

Research Article

A Power Transformer Fault Diagnosis Method Based on Random Vector Functional-Link Neural Network

Qian Wang ¹, Shinan Wang ¹, Rong Shi,² and Yong Li³

¹School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710054, China

²State Grid Shaanxi Electric Power Company Economic Research Institute, Xi'an, Shaanxi 710065, China

³Trinity International Ltd., Chaoyang, Beijing 100022, China

Correspondence should be addressed to Qian Wang; qianqian82@126.com

Received 12 October 2020; Revised 29 December 2020; Accepted 10 March 2021; Published 23 March 2021

Academic Editor: Haoran Zhang

Copyright © 2021 Qian Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The random vector functional link (RVFL) network is suitable for solving nonlinear problems from transformer fault symptoms and different fault types due to its simple structure and strong generalization ability. However, the RVFL network has a disadvantage in that the network structure, and parameters are basically determined by experiences. In this paper, we proposed a method to improve the RVFL neural network algorithm by introducing the concept of hidden node sensitivity, classify each hidden layer node, and remove nodes with low sensitivity. The simplified network structure could avoid interfering nodes and improve the global search capability. The five characteristic gases produced by transformer faults are divided into two groups. A fault diagnosis model of three layers with four classifiers was built. We also investigated the effects of the number of hidden nodes and scale factors on RVFL network learning ability. Simulation results show that the number of implicit layer nodes has a large impact on the network model when the number of input dimensions is small. The network requires a higher number of implicit layer neurons and a smaller threshold range. The size of the scale factor has significant influence on the network model with larger input dimension. This paper describes the theoretical basis for parameter selection in RVFL neural networks. The theoretical basis for the selection of the number of hidden nodes, and the scale factor is derived. The importance of parameter selection for the improvement of diagnostic accuracy is verified through simulation experiments in transformer fault diagnosis.

1. Introduction

Urban construction and sustainable development have become the focus of modern social development. The sustainable development of a city is inseparable from the safe and stable operation of the urban power grid. An important part of city construction and planning is the power network. The power transformer is one of the most important components in the power system. A good fault diagnosis method for power transformers is very important to improve the reliability of transformers [1–7]. Dissolved gas analysis (DGA) method is an effective method for transformer fault diagnosis. DGA methods can be divided into two categories: conventional and intelligent methods, including critical gas, IEC ratio, Rogers ratio, Doernenburg ratio, etc., all of which detect faults by the amount of gas

released in the transformer oil. Fault diagnosis in these methods is based on the decisions of human experts. They explain the different gas releases due to temperature and energy variations of the fault [8]. All of these methods are relatively simple to implement, but their percentage of fault detection is quite low and may not be sufficient to reliably determine the type of fault [9]. Poor fault diagnosis results prompted researchers to look for new techniques.

With the development of artificial intelligence, machine learning techniques are also widely used to solve complex classification and regression problems in power systems. Among them are artificial neural networks, Bayesian networks, and support vector machines. Using intelligent methods can achieve a considerably higher diagnostic accuracy than traditional similar methods. However, the complex and variable relationship between power

transformer fault symptoms and different fault types makes it difficult to establish simulation models for power transformer fault diagnosis and network training.

The Artificial Neural Network (ANN) diagnostic method has a high diagnostic capability and compensates for the shortcomings of the three-ratio method for determining mixed fault types [10–16]. However, the method suffers from overfitting and nonconvergence in the learning process [17]. Bayesian multicore learning method was proposed and applied in transient stability assessment, but the method requires large-scale training data and the threshold range is chosen empirically [18–21]. Least-squares support vector machine (LS-SVM) was applied for load prediction [22–28]. The main drawback is that the kernel functions are difficult to select and must satisfy the Mercer condition [29, 30]. The key parameters directly affect the generalization performance but lack a structured approach to achieve optimal parameter selection.

The random vector function link (RVFL) network is a single hidden layer feedforward neural network [21–33]. It is widely used in deep and transfer learning [34–37]. Also, it has good potential for handling large-scale data, fast dynamic modeling, and real-time data processing [38–40]. In recent years, RVFL network models have been used in classification and regression, but there are few studies involving the number of hidden nodes and scale factor values. Neural network models with few hidden nodes cannot guarantee modeling performance, but neural network models with more hidden nodes may be overfitted leading to poor generalization performance [41–45]. Over the past few years, some constructive solutions have received considerable attention. These solutions start with a smaller network and gradually generate hidden nodes and output weights until an acceptable learning performance is achieved [46, 47].

This paper examines common shortcomings in RVFL network data modeling and transformer fault diagnosis. Radbas activation functions and uniformly distributed unbiased network models are selected. This study explains the instability of the neural network model used in numerous literatures. The importance of randomly assigning ranges to implicit parameters in RVFL networks is illustrated by an arithmetic example that points out the validity and infeasibility of two-step generation of RVFL networks with randomly assigned inputs. Furthermore, modeling data on RVFL networks with fixed range settings is risky. The main contributions of this paper are summarized as follows:

Firstly, an improved RVFL network model is proposed to introduce the concept of hidden node sensitivity. Each hidden node is classified and the hidden nodes with low sensitivity are eliminated in Section 2. Secondly, a standardization method using characteristic inputs is proposed in Section 3. The proposed method of processing the input data ensures full utilization of the sample. Finally, the theoretical basis for the number of hidden layer nodes and the scale factor values of the RVFL neural network and the value rules are presented in Section 4. It is verified with a transformer fault diagnosis model.

2. The Improved Random Vector Functional Link Algorithm

2.1. The RVFL Algorithm. The feedforward neural network (FNN) is one of the most popular algorithms in small sample machine learning [48–51]. One of the most commonly used topologies is the single hidden layer feedforward neural network (SLFN). Neurons between adjacent layers (input layer to hidden layer and hidden layer to output layer) are interconnected, but there is no interconnection between neurons in nonadjacent layers. The algorithm has been widely applied in classification and regression problems.

The random vector functional link (RVFL) network is a single hidden layer feedforward neural network (SLFN) [52]. The red line in Figure 1 indicates the direct connection between the input neuron and the output neuron. The RVFL combines the advantages of random weights and function chains and can be used to solve classification, regression, and prediction problems in many fields. In addition, RVFL networks can be combined with other learning methods to generate hybrid algorithms.

The RVFL network is a random weighted neural network. Suppose the output $y \in R$, the RVFL neural network can be described as a weighted sum of the outputs of the nodes in the L hidden layer:

$$y = \sum_{m=1}^L \beta_m \mathbf{h}_m(\mathbf{x}; \mathbf{w}_m) = \beta^T \mathbf{h}(\mathbf{x}; \mathbf{w}_1, \dots, \mathbf{w}_L). \quad (1)$$

Here β is the output weight, and the input is a d -dimensional real vector $x \in R_d$. The m^{th} transform is parameterized by the vector \mathbf{w}_m . Here, $h: x \rightarrow R$ is called a base. The sigmoid activation function is used in the simulation:

$$h_k(i, w) = f(x_k) = \exp \left\{ - \left(\sum_{m=1}^M w_{m,k} i_m^{(n)} \right)^2 \right\}. \quad (2)$$

At the beginning of the learning process, w_1, \dots, w_B are selected at first. In the RVFL neural network, parameters of the implicit function are randomly selected from a predefined probability distribution. Assuming that the basis function is continuous and stationary, the general approximation ability of RVFL can be guaranteed if L is large enough. By using this method, equation (1) can be transformed into a linear regression equation for coefficients $\beta = \beta_1, \beta_2, \dots, \beta_L$. If the required function for N samples is provided, it is called a training set, $S = \{x_i, y_i\}, i = 1, 2, \dots, N$.

$$\mathbf{H} = \begin{pmatrix} h_1(\mathbf{x}_1) & \cdots & h_L(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_N) & \cdots & h_L(\mathbf{x}_N) \end{pmatrix}. \quad (3)$$

The optimal parameter β can be obtained by using a standard regular least squares method

$$\beta^* = \arg \min_{\beta \in R^L} \frac{1}{2} \|\mathbf{H}\beta - \mathbf{Y}\|_2^2 + \frac{\lambda}{2} \|\beta\|_2^2. \quad (4)$$

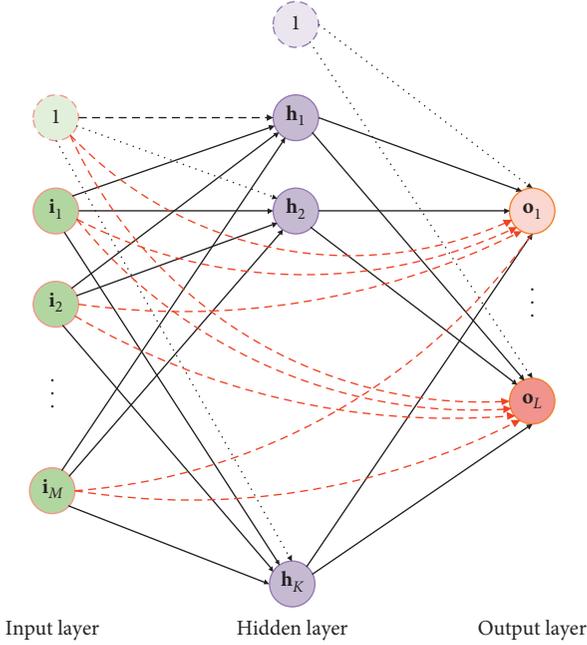


FIGURE 1: Schematic diagram of an RVFL network.

Let the gradient of $J(\beta)$ be zero, which is

$$\frac{\partial J}{\partial \beta} = \mathbf{H}^T \mathbf{H} \beta - \mathbf{H}^T \mathbf{Y} + \lambda \beta = 0. \quad (5)$$

The optimal parameter can be written as

$$\beta^* = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T \mathbf{Y}. \quad (6)$$

The parameter \mathbf{I} is the identity matrix. The inverse of the $L \times L$ matrix affects the computational complexity of RVFL. In the case of $N \ll L$, for any $\lambda > 0$, it can be simplified by the following equation:

$$(\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T = \mathbf{H}^T (\mathbf{H} \mathbf{H}^T + \lambda \mathbf{I})^{-1}. \quad (7)$$

By combining equations (6) and (7), the optimal weight β can be obtained.

$$\beta^* = \mathbf{H}^T (\mathbf{H} \mathbf{H}^T + \lambda \mathbf{I})^{-1} \mathbf{Y}. \quad (8)$$

The RVFL algorithm is simple and has short training time, but it is also weak in computational complexity and generating ability. The concept of hidden node sensitivity is introduced, and each hidden node is classified according to its sensitivity. Finally, several hidden nodes with the lowest sensitivity are removed, which leads to a simplified network structure. This method has been applied to ELM [53–55]. This paper uses the same idea to optimize and improve RVFL.

2.2. Description of the Improved RVFL. The sensitivity of hidden nodes is the space distance between the output results before and after removing hidden nodes.

$$E(l, i) = \|f_{\text{bef}} - f_{\text{aft}}\| = |h_{li}| \|\beta_l\|. \quad (9)$$

If $E(l, i)$ is larger, the greater the sensitivity. $E(l, i)$ is only based on one sample, which is not enough to measure the real state. Therefore, it needs a large amount of samples, which enables it to fully show its change law and trends. As a result, the equation could be transformed to

$$E_{\text{avg}}(l) = \frac{1}{N} \sum_{i=1}^N |h_{li}| \|\beta_l\|. \quad (10)$$

$E_{\text{avg}}(l) \rightarrow +\infty$ shows the sensitivity is the greatest. For different practical applications, the sensitivity of hidden nodes cannot be decided only by the value of $E(l, i)$ due to the uncertainty of its value. For instance, its sensitivity is very large in some programs, but very small in others. Therefore, a novel solution is proposed in this paper.

Given a training sample (x_i, y_i) , $i = 1, \dots, N$, the sensitivity based on each sample is $E(l, 1), E(l, 2), \dots, E(l, N)$. Therefore, the sensitivity of the hidden node can be represented by a point Se , where $\text{Se} = (E(l, 1), E(l, 2), \dots, E(l, N))$. In practical applications, the Se cannot be directly measured. So the Murkowski distance between Se and space original point O can be used when $q = 2$, which is the Euclidean distance.

$$d(\text{Se}, O) = \left(\sum_{i=1}^N |E(l, i) - O_i|^q \right)^{1/q} = \left(\sum_{i=1}^N |E(l, i)|^q \right)^{1/q}. \quad (11)$$

Se is similar to $E_{\text{avg}}(l)$. However, it is inconsistent with people's cognitive habits. Therefore, it can be transformed to

$$\text{Sen}(l) = e^{-\alpha/d(\text{Se}, O)} = e^{-\alpha \|\beta_l\|_q^{-1} (\sum_{i=1}^N |h_{li}|^q)^{-1/q}}. \quad (12)$$

$\text{Sen}(l) \in (0, 1)$, α is the adjustment factor governing the sensitivity distribution. $\text{Sen}(l) \rightarrow 1$ indicates higher sensitivity.

After the above detailed analysis, the description of improved RVFL is shown in Figure 2:

- (1) According to the specific problem, the threshold parameters ε and ε_{Se} are initialized first, and then the SLFN with sufficient hidden nodes is established through the RVFL. Finally, the output weight β and the output matrix of the hidden node are obtained according to equation (8).
- (2) For each input training sample, the sensitivity of each hidden node $E(s, i)$ is obtained using equation (9), and then the sensitivity $\text{Sen}(l)$ is calculated using equation (12).
- (3) For current input samples x_{N+1} , the output of SLFNS y_{N+1} can be obtained at a given threshold ε . If the error of the actual output y'_{N+1} and the expected output y_{N+1} are satisfied with $\text{SSE} (y_{N+1} - y'_{N+1}) < \varepsilon$, it can remove some hidden nodes with lower sensitivity and go to step (4). Otherwise, it should be prepared for prediction or classification based on the next new sample.

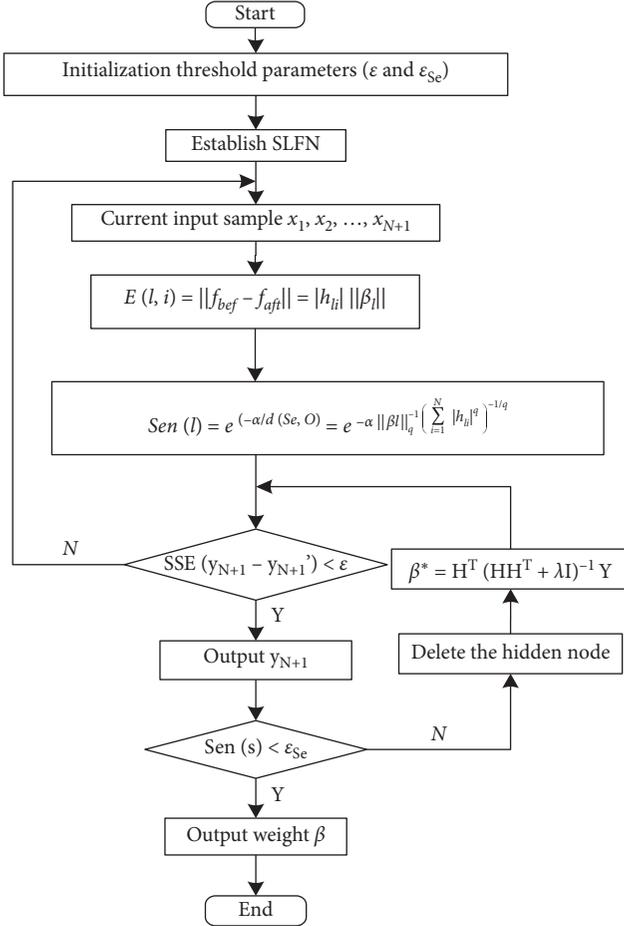


FIGURE 2: Hidden layer node improvements.

- (4) For a given threshold ε_{Se} , if the sensitivity of the S^{th} hidden node satisfies $\text{Sen}(s) < \varepsilon_{Se}$, the hidden node can be deleted.
- (5) After deleting the hidden node, update the output weight with formula (8), and continue to analyze whether there is a hidden node with low sensitivity, and go to step (3).

3. Design of Transformer Fault Diagnosis Model Based on Improved RVFL Neural Network

The fault mechanism of oil-immersed power transformers is rather complicated and there are many types of faults. In fact, no matter what the cause of the failure is, it can ultimately be attributed to two factors: thermal or electrical. Therefore, the transformer fault analysis can be divided into five states of operation: normal state, low energy discharge, high energy discharge, low temperature heat generation, and high temperature overheating. Transformer fault analysis is essentially a multi-classification problem. However, the RVFL-based classification algorithm involves only two classification problems. Therefore, it is necessary to convert multiple classifications into two classifications.

In the event of a power transformer failure, the transformer oil will decompose to produce many hydrocarbon

molecules, including hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2). Therefore, five kinds of gases mentioned above are analyzed.

3.1. Fault Diagnosis Flow. After determining the feature variables and sample data, and building a transformer fault diagnosis model, the classification model can be learned on the training set according to Figure 3. In addition, four classifiers can be constructed. The diagnosis process is shown in Figure 3:

The four RVFL classifiers are constructed in the following steps. Firstly, the activation function is chosen as a Gaussian kernel function and a suitable number of nodes in the hidden layer are selected. Then weights and biases are calculated to adjust the inputs, while the input-output relationship of the diagnostic model is obtained based on the correlation vector and the correlation function. Finally, the training is completed and the test data set is tested.

The random vector function chain neural network (RVFL) is simple in structure and training. The nonlinear problem in transformer faults is reasonably solved with a very fast convergence rate using the strong learning ability and the principle of approximating the minimum value in the RVFL network model.

3.2. Diagnostic Model Based on RVFL Network. In this paper, a diagnostic model is built using a binomial tree approach to transform the five classification problems into a three-level, four-classification problem. Each classifier is binary and the model is shown in Figure 4. Each set of input data of the RVFL neural network model is five-dimensional, representing five characteristic gases. The output data are one-dimensional. The five-dimensional inputs correspond to a set of one-dimensional outputs. The output type is normal or one of the four fault types.

The RVFL network model has three layers. It contains four classifiers and five different running states. A classifier 1 (RVFL1) in the first layer is used to diagnose whether the power transformer is in a fault or normal state. Classifier 2 (RVFL2) located in the second layer is responsible for determining whether the specific fault of the transformer in the fault state is a thermal or electrical fault. The third layer has two classifiers. Classifier 3 (RVFL3) is used to distinguish between low-energy discharge and high-energy discharge faults of transformers with electrical faults. Classifier IV (RVFL4) is responsible for the diagnosis of specific faults of power transformers in the thermal fault state, such as low temperature heating faults or high temperature heating faults. The stage value of each classifier is set up, as shown in Table 1. It should be pointed out that each classifier is independent of each other when constructing the model. Since each binary classifier is independent of each other, they can perform their own classification without interfering with each other.

3.3. Selection of Characteristic Variables. Dissolved gas samples in transformer oil usually have a large variation and dispersion due to differences in voltage levels and

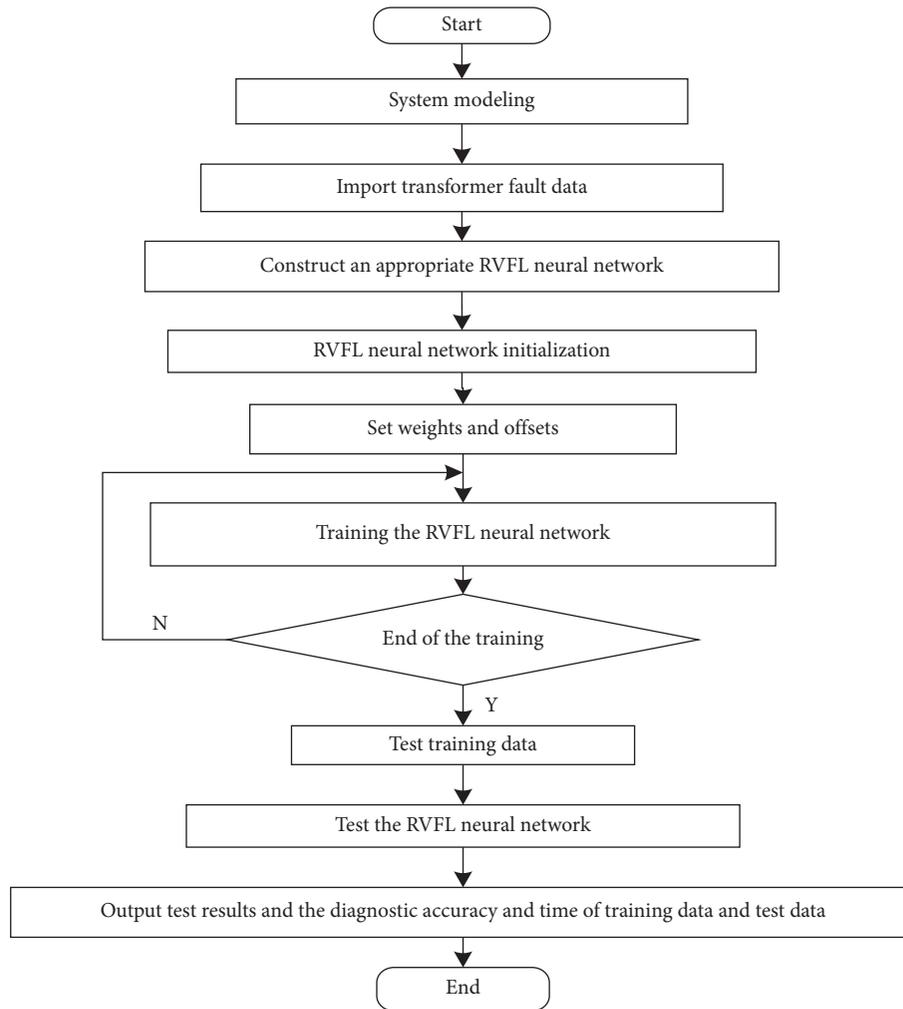


FIGURE 3: Flow chart of transformer fault diagnosis based on the RVFL network.

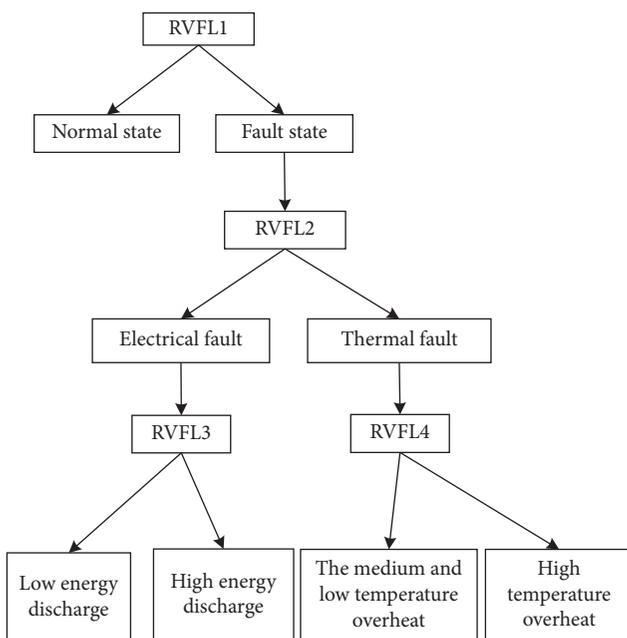


FIGURE 4: Fault diagnosis model of power transformer based on the RVFL.

transformer capacities. In some specific cases, the model is not sensitive to smaller inputs. If raw data are used directly as inputs, some important information will be missed. In order to improve the accuracy of diagnosis, different neural network models are built for different feature variables. In addition, more effective information needs to be fully utilized to reduce the effect of magnitude differences. It should be noted that the characteristic gas samples need to be normalized as inputs. Therefore, there are two scenarios for selecting the feature variables.

3.3.1. Characteristic Gas Content Ratio as a Characteristic Variable. The ratio of the separated gas content in the five transformer oils was selected as the input characteristic. Obviously, there are some differences in the order of magnitude of each group of gas content data compared to the original characteristic gas content. Using it directly as the input data of the simulation model is likely to complicate the network model structure or overfit the test results.

To investigate the diagnostic value and runtime of the test results, the training set data were tested again. The model was trained using these data and the accuracy of the

TABLE 1: Classifier value settings.

Classifier	Target value	
RVFL1	Normal state: 1	Fault target: 0
RVFL2	Electrical fault: 1	Thermal fault: 0
RVFL3	Low energy discharge: 1	High energy discharge: 0
RVFL4	Medium and low temperature overheat: 1	High temperature overheat: 0

calculated results was compared with the test set data. The input gas content was preprocessed according to the requirements of the model parameters and results. Data with larger dimensions were normalized based on the choice of scale factors. The significant effect of dimensional differences on the test results was eliminated.

The ratios of the five gases of H_2 , CH_4 , C_2H_6 , C_2H_4 , and C_2H_2 to the total gas content are selected as the characteristic variables. Because the ratio of the five characteristic gas contents is between 0 and 1, it does not need to be standardized and can be directly used as input data.

3.3.2. Standardized Processing of Characteristic Gas Content as a Characteristic Variable. There is a large difference between the data sets in terms of raw characteristic gas content. If the characteristic content is used directly as input data, it is not only complicated to operate but also the fault diagnosis accuracy is low. In this paper, the minimum distance obtained by normalized least squares method is used to normalize the raw characteristic gas content, which avoids the significant influence of input data with different dimensions on the simulation model.

$$X_{\text{new}} = C_{\min} + \frac{X - X_{\min}}{X_{\max} - X_{\min}} (C_{\max} - C_{\min}), \quad (13)$$

where X_{new} is the gas content after standardization. X_{\min} and X_{\max} are the minimum and maximum value of the gas content, respectively. C_{\min} and C_{\max} are the upper and lower bounds of the gas content standardization, respectively.

Different neural network models are built for different feature variables to obtain different diagnostic results. In order to reduce the effect caused by the size difference, more effective information can be fully utilized. Assume that we have n samples, where $X_n = \{x_{n1}, x_{n2}, x_{n3}, x_{n4}, x_{n5}\}$. Five of them represent the input quantities of five gases, namely, hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2) [56]. Each set of input data has five dimensions and the output data have one dimension. Depending on the five characteristic gases of the inputs, each output represents a fault state or a normal state of the transformer output.

The diagnostic results of the two models with feature variables as input data will be compared in the simulation. More effective information will be fully explored by analyzing the requirements and relationships of the relevant parameters of the RVFL network model under different input types.

3.4. Selection of Sample Data. In the establishment of any neural network model, in order to achieve a good fault

diagnosis effect, the most representative and universal sample data should be selected. When selecting the type of dissolved gas in the transformer oil, the sample data should contain all the data in the operating state.

In addition, try to satisfy each run with roughly the same proportion of input data. However, because of the influence of the transformer internal and external uncertainties, the real amount of dissolved gas in transformer span is greater than the training sample span. Hence, small sample training cannot meet the practical requirements of transformer fault diagnosis, and large sample size training sample spans a wide range of existence. The problem of sample dispersion will reduce the generalization ability of the neural network. Therefore, for samples with large amount of data and discrete order of magnitude of features, the input feature scaling is used to normalize the sample data.

When the network model is tested, it is also necessary to include every running state in the test data. The proportion is basically the same. The data volume of the training set and the test set are divided in a ratio of roughly 2 : 1. Each data set should contain transformer data in various operating states. And each run state has a roughly equal share of each data set.

4. Simulation Results

4.1. The Experimental Setup. To verify the effectiveness of the proposed algorithm, four independent fault detectors with no direct relationship between diagnostic results are used to identify four transformer states. The RVFL1, RVFL2, RVFL3, and RVFL4 classifiers are used to distinguish normal from fault, thermal from electrical fault, low-energy from high-energy discharge, and low-temperature from high-temperature fever, respectively. The simulation results and diagnostic accuracy α of each classifier will be introduced separately below (α = the correct number of test samples/the total number of test samples).

First, the scale factor values are determined for different sample data and input characteristics. Considering the effects of model structure and input parameters, the model produces more accurate results when diagnosing the training set. However, the model has higher error in the results when diagnosing the test set data. Therefore, it is necessary to preprocess the input data to ensure the validity of the experiment and the accuracy of the simulation results.

In this paper, the RVFL neural network normalizes the input feature data before diagnosis. Data standardization avoids large network diagnostic errors caused by large numerical differences between input data and output data. Secondly, network diagnostics are not always faster and more accurate when using a scaling range of $[-1, 1]$, and optimal performance is not always guaranteed. Therefore, in

order to achieve the desired order of magnitude of input sample data in each dimension and to ensure the validity of RVFL neural network diagnosis, simulation was carried out to study the input characteristics in two different cases, that is, the content characteristic variables of the five characteristic gases as input data and the content characteristic variables of the characteristic gases as standardized input data. After that, the number of hidden layer neurons under different sample data and input characteristics was determined. The number of nodes in the hidden layer has an important influence on the accuracy of model diagnosis during the construction of the RVFL neural network. The experiment was performed by separately changing the characteristics of two different variables as input data for the number of hidden nodes. Observe the data model of transformer fault diagnosis with the same number of hidden layer nodes in different neural networks. Record the running time of training data and test data, and diagnose the difference in accuracy of test data set and test data set diagnosis.

As a case study, the sample data set contains 162 groups of content data of five characteristic gases of transformer during operation, namely, hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2) [18]. Among them, there are 64 groups of characteristic gas content data in normal operation state and 98 groups of data in fault operation state. In this simulation experiment, 162 samples of characteristic gas content extracted from 10 kV power transformer oil were provided by our cooperative enterprise.

4.2. Analysis of Test Results

4.2.1. RVFL1 Simulation and Test Results. In order to ensure the reliability of learning, 120 sets of data were randomly selected from 162 sample data as training sets. The remaining 42 sets of data serve as test sets. When the characteristic variable is the characteristic gas content, the value of the scaling factor S is adjusted to obtain the new simulation results, as shown in Figure 5.

It can be seen from Figure 5 that the expected purpose can be achieved by adjusting the scale of the two data sets on the premise that the testing process of the training set and the test set is not affected. The total amount of the training data set is appropriately expanded, so it can be seen that the scaling factor $S = 10$ is the optimal threshold for diagnostic accuracy and running time. Therefore, when the input data are the transformer gas content, the selected scale factor $S = 10$ is the best value of the existing data set, and its random range $[-S, S]$ is $[-10, 10]$.

When the characteristic variable is the characteristic gas content, adjust the value of N , select the optimal number of hidden layer nodes, and sort out selected partial results, as shown in Figure 6.

The observation of the test data shows that starting from $N = 1$, the diagnostic accuracy of the RVFL neural network for both the training and test sets grows rapidly in a stepwise manner as the number of hidden nodes increases. When the number of hidden nodes is $13 = N < 20$, the diagnostic

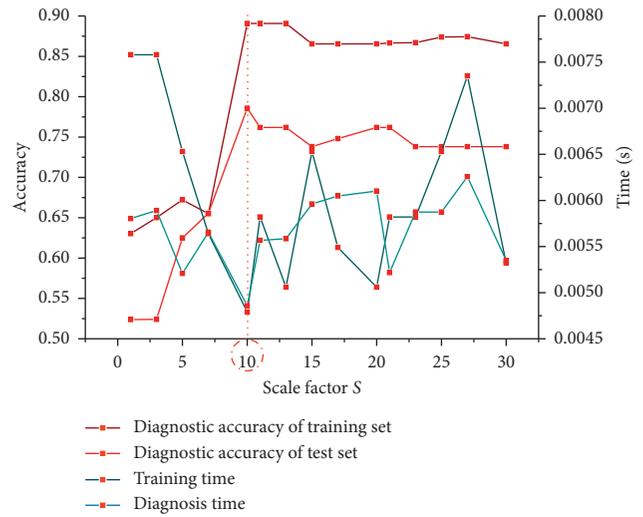


FIGURE 5: Selection of scale factor S of classifier RVFL1.

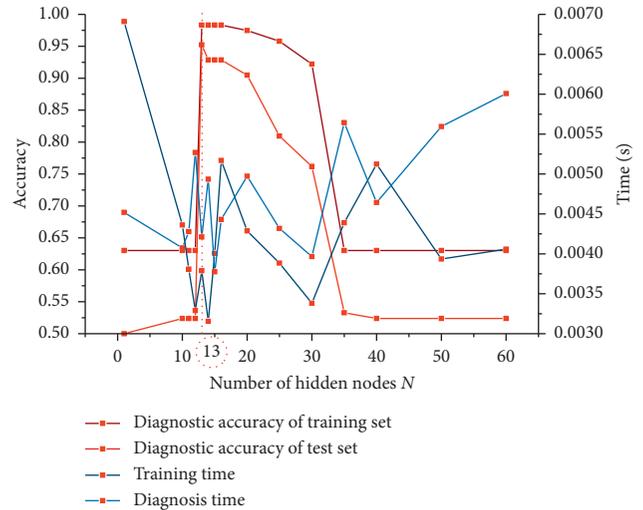


FIGURE 6: Selection of hidden layer node N in classifier RVFL1.

accuracy reaches a saturation peak for both the training and test sets, and the running time is slightly higher than the other hidden node numbers. When the hidden nodes $N \geq 20$, the network diagnostic capability gradually decreases. Considering that too many neurons in the hidden layer will complicate the network structure, prolong the convergence time, and even result in “overfitting,” we choose the number of nodes $N = 13$ in the hidden layer as the best choice for the available sample data set.

When the characteristic variable is the standardized characteristic gas content, the value of the scaling factor S is adjusted to obtain the new simulation result, as shown in Figure 7.

It can be seen that when the scaling range increases, the accuracy of the RVFL neural network diagnostic training set changes greatly, while the data of the diagnostic accuracy test set changes less. When the scaling factor was greater than 2.1, it remained basically the same. After the diagnostic accuracy of the training set data reached the peak, the scale

factor value was greater than 4, which was difficult to adjust due to the small sample data set. In the experiment, if the value of S continues to increase, the diagnostic accuracy of the training set reaches a relatively high value floating around 0.9 after $S = 500$, but the diagnostic accuracy of the test set data does not improve. At the same time, considering the impact of running time, the scale factor $S = 2.1$ is selected as the best value of the existing data set, and its random range $[-S, S]$ is $[-2.1, 2.1]$.

When the characteristic variable is the standardized characteristic gas content, adjust the value of N , select the optimal number of hidden layer nodes, and sort out selected partial results, as shown in Figure 8.

When the input feature is the normalized gas content, the diagnostic accuracy of changing the number of hidden nodes is less than that of the feature vector being the characteristic gas content. The effect of this on the diagnostic accuracy of the test set is almost negligible. At the same time, it can be concluded that when the number of nodes in the hidden layer reaches 130, the RVFL network shows an obvious “overfitting” phenomenon. The diagnostic accuracy of trained data is high but the diagnosis of untrained data is often unsatisfactory. Therefore, considering the running time and diagnostic accuracy, the number of nodes in the hidden layer $N = 100$ is chosen as the optimal value for the existing data set.

The optimal parameters tested were $N = 13$, $S = 10$ for the characteristic gas content as a feature variable and $N = 100$, $S = 2.1$ for the normalized gas content as a feature variable. N is the number of nodes in the hidden layer and S is the scale factor. The specific process is not repeated here. Table 2 shows the diagnostic results for the characteristic gas content ratio and the results after normalizing the characteristic variable gas content. The results of the partial test of the RVFL1 classifier are shown in Table 2:

Fault type 1 is the normal state, and fault state 0 is the fault state. Y_{t_t} and Y_{t_t} denote the network diagnosis results of the test set before and after the standardization process. The trY represents the true failure type of the test set. The diagnostic accuracy is 95.24% and 97.99%, respectively, when the characteristic gas content ratio and the standard gas content ratio are used as the characteristic variables. The test times are 0.004952 seconds and 0.004212 seconds.

4.2.2. RVFL2 Simulation and Test Results. The classifier RVFL2 contains a total of 98 sets of characteristic gas data for electrical transformer fault operation states. Among them, there are 55 sets of data in electrical fault operation state, and the remaining 43 sets of data in thermal fault operation state. Seventy of the 98 sample data are selected as training data. The remaining 28 groups are used as test data.

The input data transformer characteristic gas content was reduced due to the sample data when the number of hidden layer nodes was $N = 13$ and the random range was still $[-10, 10]$. The diagnostic accuracy of the training set in RVFL2 with 70 sets of sample data is 82.86% and the convergence time is 0.008460 seconds. The diagnostic accuracy of the test set in the remaining 28 sets of sample data

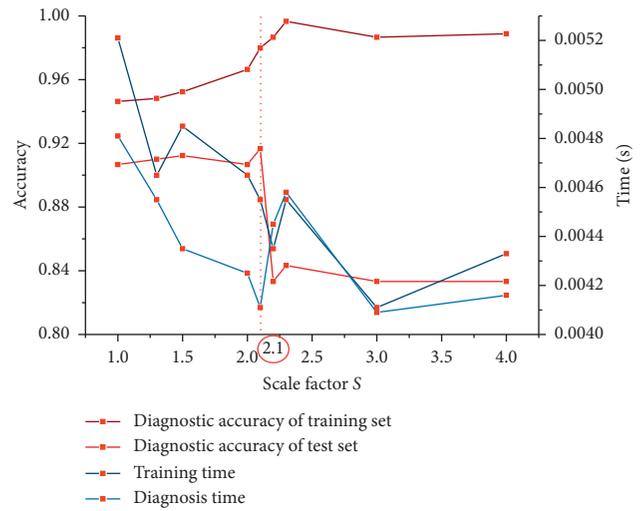


FIGURE 7: Selection of scale factor S of classifier RVFL1 after standardization.

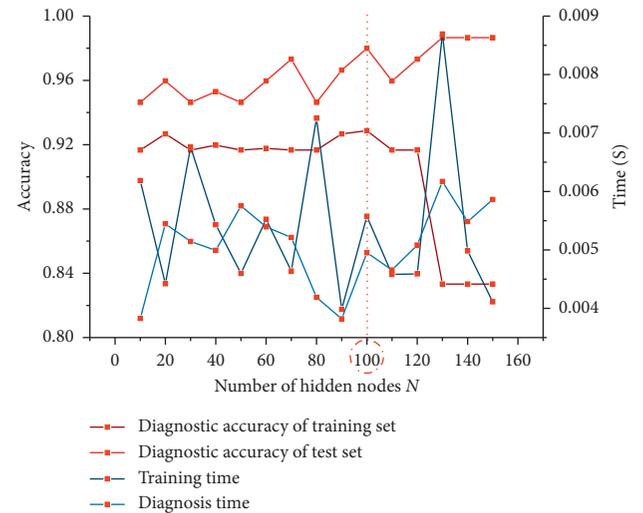


FIGURE 8: Selection of hidden layer node N in classifier RVFL1 after standardization.

TABLE 2: Comparison of RVFL1 diagnosis results under standardized input characteristics.

	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂	trY	Y _{t_t}	Y _t
1	155	581	215	293	142	0	1	0
2	477	566	734	239	116	0	0	0
3	117	5.8	7	15.9	12.8	0	0	0
4	52.5	32.9	79.3	9.73	9.8	0	0	0
5	19.8	9.6	2.29	1.17	0	1	1	1
6	21.1	8.97	2.25	1.08	0	1	0	0
7	166.6	28.31	6.72	12.4	0.3	1	1	1

in RVFL2 is 73.57%, and the convergence time is 0.005415 seconds.

This shows that the simulation output is not satisfactory. From the results, it was found that the RVFL fault diagnosis model outputs all results as “electrical fault”. The network fault diagnosis accuracy for both training data and test data

is not satisfactory. Therefore, it is necessary to adjust the relevant parameters in the model to ensure the effectiveness of the RVFL network training. Attempts to adjust the number of nodes N in the hidden layer were made, and it was found that adjusting the number of hidden nodes could not improve the accuracy of fault diagnosis and had very little effect on detection. Therefore, we adjusted the randomization range.

When the characteristic variable is the characteristic gas content, the value of the scaling factor S is adjusted to obtain the new simulation results, as shown in Figure 9.

According to the data in the above table, the randomization range is adjusted, and it is found that when the scale factor is within the range of $1 = S \leq 15$, the detection accuracy keeps increasing. In addition, after the range is exceeded, the model gradually appears as an overfitting phenomenon. Therefore, $S = 15$ is selected as the optimal scale factor value of classifier RVFL2. Adjust the random range $[-S, S]$ to $[-15, 15]$ and keep the number of hidden layer nodes N unchanged.

When the number of hidden layer nodes is $N = 13$ and the initial value $S = 10$ and the optimal value $S = 15$ are selected by the scale factor, the diagnosis results of transformer fault by RVFL network are compared. Where Y_{t1_temp} and Y_{1_temp} , respectively, represent the diagnostic output results of the network to the training set and the test set when the randomization range $[-S, S]$ is the initial value $[-10, 10]$. Y_{t2_temp} and Y_{2_temp} , respectively, represent the optimal threshold range $[-S, S]$ of debugging randomization $[-15, 15]$ and the diagnostic output results of the network to the training set and the test set. The diagnostic output results of the network to the training set and the test set are shown in Table 3.

The simulation test results with the number of hidden nodes $N = 13$ and the scale factor $S = 10$ are compared with the test results with the scale factor S adjusted to 15. When the classifier RVFL2 adjusted the parameters to $N = 13$, $S = 15$, the diagnostic accuracy of 70 sets of data in the training set of classifier RVFL2 was 94.29%, and the convergence time was 0.003247 seconds. The accuracy of 28 sets of test set data was 90.71% and the test time was 0.003544s.

The scaling range $[-15, 15]$ was set according to Table 3. The detection accuracy of the training data is improved by 11.43% and the detection accuracy of the test data is improved by 17.14% after parameter optimization. It can be seen that the parameter tuning of the RVFL2 classifier achieves the expected effect when the input feature is the characteristic gas content of the transformer.

The optimal parameters for the tested characteristic gas content as a feature variable are $N = 13$, $S = 15$, and the optimal parameters for the normalized gas content as a feature variable are still $N = 100$, $S = 2.1$. N is the number of hidden layer nodes and S is the scale factor. Table 4 shows the diagnostic results for the characteristic gas content ratio and the results after normalizing the gas content of the characteristic variables. Some test results of the RVFL2 classifier are shown in Table 4.

Failure type 1 is electrical failure, and 0 is thermal failure. The letters mean the same as above. The diagnostic accuracy

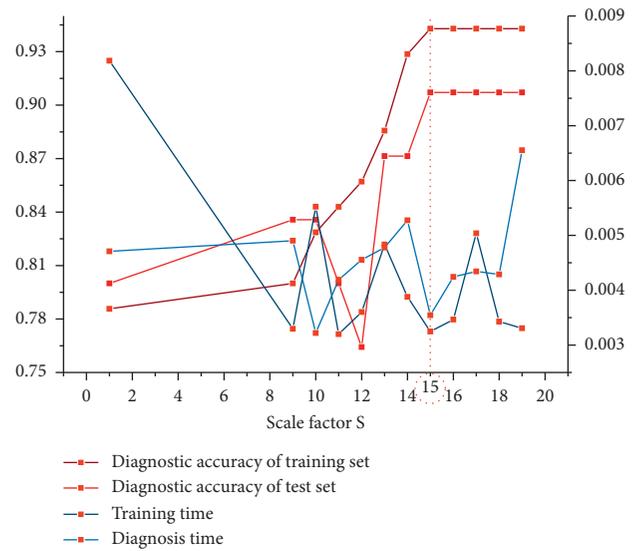


FIGURE 9: Selection of scale factor S of classifier RVFL2.

TABLE 3: Comparison of RVFL2 diagnosis results.

	Yt1_temp	Yt2_temp	testY	Y1_temp	Y2_temp	trainY
1	1	1	1	1	1	0
2	1	1	0	1	0	0
3	1	1	1	1	1	1
4	1	0	1	1	1	1
5	1	1	1	1	1	1
6	1	1	1	1	1	1
7	1	1	0	1	1	0
8	1	1	1	1	1	1
9	1	0	0	1	1	1
10	1	0	1	1	1	1
11	1	1	0	1	1	1
12	0	0	0	1	1	0
13	1	0	0	1	0	0
14	1	1	1	1	1	1
15	1	1	0	1	1	0

TABLE 4: Comparison of RVFL4 diagnosis results under standardized input characteristics.

	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂	trY	Yt_t	Y_t
1	117	5.8	7	15.9	12.8	1	1	0
2	52.5	32.9	79.3	9.73	9.8	1	0	1
3	421	135	351	27.7	374	1	1	1
4	46	98	26.3	41.3	0	0	0	0
5	25	85	178	9.6	1.4	0	0	0
6	56.4	42.7	1413	167	17.1	0	1	0
7	74.8	482	1409	297	7.5	0	1	0

was 90.71% and 93.10%, respectively. The test time was 0.002975 seconds and 0.004761 seconds, respectively.

4.2.3. *RVFL3 Simulation and Test Results.* The classifier RVFL3 contains a total of 55 groups of characteristic gas data of power transformers under the operation state of electrical faults, among which 22 groups are gas content data under

the operation state of low energy discharge, and the other 33 groups are gas content data under the operation state of high energy discharge. Take any 40 of the 55 sample data as training data. The remaining 15 groups are used as test data. The characteristic variable is the characteristic gas content, the hidden layer node $N=13$, and the scale factor is $[10, 10]$. The diagnostic model of the test data and the resultant output of the training data are 1.

At this time, it can be concluded that the accuracy rate of 40 groups of sample data in the classifier RVFL3 diagnosis training set is 67.50%, and the training time is 0.004132 seconds. The accuracy of the 15 sample data in the classifier RVFL3 diagnostic test set is 64.57%, and the test time is 0.005092 seconds. After trying to adjust the number of nodes in the hidden layer N , it is found that under the sample data of classifier RVFL3, adjusting the number of hidden nodes cannot improve the fault diagnosis accuracy of RVFL3, and the impact on the detection time is extremely small. Therefore, the adjustment scale factor is selected. The test results of classifier RVFL3 under the change of scale factor are shown in Figure 10.

According to the above data, it is found that before the scale factor $S=17$, with the expansion of the threshold range, the diagnosis time also increases rapidly. But accuracy does not change for either the training data or the test data. When the scale factor is within the range of $17=S<22$, the detection accuracy begins to rise slowly. After exceeding this range, the training effect of the model becomes significantly worse. From the above data, it is found that when the scale factor S increases, the diagnostic accuracy of the classifier RVFL3 to the training data is slightly improved. However, with the increase of the scale factor value, the diagnostic accuracy of the network for the test set data begins to decline. Therefore, the first consideration is that the diagnostic accuracy of the training set data is not greatly improved. Second, according to the figure above, the improvement of the diagnostic accuracy of training data does not improve the diagnostic performance of training data, but shows a small decline. Third, considering that the increase of scale factor S will prolong the training and detection time of the neural network, it is not applicable to the real power network. Therefore, under the condition of ensuring the diagnosis accuracy of RVFL network to the maximum extent, choosing the minimum random range will make the fault diagnosis network more practical, which is conducive to the RVFL fault detection model to judge the occurrence of faults in the first time, and reduce the fault loss to the minimum to ensure the economic efficiency of the power system.

Therefore, in the case that the number of hidden nodes $N=13$, $S=20$ are selected as the optimal scale factor values of classifier RVFL3.

As shown in Table 5, when the number of hidden layer nodes is constant, the input feature variable characteristic gas content scaling factor S is compared as its initial values 10 and 20 corresponding to the optimal value of classifier RVFL3 and the diagnosis results of transformer faults in RVFL network, respectively. Y_{t1_temp} and Y_{1_temp} , respectively, represent the diagnostic output results of the network to the training set and the test set when the

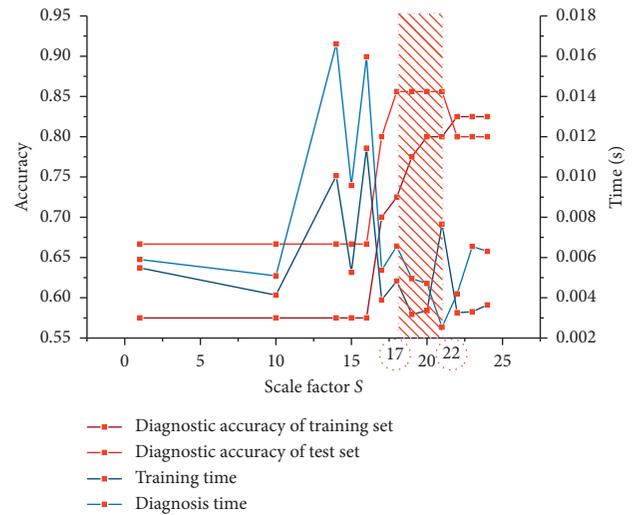


FIGURE 10: Selection of scale factor S of classifier RVFL3.

TABLE 5: Comparison of RVFL3 diagnosis results by scale factor optimization.

	Yt1_temp	Yt2_temp	testY	Y1_temp	Y2_temp	trainY
1	0	0	0	1	1	1
2	1	1	1	1	1	1
3	0	0	0	0	0	0
4	0	0	0	0	1	1
5	0	0	0	0	1	0
6	1	1	1	0	0	0
7	0	0	0	0	0	0
8	1	1	1	0	0	0
9	0	0	0	0	1	1
10	1	1	0	1	1	1
11	0	1	1	0	0	0
12	0	0	0	1	1	1
13	0	1	1	0	1	1
14	0	0	0	0	0	0
15	0	0	0	0	0	1

randomization range $[-S, S]$ is the initial value $[-10, 10]$. Y_{t2_temp} and Y_{2_temp} , respectively, represent the diagnostic output results of the network to the training set and the test set when the optimal threshold range $[-S, S]$ is set $[-20, 20]$ after debugging.

Combined with the data in the table, the simulation test results in the case of the hidden node number $N=13$ and scale factor $S=10$ before adjusting the relevant parameters of classifier RVFL3 are compared with the test results in the case of tuning scale factor S to 20.

When the relevant parameters of the classifier RVFL3 are adjusted, that is, when the number of hidden nodes $N=13$ and the scaling factor $S=20$, the diagnostic accuracy of the 40 training sets of the classifier RVFL3 is 80.3% and the training time is 0.003361 seconds. The accuracy of the 15 sample data in the classifier RVFL3 diagnostic test set is 85.6%, and the test time is 0.004706 seconds. It is clear that after parameter optimization, the detection accuracy of the network model improved by 12.8% for the training data and 21.03% for the test data. It can be seen that the parameter

tuning of the RVFL3 classifier is very successful when the input feature is the characteristic gas content of the transformer.

The input data are the standardized transformer characteristic gas content, the number of hidden layer nodes $N=100$, and the randomization range is $[-2.1, 2.1]$. At this point, the diagnostic results of the RVFL network model are acceptable for the training data set but not optimal for the test data set. The effectiveness of the network training process is poor.

At this time, the accuracy of 40 sets of sample data in the classifier RVFL3 diagnostic training set was 67.28%, and the training time was 0.004133 seconds. The accuracy of 15 sets of sample data in the RVFL3 diagnostic test set classifier was 74.67%, and the test time was 0.004709 seconds. By debugging the relevant parameters, it was found that adjusting the scaling factor did not significantly improve the test results of the classifier RVFL3 for the corresponding sample data under the input type. Therefore, we try to improve the training effect by changing the number of hidden layer nodes N .

When the scaling range remains unchanged as $[-2.1, 2.1]$, change the number of nodes in the hidden layer N to observe the changes in the simulation results of classifier RVFL3, as shown in Figure 11.

According to the data in the figure, when the scale factor S is still 2.1 and the number of hidden layer nodes is selected, $N=20$, the fault diagnosis accuracy of classifier RVFL3 reaches the highest value. When the number of hidden nodes is $N > 20$, the diagnostic accuracy of the RVFL network model to the training data still increases, but the diagnostic accuracy of the test data does drop one level. Although with the increase of the number of hidden nodes, the diagnostic accuracy of the network for the training set data is improved to some extent, the first consideration is that the diagnostic accuracy of the training set data is not greatly improved. Second, the improvement of the diagnostic accuracy of training data does not improve the diagnostic performance of test data but shows a small decline. Third, considering that the increase in the number of hidden layer nodes will prolong the training and detection time of the neural network, it can be seen that it is not applicable to the real power network.

The results show that on the premise of ensuring the transformer fault diagnosis of the RVFL neural network, choosing the number of hidden nodes N can effectively reduce the complexity of the network model. This method is helpful to determine the RVFL fault detection model to reduce the loss degree for the first time and ensure the safety of the power system. By analyzing the relationship between the number of hidden nodes in Figure 11 and the test accuracy and the length of test time, the number of hidden layer nodes $N=20$ is finally selected as the optimal value of this classifier.

As shown in Table 6, the scale factor is $[-2.1, 2.1]$, and the number of nodes in the hidden layer is used as the initial optimum when the normalized characteristic gas content is the input feature compared with the optimum value of the classifier RVFL3. $Yt1_temp$ and $Y1_temp$ denote the

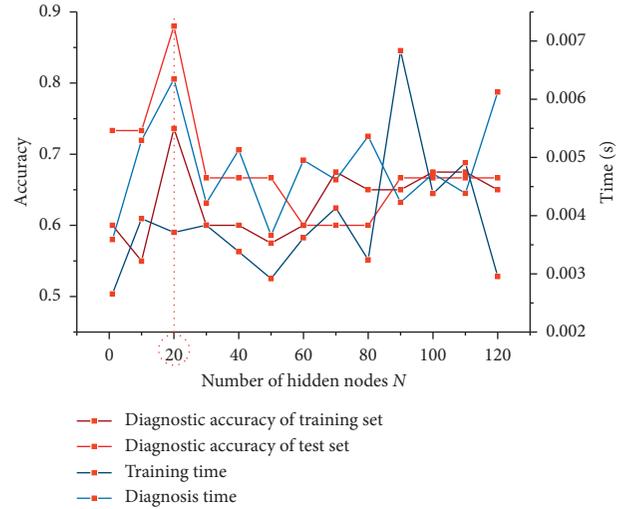


FIGURE 11: Selection of hidden layer node N in classifier RVFL3 after standardization.

TABLE 6: Hidden node optimization RVFL3 diagnosis comparison.

	Yt1_temp	Yt2_temp	testY	Y1_temp	Y2_temp	trainY
1	1	0	0	1	0	1
2	0	0	0	0	0	0
3	0	0	1	1	1	1
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	1	1	1	1	0	1
7	1	0	0	0	0	0
8	1	1	1	0	0	1
9	0	0	0	1	1	0
10	0	0	1	0	0	0
11	0	0	0	1	1	1
12	0	1	0	0	0	0
13	1	1	1	0	0	1
14	0	0	0	0	0	0
15	1	0	0	1	0	1

diagnostic outputs of the network for the training and test sets, respectively, when the number of hidden nodes N is an initial value of 100. $Yt2_temp$ and $Y2_temp$ represent the diagnostic output of the network for both the training and test sets, respectively. The optimal value of 20 is obtained when the number of hidden nodes is optimized.

When the relevant parameters of the classifier RVFL3 are adjusted, that is, when the number of hidden nodes is $N=20$ and the scale factor is $S=2.1$, the accuracy rate of the 40 groups of sample data in the diagnostic training set of the classifier RVFL3 is 73.62% and the training time is 0.003713 seconds. The accuracy of the 15 sample data in the classifier RVFL3 diagnostic test set is 88.00% and the test time is 0.006346 seconds.

Obviously, after parameter optimization, the detection accuracy of network model training data increases by 6.34%, and the detection accuracy of test data increases by 13.33%. It can be seen that the parameter adjustment of classifier RVFL3 is very successful when the input feature is the normalized characteristic gas content of transformer.

Through the test, the characteristic gas content is obtained as the best parameter of the characteristic variable, $N=13$, $S=20$. After standardized treatment on characteristic variables, the optimal gas content parameters were $N=20$ and $S=2.1$. Where N is the number of hidden layer nodes, and S is the scale factor. The diagnostic results shown in Table 7 are the characteristic gas content ratio and the standardized treatment of gas content as the characteristic variables, respectively. Some test results for the RVFL3 classifier are shown in Table 7.

Fault 1 is low energy discharge, and fault 0 is high energy discharge. The letters mean the same as above. The diagnostic accuracy is 85.6% and 88%, respectively. The test times are 0.003132 seconds and 0.006254 seconds, respectively.

4.2.4. RVFL4 Simulation and Test Results. The classifier RVFL4 contains a total of 43 groups of characteristic gas data of power transformers under thermal failure operation state, among which 14 groups are gas content data under medium and low temperature heating operation state, and the remaining 27 groups are gas content data under high temperature heating operation state. Take 35 groups of gas characteristic variables as training data. The remaining 8 groups of characteristic variables are test data.

When the input data are the characteristic gas content of the transformer, that is, when the number of hidden layer nodes is $N=13$ and the randomization range is still $[-10, 10]$, the simulation results of the classifier RVFL4 are good. When the input data are the standardized characteristic gas content and the parameters are not optimized, that is, the number of hidden layer nodes $N=100$ and the randomization range is $[-2.1, 2.1]$, the detection accuracy of classifier RVFL4 for 35 groups of sample data in the training set is 72.14%, and the training time is 0.002364 seconds. However, the diagnostic accuracy of the classifier RVFL4 for 8 groups of sample data in the test set is 62.50% and the test time is 0.001845 seconds.

It can be seen that the RVFL network model has an ideal simulation result for the training set but has low simulation accuracy for the test set. According to the model test results, it can be found that the diagnostic accuracy of the training set data is much smaller than that of the test set by 9.64%. Therefore, the training process of the network model is basically ineffective at the current parameter settings.

According to the above assumptions, it is found through simulation and debugging related parameters that under the input type, for the corresponding sample data, the adjustment scale factor S does not significantly improve the test results of the classifier RVFL4. Therefore, by changing the value of the hidden layer node N , a certain rule can be observed. The relationship between the simulation results of classifier RVFL4 and the number of hidden layer nodes is shown in Figure 12:

As shown in the figure, it can be seen that when the number of hidden layer nodes $N=1$, the test results for both the training data and the test data set are the best, and the fault diagnosis accuracy is also the highest. When the

TABLE 7: Comparison of RVFL3 diagnosis results under standardized input characteristics.

	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂	trY	Y _{t_t}	Y _{_t}
1	26.5	29.8	69.3	0	8.6	1	1	1
2	34.4	21.9	45	3.2	19.6	1	0	1
3	316	36.1	109	37.3	440	1	1	1
4	29	7.52	14.8	5.5	5.4	0	0	0
5	6.3	7.87	3.03	6.97	11.1	0	0	1
6	413	6.8	4.6	40.4	8.15	0	1	0
7	23.4	14.3	9.71	7.12	3.44	0	0	0

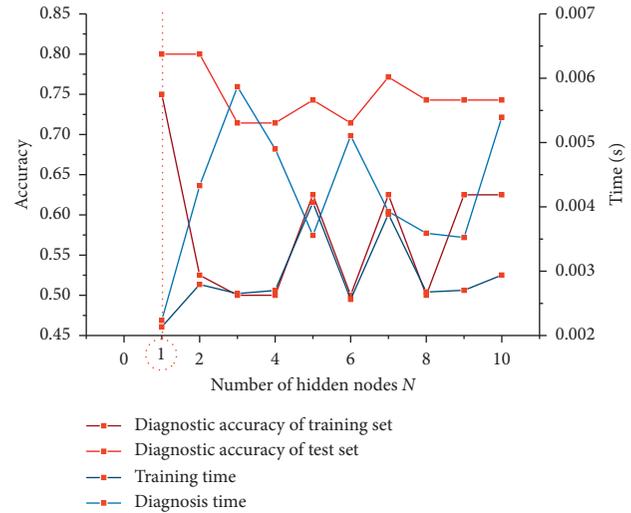


FIGURE 12: Selection of the hidden layer node N in classifier RVFL4 after standardization.

number of nodes in the hidden layer is $N > 2$, the diagnostic accuracy of the network for the training data set is basically stable, while the diagnostic accuracy of the data for the test set is also gradually decreasing, and the network gradually shows the phenomenon of overfitting. Because the increase of the number of hidden layer nodes prolong the training and detection time of the neural network, it is not applicable to the real power network. Therefore, in order to ensure the diagnostic accuracy of the RVFL network to the maximum extent, the least hidden layer neurons are selected. The fault type can be diagnosed and processed in the shortest time with the lowest model complexity, and the loss can be reduced to the lowest degree. Therefore, the number of hidden layer nodes $N=1$ is selected as the optimal value of classifier RVFL4.

As shown in Table 8, the diagnosis results of transformer faults by the RVFL network are compared under two different conditions, in which the initial value of 100 is selected for the number of hidden layer nodes and the optimal value 1 corresponding to classifier RVFL4 is selected. Y_{t1_temp} and Y_{1_temp} represent the diagnostic output of the network for the training and test sets, respectively, of classifier RVFL4 when the number of hidden nodes N is the initial value of 100. Y_{t2_temp} and Y_{2_temp} represent the diagnostic output of the network for the training and test sets, respectively, when the number of hidden nodes is adjusted and the optimal value of 1 is chosen with the same scaling range.

TABLE 8: Hidden node optimization RVFL4 diagnosis comparison.

	Yt1_temp	Yt2_temp	testY	Y1_temp	Y2_temp	trainY
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	1	1	1
4	0	1	1	0	0	1
5	0	0	0	0	1	1
6	1	0	1	0	0	0
7	0	1	1	0	0	0
8	0	0	1	0	0	1

The comparison of the data in Table 8 illustrates the simulation test results of the classifier RVFL4 for both the test data and the training data. The simulation result data when the number of hidden nodes N is set to 1 and the scale factor S remains 2.1 are compared with the simulation data when the number of hidden nodes $N=100$ and the scale factor $S=2.1$. The accuracy of the diagnostic training set classifier RVFL4 for 35 sets of sample data in RVFL4 with the optimized number of hidden nodes $N=1$ and scale factor $S=2.1$ is 75.00%, and the training time is 0.002135 seconds. The accuracy of classifier RVFL4 for diagnostic test set of 8 sample data is 80.00%, and the test time is 0.002235 seconds.

It can be found that by optimizing the number of hidden nodes in the classifier RVFL4, the diagnostic accuracy of the RVFL neural network for training data is improved by 2.86%, and the diagnostic accuracy of test data is increased by 17.50%. It can be seen that the parameter adjustment of classifier RVFL4 is successful when the input feature is the transformer gas content ratio.

The simulation experiment shows that the optimal parameter of characteristic gas content as the characteristic variable is $N=1$, $S=2.1$, and the optimal parameter of characteristic variable after standardized treatment is $N=13$, $S=10$. Where N is the number of hidden layer nodes and S is the scale factor. The diagnostic results shown in Table 9 are the characteristic gas content ratio and the standardized treatment of gas content as the characteristic variable, respectively. Some test results shown in Table 9.

Fault type 1 is a medium to low temperature heating fault, and fault type 0 is a high temperature heating fault. The meaning of these letters is the same as above. The correct diagnosis rate is 75.7% and 80.00%, respectively. The convergence times are 0.003525 and 0.002235 seconds, respectively.

4.3. Comparison with the Five Classification Methods

4.3.1. RVFL Simulation and Test Results. A classifier is constructed to classify the five operating conditions of transformers using the available 162 sets of transformer gas characteristic data. There are 64 sets of transformer normal working condition operation data, 22 sets of low energy discharge fault data, 33 sets of high energy discharge data, 14 sets of medium and low temperature overheating fault data, and 27 sets of high temperature overheating fault data. The randomly selected 120 sets of data are used as the training set, and the remaining 42 sets are used as the test set. Considering

TABLE 9: Comparison of RVFL4 diagnosis results under standardized input characteristics.

	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂	trY	Yt_t	Y_t
1	18	20	12	6	2	1	1	1
2	46	98	26.3	41.3	0	1	0	1
3	25	85	178	9.6	1.4	1	1	1
4	155	581	215	293	142	0	0	0
5	477	566	734	239	116	0	1	1
6	56.4	42.7	1413	167	17.1	0	1	0
7	74.8	482	1409	297	7.5	0	0	0

the influence of input characteristics and related parameter values on the simulation test results, the unification of parameter values with input characteristics of order of magnitude will improve the network diagnosis function.

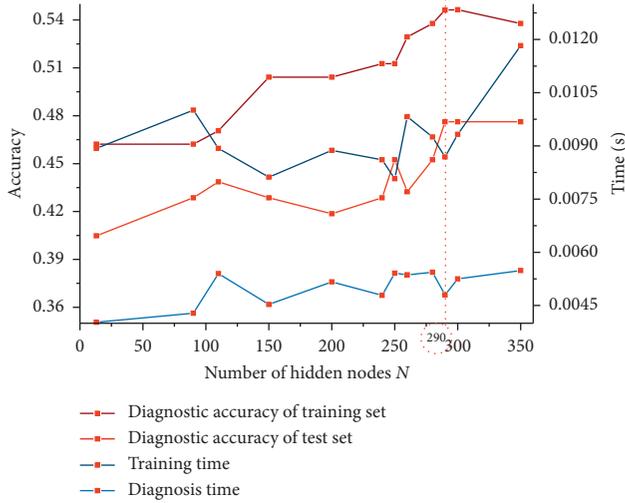
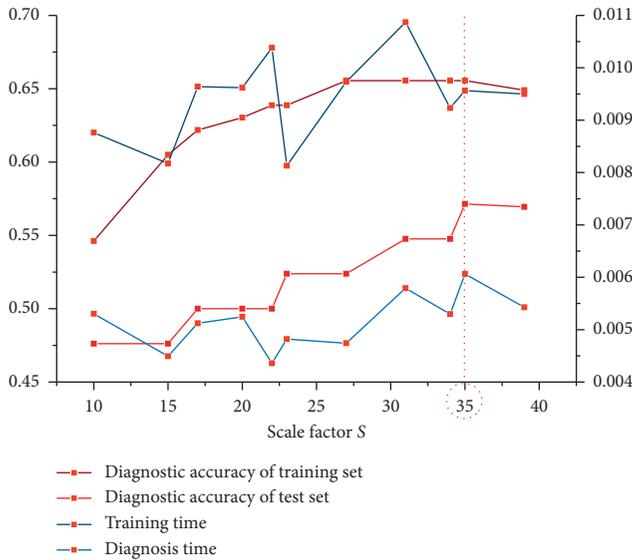
When the input feature is the gas content extracted from transformer oil, and when the scaling range $[-S, S]$ is still $[-10, 10]$, the influence of the numerical selection of the number of hidden layer nodes N on the five classifiers is first considered, as shown in Figure 13:

With the gradual increase in the number of nodes in the hidden layer, the fault diagnosis accuracy of the network steadily improves for both training and test data. The convergence speed increases slightly when the network structure tends to be complex. The accuracy of RVFL neural network for training set and test set data diagnosis is the highest when the number of hidden nodes $N=290$. When $N=290$, the diagnostic accuracy of the training set decreases slightly and then continues to increase. Meanwhile, the diagnostic accuracy of the neural network for the test set data will remain at the peak after a sharp drop until the neural network overfitting occurs. Therefore, in order to minimize the complexity of network construction, shorten the convergence time and improve the learning effect of the model; the number of hidden layer nodes $N=290$ is selected as the characteristic input for five characteristic gas contents of the transformer, and the optimal value of hidden nodes in the overall classification is obtained.

The parameter scale factor S has a significant effect on the gas content of the relative dispersion of the input features. The effect of different values of the scale factor S on the overall classification simulation test results is shown in Figure 14.

The scale factor range increases with the classification accuracy of the training data and test data of the RVFL network model, and the diagnostic accuracy of the training data network slowly steps up. The accuracy of both training data and test data showed an overall increasing trend. When the value of the scale factor is in the range of $23 = S \leq 39$, the detection accuracy starts to rise slowly. Beyond this range, the training effect becomes significantly worse.

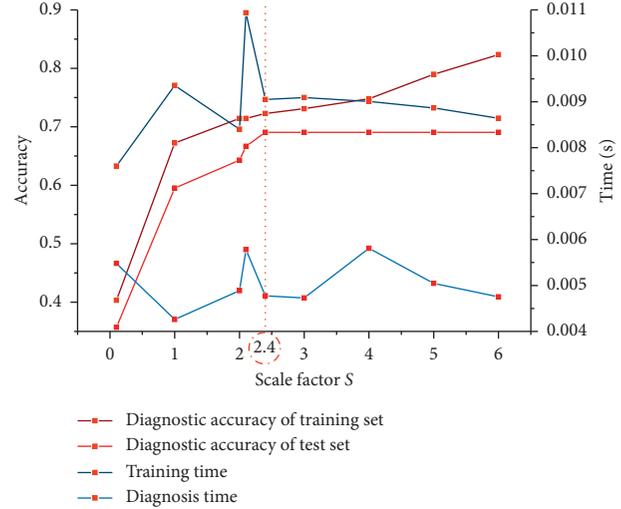
Therefore, choosing the smallest scale factor value $S=35$ under the condition of ensuring the diagnostic accuracy of RVFL will make the fault diagnosis network more practical. This facilitates the RVFL fault detection model to determine the occurrence of faults in the first time and reduce the fault loss to a minimum. $S=35$ was chosen as the optimal scale factor value for the overall classification of the model, and the random range $[-S, S]$ was adjusted to $[-35, 35]$.

FIGURE 13: Selection of hidden layer node N in classifier RVFL.FIGURE 14: Selection of scale factor S of classifier RVFL.

When the input feature is the characteristic gas content of the transformer after standardization, and when the number of hidden layer neurons N is still 100, the influence of the value selection of scale factor S on the five classifiers is first considered, as shown in Figure 15.

From Figure 15, it can be seen that the fault diagnosis accuracy of both training and test data has increased but the change in model convergence time is small. When the scale factor $[S, S]$ is $[2.4, 2.4]$, the diagnostic accuracy set saturation of the network model for the test data is fixed at 69.05% without rising. If we continue to debug the scale factor, we can find that the diagnostic accuracy of both network data and test data decreases significantly when $S > 6$. Therefore, $S = 2.4$ is selected as the optimal scale factor value of the five classifiers. The random range $[-S, S]$ is adjusted to $[-2.4, 2.4]$.

Then, under the condition that the scaling range $[-S, S]$ is $[-2.4, 2.4]$, the number of hidden layer nodes N , which has

FIGURE 15: Selection of scale factor S of classifier RVFL after standardization.

a great influence on the input characteristics, is studied. As shown in Figure 16, it is the influence of different values of hidden node N on the simulation test results of five categories.

From the data in the figure, it can be seen that the highest fault diagnosis accuracy is achieved for the five classifiers with the selected number of hidden layer nodes $N = 200$ when the scale factor S is 2.4. When the number of hidden nodes is in the range of $200 = N < 250$, the network saturates the diagnostic accuracy of the test data set. With the increasing number of hidden nodes N , the diagnostic accuracy of the network on the training set data is improved to some extent. However, the diagnostic accuracy of the training set data is not greatly improved. Secondly, the improvement of the diagnostic accuracy of the training data did not improve the diagnostic performance of the test data but slightly decreased. In addition, the increase in the number of nodes in the hidden layer prolongs the training and detection time of the neural network. This indicates that it is not applicable in the real grid.

While ensuring the diagnostic accuracy of the RVFL neural network, improving the number of hidden nodes can effectively reduce the complexity of the network model. This method can reduce the loss of RVFL fault diagnosis model and ensure the safety of power system. By analyzing the relationship between the number of hidden nodes in Figure 16 and the test accuracy and the length of test time, the number of hidden layer nodes $N = 200$ is finally selected as the optimal value of this classifier.

The optimal parameters tested for the characteristic gas content as a characteristic variable are $N = 13$, $S = 10$, and the optimal parameters tested for the standardized gas content as a characteristic variable are $N = 100$, $S = 2.1$. N is the number of hidden layer nodes, and S is the scale factor. Some of the diagnostic results are shown in Table 10 for the characteristic gas content ratio and the results after normalizing the gas content of the characteristic variables.

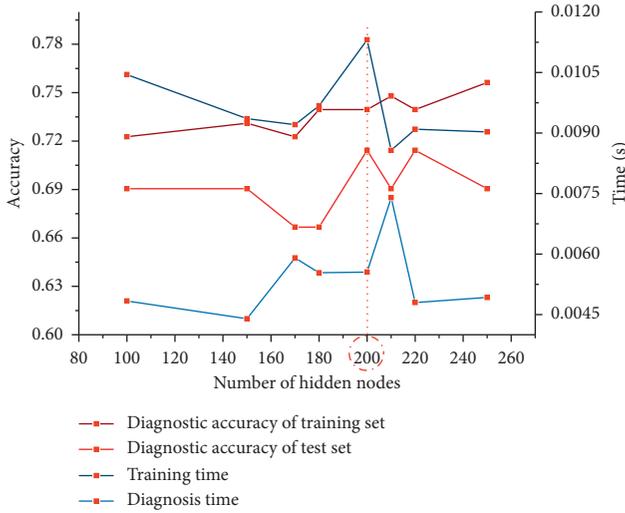


FIGURE 16: Selection of hidden layer node N in classifier RVFL after standardization.

TABLE 10: Comparison of RVFL diagnosis results under standardized input characteristics.

	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂	trY	Y _{t_t}	Y _{_t}
1	20.9	8.85	2.47	1.04	0	1	1	1
2	22.5	8.5	7.3	1.2	8.4	3	3	4
3	19.8	8.18	2.47	1.11	0	1	3	1
4	155	581	215	293	142	4	4	2
5	20.63	9.12	2.37	1.07	0	1	2	1
6	2054	370	535	26	805	2	2	2
7	20.8	8.97	2.39	1.13	0	1	3	1
8	4430	941	2475	76	55585	2	3	2
9	11	31	52	26	0	5	5	4
10	18.2	9.56	2.27	1.07	0	1	1	4
11	179	413	544	27	473	2	4	2
12	143	305	721	78	2	4	4	1
13	18.8	9.32	2.49	1.12	0	1	1	1
14	5.5	9.6	50	3.9	1.5	5	2	1
15	17	21	156	11	4.1	4	4	2
16	18.9	8.86	2.48	0.99	0	1	2	1
17	246	432	674	141	7.2	4	4	4
18	20	9.76	2.42	1.06	0	1	5	1
19	20.3	8.38	2.38	1.15	0	1	1	1
20	21.5	9.45	2.5	1.18	0	1	1	2

The data in the table show the diagnosis results of test set and training set, respectively, by RVFL network. Among them, fault type 1 represents normal operation state, 2 represents high energy discharge fault, 3 represents low energy discharge fault, 4 represents high temperature heating fault, and 5 represents medium and low temperature heating fault. Y_{t_t} and $Y_{_t}$ denote the fault diagnosis results of the test set before and after normalization, respectively. TrY denotes the true failure type of the tester. The correct diagnosis rate is 51.14% and 71.43%, respectively. The convergence time is 0.005356 seconds and 0.006428 seconds, respectively.

In summary, the fault diagnosis accuracy of the RVFL neural network for overall classification is much lower than that of the fault detection model consisting of four binary

classifiers. The convergence rate of the overall classification, which is nearly twice faster than that of the step classification, is not favorable for transformer fault diagnosis in real power grids. Meanwhile, the weight scaling effect of the classification model network becomes worse for the input feature of gas content. The diagnostic accuracy of the training set with standardised gas content as input feature is 8.4% higher than that with gas content as input. When the input feature is normalized gas content, the diagnostic accuracy of the test data is 14.29% higher than that when the input feature is gas content ratio.

4.3.2. *Analysis of Test Results.* In part B, a three-layer, four-classifier model is constructed for five different operating states of the power transformer. Under two different types of input characteristics, four different sets of sample data are obtained by experimenting on the test set and training set data through the classifier. Two different types of feature variables, pre-standardized and post-standardized gas content, are taken as input data, and the feature scaling of different input sample data is performed according to the optimal threshold range, so that the feature data of different dimensions can be compared. At the same time, it speeds up the search for the optimal solution and greatly improves the accuracy of fault diagnosis.

In the five-classification model, the number of data types for the five samples varies considerably, resulting in an inadequate theoretical basis for the weighting values of a particular fault type when the training set data is small. The network needs to acquire more fault information and data volumes to learn, so that a significant increase in convergence time can be observed. Not surprisingly, the fault diagnostic accuracy of the RVFL neural network is lower than that of the fault detection model consisting of four dichotomous classifiers. In addition, the convergence of the five-classifier algorithm is slower than the stepwise classification, which is not conducive to the diagnosis of transformer faults in real power grids. Meanwhile, the weight scaling of the network is poor for the classification model with gas content as the input feature.

Thus, the three-layer, four-type classification model constructed in this paper for five different types of power transformer faults is superior to the five-class classification processing of the sample data.

5. Conclusion

In this paper, the relationship between RVFL neural network-based transformer fault diagnosis and the values of the associated parameters is investigated. The method yields stable ranges and diagnostic accuracies of the relevant parameters of the network model under two different input features and data sets, and the optimal range of values is discussed. To improve the diagnostic accuracy of the model, the traditional RVFL neural network model is modified and the input data are standardized. The method also reduces the complexity of the network model and simplifies the network model. The experiments validate the validity of the

parameters associated with the RVFL neural network. After adjusting the parameters, it was found that RVFL2 improved the diagnostic accuracy by 11.43% for the training set and 17.14% for the test set. The diagnostic accuracy of RVFL3 standardized on the training set improved by 6.34% and the diagnostic progress of the test set improved by 13.33%. Diagnostic accuracy improved by 2.86% for the RVFL4 training set and by 17.5% for the test set. The diagnostic accuracy of each classifier is substantially improved. Adjusting the number of hidden nodes when the input sample data are uniform can improve the accuracy of network diagnostics. When the input sample data are scattered, adjusting the scale range can improve the accuracy of network testing. For different sample data, the network testing accuracy is more sensitive to the change of scale factor than the change of the number of hidden nodes under the premise of ensuring the validity of model training.

This paper describes the theoretical basis for the number of implied nodes and scale factor assignment in the RVFL network model. It effectively improves the diagnostic accuracy of transformer faults. This study is of reference value for the analysis of other neural network engineering applications.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant No. 51507140); China Scholarship Council (CSC) State Scholarship Fund International Clean Energy Talent Project (Grant No. [2018]5046); State Key Laboratory of Electrical Insulation and Power Equipment (Grant No. EIPE17209); Natural Science Basic Research Plan of Shaanxi Province (Grant No. 2018JM5041); Operation Fund of Guangdong Key Laboratory of Clean Energy Technology (Grant No. 2014B030301022); Open Research Fund of Jiangsu Collaborative Innovation Center for Smart Distribution Network; and Nanjing Institute of Technology (Grant No. XTCX201703).

References

- [1] K. Wu, J. Kang, and K. Chi, "Power transformer fault diagnosis method based on improved multi-classification algorithm and related vector machine," *High Voltage Technology*, vol. 42, no. 9, pp. 3011–3017, 2016.
- [2] A. Abu-Siada and S. Hmood, "A new fuzzy logic approach to identify power transformer criticality using dissolved gas-in-oil analysis," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 401–408, 2015.
- [3] H. Wu, X. Li, and D. Wu, "RMP neural network based dissolved gas analyzer for fault diagnostic of oil-filled electrical equipment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 18, no. 2, pp. 495–498, 2011.
- [4] J. Fan, Z. Liu, A. Meng et al., "Characteristics of tin oxide chromatographic detector for dissolved gases analysis of transformer oil," *IEEE Access*, vol. 7, pp. 94012–94020, 2019.
- [5] C. Xiang, Z. Huang, J. Li, Q. Zhou, and W. Yao, "Graphic approaches for faults diagnosis for Camellia insulating liquid filled transformers based on dissolved gas analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 25, no. 5, pp. 1897–1903, 2018.
- [6] J. Jiang, M. Zhao, G.-M. Ma et al., "TDLAS-based detection of dissolved methane in power transformer oil and field application," *IEEE Sensors Journal*, vol. 18, no. 6, pp. 2318–2325, 2018.
- [7] Standard IEC 60599, *Guide for the Interpretation of Dissolved Gas Analysis and Gas Free*, International Electrotechnical Commission, Geneva, Switzerland, 2007.
- [8] J. Faiz and M. Soleimani, "Dissolved gas analysis evaluation in electric power transformers using conventional methods a review," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 2, pp. 1239–1248, 2017.
- [9] H.-C. Sun, Y.-C. Huang, and C.-M. Huang, "Fault diagnosis of power transformers using computational intelligence: a review," *Energy Procedia*, vol. 14, no. 2, pp. 1226–1231, 2012.
- [10] B. Zhao, X. Zhang, H. Li, and Z. Yang, "Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions," *Knowledge-Based Systems*, vol. 199, Article ID 105971, 2020.
- [11] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mechanical Systems and Signal Processing*, vol. 110, pp. 349–367, 2018.
- [12] Z. Meng, X. Guo, Z. Pan, D. Sun, and S. Liu, "Data segmentation and augmentation methods based on raw data using deep neural networks approach for rotating machinery fault diagnosis," *IEEE Access*, vol. 7, pp. 79510–79522, 2019.
- [13] J. Liang, T. Jing, H. Niu, and J. Wang, "Two-terminal fault location method of distribution network based on adaptive convolution neural network," *IEEE Access*, vol. 8, pp. 54035–54043, 2020.
- [14] S. Chen, H. Ge, J. Li, and M. Pecht, "Progressive improved convolutional neural network for avionics fault diagnosis," *IEEE Access*, vol. 7, pp. 177362–177375, 2019.
- [15] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Applied Soft Computing*, vol. 100, Article ID 106724, 2020.
- [16] H. Zhao, S. Zuo, M. Hou et al., "A novel adaptive signal processing method based on enhanced empirical wavelet transform technology," *Sensors*, vol. 18, no. 10, Article ID 3323, 2018.
- [17] J. Yu, H. Huang, H. Chen et al., "Application of neural network in transformer fault diagnosis research," *Control Engineering*, vol. 25, no. 10, pp. 1898–1904, 2008.
- [18] Y. Liu, T. Yan, L. Jiang et al., "An improved neural network algorithm optimized by genetic algorithm and its application in power transformer fault diagnosis," *Electronic Technology*, vol. 46, no. 10, pp. 54–56, 2017.
- [19] S. Wang, S. Zhang, T. Wu et al., "A first-order Markov dynamic Bayesian network classifier with continuous

- attributes,” *Knowledge-Based Systems*, vol. 195, Article ID 105638, 2020.
- [20] G. Chen, M. Liu, and J. Chen, “Frequency-temporal-logic-based bearing fault diagnosis and fault interpretation using Bayesian optimization with Bayesian neural networks,” *Mechanical Systems and Signal Processing*, vol. 145, Article ID 106951, 2020.
- [21] K. Jia, T. Feng, Q. Zhao, C. Wang, and T. Bi, “High frequency transient sparse measurement-based fault location for complex DC distribution networks,” *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 312–322, 2020.
- [22] C. Zhang and Z. Lin, “Transformer fault diagnosis based on normal cloud model and improved Bayesian classifier,” *Electrical Measurement & Instrumentation*, vol. 54, no. 4, pp. 50–56, 2017.
- [23] Y. Li and X. Gu, “Application of online SVR in very short-term load forecasting,” *International Review of Electrical Engineering*, vol. 8, no. 1, pp. 277–282, 2013.
- [24] X. W. Liang, A. P. Jiang, T. Li, Y. Y. Xue, and G. T. Wang, “LR-SMOTE—an improved unbalanced data set oversampling based on K-means and SVM,” *Knowledge-Based Systems*, vol. 196, Article ID 105845, 2020.
- [25] R. Gao, L. Zhang, and X. Liu, “Short-term load forecasting based on least square support vector machine combined with fuzzy control,” in *Proceedings of the 10th World Congress on Intelligent Control and Automation*, pp. 1048–1051, Beijing, China, July 2012.
- [26] L. Limei and H. Xuan, “Study of electricity load forecasting based on multiple kernels learning and weighted support vector regression machine,” in *Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC)*, pp. 1421–1424, Chongqing, China, May 2017.
- [27] M. Ali, Z. A. Khan, S. Mujeeb, S. Abbas, and N. Javaid, “Short-term electricity price and load forecasting using enhanced support vector machine and K-nearest neighbor,” in *Proceedings of the 2019 Sixth HCT Information Technology Trends (ITT)*, pp. 79–83, Ras Al Khaimah, UAE, November 2019.
- [28] G. Mitchell, S. Bahadoorsingh, N. Ramsamooj, and C. Sharma, “A comparison of artificial neural networks and support vector machines for short-term load forecasting using various load types,” in *Proceedings of the 2017 IEEE Manchester PowerTech*, pp. 1–4, Manchester, UK, June 2017.
- [29] Y. Li, “Short-term load forecasting based on LS-SVM optimized by BCC algorithm,” in *Proceedings of the 2009 15th International Conference on Intelligent System Applications to Power Systems*, pp. 1–5, Curitiba, Brazil, November 2009.
- [30] L. Shi and P. Yu, “Application of fuzzy support vector machine in transformer fault diagnosis,” *Electrical Measurement & Instrumentation*, vol. 52, no. 8, pp. 115–119, 2015.
- [31] L. Zhang, H. Ma, N. Jiang, T. Gao, and H. Xu, “Recognition and location of transformer winding looseness based on singular value spectrum entropy and SVM,” *Power System Protection and Control*, vol. 45, no. 18, pp. 69–75, 2017.
- [32] X. Xue, W. Sun, J. Wang et al., “RVFL-LQP: RVFL-based link quality prediction of wireless sensor networks in smart grid,” *IEEE Access*, vol. 8, pp. 7829–7841, 2020.
- [33] Y. Zhang, B. Li, Y. Wang, and Q. Tian, “The forecast of the temperature in subway station based on RVFL neural network,” in *Proceedings of the 2018 11th International Congress on Image and Signal Processing, Bio Medical Engineering and Informatics (CISP-BMEI)*, pp. 1–5, Beijing, China, October 2018.
- [34] T. Han, C. Liu, W. Yang, and D. Jiang, “Deep transfer network with joint distribution adaptation: a new intelligent fault diagnosis framework for industry application,” *ISA Transactions*, vol. 97, pp. 269–281, 2020.
- [35] T. Han, C. Liu, W. Yang, and D. Jiang, “Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions,” *ISA Transactions*, vol. 93, pp. 341–353, 2019.
- [36] K. Patan and M. Patan, “Neural-network-based iterative learning control of nonlinear systems,” *ISA Transactions*, vol. 98, pp. 445–453, 2020.
- [37] Y. Wu, Y. Fang, S. Shang, J. Jin, L. Wei, and H. Wang, “A novel framework for detecting social bots with deep neural networks and active learning,” *Knowledge-Based Systems*, vol. 211, Article ID 106525, 2021.
- [38] C. Cui and D. Wang, “High dimensional data regression using lasso model and neural networks with random weights,” *Information Sciences*, vol. 372, pp. 505–517, 2016.
- [39] S. Scardapane, D. Wang, M. Panella, and A. Uncini, “Distributed learning for random vector functional-link networks,” *Information Sciences*, vol. 301, pp. 271–284, 2015.
- [40] D. Wang, “Editorial: randomized algorithms for training neural networks,” *Information Sciences*, vol. 364–365, pp. 126–128, 2016.
- [41] W. F. Schmidt, M. A. Kraaijveld, and R. P. Duin, “Feedforward neural networks with random weights,” in *Proceedings of the 11th IAPR International Conference on Pattern Recognition. Vol. II: Conference B: Pattern Recognition Methodology and Systems*, pp. 1–4, The Hague, Netherlands, September 1992.
- [42] A. Elisseeff, T. Evgeniou, and M. Pontil, “Stability of randomized learning algorithms,” *Journal of Machine Learning Research*, vol. 6, no. 1, pp. 55–79, 2005.
- [43] D. Husmeier, “Random vector functional link (RVFL) networks,” *Neural Networks for Conditional Probability Estimation: Forecasting Beyond Pint Predictions*, Springer, vol. 6, London, UK, 1999.
- [44] B. Igel and Y.-H. Pao, “Stochastic choice of basis functions in adaptive function approximation and the functional-link net,” *IEEE Transactions on Neural Networks*, vol. 6, no. 6, pp. 1320–1329, 1995.
- [45] K. Luo, Z. Y. Jia, and W. Gao, “Strong and weak stability of randomized learning algorithms,” in *Proceedings of the 2012 IEEE 14th International Conference on Communication Technology*, pp. 9–11, Chengdu, China, November 2012.
- [46] Y. LeCun, L. Bottou, G. B. Orr, and K. R. Müller, “Efficient backprop,” in *Lecture Notes in Computer Science*, vol. 1524, pp. 9–50, Springer, Berlin, Germany, 1998.
- [47] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [48] Y.-H. Pao, S. M. Phillips, and D. J. Sobajic, “Neural-net computing and the intelligent control of systems,” *International Journal of Control*, vol. 56, no. 2, pp. 263–289, 1992.
- [49] E. Klinefelter and J. A. Nanzer, “Interferometric microwave radar with a feedforward neural network for vehicle speed-over-ground estimation,” *IEEE Microwave and Wireless Components Letters*, vol. 30, no. 3, pp. 304–307, 2020.
- [50] P. Musikawan, K. Sunat, Y. Kongsorot, P. Horata, and S. Chiewchanwattana, “Parallelized metaheuristic-ensemble of heterogeneous feedforward neural networks for regression problems,” *IEEE Access*, vol. 7, pp. 26909–26932, 2019.
- [51] G. Han, Q. Cheng, X. Sun, L. Li, and W. Di, “A biological mechanism based structure self-adaptive algorithm for feedforward neural network and its engineering applications,” *IEEE Access*, vol. 7, pp. 25111–25122, 2019.

- [52] L. Zhang and P. N. Suganthan, "A comprehensive evaluation of random vector functional link networks," *Information Sciences*, vol. 367-368, pp. 1094–1105, 2016.
- [53] L. Cui, H. Zhai, B. Wang et al., "Improved extreme learning machine based on the sensitivity analysis," *IOP Conference Series Materials Science and Engineering*, vol. 320, no. 1, Article ID 012015, 2018.
- [54] H. Han, L. Gan, and L. He, "Improved variations for extreme learning machine: space embedded ELM and optimal distribution ELM," in *Proceedings of the 2017 20th International Conference on Information Fusion (Fusion)*, pp. 1–8, Xi'an, China, July 2017.
- [55] L. Yang, Q. Zhao, and Y. Jing, "Channel equalization and detection with ELM-based regressors for OFDM systems," *IEEE Communications Letters*, vol. 24, no. 1, pp. 86–89, 2020.
- [56] J. B. Leite and J. R. S. Mantovani, "Distribution system state estimation using the Hamiltonian cycle theory," in *Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM)*, p. 1, Boston, MA, USA, July 2016.