IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 17 March 2023, accepted 3 April 2023, date of publication 7 April 2023, date of current version 12 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3265469

RESEARCH ARTICLE

Fault Diagnosis Method for Power Transformers Based on Improved Golden Jackal Optimization Algorithm and Random Configuration Network

WANJIE LU[1,2], CHUN SHI[1], HUA FU[1], AND YAOSONG XU[1]
[1]School of Electrical Control, Liaoning Technical University, Huludao, Liaoning 125000, China
[2]School of Mechanical Engineering, Liaoning Technical University, Fuxin, Liaoning 123000, China

Corresponding author: Chun Shi (schun97@163.com)

This work was supported in part by the Basic Scientific Research Project of Universities in Liaoning Province under Grant LJKZ0352, and in part by the National Natural Science Foundation of China under Grant 51974151 and Grant 51204160.

ABSTRACT The problem of the low accuracy of Dissolved Gas Analysis (DGA) in diagnosing transformer faults is addressed by proposing an Improved Golden Jackal Optimization (IGJO) based Stochastic Configuration Network (SCN) method. The method of transformer fault diagnosis based on IGJO optimized SCN is proposed. Firstly, Kernel Principal Component Analysis (KPCA) is used to reduce the dimensionality of the gas data and extract the effective feature quantities. Secondly, the L2 parametric penalty term is introduced into the SCN to improve the generalisation ability of SCN in practical applications. The elite backward learning and golden sine algorithms are incorporated into the golden jackal algorithm, and the IGJO performance is tested using 13 typical test functions, demonstrating that the IGJO has greater stability and merit-seeking capability. The penalty term coefficient C of the SCN is optimised using the IGJO to develop a transformer fault diagnosis model with an Improved Golden Jackal algorithm optimised Random Configuration Network (IGJO-SCN). Finally, the feature quantities extracted by KPCA are used as the input set of the model and the different transformer fault diagnosis models are simulated and validated. The results show that the IGJO-SCN has higher transformer fault diagnosis accuracy compared to other models.

INDEX TERMS Transformer, fault diagnosis, dissolved gas analysis, improved golden jackal optimization algorithm.

I. INTRODUCTION

As an important electrical equipment in the power system, power transformers bear the heavy responsibility of voltage boost or voltage reduction [1]. Once a transformer failure occurs, it will seriously threaten the normal operation of the power system, so it is of great significance to improve the accuracy of transformer fault diagnosis for the stability and reliability of power system operation [2], [3].

Currently, dissolved gas analysis in oil has become a viable method for transformer fault diagnosis in the power industry [4]. Based on the principles of Dissolved Gas Analysis (DGA), many scholars have proposed a series of methods for fault diagnosis of transformers, such as the codeless ratio method [5], IEC triple ratio method [6], Duval triangle method [7], etc. Although these ratio discriminatory methods are simple, they often suffer from missing codes and lead to The accuracy of fault diagnosis is not high. Combined with mathematical models and theories, some scholars have also proposed transformer fault diagnosis methods based on grey correlation theory [8], fuzzy theory [9], rough set theory [10], Bayesian theory [11], and other aspects, but these methods generally have the disadvantage of low correct classification rate.

In recent years, with the development of artificial intelligence, a series of intelligent diagnosis methods using dissolved gas concentration values in oil as characteristic quantities have been generated. The literature [12] uses artificial neural networks for fault diagnosis, but the method has a long training time and is prone to fall into local optimal

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

solutions in the process of finding the best solution; the expert system used in the literature [13] has poor autonomous learning capability and is not suitable for diagnosing overly complex systems. The literature [14] uses Support Vector Machine (SVM) for fault diagnosis, which can well solve the common problems such as small samples and non-linear classification, but the selection of its two parameters, i.e. penalty factor and kernel function parameters, has a great impact on the classification accuracy of SVM. A transformer fault diagnosis model based on Improved Grey Wolf Optimization (IGWO) and Probabilistic Neural Network (PNN) was proposed in the literature [15]. The accuracy of transformer fault diagnosis was significantly improved. However, the authors did not study the initialisation of Grey Wolf Optimization (GWO) and the improved update strategy alone does not consistently improve the GWO optimisation performance. The literature [16] proposes a DGA-based Semisupervised Autoencoder with an Auxiliary Task (SAAT) for power transformer fault diagnosis. The results show that the SAAT-based fault diagnosis method has good diagnostic performance. The proposed health feature space helps to monitor the health status of the transformer intuitively. The above method only reflects the health condition of the transformer and cannot identify specific fault types, therefore the proposed method cannot be applied for accurate fault identification. An intelligent monitoring and classification algorithm based on Frequency Response Analysis (FRA) is proposed in the literature [17] for detecting transformer winding faults. The model is able to accurately identify various winding states in transformers. However, they did not investigate the disadvantages of Artificial Neural Network (ANN), whose parameters are difficult to determine, or the inability to optimise the diagnostic accuracy of the proposed model using ANN alone.

Synthesizing the above research methods in the literature [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], there are difficulties in solving multi-classification problems with SVM, while ANN requires a large amount of sample data, the model validity is not strong, and the stability and robustness of fuzzy theory is poor. Compared with other neural network architectures, SCN not only has good generalization characteristics of stochastic learning, but also the number of nodes in its implicit layer is generated gradually based on the supervision mechanism, which is a good solution to the problem that the number of nodes in the implicit layer of RVFL model is difficult to determine, and also has more extensive applications in the field of fault diagnosis. Therefore, the random configuration network is chosen as the basic model in this paper.

The value of the penalty term coefficient C of the SCN is also very elaborate and the parameters can be optimized using an optimization algorithm. Common optimisation algorithms include the particle swarm algorithm [18], the artificial bee colony algorithm [19] and the grey wolf algorithm [20]. The Golden Jackal Optimisation (GJO) algorithm, first proposed in 2022, is able to search the objective space better and has a higher success rate in solving optimisation problems [21]. Compared with traditional optimization algorithms, the Golden Jackal optimization algorithm has significant advantages in terms of global search capability and search accuracy, but similar to traditional methods, the drawback that the GJO is highly susceptible to local optimal solutions due to the reduction of population diversity in the late iteration still exists. Due to the high dimensionality, nonlinearity and complexity of the original transformer fault data, higher diagnostic accuracy cannot be achieved with a single machine learning approach, so it is necessary to reduce the dimensionality of the fault data.

Three shortcomings of SCN-based transformer fault diagnosis are summarised: 1) the SCN may be overfitted when the number of training samples is small and the value of the penalty coefficients is very delicate; 2) the noise of transformer fault data reduces the stability of the model; 3) the study of the optimization algorithm is not targeted and cannot significantly improve the optimization performance. Therefore, this paper proposes a transformer fault diagnosis method based on KPCA and IGJO-SCN methods. It is worth noting that the innovations and contributions of this paper are mainly divided into the following four improvement methods. Firstly, KPCA is used to extract the features of the DGA data to reduce the influence of noise on the diagnosis results. In addition, the GJO can be improved by the following two methods to obtain the IGJO. elite backward learning is proposed to improve the initial diversity of golden jackal populations. And a golden sine strategy is proposed to improve the convergence speed and accuracy of the GJO. The IGJO can then be obtained from the two improved methods mentioned above, and benchmark functions are used to test the optimization performance of the IGJO and other algorithms. The results show that the GJO has the best optimization performance. Finally, the introduction of L2 parametric penalty terms in SCNs is proposed to address the shortcomings of stochastic configuration networks. The elite backward learning, golden sine strategy is incorporated into the traditional golden jackal algorithm, and the performance of the improved golden jackal optimization algorithm is tested using a typical test function, demonstrating that the IGJO has better search capability and optimization finding accuracy. Afterward, the IGJO is used to dynamically optimize the parameters of the random configuration network, and the dimensionality-reduced feature quantities are fed into fault diagnosis models such as IGJO-SCN, GJO-SCN, WOA-SCN, MPA-SCN, and SCSO-SCN. The results show that the proposed method has the highest diagnostic accuracy, the shortest diagnostic time, the best effectiveness, and the strongest significance. These results can prove that the proposed method has the best diagnostic performance.

II. FAULT FEATURE EXTRACTION

In traditional intelligent diagnosis methods, the characteristic gases generally used include H_2, CH_4, C_2H_4, C_2H_6, C_2H_2, and the total hydrocarbon TH, which is the sum of the four gases CH_4, C_2H_4, C_2H_6, and C_2H_2. The impact of data

pre-processing on the correct fault diagnosis rate is ignored when these six gases are used as input parameters. Therefore, in this paper, these 6 gases are paired in pairs to generate a DGA ratio that may be used for transformer fault diagnosis. Since 2 volume fraction ratios of the same gas, e.g. H_2/CH_4 and CH_4/H_2, provide the same discriminability to the classifier, only 15 gas ratio combinations are listed, as shown in Table 1.

TABLE 1. Ratio of dissolved gases in oil.

Number	Volume to fraction ratio
X_1	H_2/CH_4
X_2	H_2/C_2H_4
X_3	H_2/C_2H_6
X_4	H_2/C_2H_2
X_5	CH_4/C_2H_4
X_6	CH_4/C_2H_6
X_7	CH_4/C_2H_2
X_8	C_2H_4/C_2H_6
X_9	C_2H_4/C_2H_2
X_{10}	C_2H_6/C_2H_2
X_{11}	H_2/TH
X_{12}	CH_4/TH
X_{13}	C_2H_4/TH
X_{14}	C_2H_6/TH
X_{15}	C_2H_2/TH

KPCA is a non-linear feature extraction method that effectively eliminates redundancy and spatial correlation between data and extracts non-linear feature primitives containing the main data information. It makes up for the shortcomings of traditional PCA algorithms in processing non-linear data, and is commonly used in the detection of anomalies in industrial control systems.

When processing the non-linear DGA data, considering the high dimensional characteristics of the data, KPCA is used to reduce the dimensionality of the DGA data. KPCA uses kernel functions to achieve dimensionality reduction of non-linear data in a high-dimensional space, avoiding the problem of linear indistinguishability of the feature vectors. Therefore, in this paper, KPCA is used for feature extraction of 15 feature ratios, and the results are shown in Figure 1.

As can be seen from Figure 1, the first six principal components correspond to eigenvalues greater than one, and the cumulative contribution rate has exceeded 95%, which can obtain the majority of the required feature information, so the dimensionality of the feature variable after dimensionality reduction is 6.

To illustrate the superiority of KPCA, it was compared with the traditional DGA-based transformer fault diagnosis methods (Rogers four-ratio and IEC three-ratio methods). The features of each DGA method were input into SCN for fault diagnosis, and the results of the comparison of model feature input and correct diagnosis rate are shown in Table 2.

The data before and after the dimensionality reduction of KPCA were used for fault diagnosis by SCN, and the running time and correct diagnosis rate were compared before and after the dimensionality reduction, as shown in Table 3.

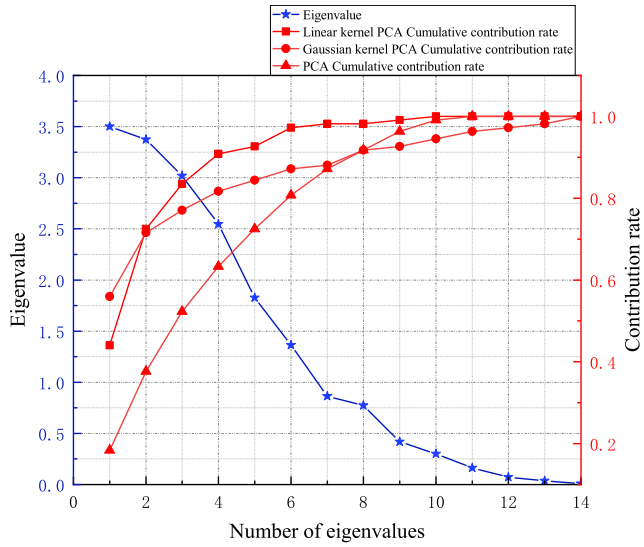


FIGURE 1. Correlation coefficient matrix eigenvalues and cumulative contribution rates.

TABLE 2. Comparison of fault diagnostic accuracy of different DGA methods.

Method	Feature input	Correct diagnosis rate
Rogers Four Ratio Method	CH_4/H_2, C_2H_6/CH_4, C_2H_4/C_2H_6, C_2H_2/C_2H_4	71.08
IEC Triple Ratio Method	C_2H_4/H_2, C_2H_4/C_2H_6, C_2H_2/C_2H_4	77.03
KPCA	H_2/CH_4, H_2/C_2H_4, H_2/C_2H_6, H_2/C_2H_2, CH_4/C_2H_4, CH_4/C_2H_6	90.56

TABLE 3. KPCA extraction before and after.

Parameters	KPCA before downscaling	KPCA after downscaling
Running time /s	34.256	13.225
Correct diagnosis rate /%	75.23	90.56

III. GOLDEN JACKAL ALGORITHM

The Golden Jackal optimization (GJO) algorithm is a metaheuristic algorithm proposed in 2022 based on the process of prey predation by pairs of golden jackals, which has the characteristics of fast optimisation and good convergence. The GJO algorithm is designed and optimised by modeling the specific hunting strategy of the paired golden jackal, and the algorithm is optimised as follows.

Stage 1: Exploration stage (prey search stage). The golden jackal searches and waits for its prey according to its nature and instincts, with the male leading and the female following the male. The couple follow their prey and update their position according to equation (1).

$$\begin{cases} Y_1(t) = Y_M(t) - E.\left|Y_M(t) - rl.\Pr ey(t)\right| \\ Y_2(t) = Y_{FM}(t) - E.\left|Y_{FM}(t) - rl.\Pr ey(t)\right| \end{cases} \tag{1}$$

In equation (1): $\Pr ey(t)$ is the location of the prey; $Y_M(t)$ and $Y_{FM}(t)$ are the current male and female jackal positions

TABLE 4. pseudo code of GJO.

```
Algorithm:GJO
Initialize the random prey and its fitness Yi(i=1,2,…,N)
While(t<T)
    Calculate the fitness values of preys
    Y1 = best prey(Male Jackal position)
    Y2 = second best prey(Female Jackal position)
    for(each prey)
      Update the escape energy E according to Eq.(2)
      Update rl according to Eq.(3)
      if(|E|>=1)
      Update the prey position according to Eq.(1)(4)
      if(|E|<1)
      Update the prey position according to Eq.(4)(5)
    end for
    t=t+1
end while
```

respectively; $Y_1(t)$ and $Y_2(t)$ are the positions of male and female jackals respectively after changing with the prey; E is the energy change of the prey when avoiding the grasp of the golden jackal, E is determined by equation (2)

$$\begin{cases} E = E_1 * E_0 \\ E_0 = 2 * rand - 1 \\ E_1 = c_1 * (1 - (t/T)) \end{cases} \tag{2}$$

In equation (2): E_0 is the initial energy of the prey; t is the current number of iterations; T is the maximum number of iterations; c_1 is a constant with a constant value of 1.5; E_1 gradually decreases from 1.5 to 0 during the iterations, indicating that the energy of the prey gradually decreases during the capture process. $rl.\Pr ey(t)$ is a function that simulates the trajectory of the prey where rl is determined by equation (3)

$$rl = 0.05^{*}LF(y) \tag{3}$$

In equation (3), $LF(y)$ is the levy flight distribution function.

Finally, the prey position is updated by equation (4)

$$Y(t+1) = (Y_1(t) + Y_2(t))/2 \tag{4}$$

Stage 2: The hunting stage (encirclement and capture of prey). When the prey is discovered by the golden jackal and starts to hide from the golden jackal, the energy gradually decreases during the hiding process and the male and female jackals start to surround the prey and capture it. This process is mathematically modelled as

$$\begin{cases} Y_1(t) = Y_M(t) - E.\,|rl.Y_M(t) - \Pr ey(t)| \\ Y_2(t) = Y_{FM}(t) - E.\,|rl.Y_{FM}(t) - \Pr ey(t)| \end{cases} \tag{5}$$

All variables have the same meaning as in equation (1), and finally the prey position is updated by equation (4).

The condition for the golden jackal to switch from the exploration phase to the hunting phase is determined by E. When $|E| > 1$, the golden jackal performs the exploration phase in different areas, and when $|E| < 1$, the golden jackal surrounds and attacks the prey to capture it. The pseudo code of GJO is follows Table 4.

IV. IMPROVED GOLDEN JACKAL ALGORITHM

A. ELITE REVERSE LEARNING STRATEGIES

The main idea of Opposition-Based Learning (OBL) [22] is to simultaneously compute and determine the candidate solution and the corresponding inverse solution, and select the best candidate solution from them by its fitness value. The OBL strategy can improve the population diversity more effectively and prevent the algorithm from converging too early.

Definition 1 (Inverse Solution): A feasible solution in a population is $\boldsymbol{X} = (x_1, x_2, \cdots, x_D)$, $x_j \in [a_j, b_j]$, $1 \le j \le D$, then its inverse solution is $\overline{\boldsymbol{X}} = (\overline{x}_1, \overline{x}_2, \cdots, \overline{x}_D)$, $\overline{x}_j = r(a_j + b_j) - x_j$, and r is a random number obeying a uniform distribution of [0,1].

OBL can better extend the search range of the population and improve the performance of the algorithm. However, OBL has a certain degree of randomness in generating its opposite solution. Each randomly generated candidate has a 50% probability of being far away from the optimal solution of the problem compared to its opposite individual.

The main idea of Elite Opposition-Based Learning (EOBL) [23], which aims to improve the solution quality of reverse learning, is to generate more promising solutions by evaluating the opposite solutions of elite solutions. The opposite solution is more likely to be located where the global optimum is located. This mechanism has been successfully applied to improvements of the Whale algorithm (WOA) [24], the Grey Wolf algorithm (GWO) [25] and the Harris Hawk algorithm (HHO) [26], among others.

Definition 2 (Elite Inverse Solution): The extreme value point of an ordinary individual in the population is defined as the corresponding elite individual, $X_i^{\mathrm{E}} = \left(X_{i,1}^{\mathrm{E}}, X_{i,2}^{\mathrm{E}}, \cdots, X_{i,D}^{\mathrm{E}}\right)$ $(i = 1, 2, \cdots, N)$ in the population, then its elite inverse solution $\overline{X}_i^{\mathrm{E}} = \left(\overline{X}_{i,1}^{\mathrm{E}}, \overline{X}_{i,2}^{\mathrm{E}}, \cdots, \overline{X}_{i,D}^{\mathrm{E}}\right)$ is defined as shown in equation (6):

$$\overline{X}_{i,j}^{E} = \delta \times (lb_j + ub_j) - E_{i,j}^{\mathrm{E}} \tag{6}$$

where, δ is a random number of [0,1]; $X_{i,j}^{\mathrm{E}} \in [lb_j, ub_j]$; $lb_j = \min\{X_{i,j}\}$, $ub_j = \max\{X_{i,j}\}$, are the lower and upper bounds of the dynamic boundary, respectively. If $\overline{X}_{i,j}^{E}$ exceeds the bound, then the mathematical model is corrected using the stochastic method as shown in equation (7):

$$\overline{X}_{i,j}^{E} = rand(lb_j + ub_j) \tag{7}$$

B. GOLDEN SINE PREDATION MECHANISM

The golden sine algorithm (gloden-SA) [27] is a metaheuristic algorithm proposed by Tanyildizi in 2017. Based on the relationship between the sine function and the unit circle, gloden-SA can traverse all points on the sine function, i.e. traverse the entire unit circle. The scanning of the whole unit circle is similar to the search of the search space in an optimisation problem. At the same time, the introduction of the golden partition coefficient in its position updating

process makes each iteration both narrow the search area and traverse the excellent solution region, which improves the local mining ability and solution accuracy while speeding up the convergence of the algorithm.

The core process of the gloden-SA algorithm is the way in which its positions are updated. First, n individual positions are randomly generated, each corresponding to an initial solution, and the position is updated for each individual by means of equation (8).

$$x_i^{t+1} = x_i^t \cdot |\sin(r_1)| - r_2 \cdot \sin(r_1) \cdot \left|c_1 P_i^t - c_2 x_i^t\right| \quad (8)$$

where: $x_i^t = \left[x_{i,1}^t, x_{i,2}^t, \cdots, x_{i,d}^t\right]$, , $((i = 1, 2, \cdots, n)$, $(t = 1, 2, \cdots, t_{\max}))$, $x_{i,d}^t$ denote the position of the ith individual in the dth dimension at the tth iteration; P_i^t is the global optimal position at the tth iteration; r_1, r_2 are random numbers belonging to [0, 2π] and [0, π] respectively, r_1 determines the distance of movement of individual i at the next iteration, and r_2 determines the direction of movement of individual i at the next iteration [18]; c_1, c_2 are introduced as golden partition coefficients into the position update formula, $c_1 = a(1-g)+bg$, $c_2 = ag+b(1-g)$, the initial values of a and b are -π, π respectively, and the golden partition numbers $g = (\sqrt{5}-1)/2$, c_1, c_2 narrow the search space and guide the individual at the current position to the global optimal position during the iteration.

C. IGJO ALGORITHM

In response to the shortcomings of the basic golden jackal optimisation algorithm, the paper proposes an elite reverse learning mechanism with golden sine for Golden Jackal Optimisation (IGJO). The basic GJO algorithm uses a random initialisation method to initialise the population without a priori knowledge, which is prone to the problem of poor diversity of the golden jackal population. The quality of the initial population has a large impact on the algorithm's optimisation search performance, and a good initial population is beneficial to global optimisation search. Therefore, the thesis introduces an elite backward learning mechanism to improve the golden jackal algorithm, using a population selection method to rank the current golden jackal population and its inverse population by their fitness values, and select the best golden jackal individuals as the new generation individuals, so as to improve the population quality of the GJO algorithm. Firstly, the GJO algorithm incorporates the EBOL strategy to increase the diversity of the initialized population, increase the search space and lay the foundation for global optimization; secondly, at each iteration, the EBOL strategy can generate the inverse solution far away from the local extremum, guiding the GJO algorithm to jump out of the local extremum and enhancing the ability of the algorithm to search globally. In addition, the EBOL strategy uses a dynamic boundary tracking search mode to locate individuals in a progressively smaller search region, improving the convergence accuracy and speed of the GJO algorithm.

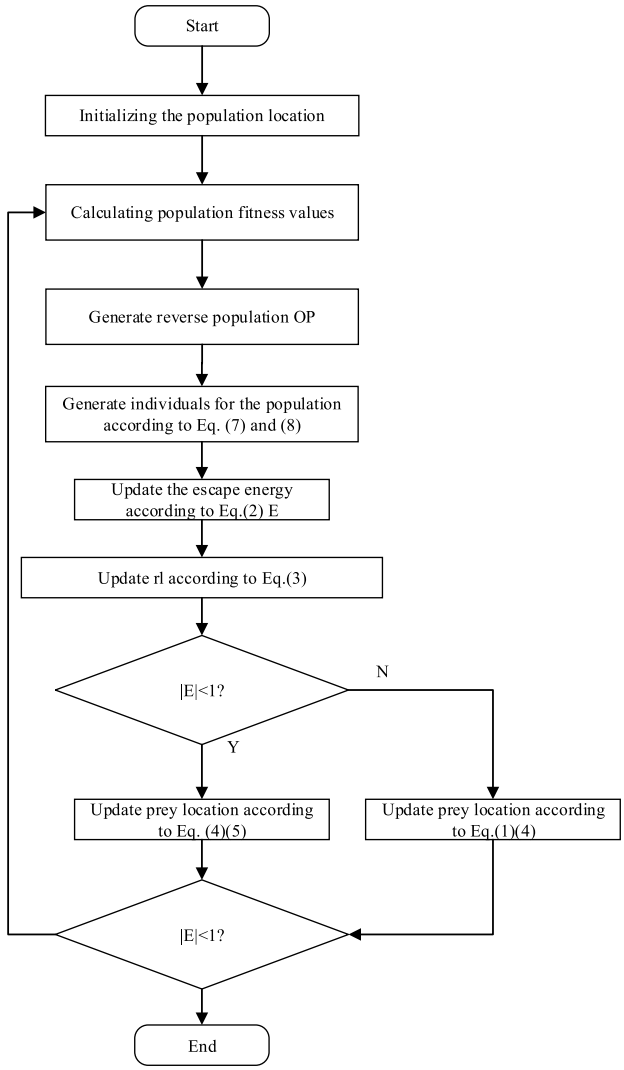


FIGURE 2. IGJO flow chart.

The position of the prey is the guiding mark of the movement of the golden jackal individuals. In the golden jackal algorithm incorporating the golden sine algorithm, the golden jackal moves in a golden sine manner to arrest the prey, and in the process of each iteration, the ordinary individual will exchange information with the best individual, fully absorbing the information of the position gap with the best individual to improve the algorithm's performance and accuracy of the search for the best. In addition, the search area is gradually reduced according to the golden partition coefficient, and the update distance and direction of the golden jackal individuals are adjusted according to r1 and r2 to guide the individuals to quickly approach the optimal value in the high-quality solution area, which optimizes the traditional GJO algorithm's search method and improves the algorithm's search speed and exploitation capability. Figure 2 is the flowchart of IGJO.

D. TIME COMPLEXITY ANALYSIS OF IGJO

The time complexity of the original golden jackal algorithm is O(N × D × T). In the IGJO algorithm, the time complexity of

the elite reverse learning strategy is O(N × D), the time complexity of calculating the fitness value is O(N × D), and the time complexity of calculating the individual position update is O(N × D). The time complexity of the IGJO algorithm is the same as that of the basic GJO algorithm, indicating that the 2 improved strategies do not increase the computational burden of the golden jackal algorithm.

V. SIMULATION EXPERIMENTS AND ANALYSIS OF RESULTS

A. INITIAL PARAMETER SETTING

In this paper, Whale Optimization Algorithm (WOA) [28], the Sand Cat Swarm Optimization (SCSO) [29], Marine Predators Algorithm (MPA) [30], Sine Cosine Algorithm (SCA) [31], Salp Swarm Algorithm (SSA) [32] and Golden Jackal Optimization Algorithm (GJO) were selected for comparison with the elite reverse learning and golden sine golden jackal algorithm IGJO. The initial population size was set to 30 and the maximum number of iterations was set to 500. the parameters of the above algorithms are shown in Table 5.

TABLE 5. Parameter settings.

Algorithm	Parameters
WOA	The parameter α changed from 2 to 0, $b = 1$
SCSO	The sensitivity range rG changed from 2 to 0, R changed from $-2rG$ to $2rG$
MPA	$FADs = 0.2$ $P = 0.5$
SCA	Constant $\alpha = 2$ and probability update was 0.5
SSA	Probability update was 0.5
GJO	$C1$ variable is varied from 1 to 2
IGJO	$C1$ variable is varied from 1 to 2

B. TEST FUNCTIONS

In this paper, 13 commonly used benchmarking functions are selected to test the performance of each algorithm, and the relevant properties of the functions are shown in Table 6.

C. ANALYSIS OF EXPERIMENTAL RESULTS

Each of the above algorithms was run 30 times on each benchmark test function to obtain the average value, and the experimental results are shown in Table 7.

As can be seen from Table 7, for the 12 test functions, IGJO outperforms WOA, SCSO, MPA, SCA, SSA and GJO in terms of the optimal value, mean and standard deviation. For the test functions F1 to F4, IGJO can calculate the optimal value of 0. The optimal value and average value are hundreds of orders of magnitude higher than the other five algorithms, and the standard deviation is 0, so the robustness is better. For the test functions F6 and F7, IGJO outperforms the other six algorithms. For functions F8 to F13, the accuracy and robustness of the IGJO optimization are optimal. The above shows that the two methods, elite backward learning and golden sine, are effective in optimizing the GJO.

Figures 3 and 4 show the convergence curves of the test functions used, which more clearly demonstrate the effect of the IGJO on the search for an optimum. From Figures 3 and 4, it can be seen that the convergence accuracy and convergence speed of the improved IGJO are better than those of WOA, SCSO, MPA, SCA, SSA, and GJO, indicating that the IGJO with the introduction of two strategies outperforms the single-strategy improved Golden Jackal algorithm and has the excellent merit-seeking capability.

A box plot is a statistical chart that uses five statistics in the data: minimum, upper quartile, median, lower quartile, and maximum to describe the data. The top and bottom segments of a box plot represent the maximum and minimum values of the data respectively. The top and bottom segments of the box plot represent the third quartile and first quartile respectively. The thick line in the middle of the box plot represents the median of the data. It provides a visual indication of the outliers, the dispersion of the distribution, and the symmetry of the data. The stability of the algorithms was statistically analyzed using the box plots, where the WOA, SCSO, MPA, GJO, and IGJO algorithms were selected for comparative testing and the results of 30 independent experiments were recorded. Figure 5 shows the statistical results of the algorithms on F1, F5, F7, and F10. On all the functions tested, the deviation between the minimum and maximum values obtained by the proposed IGJO algorithm is relatively small, while the deviation between the minimum and maximum values obtained by the WOA algorithm on F1 is larger and the deviation between the minimum and maximum values obtained by the MPA algorithm on F7 and F10 is larger. This indicates that the proposed IGJO is stable and can keep the deviations within a reasonable range with high performance.

D. STATISTICAL TESTS AND MAE RANKING

In the comparison of the computational power of the above seven algorithms, it is not sufficient to use only the three indicators of the best value, the worst value, the mean and the standard deviation to evaluate them, so the Wilcoxon rank sum test was used in this paper to test the results, as shown in Table 8. When $p < 0.05$, it means that there is a difference between the two algorithms in terms of the effect of finding the best value, and vice versa.

As can be seen from Table 8, for all functions except F9, F10 and F11, the p-values between IGJO and the other algorithms are much less than 0.05, indicating that IGJO has better performance in finding the optimum. The presence of NaN in Table 7 indicates that the corresponding algorithms have found the global optimum.

Mean Absolute Error (MAE) [33] represents the average of the absolute values of the differences between the results and the actual values. The mathematical definition of MAE is shown in equation (14):

$$MAE = \frac{\sum_{i=1}^{n} \mid m_i - k_i \mid}{n} \tag{9}$$

TABLE 6. Test functions.

Function	Dim	Range	fmin
$F_1(x)=\sum_{i=1}^{n}x_i^2$	30	[-100,100]	0
$F_2(x)=\sum_{i=1}^{n}\lvert x_i\rvert+\prod_{i=1}^{n}\lvert x_i\rvert$	30	[-10,10]	0
$F_3(x)=\sum_{i=1}^{n}(\sum_{j=1}^{n}x_j)^2$	30	[-100,100]	0
$F_4(x)=\max_i\{\lvert x_i\rvert,1\le i\le n\}$	30	[-100,100]	0
$F_5(x)=\sum_{i=1}^{n-1}[100(x_{i+1}-x_i^2)^2+(x_i-1)^2]$	30	[-30,30]	0
$F_6(x)=\sum_{i=1}^{n}(x_i+0.5)^2$	30	[-100,100]	0
$F_7(x)=\sum_{i=1}^{n}ix_i^4+rand[0,1)$	30	[-1.28,1.28]	0
$F_8(x)=\sum_{i=1}^{n}x_is^2-10cos(2\pi x_i)+10$	30	[-500,500]	-418.9829xDim
$F_9(x)=\sum_{i=1}^{n}x_i^2-10cos(2\pi x_i)+10$	30	[-5.12,5.12]	0
$F_{10}(x)=-20\exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2})-\exp(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i))+20+e$	30	[-32,32]	0
$F_{11}(x)=\frac{1}{4000}\sum_{i=1}^{n}x_i^2-\prod_{i=1}^{n}\cos(\frac{x_i}{\sqrt{i}})+1$	30	[-600,600]	0
$F_{12}(x)=\frac{\pi}{n}\left\{10\sin(\pi y_1)+\sum_{i=1}^{n}(y_i-1)^2[1+10\sin^2(\pi_{i+1})]\right\}+$ $\sum_{i=1}^{n}u(x_i,10,100,4)$ $y_i=1+\frac{x_i+1}{4},u(x_i,a,k,m)=\begin{cases}k(x_i-a)^m & x_i>a\\ 0 & -a<x_i<a\\ k(-x_i-a)^m & x_i<-a\end{cases}$	30	[-50,50]	0
$F_{13}(x)=0.1\times\{10\sin^2 3\pi x_i+\sum_{i=1}^{n}(x_i-1)^2[1+\sin^2(3\pi x_i+1)]$ $+(x_n-1)^2[1+\sin^2(3\pi x_n)]\}+\sum_{i=1}^{n}u(x_i,5,100,4)$	30	[-50,50]	0

where m_i is the average of the optimal solutions computed for each algorithm on each test function, k_i is the theoretical optimal value of each test function, and n is the number of test functions. The smaller the value of MAE, the higher the computational performance of the algorithm. The results of the ranking of the algorithms according to the MAE values are shown in Table 9.

From Table 8, we can see that the MAE value of IGJO is the smallest, which means that IGJO has better performance in finding the best.

VI. REGULARISED RANDOM CONFIGURATION NETWORK

The stochastic configuration network (SCN) is a powerful class of stochastic learning models with a stronger generalization performance than traditional stochastic learning models, as its hidden layer structure can be generated adaptively based on training effects [34]. The basic idea is to start with a smaller network and then gradually add new hidden nodes with random parameters until an acceptable tolerance is reached.

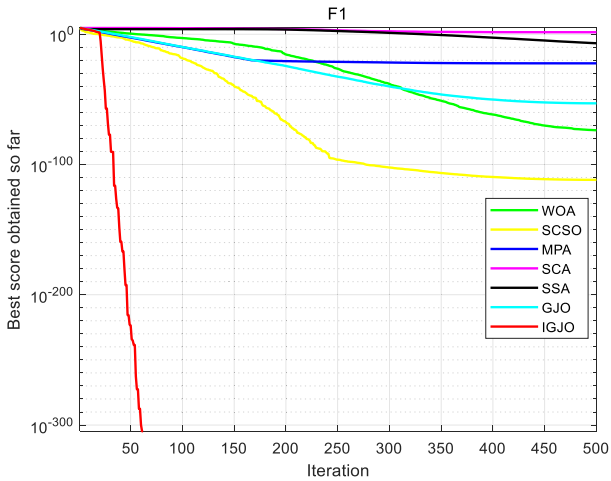
For a dataset $D=\{(x_i,y_i)\}$, $i=1,\cdots,I$, $x_i\in R^{1\times d}$, $y_i\in R^{1\times m}$; where x_i denotes the dataset feature attribute data and y_i denotes the data label attribute.

$$f_{L-1}(\boldsymbol{x})=\sum_{l=1}^{L-1}\beta_l\sigma_l(\boldsymbol{w}_l^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}_l) \tag{10}$$

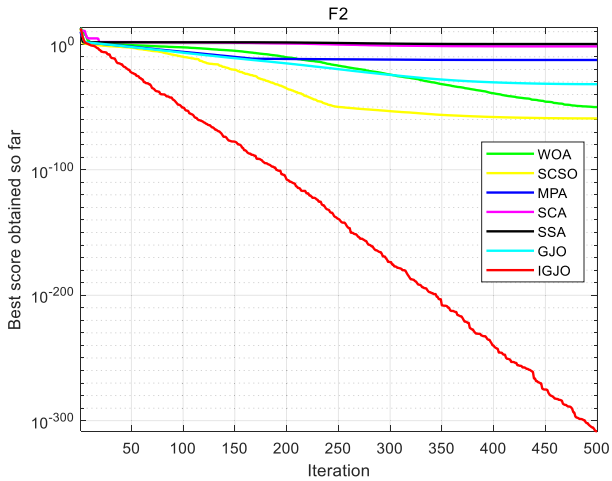
where: $\beta_l=[\beta_{l,1},\beta_{l,2},\cdots,\beta_{l,n}]$- output weight; L- number of layers of the neural network, a positive integer; $\sigma_l(\cdot)$- activation function of the lth hidden neuron; $\mathbf{w}_l\in[-v,v]^{m\times d}$- the weight of the lth hidden neuron; $b_l\in[-v,v]^{m\times d}$- bias of

TABLE 7. Statistical results of the optimized standard functions.

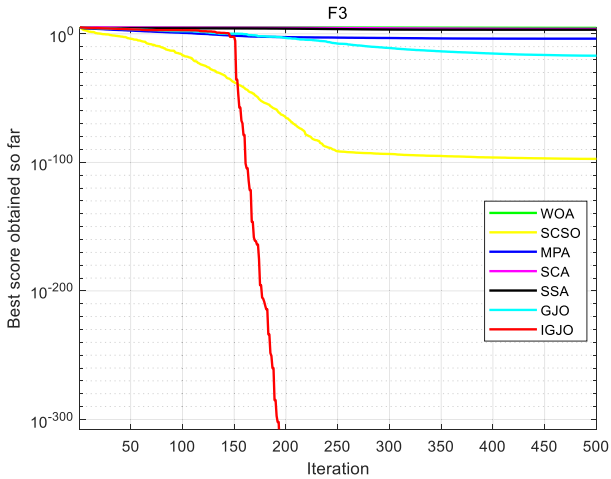
Function	Performance	WOA	SCSO	MPA	SCA	SSA	GJO	IGJO
F1	Best	1.11E-82	1.23E-126	1.71E-24	8.65E-02	2.11E-08	1.25E-56	0.00E+00
	Wrost	4.53E-73	4.80E-111	2.03E-22	4.58E+02	5.01E-07	1.19E-52	0.00E+00
	Ave	2.07E-74	1.61E-112	4.57E-23	3.52E+01	1.18E-07	9.12E-54	0.00E+00
	Std	8.42E-74	8.76E-112	4.99E-23	8.87E+01	1.08E-07	2.63E-53	0.00E+00
F2	Best	1.44E-58	1.23E-66	1.21E-15	4.08E-04	7.49E-02	5.65E-34	0.00E+00
	Wrost	2.55E-49	1.27E-58	2.36E-12	1.45E-01	6.06E+00	5.69E-32	1.30E-307
	Ave	9.22E-51	8.24E-60	3.12E-13	2.13E-02	1.64E+00	1.63E-32	0.00E+00
	Std	4.64E-50	2.79E-59	4.39E-13	3.45E-02	1.23E+00	1.51E-32	0.00E+00
F3	Best	1.84E+04	7.44E-111	1.23E-07	8.53E+02	2.45E+02	3.05E-23	0.00E+00
	Wrost	7.27E+04	1.45E-96	1.02E-03	1.95E+04	3.36E+03	1.54E-16	0.00E+00
	Ave	4.37E+04	5.13E-98	1.70E-04	7.81E+03	1.60E+03	9.85E-18	0.00E+00
	Std	1.24E+04	2.64E-97	2.84E-04	4.72E+03	1.00E+03	2.87E-17	0.00E+00
F4	Best	2.42E+00	6.14E-56	8.09E-10	1.13E+01	3.70E+00	1.72E-18	0.00E+00
	Wrost	9.13E+01	3.11E-50	1.08E-08	5.57E+01	1.68E+01	4.31E-15	0.00E+00
	Ave	4.43E+01	3.59E-51	4.00E-09	3.33E+01	1.14E+01	4.26E-16	0.00E+00
	Std	2.98E+01	8.15E-51	2.37E-09	1.39E+01	3.04E+00	8.51E-16	0.00E+00
F5	Best	2.72E+01	2.62E+01	2.45E+01	1.73E+02	2.56E+01	2.63E+01	3.85E-03
	Wrost	2.88E+01	2.89E+01	2.62E+01	4.89E+05	7.03E+03	2.88E+01	3.41E-01
	Ave	2.79E+01	2.81E+01	2.52E+01	6.80E+04	4.29E+02	2.78E+01	1.30E-01
	Std	5.53E-01	8.70E-01	4.10E-01	1.15E+05	1.27E+03	7.74E-01	9.81E-02
F6	Best	5.22E-02	7.60E-01	1.37E-08	4.44E+00	3.35E-08	1.61E+00	6.93E-05
	Wrost	9.47E-01	2.76E+00	2.37E-07	1.34E+02	4.91E-07	3.50E+00	1.14E-03
	Ave	4.36E-01	1.97E+00	5.16E-08	1.97E+01	1.16E-07	2.62E+00	6.17E-04
	Std	2.31E-01	5.22E-01	4.70E-08	2.87E+01	9.57E-08	4.49E-01	2.38E-04
F7	Best	2.07E-04	4.92E-06	3.52E-04	4.63E-03	5.07E-02	1.78E-05	3.26E-06
	Wrost	2.07E-02	1.52E-03	2.77E-03	2.06E+00	2.92E-01	2.96E-03	3.25E-04
	Ave	5.16E-03	1.47E-04	1.37E-03	1.58E-01	1.71E-01	6.37E-04	9.15E-05
	Std	6.08E-03	2.73E-04	6.19E-04	3.70E-01	6.62E-02	6.25E-04	7.57E-05
F8	Best	-1.26E+04	-9.28E+03	-1.02E+04	-4.57E+03	-9.74E+03	-6.46E+03	-6.78E+03
	Wrost	-7.71E+03	-5.51E+03	-7.98E+03	-3.38E+03	-6.27E+03	-2.44E+03	-2.90E+03
	Ave	-1.07E+04	-6.90E+03	-9.03E+03	-3.80E+03	-7.56E+03	-3.99E+03	-3.56E+03
	Std	1.78E+03	8.83E+02	5.06E+02	2.95E+02	7.17E+02	1.08E+03	1.39E+05
F9	Best	0.00E+00	0.00E+00	0.00E+00	3.45E-01	2.98E+01	0.00E+00	0.00E+00
	Wrost	0.00E+00	0.00E+00	0.00E+00	1.04E+02	1.29E+02	0.00E+00	0.00E+00
	Ave	0.00E+00	0.00E+00	0.00E+00	2.98E+01	5.71E+01	0.00E+00	0.00E+00
	Std	0.00E+00	0.00E+00	0.00E+00	3.11E+01	1.98E+01	0.00E+00	0.00E+00
F10	Best	8.88E-16	8.88E-16	4.77E-13	7.07E-02	5.68E-01	4.44E-15	8.88E-16
	Wrost	7.99E-15	8.88E-16	3.44E-12	2.03E+01	4.34E+00	7.99E-15	8.88E-16
	Ave	4.56E-15	8.88E-16	1.38E-12	1.12E+01	2.64E+00	7.40E-15	8.88E-16
	Std	2.72E-15	0.00E+00	7.20E-13	9.57E+00	9.24E-01	1.35E-15	0.00E+00
F11	Best	0.00E+00	0.00E+00	0.00E+00	3.14E-02	5.98E-04	0.00E+00	0.00E+00
	Wrost	0.00E+00	0.00E+00	0.00E+00	2.11E+00	3.89E-02	0.00E+00	0.00E+00
	Ave	0.00E+00	0.00E+00	0.00E+00	9.75E-01	1.48E-02	0.00E+00	0.00E+00
	Std	0.00E+00	0.00E+00	0.00E+00	3.86E-01	9.66E-03	0.00E+00	0.00E+00
F12	Best	5.54E-03	4.26E-02	1.66E-09	6.82E-01	4.09E+00	7.46E-02	1.29E-05
	Wrost	4.91E-02	3.29E-01	1.26E-05	1.15E+06	1.77E+01	3.27E-01	4.58E-02
	Ave	2.03E-02	1.21E-01	6.78E-07	4.31E+04	8.54E+00	1.92E-01	3.83E-03
	Std	1.03E-02	6.54E-02	2.60E-06	2.10E+05	3.40E+00	7.15E-02	9.38E-03
F13	Best	1.58E-01	1.51E+00	4.52E-08	4.67E+00	1.42E-02	1.27E+00	1.66E-05
	Wrost	1.05E+00	2.88E+00	8.73E-02	2.11E+05	4.93E+01	2.12E+00	1.23E-02
	Ave	4.98E-01	2.38E+00	1.34E-02	1.61E+04	1.73E+01	1.73E+00	1.59E-03
	Std	2.34E-01	3.78E-01	2.22E-02	5.01E+04	1.53E+01	2.11E-01	2.24E-03



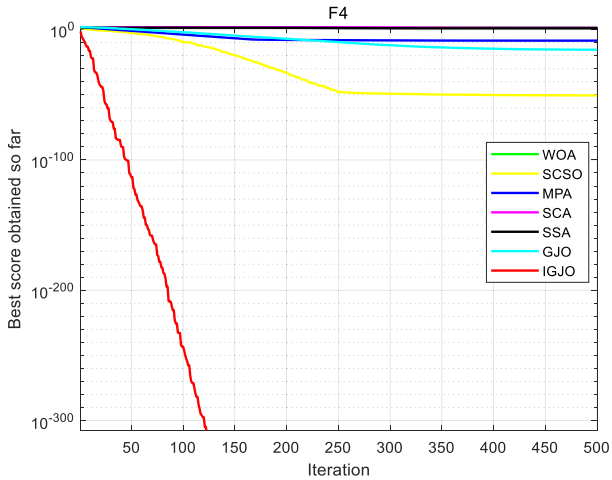
(a) Convergence curves of F1



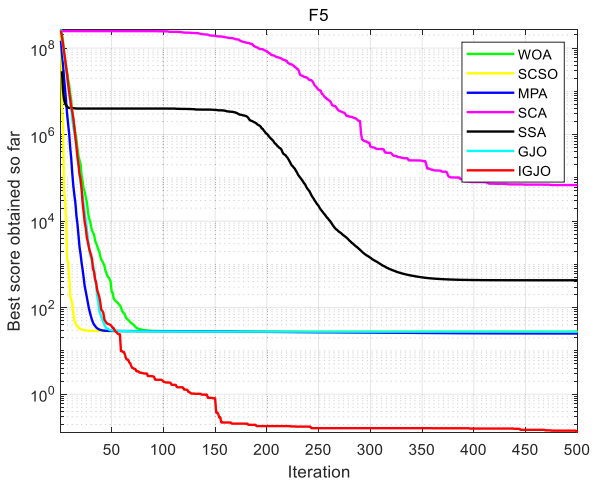
(b) Convergence curves of F2



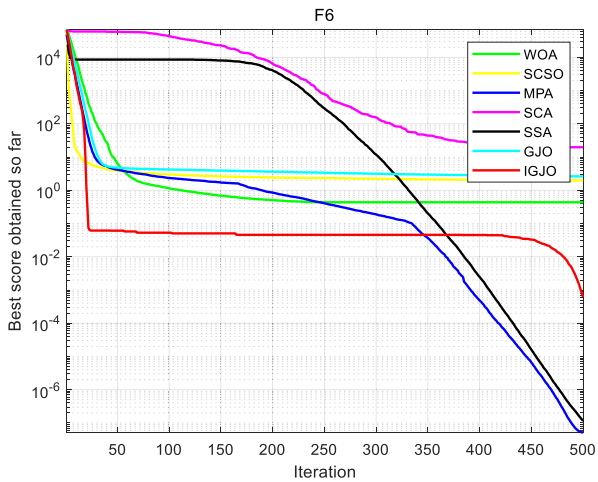
(c) Convergence curves of F3



(d) Convergence curves of F4

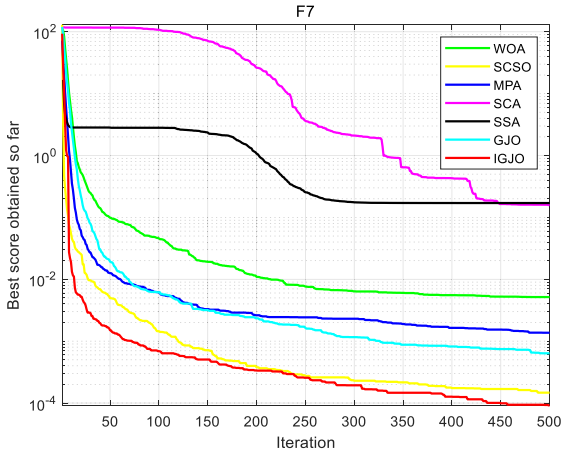


(e) Convergence curves of F5

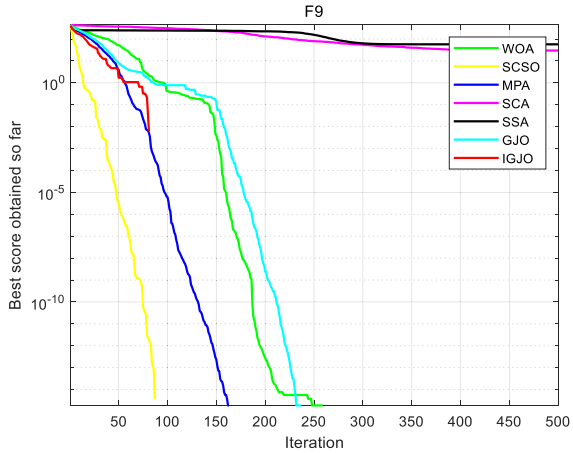


(f) Convergence curves of F6

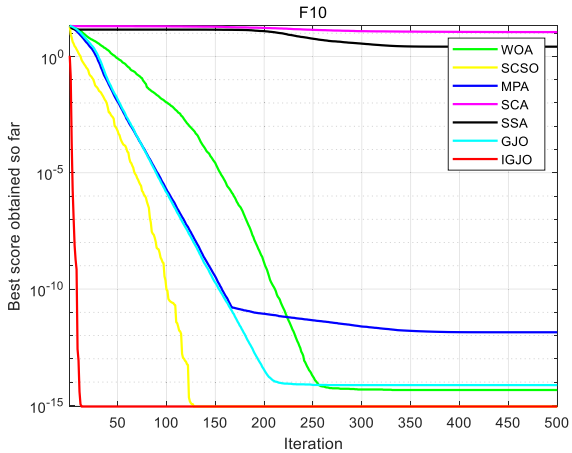
FIGURE 3. **Simulation experiment results.**



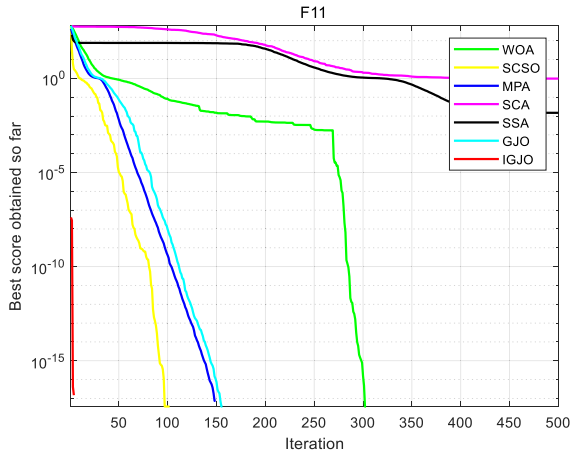
(g) Convergence curves of F7



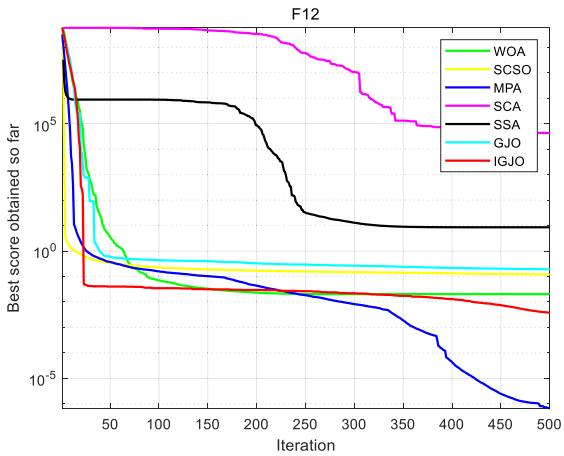
(h) Convergence curves of F9



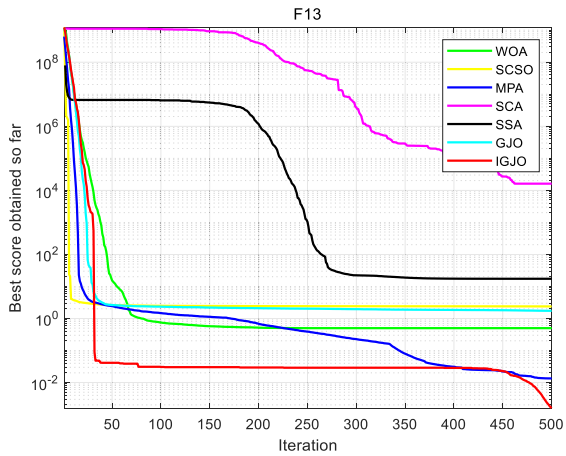
(i) Convergence curves of F10



(j) Convergence curves of F11



(k) Convergence curves of F12



(l) Convergence curves of F13

FIGURE 4. Simulation experiment results.

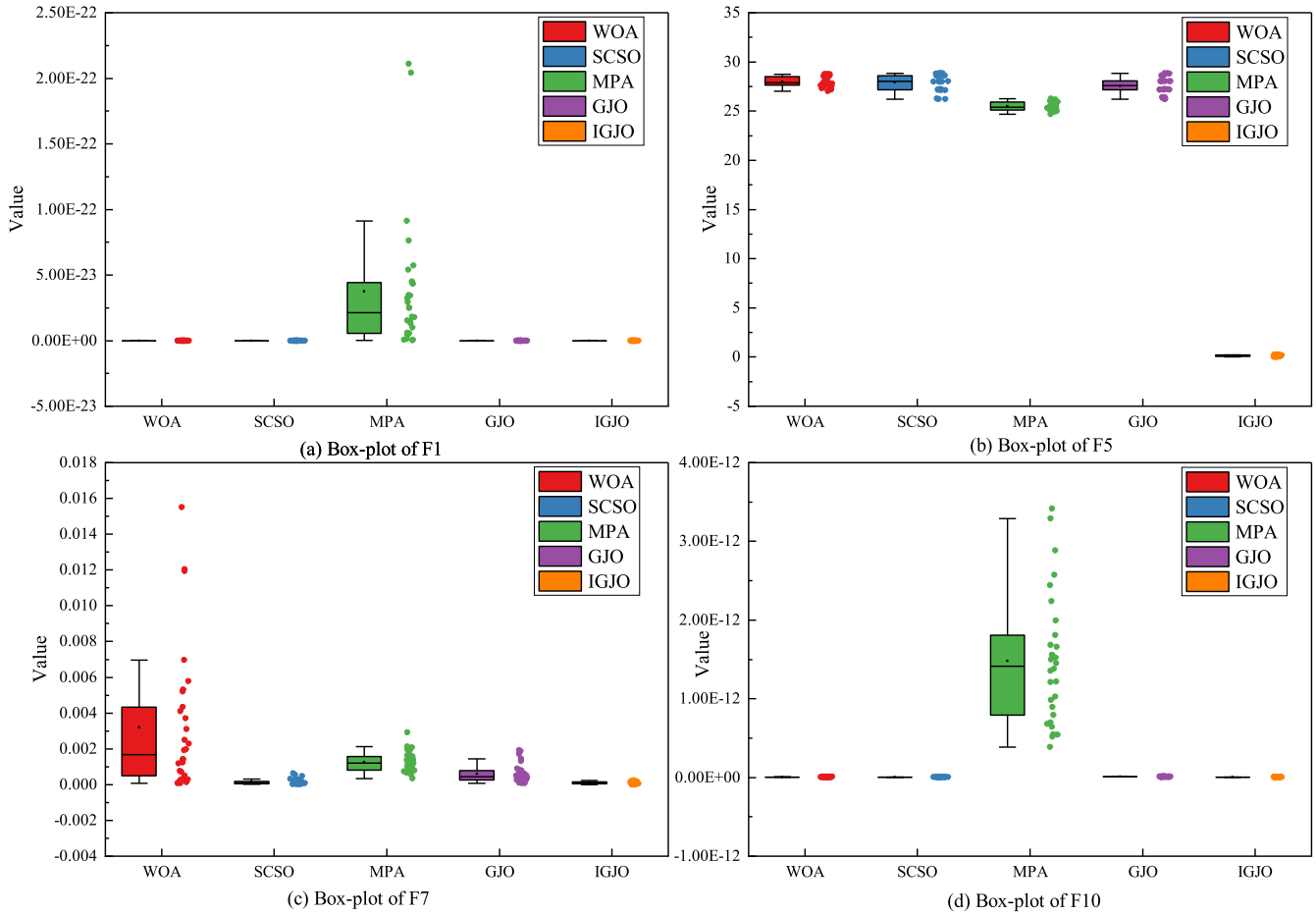


FIGURE 5. Box plot statistics results.

the lth hidden l neuron; $\boldsymbol{e}_{L-1} = f - f_{L-1} = \left[e_{L-1,1}, \cdots e_{L-1,m}\right]$ - Residuals of the L-1st hidden layer node; f -Actual label type.

If the specified error tolerance is not reached, the model will generate new hidden layer nodes under the constraints and satisfy the trend of decreasing deviation as the number of nodes increases, eventually achieving $\lim_{L\to\infty} \| \boldsymbol{f} - \boldsymbol{f}_L \| = 0$. where: $\boldsymbol{f}_L = \sum_{i=1}^{L} \boldsymbol{\beta}_L \boldsymbol{\sigma}_L$.

The basic structure of the SCN is shown in Figure 6.

At this point, the model outputs the following weights based on the updated model:

$$[\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*, \boldsymbol{\beta}_3^*, \cdots \boldsymbol{\beta}_L^*] = \arg\min_{\boldsymbol{\beta}} \| f - \sum_{l=1}^{L} \boldsymbol{\beta}_l \boldsymbol{\sigma}_l \| \tag{11}$$

where: $\boldsymbol{\beta}_l^* = \left[\boldsymbol{\beta}_{l,1}^*, \boldsymbol{\beta}_{l,2}^*, \cdots, \boldsymbol{\beta}_{l,m}^*\right]$. When the number of training samples is small, in order to avoid the phenomenon of overfitting in the SCN, the author introduces the L2 parametric penalty term in the objective function of the model; at the same time, the empirical risk and structural risk are minimized, so as to improve the generalization performance of the network. At this point, the objective function of the

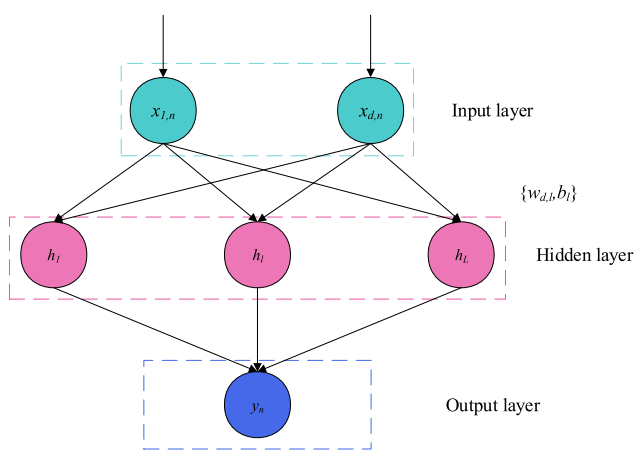


FIGURE 6. Basic structure of a random configuration network.

SCN is improved as follows:

$$[\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*, \boldsymbol{\beta}_3^*, \cdots \boldsymbol{\beta}_L^*] = \arg\min_{\boldsymbol{\beta}} \left(\left\| f - \sum_{l=1}^{L} \boldsymbol{\beta}_l \boldsymbol{\sigma}_l \right\| + C \|\beta\| \right) \tag{12}$$

TABLE 8. Wilcoxon rank sum test results.

function	WOA	SCSO	MPA	SCA	SSA	GJO
F1	1.21E-12	1.21E-12	1.21E-12	1.21E-12	1.21E-12	1.21E-12
F2	6.48E-12	6.48E-12	6.48E-12	6.48E-12	6.48E-12	6.48E-12
F3	1.21E-12	1.21E-12	1.21E-12	1.21E-12	1.21E-12	1.21E-12
F4	1.21E-12	1.21E-12	1.21E-12	1.21E-12	1.21E-12	1.21E-12
F5	3.02E-11	3.02E-11	3.02E-11	3.02E-11	3.02E-11	3.02E-11
F6	3.02E-11	3.02E-11	3.02E-11	3.02E-11	3.02E-11	3.02E-11
F7	4.98E-11	0.71719	3.02E-11	3.02E-11	3.02E-11	1.47E-07
F8	1.21E-12	1.21E-12	1.21E-12	1.21E-12	1.21E-12	1.21E-12
F9	NaN	NaN	NaN	1.21E-12	1.21E-12	NaN
F10	1.26E-08	NaN	1.21E-12	1.21E-12	1.21E-12	1.18E-13
F11	NaN	NaN	NaN	1.21E-12	1.21E-12	NaN
F12	1.31E-08	3.34E-11	3.02E-11	3.02E-11	3.02E-11	3.02E-11
F13	3.02E-11	3.02E-11	9.23E-01	3.02E-11	3.02E-11	3.02E-11

TABLE 9. Sort results.

Algorithm	MAE	Rank
IGJO	2.74E+02	1
GJO	3.10E+02	2
SCSO	5.33E+02	3
MPA	6.96E+02	4
SSA	7.45E+02	5
WOA	4.19E+03	6
SCA	1.07E+04	7

where: C - the penalty term weighting factor of the model.

The output weights are defined according to the least squares method as:

$$\boldsymbol{\beta}^* = (\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}} + \lambda \boldsymbol{I})^{-1}\boldsymbol{G}\boldsymbol{T} \tag{13}$$

where: $\boldsymbol{G} = \sigma\left(\boldsymbol{w}^{\mathrm{T}} \cdot \boldsymbol{x} + \boldsymbol{b}\right)$ - implied layer output.

Where: $\boldsymbol{\beta}^* = \left[\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*, \boldsymbol{\beta}_3^*, \cdots \boldsymbol{\beta}_L^*\right]^{\mathrm{T}}$.

VII. TRANSFORMER FAULT DIAGNOSIS EXAMPLE ANALYSIS

The data in this paper was obtained from a grid in the northeast of the State Grid and the IEC TC 10 database. The data set contains six different operating states of transformers: high energy discharge, low energy discharge, partial discharge, high-temperature overheating, medium and low-temperature overheating, and normal, each operating state contains five fault gas characteristics such as H_2, C_2H_2, C_2H_4, C_2H_6 and CH_4, of which 720 data sets were randomly selected and scaled by 3: 1 ratio for training and testing. The six operating condition types were numbered 1-6: low energy discharge (D1), high energy discharge (D2), partial discharge (PD), medium to low-temperature overheating (T1), high-temperature overheating (T2), and normal condition (NC). The composition of the training and test sample data is shown in Table 10, and some data are given in Table 11.

TABLE 10. Fault sample data composition.

Fault type	Code	Training sets	Test sets	Total sample
D1	1	111	37	148
D2	2	81	27	108
PD	3	96	32	128
T1	5	78	26	104
T2	4	102	34	136
NC	6	72	24	96
Total		540	180	720

TABLE 11. Fault data.

H2	CH4	C2H6	C2H4	C2H2	Type
59	10.4	4	10	12.7	T1
30	7.4	8.5	1.8	19	T1
176	206	47.7	75.7	68.7	T1
3.1	0.5	2.9	4.5	42.8	T2
32.5	28.7	6.6	16.2	108.5	T2
30.6	29.9	6.8	16.1	99.4	T2
242	6	1	4	0	PD
565	93	34	47	0	PD
538.3	12.6	8.7	14.1	0.3	PD
27	90	63	42	0.2	D1
101	169	595	34	0	D1
98	58	43	37	0	D1
73	520	140	1200	6	D2
5.5	35.9	28.1	156	0	D2
12	23	9	98	0	D2
42	33.6	14.8	9	0.6	NC
33.7	3	33.2	27.7	2.5	NC
39.2	24.5	18.4	11.4	6.5	NC

To avoid redundancy and overlap of fault feature information in the original samples, as well as to improve the efficiency and accuracy of the algorithm, and because a single fault feature cannot fully reflect the intrinsic connection with different state types, the five single fault feature data were extended, while the transformer fault features were dimensionally reduced using KPCA to obtain a new fault feature set consisting of six principal components. The new fault feature set was input into the model for training and testing respectively, and the diagnostic accuracy of the different models was analyzed. The simulation experiment platform for this paper is a computer with Windows 10 operating system, 16G of memory, and Matlab 2020b as the running environment.

TABLE 12. Comparison of the average accuracy of differentdiagnostic models.

Fault type	Average accuracy rate /%					
	PNN	RF	SVM	ELM	XGBoost	SCN
D1	78.38	81.08	83.78	81.08	83.78	91.89
D2	74.07	77.78	85.19	85.19	88.89	88.89
PD	78.13	81.25	84.38	84.38	84.38	87.50
T1	80.77	80.77	84.62	80.77	88.46	88.46
T2	79.41	82.35	85.29	79.41	82.35	88.24
NC	75.00	79.17	83.33	87.50	91.67	95.83
Combined diagnosis rate /%	77.78	80.56	84.44	82.78	86.11	90.00

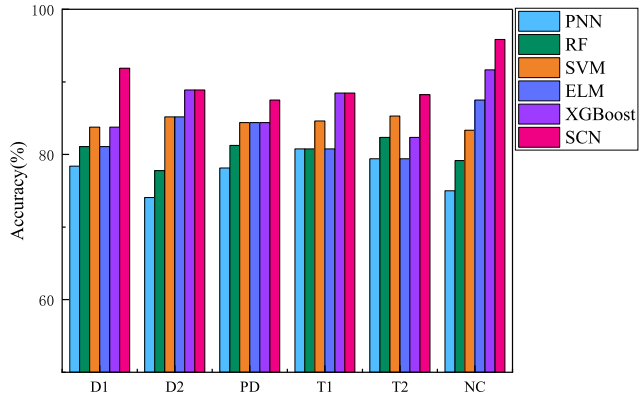


FIGURE 7. Comparison of the accuracy of different diagnostic models.

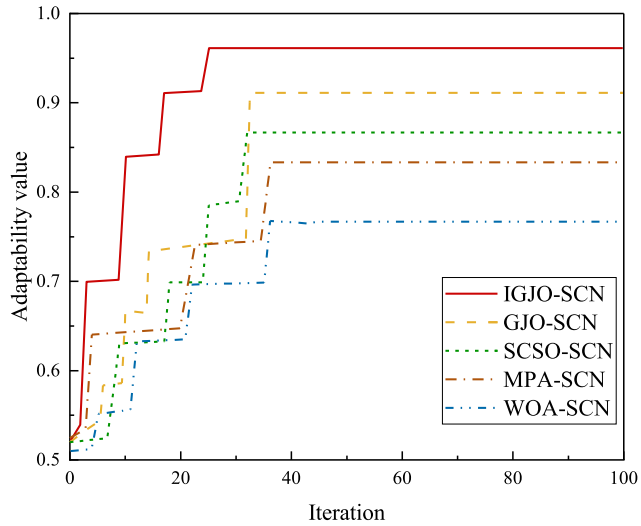


FIGURE 8. Adaptation change curve.

A. COMPARISON OF DIAGNOSTIC PERFORMANCE OF DIFFERENT MODELS WITH SCN

To verify the effectiveness and superiority of SCN, five common machine learning models, including Probabilistic Neural Network (PNN), Random Forest (RF), Support Vector Machine (SVM), Extreme Learning Machine (ELM), and eXtreme Gradient Boosting (XGBoost), were used as comparisons in this paper. The feature datasets obtained from KPCA dimensionality reduction were input into different models for training and classification, and the average accuracy was obtained by running the different models 10 times respectively, and the results are shown in Table 12.

As can be seen from Table 12 and Figure 7, the correct diagnosis rate using the SCN model was significantly higher than that of other models in the low-energy discharge, partial discharge, high-temperature overheating, and normal operation states; the correct diagnosis rate using XGBoost and SCN was the same in the high-energy discharge and medium-to-low-temperature overheating states. The combined diagnostic rate using the SCN model was the highest at 90%, while the accuracy rates using the PNN, RF, SVM, ELM and XGBoost models were 77.78%, 80.56%, 84.44%, 82.78%, and 86.11% respectively, all of which were lower than the SCN model. This indicates that the diagnostic performance of the SCN model is significantly better than other similar models.

B. COMPARISON OF ALGORITHM OPTIMISATION

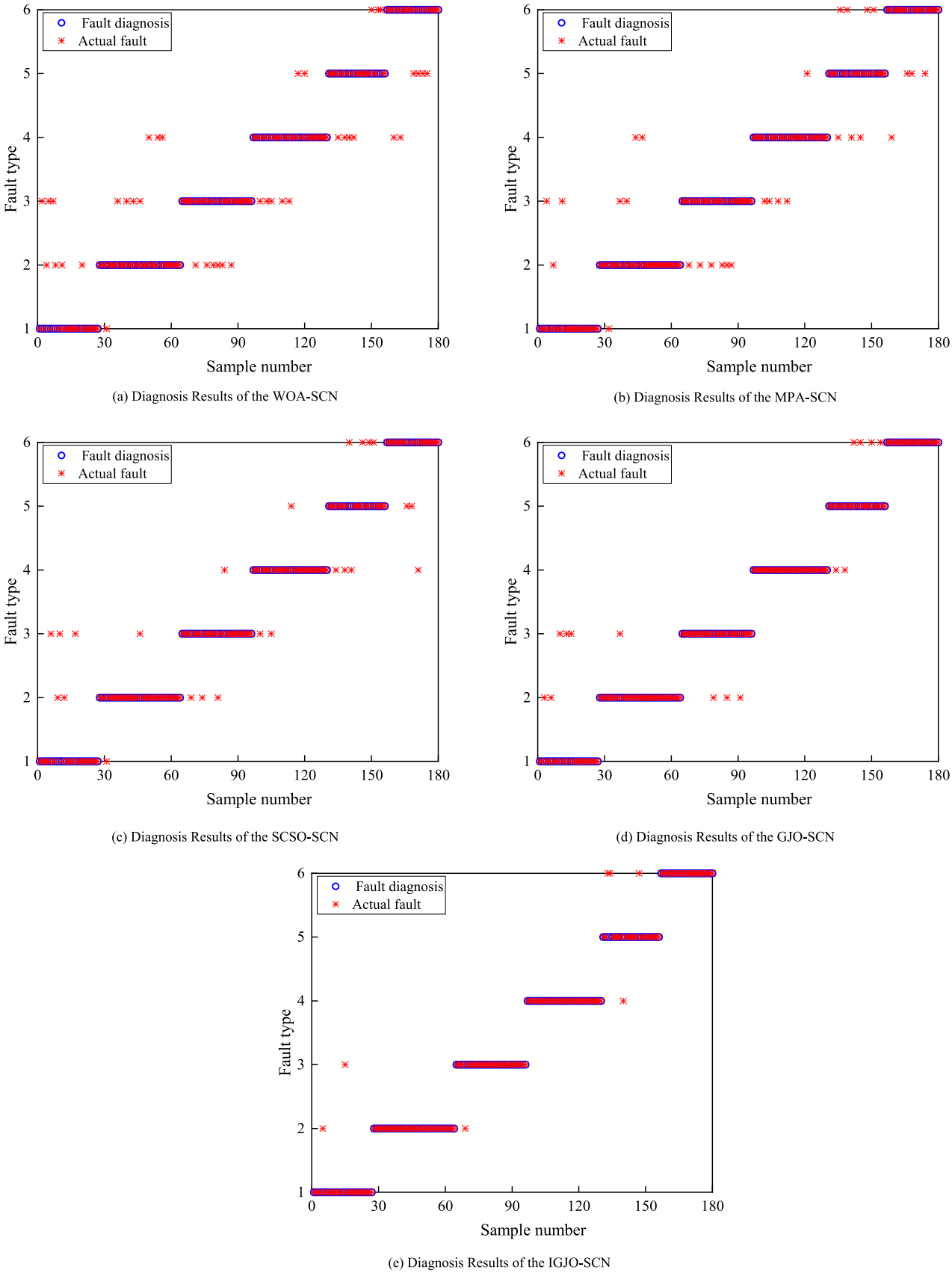
As the penalty term coefficients of the SCN have a significant impact on the training learning effect, the IGJO was used to find the optimization of the two hyperparameters of the LSSVM and to compare the optimization results with those of the WOA, SCSO, MPA, and GJO-SCN. The iterative curve with KPCA reduced data as the model input and diagnostic accuracy as the fitness value is shown in Figure 8.

As can be seen from Figure 8, the WOA algorithm is not effective in optimizing the parameters, resulting in the lowest fitness. The standard GJO algorithm reaches the global optimum after 32 iterations due to its weak global search capability and weak local exploitation capability. SCSO and MPA reach the global optimum in 31 and 36 iterations respectively and have low fitness. The IGJO with the EOBL and golden sine strategies converge in only 25 iterations and has the highest fitness value. This shows that it is feasible to use the IGJO to optimize the SCN-related hyperparameters to build a transformer fault diagnosis model.

C. COMPARISON OF THE DIAGNOSTIC PERFORMANCE OF DIFFERENT HEURISTIC ALGORITHMS FOR OPTIMISING SCN

To verify the superiority of the IGJO superior SCN diagnostic model, the WOA-SCN, SCSO-SCN, MPA-SCN, GJO-SCN, and IGJO-SCN models were compared and analyzed. The average accuracies of the different diagnostic methods are shown in Table 13.

As can be seen from Figure 9 and Table 13, the IGJO-SCN has the highest fault accuracy rate, where the diagnostic accuracy rates for low energy discharge, high-temperature overheating, and normal operating conditions are 100%, and the diagnostic accuracy rates for high energy discharge,



(a) Diagnosis Results of the WOA-SCN

(b) Diagnosis Results of the MPA-SCN

(c) Diagnosis Results of the SCSO-SCN

(d) Diagnosis Results of the GJO-SCN

(e) Diagnosis Results of the IGJO-SCN

FIGURE 9. **Diagnosis results.**

TABLE 13. Average accuracy of different diagnostic methods for different faults/%.

Fault type	Average accuracy rate /%				
	WOA-SCN	MPA-SCN	SCSO-SCN	GJO-SCN	IGJO-SCN
D1	78.38	86.49	94.59	97.30	100
D2	74.07	88.89	81.48	77.78	92.59
PD	81.25	81.25	87.50	90.63	96.88
T1	69.23	73.08	73.08	76.92	84.62
T2	79.41	85.29	91.18	100	100
NC	75.00	83.33	87.50	100	100
Combined diagnosis rate /%	76.67	83.33	86.67	91.11	96.11

partial discharge, and medium and low-temperature overheating are 92.59%, 96.88%, and 84.62% respectively, and the correct diagnostic rates for all six fault types are above 84%, for a total The fault diagnosis accuracy rate reached 96.11%. In addition, the fault diagnosis accuracy of GJO-SCN, SCSO-SCN, MPA-SCN, and WOA-SCN was 91.11%, 86.67%, 83.33%, and 76.67% respectively, and the fault diagnosis accuracy of IGJO-SCN was improved by 5%, 9.44%, 12.78%, and 19.44% respectively compared with the above four methods. The results indicate that the SCN fault diagnosis model optimized by IGJO has higher reliability after KPCA dimensionality reduction.

VIII. CONCLUSION

1) This paper introduces the L2 parametric penalty term in the random configuration network, which improves the generalization ability of SCN in practical applications.

2) Using KPCA to reduce the dimensionality of 15 features, the resulting six features can express valid information about transformer faults and avoid redundant information, while facilitating the learning and training of SCN models. The running time is improved by nearly 20s and the correct rate is improved by nearly 15% compared to that before dimensionality reduction. The complexity and running time of the model are reduced and the accuracy of the model is improved.

3) The IGJO algorithm overcomes the drawback that the GJO algorithm is prone to fall into local optimality, and has better stability and superiority-seeking ability compared with the WOA, SCSO, MPA, SCA and SSA algorithms.

4) The simulation results show that the parameter optimization of SCN using the IGJO algorithm can effectively improve the fault diagnosis accuracy of the transformer by 96.11%. the fault diagnosis accuracy of GJO-SCN, SCSO-SCN, MPA-SCN, and WOA-SCN is 91.11%, 86.67%, 83.33%, and 76.67% respectively. The fault diagnosis accuracy of IGJO-SCN improved by 5%, 9.44%, 12.78%, and 19.44% respectively compared to the above four methods. The IGJO-SCN model proposed in this paper verifies that its diagnostic performance is excellent, but there is still room for improvement in the selection of data features. Due to the small influence of gases such as CO and CO2, only five of the characteristic gases are selected as the original data in this paper, and the influence of these gases needs to be considered when more detailed fault diagnosis classification is carried out later, and more combinations of ratios can also be used to enrich the characteristic In addition, since different combination models have different diagnostic rates for different fault types, it is possible to try to compare the fault diagnostic rates of multiple models and select the optimal model to diagnose the corresponding faults.

REFERENCES

[1] C. Zhang, Y. He, B. Du, L. Yuan, B. Li, and S. Jiang, "Transformer fault diagnosis method using IoT based monitoring system and ensemble machine learning," *Future Gener. Comput. Syst.*, vol. 108, pp. 533–545, Jul. 2020.

[2] D. Jiejie, S. Hui, and S. Gehao, "Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 5, pp. 2828–2835, Oct. 2017.

[3] W. Lu, C. Shi, H. Fu, and Y. Xu, "Research on transformer fault diagnosis based on ISOMAP and IChOA-LSSVM," *IET Electr. Power Appl.*, vol. 2023, pp. 1–15, Oct. 2023.

[4] Y. Zhang, H. Wei, R. Liao, Y. Wang, L. Yang, and C. Yan, "A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed transformers," *J. Elect. Eng. Technol.*, vol. 12, no. 2, pp. 830–839, Mar. 2017.

[5] S. I. Ibrahim, S. S. M. Ghoneim, and I. B. M. Taha, "DGALab: An extensible software implementation for DGA," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 18, pp. 4117–4124, Oct. 2018.

[6] H. Malik and S. Mishra, "Application of gene expression programming (GEP) in power transformers fault diagnosis using DGA," *IEEE Trans. Ind. Appl.*, vol. 52, no. 6, pp. 4556–4565, Nov. 2016.

[7] *IEEE Approved Draft Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*, document IEEE PC57.104/D6.2, Mar. 2019.

[8] C. Jin-Qiang, "Fault prediction of a transformer bushing based on entropy weight TOPSIS and gray theory," *Comput. Sci. Eng.*, vol. 21, no. 6, pp. 55–62, Nov. 2019.

[9] C. Guo, B. Wang, Z. Wu, M. Ren, Y. He, R. Albarracín, and M. Dong, "Transformer failure diagnosis using fuzzy association rule mining combined with case-based reasoning," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 11, pp. 2202–2208, Jun. 2020.

[10] P. Lu, W. Li, and D. Huang, "Transformer fault diagnosis method based on graph theory and rough set," *J. Intell. Fuzzy Syst.*, vol. 35, no. 1, pp. 223–230, Jul. 2018.

[11] J. I. Aizpurua, V. M. Catterson, B. G. Stewart, S. D. J. McArthur, B. Lambert, B. Ampofo, G. Pereira, and J. G. Cross, "Power transformer dissolved gas analysis through Bayesian networks and hypothesis testing," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 2, pp. 494–506, Apr. 2018.

[12] A. Zou, R. Deng, Q. Mei, and L. Zou, "Fault diagnosis of a transformer based on polynomial neural networks," *Cluster Comput.*, vol. 22, no. S4, pp. 9941–9949, Jul. 2019.

[13] M. Žarković and Z. Stojković, "Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics," *Electr. Power Syst. Res.*, vol. 149, pp. 125–136, Aug. 2017.

[14] J. Z. Li, Q. G. Zhang, K. Wang, and J. Y. Wang, "Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 2, pp. 1198–1206, Apr. 2016.

[15] Y. Zhou, X. Yang, L. Tao, and L. Yang, "Transformer fault diagnosis model based on improved gray wolf optimizer and probabilistic neural network," *Energies*, vol. 14, no. 11, p. 3029, May 2021.

[16] S. Kim, S.-H. Jo, W. Kim, J. Park, J. Jeong, Y. Han, D. Kim, and B. D. Youn, "A semi-supervised autoencoder with an auxiliary task (SAAT) for power transformer fault diagnosis using dissolved gas analysis," *IEEE Access*, vol. 8, pp. 178295–178310, 2020.

[17] M. Tahir and S. Tenbohlen, "Transformer winding condition assessment using feedforward artificial neural network and frequency response measurements," *Energies*, vol. 14, no. 11, p. 3227, May 2021.

[18] H. A. Illias and W. Z. Liang, "Identification of transformer fault based on dissolved gas analysis using hybrid support vector machine-modified evolutionary particle swarm optimisation," *PLoS ONE*, vol. 13, no. 1, Jan. 2018, Art. no. e0191366.

[19] X. Huang, Y. Zhang, J. Liu, H. Zheng, and K. Wang, "A novel fault diagnosis system on polymer insulation of power transformers based on 3-stage GA–SA–SVM OFC selection and ABC–SVM classifier," *Polymers*, vol. 10, no. 10, p. 1096, Oct. 2018.

[20] X. Huang, X. Huang, B. Wang, and Z. Xie, "Fault diagnosis of transformer based on modified grey wolf optimization algorithm and support vector machine," *IEEJ Trans. Electr. Electron. Eng.*, vol. 15, no. 3, pp. 409–417, Mar. 2020.

[21] N. Chopra and M. M. Ansari, "Golden jackal optimization: A novel nature-inspired optimizer for engineering applications," *Expert Syst. Appl.*, vol. 198, Jul. 2022, Art. no. 116924.

[22] M. Khishe, "Greedy opposition-based learning for chimp optimization algorithm," *Artif. Intell. Rev.*, vol. 2022, pp. 1–31, Dec. 2022.

[23] B. S. Yildiz, N. Pholdee, S. Bureerat, A. R. Yildiz, and S. M. Sait, "Enhanced grasshopper optimization algorithm using elite opposition-based learning for solving real-world engineering problems," *Eng. Comput.*, vol. 38, no. 5, pp. 4207–4219, Oct. 2022.

[24] Z. Xiao and S. Liu, "Study on elite opposition-based golden-sine whale optimization algorithm and its application of project optimization," *Electron. Lett.*, vol. 47, no. 10, pp. 2177–2186, 2019.

[25] Y. Yuan, X. Mu, X. Shao, J. Ren, Y. Zhao, and Z. Wang, "Optimization of an auto drum fashioned brake using the elite opposition-based learning and chaotic k-best gravitational search strategy based grey wolf optimizer algorithm," *Appl. Soft Comput.*, vol. 123, Jul. 2022, Art. no. 108947.

[26] R. Sihwail, K. Omar, K. Ariffin, and M. Tubishat, "Improved Harris hawks optimization using elite opposition-based learning and novel search mechanism for feature selection," *IEEE Access*, vol. 8, pp. 121127–121145, 2020.

[27] E. Tanyildizi and G. Demir, "Golden sine algorithm: A novel math-inspired algorithm," *Adv. Elect. Comput. Eng.*, vol. 17, no. 2, pp. 71–78, 2017.

[28] G. Kaur and S. Arora, "Chaotic whale optimization algorithm," *J. Comput. Des. Eng.*, vol. 5, no. 3, pp. 275–284, Apr. 2018.

[29] A. Seyyedabbasi and F. Kiani, "Sand Cat swarm optimization: A nature-inspired algorithm to solve global optimization problems," *Eng. Comput.*, vol., 2022, pp. 1–25, Apr. 2022.

[30] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, "Marine predators algorithm: A nature-inspired Metaheuristic," *Expert Syst. Appl.*, vol. 152, Aug. 2020, Art. no. 113377.

[31] L. Abualigah and A. Diabat, "Advances in sine cosine algorithm: A comprehensive survey," *Artif. Intell. Rev.*, vol. 54, pp. 2567–2608, Jan. 2021.

[32] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp swarm algorithm: A bio-inspired optimizer for engineering design problems," *Adv. Eng. Softw.*, vol. 114, pp. 163–191, Dec. 2017.

[33] E. Nabil, "A modified flower pollination algorithm for global optimization," *Expert Syst. Appl.*, vol. 57, pp. 192–203, Sep. 2016.

[34] H. Wu, A. Zhang, Y. Han, J. Nan, and K. Li, "Fast stochastic configuration network based on an improved sparrow search algorithm for fire flame recognition," *Knowl.-Based Syst.*, vol. 245, Jun. 2022, Art. no. 108626.

WANJIE LU was born in Songyuan, Jilin, China, in 1979. She received the B.Sc. degree in electrical technology from Liaoning Technical University, in 2004, the M.Sc. degree in mechanical and electronic engineering, in 2007, and Ph.D. degree in mechanical design, in 2019, respectively.

She is an Associate Professor with Liaoning Technical University. She is the author of nine teaching books, 11 articles, and two inventions. Her current research interests include signal detection and processing, the intelligent monitoring of equipment, fault diagnosis, and the prediction of equipment.

CHUN SHI was born in Liaocheng, Shandong, China, in 1997. He is currently pursuing the master's degree with the Liaoning University of Engineering and Technology. His current research interests include intelligent detection, data fusion technology, and deep learning.

HUA FU was born in Fuxin, Liaoning, China, in 1962. She received the B.S. degree in electrical automation, the M.S. degree in power electronics and power transmission, and the Ph.D. degree in mining engineering from Liaoning Technical University, Huludao, China, in 1984, 1996, and 2002, respectively. She was a Senior Visiting Scholar with the University of Michigan, USA, in 2008, and with the University of Technology Sydney, Australia, in 2016. She has been a Professor with Liaoning Technical University, Liaoning, China, since 1984. She is the author of six monographs, 21 teaching books, 120 articles, and 34 inventions. Her current research interests include coal mine gas monitoring, safety monitoring and surveillance, and intelligent detection and control.

YAOSONG XU was born in Cangzhou, Hebei, China, in 1979. He received the B.S. degree in detection technology and instrumentation and the M.S. degree in power electronics and power transmission from Liaoning Technical University, Huludao, China, in 2001 and 2004, respectively, and the Ph.D. degree in precision instruments and machinery from Tianjin University, Tianjin, China, in 2012. He is a Teacher with the Department of Measurement and Control Technology and Instrumentation, Liaoning Technical University. He is the author of three books, more than 30 articles, and ten inventions. His current research interests include power quality monitoring, grid load prediction, and electrical equipment fault diagnosis.