Novel Probabilistic Neural Network Models Combined with Dissolved Gas Analysis for Fault Diagnosis of Oil-Immersed Power Transformers

Yichen Zhou, Lingyu Tao, Xiaohui Yang,* and Li Yang

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Fault diagnosis technology of power transformers is essential for the stable operation of power systems. Fault diagnosis technology based on dissolved gas analysis (DGA) is one of the most commonly used methods. However, due to the lack of fault information, traditional DGA fault diagnosis techniques are difficult to meet increasing power demand in terms of accuracy and efficiency. To address this problem, this paper proposes a novel fault diagnosis model for oil-immersed transformers based on International Electrotechnical Commission (IEC) ratio methods and probabilistic neural network (PNN) optimized with the modified moth flame optimization algorithm (MMFO). PNN as a radial neural network has good utility and is often used in classification models, but its classification performance is easily affected by the smoothing factor (σ) of the hidden layer and is not stable. This paper addresses this issue using the MMFO to optimize the smoothing factor, which effectively improves the classification accuracy and robustness of PNN. The proposed method was validated by conducting the experiments with the real data collected from transformers. Experimental results show that the MMFO-PNN model improves the fault diagnosis accuracy rate from 70.65 to 99.04%, which is higher than other power transformer fault diagnosis models.
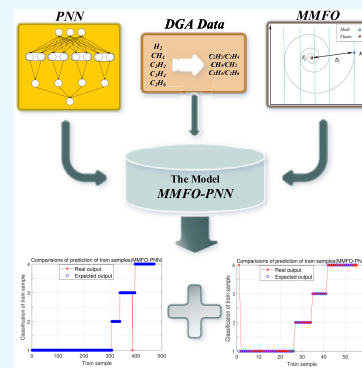
INTRODUCTION

An oil-immersed power transformer is one of the most important high-voltage devices in a transmission and transformer system, and its operational status determines whether the power grid can be reliably supplied.[1] Failure to detect power transformer faults in a timely and accurate manner will have a serious negative impact on grid paralysis and damage the normal development of social economy.[2] Therefore, the study of power transformer fault diagnosis is of great importance for the security and reliability of the power network.[3]

Over the years, dissolved gas analysis (DGA) has become the most popular method to identify the initial fault of a transformer.[4] Normal operation of oil-immersed transformers due to insulation aging cracking and other reasons will produce a very small amount of gas: hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2). When a power transformer fails or has a potential fault, the content of various gases dissolved in the oil can change significantly. Therefore, the composition of dissolved gases in transformer oil can reflect the operating condition of power transformers to a large extent.[5] At present, traditional power transformer fault diagnosis methods such as the three-ratio method, Rogers method, and Dornerburg method recommended by International Electrotechnical Commission (IEC) have been developed based on DGA data.[6] However, the composition of gas components generated by oil-immersed power transformer faults is complex and the distribution characteristics are difficult to speculate, so it is difficult to map a precise relationship between the gas content or content ratio in oil and transformer fault type.[7]

At the same time, with the continuous development of artificial intelligence technology, the advantages of artificial intelligence technology in the calculation, classification, and prediction have been found by many researchers in other fields and widely used in their own research work. For example, multilayer perceptron artificial neural network (MLP-ANN) is used to predict the solubility of hydrogen sulfide at different temperatures, pressures, and concentrations, which shows good prediction performance;[8] the least-squares support vector machine (LS-SVM) model was established by Ahmadi and Ahmadi[9] to predict the solubility of CO_2 brine. The average relative absolute deviation between the model prediction and the experimental data is 0.1%; in the field of liquid production of condensate gas reservoir, a hybrid model of particle swarm optimization and artificial neural network[10] is established.
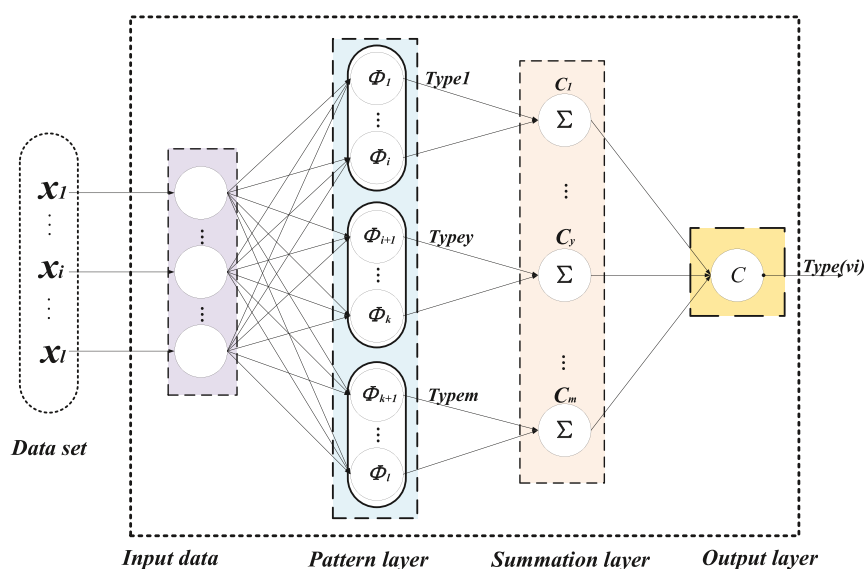
18084

Figure 1. Topology of a probabilistic neural network.

Compared with the traditional scheme, the intelligent model has superior performance in determining the dew point pressure of the condensate gas reservoir.

These research studies show that the artificial intelligence method has achieved good results in different fields. To solve the problem of low accuracy and efficiency of traditional DGA fault diagnosis technology and meet the needs of the increasingly complex power grid system, it is particularly important to adopt higher performance intelligent fault diagnosis technology in today's growing power demand. Combined with artificial intelligence techniques, establishing a transformer fault classification model based on DGA data set is the basis of intelligent fault diagnosis in transformers. Su et al.[11] established a fault diagnosis model based on the fuzzy logic technique, which can diagnose multiple faults in transformers and can quantitatively represent the likelihood of each fault. Dhini et al.[12] established a support vector machine (SVM)-based oil immersion transformer fault diagnosis model with a higher recognition rate compared to traditional methods. Jiang et al.[13] and Yu et al.[14] established transformer fault diagnosis models based on the hidden Markov model (HMM) and k-nearest neighbor (KNN) algorithms, respectively, with faster diagnostic efficiency and showed superiority in dynamic fault prediction as well as real-time monitoring of operating conditions.

Although these methods can improve the accuracy of fault diagnosis to a certain extent, there are still some shortcomings. For example, Su[11] applied the fuzzy logic technique to the field of fault diagnosis, which has the advantages of convenience and intuitive effect. However, the disadvantage is that in the face of large data sets, the classification effect will decline sharply; the SVM-based fault diagnosis model[12] is limited by the limitations of SVM itself in the multiclassification problem, and it does not perform well in the face of complex high-dimensional data; although the KNN fault diagnosis model established by Yu[14] improves the work efficiency of KNN algorithm, it does not solve the shortcomings of low fault tolerance rate of KNN to training data and is easy to fall into the dimension disaster, which leads to the weak generalization of the model; the fault diagnosis method based on the hidden Markov model (HMM) proposed by Jiang[13] is difficult to guarantee the accuracy and stability in the medium- and long-term prediction of fault data. Compared

with other artificial intelligence methods, artificial neural network (ANN) can significantly improve the accuracy of fault diagnosis.[15] ANN-based fault diagnosis models for power transformers are trained by sampling data of power transformers under different operating conditions, and the connection weights and biases (significant parameters) of the network model are continuously adjusted during the training process to finally establish the corresponding mapping relationship between specific fault features and fault types.[16] The researchers are integrating neural network-based, deep learning methods with transformer fault diagnosis techniques. Huang et al.[17] proposed an evolutionary neural network approach for power transformer fault diagnosis. Based on the proposed evolutionary algorithm, the neural network automatically adjusts the network parameters (connection weights and deviation terms) to obtain the best model. Meng and Dong et al.[18] proposed a radial basis function neural network (RBFNN) based on a hybrid adaptive training method for fault diagnosis of power transformers. The method is able to generate RBFNN models based on fuzzy c-means (FCM) and quantum-inspired particle swarm optimization (QPSO), which allows automatic configuration of the network structure and 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. Dai et al.[19] 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. 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. Huang et al.[20] proposed a transformer fault diagnosis method based on the gray wolf optimization (GWO) algorithm to optimize the hybrid kernel extreme learning machine (KELM); using the GWO algorithm can optimize the parameters of the hybrid kernel function, while using logistic chaos mapping can generate the initial population parameters of the GWO algorithm to avoid the adverse effects of
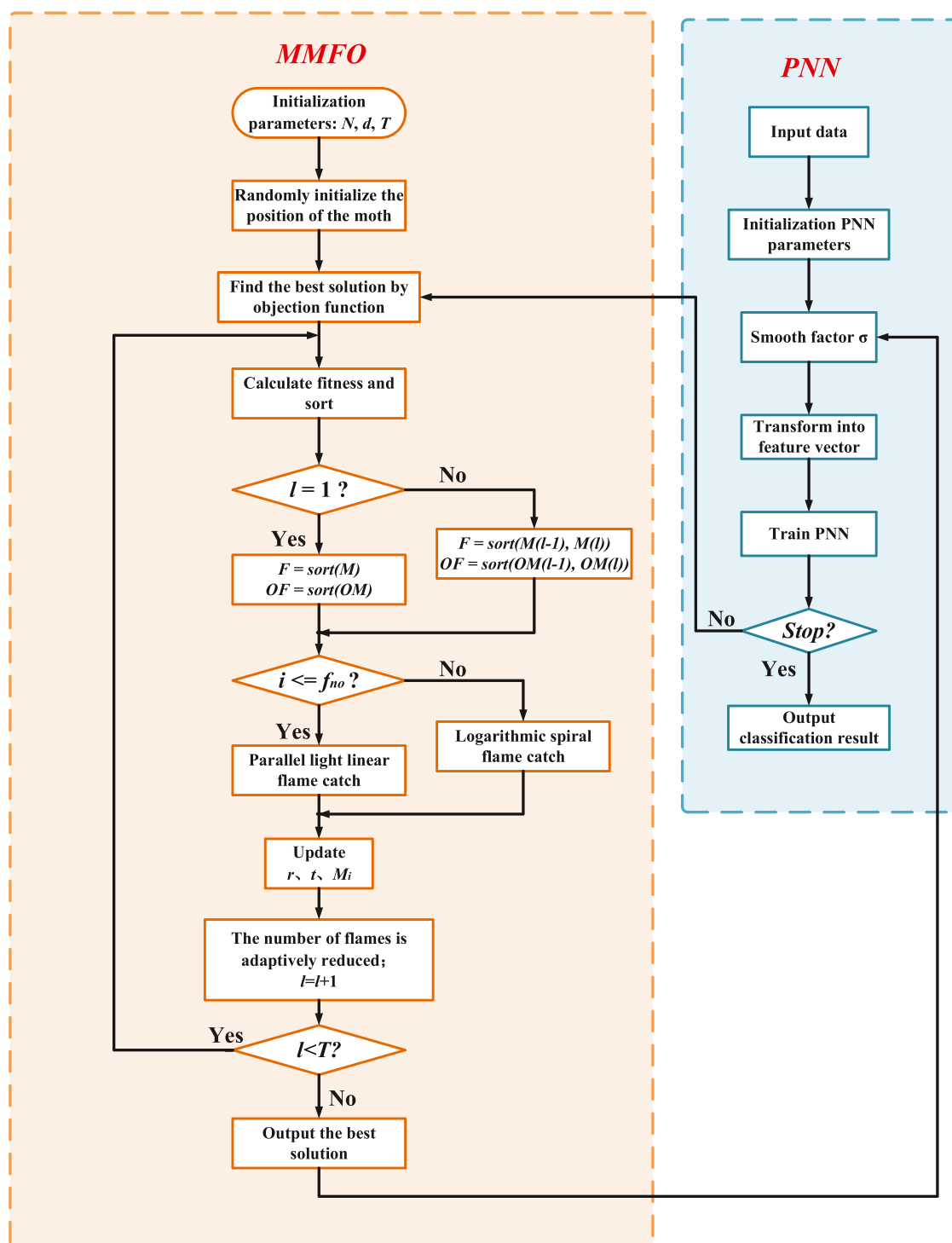
Figure 2. Diagram of the proposed MMFO-based PNN for fault diagnostics.

too fast convergence on the optimization results and effectively improve the classification accuracy.

Although the evolutionary neural network model proposed by Huang[17] can automatically adjust the network parameters, the convergence ability of the evolutionary algorithm is insufficient, and it is easy to fall into the local optimum, which limits the accuracy of the classification model; quantum-inspired particle swarm optimization (QPSO) proposed by Meng[18] can solve the problem of slow convergence of PSO. However, when the data sample is large, the RBFNN has the disadvantage of complex

structure and huge computation; the classification accuracy of fault diagnosis model based on DBN is very high,[19] but it needs a lot of fault data for network training, and the classification performance is not stable in the case of a small amount of data; the method proposed in the literature[20] is very effective for KELM optimization, but the efficiency and accuracy need to be improved.

This paper chooses the probabilistic neural network (PNN) with certain advantages in fault classification as the basic classifier of the fault diagnosis model. Since the classification

performance of PNN is susceptible to network parameters (such as smoothing factor and connection weight), the classification performance of PNN can be improved by optimizing network parameters.[21,22] We optimized the smoothing factor of PNN by improved moth flame optimization (MMFO) and established an MMFO-PNN fault diagnosis model. Compared with other fault diagnosis methods, MMFO-PNN has higher classification accuracy and efficiency in DGA data classification and has strong engineering practicability.

This paper is organized as follows: after the Introduction, Proposed Method describes the proposed machine learning theoretical approach, Implementation and Experimental Setup describes the establishment of a simulation model for transformer fault diagnosis, Experimental Results and Discussion presents and discusses the experimental results, and the final section draws conclusions.

■ PROPOSED METHOD

In this section, we present the proposed method for transformer fault diagnosis. We first discuss IEC ratio methods and the improved moth flame optimization algorithm and then detail the MMFO-based PNN model for fault diagnostics.

IEC Ratio Methods. This work focuses on the IEC three-ratio method. The IEC three-ratio method uses five gases: H_2, CH_4, C_2H_2, C_2H_4, and C_2H_6. These gases are produced in three gas ratios: C_2H_2/C_2H_4, CH_4/H_2, and C_2H_4/C_2H_6. It provides the values of the three key gas ratios corresponding to the suggested fault diagnosis. When the critical gas ratio exceeds a given threshold, the power transformer may fail early. Therefore, these given values can be used to detect early failures of power transformers.

Modified Moth Flame Optimization Algorithm. Moth flame optimization algorithm is a natural heuristic optimization algorithm, which was proposed by Mirjalili[23] in 2015. There are two important parts in MFO: moth and flame. The position of one moth corresponds to a solution to the problem. Flame stores all of the optimal solutions of moth population so far. Compared with other metaheuristics, MFO has the advantages of simple structure, good robustness, and easy implementation.[24]

The MFO optimization process can be summarized into three phases:

(a) Randomly generated moth positions in the search space: The initialization of the position of each moth in the moth population is implemented in MFO as follows

$$M_i = (ub_i - lb_i) \cdot \text{rand}() + lb_i \quad (1)$$

where the two matrices, ub and lb, define the upper and lower bounds of the variables, respectively.

(b) Adaptive reduction of the number of flames: The equation for the adaptive reduction of the number of flames is as follows

$$f_{no} = \text{round}\left(N - l\frac{N-1}{T}\right) \quad (2)$$

where f_{no} represents the number of flames, N represents the number of moth populations, l is the number of current iterations, and T is the maximum number of iterations.

(c) Position update: The standard MFO chooses a logarithmic spiral update mechanism, which updates the flame list based on the best position at each iteration, enhancing the ability of spatial search. The logarithmic spiral update mechanism is formulated as follows

$$M_i(l+1) = \begin{cases} D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_i(l), & i \leq f_{no} \\ D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_{f_{no}}(l), & i > f_{no} \end{cases} \quad (3)$$

where b is the constant for constructing the logarithmic helix trajectory; after arranging the moth positions in ascending order according to their fitness values in the l-1th generation, the first f_{no} moth positions are taken as the list of lth generation flame positions, that is, $F(l)$, and $F_i(l)$ is the ith flame in the list; then, $F_{f_{no}}(l)$ denotes the flame position with the worst fitness. D_i is calculated as follows

$$D_i = \begin{cases} |F_i - M_i|, & i \leq f_{no} \\ |F_{f_{no}} - M_i|, & i > f_{no} \end{cases} \quad (4)$$

where t denotes the proximity of the moth to the location of the flame, which is determined by the following equation

$$\begin{cases} t = (a-1) \cdot \text{rand}() + 1 \\ a = -1 + l \cdot \dfrac{-1}{T} \end{cases} \quad (5)$$

Through the description of the above three stages, we summarize the whole optimization process of MFO as follows:

Step 1: Setting the algorithm parameters: population size N, dimension d, maximum number of iterations T.

Step 2: Initializing randomly the moth's position in the search space by eq 1 and recording.

Step 3: Calculating the fitness of each moth position, ranking the moth positions according to the fitness, and recording the better solution among them as the next generation of flame positions (the first generation of flame positions is generated from the first generation of moth positions).

Step 4: Performing adaptive reduction of the number of flames by eq 2.

Step 5: Updating D_i and t by eqs 4 and 5 and finally the individual moth positions by eq 3.

Step 6: Checking if l is greater than T. If not, return to Step 3; if yes, stop iteration and output the result.

In the traditional moth flame algorithm, the moth approximates along a logarithmic helix trajectory for any flame, which can enhance the local search capability of the algorithm, but it is also easy to fall into the local optimum. To speed up the convergence speed and avoid the model falling into the local optimum, we introduce the chaotic operator after initialization and use the parallel light straight-line trajectory instead of the traditional logarithmic spiral trajectory to update the position. The pseudocode of MMFO is shown in Algorithm 1.

Algorithm 1 The modified MFO

1: **Initialization parameters:** population size N, dimension d, maximum number of iterations T;
2: Randomly initialize the position of the moth in the search space by Equation (1);
3: **while** $l <= T$ **do**
4:     Update number of flames by Equation (2)
5:     $OM = FitnessFunction(M)$
6:     **if** $l == 1$ **then**
7:         $F = sort(M); OF = sort(OM);$
8:     **else**
9:         $F = sort(M(l-1), M(l));$
10:        $OF = sort(OM(l-1), OM(l));$
11:    **end if**
12:    Record the first flame as the optimal individual;
13:    **for** $i = 1 : N$ **do**
14:        **if** $i <= f_{no}$ **then**
15:            Update c_n by Equation (7)
16:            Update M_i by Equation (6)
17:        **else**
18:            Update t by Equation (5)
19:            Calculate D_i by Equation (8)
20:            Update M_i by Equation (8)
21:        **end if**
22:    **end for**
23: **end while**
Output: Best classification and predication results

Because moth individuals search for flame individuals relatively independently without too much information, we select the top three relatively good individuals after adaptation ranking to perform a chaotic linear combination with each individual in the moth population and average them out to obtain the updated positions

$$M_i \atop i=1,2,3,...,N = \frac{1}{3} \sum_{x=1}^{3} (M_x - xc_n \cdot |2c_n M_x - M_i|) \tag{6}$$

where c_n is the chaotic mapping optimization operator, which is ergodic and stochastic and often achieves better results using the chaos mapping optimization operator instead of random numbers in the algorithm.[25] The introduction of the chaos operator in the position update process of MFO can effectively reduce the probability of falling into the local optimum and thus improve the global search capability, which is calculated as follows

$$c_n = 4c_{n-1}^3 - 3c_{n-1} \tag{7}$$

where M_x represents the best individual and M_i represents each individual in the moth population.

As the number of flames adaptively decreases, when the number of flames is smaller than the number of moths, the traditional logarithmic helix is used for position updating

$$\begin{cases} M_i \atop i=1,2,3,...,N = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_{f_{no}} \\ D_i = |F_{f_{no}} - M_i| \end{cases} \tag{8}$$

the whole update process can be expressed as (the meaning of the parameters in eq 9 is the same as described in the previous section)

$$M_i(l+1) = \begin{cases} \frac{1}{3} \sum_{x=1}^{3} (M_x(l) - xt_n \cdot |2t_n M_x(l) & i \leq f_{no} \\ \quad - M_i(l)|), \\ \frac{1}{3} \sum_{x=1}^{3} (M_x(l) - xt_n \cdot |2t_n M_x(l) & i > f_{no} \\ \quad - M_i(l)|), \end{cases} \tag{9}$$

PNN Optimized by MMFO. Probabilistic neural network (PNN) is a kind of neural network with a simple structure and

wide application.[26] It is also a radial basis function feedforward neural network based on the Bayesian decision theory.

As shown in Figure 1, the structure of PNN is a parallel four-layer structure: input layer, mode layer, summation layer, and output layer. The input layer receives values from the training samples, converts them into feature vectors, and then assigns them to the network. The number of neurons in the input layer is equal to the number of sample vector dimensions. The pattern layer is calculated using the Euclidean distance between the training sample feature vectors and the center of the free base to match the input feature vectors to various types of relationships in the training set. The output of each pattern unit is

$$\Phi_{ij}(x) = \frac{1}{(2\pi)^{1/2} \sigma^d} e^{-(X-x_{ij})^T(X-x_{ij})/\sigma^2} \tag{10}$$

where $X = [x_1, x_2, ..., x_n]^T$, $n = 1, 2, ..., l$. l denotes the total number of training patterns. d denotes the feature vector dimension. X_{ij} is the jth center of the ith class of training samples, and σ is the smoothing factor. The summation layer performs a weighted average of the outputs of the neurons in the pattern layers belonging to the same type. The formula is as follows

$$v_i = \frac{\sum_{j=1}^{L} \Phi_{ij}}{L} \tag{11}$$

where v_i is the output type of class i and L is the number of neurons of class i.

The type corresponding to the maximum output in the output layer is the output type. The result is as follows

$$type(v_i) = \arg\max(v_i) \tag{12}$$

Due to the limitation of PNN itself, the smoothing factor σ has a great influence on the computational process of the input layer output to the hidden layer. If the value of σ is not chosen appropriately, too large or too small values can make the network converge too fast or easily fall into local optimum, allowing the classification accuracy to drop dramatically.[27] The MMFO described above is advantageous in global search capability and robustness compared with other traditional optimization algorithms. The classification performance of PNN can be greatly improved by MMFO optimization σ.

In the proposed MMFO-PNN model, the input feature vector can be expressed as

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{l1} & x_{l2} & \cdots & x_{ln} \end{bmatrix} = [x_1, x_2, ..., x_n] \tag{13}$$

The steps for classifying the input feature vectors by the MMFO-PNN model are represented in Figure 2, which are described as follows:

Step 1: Set the initial training sample X.
Step 2: Initialize the probabilistic neural network by randomly defining the set of smoothing factors as follows

$$\sigma = (\sigma_1, \sigma_2, \sigma_{3}, ..., \sigma_n)^T \tag{14}$$

Step 3: Initialize the MMFO algorithm parameters: population size N, dimensionality d, maximum number of iterations T, and initialize the fitness function $f(x)$. It is worth noting that in our model, the mean square error (MSE)
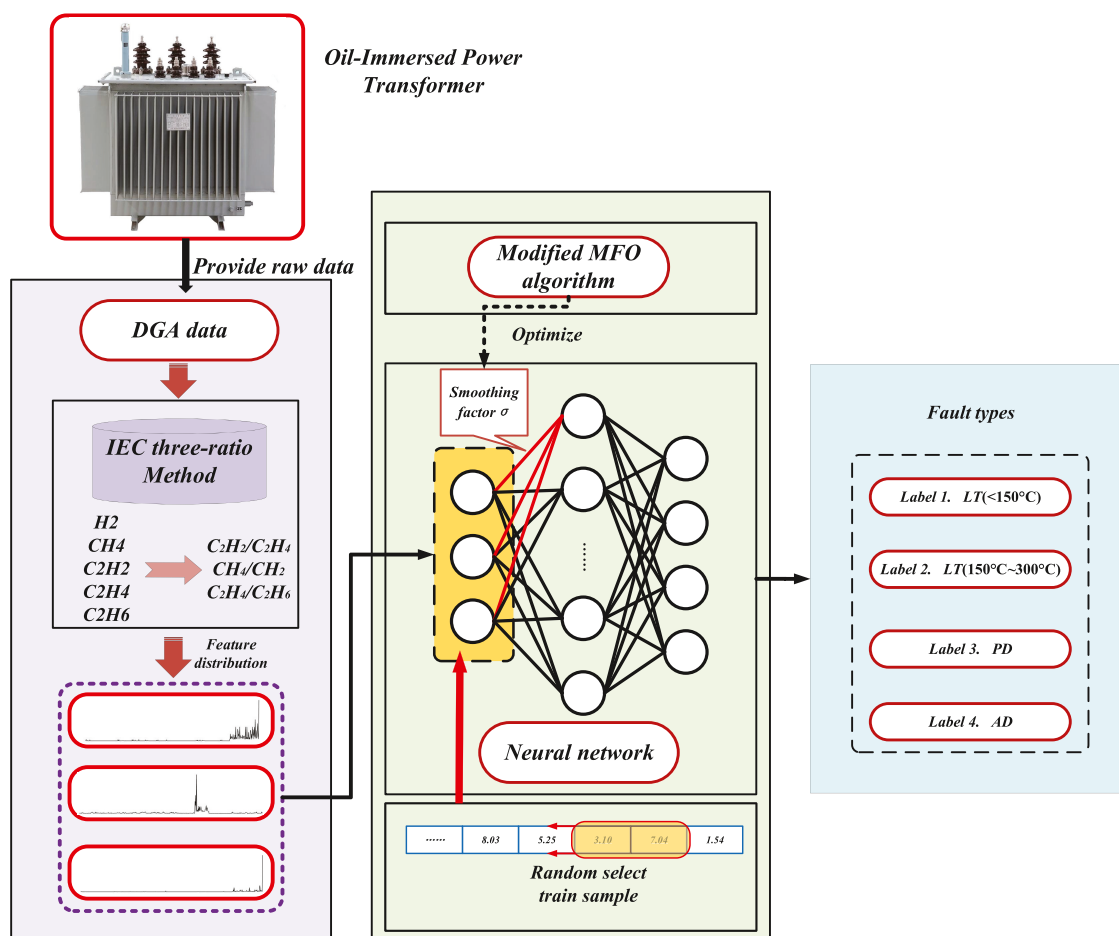
Figure 3. Implemented framework of the power transformer fault diagnosis.

Table 1. Some Real Data from China Electric Power Research Institute Diagnosing Power Transformer Fault Types by the DGA Method[a]

	dissolved gas (μL/L)					
fault type	CH_4	C_2H_2	C_2H_4	C_2H_6	TH	sources
LT (<150 °C)	83	53	13	1.2	150.2	Jiujiang PSC
LT (150−300 °C)	6.5	98	16	1.5	122	Fuzhou PSC
LT (150−300 °C)	193	191	28	16	428	Yingtan PSC
LT (150−300 °C)	12	46	11	1.8	70.8	Nanchang PSC
LT (150−300 °C)	3.5	31	8.2	1	43.7	Yichun PSC
AD	61	307	105	6	479	Yingtan PSC

[a]PSC is the power supply company, TH is the total hydrocarbon, the temperature is 25 °C, the humidity is 50.

is set as the fitness value, and the corresponding fitness function can be expressed as

$$f(x) = \frac{1}{N} \sum_{i=1}^{N} (Y_i - O_i)^2 \tag{15}$$

where Y_i is the actual output after network training and O_i is the theoretically expected output.

Step 4: Initialize the position of the smoothing factor randomly and record it by eq 1.

Step 5: Calculate the fitness of each smoothing factor position by eq 15 and record the current optimal solution position.

Step 6: Sort the positions in ascending order according to the size of the fitness and select the better solution as the

next-generation flame position (if $l = 1$, then as the contemporary flame position).

Step 7: Update f_{no} by eq 2.

Step 8: Update the smoothing factor positions by eq 9.

Step 9: Continue to the next step if the maximum iterations condition ($l < T$) is satisfied; otherwise, return to Step 5.

Step 10: Fed the optimized smoothing factor σ into the PNN network for training to obtain the best PNN fault diagnosis model.

Step 11: Fed the test samples into the network instead of the training samples to get the corresponding data.

Table 2. Coding Format for Different Fault Types

fault type	LT (<150 °C)	LT (150−300 °C)	PD	AD
coding format	1	0	0	0
	0	1	0	0
	0	0	1	0
	0	0	0	1

Table 3. Partial Sample Data

dissolved gas (μL/L)			
C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6	fault type
0.07143	0.54667	0.03571	LT (<150 °C)
0.03333	0.5	0.03333	LT (<150 °C)
0.07831	0.14945	0.44828	LT (<150 °C)
0.01583	1.68618	0.00179	LT (150−300 °C)
0.04878	1.16761	0.025	LT (150−300 °C)
0.0117	1.13993	0.00514	LT (150−300 °C)
0.05556	0.07524	0.05882	PD
0	0.08595	0	PD
0.06667	0.06754	0.2191	PD
0.16667	0.26891	0.06818	AD
0.27372	0.12839	0.3012	AD
0.26642	0.1946	0.23175	AD

Table 4. Parameters Settings of Different Methods

methods	parameters settings
MMFO-PNN	$N = 10$, $T = 20$
Sa-PNN	$s_{min} = 0.1$, $s_{max} = 0.8$, $s_{inter} = 4.9$, $T = 20$
GA-PNN	$P_m = 0.01$, $P_c = 0.7$, $N = 10$, $T = 20$
BA-PNN	$A = 0.5$, $r = 0.5$, $N = 10$, $T = 20$
PSO-PNN	$c_1 = 1.49445$, $c_2 = 1.49445$, $T = 20$
GA-SVM	$P_m = 0.01$, $P_c = 0.9$, $N = 10$, $T = 20$, $k_{max} = 10^3$, $k_{min} = 10^{-3}$, $p_{max} = 10^3$, $p_{min} = 10^{-3}$
GWO-hybrid KELM	$N = 10$, $T = 20$
SaE-ELM	$N = 10$, $T = 20$, strategy = 1, numst = 4, hidden number = 10
MBA-BP	$A = 0.5$, $r = 0.5$, $N = 10$, $T = 20$, hidden number = 10
MCS-BP	$Pa = 0.25$, $N = 10$, $T = 20$, hidden number = 10
PNN	$\sigma = 0.07$
IEC	non

■ IMPLEMENTATION AND EXPERIMENTAL SETUP

Model Implementation. In this section, we discuss the implementation of the proposed MMFO-based PNN method. Figure 3 shows the implemented framework for fault diagnosis. First, a series of DGA data are sampled during the actual operation of the oil-immersed power transformer, and a portion of the DGA data is randomly selected as training samples and input into the MMFO-PNN model to optimize the training of the neural network and output its fault-type classification results.

Then, the remaining DGA data, i.e., the test samples, are used to test the established neural network and verify its effectiveness.

According to the dissolved gas content in power transformer oil, oil-immersed power transformer faults can be classified into four categories, namely, low-temperature (LT) overheating (<150 °C), low-temperature overheating (LT) (150−300 °C), partial discharge (PD), and arc discharge (AD).

Table 1 shows some of the real data of the DGA method used by the China Electric Power Research Institute to determine the fault types of oil-immersed power transformers. In addition, to facilitate the training of the probabilistic neural network, the four fault types of power transformers will be coded in the form shown in Table 2.

Data Collection and Preprocessing. Considering the influence of temperature, humidity, and other parameters on power transformer fault diagnosis, we collected several groups of real data of various gas contents in oil of oil-immersed power transformers from various power supply companies and substations in Jiangxi Province. For each group of gas data, we selected some characteristic gas contents (C_2H_2, C_2H_4, CH_4, H_2, C_2H_4, C_2H_6) dissolved in the oil as the main basis for fault-type determination. After screening all of the gas data and processing with the IEC three-ratio method, 525 sets of valid DGA data were obtained, including 333 sets of low-temperature over-heating (LT) (<150 °C), 39 sets of low-temperature over-heating (LT) (150−300 °C), 65 sets of partial discharge (PD), and 88 sets of arc discharge (AD), where 470 sets of data were used as training samples and the remaining 55 sets of data were used as test samples. Some of the DGA data are shown in Table 3.

Experimental Setting. To effectively evaluate the performance of our proposed method in fault diagnosis of oil-immersed power transformer, we compare MMFO-PNN with 11 other methods. It includes five types of PNN-based fault diagnosis models, namely, PNN, particle swarm optimization (PSO)-PNN, bat algorithm (BA)-PNN,[27] genetic algorithm (GA)-PNN, and adaptive (Sa)-PNN;[28] it also includes six fault diagnosis methods that have been proposed in previous work and verified by experiments, which are self-adaptive evolutionary extreme learning machine (SaE-ELM),[29] modified bat algorithm (MBA)-BP,[30] gray wolf optimizer optimized hybrid kernel extreme learning machine (GWO-hybrid ELM),[20] genetic algorithm (GA)-SVM,[31] modified cuckoo search (MCS)-BP,[32] and IEC three-ratio method (hereinafter referred to as IEC).[33] Through the MATLAB 2018a simulation platform based on the same DGA data set training and testing to compare their performance. The parameter settings of these 11 methods are shown in Table 4.

■ EXPERIMENTAL RESULTS AND DISCUSSION

The results of accuracy comparison of different methods based on PNN are shown in Table 5 with 86.09% for PNN and 92.34,

Table 5. Comparison of Different Methods Based on PNN

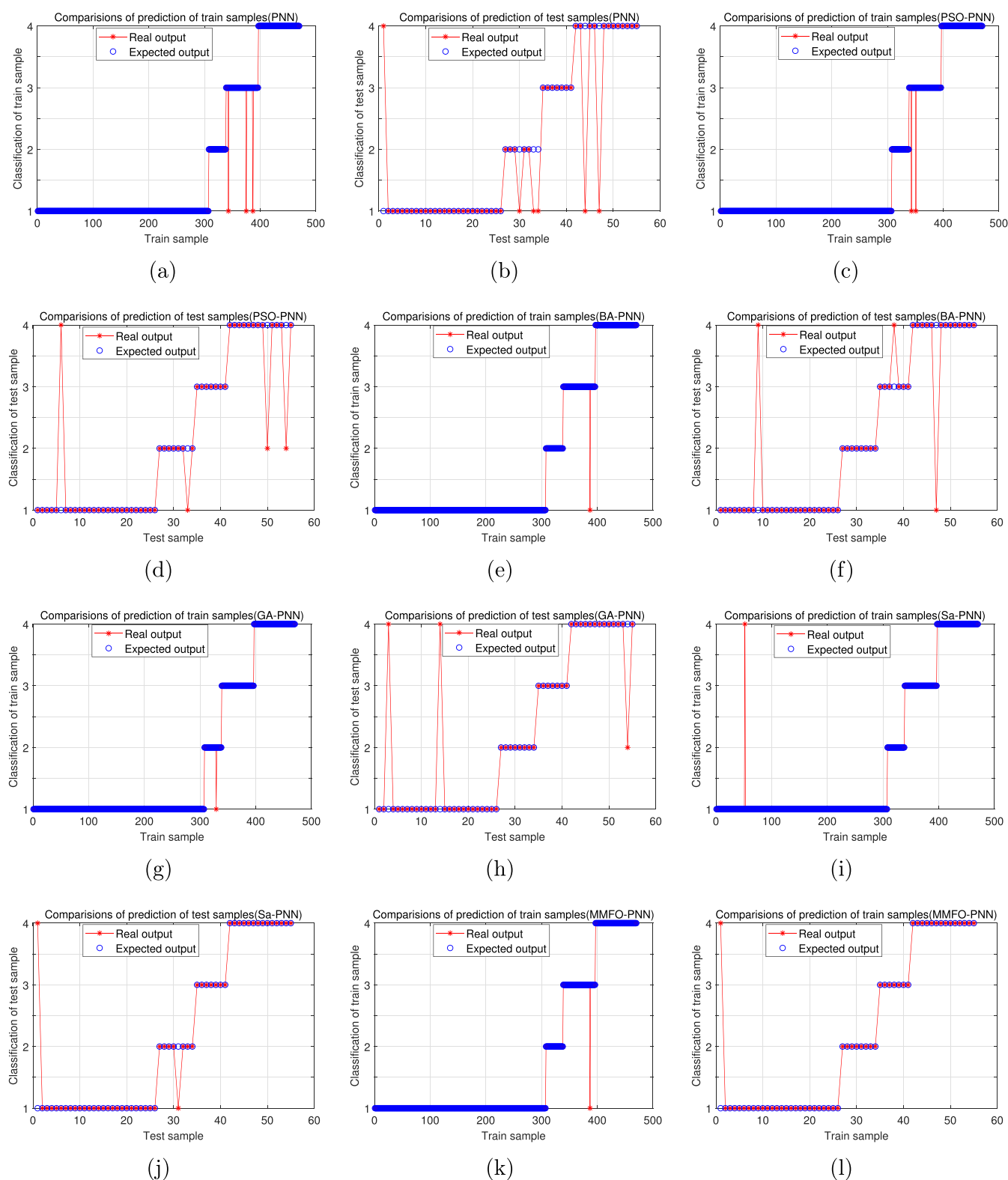| fault type	accuracy (%)					
	MMFO-PNN	PNN	PSO-PNN	BA-PNN	GA-PNN	Sa-PNN
LT (<150 °C)	96.15	96.15	96.15	96.15	88.46	96.15
LT (150−300 °C)	100.00	62.50	87.50	100.00	100.00	87.50
PD	100.00	100.00	100.00	85.71	100.00	100.00
AD	100.00	85.71	85.71	92.86	92.86	100.00
average	99.04	86.09	92.34	93.68	95.33	95.91

Figure 4. Classification results of different models. (a), (c), (e), (g), (i), and (k) represent the results of train sample classification for different methods. (b), (d), (f), (h), (j), and (l) are the results of test sample classification for different methods.

93.68, 95.33, and 95.91% for the remaining four improved algorithms of PNN (PSO-PNN, BA-PNN, GA-PNN, and Sa-PNN), respectively. The MMFO-PNN algorithm achieves 96.15% (25/26) accuracy for LT (<150 °C) fault type, 100% accuracy for the remaining three fault types, and 99.04% average accuracy. MMFO-PNN has the highest accuracy and average

accuracy for each fault type, which is significantly better than the remaining five algorithms. Even in the AD fault type where the accuracy is generally low, the accuracy of MMFO-PNN reaches 100%. It fully shows that MMFO is better than PSO, BA, GA, Sa, and other optimization methods for PNN optimization,

Table 6. Fault Diagnosis Based on the IEC Three-Ratio Method

fault type	C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6
LT (<150 °C)	<0.1	0.1−1	1−3
LT (150−300 °C)	<0.1	≥1	<1
PD	<0.1	<0.1	<1
AD	0.1−3	<1	NS

demonstrating the superiority of MMFO-PNN in power transformer fault diagnosis.

To show the prediction results of various PNN models more intuitively, we plotted the classification results of different algorithms, as shown in Figure 4, where a, c, e, g, i, and k are the classification results of training samples and b, d, f, h, j, and l are the classification results of test samples. In Figure 4 (each sample on the x-axis in the subgraph represents a faulty transformer; 1, 2, 3, and 4 on the y-axis are the labels of each type of fault, respectively), we can visually see that the classification output of the remaining five algorithms is not satisfactory; especially in the first, second, and fourth categories of faults, there are several classification errors, but from the k and l subfigures, we can see that MMFO-PNN has a high accuracy for these classification faults; only in the first category of faults, there is a classification error; the other three faults are all correctly classified. The excellent classification performance of MMFO-PNN in the test and classification samples shows that the stability and classification performance of the MMFO-PNN algorithm itself are better than those of the other algorithms, and it is more applicable and reliable in the field of power transformer fault diagnosis.

The above comparison experiments show that MMFO optimizes PNN better and has better convergence capability than other intelligent algorithms. Next, to demonstrate that the MMFO-PNN model has strong practicality in the field of transformer fault diagnosis, we let the MMFO-PNN model be compared with a variety of transformer fault diagnosis methods proposed by previous authors (including empirical methods and methods in the field of artificial intelligence) to prove the superiority of the proposed model from several perspectives. Among them, MCS-BP, MBA-BP, SaE-ELM, GWO-hybrid KELM, and GA-SVM are the fault diagnosis methods proposed by the researchers combined with artificial intelligence techniques. Through previous experiments, they were proved to be effective methods in the field of transformer fault diagnosis. The IEC three-ratio method is a traditional and commonly used fault diagnosis method for oil-immersed transformers. It is essentially an empirical discrimination method in which the researcher makes an empirical judgment of the fault type based on the range of dissolved gas contents sampled. For the existing fault types in this paper, the IEC three-ratio method is identified, as shown in Table 6.

The comparison of the diagnostic accuracy of MMFO with the above methods is shown in Table 7, where the diagnostic simulation results of the training set and test set of the artificial intelligence class methods are shown in Figure 5.

The comparison results shown in Table 7 indicate that MMFO-PNN still has a significant advantage in terms of accuracy compared with the other six methods. At the same time, combining Tables 5 and 7, we can also see that accuracy of the IEC three-ratio method is much lower than that of AI class methods, which shows that the empirical discriminative approach is not effective in the face of complex concentrated data. Among the artificial intelligence methods, the accuracy of GA-SVM is relatively low. Li and Zhang et al. used the genetic algorithm to filter the parameters of SVM for optimization, which significantly improved the classification accuracy of SVM, but there is still a gap between the SVM classifier and the neural network-based methods for multiclassification problems.

For engineering-type problems like transformer fault diagnosis, efficiency is also an important indicator of fault diagnosis model performance. We use the running time of various fault diagnosis algorithms as a measure of efficiency (simulation platform, MATLAB 2018a; computational platform, i7-9750CPU@2.60GHz).

Also, we use the error rate to measure the ability of fault diagnosis models to identify fault types. As can be seen from Table 8, the time consumption of fault diagnosis models based on artificial intelligence methods is generally higher than that of IEC three-ratio methods (except PNN). This is in accordance with the law of objective facts because machine learning requires sufficient training in identifying the same number of test sets, while the IEC three-ratio method based on empirical discrimination does not require training and can directly discriminate fault types based on the range of the three gas ratios. However, in contrast, the IEC three-ratio method has the highest error rate of all methods. Here, we can also analyze the characteristics of the traditional empirical fault diagnosis method: although it is easy and fast to use, the error rate is too high to meet the needs of today's industry. In contrast, MMFO-PNN has the lowest error rate and takes only a little more time than PNN and IEC three-ratio methods. Considered together, MMFO-PNN is the fault diagnosis method with the best engineering practical performance among all of the methods involved in the comparison.

The error rate is one of the simplest indicators to discriminate the classifier performance. However, the error rate can only calculate the percentage of cases with wrong judgments among all cases, and it does not reflect how the cases with wrong judgments are classified wrong. Therefore, we use the mean square error (MSE) to reflect the dispersion of the classification results. The MSEs of the training and test sets for different methods are shown in Table 9. It is worth noting that since the IEC three-ratio method is not a machine learning class method,

Table 7. Comparison Results of Different Methods

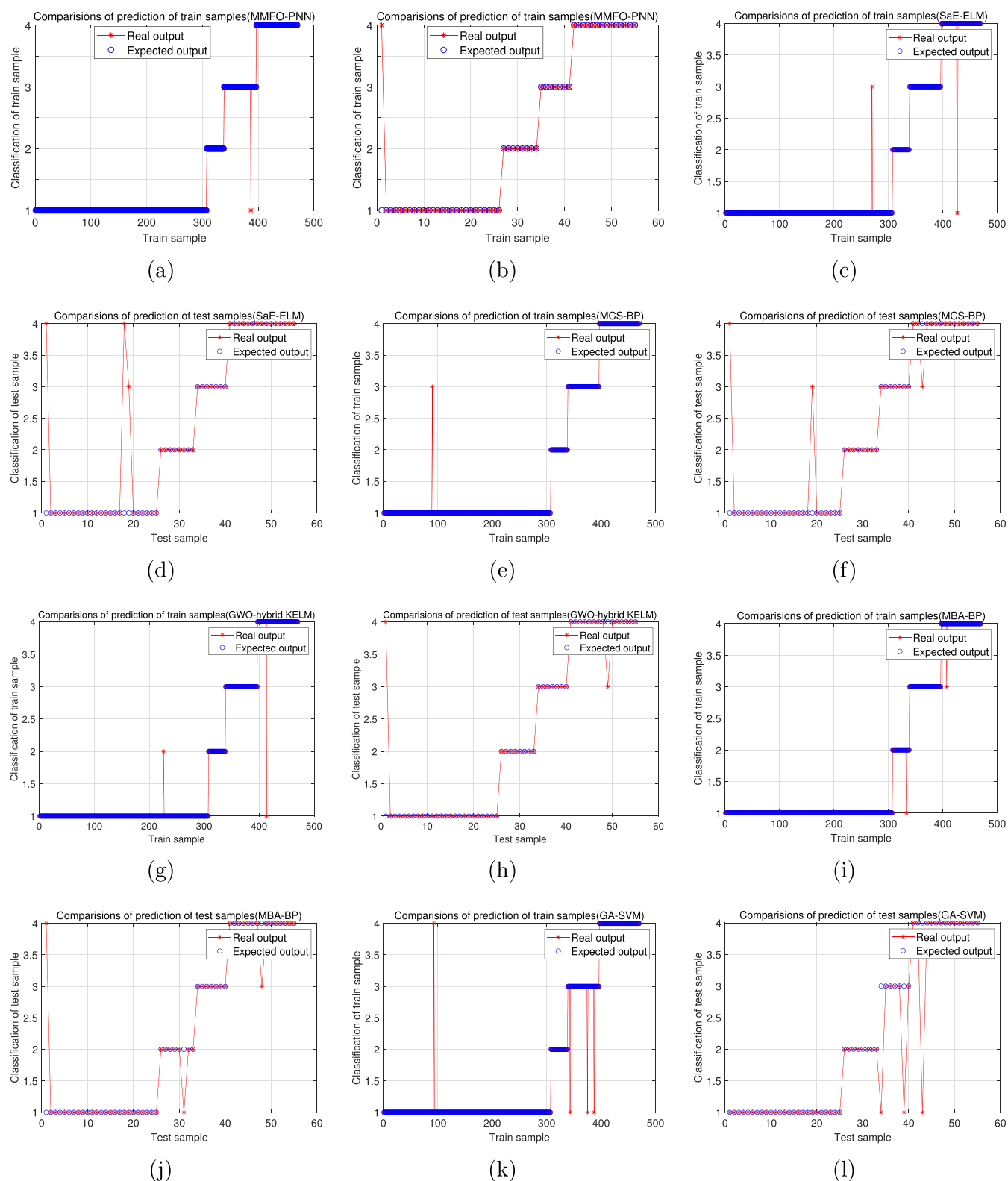| fault type	accuracy (%)						
	MMFO-PNN (%)	SaE-ELM (%)	GWO-hybrid KELM (%)	MCS-BP (%)	MBA-BP (%)	GA-SVM (%)	IEC (%)
LT (<150 °C)	96.15	88.00	96.00	92.00	96.00	100.00	77.17
LT (150−300 °C)	100.00	100.00	100.00	100.00	87.50	100.00	97.44
PD	100.00	100.00	100.00	100.00	100.00	71.43	18.47
AD	100.00	100.00	93.33	93.33	93.33	93.33	100.00
average	99.04	97.00	97.33	96.33	94.21	91.19	70.65

Figure 5. Classification results of different models. (a), (c), (e), (g), (i), and (k) represent the results of train sample classification for different methods. (b), (d), (f), (h), (j), and (l) are the results of test sample classification for different methods.

there is no need to divide the training and test sets, but to ensure the fairness of the experiments, we use the same training set data and test set data as the machine learning class method for the classification experiments when testing the performance of the IEC three-ratio method.

It can be seen in Table 9 that the MSE of MMFO-PNN is the smallest among all methods when identifying the test set data. Therefore, it can be seen that, compared with other methods, MMFO-PNN has strong robustness and generalization ability.

Table 8. Efficiency and Error Rate of Different Methods

method	time (s)	error rate (%)
MCS-BP	18.0127	5.4545
GWO-hybrid KELM	16.3495	3.636
GA-SVM	15.8398	5.4545
MBA-BP	24.2396	5.4545
BA-PNN	14.9398	5.4545
SaE-ELM	13.3365	5.4545
GA-PNN	11.2378	5.4545
PSO-PNN	9.3433	7.273
Sa-PNN	8.3158	3.636
MMFO-PNN	7.2763	1.818
IEC	3.7653	24.76
PNN	2.9606	10.91

Table 9. Mean Square Error of Different Methods

algorithms	MSE of train sample	MSE of test sample
MMFO-PNN	0.0085	0.1636
Sa-PNN	0.0191	0.1818
GWO-hybrid KELM	0.0213	0.1818
MBA-BP	0.0043	0.2000
MCS-BP	0.0085	0.2545
GA-SVM	0.0447	0.3091
PSO-PNN	0.0170	0.3273
BA-PNN	0.0085	0.3455
GA-PNN	0.0021	0.4000
SaE-ELM	0.0468	0.4000
PNN	0.0255	0.5455
IEC	0.4787	0.6545

Table 10. Comparison of Different Machine Learning Methods with the F1-Score

| method | Marco F1-score | | | | |
	LT (%) (<150 °C)	LT (%) (150−300 °C)	PD (%)	AD (%)	average (%)
MMFO-PNN	98.04	100.00	100.00	96.55	98.65
Sa-PNN	96.15	93.33	100.00	96.55	96.51
GWO-hybrid KELM	97.96	100.00	93.33	93.75	96.26
SaE-ELM	93.62	100.00	94.12	94.44	95.54
GA-PNN	95.83	94.12	100.00	89.66	94.90
MCS-BP	95.83	100.00	90.00	93.75	94.90
BA-PNN	96.15	100.00	92.31	89.66	94.53
MBA-BP	96.15	93.33	93.33	93.75	94.14
GA-SVM	94.34	93.33	90.91	96.55	93.78
PSO-PNN	96.15	82.35	100.00	88.89	91.85
PNN	89.29	76.92	100.00	88.89	88.77

Considering the unbalanced nature of the experimental data (far more fault data of the low-temperature overheating (LT) (<150 °C) type than other types of faults), we use the metric F1-score to evaluate the performance of different fault diagnosis methods. The Marco F1-score for different methods is shown in Table 10.

When the data samples are unbalanced, the F1-score is more indicative of the practical performance of the classifier compared to the accuracy. Table 10 shows that the Marco F1-score of MMFO-PNN is the highest among all of the fault diagnosis methods, showing the excellent performance of MMFO-PNN in the field of transformer fault diagnosis.

Table 11. Results of Fivefold Cross Validation of Different Machine Learning Models (%)

model	fold 1	fold 2	fold 3	fold 4	fold 5	average accuracy
MMFO-PNN	99.68	99.64	100.00	95.80	99.88	98.40
PNN	78.03	73.89	77.42	82.41	75.00	77.35
PSO-PNN	94.73	95.69	91.25	93.10	93.75	93.70
BA-PNN	93.73	93.98	90.97	93.94	97.17	93.96
GA-PNN	95.38	91.71	98.08	90.36	97.62	94.63
Sa-PNN	93.44	92.39	85.88	93.03	91.16	91.18
SaE-ELM	96.46	97.68	95.59	94.92	91.32	95.19
GWO-hybrid KELM	98.33	98.36	97.67	96.19	97.84	97.68
MCS-BP	97.05	95.43	96.15	93.12	95.86	95.52
MBA-BP	93.12	99.71	92.84	94.58	95.55	95.16
GA-SVM	89.03	87.34	92.05	86.14	90.06	88.92

To evaluate the predictive performance of the proposed MMFO-PNN fault diagnosis model, we performed fivefold cross validation on the proposed model and other machine learning models. Nonrepetitive sampling divides the gas ratio data into five randomly. Each time, one of them is selected as the test set, and the remaining four are used as the training set for model training, and the average prediction accuracy of each test set is calculated. After repeating the above process five times, the average of the five sets of test results is calculated as an estimate of the model accuracy. The results are shown in Table 11.

The results show that the performance of the proposed MMFO-PNN model in fivefold cross validation is still better than other machine learning models. The average accuracy of some models after cross validation is lower than the diagnostic accuracy shown in Table 5, but the results of cross validation can better reflect the true performance of the model.

Figure 6 shows the fitness curves of MMFO-PNN when the maximum number of iterations is set to 4, 10, 80, and 100, respectively, and Figure 7 shows its corresponding average classification accuracy. It can be seen that MMFO-PNN converges quickly, and when the maximum number of iterations is 10, it converges to the optimum in the seventh generation, with rapid classification and accurate results, which fully reflects the engineering practicality of MMFO-PNN. When the maximum number of iterations is 80, the sixth generation converges to the optimum. It can also be seen that the initial error of MMFO-PNN is very small and converges to the optimum in just one time, which fully demonstrates its excellent global search and convergence ability. In addition, when the maximum number of iterations is 80, even if it is optimal in the sixth generation, it does not fall into overfitting afterward, which affects the accuracy. Only when the maximum number of iterations is 100, the 85th generation fitness is 0 and the accuracy decreases from 99.038 to 58.036%.

The performance of machine learning algorithm models is often very sensitive to the input parameters of the model. The adjustable input parameters of the MMFO-PNN model we built are the number of populations (N) and the maximum number of iterations (T). In the previous comparison experiments, we uniformly set T to 20 and N to 10 for all models requiring input T and N. The purpose is to control the variables and ensure the fairness of the experiments. The results of the previous experiments (Table 8) show that MMFO-PNN has superior diagnostic performance compared to other models under fair conditions. In this section, we will focus on the effect of input parameters on the performance of the MMFO-PNN model by
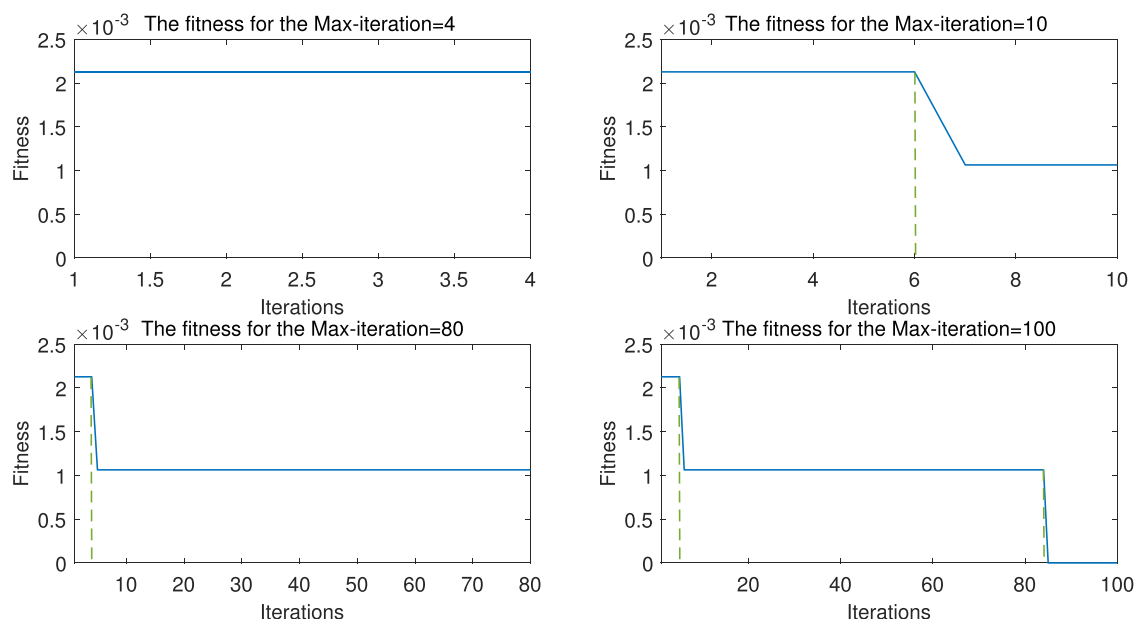
Figure 6. Fitness curves of different iteration times.
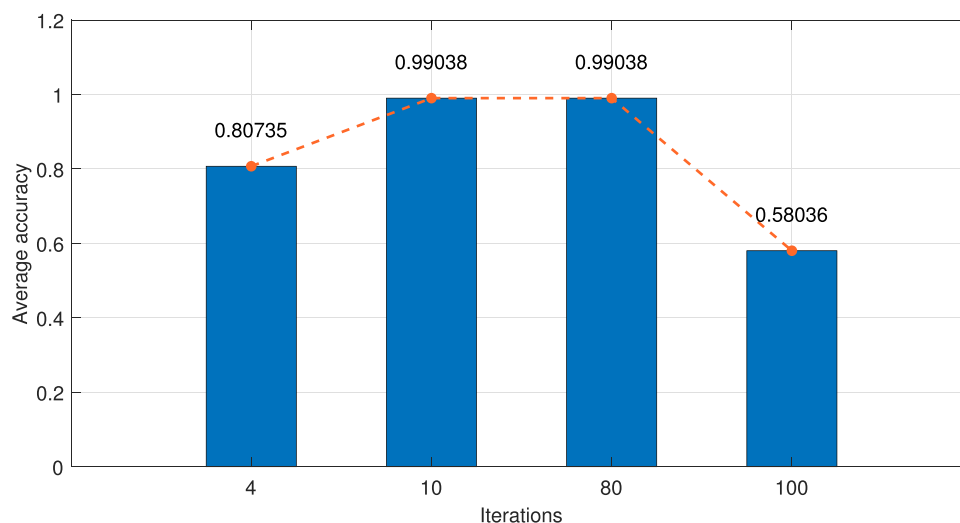

Figure 7. Average accuracy of different iterations.

adjusting the input parameters to make the MMFO-PNN model optimal.

Figures 6 and 7 only show the effect of T on accuracy and do not discuss the effect of N on the overall model performance. To observe the model performance as a whole under different input parameters, we plot the three-dimensional (3D) plot shown in Figure 8 to investigate the effect on MMFO-PNN when the only two input parameters, N and T, are varied simultaneously. We use the fitness defined by eq 15 to judge the model performance; the lower the value, the better the model performance. To give the reader a clearer understanding of the 3D plot shown in Figure 8a, we also plot its corresponding top view, which is Figure 8b. In Figure 8, the bluer the color, the smaller its fitness, and the better the performance of the model. As can be seen, T that is too large leads to a surge in fitness, while T that is too small and N that is small may also lead to a larger fitness. In addition, if N is relatively large, then fitness will increase more quickly with the number of iterations. Also, if N is too small, then fitness will appear to be larger due to a small number of

iterations, or even always larger no matter how the number of iterations changes. Combining the above phenomena, we can learn that for both N and T, neither can be too large or too small; otherwise, the model cannot be put in an optimal state. Moreover, if N is chosen appropriately, the fault tolerance for the choice of T will also increase.

To show more clearly the effect of T and N each on the fitness of the MMFO-PNN model, as shown in Figures 9 and 10, we investigate the effect of another input parameter on the model at $T = 10$ and $N = 3$, respectively. It can be seen that neither too large nor too small is good for either N or T, which will lead to the degradation of the model performance. Therefore, the choice of model input parameters is very important to the final learning effect of the model, which is also a common problem in the field of machine learning. Through the above findings, considering that the fault diagnosis model needs to balance the error rate and efficiency, $N = 3$ and $T = 10$ can be chosen as the final input parameters for the established MMFO-PNN model.
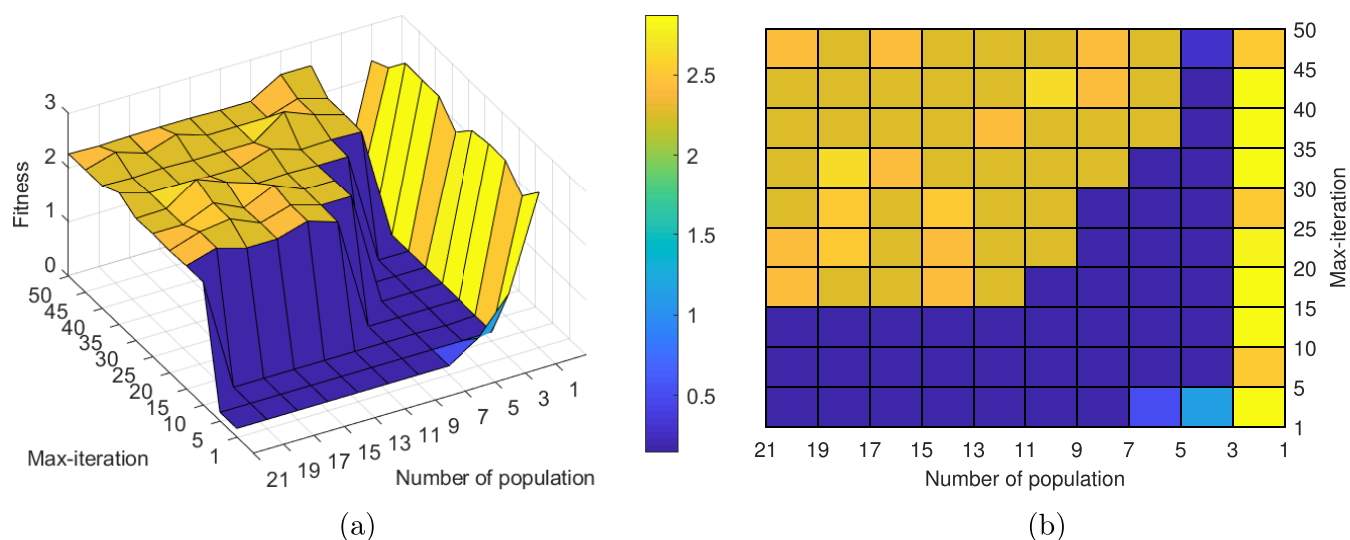
Figure 8. Three-dimensional diagram of the change in fitness: (a) main view and (b) top view.
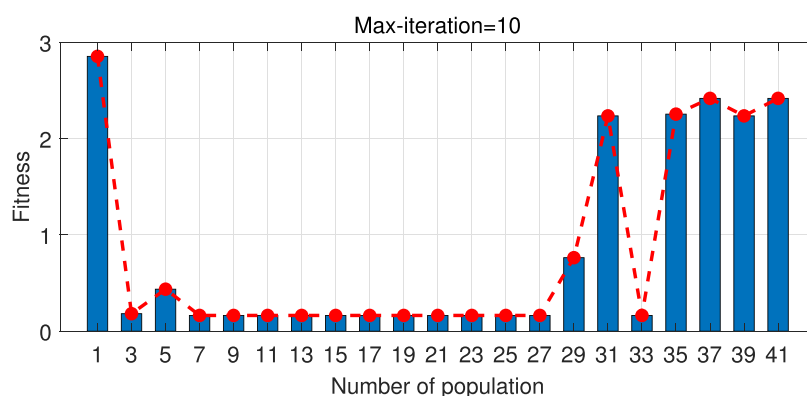

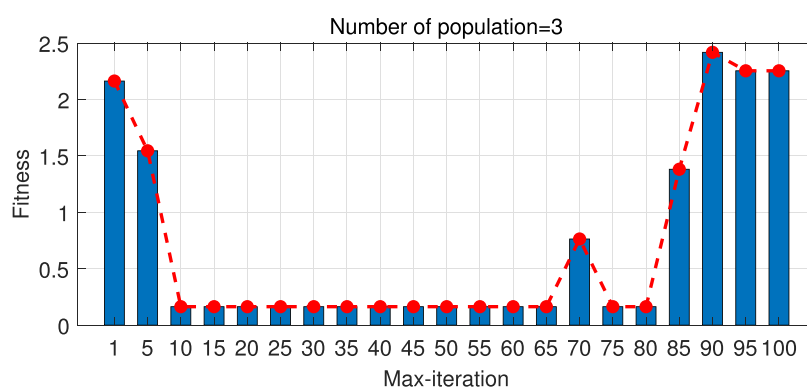**Figure 9.** Trends in fitness values with the number of population (max iteration = 10).

Figure 10. Trends of fitness values with max iteration (number of population = 3).

■ CONCLUSIONS

In this paper, we proposed an MMFO-based PNN as a fault diagnosis method for power transformers using MMFO to optimize the smoothing factor (σ), which is crucial to the performance of PNN, to improve the performance of PNN. For the optimization of PNN, the MMFO algorithm can search the solution space to a greater extent than other optimization algorithms, enhancing the global search and convergence to find a better global optimal solution. We validated these algorithms using real data collected from transformers by evaluating the performance of the algorithmic models. The experimental results show that the proposed MMFO outperforms other algorithms, can effectively enhance the global optimal solution search performance, and has good stability to overcome the perturbation of noisy data, thus improving the fault determination accuracy.

In addition, the developed technique is also applicable to other engineering fields, such as sensor and diesel machine fault diagnosis.

AUTHOR INFORMATION

Corresponding Author

Xiaohui Yang − *College of Information Engineering, Nanchang University, Nanchang 330031, China;* ⦿ orcid.org/0000-0001-5439-4787; Phone: +86-139-7094-145; Email: yangxiaohui@ncu.edu.cn

Authors

Yichen Zhou − *College of Qianhu, Nanchang University, Nanchang 330031, China*

Lingyu Tao − *College of Information Engineering, Nanchang University, Nanchang 330031, China*

Li Yang − *College of Information Engineering, Nanchang University, Nanchang 330031, China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.1c01878

Author Contributions

Conceptualization, X.Y. and Y.Z.; methodology, Y.Z.; software, L.T.; validation, Y.Z., L.T., and X.Y.; formal analysis, L.Y.; investigation, Y.Z.; resources, X.Y.; writing—original draft preparation, Y.Z. and L.T.; writing—review and editing, Y.Z., L.T., and X.Y.; visualization, Y.Z.; supervision, X.Y. All authors have read and agreed to the published version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

ABBREVIATIONS

AD	arc discharge
ANN	artificial neural network
BA	bat algorithm
DBN	deep belief network
DGA	dissolved gas analysis
ELM	extreme learning machine
FCM	fuzzy c-means
GA	genetic algorithm
GWO	gray wolf optimization
HMM	hidden Markov model
IEC	International Electrotechnical Commission
KELM	kernel extreme learning machine
KNN	*k*-nearest neighbor
MLP	multilayer perceptron
MBA	modified bat algorithm
MCS	modified cuckoo search
MMFO	modified moth flame optimization
PD	partial discharge
PNN	probabilistic neural network
PSC	power supply company
PSO	particle swarm optimization
QPSO	quantum-inspired particle swarm optimization
RBFNN	radial basis function neural network
Sa	self-adaptive
SaE	self-adaptive evolutionary
SVM	support vector machine
σ	smoothing factor

REFERENCES

(1) Yu, S.; Zhao, D.; Chen, W.; Hou, H. *Oil-immersed Power Transformer Internal Fault Diagnosis Research Based on Probabilistic Neural Network.* ANT/SEIT, 2016.

(2) Chen, W.; Pan, C.; Yun, Y.; Liu, Y. Wavelet networks in power transformers diagnosis using dissolved gas analysis. *IEEE Trans. Power Delivery* **2009**, *24*, 187−194.

(3) Kim, J.-W.; Park, B.; Jeong, S. C.; Kim, S. W.; Park, P. Fault diagnosis of a power transformer using an improved frequency-response analysis. *IEEE Trans. Power Delivery* **2005**, *20*, 169−178.

(4) Wang, X.; Wang, Z.; Liu, Q.; Dyer, P. Dissolved gas analysis of thermal faults in transformer liquids simulated using immersed heating method. *IEEE Trans. Dielectr. Electr. Insul.* **2018**, *25*, 1749−1757.

(5) Kari, T.; Gao, W.; Zhao, D.; Zhang, Z.; Mo, W.; Wang, Y.; Luan, L. An integrated method of ANFIS and Dempster-Shafer theory for fault diagnosis of power transformer. *IEEE Trans. Dielectr. Electr. Insul.* **2018**, *25*, 360−371.

(6) Cong, H.; Du, S.; Li, Q.; Liu, J. Electro-thermal fault diagnosis method of rapo vegetable oil transformer based on characteristic gas and ratio criterion. *IEEE Access* **2019**, *7*, 101147−101159.

(7) Kirkbas, A.; Demircali, A.; Koroglu, S.; Kizilkaya, A. Fault diagnosis of oil-immersed power transformers using common vector approach. *Electr. Power Syst. Res.* **2020**, *184*, No. 106346.

(8) Amedi, H.; Baghban, A.; Ahmadi, M. Evolving machine learning models to predict hydrogen sulfide solubility in the presence of various ionic liquids. *J. Mol. Liq.* **2016**, *216*, 411−422.

(9) Ahmadi, M. A.; Ahmadi, A. Applying a sophisticated approach to predict CO2 solubility in brines: application to CO2 sequestration. *Int. J. Low-Carbon Technol.* **2016**, *11*, 325−332.

(10) Ahmadi, M.; Ebadi, M.; Yazdanpanah, A. Robust intelligent tool for estimating dew point pressure in retrograded condensate gas reservoirs: Application of particle swarm optimization. *J. Pet. Sci. Eng.* **2014**, *123*, 7−19.

(11) Su, Q.; Mi, C.; Lai, L.; Austin, P. A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer. *IEEE Trans. Power Syst.* **2000**, *15*, 593−598.

(12) Dhini, A.; Surjandari, I.; Faqih, A.; Kusumoputro, B. In *Intelligent Fault Diagnosis for Power Transformer Based on DGA Data Using Support Vector Machine (SVM)*, 2018 3rd International Conference on System Reliability and Safety (ICSRS), 2018; pp 294−298.

(13) 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 Delivery* **2019**, *34*, 1393−1400.

(14) Yu, H.; Wu, Q.; Lu, Y.; Hu, C.; Wang, Y.; Liu, G. In *Research on Fault Diagnosis of Power Transformer Equipment Based on KNN Algorithm*, International Conference on Mechatronics and Intelligent Robotics, 2017; pp 172−176.

(15) Zhang, Y.; Ding, X.; Liu, Y.; Griffin, P. An artificial neural network approach to transformer fault diagnosis. *IEEE Trans. Power Delivery* **1996**, *11*, 1836−1841.

(16) Zhang, W.; Yang, X.; Deng, Y.; Li, A. An Inspired Machine-Learning Algorithm with a Hybrid Whale Optimization for Power Transformer PHM. *Energies* **2020**, *13*, No. 3143.

(17) Huang, Y. Evolving neural nets for fault diagnosis of power transformers. *IEEE Trans. Power Delivery* **2003**, *18*, 843−848.

(18) 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.

(19) 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.

(20) Huang, X.; Wang, X.; Tian, Y. In *Research on Transformer Fault Diagnosis Method based on GWO Optimized Hybrid Kernel Extreme Learning Machine*, 2018 Condition Monitoring and Diagnosis (CMD), 2018; pp 1−5.

(21) Song, T.; Jamshidi, M. M.; Lee, R. R.; Huang, M. A modified probabilistic neural network for partial volume segmentation in brain MR image. *IEEE Trans. Neural Netw.* **2007**, *18*, 1424−1432.

(22) Kusy, M.; Zajdel, R. Probabilistic neural network training procedure based on Q (0)-learning algorithm in medical data classification. *Appl. Intell.* **2014**, *41*, 837−854.

(23) Mirjalili, S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Syst.* **2015**, *89*, 228−249.

(24) Ma, L.; Wang, C.; gang Xie, N.; Shi, M.; Ye, Y.; Wang, L. Moth-flame optimization algorithm based on diversity and mutation strategy. *Appl. Intell.* **2021**, 1−37.

(25) Yu, Y.; Gao, S.; Cheng, S.; Wang, Y.; Song, S.; Yuan, F. CBSO: a memetic brain storm optimization with chaotic local search. *Memet. Comput.* **2018**, *10*, 353−367.

(26) Specht, D. F. Probabilistic neural networks. *Neural Networks* **1990**, *3*, 109−118.

(27) 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.

(28) Yi, J.-H.; Wang, J.; Wang, G.-G. Improved probabilistic neural networks with self-adaptive strategies for transformer fault diagnosis problem. *Adv. Mech. Eng.* **2016**, *8*, 1−13.

(29) Li, S.; Wu, G.; Gao, B.; Hao, C.; Xin, D.; Yin, X. Interpretation of DGA for transformer fault diagnosis with complementary SaE-ELM and arctangent transform. *IEEE Trans. Dielectr. Electr. Insul.* **2016**, *23*, 586−595.

(30) Dong, H.; Yang, X.; Li, A.; Xie, Z.; Zuo, Y. Bio-Inspired PHM Model for Diagnostics of Faults in Power Transformers Using Dissolved Gas-in-Oil Data. *Sensors* **2019**, *19*, No. 845.

(31) Li, J.; Zhang, Q.; Wang, K.; Wang, J.; Zhou, T.; Zhang, Y. Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine. *IEEE Trans. Dielectr. Electr. Insul.* **2016**, *23*, 1198−1206.

(32) Li, A.; Yang, X.; Dong, H.; Xie, Z.; Yang, C. Machine Learning-Based Sensor Data Modeling Methods for Power Transformer PHM. *Sensors* **2018**, *18*, No. 4430.

(33) Liu, Z.; Song, B.; Li, E.; Mao, Y.; Wang, G. Study of "code absence" in the IEC three-ratio method of dissolved gas analysis. *IEEE Electr. Insul. Mag.* **2015**, *31*, 6−12.