

Article

Machine Learning-Based Sensor Data Modeling Methods for Power Transformer PHM

Anyi Li ^{1,2}, Xiaohui Yang ^{1,*}, Huanyu Dong ^{1,2}, Zihao Xie ¹ and Chunsheng Yang ³

¹ College of Information Engineering, Nanchang University, Nanchang 330031, China; 6101116067@email.ncu.edu.cn (A.L.); 6002115114@email.ncu.edu.cn (H.D.); 6101116073@email.ncu.edu.cn (Z.X.)

² College of Qianhu, Nanchang University, Nanchang 330031, China

³ National Research Council Canada, Ottawa, ON K1S 5B6, Canada; Chunsheng.Yang@nrc-cnrc.gc.ca

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

Received: 16 November 2018; Accepted: 13 December 2018; Published: 14 December 2018



Abstract: An emerging prognostic and health management (PHM) technology has recently attracted a great deal of attention from academies, industries, and governments. The need for higher equipment availability and lower maintenance cost is driving the development and integration of prognostic and health management systems. PHM models depend on the smart sensors and data generated from sensors. This paper proposed a machine learning-based methods for developing PHM models from sensor data to perform fault diagnostic for transformer systems in a smart grid. In particular, we apply the Cuckoo Search (CS) algorithm to optimize the Back-propagation (BP) neural network in order to build high performance fault diagnostics models. The models were developed using sensor data called dissolved gas data in oil of the power transformer. We validated the models using real sensor data collected from power transformers in China. The results demonstrate that the developed meta heuristic algorithm for optimizing the parameters of the neural network is effective and useful; and machine learning-based models significantly improved the performance and accuracy of fault diagnosis/detection for power transformer PHM.

Keywords: machine learning; effective cuckoo search; BP neural network; IEC-three ratio method; power transformer PHM; fault diagnosis

1. Introduction

Prognostic and System Health Management (PHM) generally provides capabilities such as fault detection, fault prediction, and component life tracking to assess product reliability. PHM technologies include sensing, anomaly detection, diagnosis, prediction and decision support for intelligent machinery maintenance and health operation. Taking advantage of advances in sensor technologies, PHM enables a pro-active fault prevention strategy through continuously monitoring the health of complex systems. A power transformer is a piece of equipment that is of great importance to the electronic system. Thus, its performance can have a great impact on the power grid [1–3]. Power transformer aging is an important factor leading to grid failure, which can also cause three main fault types in transformers: electrical, mechanical, and thermal failure. Among them, mechanical failure ranks first [4,5]. Therefore, it is critical to improve the accuracy of fault diagnosis of power transformers [6,7].

Some traditional methods for fault diagnosis of transformers such as dissolved gas analysis (DGA) [8–10], short circuit reactance (SCR) [11], and frequency response analysis (FRA) [12] have been widely used in industries. Nevertheless, these methods were limited by the low accuracy of fault diagnosis when the component of the dissolved gas in oil is complicated. High-dimensional fault data

on power transformers can lead to the nonlinearity of the whole system, and FRA and SCR in this condition cannot find the real locations of fault, and also cannot provide the information about the types of transformers [11].

The methods of power transformer faults' diagnosis [13] include mainly the International Electrotechnical Commission (IEC) four-ratio and the three-ratio method, characteristic gas method and so on. However, these methods generate large errors in the diagnosis of power transformers. The accuracy will be greatly reduced when the sample data are too small or there are some outliers the samples. Therefore, artificial intelligence technology with excellent performance is desired to be used in transformer fault diagnosis. Intelligent algorithms based on the DGA data are the widely-used methods in transformer fault diagnosis, especially the back-propagation (BP) neural network [14,15]. The BP neural network can be utilized to find the connection weights and bias to implement accurate diagnostic methods or models for DGA. The updated parameters of BP neural network follow the rule of gradient descending to avoid mistaking the parameters as the optimal parameters.

Nowadays, many smart optimization algorithms and machine learning algorithms have been applied to different domains such as power transformers since these methods have great fault diagnosis performance. There are plenty of power transformer fault diagnoses and other cutting-edge research. In the fault diagnosis of power transformers, various intelligent and machine learning methods are used to detect the state of transformers.

As for power transformer fault diagnosis, Khmais et al. [16] developed a fault classification method of power transformer based on support vector machine (SVM) using train data to build a multi-layer SVM classifier. This classifier has superior performance in identifying transformer fault types. Li et al. [17] presented an intelligent method for power transformer fault diagnosis based on selected gas ratio and SVM. They used a genetic algorithm (GA) to obtain the optimal dissolved gas ratio (ODGR) for DGA ratio selection and support vector machine parameter optimization. Three and four-digit coding with faulty information and fuzzy logic is used to improve the result by Hooshmand et al. [18]. The method has been applied to the diagnosis of dissolved oil in the transformer. Wang [19] developed a new transformer fault diagnosis method based on a probabilistic neural network (PNN) and dissolved gas analysis. A hybrid evolutionary algorithm based on particle swarm optimization (PSO) and BP is used to optimize the parameters of PNN. In order to solve the problem of power transformer accidents, Trappey et al. [20] developed an intelligent engineering asset management system. Data-driven models are used to detect potential faults in transformers. The Principal component analysis (PCA) and BP-Artificial Neural Network (BP-ANN) are used as prediction models to carry out this task. Zheng et al. [21] proposed a transformer solubility prediction method based on PSO and least squares support vector machine (LS-SVM). The results demonstrated that the method is superior to BPNN, Generalized Regression Neural Network (GRNN), Radial Basis Function Neural Network (RBFNN) and Support Vector Regression (SVR) methods.

With regard to another piece of equipment detected by novel methods, Zhou et al. [22] presented a method of intelligent fault diagnosis based on ontology and FMECA (Failure Mode, Effects and Critically Analysis) for the fault diagnosis of wind turbines. This method realizes the knowledge sharing between deep knowledge and shallow knowledge, improves the fault diagnosis ability and makes a better decision for the diagnosis system. In order to improve the efficiency and accuracy of transient probability analysis of flexible mechanisms, a dynamic network method (DNNM) based on Improved PSO/Bayesian regularization (BR) is proposed by Song et al. [23]. The results show that the method improves the computational efficiency and provides a meaningful insight for flexible mechanisms. In order to address the problem of rolling bearing tip under complex working conditions, it is often affected by mechanical and electrical system faults. Therefore, Lu et al. [24] proposed a deep learning method based on a convolutional neural network (CNN). Evssukoff and Gentil [25] proposed a recursive neural fuzzy system for fault detection and isolation in nuclear reactors. It generates good performance in detecting and isolating various security related faults.

This paper is extended from the DPDC 2008 conference, entitled, “Cuckoo Search Optimized NN-based Fault Diagnosis Approach for Power Transformer PHM” [26]. Based on the recommendation from the SPSC 2008 committee, we extensively rewrote the paper by extending the experiments and providing more validation results obtained from real transformer sensor data collected in a smart grid. The main contributions are as follows: (1) to develop machine learning-based models for transformer PHM, we proposed a novel method to enhance the cuckoo search algorithm for optimizing the parameters of multi-layer back-propagation neural network for fault diagnosis of a power transformer. (2) We introduce the important factors such as improvement rate (IR) to update the function of cuckoos (solutions). (3) Given the mutation of the process of finding optimal solution, we consider that the mutation of solution x , which is controlled by mutation probability P_m . (4) We evaluated the developed machine learning-based PHM models by using the real operational data collected from power transformers in a smart grid. The results demonstrated the high performance of the PHM models for transformer fault diagnosis.

The paper is organized as follows. After the Introduction section, Section 2 presents the machine learning-based method, using a Cuckoo search algorithm to optimize the BP neural network for power transformer fault diagnosis. Section 3 introduces the developed machine learning-based model for power transformer fault diagnosis; Section 4 presents the experiments and the results; Section 5 discusses the results and draws the conclusions.

2. Methods

2.1. Modified Cuckoo Search (MCS) Algorithm

Cuckoo Search Algorithm (CS) is a nature-inspired meta heuristic algorithm which imitates parasitic brood behavior of cuckoos [27]. To simulate the behavior of cuckoo nesting, the CS algorithm sets three rules. The cuckoo produces an egg each time, which represents a solution to the problem, and randomly places the eggs in a nest for hatching. In addition, the number of nests is fixed and set a value $P_a \in (0, 1)$ to describe the probability that the nest owner finds the that the egg is a foreign egg. CS is enhanced by the Levy flight so that CS can explore global space and local space of solution and combine them with local search and global search mechanisms that make itself efficient [28]. In addition, important parameters P_a and step-size α of CS algorithm in fine-tuning of solution vectors are used to adjust the convergence rate of the algorithm. However, the standard CS algorithm uses a constant value for these parameters by the experience. Unquestionable parameter setting and constant parameters during iterations will decrease the performance of CS algorithm [29].

Thus, in order to improve the ability and overcome disadvantages, a modified Cuckoo Search Algorithm (MCS) is proposed in [30], which the main task is to implement the iterative process in which parameters P_a and α are updated via function in the appropriate range.

In order to use feedback information during evolution, parameters P_a and α are set as proportional to the improvement rate (IR). In addition, the IR can be computed by

$$IR = \frac{\sum_{i=1}^{NN} NI_i}{NN}, \quad (1)$$

where

$$NI_i = \begin{cases} 1, & f(x_i^{t+1}) < f(x_i^t), \\ 0, & otherwise, \end{cases} \quad (2)$$

where NI is the number of improvement of solutions. f is the fitness function we set. NN is the total population size.

The discovery probability P_a and step size α are dynamically updated as follows:

$$P_a = P_{amin} + (P_{amax} - P_{amin}) \cdot IR^m, \quad (3)$$

$$\alpha = \alpha_{min} + (\alpha_{max} - \alpha_{min}) \cdot IR^n, \quad (4)$$

where α_{max} and α_{min} are the maximum and minimum values of step size α , respectively. P_{amax} and P_{amin} are the maximum and minimum values of discovery probability P_a , respectively. m and n are nonlinear factors for adjusting the speed of change of the control parameters.

There are two different strategies in MCS for exploration and exploitation. The first strategy uses Mantegna's algorithm [31] as follows:

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda), \quad (5)$$

where

$$L(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi\lambda/2)}{\pi} \frac{1}{s^{1+\lambda}}, \quad (s \gg s_0 > 0). \quad (6)$$

Here, L is the characteristic scale of the data set. α in Equation (5) is the step size. λ is the Levy exponent which controls the scale of distribution. s is the step size that can be computed as follows:

$$s = \frac{U}{|V|^{1/\lambda}}, \quad (7)$$

where $U \sim N(0, \sigma^2)$, $V \sim N(0, 1)$. In addition, σ^2 can be calculated as follows:

$$\sigma^2 = \left[\frac{\Gamma(1+\lambda)}{\lambda \Gamma((1+\lambda)/2)} \cdot \frac{\sin(\pi\lambda/2)}{2^{(\lambda-1)/2}} \right]^{1/\lambda}. \quad (8)$$

The second strategy is to attract the closest individuals of the current solution and conduct a global random walk.

These two strategies are randomly selected by the switching probability P_c , and the second strategy is described as follows:

$$x_i^{t+1} = x_i^t + t_1 \cdot (x_{q1}^t - x_i^t) + T \cdot (x_{q2}^t - x_{q3}^t), \quad (9)$$

where the integers $q1$, $q2$ and $q3$ represent three mutually different indices randomly selected in the range $[1, 2, \dots, NN]$, which are different from the integer i . T is scaling factor and t_1 is a random number within the interval $[0, 1]$.

In addition, to strengthen the global search capability of MCS, a mutation strategy is also introduced through mutation probability P_m . In addition, the mutation is as follows:

$$x_{ik}^{t+1} = x_{ik}^t + t_2 \cdot (x_{ik}^t - x_{jk}^t), \quad (10)$$

where t_2 is a random number in the interval $[-1, 1]$. i and j are different integers selected within the range $[1, 2, \dots, NN]$. k is an integer within the range $[1, 2, \dots, D]$, D is the solution space dimension.

In addition, the parameter P_a is to judge the probability of hosts finding exotic birds' eggs. It can determine whether to generate the next new nest. At this time, the location update equation is given by

$$x_i^{t+1} = x_i^t + r_1 \cdot (x_{q1}^t - x_{q2}^t) + r_2 \cdot (x_{best}^t - x_{q3}^t), \quad (11)$$

where the integers $q1$, $q2$ and $q3$ denote three different integers. r_1 and r_2 are randomly generated numbers in $[0, 1]$.

The initial location of MCS can be expressed as:

$$x_i = Lb + rand \times (Ub - Lb), \quad (12)$$

where the Ub and Lb are the upper and lower bounds of the search space, respectively.

The pseudo-code of modified CS algorithm is shown as Algorithm 1.

Algorithm 1: Pseudo-code of the modified CS algorithm

```

1 Begin Input: PS: the population size
2  $t = 0$ ;
3 Generate the initial population  $x_t$  by Equation (12);
4 Update objective function  $f(x_t)$  and define  $FES \leftarrow PS$ ;
5 while  $FES < Max\ FES$  do
6    $t \leftarrow t + 1$ ;
7   Set the number of improvement  $P \leftarrow 0$ ;
8   for  $i = 1$  to  $PS$  do
9     if  $rand > P_c$  then
10      Obtain the new solution  $u_{t_i}$  by Equation (5);
11     else
12      Obtain the new solution  $u_{t_i}$  by Equation (9);
13     end
14     if  $rand < P_m$  then
15      Record the mutation by Equation (10);
16     end
17     Update the function value  $f(u_{t_i})$ ;
18     if  $f(u_{t_i}) < f(x_{t_i})$  then
19       $NI \leftarrow NI + 1$ ;
20     end
21      $FES \leftarrow FES + 1$ ;
22   end
23   Compute the best rate IR by Equation (1);
24   Compute the step size  $\alpha$  by Equation (3);
25   Compute the discovery probability  $P_a$  by Equation (4);
26   for  $i = 1$  to  $PS$  do
27     Obtain the new solution  $u_{t_i}$  by Equation (11);
28     Update the function value  $f(u_{t_i})$ ;
29      $FES \leftarrow FES + 1$ ;
30   end
31 end

```

Output: Best classification and predication results

2.2. Back-Propagation (BP) Neural Network

Back-propagation (BP) neural network is a multi-layer feed-forward neural network, which belongs to an uncertain nonlinear mathematical model [32–34]. The BP network consists of an input layer, hidden layer and output layer. The two processes of forward propagation and back propagation are of great importance to the BP neural network [35,36]. The BP network can have better performers in classification and prediction because of the combination of these two processes. In the forward propagation, the data are passed through the input layer and combined with the hidden layer weights and thresholds to calculate layer by layer, and finally reach the output layer to obtain the classification result. In back propagation, when the output in the output layer does not comply with

expectations, the error signal will propagate back. It uses an error gradient descent algorithm to reduce the mean square error (MSE) between the network output value and the actual output value, and the network adjusts the weights and thresholds layer by layer from the output layer to all hidden layers. Finally, the corrected result is output to the output layer.

- Feed-forward

After recording the input value vector x , the activation a^l in the input layer l can be computed in a simple and compact vectorized form:

$$a^l = f(w^l a^{l-1} + b^l) \quad (l = 2, 3, \dots, L), \quad (13)$$

where w^l and b_j^l are the weight and the bias between the $(l - 1)$ th and the l th layer.

To set the corresponding activation, this paper uses the most popular sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (14)$$

The quadratic error criterion function of sample n is C :

$$C = \frac{1}{2n} \sum_x \|y(x) - a^L(x)\|^2. \quad (15)$$

- Back-Propagation

While reaching the layer L , the output error δ^L can be calculated by

$$\delta^L = \nabla_a C \odot \sigma'(z^L). \quad (16)$$

$\nabla_a C$ contains the rate of C changing. \odot denotes the entry-wise product of two vectors. Subsequently, the error in the next layer is

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l), \quad (17)$$

where $(w^{l+1})^T$ is the transpose for the $(l + 1)$ th layer. $\odot \sigma'(z^l)$ is the Hadamard product within the interval.

According to the error gradient descent method, the threshold can be calculated as follows:

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l. \quad (18)$$

Any weight in the network is:

$$\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l. \quad (19)$$

By combining Label (11) with Label (12), the error goes backward through the activation function in layer l .

The BP neural network model structure can be seen in Figure 1.

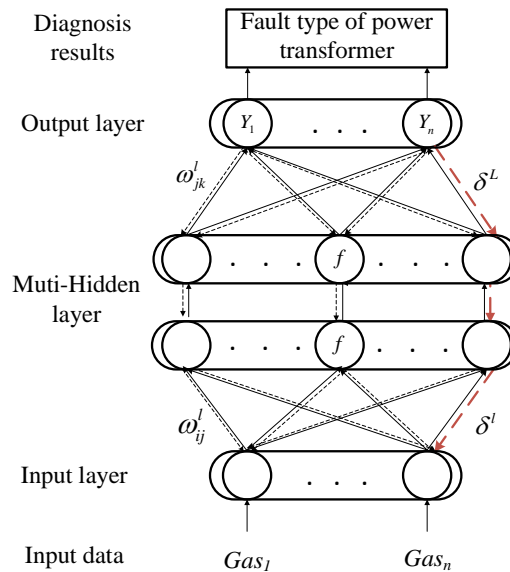


Figure 1. BP neural network model structure.

2.3. MCS Optimized BP Neural Network (MCS-BP)

The fault diagnosis of a power transformer based on an MCS optimized BP neural network can be used as a comprehensive diagnosis platform, which combines the data of gas in oil with the detection system, and then obtains good results by supervised learning methods.

As shown in Figure 2, MCS optimizes the block diagram of BP neural network. The following are the main steps:

Step 1: At first, use an IEC three-ratio method to process the features of DGA data of power transformer.

Step 2: Randomly choose the different types of faults of power transformers into the neural network.

Step 3: Initialize the parameters of the BP neural network.

Step 4: Initialize the modified cuckoo search size x_i ($i = 1, 2, \dots, N$), population size N , switching probability P_c , mutation probability P_m , and value of step size α_0 , maximum value of step size α_{max} , minimum value of step size α_{min} , maximum value of discovery probability P_{amax} , minimum value of discovery probability P_{amin} , nonlinear factor m and n , scaling factor F and the fitness function $f(x)$. The fitness function we used in this paper is the mean square error (MSE), as follows:

$$f(x) = \frac{1}{2n} \sum_{i=1}^N (Y_i - O_i)^2, \quad (20)$$

where Y_i is the measure value and O_i is the predicted result.

Step 5: Calculate the fitness value of the initial nest via the fitness function, and then select the current optimal solution in the solution space.

Step 6: Generate a random number $K1$ and compare with P_c . Compare $K1$ and P_c , if $K1 > P_c$, update nests x_i^{t+1} via Equation (5), otherwise by Equation (9).

Step 7: Generate a random number $K2$ and compare with P_m of MCS. If $K2 < P_m$, perform the mutation via Equation (10); otherwise, it is unchanged.

Step 8: Calculate the updated solution's fitness value and update the discovery probability P_a and the step size α via Equation (3) and Equation (4).

Step 9: Generate a random number $K3$ and compare with P_a . If $K3 > P_a$, update nests x_i^{t+1} via Equation (11), or do not change. Compare the last fitness values with new birds' nests, keep the optimal bird's nest as the contemporary best nest x_b .

Step 10: If it can reach the maximum iteration condition, proceed to the next step, or return to Step 6.

Step 11: Substitute the optimized weights and bias of the BP neural network.

Step 12: Input the test set into the trained BP neural network to get the classification output.

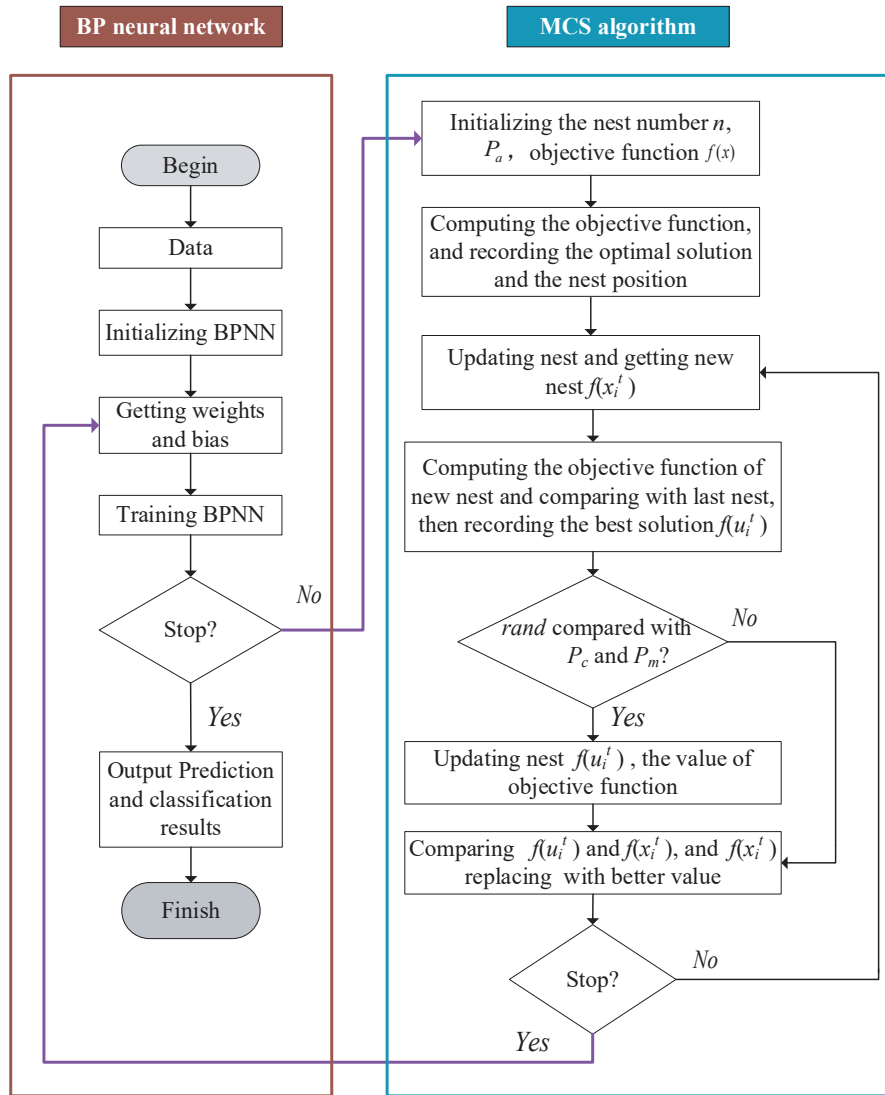


Figure 2. The flowchart of MCS-BP.

3. MCS-BP for Power Transformer Fault Diagnosis Platform

In this paper, power transformer fault diagnosis is mainly divided into four parts: data collection and preprocessing, segmentation of data set, neural network model-training, and comparison between test set output and train set output, as is shown as Figure 3.

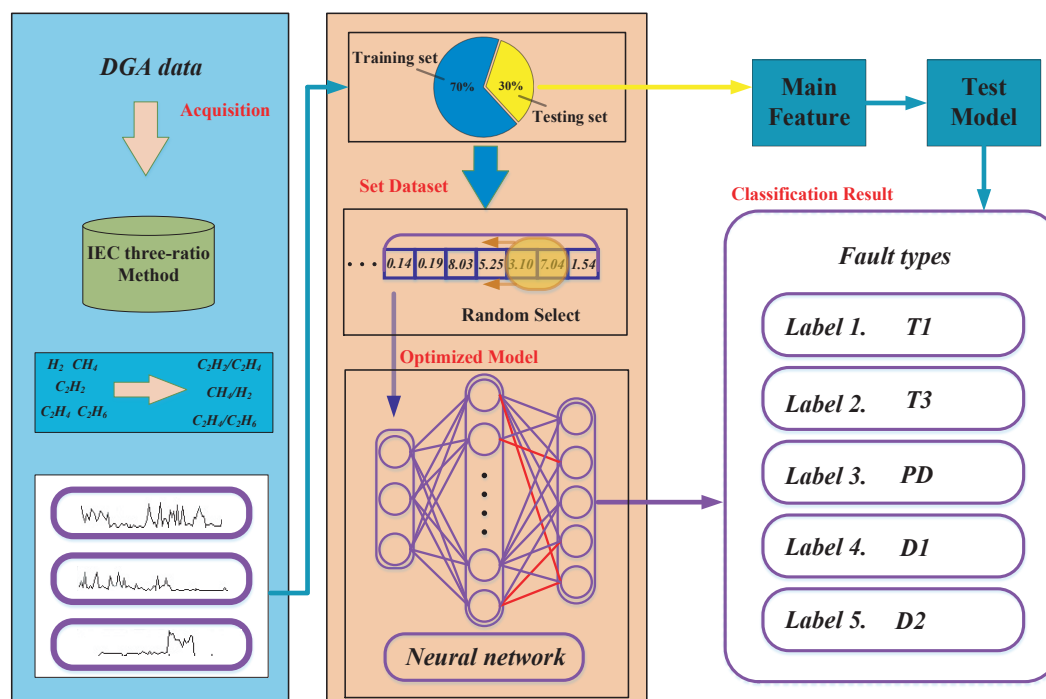


Figure 3. Structure flow of the power transformer fault diagnosis process.

In Figure 3, firstly power transformer DGA data will be processed in feature selection via the IEC three-ratio method. This procession can be seen in Table 1. Then, 70% of the data can be used in the training model, which has been sorted randomly to ensure the train set and test set containing all types of faults. The other 30% of the data is utilized to test the optimized model. In this study, we test five types of faults of power transformers, which are the thermal faults $T > 700$ °C, thermal faults $T < 300$ °C, high energy discharge, low energy discharge and partial discharge. It can be seen as Table 2, and each group of data is balanced. There are 109 sets of data.

Table 1. Diagnosis using the three-ratio method (IEC 60599) [37].

Fault Type	C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6
PD	<0.1	<0.1	<0.2
D1	>1	0.1–0.5	>1
D2	0.6–2.5	0.1–1	>2
T1	NS	>1/NS	<1
T2	<0.1	>1	1–4
T3	<0.2	>1	>4

Table 2. Fault type used in analysis.

NO.	Fault Type	Fault Type Code
Fault 1	Thermal faults $T > 700$ °C	T3
Fault 2	Thermal faults $T < 300$ °C	T1
Fault 3	High energy discharge	D2
Fault 4	Low energy discharge	D1
Fault 5	Partial discharge	PD

Through this optimization model, the potential faults of power transformers can be predicted and classified.

4. Experimental Setup and Results

4.1. Experimental Setup

In order to evaluate the performance of the proposed method in power transformer fault diagnosis, we obtain real-world data to implement experiments. By IEC three-ratio processing 109 sets of DGA data, the feature-filtered data sets are obtained, some of which are shown in Table 3. In this paper, the neural network is used as the basic classifier, so the encoding method of data is shown in Table 4.

Table 3. Statistical data of partial samples.

C_2H_4	H_2	C_2H_6	Fault Type
0.019	0.0899	2.157	T1
0.029	0.231	2.654	T1
0.0246	0.9655	8.2797	T3
0.0541	1.2551	8.9697	T3
1.38	0.211	5.396	D2
0.12	0.438	5.664	D2
8.097	2.694	1.752	D1
8.382	2.708	1.768	D1
0	0.041	0.149	PD
0.088	0.052	0.099	PD

The MCS-BP method is compared with other excellent predictive classifiers BP, CS-BP, Multi-Verse Optimizer-Multi-Layer Perceptron (MVO-MLP), PSO-BP, GA-BP, PNN and SVM, respectively. Firstly, the accuracy and error rate of the algorithm of MCS-BP is compared with CS-BP and BP, and the superiority of the algorithm of MCS in the optimization of a neural network is proved. Secondly, by comparing the MCS-BP algorithm with other machine learning algorithms and optimization algorithms, it is proved that the algorithm has strong robustness and classification performance.

Table 4. Output target coding of different faults.

	T3	T1	D2	D1	PD
Coding format	1	0	0	0	0
	0	1	0	0	0
	0	0	1	0	0
	0	0	0	1	0
	0	0	0	0	1

4.2. Experimental Results

Firstly, the MCS-BP algorithm and CS-BP algorithms are compared. In this experiment, the parameters of the BP neural network are the same, its structure is 3-4-5, three-dimensional input, five-dimensional output, for MCS and CS, the MCS algorithm sets the Levy exponent λ to 1.5, the step size α_0 to 0.1, α_{max} to 0.5, α_{min} to 0.05, the discovery probability P_{amax} to 0.5, P_{amin} to 0.1, P_c to 0.3, P_m to 0.3, m to 0.5, and n to 0.5. The CS algorithm sets the discovery probability P_a to 0.25, λ to 1, and the step size α to 0.4. The accuracy of classification is shown in Table 5. For five types of fault outputs, MCS-BP has a high recognition rate of 97.14% with asterisk, and the classification recognition rate of different fault types is higher than or equal to the BP and CS-BP algorithms.

As shown in Table 6, the training set and test set MSE of MCS-BP are the smallest. Compared with the other two algorithms, the performance of MCS is better than that of normal BP and standard CS.

Table 5. The comparison of basic methods (* means the best result in the table).

Fault Type	Accuracy Rate (%)		
	BP	CS-BP	MCS-BP
T3	100.00	100.00	100.00
T1	100.00	85.71	100.00
D2	85.71	85.71	85.71
D1	100.00	100.00	100.00
PD	0.00	100.00	100.00
Total	77.14	94.29	97.14 *

Table 6. Comparison of sample errors.

Model	MSE of Train Sample	MSE of Test Sample
BP	0.0330	0.1571
CS-BP	0.0053	0.0220
MCS-BP	0.0058	0.0204

Next, we use many different algorithms to further evaluate the classification performance of MCS-BP. The GA sets the crossover probability to 0.7 and the mutation probability to be 0.01. The PSO algorithm sets the maximum speed to 1, the minimum speed to -1 , the solution space is $[-5, 5]$ and the learning factor to 1.49445. The recognition rate is shown in Table 7. From Table 7, it can be found that the recognition rate and Micro F1-score of MCS-BP is obviously higher than that of other algorithms, and for fault type T1 and PD, the classification recognition rate of other algorithms is very low, and the diagnosis rate of MCS-BP is still very high. The error rate comparison is shown in Table 6, from which we can see that the MSE of MCS-BP is still the smallest, indicating the excellent performance of MCS-BP. From Table 8, we further know that the performance of the classification of MCS-BP is better than other models via each fault type F1-score and the Macro F1-score (98.46%).

In order to further evaluate the performance of MCS-BP, we plot the output results of different algorithms, such as Figure 4, Figure 4a,c,e,g for the output of training set and the layout of predicted classification results, and Figure 4b,d,f,h for the output of test set classification. Its test set classification effect is not ideal; however, the MCS-BP (a), (b) shows a good recognition rate, and (b) shows that only the third fault type judgment is an error.

Table 7. The comparison of different methods (* means the best result in the table).

Fault Type	Accuracy Rate (%)					
	MCS-BP	MVO-MLP	PSO-BP	GA-BP	PNN	SVM
T3	100.00	100.00	100.00	100.00	83.33	83.33
T1	100.00	71.43	85.71	57.14	85.71	28.57
D2	85.71	100.00	85.71	100.00	100.00	71.43
D1	100.00	100.00	100.00	100.00	66.67	100.00
PD	100.00	85.71	85.71	85.71	85.71	85.71
Total	97.14 *	91.43	91.43	88.57	78.57	73.81

Table 8. The comparison of different methods with F1-score (* means the best result in the table).

Fault Type	Macro F1-Score (%)					
	MCS-BP	MVO-MLP	PSO-BP	GA-BP	PNN	SVM
T3	100.00	100.00	100.00	100.00	100.00	90.91
T1	100.00	92.31	60.00	92.30	44.44	44.44
D2	92.30	100.00	100.00	92.30	100.00	83.33
D1	100.00	100.00	100.00	100.00	100.00	100.00
PD	100.00	92.30	92.30	92.30	92.30	92.31
Macro F1-score	98.46 *	96.92	90.46	95.38	87.35	82.20

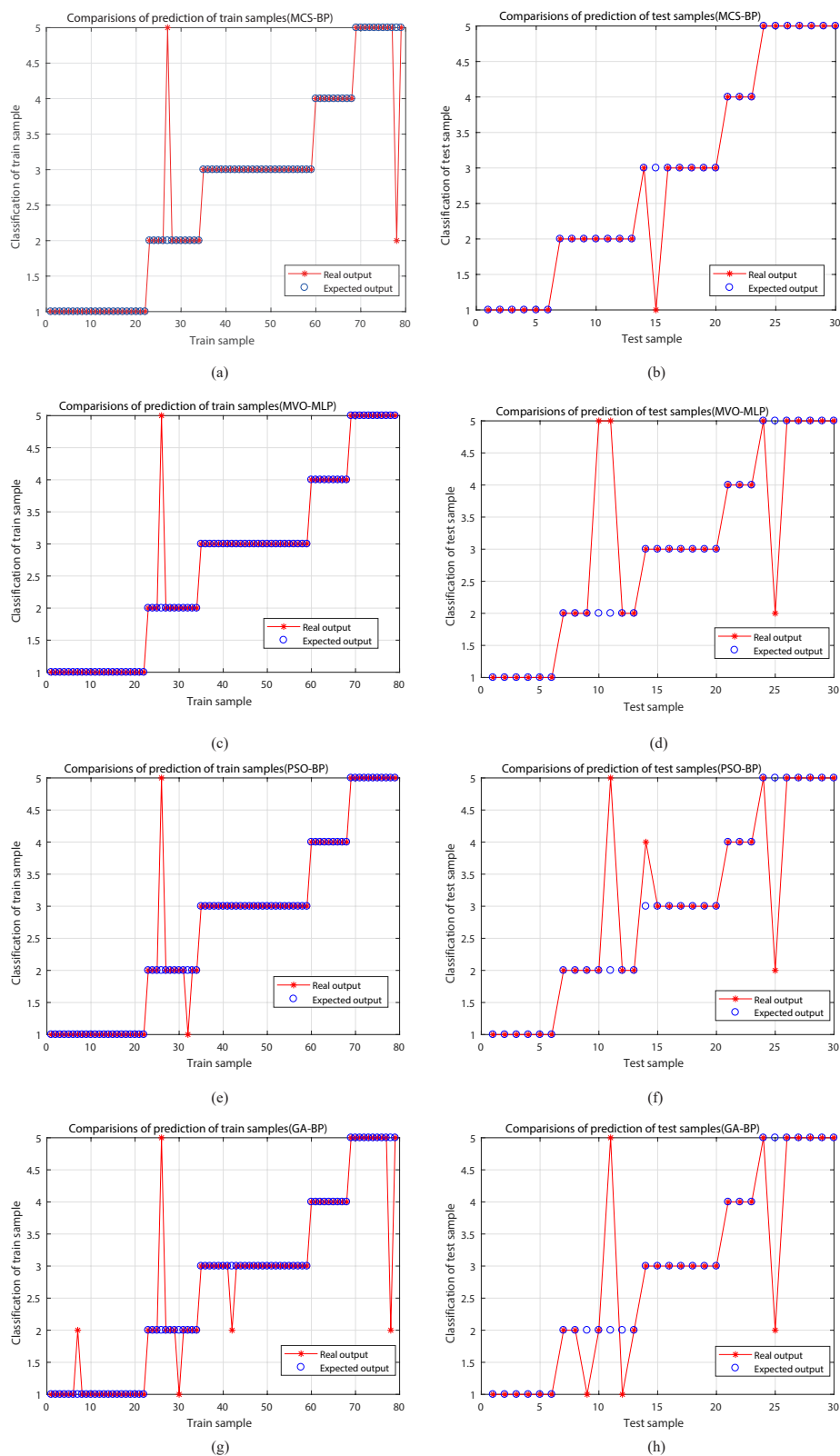


Figure 4. The classification results of different models. (a), (c), (e) and (g) represent the results of train sample classification for different methods, respectively. (b), (d), (f) and (h) are the results of test sample classification for different methods, respectively.

By Figure 4, we know that the developed method is better than other algorithms with respect to stability as the test data results can reflect that our model won't fall into the problem of over-fitting,

and both the train sample and test samples have great classification results. It indicates that this model is helpful for fault diagnosis of power transformer because it can give a suitable decision for which the fault type of power transformer is contained.

In addition, to clarify the optimization effect of MCS on BP during the iteration process, we consider the variation of the fitness (error) of the MCS, as shown in Figure 5. From generation 1 to generation 50, the fitness value of MCS in the process of optimizing neural network decreases rapidly in generations 1 to 4, which shows that the efficiency of global optimization and local optimization of MCS is very high. Then, it falls into a local extremum in generation 18 or so. However, after a period of iteration, it jumps out of a local extremum point and continues to search for optimization. In addition, we completed the Receiver Operating Characteristic (ROC) Curve, which is shown in Figure 6, and obtain the value area under the curve (AUC). Both of them can prove that our model has great performance of classification for power transformers. For Figure 6a, we can see that each fault class can be classified well.

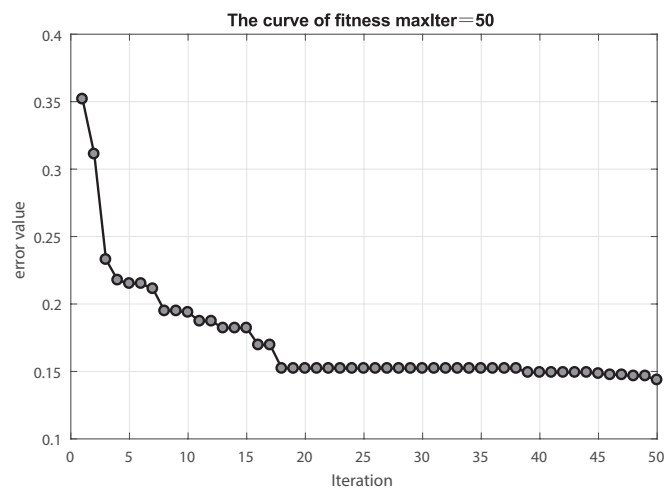


Figure 5. The curve of fitness of MCS-BP.

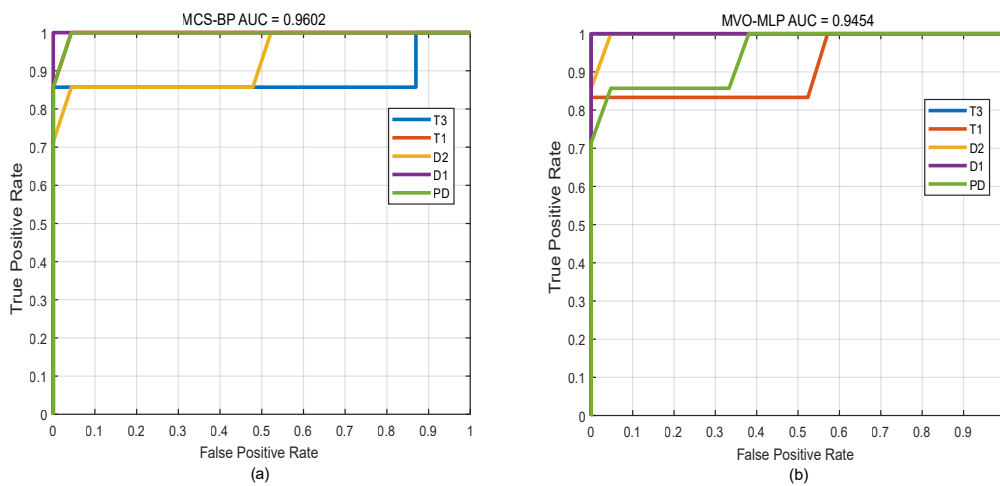


Figure 6. Cont.

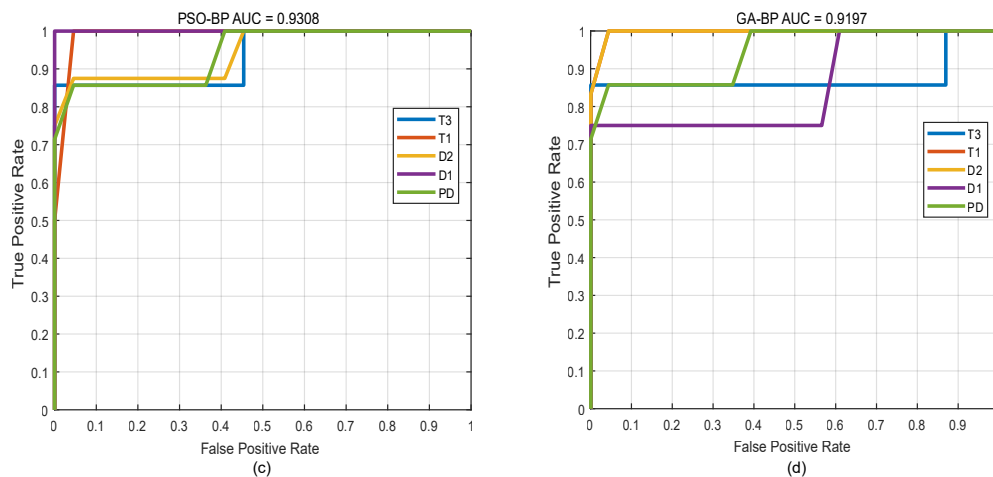


Figure 6. The curve of ROC-AUC of different models. (a–d) represent the receiver operating characteristic curve with AUC for MCS-BP, MVO-MLP, PSO-BP and GA-BP, respectively.

5. Conclusions

In this paper, we propose a machine learning-based method, CS optimized BP neural network model for power transformer fault diagnosis. This algorithm can adjust the search step of solution space adaptively to find a better global optimal solution, and the fitness value of each solution is utilized to build the mutation probability to avoid local convergence. In addition, the MCS enhances the exploitation capacity and convergence rate. We conducted the experiments to validate the developed models by using 109 sets of real-world data collected from power transformers. Compared with other algorithms, experimental results show that the MCS method we developed outperformed other methods and can converge to the optimal solution for most test cases.

To validate the machine learning-based models or methods for fault diagnosis, more extensive experiments and more advanced metrics and evaluation tools are in high demand. This will be our future work. We will continue to enhance the performance of the algorithms and models and evaluate the performance of the models under different circumstances of the error rate and operating efficiency, using other evaluation tools and metrics.

Author Contributions: A.L. designed the whole method for research. A.L., Z.X. and H.D. wrote the draft. A.L. and Z.X. contributed to the experiment. C.Y. gave a detailed revision. X.Y. provided important guidance for this paper. All authors have read and approved the final manuscript.

Funding: This work was supported in part by the National Science Foundation of China (51765042, 61463031, 61662044, 61773051), the Jiangxi Provincial Department of Science and Technology JXYJG-2017-02.

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

Abbreviations

PHM	Prognostic and health management
MCS	Modified Cuckoo Search Algorithm
BP	Back-propagation
MCS-BP	Modified Cuckoo Search Algorithm optimized Back-propagation neural network
DGA	Dissolved gas analysis
GA	Genetic algorithm
PCA	Principal component analysis
PSO	Particle Swarm Optimization
IR	improvement rate
α	step-size
P_a	discovery probability

f	fitness function
λ	Lévy exponent
P_c	switching probability
P_m	mutation probability
a^l	activation functions
C	cost function
σ^L	output error
MVO	multi-verse optimizer
MLP	multi-layer perceptron
PNN	probability neural network
MSE	mean square error
SVM	support vector machine
x_i^t	the nest in t generation
x_i^{t+1}	the nest in $t + 1$ generation

References

1. Wang, T.; He, Y.; Li, B.; Shi, T. Transformer Fault Diagnosis Using Self-powered RFID Sensor and Deep Learning Approach. *IEEE Sens. J.* **2018**. [[CrossRef](#)]
2. Tang, S.; Hale, C.; Thaker, H. Reliability modeling of power transformers with maintenance outage. *Syst. Sci. Control Eng. Open Access J.* **2014**, *2*, 316–324. [[CrossRef](#)]
3. Zeng, W.; Yang, Y.; Gan, C.; Li, H.; Liu, G. Study on Intelligent Development of Power Transformer On-Line Monitoring Based on the Data of DGA. In Proceedings of the Power and Energy Engineering Conference (APPEEC), Wuhan, China, 25–28 March 2011; pp. 1–4.
4. Abu-Elanien, A.E.B.; Salama, M.M.A.; Ibrahim, M. Calculation of a Health Index for Oil-Immersed Transformers Rated Under 69 kV Using Fuzzy Logic. *IEEE Trans. Power Deliv.* **2012**, *27*, 2029–2036. [[CrossRef](#)]
5. Mauntz, M.; Peuser, J. Continuous condition monitoring of high voltage transformers by direct sensor monitoring of oil aging for a stable power network. In Proceedings of the IEEE Conference on Diagnostics in Electrical Engineering (Diagnostika), Pilsen, Czech Republic, 6–8 September 2016; pp. 1–4.
6. Bakshi, A.; Kulkarni, S.V. Eigenvalue Analysis for Investigation of Tilting of Transformer Winding Conductors Under Axial Short-Circuit Forces. *IEEE Trans. Power Deliv.* **2011**, *26*, 2505–2512. [[CrossRef](#)]
7. Rybel, T.D.; Singh, A.; Vandermaar, J.A.; Wang, M.; Marti, J.R.; Srivastava, K.D. Apparatus for Online Power Transformer Winding Monitoring Using Bushing Tap Injection. *IEEE Trans. Power Deliv.* **2009**, *24*, 996–1003. [[CrossRef](#)]
8. Khan, S.A.; Eqbal, M.D.; Islam, T. A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models. *IEEE Trans. Dielectr. Electr. Insul.* **2015**, *22*, 590–596. [[CrossRef](#)]
9. Zhou, Q.; Wang, S.; An, W.; Sun, C.; Xie, H.; Rao, J. Power transformer fault diagnosis based on DGA combined with cloud model. In Proceedings of the 2014 International Conference on High Voltage Engineering and Application (ICHVE), Poznan, Poland, 8–11 September 2014; pp. 1–4.
10. 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, 21–24 November 2004; Volume 100, pp. 444–447.
11. Palani, A.; Santhi, S.; Gopalakrishna, S.; Jayashankar, V. Real-time techniques to measure winding displacement in transformers during short-circuit tests. *IEEE Trans. Power Deliv.* **2008**, *23*, 726–732. [[CrossRef](#)]
12. Ahmed, I. Use of Frequency Response Analysis to Detect Transformer Winding Movement. Ph.D. Thesis, Murdoch University, Hong Kong, China, 2013.
13. Waghmare, M.H.V. Modeling of Transformer DGA using IEC & Fuzzy Based Three Gas Ratio Method. *Int. J. Eng. Res. Technol.* **2014**, *3*, 1149–1152.
14. Shen, M.K.; Huang, Z.Y.; Wang, Z.H.; Zhou, J.H. Prediction of coal ash deformation temperature based on Cuckoo Search and BP Neural Network. *J. Fuel Chem. Technol.* **2014**, *52*, 89–90.

15. Pratimsarangi, P.; Sahu, A.; Panda, M. A Hybrid Differential Evolution and Back-Propagation Algorithm for Feedforward Neural Network Training. *Int. J. Comput. Appl.* **2013**, *84*, 1–9. [[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. 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. [[CrossRef](#)]
18. Wagh, N.; Deshpande, D.M. Investigations on incipient fault diagnosis of power transformer using neural networks and adaptive neurofuzzy inference system. *Appl. Comput. Intell. Soft Comput.* **2014**, *2014*, 135–143. [[CrossRef](#)]
19. Wang, X.; Wang, T.; Wang, B. Hybrid PSO-BP Based Probabilistic Neural Network for Power Transformer Fault Diagnosis. In Proceedings of the International Symposium on Intelligent Information Technology Application, Nanchang, China, 21–22 November 2009; pp. 545–549.
20. Trappey, A.J.C.; Trappey, C.V.; Ma, L.; Chang, J.C.M. Intelligent engineering asset management system for power transformer maintenance decision supports under various operating conditions. *Comput. Ind. Eng.* **2015**, *84*, 3–11. [[CrossRef](#)]
21. Zheng, H.; Zhang, Y.; Liu, J.; Wei, H.; Zhao, J.; Liao, R. A novel model based on wavelet LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers. *Electr. Power Syst. Res.* **2018**, *155*, 196–205. [[CrossRef](#)]
22. Zhou, A.; Yu, D.; Zhang, W. A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA. *Adv. Eng. Inform.* **2015**, *29*, 115–125. [[CrossRef](#)]
23. Song, L.K.; Fei, C.W.; Bai, G.C.; Yu, L.C. Dynamic neural network method-based improved PSO and BR algorithms for transient probabilistic analysis of flexible mechanism. *Adv. Eng. Inform.* **2017**, *33*, 144–153. [[CrossRef](#)]
24. Lu, C.; Wang, Z.; Zhou, B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Adv. Eng. Inform.* **2017**, *32*, 139–151. [[CrossRef](#)]
25. Evsukoff, A.; Schirru, R. Neuro-fuzzy systems for fault detection and isolation in nuclear reactors. *Adv. Eng. Inform.* **2005**, *19*, 55–66. [[CrossRef](#)]
26. Yang, X.; Li, A.; Dong, H.; Yang, C. Cuckoo Search Optimized NN-based Fault Diagnosis Approach for Power Transformer PHM. In Proceedings of The International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC 2018), Xi'an, China, 15–17 August 2018.
27. Yang, X.S.; Deb, S. *Eagle Strategy Using Levy Walk and Firefly Algorithms for Stochastic Optimization*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 101–111.
28. Yang, X.S.; Deb, S. Cuckoo search: Recent advances and applications. *Neural Comput. Appl.* **2014**, *24*, 169–174. [[CrossRef](#)]
29. Valian, E.; Mohanna, S.; Tavakoli, S. Improved cuckoo search algorithm for feedforward neural network training. *Int. J. Artif. Intell. Appl.* **2011**, *2*, 36–43.
30. Cheng, J.; Wang, L.; Xiong, Y. Modified cuckoo search algorithm and the prediction of flashover voltage of insulators. *Neural Comput. Appl.* **2018**, *30*, 355–370. [[CrossRef](#)]
31. Yang, X.S. *Metaheuristic Optimization: Nature-Inspired Algorithms and Applications*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 405–420.
32. Han, X.; Xiong, X.; Fu, D. *A New Method for Image Segmentation Based on BP Neural Network and Gravitational Search Algorithm Enhanced by Cat Chaotic Mapping*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2015; pp. 855–873.
33. Liu, Y.K.; Xie, F.; Xie, C.L.; Peng, M.J.; Wu, G.H.; Xia, H. Prediction of time series of NPP operating parameters using dynamic model based on BP neural network. *Ann. Nucl. Energy* **2015**, *85*, 566–575. [[CrossRef](#)]
34. Liu, N.; Yang, H.; Li, H.; Yan, S.; Zhang, H.; Tang, W. BP artificial neural network modeling for accurate radius prediction and application in incremental in-plane bending. *Int. J. Adv. Manuf. Technol.* **2015**, *80*, 971–984. [[CrossRef](#)]
35. Wang, H.; Kong, C.; Li, D.; Qin, N.; Fan, H.; Hong, H.; Luo, Y. Modeling Quality Changes in Brined Bream (*Megalobrama amblycephala*) Fillets During Storage: Comparison of the Arrhenius Model, BP, and RBF Neural Network. *Food Bioprocess Technol.* **2015**, *8*, 2429–2443. [[CrossRef](#)]

36. Wang, Q.; Li, Y.; Diao, M.; Gao, W.; Qi, Z. Performance enhancement of INS/CNS integration navigation system based on particle swarm optimization back propagation neural network. *Ocean Eng.* **2015**, *108*, 33–45. [[CrossRef](#)]
37. Duval, M.; Depabla, A. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. *IEEE Electr. Insul. Mag.* **2002**, *17*, 31–41. [[CrossRef](#)]



© 2018 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/>).