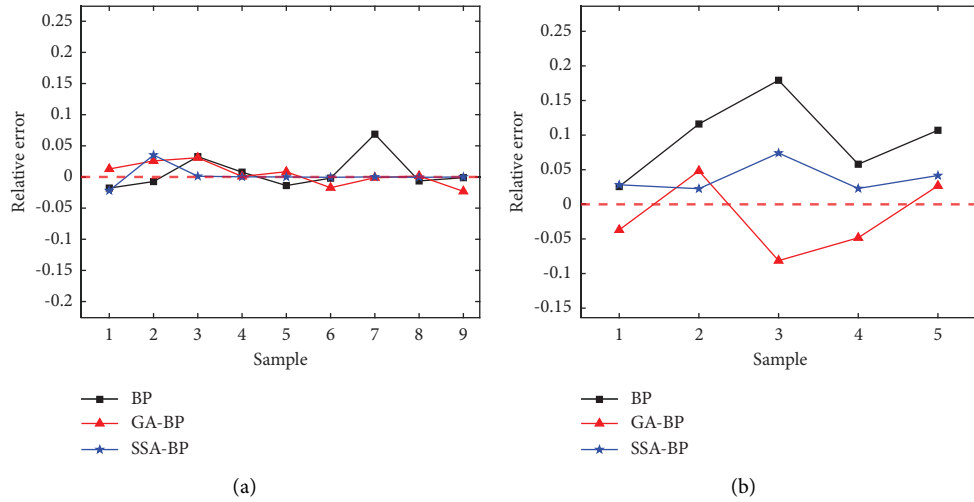


FIGURE 10: Comparison diagram of relative error of the model. (a) Training relative error. (b) Prediction relative error.

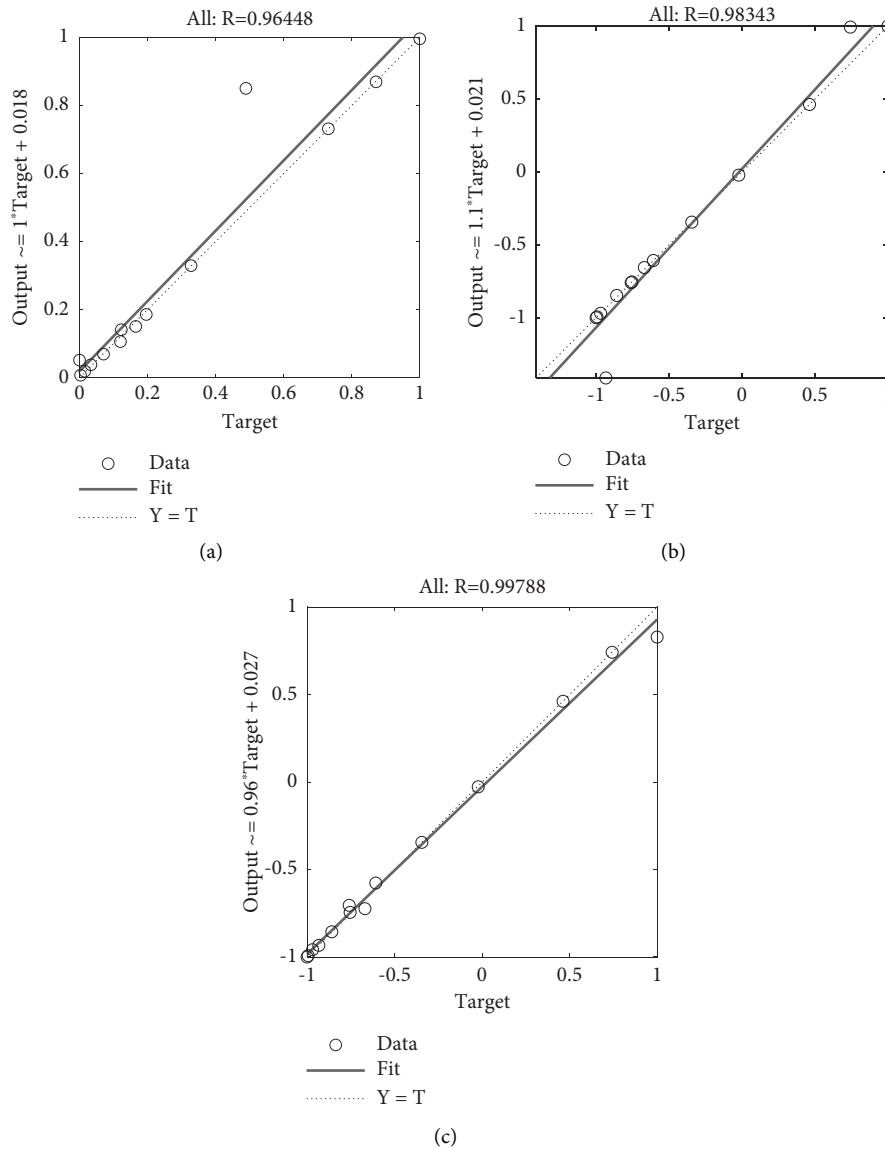


FIGURE 11: Regression curves for BP, GA-BP, and SSA-BP neural network. (a) BP neural network. (b) GA-BP model. (c) SSA-BP model.

TABLE 4: Comparison of rice-fish farming production prediction error results between BP neural network, GA-BP, and SSA-BP models.

Predictive models	MAE	MSE	RMSE
BP	164573.50	43858270178	209423.70
GA-BP	84792.11	10657362996	103234.50
SSA-BP	71477.19	8087771253	89932.04

where  $N$  is the number of samples,  $y(t)$  is the true value of rice-fish farming production, and  $\hat{y}(t)$  is the predicted value of rice-fish farming production.

Table 4 lists the errors in predicting rice-fish yield using the BP neural network, GA-BP model, and SSA-BP model. The relative MAE, MSE, and RMSE of the BP neural network and GA-BP model versus the SSA-BP model were 93096.31, 35770498925, and 119491.66 and 13314.92, 2569591743, and 13302.46, respectively, which is an improvement of 56.57%, 81.56%, and 57.06% and 15.70%, 24.11%, and 12.89%. The results show that the SSA-BP model prediction model has lower performance indicators than the BP neural network and GA-BP model, the prediction accuracy is higher, and the prediction error is smaller, which verifies that the SSA-BP proposed in this paper has good prediction ability for rice-fish yield.

However, when testing the samples, the SSA-BP model performed the same as the samples with a high level of large errors in the BP neural network and GA-BP models. According to the sample query, this significant error point appeared in 2018. Looking back at 2018, with its record-breaking weather extremes, the aquaculture business has been greatly impacted by different changes in the ecological environment, which has led to the increased diversity of factors affecting aquaculture productivity in that year. Climate change, resource constraints, and the COVID-19 pandemic will all provide substantial hurdles to production forecasting in the upcoming years [21]. Therefore, it may be challenging to accurately predict the production of paddy-farmed fish using basic production factors, and the SSA-BP prediction model needs to be continuously improved to enhance its operational performance and prediction accuracy. Consequently, the data-driven machine learning strategy continues to be an effective yield forecast tool overall and can also be applied in the future in engineering fields such as path prediction, image segmentation, and other related disciplines.

## 4. Conclusions

In this study, we briefly introduced the latest development of rice-fish farming in China and proposed a yield prediction model using the SSA algorithm to optimize the BP neural network for Chinese rice-fish farming, which overcame the shortcomings of the traditional BP neural network such as easy to fall into local optimum and slow convergence and was able to train the overall samples quickly and obtain accurate prediction results. The SSA-BP prediction model can effectively obtain the mapping relationship between the four main production factors and rice-fish yield, and the fitting accuracy is better than that of the BP neural network

and GA-BP model. The experiment shows that the average relative error of the SSA-BP prediction model is 3.79%, which is 61.01% and 16.15% lower than that of the BP neural network and GA-BP model, respectively, indicating that the optimization effect of SSA algorithm is obvious, and it can provide some theoretical suggestions for the future industrialization application.

The SSA-BP model developed based on this prediction plays a strong role in small sample data, but individual data points may have large errors. In addition, only four main production factors related to rice-fish production were considered in this paper, and we will collect as many data samples as possible and consider more influencing factors, such as geologic hazards, aquatic product trade, and aquatic product processing, in our future research in order to expand the training dataset for a more comprehensive and reasonable prediction of rice-fish production in rice farming.

## Data Availability

The data are from the China Fishery Statistical Yearbook 2009–2022.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Authors' Contributions

J.L.W. and G.R.Z. were responsible for software. J.L.W. and C.C.H. were responsible for data curation. J.L.W. and C.H.F. were responsible for formal analysis. G.R.Z. was responsible for original draft preparation. G.R.Z. and M.S.X. were responsible for review and editing. M.S.X. was responsible for funding acquisition. X.C.W. and K.K.W. were responsible for supervision. J.G.M. and P.F.H. were responsible for project administration.

## Acknowledgments

The authors are very grateful to the Institute of Fluid Equipment and Inspection Technology, China Jiliang University, for supporting this study and all the students and teachers who helped during the investigation, as well as the financial support from the Fundamental Research Funds for the Provincial Universities of Zhejiang under grant no. 2023YW44.

## References

- [1] G. Aditya, S. Pal, and G. Saha, "An assessment of fish species assemblages in rice fields in West Bengal, India: implications for management," *Journal of Applied Ichthyology*, vol. 26, no. 4, pp. 535–539, 2010.
- [2] S. A. Noorhosseini-Niyaki and F. Bagherzadeh-Lakani, *Ecological and Biological Effects of Fish Farming in rice fields, Regional Congress of Sustainable Management Science-Based in Agriculture and Natural Resources*, Gorgan University of Agricultural Sciences and Natural Resources, Iran, 2011.

- [3] E. T. Yeh and F. Li, "Of crayfish, rice, and anxiety: agricultural modernization in chongzhou, sichuan," *Global Food History*, vol. 8, no. 3, pp. 232–253, 2022.
- [4] L. Hu, J. Tang, J. Zhang et al., "Development of rice-fish system: today and tomorrow," *Chinese Journal of Eco-Agriculture*, pp. 268–275, 2015.
- [5] J. Pretty, C. Toulmin, and S. Williams, "Sustainable intensification in African agriculture," *International Journal of Agricultural Sustainability*, vol. 9, no. 1, pp. 5–24, 2011.
- [6] U. Junta, L. Newiduum, A. Opuiyo, and I. Browndi, "Predictive analysis of urban planning for through the operation of artificial cloud network," *International Journal of Science and Advanced Technology*, vol. 62, pp. 622–627, 2022.
- [7] E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks," *SN Applied Sciences*, vol. 1, no. 9, pp. 982–1010, 2019.
- [8] S. Zai, D. Guo, J. Wen, H. Luo, and R. Dong, "Study on BP neural network model for crop yield prediction," *YELLOW RIVER*, vol. 32, pp. 71–72, 2010.
- [9] B. Li, Y. Zhang, S. Zhang, and W. Li, "Prediction of grain yield in Henan Province based on grey BP neural network model," *Discrete Dynamics in Nature and Society*, vol. 2021, pp. 1–13, Article ID 9919332, 2021.
- [10] A. Abraham and V. Rohini, "A particle swarm optimization-backpropagation (PSO-BP) model for the prediction of earthquake in Japan," *Emerging Research in Computing, Information, Communication and Applications*, Springer, Cham, 2019.
- [11] M. Jahangiri and F. Razaghian, "Fault detection in analogue circuits using hybrid evolutionary algorithm and neural network," *Analog Integrated Circuits and Signal Processing*, vol. 80, no. 3, pp. 551–556, 2014.
- [12] B. Jia, W. Zhu, R. Wang, S. Gao, Y. Hu, and H. Wang, "The assessment of storm surge disaster loss based on SSA-BP neural network model," *Marine Forecasts*, vol. 39, pp. 50–58, 2022.
- [13] C. S. W. Ng, A. J. Ghahfarokhi, and M. N. Amar, "Well production forecast in Volve field: application of rigorous machine learning techniques and metaheuristic algorithm," *Journal of Petroleum Science and Engineering*, vol. 208, 2022.
- [14] B. Ruan, "Prediction of stock market by BP neural network model," *Journal of Physics: Conference Series*, vol. 1744, no. 4, p. 042232, 2021.
- [15] Ö. Can, T. Baklacioglu, E. Özturk, and O. Turan, "Artificial neural networks modeling of combustion parameters for a diesel engine fueled with biodiesel fuel," *Energy*, vol. 247, 2022.
- [16] B. Alankar, N. Yousf, and S. U. Ahsaan, "Predictive analytics for weather forecasting using back propagation and resilient back propagation neural networks," *New Paradigm in Decision Science and Management*, Springer, Cham, 2020.
- [17] A. Jahani and M. Saffariha, "Human activities impact prediction in vegetation diversity of Lar National Park in Iran using artificial neural network model," *Integrated Environmental Assessment and Management*, vol. 17, no. 1, pp. 42–52, 2021.
- [18] X. Deng, T. Xu, and R. Wang, "Risk evaluation model of highway tunnel portal construction based on BP fuzzy neural network," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–16, 2018.
- [19] M. Xu, G. Zeng, D. Wu et al., "Structural optimization of jet fish pump design based on a multi-objective genetic algorithm," *Energies*, vol. 15, no. 11, p. 4104, 2022.
- [20] J. H. Holland, "Genetic algorithms," *Scientific American*, vol. 267, no. 1, pp. 66–72, 1992.
- [21] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems Science and Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.
- [22] M. Caixia, W. Di, and L. Yu, "BP neural network for satellite clock bias prediction based on sparrow search algorithm," *Journal of Geodesy and Geodynamics*, vol. 42, pp. 125–131, 2022.
- [23] G. Wang, W. Wang, K. Cheng et al., "Hyperspectral prediction of gelatinization characteristics of millet flour based on sparrow search Optimized BP algorithm," *Food Science*, pp. 1–10, 2021.
- [24] E. W. Erliani, M. Marto, and E. Tanjaya, "Fishing equipment technical efficiency for fisheries production in banda naira," in *Proceedings of the IOP Conference Series: Earth and Environmental Science*, Southeast Maluku, Indonesia, October 2020.
- [25] X. Li, D. Chen, X. Cao, Z. Luo, and M. Webber, "Assessing the components of, and factors influencing, paddy rice water footprint in China," *Agricultural Water Management*, vol. 229, 2020.