

Article

An Inspired Machine-Learning Algorithm with a Hybrid Whale Optimization for Power Transformer PHM

Wei Zhang ¹, Xiaohui Yang ^{1,*} , Yeheng Deng ¹ and Anyi Li ²

¹ School of Information Engineering, Nanchang University, Nanchang 330031, China; 6105117189@email.ncu.edu.cn (W.Z.); 6104117022@email.ncu.edu.cn (Y.D.)

² College of Qianhu, Nanchang University, Nanchang 330031, China; 6101116067@email.ncu.edu.cn

* Correspondence: yangxiaohui@ncu.edu.cn; Tel.: +86-139-7094-1450

Received: 22 May 2020; Accepted: 12 June 2020; Published: 17 June 2020



Abstract: The burgeoning prognostic and health management (PHM) engineering technology with superior performance has lately received extensive attention in the academic circle. Nevertheless, the various types of faults of the power transformer often lead to less accurate predictions and the instability of the power system. To address these problems, a power transformer PHM model with a hybrid machine learning method-approach is proposed in this paper. The model uses intelligent sensors to obtain dissolved gas analysis (DGA) data for fault diagnosis of the power transformer system, so as to compress the complexity of features (gas types) in the power transformer. In particular, to enhance the robustness of the model, we adopt a modified differential evolution whale optimization algorithm (MDE-WOA) to optimize the probabilistic neural network (PNN), namely, the classification performance of the model is improved by updating the smoothing factor (σ) of PNN. In addition, compared with other optimization algorithms, the MDE-WOA algorithm has a lower complexity and more stable optimization process. Finally, we evaluate this model with real world data from the power transformer sensor in Jiangxi province, China. The results indicated that the proposed algorithm could achieve the highest diagnostic accuracy in the fourth iteration, its accuracy having reached 98.86%. Therefore, the proposed PNN parameter optimization meta heuristic algorithm could effectively enhance the accuracy and efficiency of the power transformer fault diagnosis.

Keywords: hybrid whale optimization; probabilistic neural network; machine learning; power transformer system; fault diagnosis

1. Introduction

As a major equipment of the power system in an era of increasing power demand [1–6], the power transformer is indispensable to the transmission of electric energy, the connection between the main system and each sub-system of the power grid, so it is obvious that their performance directly affects the reliable operation of the power grid [7]. With the rapid development of high-voltage and ultra-high-voltage transmission technologies, the capacity of the power grid is continuously increasing, while the coverage is persistently expanding. Consequently, if the fault of the power transformer is not timely and accurately detected, it will have a seriously negative effect on the power grid paralysis hurting the normal development of social economy [8]. Therefore, the study on fault diagnosis of power transformers is highly significant to the development of the power system.

At present, oil-immersed power transformers are used primarily in the power grid. Due to the influence of environmental factors such as electricity, machinery, and chemistry, the mineral oil and insulating cellulose paper inside the traditional power transformer will gradually undergo qualitative change, which will produce carbon monoxide (CO), carbon dioxide (CO₂), and a series

of low molecular hydrocarbons such as hydrogen (H_2), methane (CH_4), ethane (C_2H_6), and other gases. When potential faults exist in power transformers, the contents of various gases will change significantly and gradually dissolve into oil. Therefore, the composition of dissolved gas in transformer oil can reflect the operation state of a power transformer to a great extent [9]. Currently, dissolved gas analysis (DGA) has been extensively used as an effective approach in power transformer fault diagnosis [10–12]. Based on DGA data, traditional power transformer fault diagnosis methods [13], for instance characteristic gas method, three-ratio method, and an improved three-ratio method have been developed by the international electrotechnical commission (IEC). However, due to the complex gas generation mechanism of the power transformer, there is no clear corresponding relationship between gas content and ratio in oil and fault type, the traditional fault detection methods of the power transformer often rely on the diagnosis experience of experts and are difficult to be realized through the program. When the sample data tested is too little or there are abnormal values in the sample data tested, the accuracy of the test cannot meet the requirements of industrial production [14].

Recently, aiming at correcting the flaws of the traditional method of fault diagnosis for the above-mentioned power transformer, more and more scholars have tried to apply artificial intelligence (AI) methods to the power transformer fault detection model based on DGA dataset, such as fuzzy theory [15], support vector machine (SVM) [16], and artificial neural network (ANN) [17–20]. The power transformer fault diagnosis approach with artificial neural network has been widely applied. It studies the sampled data of the power transformer under different working conditions, continuously adjusting the connection weights and bias (significant parameters) of the network model, establishing the corresponding mapping relationship between specific fault characteristics and fault types and a fault diagnosis model. Therefore, the application of ANN improves the accuracy of fault diagnosis. However, it has the disadvantages of a quite small convergence area and easy to fall into local optimal. In addition, many intelligent optimization algorithms with excellent fault diagnosis effects are also widely applied in power transformer fault diagnosis, such as particle swarm optimization algorithm (PSO) [21], cuckoo search (CS) algorithm [22], etc.

For different studies in the fields of fault detection and prediction in power transformer, Zhang, Y. et al. [23] presented a novel neural network with two steps. They enhanced the accuracy of fault detection by using two artificial neural networks to detect the fault type and the condition of cellulose respectively. Dong, M. et al. [14] proposed a power transformer fault diagnosis model using SVM as hierarchical decision making. The experimental results indicate that this model can settle the problem of parameter selection of support vector classifier and has strong generalization.

Based on the strong ability of deep learning, system features can be obtained from a small amount of sample data and represent complex relationships, therefore, the research of fault diagnosis based on deep learning has attracted many scholars' attention. For example, Zhang, C. et al. [24] developed a deep learning method for fault diagnosis of rotating equipment. By setting appropriate network parameters, the extraction time of fault feature data can be saved. In addition, this method can classify faults accurately even when the number of sample data is small. On the other hand, the method has some defects, that is, the convergence speed is slow. Zhang, L. et al. [25] used deep belief network for the fault classification and identification of a vehicle transmission system. Firstly they propose calculating the spectrum of the original signal, then carrying out data fusion, and finally establishing a pattern recognition model based on deep learning. The classification results indicate that the method has a good recognition accuracy. The main advantages of this method are as follows: First, based on deep learning, it can extract features from the spectrum of sample data. Second, this method can combine the sample data of multiple sensors to extract features. Compared with the sample data of a single sensor, its data structure is more complete, which leads the classification to become more accurate. However, this method still has a shortcoming, that is, the model structure is complex, which requires the model to take a long time to train the sample data. Ji, X. et al. [26] proposed a new method for power transformer fault diagnosis utilizing deep learning and soft maximum classification. The method utilizes the superposition of the encoder and soft maximum return to the

power transformer fault detection and prediction model, using tagged without supervision and the training of a mass of samples, by the method of step k contrast differences, the parameters of the fault diagnosis model are optimized, and using the supervised algorithm to adjust the parameters of the fault diagnosis model, then the soft biggest regression method is used to determine the fault type of power transformers. Finally, through the comparative analysis, the accuracy and adaptability of this method for fault detection and prediction is superior to the methods of back-propagation neural network and SVM. In a word, deep learning has been gradually applied to the field of fault diagnosis.

It can be seen from the previous paragraphs that ANN has the advantages of a high classification accuracy and strong parallel distributed processing ability. As a branch of ANN, probabilistic neural network (PNN) not only has the advantages of ANN, but also has the advantages of easy training, fast convergence speed, and arbitrary nonlinear approximation. Thus based on the advantages of PNN, our use PNN to build the basic model of power transformer fault detection and prediction.

In view of the status quo of the power transformer fault diagnosis technology mentioned above, this paper proposes a new power transformer fault diagnosis method to improve the efficiency and accuracy of diagnosis. In addition, the other purpose of this paper is to provide a new way of thinking for the research of diagnosis methods combined with artificial intelligence technology. The main contributions of this paper are as follows: First of all, one embed modified differential evolution (MDE) operator into the whale optimization algorithm (WOA) based on life mechanism to overcome the vulnerability of WOA to drop into local optimum. Secondly, the structure parameter of PNN is optimized by using the combinatorial optimization algorithm, which leads the detection ratio of power transformer fault diagnosis to a higher level. Finally, a fault diagnosis model of power transformer is constructed, which provides a new idea for the development of fault technology.

Following introduction, Section 2 introduces the proposed method and describes the prognostic and health management (PHM) model of the power transformer based on the method; Section 3 describes the process of the experiment; Section 4 introduces and analyses the experimental results; Section 5 discusses our research work of this time; and Section 6 draws the conclusions.

2. Methods

2.1. Whale Optimization Algorithm

WOA is a swarm optimization algorithm following a kind of special hunting way of humpback whales, developed by Mirjalili et al. [27]. WOA simulates the hunting behavior of humpback whales in the natural world including the whales search, encircle attack prey, and so on the process to achieve optimization search. Lots of studies in the past showed that WOA has advantages of simple principle, easy implementation, and fewer parameter settings [28–30]. The algorithm consists of three stages: Random search for prey, encircle prey, and bubble-net attack. First, whales hunt for prey at random. In this process, groups of whale search for better prey by moving away from each other. The mathematical model of this process is described as:

$$X(t+1) = X_{rand}(t) - A \cdot |C \cdot X_{rand}(t) - X(t)|. \quad (1)$$

Here, $X_{rand}(t)$ is the position of randomly selected whale individuals in the whale population, t is the current number of iterations, and A and C are coefficient vectors. At this stage, the algorithm sets $|A| \geq 1$ to reach the location of the search agent away from the reference whale, so as to achieve the purpose of exploring a broader field.

The coefficient vectors A and C can be found as:

$$\begin{cases} A = 2a \cdot r - a \\ C = 2r \\ a = 2 - \frac{2t}{t_{max}} \end{cases} \quad (2)$$

Here, r is a random vector between 0 and 1, and t_{max} is the maximum number of iterations. From the above equation, it can be observed that there is a linear decrease from 2 to 0 with an increase of the number of iterations. Then, the whales move closer to the prey after they have found it. Namely, WOA's shrinkage enveloping mechanism, in which the mathematical formula of individual position updating of whales is as follows:

$$X(t+1) = X_{g_best} - A \cdot |C \cdot X_{g_best} - X(t)|. \quad (3)$$

It is worth noting that $|A|$ is less than 1, X_{g_best} is the current best search agent, if $X(t+1)$ has a better positional advantage, X_{g_best} will automatically update the current position to surround the prey. Finally, the whales use spiraling and narrowing enclosure to achieve prey hunting. The mathematical model to realize this process is:

$$\begin{cases} X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X_{g_best} \\ D' = |X_{g_best} - X(t)|. \end{cases} \quad (4)$$

The D' is the distance between the individual whale and prey before it updates its position. b is a constant that determines the shape of the helix, and l is a random value between 0 and 1. It is important to note that in the WOA algorithm, in order to ensure that whales encircle prey and spiral upward simultaneously, it is assumed that the probability of both is equal, i.e., $p = 0.5$.

From the above description, the WOA algorithm may be roughly described as follows:

Step 1: Set algorithm parameters, namely the total whale group size N , the maximum number of iterations t_{max} , and the dimension dim ;

Step 2: Generate the initial individuals randomly and record their current position;

Step 3: Calculate the fitness value $f(X_i)$ of each individual and preserve the current optimal solution with its features as position;

Step 4: Judge whether the updating process finished: If $t = t_{max}$, output the optimal solution and end; if $t < t_{max}$, update a , A , and C according to Equation (2);

Step 5: Generate a random number p between $[0,1]$. If $p \geq 0.5$, update the individual position according to Equation (4), and turn into Step 3. If $p < 0.5$, then determine the size of $|A|$ and 1: If $|A| \geq 1$, update the individual position via Equation (1) and turn into Step 3; if $|A| < 1$, update the individual position by Equation (3) and go to Step 3.

Note: Fitness value $f(X_i)$ refers to the objective function value calculated during iteration.

2.2. Hybrid Whale Optimization Algorithm with Modified Differential Evolution Operators

In WOA, the Equation (1) requires whale populations to separate from prey and randomly move with different individuals in the beginning of the iteration. This process makes gives WOA a good global optimization ability. However, as a general swarm intelligence optimization algorithm, WOA also has common shortcomings. Increasingly the number of iterations, the population will continue to move closer to an optimal individual region, thus losing the opportunity to explore other locations in the space, which will cause a loss of diversity in the population. According to Equation (2), a decreases linearly with the increase of the number of iterations. It is this linear reduction that leads to $|A| < 1$ and the positions of all search agents in the algorithm can only be updated by the Equations (3) and (4) in the later period of iteration, leading the algorithm to easily fall into the local optimization. Therefore, we propose a whale optimization algorithm based on the modified differential evolution (MDE) operator to address the problem of easily falling into local optimality (the pseudocode for MDE-WOA is shown in Algorithm 1).

Algorithm 1: MDE-WOA

```

1 Initialize whales population ( $i=1,2, \dots, NP$ ), scaling factor  $F$ , crossover rate  $Cr$ , lifespan  $S$ ;
2 Compute the fitness of each search factor (solution),  $X_{g\_best}$  = the best search factor,  $s = 0$ ;
3 while  $t < \text{maximum iterations}$  do
4   for each search factor do
5     Updating  $a, A, C, l$ , and  $p$  if  $p < 0.5$  then
6       if  $(|A| < 1) \& (s < S)$  then
7         Update the position of the current search factor by Equation (3);
8       else if  $(|A| \geq 1) \& (s < S)$  then
9         Select a random search agent;
10        Update the position of the current search factor by Equation (1);
11       else if  $(s = S)$  then
12         Local and global neighborhood-based mutations;
13         Generate a donor vector by Equation (8);
14         Crossover;
15         Generate a trial vector by Equation (9);
16         Greedy selection;
17         Evaluate the trial vector and update the position of the current search by
            Equation (10)
18       end
19     end
20   end
21   else if  $p \geq 0.5$  then
22     Update the position of the current search by Equation (4);
23   end
24   Check whether the solution stays within the search space ;
25   Compute the fitness of each search factor, update  $X_{g\_best}$  if there is a better solution;
26 end
27    $t = t + 1$ ; update  $s$  by Equation (5)
28 end

```

In the MDE-WOA, MDE shares a population with WOA, and the improved differential evolution operator is used as a component of WOA based on the lifetime mechanism. The use of the lifetime mechanism determines when the improved differential evolution operator is embedded in WOA. In this paper, S is taken as the life span of the individual and the current age of the individual is s . Here is the formula for updating s :

$$s = \begin{cases} s + 1 & \text{if } (\delta_{X_{g_best}} = 0) \\ s & \text{if } (\delta_{X_{g_best}} < 0). \end{cases} \quad (5)$$

Here, t is the number of iterations, $\delta_{X_{g_best}} = f(X_{g_best}, t) - f(X_{g_best}, t - 1)$. When $s = S$, it means that X_{g_best} has not been updated for s times, and then Equation (3) will be optimized by the modified differential evolution strategy.

In this paper, MDE operator is embedded into WOA by using the concept of neighborhood mutation operator. Similar to the traditional differential evolution algorithm, MDE is mainly composed of mutation, crossover, and selection. In the mutation operation, MDE combines the local model with the global model and adds a weighting factor to obtain the desired donor vector. The local donor

vector is composed of the optimal solution in the neighborhood of $X_{i,t}$ and two vectors randomly selected. This model showed as:

$$L_{i,t} = X_{i,t} + \alpha_l \cdot (X_{l_best,t} - X_{i,t}) + \beta_l \cdot (X_{p,t} - X_{q,t}) \quad (6)$$

where $X_{l_best,t}$ is the best solution in the neighborhood of $X_{i,t}$, $p, q \in [i - k, i + k] (p \neq q \neq i)$, here k is a non-zero integer number in the codomain $[1, (NP - 1)/2]$ (NP is the population size). α_l and β_l are disturbances selected randomly based on the fixed scaling factor $F, \alpha_l = \beta_l = \lambda \times rand(NP, D) + F$. The increase of the disturbance can reduce the chance of the solution dropping into the local optimum.

Similarly, the mathematical model of the global donor vector can be expressed as:

$$G_{i,t} = X_{g_best,t} + F \cdot (X_{g_best,t} - X_{i,t}) + F \cdot (X_{r_1,t} - X_{r_2,t}). \quad (7)$$

Here, $X_{g_best,t}$ is the best vector captured in the $i - th$ iteration. r_1 and r_2 are random numbers on a whole scale. The first term of the model uses the global optimal vector to replace $X_{i,t}$ to enhance the performance of convergence. Finally, the local donor vector is combined with the global donor vector to obtain the final donor vector, which can be expressed as:

$$V_{i,t} = \omega \cdot G_{i,t} + (1 - \omega) \cdot L_{i,t} \quad (8)$$

where ω is the weighting factor, which ranges from 0 to 1. To reduce the parameters and control the balance, ω is set here as the middle value of its range, namely 0.5.

After the donor vector was obtained by mutation operation, crossover operation was carried out to further boost the diversity of the population. In existing differential evolution algorithms, exponential crossover and binomial crossover are widely used. We adopt the binomial crossing approach, which is introduced as:

$$U_{i,j} = \begin{cases} V_{i,j} & (rand \leq Cr \text{ or } j = j_{rand}) \\ X_{i,j} & other. \end{cases} \quad (9)$$

Here $j_{rand} \in [1, 2...D]$ is a random dimension index, ensuring that the test vector $U_{i,j}$ has at least one element provided by the mutation vector $V_{i,j}$, and Cr controls the crossover probability.

Selection operation is to compare the experimental individuals generated by mutation and crossover operation with the target individuals, and then the better individuals are selected to enter the next generation of the population. The selection process can be described as:

$$X_{i,t+1} = \begin{cases} U_{i,t} & f(U_{i,t}) \leq f(X_{i,t}) \\ X_{i,t} & f(U_{i,t}) > f(X_{i,t}). \end{cases} \quad (10)$$

According to the equation, if the evaluation value of $U_{i,t}$ of the test individual is less than or equal to that of the corresponding target individual, then $U_{i,t}$ of the test individual will replace the corresponding target individual and enter the next generation of the population; otherwise, the individual $X_{i,t}$ will remain unchanged.

2.3. Overview of the Probabilistic Neural Network (PNN)

PNN, a feedforward neural network with the radial basis function (RBF), presented by Dr. Specht in 1989 [31]. The application of the Bayesian decision theory and RBF in PNN and the consideration of the cross effect of different pattern types give it a certain competitive strength over other neural network models. When there is an increasing amount of enormous data, PNN is capable of converging to the Bayesian classifier without falling into local minima. Additionally, PNN is popular in pattern classification and fault detection and prediction.

Different from the structure of the back propagation (BP) neural network, PNN is typically a parallel 4-layer structure, indicated in Figure 1. The function of each layer and corresponding equation are described as follows:

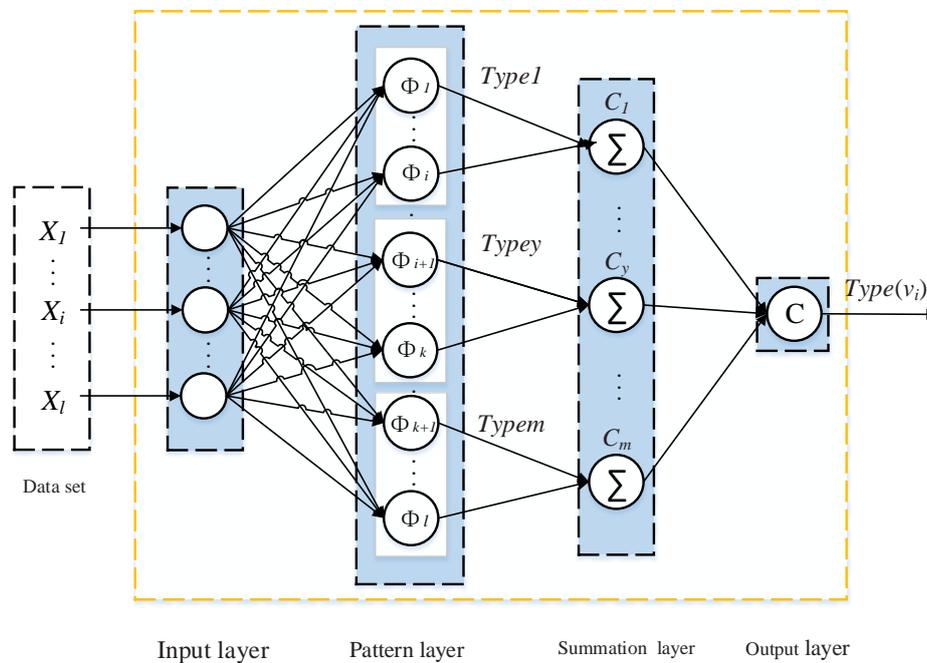


Figure 1. Probabilistic neural network (PNN) structure.

The input layer is made use of a pre-processing data set of the training sample and transmit characteristics of the sample to the network, so the number of its neurons should be the same as the dimension of all the sample.

For the pattern layer, the Euclidean distance between the feature vector of training sample X and radial center x_{ij} is used to realize the matching between the input feature vector and various types of training set. It can be expressed as follows:

$$\Phi = \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} e^{-\frac{(X-x_{ij})^T(X-x_{ij})}{2\sigma^2}}. \tag{11}$$

Here, $X = [x_1, x_2, x_3, \dots, x_n]^T, n = 1, 2, \dots, l$. l is for all types of training, d is the dimension of eigenvector, x_{ij} is the j -th center of the i -th training sample, and σ is a smoothing factor. The function of summation layer is weighted to average the output of the same type of pattern layer. It is expressed as:

$$v_i = \frac{\sum_{j=1}^L \Phi_{ij}}{L}. \tag{12}$$

Here, v_i is the output of class i neurons, and L is the number of class i neurons. The type corresponding to maximum output in the summation layer is the output type of the output layer, and its equation:

$$Type(v_i) = arg \max(v_i). \tag{13}$$

2.4. PNN Optimized by MDE-WOA Power Transformer PHM Model

For the defect of PNN, the hidden layer of calculation by the smoothing factor (σ) great influence. If the σ incorrect value is too large or too small, the network convergence falls into local optimum too quickly or easily. As an improved intelligent optimization algorithm, MDE-WOA has strong global optimization and rich population diversity. It can be extracted by selecting a suitable σ number set, to improve the performance of PNN.

In this model, the input data are as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{l1} & x_{l2} & \cdots & x_{ln} \end{bmatrix} = [x_1, x_2, \cdots, x_n]. \tag{14}$$

The flow chart of the PNN network model optimized by MDE-WOA is shown in Figure 2, and the specific steps can be summarized as follows:

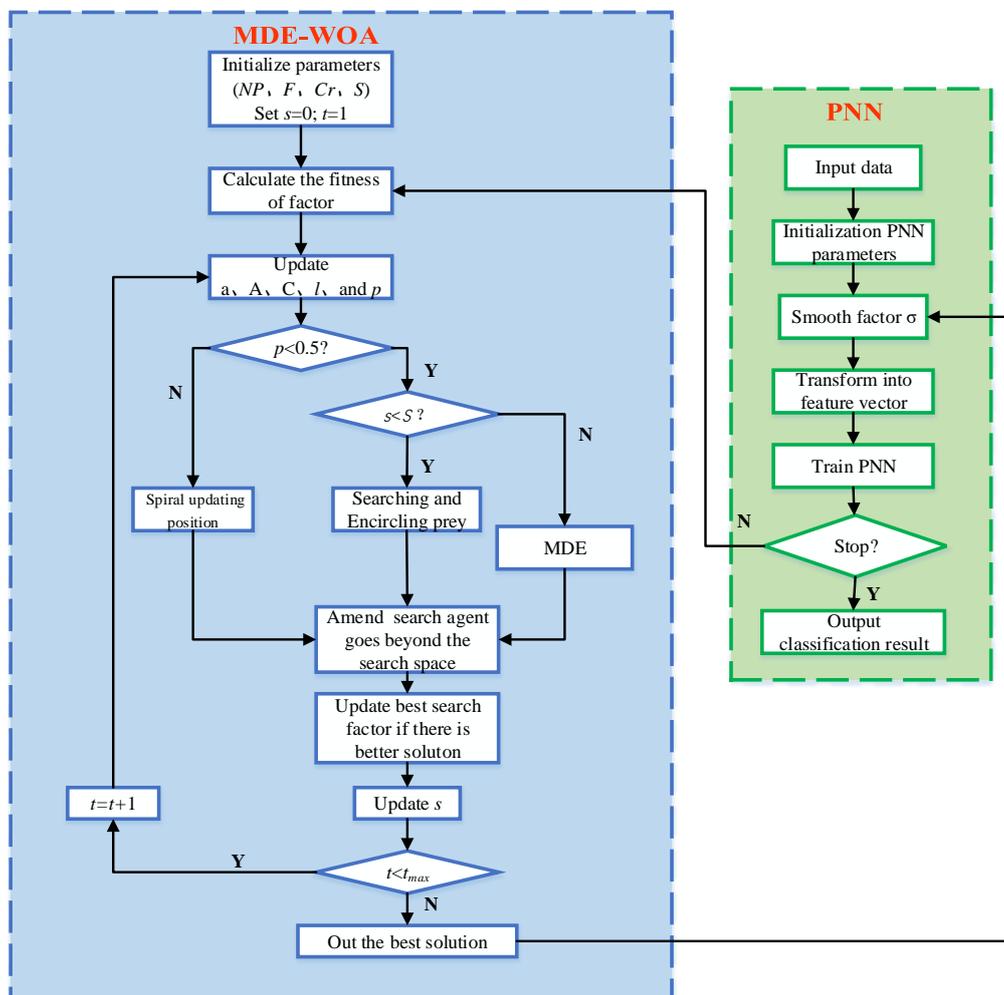


Figure 2. A modified differential evolution whale optimization algorithm (MDE-WOA) flow-chart.

Step 1: Randomly generate initialization sample X ;

Step 2: Initialize the parameters and structures of PNN and define the random smoothing factor as:

$$d = (d_1, d_2, d_3, \cdots, d_n)^T \tag{15}$$

Step 3: Set the current life $s = 0$ and the current number of iterations $t = 1$. Initialize the size (NP), proportional factor (F), cross control parameter (Cr), life span (S) of the whale population, and the fitness function $f(x)$. It is worth noting that the mean square error (MSE) is taken as the corresponding value of fitness function in our study.

$$f(x) = \frac{1}{NP} \sum_{i=1}^{NP} (Y_i - O_i)^2 \quad (16)$$

Here, Y_i is actual results and O_i is the expected result.

Step 4: Compute the fitness value of factor and record the position of the optimal individual;

Step 5: Update algorithm parameters: a, A, C, l , and p ;

Step 6: Determine the size relationship between random number p and 0.5 between $[0,1]$: If $p \geq 0.5$, the factor updates position by spiraling through Equation (4). If $p < 0.5$, the size relation between s and S is determined: If $s < S$, the current search agent searches and encircles the prey, and updates the position via Equations (1) and (3) respectively. If $s = S$, the MDE operator is introduced to optimize the search strategy;

Step 7: Compute the fitness value of the factor again, and update the best search factor if there is a better solution;

Step 8: Update the current life according to Equation (5);

Step 9: When the number of iterations t reaches the maximum number of iterations t_{max} and other parameters of the algorithm reach the preset conditions, the algorithm goes to the next step; if not, it returns to step 5;

Step 10: Optimal search agent instead of PNN in training smoothing factor σ to gain better fault diagnosis model;

Step 11: Test samples are substituted into the network to obtain the corresponding analysis data.

3. Experimental Setup

3.1. Data Collection

In terms of the effect of transformer model capacity, the environmental humidity and temperature on transformer performance, this paper collected sorts of gas data from real power transformer equipment oil from power supply companies (PSC) in the Jiangxi province, China in 2019 as experimental data samples. After screening all the data, the featured gas content data samples of 555 were obtained from the power transformer, including 65 cases of partial discharge (PD), 361 cases of low temperature overheating (LT) ($<150^\circ\text{C}$), 40 cases of low temperature overheating (LT) (150°C – 300°C), and 89 cases of arc discharge (AD). For these data samples, 400 sets of data were used as the training set in this paper, and the remaining data sets were used as our test set.

The power transformer studied is a kind of power transformer with dissolved gas data in oil for fault diagnosis. The corresponding diagram is shown in Figure 3.



Figure 3. Power transformer equipment studied in our study.

The collected 555 sets of power transformer fault data were simulated by MATLAB (R2019a), and the simulation results can be seen in Figure 4. It can be clearly seen from the three subgraphs that each fault type has its own ratio data distribution. In addition, combined with the collected fault types for data analysis, LT (150 °C–300 °C) and AD have obvious differences in the ratio data distribution compared with other fault types, however, for LT (<150 °C) and PD the distribution of three ratio data are close, which is a challenge of power transformer fault classification technology.

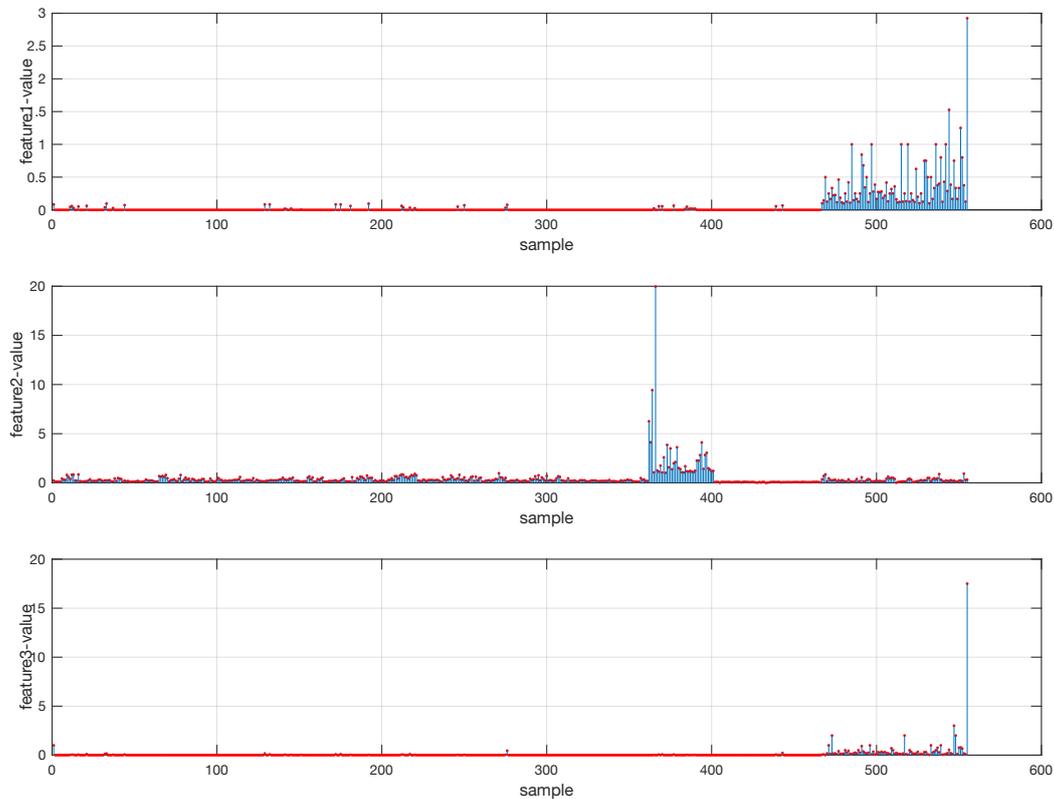


Figure 4. The distribution features of dissolved gas analysis (DGA) dataset in our study.

According to the ratio of dissolved gas content in transformer oil, the corresponding fault type can be obtained. Table 1 is part of the original data of the power transformer fault type judged by China electric power research institute with the DGA method.

Table 1. Part of the original data of power transformer fault type judged by China electric power research institute with the DGA method.

Sample Sources	Type of Fault	Dissolved Gas, $\mu\text{L/L}$				
		C_2H_6	TH	CH_4	C_2H_2	C_2H_4
Jiujiang PSC	LT (<150 °C)	1.2	150.2	83	53	13
Nanchang PSC	LT (150 °C–300 °C)	1.8	70.8	12	46	11
Fuzhou PSC	LT (150 °C–300 °C)	1.5	122	6.5	98	16
Yichun PSC	LT (150 °C–300 °C)	1	43.7	3.5	31	8.2
Yingtang PSC	AD	6	479	61	307	105

PSC means power supply company, TH is the total hydrocarbon of transformer oil and during the judge, the normal operational temperature is 25 °C and the setting humidity is 50%.

3.2. Algorithm Setting

To further assess the stability and performance of the MDE-WOA-PNN algorithm for power transformer, we adopt different methods to compare our model with BA-BP, MCS-BP [32], GA-BP, and so on. The setting of these algorithms can be seen in Table 2.

Table 2. Parameter setting in our study of various approaches.

Method Number	Approach	Parameters Settings
1	MDE-WOA	$t_{max} = 4, NP = 3, F = 0.8, Cr = 0.9, S = 110$
2	BA-BP	$A = 0.5, r = 0.5, n = 20$
3	CS-BP	$Pa = 0.25, n = 20$
4	GA-BP	$N = 20, Pm = 0.01, Px = 0.7$
5	PNN	Non

BA-BP means BP optimized by Bat algorithm (BA). CS-BP means BP optimized by Cuckoo search (CS) algorithm. GA-BP means BP optimized by Genetic algorithm (GA)

4. Experimental Results

To assess the effectiveness of our approach in the power transformer fault diagnosis, we compared the classification accuracy of the method with four methods (BA-BP, CS-BP, GA-BP, and PNN). We used MATLAB for simulation experiments. The classification accuracy calculated by the experiment is shown in Table 3.

Table 3 demonstrates that the accuracy of fault detection and prediction of MDE-WOA-PNN model was best among all of the diagnostic models, which further indicates that MDE-WOA obviously improved optimizing PNN model. With regard to four types of fault diagnosis results, for instance MDE-WOA-PNN, LT (<150 °C) was 100% (106/106), low temperature overheating (LT) 150 °C–300 °C was 100% (13/13), partial discharge (PD) was 100% (14/14), and arc discharge (AD) was 95.46% (21/22). Therefore, compared with other diagnostic models, this model is more suitable for power transformer fault detection and prediction.

Table 3. The comparison of different methods.

Ranking	Method	Successful	LT (<150 °C)	LT 150 °C–300 °C	PD	AD
		Accuracy %	Accuracy %	Accuracy %	Accuracy %	Accuracy %
1	MDE-WOA	98.86	100	100	100	95.46
2	BA-BP	96.7	99.06	92.31	100	95.45
3	CS-BP	96.31	94.34	100	100	90.91
4	GA-BP	93.3	99.06	92.31	100	81.82
5	PNN	84.58	100	100	42.86	95.45

Among them, the BA-BP and CS-BP algorithm results refer to another piece of our team's literature on similar research [33].

As another important diagnosis index of the model, MSE can directly express the error between PNN output and ideal output of the model. Therefore, to explore the superiority of our presented method, we compare MSE with the above four methods. As can be seen from Table 4 the MSE of the test set of MDE-WOA-PNN model was minimal. When using MDE-WOA to optimize PNN and the test sample was used as the input of PNN, the MSE of this test set was only 0.058. The performance was far superior to other models. Due to the existence of some noise data, the MSE performance of training samples was not very excellent. However, combined with Table 3, we know that the model still obtained a competitive diagnostic accuracy, which also proved that this model had a very high robustness from the side. In addition, to know more about the impact of the number of iterations on the MSE of the proposed model, we selected iteration times of 2, 4, 6, and 8 to conduct the exploratory experiment. The experimental results are given in Figure 5.

Table 4. A comparison mean square error (MSE) error of different algorithms.

Method Number	Approach	MSE (Training Set)	MSE (Test Set)
1	MDE-WOA	0.125	0.058
2	BA-BP	0.025	0.131
3	CS-BP	0.0075	0.1548
4	GA-BP	0.005	0.1903
5	PNN	0.4975	0.2645

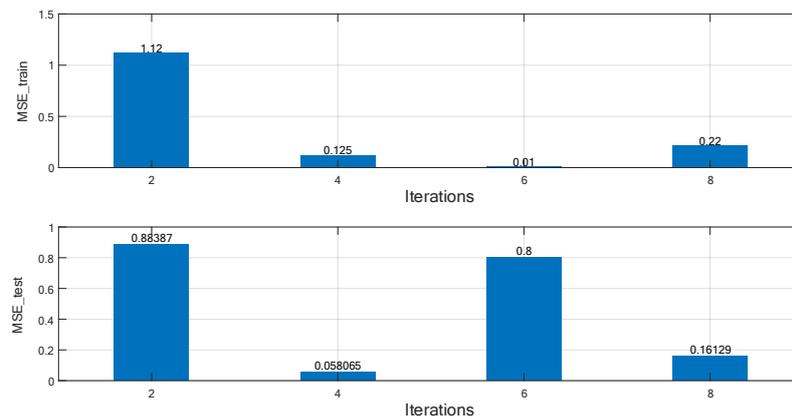


Figure 5. MSE between training sample and test sample.

To explore the impact of the number of iterations on the diagnostic accuracy of the the presented algorithm, we calculated the accuracy of the model’s fault diagnosis when the number of iterations was 2, 4, 6, and 8, respectively. The variation of fault diagnosis accuracy of various types of test samples in the network model is shown in Figure 6.

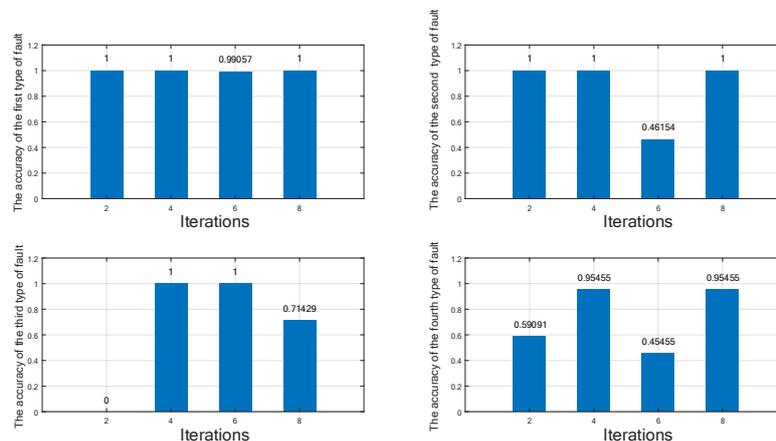


Figure 6. Diagnostic accuracy of various faults.

Figure 7 suggests the classification results of MDE-WOA-PNN training results and test sets. As can be seen from the figure, when the iteration was 4, the diagnostic accuracy of the test set of the model was highest. Therefore, it can be seen from the analysis that when the number of iterations was 2, there was an under fitting phenomenon in the network model; when the number of iterations was 6 or 8, the phenomenon of over fitting existed in the network model, which proves the efficiency of the developed approach.

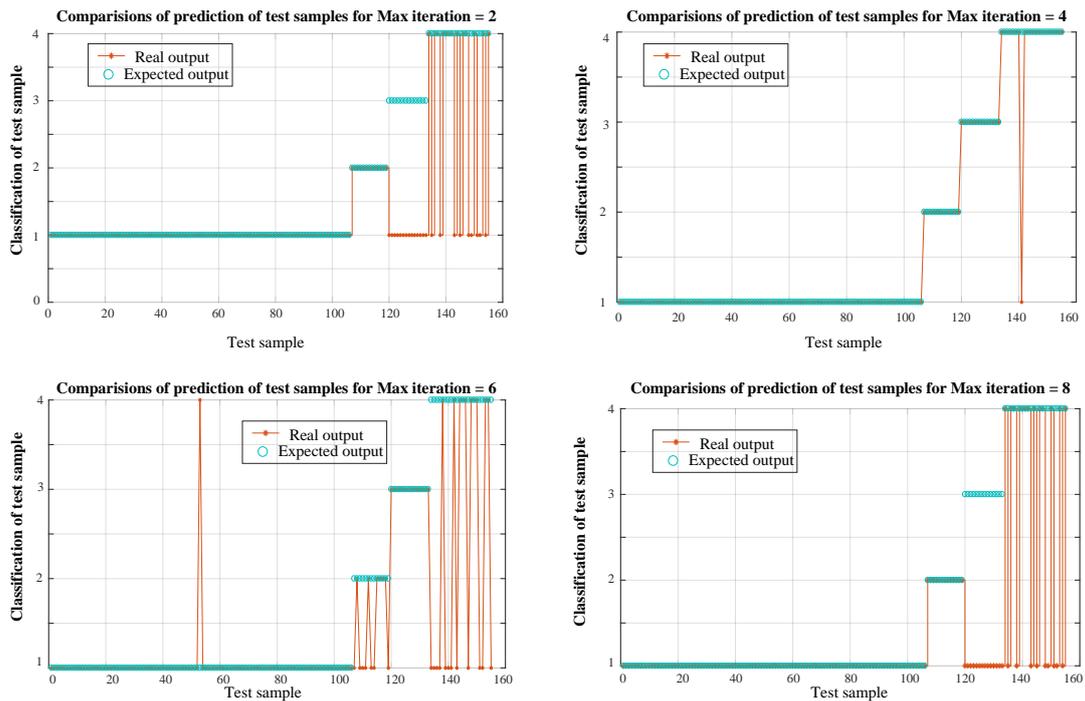


Figure 7. MDE-WOA-PNN training result classification and test set classification result.

The variation of the average accuracy with the number of iterations is given in Figure 8. According to the comprehensive analysis, MDE-WOA-PNN has a very high average accuracy for the problem studied in this paper, up to 98.86%. Therefore, MDE-WOA-PNN was quite appropriate for fault detection and prediction of power transformer.

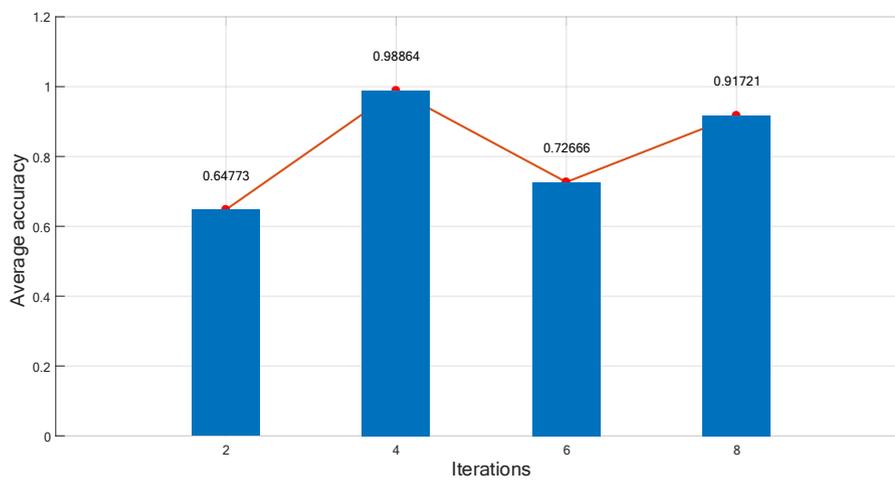


Figure 8. The average accuracy varies with the number of iterations.

The value and variance of the optimal search agent for MDE-WOA-PNN algorithm are given in Figure 9. Analysis together with Figure 8 indicates that the model had the highest diagnostic accuracy when the value of the best search agent was 0.047265.

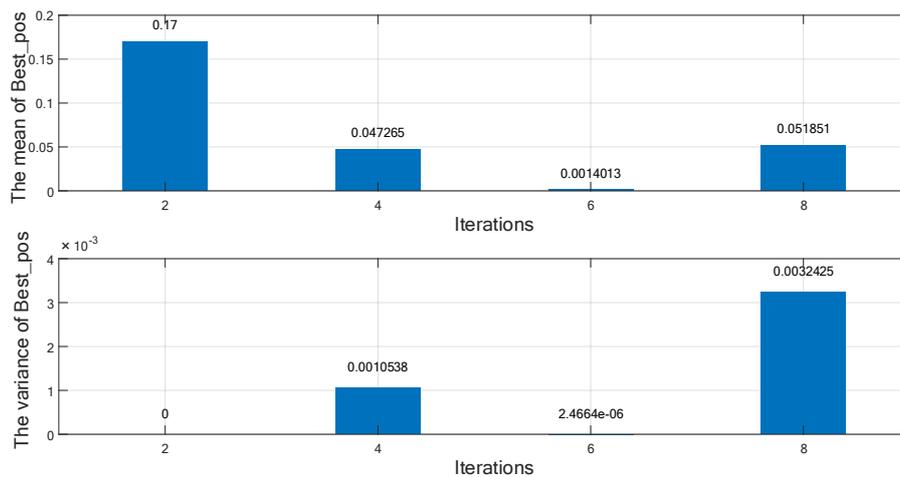


Figure 9. The value and variance of the optimal search agent for the MDE-WOA-PNN algorithm.

The fitness curve of this model shown in Figure 10 shows the convergence speed of MDE-WOA-PNN algorithm was very fast. Meanwhile, it can be seen that the proposed method could jump out of local optimality quickly, which shows the high efficiency of the algorithm. Besides, it is worth noting that the initial error of the algorithm was small, indicating that the initial value of the algorithm was close to the global optimal value.

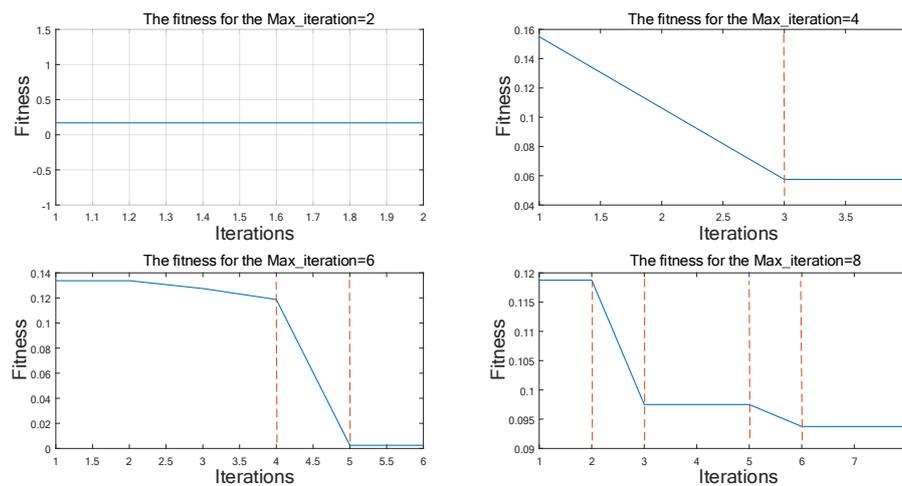


Figure 10. The adaptive curve of different iteration times.

5. Discussion

In this work, a PHM model for a power transformer was established via PNN. An intelligent optimization algorithm MDE-WOA was introduced to optimize unknown parameters, i.e., smoothing factor (σ) in the PNN model, so as to enhance the performance of the PNN algorithm. Optimized by the MDE-WOA algorithm, the global convergence of PNN network was significantly enhanced and could jump out from local optimal quickly, which made the network more efficient. Compared with other optimization algorithms, the efficiency of this model was superior. In the process of optimizing PNN by the MDE-WOA algorithm, the optimization effect was not affected when the initial parameters changed slightly, which could not be realized by many algorithms. Furthermore, this paper proved that the MDE-WOA algorithm could accelerate the speed of the convergence and efficiency of the network and then up-grading the fault prediction ratio of power transformers. Therefore, the MDE-WOA-PNN model had an obvious superiority in fault detection and prediction of the power transformer.

6. Conclusions

In the course of this study, there was no study on the influence of the number of sample data on the model diagnosis results. In addition, MDE-WOA-PNN, as a novel power transformer fault diagnosis method also had some defects, that is, the parameter setting had a greater impact on the performance of the model. Compared with other methods of similar publications, such as the power transformer fault diagnosis method combining hypersphere multiclass SVM and improved D-S evidence theory proposed by Shang, H. et al. [34] and the traditional diagnosis method of BA-BP, the MDE-WOA-PNN not only had a higher accuracy and faster convergence speed, but also improved the defect whereby traditional fault diagnosis method is easy to fall into local optimum.

For future work, our team will increase the experimental sample data to further study the impact of the number of experimental sample data on the accuracy of fault diagnosis. Meanwhile, we will improve the intelligent optimization algorithm to adapt to a more complex fault diagnosis. It is worth noting that the algorithm is also applicable to fault diagnosis of diesel engines and sensors.

Author Contributions: Methodology, W.Z. and A.L.; Conceptualization, W.Z.; Data curation, Y.D.; Software, W.Z. and A.L.; Validation, X.Y.; Writing—original draft, W.Z.; Writing—review and editing, W.Z. and X.Y. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This work was supported in part by the National Science Foundation of China (51765042, 61773051, 61963026).

Conflicts of Interest: There are no conflict from all authors of this paper.

Abbreviations

PHM	prognostic and health management
DGA	dissolved gas analysis
MDE-WOA	modified differential evolution whale optimization algorithm
PNN	probabilistic neural network
RBF	radial basis function
σ	smoothing factor
X_{g_best}	the best search agent
f	fitness function
MSE	mean square error
LT	low temperature and overheating
AD	arc discharge
PD	partial discharge
PSO	particle swarm optimization
CS	cuckoo search

References

1. Yang, H.T.; Liao, C.C. Adaptive fuzzy diagnosis system for dissolved gas analysis of power transformers. *IEEE Trans. Power Deliv.* **1999**, *14*, 1342–1350. [[CrossRef](#)]
2. Yu, S.; Zhao, D.; Chen, W.; Hou, H. Oil-immersed Power Transformer Internal Fault Diagnosis Research Based on Probabilistic Neural Network. *Procedia Comput. Sci.* **2016**, *83*, 1327–1331. [[CrossRef](#)]
3. Akbari, A.; Setayeshmehr, A.; Borsi, H.; Gockenbach, E.; Fofana, I. Intelligent agent-based system using dissolved gas analysis to detect incipient faults in power transformers. *IEEE Electr. Insul. Mag.* **2010**, *26*, 27–40. [[CrossRef](#)]
4. Duval, M. A review of faults detectable by gas-in-oil analysis in transformers. *IEEE Electr. Insul. Mag.* **2002**, *18*, 8–17. [[CrossRef](#)]
5. Transformers Committee. *IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*; Institute of Electrical & Electronics Engineers, Inc.: Piscataway, NJ, USA, 1992.
6. Wei, C.; Tang, W.; Wu, Q. Dissolved gas analysis method based on novel feature prioritisation and support vector machine. *IET Electr. Power Appl.* **2014**, *8*, 320–328. [[CrossRef](#)]

7. Yang, X.; Chen, W.; Li, A.; Yang, C.; Xie, Z.; Dong, H. BA-PNN-based methods for power transformer fault diagnosis. *Adv. Eng. Inform.* **2019**, *39*, 178–185. [[CrossRef](#)]
8. Jian, L.; Caixin, S. *Study on the Models of Fault Diagnosis about Power Transformer Based on Dissolved Gases Analysis*; Chongqing University: Chongqing, China, 2001; pp. 26–28.
9. Kari, T.; Gao, W.; Zhao, D.; Zhang, Z.; Le, 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. [[CrossRef](#)]
10. Ghoneim, S.S.; Taha, I.B. A new approach of DGA interpretation technique for transformer fault diagnosis. *Int. J. Electr. Power Energy Syst.* **2016**, *81*, 265–274. [[CrossRef](#)]
11. Kim, Y.; Lee, S.; Seo, H.; Jung, J.; Yang, H. Development of dissolved gas analysis (DGA) expert system using new diagnostic algorithm for oil-immersed transformers. In Proceedings of the 2012 IEEE International Conference on Condition Monitoring and Diagnosis, Bali, Indonesia, 23–27 September 2012; pp. 365–369.
12. Sarma, D.S.; Kalyani, G. ANN approach for condition monitoring of power transformers using DGA. In Proceedings of the 2004 IEEE Region 10 Conference TENCON 2004, Chiang Mai, Thailand, 24 November 2004; Volume 100, pp. 444–447.
13. Waghmare, H.; Kulkarni, P. Modeling of transformer DGA using IEC & fuzzy based three gas ratio method. *Int. J. Eng. Res. Technol. IJERT* **2014**, *3*, 1149–1152.
14. Dong, M.; Xu, D.; Li, M.; Yan, Z. Fault diagnosis model for power transformer based on statistical learning theory and dissolved gas analysis. In Proceedings of the Conference Record of the 2004 IEEE International Symposium on Electrical Insulation, Indianapolis, IN, USA, 19–22 September 2004; pp. 85–88.
15. Ghoneim, S.S.M. Intelligent prediction of transformer faults and severities based on dissolved gas analysis integrated with thermodynamics theory. *IET Sci. Meas. Technol.* **2018**, *12*, 388–394. [[CrossRef](#)]
16. Bacha, K.; Souahlia, S.; Gossa, M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electr. Power Syst. Res.* **2012**, *83*, 73–79. [[CrossRef](#)]
17. Han, X.H.; Xiong, X.; Duan, F. A new method for image segmentation based on BP neural network and gravitational search algorithm enhanced by cat chaotic mapping. *Appl. Intell.* **2015**, *43*, 855–873. [[CrossRef](#)]
18. Castro, A.R.G.; Miranda, V. Knowledge Discovery in Neural Networks With Application to Transformer Failure Diagnosis. *IEEE Trans. Power Syst.* **2005**, *20*, 717–724. [[CrossRef](#)]
19. Guardado, J.L.; Naredo, J.L.; Moreno, P.; Fuente, C.R. A Comparative Study of Neural Network Efficiency in Power Transformers Diagnosis Using Dissolved Gas Analysis. *IEEE Trans. Power Deliv.* **2001**, *21*, 643–647. [[CrossRef](#)]
20. Wang, X.; Wang, T.; Wang, B. Hybrid PSO-BP based probabilistic neural network for power transformer fault diagnosis. In Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application, Shanghai, China, 20–22 December 2008; Volume 1, pp. 545–549.
21. Long, Q.; Guo, S.; Li, Q.; Sun, Y.; Li, Y.; Fan, Y. Research of converter transformer fault diagnosis based on improved PSO-BP algorithm. *IOP Conf.* **2017**, *231*, 012015. [[CrossRef](#)]
22. Wang, Y.; Zhang, L. Transformer fault diagnosis based on back-propagation neural network optimized by cuckoo search algorithm. In Proceedings of the 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE), Beijing, China, 17–19 August 2017; pp. 383–386.
23. Zhang, Y.; Ding, X.; Liu, Y.; Griffin, P.J. An artificial neural network approach to transformer fault diagnosis. *IEEE Trans. Power Deliv.* **2002**, *11*, 1836–1841. [[CrossRef](#)]
24. Zhang, C.; Xu, L.; Li, X.; Wang, H. A method of fault diagnosis for rotary equipment based on deep learning. In Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 26–28 October 2018; pp. 958–962.
25. Zhang, L.; Gao, H.; Wen, J.; Li, S.; Liu, Q. A deep learning-based recognition method for degradation monitoring of ball screw with multi-sensor data fusion. *Microelectron. Reliab.* **2017**, *75*, 215–222. [[CrossRef](#)]
26. Ji, X.; Zhang, Y.; Sun, H.; Liu, J.; Zhuang, Y.; Lei, Q. Fault diagnosis for power transformer using deep learning and softmax regression. In Proceedings of the IEEE 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 2662–2667.
27. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [[CrossRef](#)]
28. Luo, J.; Shi, B. A hybrid whale optimization algorithm based on modified differential evolution for global optimization problems. *Appl. Intell.* **2019**, *49*, 1982–2000. [[CrossRef](#)]
29. Zhou, Y.; Ling, Y.; Luo, Q. Lévy Flight Trajectory-Based Whale Optimization Algorithm for Global Optimization. *IEEE Access* **2017**, *5*, 6168–6186.

30. Kumar, C.; Rao, R.S. A novel global MPP tracking of photovoltaic system based on whale optimization algorithm. *Int. J. Renew. Energy Dev.* **2016**, *5*, 225–232. [[CrossRef](#)]
31. Specht, D.F. Probabilistic neural networks. *Neural Netw.* **1990**, *3*, 109–118. [[CrossRef](#)]
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*, 4430. [[CrossRef](#)]
33. Yang, X.; Chen, W.; Li, A.; Yang, C. A Hybrid machine-learning method for oil-immersed power transformer fault diagnosis. *IEEJ Trans. Electr. Electron. Eng.* **2020**, *15*, 501–507. [[CrossRef](#)]
34. Shang, H.; Xu, J.; Zheng, Z.; Qi, B.; Zhang, L. A Novel Fault Diagnosis Method for Power Transformer Based on Dissolved Gas Analysis Using Hypersphere Multiclass Support Vector Machine and Improved D–S Evidence Theory. *Energies* **2019**, *12*, 4017. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).