

## Article

# A Power Transformer Fault Diagnosis Method Based on Improved Sand Cat Swarm Optimization Algorithm and Bidirectional Gated Recurrent Unit

Wanjie Lu <sup>1,2</sup>, Chun Shi <sup>1,\*</sup>, Hua Fu <sup>1</sup> and Yaosong Xu <sup>1</sup><sup>1</sup> School of Electrical Control, Liaoning Technical University, Huludao 125000, China<sup>2</sup> School of Mechanical Engineering, Liaoning Technical University, Fuxin 123000, China

\* Correspondence: 472120439@stu.lntu.edu.cn

**Abstract:** The bidirectional gated recurrent unit (BiGRU) method based on dissolved gas analysis (DGA) has been studied in the field of power transformer fault diagnosis. However, there are still some shortcomings such as the fuzzy boundaries of DGA data, and the BiGRU parameters are difficult to determine. Therefore, this paper proposes a power transformer fault diagnosis method based on landmark isometric mapping (L-Isomap) and Improved Sand Cat Swarm Optimization (ISCSO) to optimize the BiGRU (ISCSO-BiGRU). Firstly, L-Isomap is used to extract features from DGA feature quantities. In addition, ISCSO is further proposed to optimize the BiGRU parameters to build an optimal diagnosis model based on BiGRU. For the ISCSO, four improvement methods are proposed. The traditional sand cat swarm algorithm is improved using logistic chaotic mapping, the water wave dynamic factor, adaptive weighting, and the golden sine strategy. Then, benchmarking functions are used to test the optimization performance of ISCSO and the four algorithms, and the results show that ISCSO has the best optimization accuracy and convergence speed. Finally, the fault diagnosis method based on L-Isomap and ISCSO-BiGRU is obtained. Using the model for fault diagnosis, the example simulation results show that using L-ISOMP to filter and downscale the model inputs can better improve model performance. The results are compared with the SCSO-BiGRU, WOA-BiGRU, GWO-BiGRU, and PSO-BiGRU fault diagnosis models. The results show that the fault diagnosis rate of ISCSO-BiGRU is 94.8%, which is 11.69%, 10.39%, 7.14%, and 5.9% higher than that of PSO-BiGRU, GWO-BiGRU, WOA-BiGRU, and SCSO-BiGRU, respectively, and validate that the proposed method can effectively improve the fault diagnosis performance of transformers.

**Keywords:** power transformer; fault diagnosis; landmark isometric mapping; bidirectional gated recurrent unit; improved sand cat swarm optimization algorithm



**Citation:** Lu, W.; Shi, C.; Fu, H.; Xu, Y. A Power Transformer Fault Diagnosis Method Based on Improved Sand Cat Swarm Optimization Algorithm and Bidirectional Gated Recurrent Unit. *Electronics* **2023**, *12*, 672. <https://doi.org/10.3390/electronics12030672>

Academic Editor: Mohamed Benbouzid

Received: 7 January 2023  
Revised: 26 January 2023  
Accepted: 26 January 2023  
Published: 29 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The power transformer is a key piece of equipment in the power grid [1,2], and its working stability has a great impact on the safety of the grid. Once a transformer fails, it will cause great damage to the country's economy and property. In order to ensure safe and stable operation of the power system, it must be accurately diagnosed [3].

In the event of a transformer fault, there is a large amount of H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, and other gases present in the insulating oil, and the composition of these gases has a strong non-linear relationship with the type of fault [4]. Therefore, dissolved gas analysis (DGA) techniques are widely used in transformer fault monitoring and diagnosis [5]. Traditional fault diagnosis methods mainly include the ratio method [6], key gas method, triangle method [7], and pentagon method [8]. Although they are simple and effective, these methods still have many problems, such as inconsistent diagnostic results and low accuracy, which reduce the reliability of fault analysis. In recent years, artificial intelligence techniques based on neural networks [9], support vector machines [10,11] (SVM), extreme learning machines (ELM), etc., combined with DGA analysis have become a research hotspot

for experts at home and abroad. Although these methods can improve the accuracy of fault diagnosis to a certain extent, there are still some shortcomings. For example, the SVM-based fault diagnostic model [12] is constrained by the multiclassification problem constraints of the SVM itself, and therefore, does not function effectively in the presence of complex high-dimensional data. Bazan et al. [13] proposed a two-stage approach for three-phase induction motor diagnosis based on mutual information measures of the current signals, principal component analysis, and intelligent systems. This offers the possibility of reducing the amount of information required, but with reduced accuracy. Although Yu [14]'s developed KNN fault diagnosis model increases the work efficiency of the KNN algorithm, it does not address the issues of the poor fault tolerance rate of the KNN to training data and is easily prone to dimension disaster, which results in weak generalization of the model. Accuracy and stability in the medium- and long-term prediction of fault data cannot be guaranteed by the fault diagnosis method based on the hidden Markov model (HMM) presented by Jiang [15]. Compared with other artificial intelligence methods, artificial neural networks (ANN) can significantly improve the accuracy of fault diagnosis. The connection weights and biases (significant parameters) of the network model are continuously adjusted during the training process to ultimately establish the corresponding mapping relationship between specific fault features and fault types for ANN-based fault diagnosis models for power transformers [16]. Researchers are integrating neural network-based, deep learning methods with transformer fault diagnosis techniques. An evolving neural network method for power transformer defect diagnostics was put out by Huang et al. [17]. The neural network automatically modifies the network parameters (connection weights and deviation terms) based on the suggested evolutionary strategy to produce the optimal model. Meng and Dong et al. [18] proposed a radial basis function neural network (RBFNN) based on a hybrid adaptive training method for the fault diagnosis of power transformers. This method is able to generate RBFNN models based on fuzzy cmeans (FCM) and quantum-inspired particle swarm optimization (QPSO), which allows for automatic configuration of the network structure and the acquisition of model parameters. Compared to conventional neural networks, using these methods, the number of neurons, the center and radius of the hidden layer activation function, and the output connection weights can be automatically calculated. The classification accuracy of RBFNN is significantly improved. This offers the possibility of reducing the amount of information required, but with reduced accuracy. Burriel et al. [19] proposed an automatic system based on neural networks for generating optimized expert diagnostic systems for fault detection when the machine works under transient conditions. Dai et al. [20] proposed a deep belief network (DBN)-based transformer fault diagnosis method. By analyzing the relationship between dissolved gas in transformer oil and fault type, the noncoding ratio of gas is determined as the feature parameter of the DBN model. The DBN adopts a multilayer multidimensional mapping method to extract more detailed fault-type differences and proves, through experiments, that this method can effectively improve the accuracy of fault diagnosis. In order to improve the hybrid kernel extreme learning machine (KELM), Huang et al. [21] proposed a transformer fault diagnosis method based on the gray wolf optimization (GWO) algorithm. The GWO algorithm can be used to optimize the parameters of the hybrid kernel function, and logistic chaos mapping can be used to generate the initial population parameters of the GWO algorithm to prevent the negative effects of convergence that is too fast on the optimization results and effectively improve the classifier performance. Although Huang [17]'s evolutionary neural network model can automatically update the network parameters, the evolutionary algorithm's capacity to converge is limited, and it is easy to fall into the local optimum, which reduces the classification model's accuracy; Meng [18]'s proposal of quantum-inspired particle swarm optimization (QPSO) can address the issue of PSO's delayed convergence. However, RBFNN's complex structure and extensive calculation are disadvantages when the data sample is large; the classification accuracy of the fault diagnosis model based on DBN is very high [20], but it needs a lot of fault data for network training, and the classification

performance is not stable in the case of small amounts of data; the method proposed in the literature [21] is very effective for KELM optimization, but its efficiency and accuracy need to be improved.

Recurrent neural networks (RNN) have achieved good performance in diagnostic models based on intelligent computing. The long short-term memory neural network (LSTM) improves the structure of the recurrent neural network. LSTM is a special kind of RNN which is used to solve the problem of gradient vanishing and gradient exploding. The principle of BiGRU is similar to that of LSTM, which simplifies the gating structure by combining the forget gate and the input gate into an 'update gate', has fewer parameters than LSTM, and can achieve functions equivalent to LSTM in some applications. BiGRU combines the unit state and the hidden state. The BiGRU network has a simpler structure than LSTM, so it needs less parameter adjustment, and has faster training speed and better prediction performance than LSTM. Therefore, BiGRU is used in this paper to construct a transformer fault diagnosis model. The inaccurate setting of hyperparameters in BiGRU can cause inefficiencies in transformer fault identification. The manual finding of hyperparameters, on the other hand, requires extensive expertise and a lot of experimentation and can therefore be optimized by intelligent optimization algorithms [22].

As a non-linear dimensionality reduction algorithm, Isomap is a good solution to non-linear problems. However, the increase in the number of samples greatly increases the computational complexity of the Isomap algorithm. Therefore, Silva and Tenenbaum et al. proposed the landmark equidistant mapping (L-Isomap) algorithm. Compared with the Isomap algorithm, L-Isomap has a faster computational speed and wider application range and can represent the low-dimensional features of high-dimensional data well.

Three shortcomings of transformer fault diagnosis based on BiGRU are summarized: (1) a single fault diagnosis model cannot greatly improve the fault diagnosis performance; (2) the noise of the transformer fault data will reduce the stability of the model; (3) research on optimization algorithms is not targeted and cannot significantly improve optimization performance. Thus, a transformer fault diagnosis method based on the L-Isomap and ISCSO-BiGRU methods is proposed in this paper. It is noteworthy that the innovations and contributions of this paper are mainly divided into the following five improved methods. First, L-Isomap is used to extract the features of DGA data to reduce the influence of noise on the diagnosis results. In addition, SCSSO can be improved by the following four methods to obtain the ISCSO. A logistic is proposed to improve the initial diversity of the sand cat population. A strategy with improved water wave dynamics factors, adaptive weights, and golden sine is introduced to improve the SCSSO. Then, it is noteworthy that ISCSO can be obtained from the above four improved methods, and the benchmark functions are used to test the optimization performance of ISCSO and the other algorithms. The results show that ISCSO has the best optimization performance. Finally, ISCSO is used to optimize the relevant hyperparameters of BiGRU. The important feature quantities selected by the L-ISOMP algorithm are input to the BiGRU optimized by the ISCSO algorithm for transformer fault identification and compared with the conventional DGA method to verify the enhancement effect of L-ISOMP on model performance. Finally, by comparing the analysis with other transformer fault diagnosis models, it is verified that the model in this paper has a higher accuracy rate.

## 2. Landmark Isomap Feature Mapping Algorithm

The Isomap algorithm is based on Multidimensional Scaling (MDS) analysis. Its basic idea is to use the geodesic distance between data points to describe the geometric properties between data, and on this basis, it establishes a mapping relationship between the geodesic and downscaled data and the original data, so as to downscale the high-dimensional data [23,24]. The Isomap-based L-Isomap algorithm, which is faster, can better express the low-dimensional characteristics of the high-dimensional data. Let the number of samples be  $N$  and Landmarks be  $n$  ( $n \ll N$ ); if the target dimensionality reduction is  $d$ , then Landmarks

is only larger than  $d + 1$ . The calculation steps of L-Isomap are as follows (the flow of data processing is shown in Figure 1).

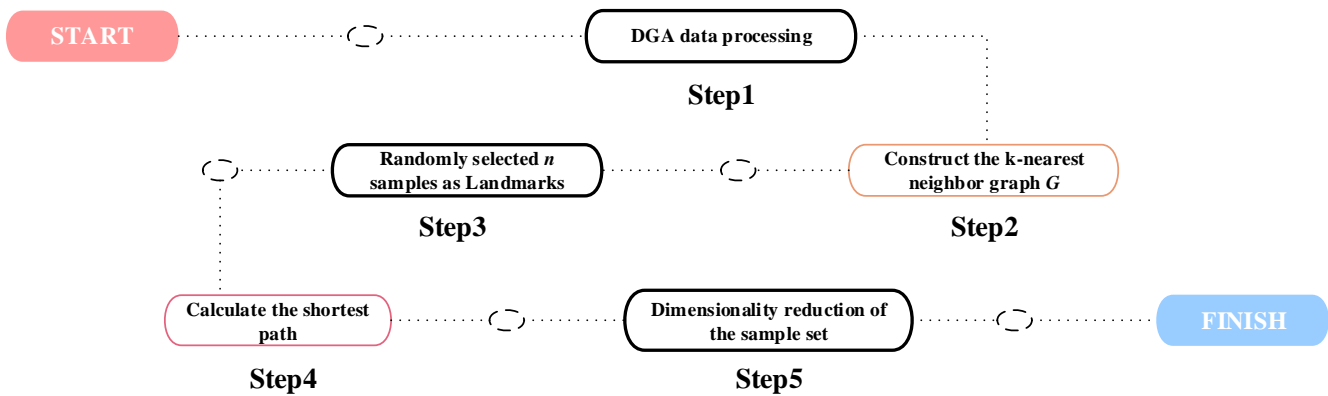


Figure 1. Data processing flow of L-Isomap.

- (1) Construct the k-nearest neighbor graph  $G$ .
- (2) Randomly select  $n$  samples as Landmarks.
- (3) Calculate the shortest path between the distance of sample points  $N$  to Landmarks  $n$  to obtain the matrix.
- (4) Apply LMDS to reduce the dimensionality of the sample set. The  $d$ -dimensional embedding coordinates of the Landmarks  $n$  are first calculated using MDS as follows.

$$B_n = -\frac{1}{2}H_n\Delta_nH_n \tag{1}$$

where  $H_n$  is the mean-centering matrix, defined by  $H_n = \delta_{ij} - 1/n$ ,  $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ .  $\Delta_n$  is the square matrix of the geodetic distance matrix  $D_{n \times n}$  between the Landmarks  $n$ ,  $\Delta_n = D_{n \times n}^2$ .

Calculating the largest eigenvalues  $d$  and eigenvectors of  $B_n$ , the low-dimensional embedding coordinates of the Landmarks points are

$$Y_{Landmarks} = \left[ \sqrt{\lambda_1} \cdot \vec{v}_1^T, \sqrt{\lambda_2} \cdot \vec{v}_2^T, \dots, \sqrt{\lambda_d} \cdot \vec{v}_d^T \right]^T \tag{2}$$

where  $\lambda_i$  and  $v_i$  are the eigenvalues and eigenvectors, respectively.

Compute the low-dimensional embedding coordinates of the non-Landmarks  $y_a$ .

$$y_a = -\frac{1}{2}Y_{Landmarks}^\# \left( \vec{\sigma}_a - \vec{\sigma}_\mu \right) \tag{3}$$

where  $Y_{Landmarks}^\# = \left[ \vec{v}_1^T / \sqrt{\lambda_1}, \vec{v}_2^T / \sqrt{\lambda_2}, \dots, \vec{v}_d^T / \sqrt{\lambda_d} \right]$ ,  $\vec{\sigma}_a = [d_{a1}^2 \quad d_{a2}^2 \quad \dots \quad d_{an}^2]^T$ ,  $\vec{\sigma}_\mu = \frac{(\vec{\sigma}_1 + \dots + \vec{\sigma}_n)}{n}$ .

### 3. Sand Cat Swarm Algorithm

Sand Cat Swarm Optimization [25,26] (SCSO) is a new meta-heuristic algorithm proposed by Amir Seyyedabbasi and Farzad Kiani in 2022, and is an intelligent optimization algorithm that mimics the survival behavior of sand cats in nature.

The sand cat is able to detect low-frequency noise to locate prey either above or below ground. The algorithm considers the optimal value in the exploration space as the prey, and the search agent continuously explores the search space through location updates, eventually moving closer to the area where the optimal value is located. The



SCSO algorithm is designed with a prey search mechanism and a prey attack mechanism. The search prey mechanism can simulate the process of sand cats searching for prey.

During the exploration phase, the sand cat can sense low frequencies below 2 kHz. In the mathematical model,  $r_G$  represents a general range of sensitivity that decreases linearly from 2 to 0, according to the working principle of the algorithm.  $S_M$  models the auditory properties of the dune cat with an assumption of 2, defined as follows.

$$\vec{r}_G = s_M - \left( \frac{S_M \times iter_c}{iter_{max}} \right) \tag{4}$$

The main parameter for the transition between the exploration and development phases is  $R$ , which is defined as follows:

$$\vec{R} = 2 \times \vec{r}_G \times rand(0,1) - \vec{r}_G \tag{5}$$

The search space is randomly initialized between the defined boundaries. During the search step, each current search agent's position is updated based on a random position. In this way, the search agent is able to explore new spaces in the search space. To avoid falling into a local optimum, the sensitivity range for each sand cat is different, defined as:

$$\vec{r} = \vec{r}_G \times rand(0,1) \tag{6}$$

In the SCSO algorithm, the sand cat updates its position based on the optimal solution, its current position, and its sensitivity range; searches for other possible best prey positions; and can find a new local optimum in a new search area, obtaining a position that lies between the current position and the prey position, while the randomness ensures that the algorithm has low running costs and complexity. The mathematical modeling of the above search process is as follows:

$$\vec{Pos}(t+1) = \vec{r} \cdot (\vec{Pos}_b(t) - rand(0,1) \cdot \vec{Pos}_c(t)) \tag{7}$$

where:  $Pos_b$  is the optimal solution,  $Pos_c$  is one's current position, and  $r$  is the sensitivity range. The SCSO algorithm attacks prey at the end of the prey search, and the prey attack mechanism for the sand cat population is described as:

$$\vec{Pos}_{md} = \left| rand(0,1) \cdot \vec{Pos}_b(t) - \vec{Pos}_c(t) \right| \tag{8}$$

$$\vec{Pos}(t+1) = \vec{Pos}_b(t) - \vec{r} \cdot \vec{Pos}_{md} \cdot \cos(\theta) \tag{9}$$

where:  $Pos_b$  is the best position,  $Pos_c$  is the current position, and  $Pos_{md}$  is the random position.

$$\vec{Pos}(t+1) = \begin{cases} \vec{Pos}_b(t) - \vec{Pos}_{md} \cdot \cos(\theta) \cdot \vec{r} & |R| \leq 1 \\ \vec{r} \cdot \left( \vec{Pos}_b(t) - rand(0,1) \cdot \vec{Pos}_c(t) \right) & |R| > 1 \end{cases} \tag{10}$$

Equation (10) represents the update to each sand cat's position during the exploration and exploitation phase. When  $R \leq 1$ , the sand cats are guided to attack their prey; otherwise, the sand cats are tasked with finding new possible solutions in the global area.

#### 4. Improved Sand Cat Swarm Optimization

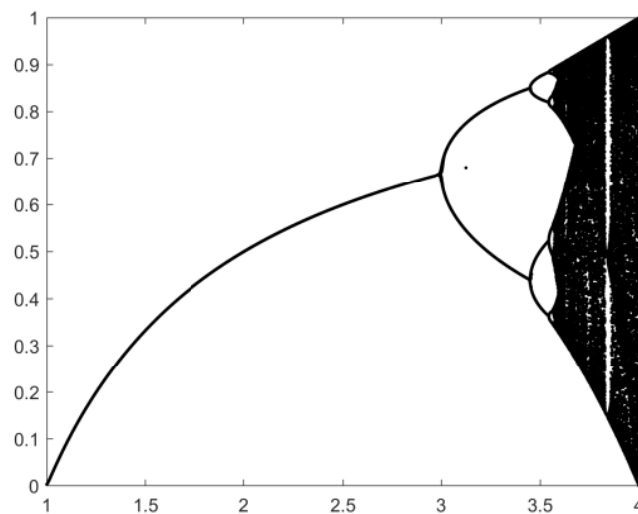
##### 4.1. Logistic Chaos Mapping

The initial position of the individuals in the population plays a key role in the optimization effect of the population intelligence algorithm itself [27]. Commonly used chaotic mappings are logistic mapping, tent mapping, Henon mapping, Chebyshev mapping, and combinatorial chaos mapping. In this paper, the logistic mapping is chosen because it has

better ergodicity, autocorrelation, and mutual correlation than other chaotic mappings. For this reason, the population is initialized using a logistic chaos mapping [28], which allows for a more balanced distribution of the population, thus improving the convergence and optimization accuracy of the algorithm. The formula is as follows.

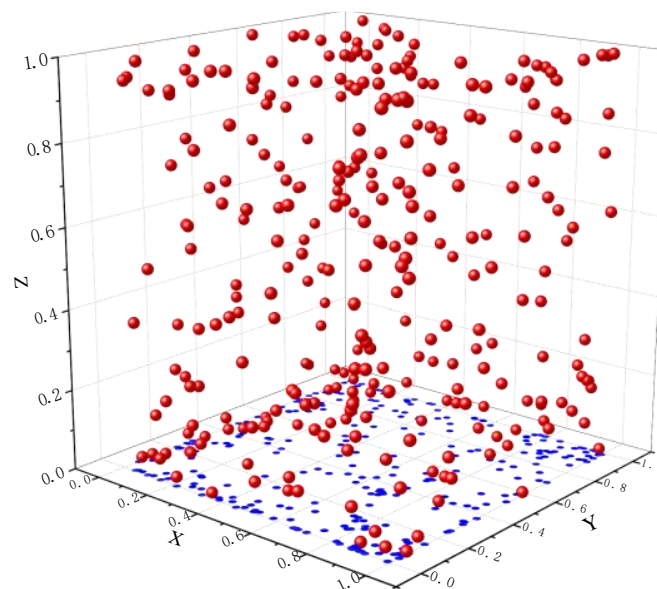
$$x_{k+1} = \mu x_k(1 - x_k) \quad (11)$$

where  $\mu \in (0, 4]$ ,  $k$  is the number of iterations, and  $x \in (0, 1)$ . With the initial value of  $x_0$ , the premise of the chaotic state is that  $\mu$  takes  $[3.5699, 4.0]$  and the sequence of chaotic states is randomly non-convergent. As  $\mu$  varies, the distribution of the logistic chaotic sequence is shown in Figure 2.



**Figure 2.** Sequence distribution under different bifurcation parameters.

In Figure 2, it can be seen that the population randomness is best when  $\mu = 4$ . Therefore, the logistic chaotic sequence at  $\mu = 4$  was chosen to initialize the population, and the number of iterations was set to 300. The spatial distribution of the logistic chaotic sequence is shown in Figure 3.



**Figure 3.** Spatial distribution of logistics chaotic mappings.

#### 4.2. Water Wave Dynamic Factor

The SCSO algorithm, in which  $rG$  decreases linearly from 2 to 0 as the iterative process proceeds, does not adapt well to complex multi-peaked multivariate functions and is prone to low accuracy. Therefore, a water wave dynamic evolution factor is introduced to take advantage of the uncertainty of the water wave dynamics so that the algorithm can better adapt to complex functions and improve the probability of finding a good solution. The uncertainty of using the dynamics of water waves allows the population to search over a wider area, reducing the blindness of other individuals following, enhancing in-formation exchange and learning between populations, maintaining population diversity, effectively avoiding convergence, and thus, improving the ability of the algorithm to jump out of the local optimum. At the same time, a control factor  $k$  is added to control the decreasing magnitude of  $rG$ , and the mathematical model is described as shown in Equation (12):

$$rG = 2 \cdot s \cdot \exp\left(\frac{-t}{T}\right)^k \cdot r \tag{12}$$

where:  $s$  is a random integer and  $s \in [-1, 1]$ ;  $r$  is a random function of  $r$ ;  $r \in [0, 1]$ ; and  $k \in [1, 3]$ . The larger  $k$  is, the smaller the decreasing  $rG$ , and the opposite is true when  $k$  is larger. A comparison of the original  $rG$  and the water wave dynamic factor with the number of iterations is shown in Figures 4 and 5:

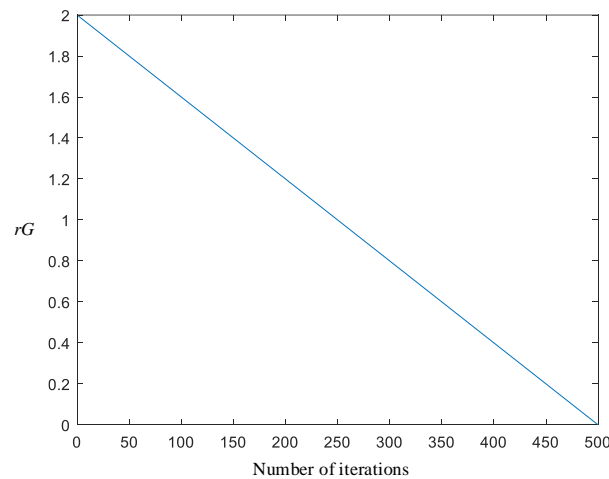


Figure 4. The original  $rG$  iteration curve.

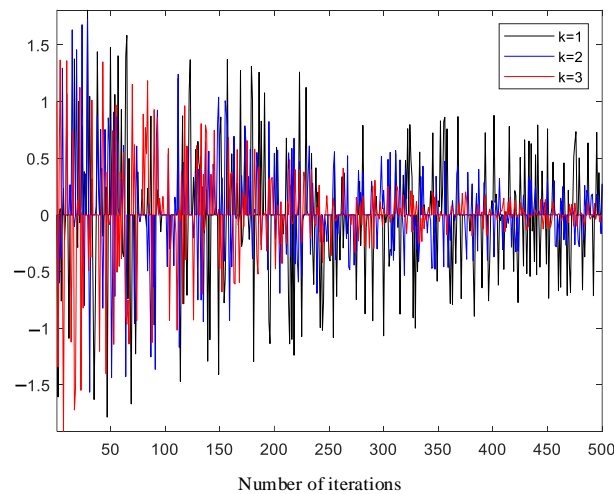


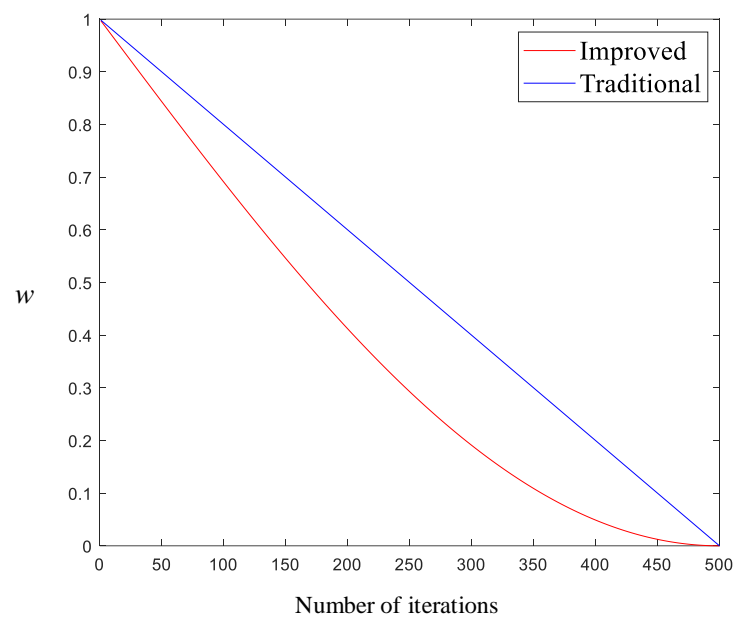
Figure 5. Water wave dynamic factors for different values of  $k$ .

#### 4.3. Adaptive Weights

Suitable inertia weights can effectively coordinate the transition between the global search and local exploitation of the algorithm, and can improve the convergence speed and the accuracy of the algorithm's search for the objective optimization function. The sand cat performed a local search according to Equation (7) when performing a local search for the optimum. In this region, when the sand cat searches according to the vicinity of Equation (7), it can only search for an optimum near the optimal solution and cannot perform better optimization. For this reason, a new weighted adaptive algorithm is introduced in this paper, which enhances the local optimization of the sand cat using the minimum adaptive weighting to adjust the current best sand cat position as it approaches the target. The adaptive weighting formula is shown in Equation (13) and the improved formula of Equation (7) is shown in Equation (14). Figure 6 shows a function image of the inertia weight before and after improvement.

$$w = \sin\left(\frac{\pi \cdot t}{2 \cdot iter_{\max}} + \pi\right) + 1 \quad (13)$$

$$\vec{Pos}(t+1) = w \cdot \vec{r} \cdot (\vec{Pos}_{bc}(t) - rand(0,1) \cdot Pos_c(t)) \quad (14)$$



**Figure 6.** Adaptive weighting curves.

#### 4.4. Golden Sine

The golden sine algorithm (goldenSA) is a new intelligent algorithm proposed by Tanyildizi et al. [29] in 2017 based on an idea related to the sine function, and has the advantages of fast search speed, simple tuning parameters, and good robustness. The golden-SA algorithm uses the special relationship between the sine function and the unit circle combined with the golden partition coefficient for iterative search; by scanning the sine function, the unit circle simulation algorithm explores the search space. The concept of the golden mean was first introduced by the ancient Greek mathematician Eudoxus in the fourth century B.C. The golden mean does not require gradient information and requires only one iteration per step, while the contraction step of the golden mean is fixed. Therefore, a combination of the sine function and the golden mean can be used to find the maximum or minimum value of the function more quickly. At the same time, the iterative nature of

the golden sine search strategy prevents the algorithm from falling into local optimum. The mathematical model of the Golden-SA strategy is described in Equation (15):

$$X_i^{t+1} = X_i^t \cdot |\sin(R_1)| + R_2 \cdot \sin(R_1) \cdot |x_1 \cdot P_i^t - x_2 \cdot x_i^t| \tag{15}$$

where  $t$  is the number of iterations;  $R_1$  and  $R_2$  are random numbers with values of  $[0, 2\pi]$  and  $[0, \pi]$ , respectively, which represent the distance and direction of the next generation of individuals;  $x_1$  and  $x_2$  are the golden section coefficients, which are used to narrow the search space and lead individuals to converge to the optimal value; and  $P_i^t$  is the position of the current optimal individual. The golden section coefficients  $x_1 = a \cdot (1 - \tau) + b \cdot \tau$ ,  $x_2 = a \cdot \tau + b \cdot (1 - \tau)$ ,  $a$ , and  $b$  are the search intervals, and  $\tau$  is the golden section ratio, which takes a value of about 0.618033.

#### 4.5. ISCSO Implementation Steps

In summary of the improvements described above, the specific implementation steps for ISCSO in this paper are as follows.

Step 1: Initialize the relevant parameters of the sand cat swarm algorithm: population size  $N$ , spatial dimension  $dim$ , upper and lower bounds  $ub$ ,  $lb$ , and the maximum number of iterations  $Tm$ .

Step 2: Initialize the population with logistics according to Equation (11).

Step 3: Calculate the fitness value of each individual.

Step 4: Replace the original Equations (4) and (7) with Equations (12) and (14).

Step 5: Update the position according to Equations (9) and (14).

Step 6: Update the position according to Equation (15).

Step 7: Determine whether the maximum number of iterations is reached. If so, output the global optimal individual position; otherwise, return to step 3.

#### 4.6. Algorithm Performance Testing

Two single-peak test functions and one multi-peak test function were selected for simulation experiments to validate the optimization search effect of the ISCSO algorithm. The algorithms were compared with the standard SCSO algorithm, the whale optimization algorithm (WOA) [30], the grey wolf optimizer (GWO) [31], and particle swarm optimization (PSO) [32]. The number of iterations for all algorithms was 500, each group of algorithms performed 30 independent optimization tests for each test function, and the optimal value, the worst value, the mean, and the standard deviation were calculated separately. The experimental results are shown in Table 1. The test functions are shown in Figures 7–9, with the following equations.

$$F_1(x) = \sum_{i=1}^n |x_i| + \prod_{i=1}^n |x_i| \tag{16}$$

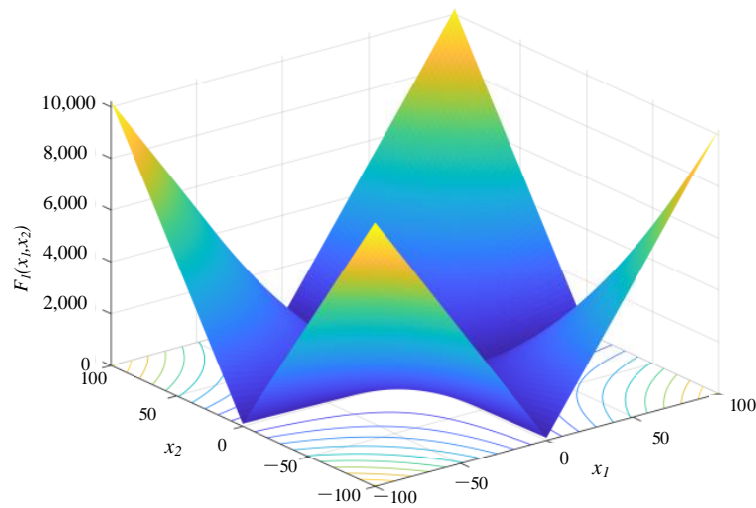
$$F_2(x) = \sum_{i=1}^n ix_i^4 + random[0, 1) \tag{17}$$

$$F_3(x) = 0.1 \left\{ \begin{array}{l} \sin^2(\pi 3x_1) + \\ \sum_{i=1}^n (x_i - 1)^2 [1 + \sin^2(3\pi x_{i+1})] + \\ \sum_{i=1}^n (x_n - 1)^2 [1 + \sin^2(2\pi x_n)] \end{array} \right\} + \sum_{i=1}^n u(x_i, 5, 100, 4) \tag{18}$$

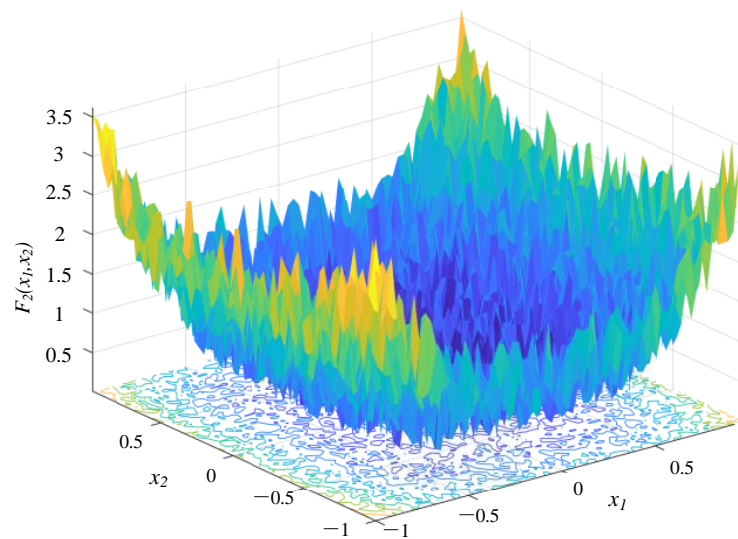


**Table 1.** Test Results for The Five Optimization Algorithms.

Function	Algorithm	Best	Worst	Ave	Std
$F_1$	SCSO	$4.68 \times 10^{-66}$	$1.06 \times 10^{-58}$	$4.35 \times 10^{-60}$	$1.96 \times 10^{-59}$
	ISCSO	0	0	0	0
	WOA	$2.63 \times 10^{-58}$	$1.15 \times 10^{-47}$	$3.86 \times 10^{-49}$	$2.09 \times 10^{-48}$
	GWO	$2.25 \times 10^{-17}$	$4.91 \times 10^{-16}$	$1.07 \times 10^{-16}$	$1.04 \times 10^{-16}$
	PSO	6.99	38.4	15.1	7.21
$F_2$	SCSO	$8.19 \times 10^{-8}$	$1.18 \times 10^{-3}$	$1.89 \times 10^{-4}$	$2.82 \times 10^{-4}$
	ISCSO	$3.79 \times 10^{-7}$	$2.55 \times 10^{-4}$	$6.92 \times 10^{-5}$	$6.36 \times 10^{-5}$
	WOA	$1.16 \times 10^{-4}$	$2.95 \times 10^{-2}$	$4.82 \times 10^{-3}$	$5.98 \times 10^{-3}$
	GWO	$1.11 \times 10^{-3}$	$5.77 \times 10^{-3}$	$2.31 \times 10^{-3}$	$1.11 \times 10^{-3}$
	PSO	0.0423	8.19	0.661	1.62
$F_3$	SCSO	1.43	2.8	2.45	0.388
	ISCSO	$7.62 \times 10^{-6}$	$1.13 \times 10^{-2}$	$8.39 \times 10^{-4}$	$2.84 \times 10^{-3}$
	WOA	0.105	1.48	0.545	0.352
	GWO	0.441	1.13	0.703	0.201
	PSO	7.87	2980	121	540



**Figure 7.** Testing function  $F_1(x)$ .



**Figure 8.** Testing function  $F_2(x)$ .

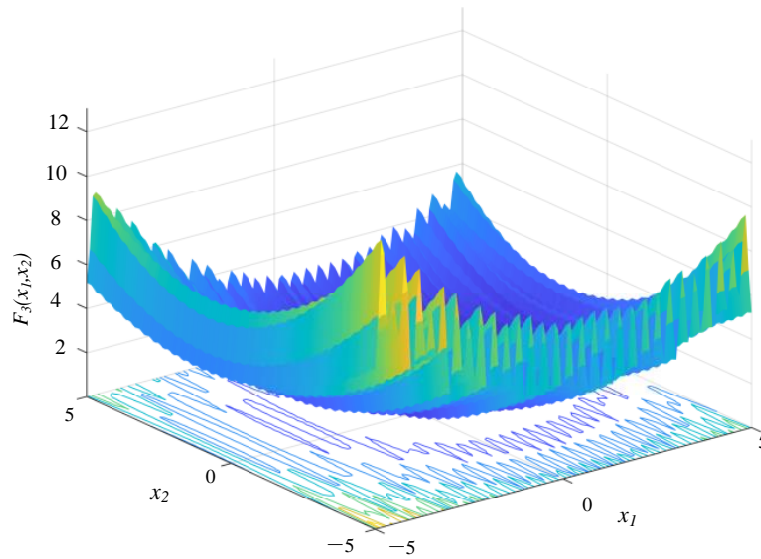


Figure 9. Testing function  $F_3(x)$ .

As can be seen in Table 1, for the single-peak test function  $F_1$ , ISCSO finds the theoretical optimum and outperforms the other four algorithms in every respect. For the single-peak test function  $F_2$ , ISCSO is slightly inferior to the native SCSO in terms of optimal values, but outperforms the SCSO and the other four algorithms in all other aspects. For the multi-peak test function  $F_3$ , ISCSO outperforms the other four algorithms in all four criteria.

As can be seen in Figures 10–12, ISCSO requires the smallest number of iterations for several algorithms to converge to the same accuracy. It shows that the introduction of multiple methods increases the proportion of high-quality individuals in the population and improves the convergence speed of the algorithm. The convergence curves of the WOA, GWO, and PSO algorithms flatten out with increasing iterations, with varying degrees of stagnation and relatively low accuracy in the search for optimal solutions.

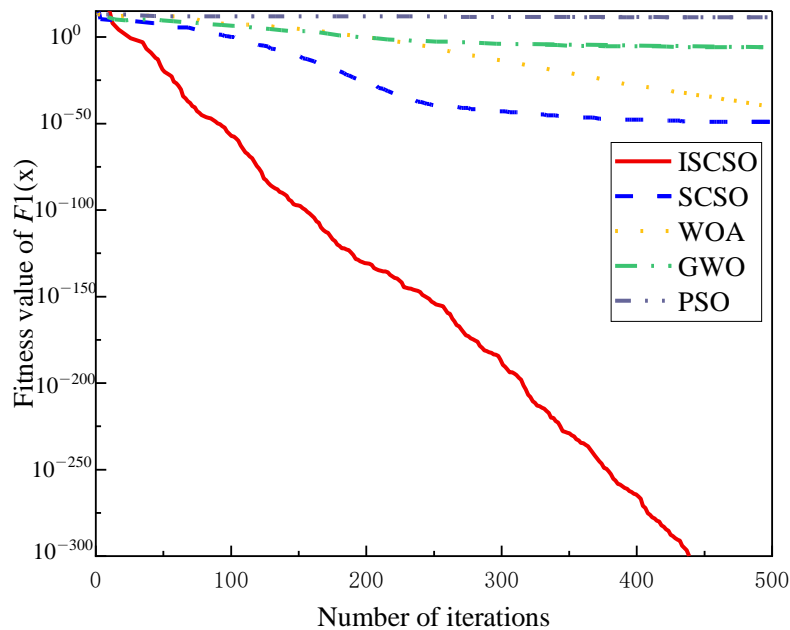


Figure 10. Comparison of the optimization search for  $F_1(x)$  for each optimization algorithm.

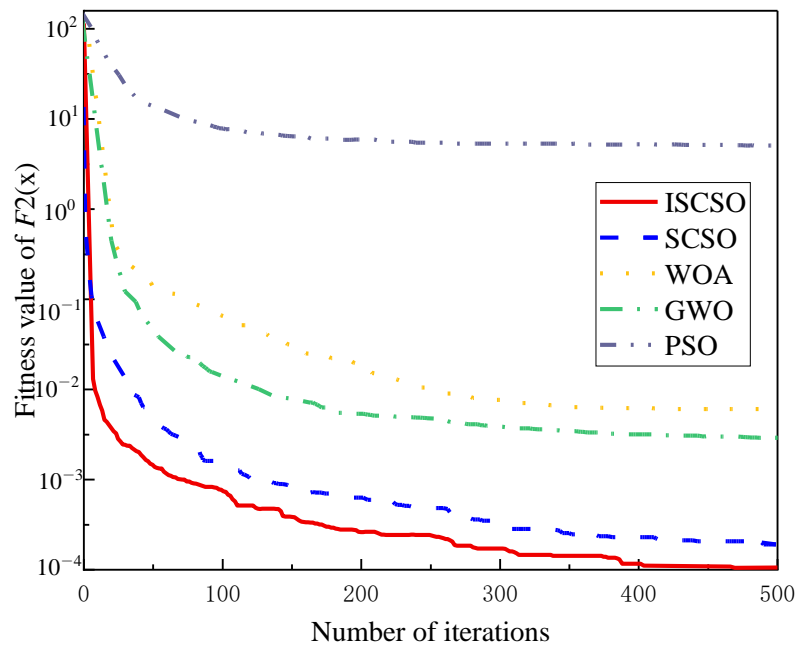


Figure 11. Comparison of the optimization search for  $F_2(x)$  for each optimization algorithm.

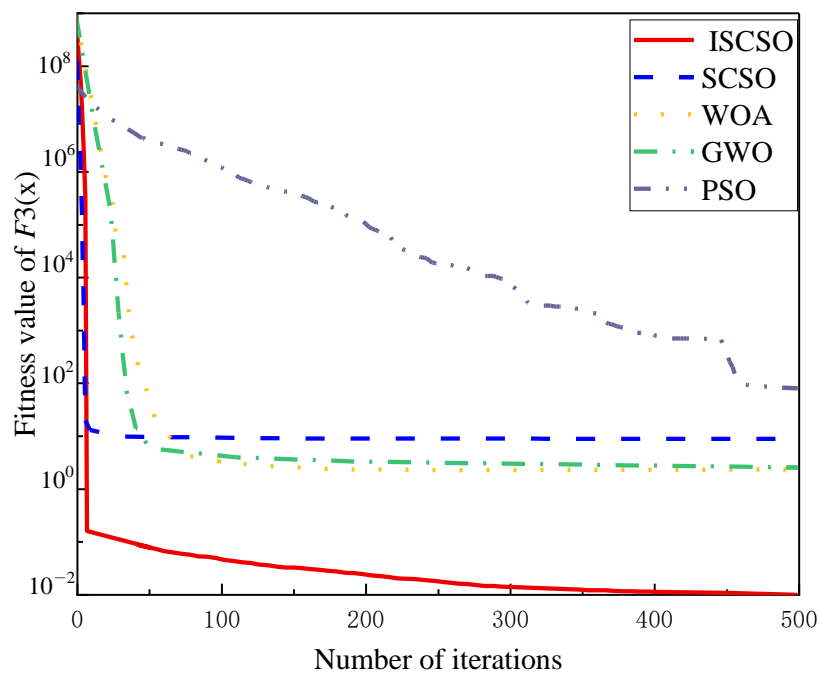


Figure 12. Comparison of the optimization search for  $F_3(x)$  for each optimization algorithm.

In summary, the ISCSO algorithm has better local extremum escape capability, overall optimization seeking synergy, and convergence performance than the standard SCSO, WOA, GWO, and PSO algorithms.

### 5. BiGRU Diagnostic Model Optimized by ISCSO Algorithm

#### 5.1. Gated Recurrent Unit Neural Networks

In early 2014, the gated recurrent unit neural network [33] (GRU) was proposed by Cho et al. In the development and improvement of RNN-LSTM-GRU, GRU inherited the advantages of the RNN model in time series computation and of the LSTM model in correlation fusion between data. It also combines the forgetting and input gates in

LSTM into an “update gate”, which shortens the computation time and improves the computation efficiency by simplifying the model structure and reducing the training parameters, avoiding the problem of gradient explosion during LSTM computation. The network structure of the GRU model is shown in Figure 13, where A represents the sigmoid activation function.

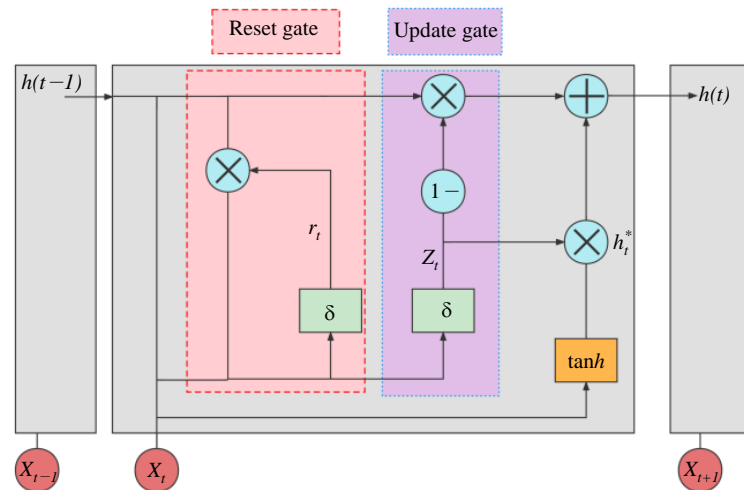


Figure 13. The network structure of the GRU model.

The updated equation for the parameters of the GRU model is

$$\begin{cases} r_t = \text{sigmoid}(W_r X_t + U_r h_{t-1} + b_r) \\ z_t = \text{sigmoid}(W_z X_t + U_z h_{t-1} + b_z) \\ h_t^* = \tanh(W X_t + r_t U h_{t-1} + b) \\ h_t = (1 - z_t) h_t^* + z_t h_{t-1} \end{cases} \quad (19)$$

where:  $X_t$  denotes the input at moment  $t$ ; sigmoid and tanh are the activation functions used to calculate the output of the hidden layer neurons;  $r$  and  $z$  denote the reset gate and update gate, respectively;  $W$  and  $U$  are both the weight matrices of GRU;  $b$  denotes the bias;  $h_t^*$  denotes the candidate hidden layer;  $h_t$  denotes the hidden layer;  $r_t$  denotes a moment in time when the reset gate allows the previously hidden state  $h_{t-1}$  to control the influence of the candidate state  $h_t^*$  so that any irrelevant information found in the future can be effectively discarded; and  $z_t$  denotes a moment in time when the update gate controls how much information in the previously hidden state  $h_{t-1}$  can be passed to the currently hidden state  $h_t$ .

The transmission state of the GRU is unidirectional from front to back and is usually used to solve the problem of the output at the current moment only being related to a single previous state variable. However, changes in the dissolved gas content of a given oil are often influenced by other gas components, as well as by external conditions and other factors. Therefore, using a bidirectional GRU model to solve the problem of interfering influencing factors will make transformer fault diagnosis more effective.

### 5.2. The BiGRU Model

The BiGRU is composed of two unidirectional GRUs superimposed on the top and bottom together, and the output is determined by the states of these two GRUs together. The model network structure is shown in Figure 14.

In Figure 14,  $X_i$  indicates the dissolved gas content in the oil and the external conditions input data;  $e_i$  indicates the vector of the input data  $X_i$ ;  $h_i$  indicates the hidden layer state in the forward direction; and  $h_i$  indicates the hidden layer state in the reverse direction.

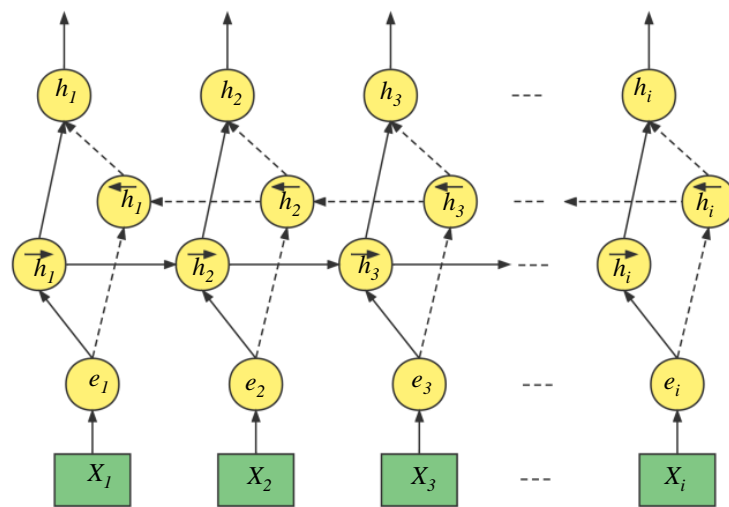


Figure 14. The network structure of the BiGRU model.

5.3. Steps for the Sand Cat Swarm Algorithm to Find the Optimal BiGRU Parameters

To improve the classification of BiGRU, the ISCSO method was used to optimize it for more accurate application in practice. There are many hyperparameters in the BiGRU model [34] which have a great impact on the accuracy of fault diagnosis in transformers: batch size (batsize), alpha, and the number of hidden layers of structural parameters and number of layer neurons (num). The search intervals for the relevant hyperparameters of BiGRU are shown in Table 2.

Table 2. BiGRU Hyperparameter Optimization Intervals.

Parameter	Scope of the Search for Excellence
batsize	[0, 300]
alpha	[0.001, 0.05]
num	[1, 100]

The steps for the sand cat swarm algorithm to find the optimal BiGRU parameters are as follows.

Step 1: Set the size of the sand cat colony to  $N$ , the dimension of the search space to  $D$ , and the maximum number of iterations to  $Tmax$ . Use the logistic chaos mapping together with the initialization method of the original SCSO algorithm to generate the initial position of the sand cat colony. Set the BiGRU training parameters batch size (batsize) and learning rate (alpha) to take a range of values. Set the range of values for the structural parameters of the number of hidden layers and number of layer neurons (num). Notate the parameter set of  $\eta = \{\text{batsize}, \text{alpha}, [\text{num}1, \dots, \text{num}n]\}$ .

Step 2: Use the L-ISOMP method to extract features from the transformer fault data, and divide the dimensionality reduction into a training set and a testing set.

Step 3: Construct the L-ISOMP-ISCSO-BIGRU transformer fault diagnosis model. Calculate the transformer fault discrimination accuracy and define it as the fitness function of the individual sand cat swarm. The accuracy of the transformer fault model is the ratio of the number of correct classifications  $n$  to the total number of samples  $m$ .

Step 4: When  $|R| < 1$ , update the position according to Equation (9). When  $|R| > 1$ , update the position according to Equation (14). Then, update the position of individuals with better fitness according to the golden sine strategy in Equation (15).

Step 5: Calculate the fitness value of the updated sand cat group individuals again. Determine whether the current fitness value is the highest value or reaches the maximum number of iterations. If the condition is satisfied, assign the optimal parameter set  $\eta_{\text{best}}$  to BiGRU. Otherwise, return to step 3.



Step 6: Construct the L-ISOMP-ISCSO-BiGRU transformer fault diagnosis model based on the optimal parameters after ISCSO optimization. After obtaining the optimal parameters, BiGRU will start the fault diagnosis and output the results, including the algorithm running time, fault classification, and accuracy.

The process of building the L-ISOMP-ISCSO-BIGRU transformer fault diagnosis model is shown in Figure 15 and the implementation framework is shown in Figure 16. The experimental parameters were set as in Table 3.

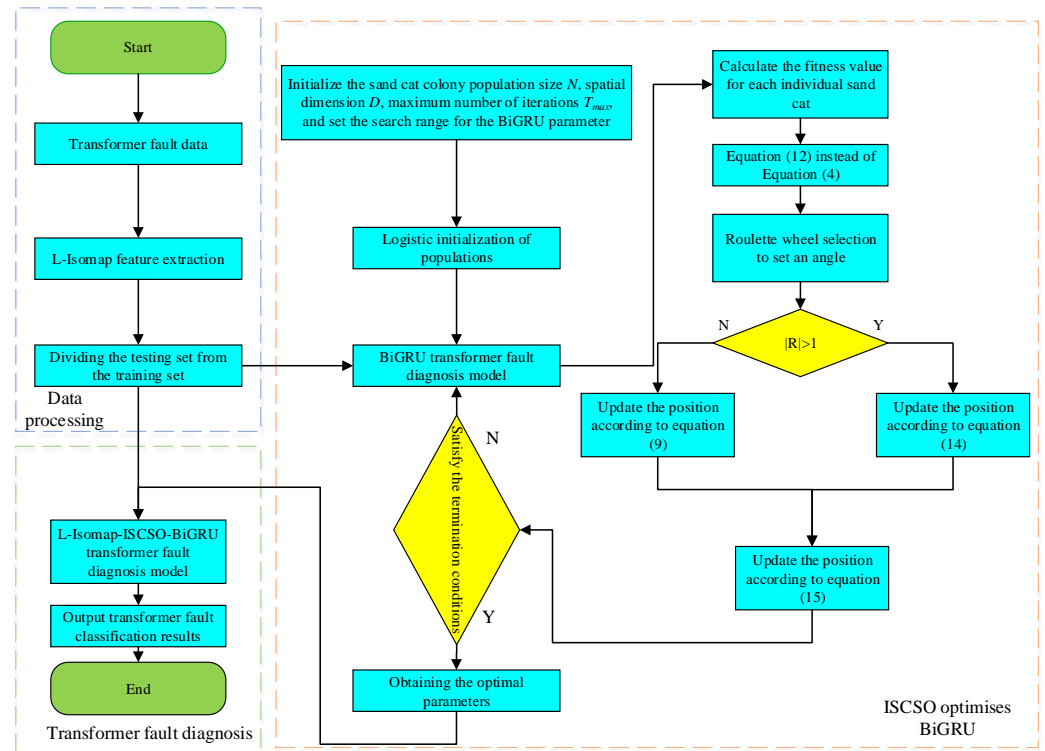


Figure 15. Fault diagnosis model based on L-ISOMP-ISCSO-BIGRU.

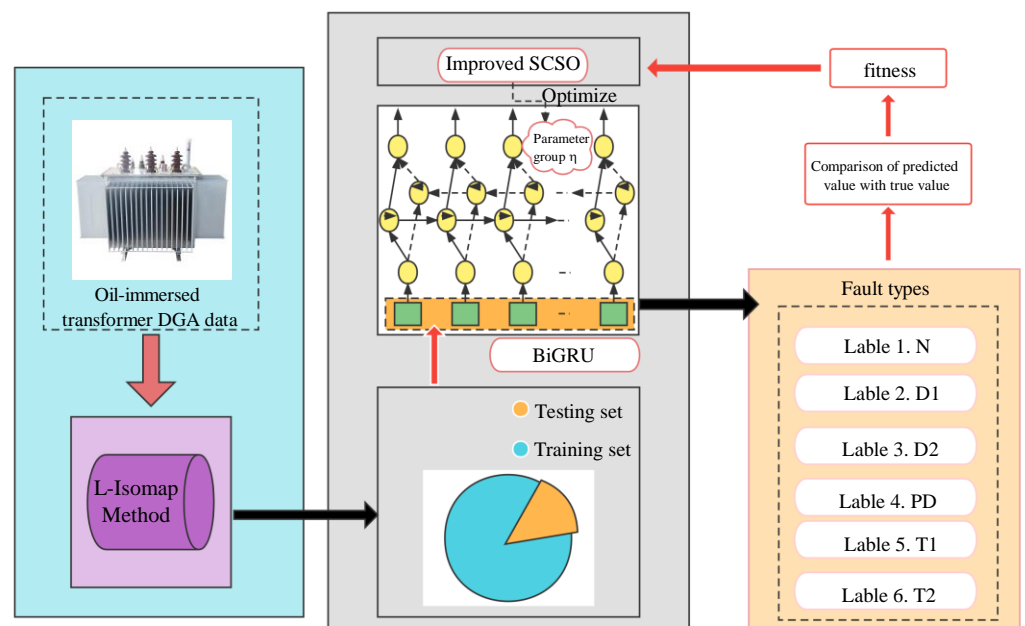


Figure 16. Implementation framework diagram of the model.

**Table 3.** Parameters Settings.

Algorithm	Parameters Settings
SCSO	The sensitivity range $rG$ changes from 2 to 0 and $R$ changes from $-2rG$ to $2rG$ , $N = 30$ , $T = 100$
ISCSO	$k = 3$ , $N = 30$ , $T = 100$
PSO	$W = 0.729$ , $C_1 = 1.49445$ , $C_2 = 1.49445$ , $N = 30$ , $T = 100$
WOA	$a$ decreases linearly from 2 to 0, $b = 1$ , $p^* = 0.5$ , $N = 30$ , $T = 100$
GWO	$a$ decreases linearly from 2 to 0, $r_1, r_2 \in [0, 1]$ , $N = 30$ , $T = 100$

**6. Example of Transformer Fault Diagnosis**

According to IEC 60599 [35], by analyzing the gas concentration of  $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ , and  $C_2H_2$  in the transformer oil, the operating status of the transformer can be judged. The data were provided by a power supply company in the northeast of China. In transformer fault diagnosis, the fault characteristics are mostly selected from the five fault-related gas content values of dissolved gas components ( $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ , and  $C_2H_2$ ) in the insulating oil. However, due to the large number of fault types in transformers and the ambiguity of the fault characteristics associated with them, judging only by their own component content, it is likely to have some impact on the fault discrimination ability [36]. Therefore, this paper uses the non-coding ratio method to construct nine characteristic parameters and combines five fault-characteristic-related gases of dissolved gas components in insulating oil, resulting in a total of 14 characteristic parameters combined to form the transformer fault raw data set. The nine characteristic parameters are:  $CH_4/H_2$ ,  $C_2H_2/C_2H_4$ ,  $C_2H_4/C_2H_6$ ,  $C_2H_2/(TH)$ ,  $H_2/(H_2 + TH)$ ,  $C_2H_4/(TH)$ ,  $C_2H_6/(TH)$ ,  $CH_4/(TH)$ , and  $(CH_4 + C_2H_4)/(TH)$ . Where TH is the total hydrocarbon, if the denominator of the ratio is 0, the denominator is modified to  $10^{-8}$  to avoid an invalid value. The original fault data for some transformers are shown in Table 4.

**Table 4.** Transformer Original Fault Data.

Characteristic Gas Composition and Content( $\mu L \cdot L^{-1}$ )					Fault Type
$H_2$	$CH_4$	$C_2H_6$	$C_2H_4$	$C_2H_2$	
45.8	36.9	7.9	7.5	0.3	Normal
49.1	12.2	0.3	3.9	4.8	Low-energy discharge
201.2	107	19.4	136.6	159.5	High-energy discharge
72.2	159	235.3	32.9	0	Medium- to low-temperature overheating
8.4	28.7	13	105	2.1	High-temperature overheating
84.3	8	0.4	7.2	0	Partial discharge

To ensure that the model has good generalization capability and to target the subjectivity of the human selection of fault data, from the acquired raw transformer fault data, 465 sets of characteristic gas-in-oil data were randomly selected in a ratio of 2:1 for training and testing. The distribution of the training and testing sample data is shown in Table 5. Normalization was carried out due to the large difference in the magnitude of the raw data, which would increase the complexity of the model calculation. Six state types were used as output features for transformer fault diagnosis (i.e., normal (N), low-energy discharge ( $D_1$ ), high-energy discharge ( $D_2$ ), partial discharge (PD), medium- to low-temperature overheating ( $T_1$ ), and high-temperature overheating ( $T_2$ )). The evaluation indicators for the experiment included the algorithm running time, fault classification, and accuracy, and the output was the transformer fault type. The details of the training set and test set sample distribution are shown in Table 5. The database was used to determine the type of fault by the parameter range of the gas ratio. The three ratio method of diagnosis recommended by IEC 60599 is shown in Table 6.

**Table 5.** Training and Testing Sample Distribution.

Working Condition	Category Tags	Total Number of Samples	Number of Training Set Samples	Test Set Sample Size
N	1	30	20	10
D <sub>1</sub>	2	66	44	22
D <sub>2</sub>	3	118	80	38
PD	4	38	26	12
T <sub>1</sub>	5	126	85	41
T <sub>2</sub>	6	87	56	31
Total		465	311	154

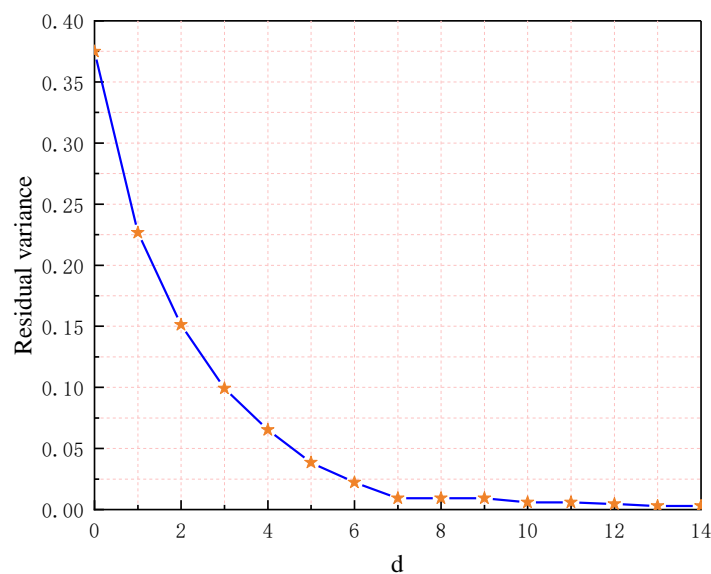
**Table 6.** DGA Diagnostic Form.

Classification	C <sub>2</sub> H <sub>2</sub> /C <sub>2</sub> H <sub>4</sub>	CH <sub>4</sub> /H <sub>2</sub>	C <sub>2</sub> H <sub>4</sub> /C <sub>2</sub> H <sub>6</sub>
D <sub>1</sub>	>1	0.1–0.5	>1
D <sub>2</sub>	0.6–2.5	0.1–1	>2
PD	NS	<0.1	<0.2
T <sub>1</sub>	NS	>1 (except NS)	<1
T <sub>2</sub>	<0.1	>1	1–4

The framework for the implementation of the ISCSO-BiGRU model for power transformer fault diagnosis is shown in Figure 16.

6.1. Comparison of the Impact of Feature Input Selection Methods on BiGRU

The samples were downsampled and analyzed using the L-Isomap method. In the L-Isomap algorithm, two key parameters must be taken into account: the nearest neighbor node  $k$  and the low-dimensional space dimension  $d$ . If the value of  $k$  is shifted too much, it will lead to a longer computing time for the model, and conversely, if the value of  $k$  is too small, the normal data will be treated as outliers by the algorithm. In this paper, through a survey of the relevant literature, the residual variance was chosen as the basis for judging  $k$  and  $d$ . As the residual variance decreases, the obtained fault feature set better reflects the characteristics of the original transformer. For the value of  $k$ , a minimum value of residual variance close to 0.01 is sufficient. As a result of the experimental tests, the residual variance satisfies the optimum condition for  $k = 10$ . The relationship between the residual variance and the dimensionality of the feature data is shown in Figure 17.



**Figure 17.** Comparison of the correctness of fault diagnosis models for different values of  $d$ .

As can be seen in Figure 17, the residual variance decreases with the increase in the dimension of the feature data. When the number of dimensions  $d$  increases to 7, the residual variance is 0.0092 and stays at around 0.01. As the dimension increases, the residual variance tends to decrease, and then, smooth out, and the fluctuation also tends to smooth out at this point. This determines that the characteristic data dimension  $d$  is 7.

Compared to L-Isomap, Isomap also achieves the above dimensionality reduction, but the computation time is much longer than that of L-Isomap. A comparison of the dimensionality reduction time is shown in Table 7. The difference in computation time between the two is three orders of magnitude, and L-Isomap is much more efficient than Isomap.

**Table 7.** Dimensionality reduction time comparison ( $d = 7$ )/s.

Number of Data Sets	Isomap Time/s	L-Isomap Time/s
150	0.24	0.00983
300	1.395	0.0124
450	3.6946	0.052

We further analyzed L-Isomap combined with BiGRU in BiGRU, KNN, ELM, SVM, and GRU using data without L-Isomap downscaling and after L-Isomap downscaling. The accuracy rates are shown in Table 8.

**Table 8.** Comparison of Diagnosis Results Before and After Dimensionality Reduction of Different Models.

	Without L-Isomap Downscaling					Downscaled via L-Isomap				
	KNN	ELM	SVM	GRU	BiGRU	KNN	ELM	SVM	GRU	BiGRU
Accuracy rate (%)	71.26	69.42	67.36	74.32	77.97	76.64	72.72	70.12	81.81	84.41

According to the results in Table 8, the prediction accuracy of the KNN, ELM, SVM, GRU, and BiGRU models is improved after dimensionality reduction of the original data by L-Isomap. The accuracy of the BiGRU model is higher than that of KNN, ELM, SVM, and GRU before and after dimensionality reduction by the L-Isomap algorithm.

### 6.2. Comparison of Multi-Model Diagnostic Results

The important features after eL-ISOMP algorithm selection were used as diagnostic model inputs. Four mainstream supervised learning models, K-nearest neighbor (KNN), ELM, SVM, and GRU, were selected for training with default parameters, and the test results were compared with the traditional BiGRU model. The results are shown in Table 9 and Figures 18 and 19. In Table 9 and Figures 18 and 19, it can be seen that the BiGRU model is more sensitive to normal and high-energy discharge in transformer fault diagnosis, and slightly less effective in partial discharge diagnosis. However, the combined fault diagnosis accuracy is the highest among the five models.

**Table 9.** Diagnostic Accuracy of Different Models.

Type of Fault	Correct Diagnosis Rate				
	KNN	ELM	SVM	GRU	BiGRU
N	0.7000	0.7000	0.9000	0.8000	0.9000
D <sub>1</sub>	0.7555	0.7272	0.6818	0.7272	0.8636
D <sub>2</sub>	0.7520	0.7368	0.7105	0.8157	0.8684
PD	0.8597	0.6666	0.6666	0.8333	0.7500
T <sub>1</sub>	0.7694	0.7073	0.6829	0.8048	0.8292
T <sub>2</sub>	0.7653	0.7741	0.6774	0.8387	0.8387
Average	0.7664	0.7272	0.7012	0.8181	0.8441

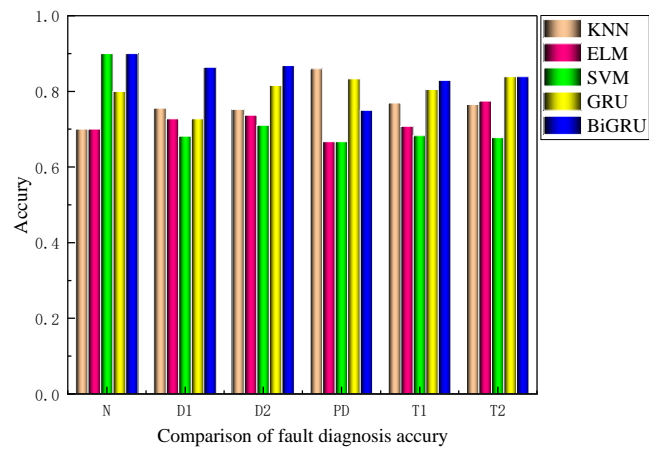


Figure 18. Diagnostic accuracy of different models.

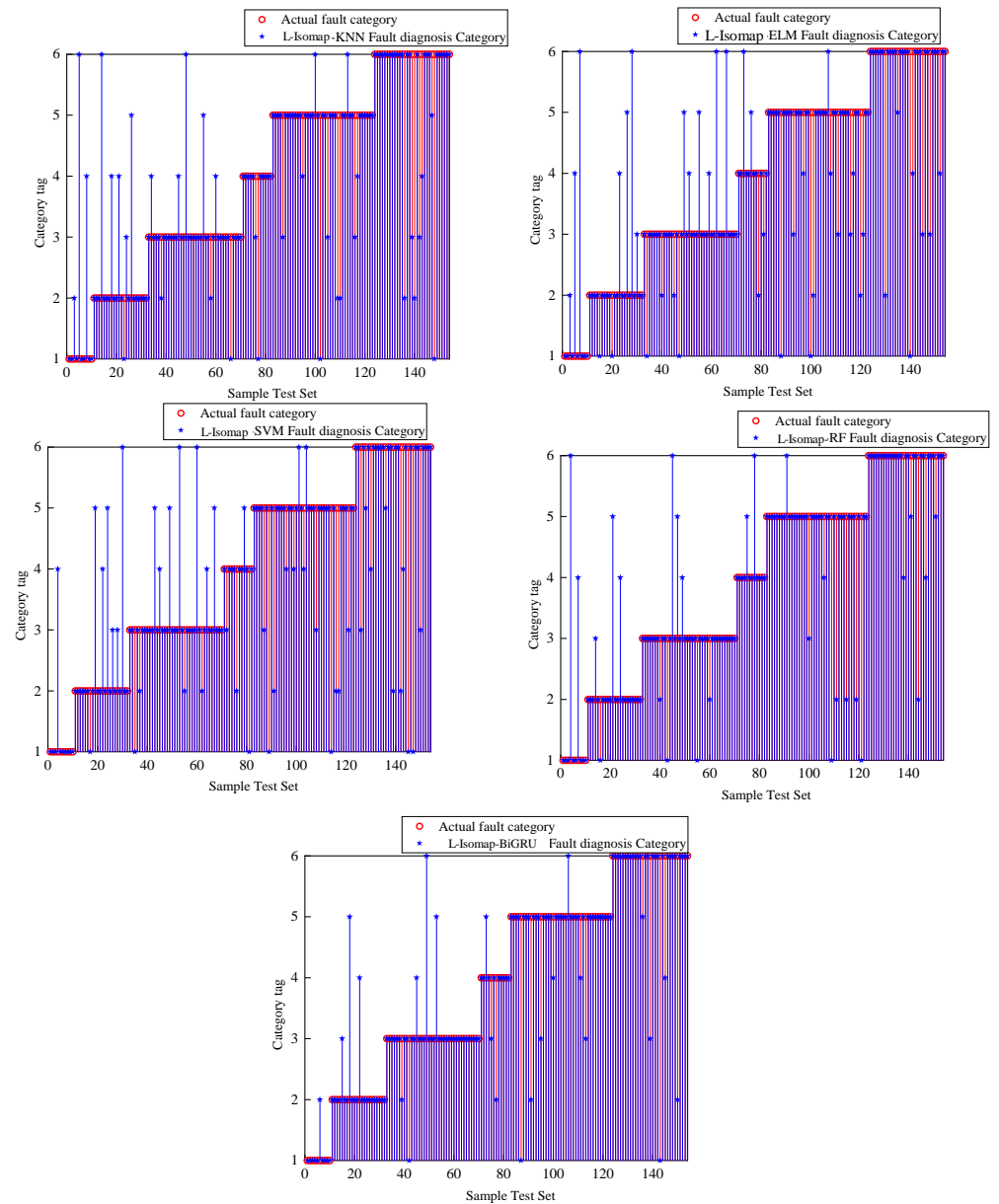


Figure 19. Adaptation change curve.



To illustrate the computational stability of BiGRU, 100 samples were randomly selected from a sample set of 465. They were divided into a training set and a testing set according to a 2:1 ratio, and then, these training samples were used to train different models, and the results were tallied. As can be seen in Table 10, the average accuracy of transformer fault diagnosis using BiGRU is 0.8502. In comparison with the traditional five main monitoring learning models, the power transformer fault diagnosis model using BiGRU is found to have the highest average accuracy. In terms of time consumption, the time consumption of fault diagnosis models based on deep learning methods is generally higher than that of machine learning methods. However, in comparison, BiGRU has the highest accuracy rate of all the methods. Although machine learning is easy and fast to use, the accuracy rate is low and does not meet the needs of today’s industry. In comparison, BiGRU has the highest accuracy and takes only a little more time than ELM and SVM. All things considered, BiGRU is the best fault diagnosis method in terms of the practical engineering performance of all the methods involved in the comparison.

Table 10. Repeat Training Results for Different Models.

Models	Highest Accuracy Rate	Minimum Accuracy Rate	Average Accuracy Rate	Time (s)
KNN	0.7356	0.7042	0.7232	26.7653
ELM	0.7914	0.7086	0.7543	13.3365
SVM	0.8030	0.7118	0.7246	15.8396
GRU	0.8312	0.7256	0.7879	23.1123
BiGRU	0.8923	0.8242	0.8502	22.8153

### 6.3. Analysis of Diagnostic Results of Different BiGRU Models

Since the hyperparameters of the BiGRU model had a great influence on its training and learning effect, the ISCSO method was used to optimize the parameter set. It was also compared with SCSO, WOA, GWO, and PSO methods. The maximum number of iterations was 100 and the adaptation degree value was based on the transformer fault diagnosis accuracy. The adaptation degree variation curve is shown in Figure 19.

As can be seen in Figure 20, PSO converges the slowest and has the lowest fitness value, reaching convergence after 55 iterations and converging the slowest. WOA improves diagnostic accuracy and convergence speed compared to PSO and GWO, reaching convergence after 32 iterations. The traditional SCSO algorithm is optimal after 31 iterations, resulting in poor local optimization performance due to its poor global optimization performance. In contrast, the improved ISCSO reaches convergence after 15 iterations and has the best adaptivity. This shows that the algorithm outperforms the other four methods in terms of global search performance and convergence speed.

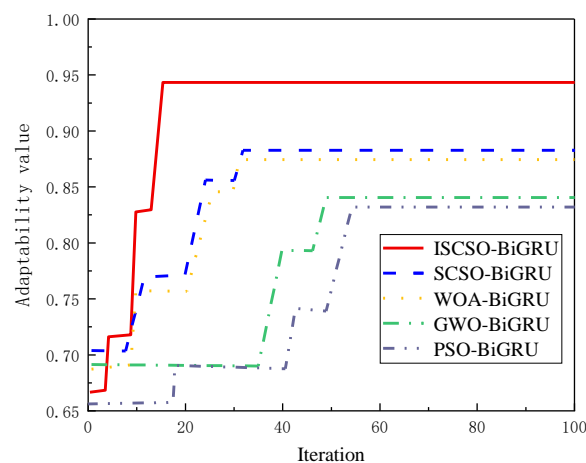


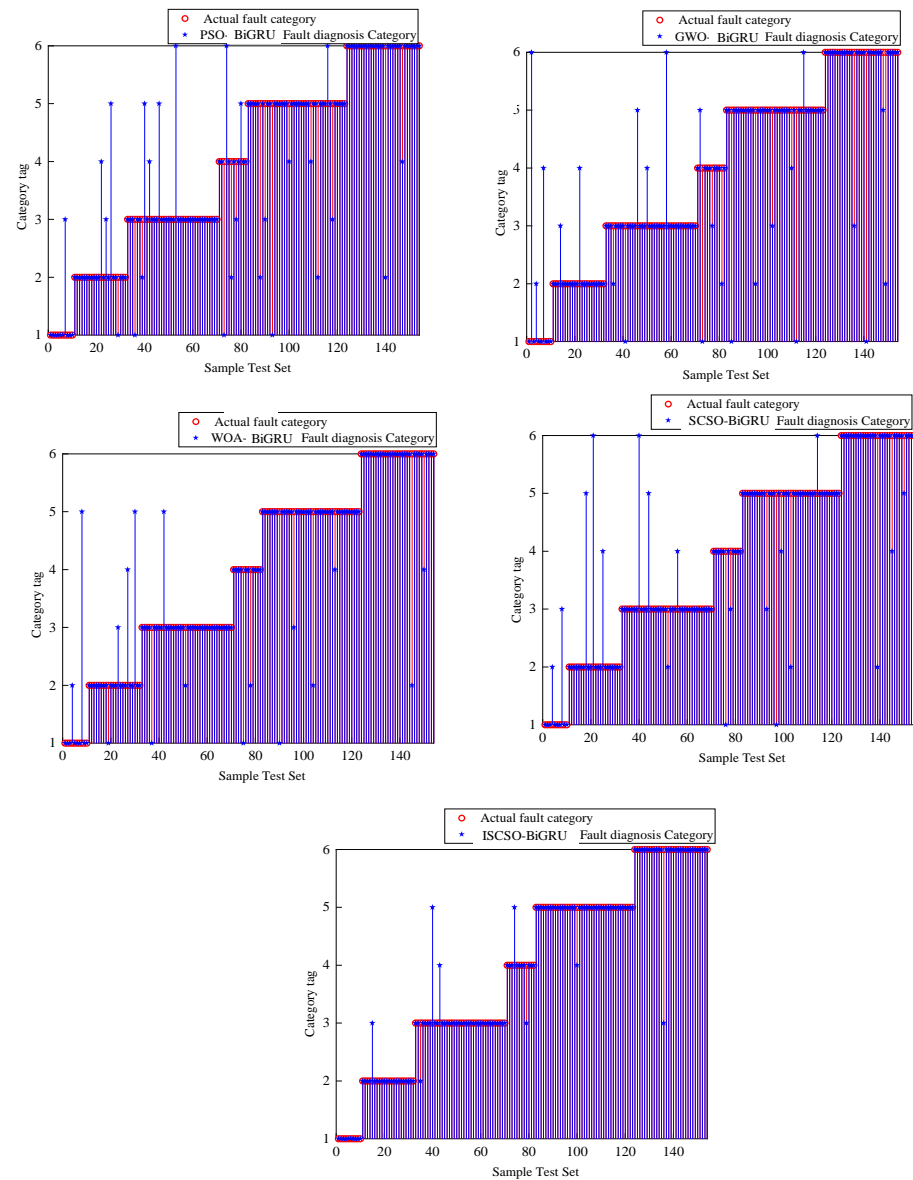
Figure 20. Fault diagnosis results for different models.

### 6.4. Comparative Analysis of Different Fault Diagnosis Models

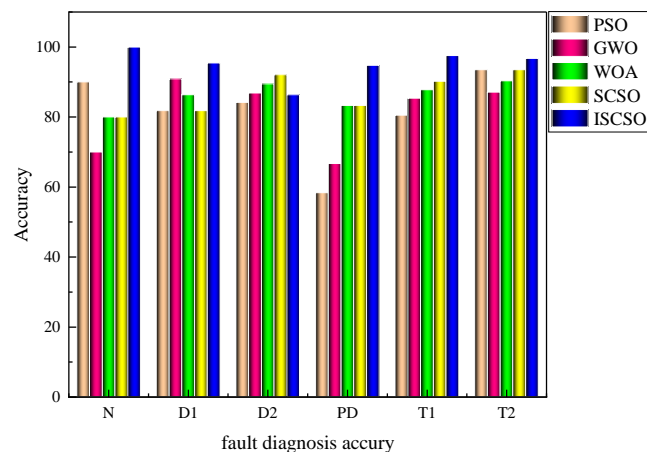
A total of 154 sets of test sample data were used for test training. The features based on L-ISOMP screening were input to BiGRU for diagnostic accuracy experiments of the fault diagnosis model. The ISCSO classification results were compared and analyzed with the SCSO-, WOA-, GWO-, and PSO-optimized BiGRU diagnostic results. The diagnostic results of each method are shown in Table 11, and the classification accuracies are shown in Figures 21 and 22.

**Table 11.** Results of Correct Transformer Fault Diagnosis.

Fault Type	PSO	GWO	WOA	SCSO	ISCSO
N	90	70	80	80	100
D <sub>1</sub>	81.81	90.90	86.36	81.81	95.45
D <sub>2</sub>	84.21	86.84	89.47	92.10	86.36
PD	58.33	66.67	83.33	83.33	94.73
T <sub>1</sub>	80.48	85.36	87.80	90.24	97.56
T <sub>2</sub>	93.54	87.09	90.32	93.54	96.77



**Figure 21.** Fault diagnosis results.



**Figure 22.** Various fault accuracy diagrams of BiGRU model.

As can be seen in Figures 21 and 22, after repeated iterations, ISCSO-BiGRU has obvious advantages in the application of transformer fault diagnosis. There were 26 fault diagnosis errors for the PSO-BiGRU model, 24 errors for GWO-BiGRU, 19 errors for WOA-BiGRU, 17 errors for SCSO-BiGRU, and 8 errors for ISCSO-BiGRU. The improved ISCSO algorithm improved the rate of correct fault diagnosis by 11.69%, 10.39%, 7.14%, and 5.9% compared with PSO, GWO, WOA, and SCSO algorithms, respectively.

After verifying and comparing the experimental results of the five models, the diagnostic accuracy of the ISCSO-BiGRU transformer fault diagnosis model proposed in this paper can reach 94.8%. In contrast, the diagnostic accuracies using the conventional PSO-BiGRU, GWO-BiGRU, WOA-BiGRU, and SCSO-BiGRU diagnostic models are 83.11%, 84.41%, 87.66%, and 88.9%, respectively. The results demonstrate that optimizing the hyperparameters of BiGRU using the ISCSO algorithm results in a higher correct fault identification rate than the PSO, GWO, WOA, and SCSO algorithms to optimize the parameters of BiGRU.

## 7. Conclusions

This paper advances a new power transformer fault diagnosis method based on the BiGRU. First, L-ISOMP is used to extract the features of DGA data. Then, aimed at the defects of SCSO, four improved methods are proposed: logistic chaotic mapping, the water wave dynamic factor, inertia weight, and the golden sine strategy. Finally, a transformer fault diagnosis method based on L-ISOMP and SCSO-BiGRU is constructed, and related experiments are used to test its diagnostic performance. Significantly, the following conclusions are obtained:

(1) Using L-ISOMP for the feature extraction of fault sample data can reduce the dimensionality of the feature vector, reduce the correlation between variables, and improve diagnostic accuracy after feature extraction. Compared with before dimensionality reduction, the filtering and dimensionality reduction of transformer fault features using the L-ISOMP method can better improve model performance.

(2) The improved sand cat swarm algorithm with logistic chaotic mapping, the water wave dynamic factor, inertia weight, and the golden sine strategy enriches the population diversity, improves the global search performance, balances the global search and local search balancing ability, and improves the defect of easily falling into the local optimum in the search process.

(3) The improved sand cat swarm algorithm ISCSO optimizes the parameters of BiGRU, which can improve the generalization ability and the rate of correct BiGRU classification. Compared and analyzed with the PSO-BiGRU, GWO-BiGRU, WOA-BiGRU, and SCSO-BiGRU fault diagnosis models, the diagnosis rates are 83.11%, 84.41%, 88.9%, and 87.66%, respectively. The ISCSO-BiGRU diagnosis rate can reach 94.8%, and can more effectively make a fast, accurate, and reliable diagnosis. It has a strong generalization ability and has certain practical significance in theoretical research and engineering.

In conclusion, the fault diagnosis method proposed in this paper has excellent diagnostic performance, can diagnose transformer faults accurately, and has high reference values. It can meet practical engineering needs, but as high accuracy requires a long training time, investigating how to further optimize the model and reduce the training time of the model with higher accuracy is the next research direction.

**Author Contributions:** Conceptualization, W.L. and C.S.; methodology, W.L.; software, C.S.; validation, W.L., C.S. and H.F.; investigation, H.F.; resources, Y.X.; data curation, C.S.; writing—original draft preparation, C.S.; writing—review and editing, W.L.; supervision, H.F.; project administration, Y.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** Basic Scientific Research project of Universities in Liaoning Province (LJKZ0352) and The National Natural Science Foundation of China (51974151) (51204160).

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request. The source of the data is not in dispute.

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

## References

- Lin, J.; Ma, J.; Zhu, J. Hierarchical Federated Learning for Power Transformer Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3520611. [\[CrossRef\]](#)
- Yang, D.; Qin, J.; Pang, Y.; Huang, T. A novel double-stacked autoencoder for power transformers DGA signals with an imbalanced data structure. *IEEE Trans. Ind. Electron.* **2021**, *69*, 1977–1987. [\[CrossRef\]](#)
- Jiang, Y.; Yin, S.; Kaynak, O. Optimized design of parity relation-based residual generator for fault detection: Data-driven approaches. *IEEE Trans. Ind. Inform.* **2020**, *17*, 1449–1458. [\[CrossRef\]](#)
- Yang, X.; Chen, W.; Li, A.; Yang, C.; Xie, Z.; Dong, H. BA-PNN-based methods for power transformer fault diagnosis. *Adv. Eng. Inform.* **2019**, *39*, 178–185. [\[CrossRef\]](#)
- Taha, I.B.; Ibrahim, S.; Mansour, D.E.A. Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in measurements. *IEEE Access* **2021**, *9*, 111162–111170. [\[CrossRef\]](#)
- IE Commission. Mineral oil-filled electrical equipment in service—Guidance on the interpretation of dissolved and free gases analysis. *IEC* **2015**, 60599, 2015.
- Duval, M. A review of faults detectable by gas-in-oil analysis in transformers. *IEEE Electr. Insul. Mag.* **2002**, *18*, 8–17. [\[CrossRef\]](#)
- Mansour, D.E.A. Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases. *IEEE Trans. Dielectr. Electr. Insul.* **2015**, *22*, 2507–2512. [\[CrossRef\]](#)
- Tao, L.; Yang, X.; Zhou, Y.; Yang, L. A novel transformers fault diagnosis method based on probabilistic neural network and bio-inspired optimizer. *Sensors* **2021**, *21*, 3623. [\[CrossRef\]](#)
- Wu, Y.; Sun, X.; Zhang, Y.; Zhong, X.; Cheng, L. A Power Transformer Fault Diagnosis Method-Based Hybrid Improved Seagull Optimization Algorithm and Support Vector Machine. *IEEE Access* **2021**, *10*, 17268–17286. [\[CrossRef\]](#)
- Tan, X.; Guo, C.; Wang, K.; Wan, F. A novel two-stage Dissolved Gas Analysis fault diagnosis system based semi-supervised learning. *High Volt.* **2022**, *7*, 676–691. [\[CrossRef\]](#)
- Dhini, A.; Surjandari, I.; Faqih, A.; Kusumoputro, B. Intelligent fault diagnosis for power transformer based on DGA data using support vector machine (SVM). In Proceedings of the 2018 3rd International Conference on System Reliability and Safety (ICRS), Barcelona, Spain, 23–25 November 2018; IEEE: Piscataway, NJ, USA; pp. 294–298.
- Bazan, G.H.; Goedtel, A.; Duque-Perez, O.; Morinigo-Sotelo, D. Multi-fault diagnosis in three-phase induction motors using data optimization and machine learning techniques. *Electronics* **2021**, *10*, 1462. [\[CrossRef\]](#)
- Yu, H.; Wu, Q.; Lu, Y.; Hu, C.; Wang, Y.; Liu, G. Research on Fault Diagnosis of Power Transformer Equipment Based on KNN Algorithm. In *Recent Developments in Mechatronics and Intelligent Robotics, Proceedings of the International Conference on Mechatronics and Intelligent Robotics (ICMIR2017), Kunming, China, 20–21 May 2017*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; Volume 2, pp. 172–176.
- Jiang, J.; Chen, R.; Chen, M.; Wang, W.; Zhang, C. Dynamic fault prediction of power transformers based on hidden Markov model of dissolved gases analysis. *IEEE Trans. Power Deliv.* **2019**, *34*, 1393–1400. [\[CrossRef\]](#)
- Zhang, Y.; Ding, X.; Liu, Y.; Griffin, P.J. An artificial neural network approach to transformer fault diagnosis. *IEEE Trans. Power Deliv.* **1996**, *11*, 1836–1841. [\[CrossRef\]](#)
- Huang, Y.C. Evolving neural nets for fault diagnosis of power transformers. *IEEE Trans. Power Deliv.* **2003**, *18*, 843–848. [\[CrossRef\]](#)
- Meng, K.; Dong, Z.Y.; Wang, D.H.; Wong, K.P. A self-adaptive RBF neural network classifier for transformer fault analysis. *IEEE Trans. Power Syst.* **2010**, *25*, 1350–1360. [\[CrossRef\]](#)
- Burriel-Valencia, J.; Puche-Panadero, R.; Martinez-Roman, J.; Sapena-Bano, A.; Pineda-Sanchez, M.; Perez-Cruz, J.; Riera-Guasp, M. Automatic fault diagnostic system for induction motors under transient regime optimized with expert systems. *Electronics* **2018**, *8*, 6. [\[CrossRef\]](#)

20. Dai, J.; Song, H.; Sheng, G.; Jiang, X. Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 2828–2835. [[CrossRef](#)]
21. Huang, X.; Wang, X.; Tian, Y. Research on transformer fault diagnosis method based on GWO optimized hybrid kernel extreme learning machine. In Proceedings of the 2018 Condition Monitoring and Diagnosis (CMD), Perth, WA, Australia, 23–26 September 2018; IEEE: Piscataway, NJ, USA; pp. 1–5.
22. Devi, A.S.; Maragatham, G.; Boopathi, K.; Rangaraj, A.G. Hourly day-ahead wind power forecasting with the EEMD-CSO-LSTM-EFG deep learning technique. *Soft Comput.* **2020**, *24*, 12391–12411. [[CrossRef](#)]
23. Silva, V.; Tenenbaum, J. Global versus local methods in nonlinear dimensionality reduction. *Adv. Neural Inf. Process. Syst.* **2002**, *15*, 721–728.
24. De Silva, V.; Tenenbaum, J.B. *Sparse Multidimensional Scaling Using Landmark Points*; Stanford University: Stanford, CA, USA, 2004; Volume 120.
25. Seyyedabbasi, A.; Kiani, F. Sand Cat swarm optimization: A nature-inspired algorithm to solve global optimization problems. *Eng. Comput.* **2022**, 1–15. [[CrossRef](#)]
26. Wu, D.; Rao, H.; Wen, C.; Jia, H.; Liu, Q.; Abualigah, L. Modified Sand Cat Swarm Optimization Algorithm for Solving Constrained Engineering Optimization Problems. *Mathematics* **2022**, *10*, 4350. [[CrossRef](#)]
27. Dokeroglu, T.; Sevinc, E.; Kucukyilmaz, T.; Cosar, A. A survey on new generation metaheuristic algorithms. *Comput. Ind. Eng.* **2019**, *137*, 106040. [[CrossRef](#)]
28. Zhang, C.; Ding, S. A stochastic configuration network based on chaotic sparrow search algorithm. *Knowl. -Based Syst.* **2021**, *220*, 106924. [[CrossRef](#)]
29. Wang, J.; Li, Y.; Hu, G.; Yang, M. An enhanced artificial hummingbird algorithm and its application in truss topology engineering optimization. *Adv. Eng. Inform.* **2022**, *54*, 101761. [[CrossRef](#)]
30. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [[CrossRef](#)]
31. Hoballah, A.; Mansour, D.E.A.; Taha, I.B. Hybrid grey wolf optimizer for transformer fault diagnosis using dissolved gases considering uncertainty in measurements. *IEEE Access* **2020**, *8*, 139176–139187. [[CrossRef](#)]
32. Yang, Z.; Zhou, Q.; Wu, X.; Zhao, Z. A novel measuring method of interfacial tension of transformer oil combined PSO optimized SVM and multi frequency ultrasonic technology. *IEEE Access* **2019**, *7*, 182624–182631. [[CrossRef](#)]
33. Hong, K.; Pan, J.; Jin, M. Transformer condition monitoring based on load-varied vibration response and GRU neural networks. *IEEE Access* **2020**, *8*, 178685–178694. [[CrossRef](#)]
34. Singh, P.; Chaudhury, S.; Panigrahi, B.K. Hybrid MPSO-CNN: Multi-level particle swarm optimized hyperparameters of convolutional neural network. *Swarm Evol. Comput.* **2021**, *63*, 100863. [[CrossRef](#)]
35. Duval, M.; DePabla, A. Interpretation of gas-in-oil analysis using new IEC publication 60,599 and IEC TC 10 databases. *IEEE Electr. Insul. Mag.* **2001**, *17*, 31–41. [[CrossRef](#)]
36. Ghoneim, S.S.; Taha, I.B. A new approach of DGA interpretation technique for transformer fault diagnosis. *Int. J. Electr. Power Energy Syst.* **2016**, *81*, 265–274. [[CrossRef](#)]

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