Transformer Fault Identification with an IF-1DCNN Based on Informative Integration of Heterogeneous Sources

Huipeng Du, Gang Wang, and Jiazhao Li

School of Electrical and Information Engineering, Beihua University, Jilin 132013, China

Correspondence should be addressed to Gang Wang; jlwg@163.com

Received 4 December 2020; Revised 14 January 2021; Accepted 7 February 2021; Published 20 February 2021

Academic Editor: Jude Hemanth

Copyright © 2021 Huipeng Du 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.

1. Introduction

Power systems are an important basic engineering tool for social and economic development; large power transformers constitute very expensive and vital components in electric power systems [1], and the working state of power transformers affects the operational stability of power systems [2, 3]. Therefore, their normal and continuous service is vital [4]. In the long-term use of transformer equipment, there are inevitable ageing and failure hidden dangers [5]. Moreover, the cost of a power transformer is high, so it is necessary to diagnose the state of the transformer and take effective measures to fix faults that will occur to ensure the reliable operation of the transformer and reduce the occurrence of faults [6]. These diagnostic methods include various electrical, chemical, mechanical, acoustic, and other diagnostic methods [7–9].

At present, dissolved gas analysis (DGA) is one of the best methods for detecting an abnormal situation in the transformer [10–12] and has been widely used to monitor the state of power transformers. The ratio of gas content in DGA is closely related to the type of transformer fault [13]. At the same time, the two different characteristic space data can reflect the operation state of the transformer from different angles [14]. Early methods for interpreting DGA data include the Doernenburg ratio method, Rogers’ ratio method, and IEC ratio method, which have been developed and validated using large datasets of equipment in service. In these methods, multiple numeric thresholds and gas boundaries are commonly set to classify features of the dissolved gas data. However, insufficient ratio combinations or “code absence” may also invalidate the interpretation of DGA [15, 16]. Therefore, the fault diagnosis accuracy rate of these methods is relatively low [12, 17].

With the rise of machine learning, machine learning-related algorithms have been applied to transformer fault diagnosis. In the early stage, neural networks [18–20], support vector machines (SVMs) [14, 21, 22], and other algorithms [23–25] have improved the accuracy of transformer fault diagnosis. Jia et al. [19] proposed a wavelet
neural network diagnosis model based on the improved artificial fish-swarm algorithm, and gas content is used as the input of the diagnosis model. Equalb et al. [20] proposed that the artificial neural network has been trained using the weighted fault gas concentration for transformer incipient fault diagnosis. Li et al. [25], based on clustering techniques, propose a new method for fault diagnosis of transformers with the DGA and, corresponding to the initialization process, calculate its membership to the reference faults by the single feature. The correlation between single feature information and fault is limited, and the hybrid feature can reflect the operation state of the transformer from different angles. In reference [22], a set of new feature combinations is selected as input from the mixed feature quantity by the genetic algorithm. However, different new feature combinations may be obtained according to different sample data. At the same time, these methods have their own shortcomings. For example, the learning speed of a neural network is poor in small-sample data, but the convergence speed is too slow in large-sample data [26]. Although SVMs work is poor in small-sample data, but the convergence of the algorithm is not high. And, the parameter optimization of the kernel function is relatively difficult. The parameter optimization problems belong to NP-Hard problems, and some intelligent optimization methods were proposed to solve the NP-Hard problems [28], such as genetic algorithm (GA), particle swarm optimization (PSO), and differential evolution (DE) [14, 19, 22, 29–31].

In recent years, the application of deep belief networks (DBNs) in transformer fault diagnosis models [32–35] has achieved high precision, but it has a disadvantage that the network structure and parameters are basically determined by experiences, and it needs the evolutionary algorithm to avoid premature convergence and improve the global search ability [36]. And, DBNs require a large amount of unlabeled data for pretraining. In reference [24, 32, 34], these models classify transformer faults based on hybrid DGA features. Convolution neural networks (CNNs) [37–39] are also used in research on the transformer fault diagnosis methods. In reference [33, 35, 37–39], single-characteristic information that can reflect the operation state of a transformer is used as the input of the diagnosis model, but the hybrid feature of multiple feature information is not considered as the input.

At present, transformer state diagnosis methods are mainly based on the single feature or hybrid feature. However, the single-characteristic information that can reflect the operation state of a transformer cannot fully reflect the transformer fault classification, and it is often difficult to make a more correct diagnosis of the fault. For the hybrid feature, these new learning models do not specially distinguish different feature space in hybrid features to train models. At the same time, research on CNNs in power transformer fault diagnosis needs further study.

This paper presents a transformer fault diagnosis model based on one-dimensional convolution neural networks (1DCNN), and this model can enable informative integration of possibly heterogeneous sources. To make the extracted features from the convolution and pooling operations in the model independent of different feature information, it is necessary to transform the mixed feature to an isomer feature and input it into the 1DCNN model. A schematic diagram of the 1DCNN based on an isomer feature (IF-1DCNN) is shown in Figure 1.

2. 1DCNN Network Model Theory

CNNs have unique network layers: convolutional layers and pooling layers. Under the interaction of the two layers, the features of the input data can be extracted automatically, and the dimension of the data features can be reduced at the same time. CNNs have many different network structures. A classic LeNet-5 CNN structure is shown in Figure 2.

2.1. 1DCNN

2.1.1. One-Dimensional Convolution. In general, the convolutional layer uses a kernel function to deal with a one-dimensional feature \(X = [x_1, x_2, x_3, \ldots, x_N]\) to map the features. The output of the one-dimensional feature is used to extract the feature of \(X\). The process can be expressed by the following formula:

\[
s(i) = W X = \sum_{m=1}^{M} x_{i-1+m} w_{m},
\]

where \(s\) is a one-dimensional output feature, \(s(i)\) is the \(i\)th output feature element, \(W\) is the convolution kernel of order \(M\), \(w_m\) is the \(m\)th element of \(W\), and \(x_{i-1+m}\) is the \((i-1+m)\)th element of input \(X\).

An activation function needs to be added after the convolution operation to introduce the nonlinear factors into the neuron node and transform the learning model into a nonlinear model. Activation functions include the sigmoid, tanh, and rectified linear unit (ReLU) functions. In this paper, the current mainstream ReLU function is used, which is a function of bionic principles. It can more effectively carry out gradient descent and backpropagation, avoid gradient disappearance, and reduce the temporal and spatial complexity. The ReLU function can be expressed by Figure 3 and formula (2).

\[
F(s(i)) = \begin{cases} 
  s(i) & s(i) > 0 \\
  0 & s(i) \leq 0
\end{cases}
\]

It can be seen from the formula that when the input is positive, the output remains the same; when the input is nonpositive, the output becomes zero, which makes the neurons in the model network have sparse activation.

Using all-0 filling to keep the edge of the output matrix unchanged, but all-1 filling will change the size of the output matrix. Formulas (3) and (4) are given to calculate the side length of the output matrix when filling with all 0s and not all 0s, respectively:

\[
W_{\text{all-zero}} = \frac{M \times (N + 2) - 1}{2},
\]

\[
W_{\text{all-one}} = \frac{M \times (N + 2) - 1}{2}.
\]
\[
\begin{align*}
\text{length}_{\text{out}} &= \text{length}_{\text{in}} - \text{length}_{\text{filter}} + 1, \\
\text{width}_{\text{out}} &= \text{width}_{\text{in}} - \text{width}_{\text{filter}} + 1.
\end{align*}
\]

The working process of the 1DCNN is shown in Figure 4. This figure shows that the input feature size is 1*7, the network filling is not all-0 filling, and the size of the convolution kernel of each layer is 1*3. The data feature changes after convolution, as shown in Figure 4.

2.1.2. One-Dimensional Pooling Layer. A pooling operation is used to remove redundant information, capture the characteristics of the input, expand perception, speed up
calculation, and prevent overfitting. Pooling adds infinitely strong apriori information to the network. As it is invariant to a small amount of translation, pooling can greatly improve the statistical efficiency of the network [40]. The most common pooling operations are average pooling and maximum (max) pooling. In one-dimensional pooling, max pooling takes the largest element in the pool area, while average pooling takes the average value of the elements in the pool area.

The max pooling formula is as follows:
\[ y = \max(R) \]
where \( R \) is the corresponding pool area.

The average pooling formula is as follows:
\[ y = \frac{1}{N} \sum_{n=1}^{N} x_n, \]
where \( n \) is the number of elements in the corresponding pool area and \( x_n \) is the \( n \)th element in the pool area.

A schematic diagram of these pooling modes is shown in Figure 5.

2.1.3. Output Layer. The convoluted and pooled neurons are tiled into a one-dimensional feature vector, which is the input of the fully connected layer. The activation function of the fully connected layer is also the ReLU function. Finally, it is used for classification. The transformer fault classification problem is a multiclass classification problem that uses the softmax function. The output of the function is a real number between 0.0 and 1.0, and the sum of the output values of the softmax function is 1. The higher the probability of a certain category, the more likely it is to classify the samples into this category. It can achieve a better performance than other classifiers.

The detailed process is shown in Figure 6. The softmax function is expressed in the following equation:
\[ y_i = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}, \]
where \( x_i \) is the \( i \)th input feature of the softmax activation function (input \( x = (x_1, x_2, \ldots, x_n) \)) and \( y_i \) is the estimated probability distribution of observation \( x \) belonging to the \( i \)th class (output \( y = (y_1, y_2, \ldots, y_n) \)).

The softmax loss function corresponding to this model is the cross-entropy loss function:
\[ \text{loss} = -\frac{1}{N} \sum_{i=1}^{N} y_i' \log y_i, \]
where \( N \) is the number of training samples and \( y_i' \) is the expected output corresponding to the input sample \( x_i \), that is, the actual label of the input.

2.2. Regularization. Compared with traditional machine learning, deep learning is more prone to overfitting, which requires regularization to improve the network model. In this way, we can build a model that performs well in the training set and has strong generalizability, as shown in:

(1) Dropout [41] is a convenient and powerful tool. In the training process of each neural network, a certain number of parameters can be ignored in the training process. In this way, each neuron must perform well, reducing the complex coadaptation between neurons. The effect is the best when the sampling probability of the hidden node is 0.5. When the sampling probability is 0.5, the number of randomly generated network structures is the largest. Generally, packet loss is used only in the fully connected layer, not the convolutional layer or pooling layer. A dropout neural network model is shown in Figure 7.

(2) L2 parameter regularization [42] can avoid a weight matrix that is too large in the network layer and can reduce the complexity of the model. Therefore, it can make the model fit more reasonable and improve the interpretability of the model.
2.3. Adaptive Learning Rate Algorithm. Various learning rate algorithms have a significant impact on model optimization. At present, commonly used optimization algorithms are Momentum [43], RMSProp [44], AdaDelta [45], etc. In this paper, the network model uses the current mainstream algorithm Adam. In the Adam algorithm, the first-order moment of the gradient is directly incorporated into the momentum. Second, the Adam algorithm includes offset correction, which makes it more robust and have good processing ability for sparse data and noisy sample data. It is also suitable for dealing with nonstationary targets [40].

2.4. Transformer Fault Diagnosis Model Based on a 1DCNN and an Isomeric Feature. The IF-1DCNN model can fuse not only different feature spaces extracted from the same inspection/monitoring data but also different feature spaces extracted from different inspection/monitoring data. Therefore, the inspection/monitoring data that can reflect the operation status of the transformer can be selected. The former, for example, fuses different feature space information extracted from DGA data that can reflect the operation status of the transformer in different aspects, such as the feature space composed of a set of characteristic variables for the dissolved gas content, a
2.5. Feature Input Selection and Processing. When a transformer breaks down, it will produce CH₄, H₂, C₂H₂, C₂H₄, C₂H₆, CO₂, CO, and N₂ [25]. Considering the current research status of the diagnosis effect, five key gases of CH₄, H₂, C₂H₂, C₂H₄, and C₂H₆ can be selected as the research object.

Therefore, the characteristic engineering reference value is the sum of the five key gas volumes corresponding to the sample data, that is, the ratio of each gas concentration, namely, CH₄/S, H₂/S, C₂H₂/S, C₂H₄/S, and C₂H₆/S, where S is the total volume of the five key gases. The contents of various dissolved gases are converted into relative contents in the range of [0, 1], which can reduce the mutual exclusion between gases and provide different characteristic information. Moreover, to reduce the difference between different characteristic gas content values and make the gas content data obey the same distribution, the original DGA data are processed by maximum and minimum normalization.

When inputting data into the model, the two-feature data are fused into an isomeric feature. The structure of the isomeric feature is different from that of the mixed feature. The isomeric feature is a two-dimensional feature, while the mixed feature is a one-dimensional feature. Every one-dimensional feature in an isomeric feature represents the data information of its own feature space. Thus, the one-dimensional convolutional layer and the pooling layer can learn and extract the characteristics of the dissolved gas content and dissolved gas content ratio, respectively. In the process of one-dimensional convoluting and pooling, the data in different feature spaces are not affected by each other. In this way, the data in different feature spaces can be convoluted and pooled to extract the features of data in different feature spaces.

In summary, the characteristic information of the gas content of transformer samples is arranged as follows:

\[ C_j = [c_j^1, c_j^2, c_j^3, c_j^4, c_j^5], \hspace{1cm} j = [1, 2, \ldots, N], \hspace{1cm} \]  

where \( N \) is the number of samples (the same as below) and \([c_j^1, c_j^2, c_j^3, c_j^4, c_j^5]\) shows the five kinds of normalized gas content values of sample \( j \). Among them, \( c_j^i = \frac{x_j^i - x_{\text{min}}^i}{x_{\text{max}}^i - x_{\text{min}}^i}, \hspace{1cm} j = [1, 2, \ldots, N], \hspace{1cm} i = [1, 2, \ldots, 5], \hspace{1cm} \)

where \( x_j^i \) is the original content value of characteristic gas \( i \) in sample \( j \) and \( x_{\text{min}}^i \) and \( x_{\text{max}}^i \) are the minimum and maximum content values, respectively, of gas \( i \) in the training sample. In addition, we need to save the values of \( x_{\text{min}}^i \) and \( x_{\text{max}}^i \) in the training set and use the values to normalize the validation set samples before testing.

Then, the characteristic information of the gas content ratio of the transformer sample is arranged as follows:

\[ R_j = [r_j^1, r_j^2, r_j^3, r_j^4, r_j^5], \hspace{1cm} j = [1, 2, \ldots, N], \hspace{1cm} \]

where \([r_j^1, r_j^2, r_j^3, r_j^4, r_j^5]\) is the set of the 5 gas content ratio values for sample \( j \) and each gas content value corresponds to the content ratio value.

\[ r_j^i = \frac{c_j^i}{\sum_{i=1}^{5} c_j^i}, \hspace{1cm} j = [1, 2, \ldots, N], \hspace{1cm} i = [1, 2, \ldots, 5], \hspace{1cm} \]

\[ \sum_{i=1}^{5} c_j^i, \hspace{1cm} j = [1, 2, \ldots, N], \hspace{1cm} i = [1, 2, \ldots, 5]. \hspace{1cm} \]

The two characteristic data sets are connected to form a 2 * 5 isomeric feature, and the final arrangement of gas characteristic information of the transformer sample can be obtained:

\[ S_j = \begin{bmatrix} s_j^1 & s_j^2 & s_j^3 & s_j^4 & s_j^5 \\ s_j^6 & s_j^7 & s_j^8 & s_j^9 & s_j^{10} \end{bmatrix}, \hspace{1cm} j = [1, 2, \ldots, N], \hspace{1cm} \]

where

\[ s_j^i = c_j^i, \hspace{1cm} j = [1, 2, \ldots, N], \hspace{1cm} i = [1, 2, \ldots, 5], \hspace{1cm} \]

\[ s_j^i = r_j^{i-5}, \hspace{1cm} j = [1, 2, \ldots, N], \hspace{1cm} i = [6, 7, \ldots, 10]. \hspace{1cm} \]

The feature processing procedure is shown in Figure 9.
as the neural network training set, and 97 are used as the test set.

2.7. IF-1DCNN Diagnostic Model Architecture. The input characteristics of the IF-1DCNN model used in this paper are the characteristics of the gas content and gas ratio of the transformer, and the input layer feature dimension is $2 \times 5$. According to the characteristics of the input feature, this paper designs a 1DCNN transformer condition evaluation model with two convolutional layers; each convolutional layer is connected to a pooling layer, and the two layers are stacked to form the convolutional structure, and the diagnosis process of the IF-1DCNN is shown in Figure 10.

The IF-1DCNN is composed of one feature extraction layer and one classification layer. In the feature extraction layer, the first convolutional structure consists of a stack of one convolutional layer (Conv-1) and an average pooling layer (Pooling-1) with a $2 \times 1$ filter, and the second convolutional structure consists of a stack of one convolutional layer (Conv-2) and an average pooling layer (Pooling-2). Each convolutional layer is connected to a pooling layer, the two layers are stacked to form a network structure, and the two network structures make the model deeper, which helps to acquire good representations of the input signals and improve the performance of the network. The feature extraction layer extracts $2 \times 5$ low-level features into $2 \times 2$ high-level features with two convolutional layers and then inputs them into the classification layer for classification. The 5-dimensional information of a single feature space has been extracted as a 2-dimensional information, and more network layers are not needed. The classification layer consists of a full-connection layer and a final output layer. The details of the parameters of the IF-1DCNN used in the experiments

![Diagram of feature processing](image)

**Figure 8:** Isomeric of different feature spaces extracted from the same inspection/monitoring data.

![Diagram of feature processing](image)

**Figure 9:** Schematic diagram of feature processing.
are summarized in Table 1. In Table 1, the kernel size is noted as $D \times W \times H$, where $D$ indicates the channel size of kernels, $W$ indicates the width of the kernel, and $H$ indicates the height of the kernel.

2.8. Application Steps of the 1DCNN Diagnostic Model.

The transformer fault diagnosis model based on a CNN is shown in Figure 11. The specific application steps are as follows:

1. The gas content and gas content ratio of the transformer are selected as the characteristic parameters of the model
2. The fault type of the power transformer is coded
3. The feature input parameters are preprocessed
4. The sample data are divided into a training set and a validation set
5. The 1DCNN fault diagnosis model of the transformer is studied and tested

2.9. Example Analysis.

In this section, the proposed fault diagnosis method is compared with other models based on a CNN to verify the effectiveness of the method. The proposed method in this paper was written in Python and run on a desktop computer with an Intel Core i7-9750H CPU and a 16 GB RAM.

2.10. Visualization of the Network Learning Process.

The 1DCNN used in this paper has two CNN modules. Each CNN module has a convolutional layer for feature extraction

and then a pooling layer to further extract the most important features from the convolutional layer and reduce the feature dimension by half. During the model training and verification steps, the learning rate was set to 0.001, and the activation function was the ReLU function. In total, 500 rounds of training were conducted. The 1DCNN learns through the training samples and inputs the validation set into the model in the training process for verification.

Because the principle of a CNN is similar to a black box, the internal working principle is difficult to explain. Therefore, to investigate the potential mechanism and present the extracted features from each layer of the IF-CNN, the t-distributed stochastic neighbor embedding (t-SNE) method was used to visualize and understand the classification effect; this paper uses 3D spatial visualization, as shown in Figure 12.

Figure 12 shows the visualization of the data feature distribution of the input layer, convolutional layer 1, convolutional layer 2, dense layer, and softmax output layer. From Figure 12, we can see that the 6 states of the transformer are confusing, and the characteristics of most of the raw data samples are mixed with each other. In convolutional layer 2, the feature extraction ability of the layer is increased, and the distance between the features of different types becomes larger, while the distance between similar tags decreases; as a result, clustering is present, which shows that the CNN can effectively extract information related to category mapping. After the training of the whole connection layer, the same types of features are more obviously gathered together. Finally, after the softmax function is calculated, we can see that the features of the same label are clustered into one group. The 3D visualization of the classification results shows that the trained model has excellent feature extraction and nonlinear mapping abilities.

As seen from Figure 12, a few samples are still misclassified. Some faults are divided into different faults according to the same index. For example, if the temperature of the transformer fault is greater than 700°C, it is divided into T12 and T3 faults, which will lead to the index critical fault being easily misclassified. At the same time, it can be found from Figure 12 that the samples of different faults of the same nature are relatively close, while the samples of different faults of different natures are far apart, which conforms to the theoretical basis.

2.11. Comparison of Different Feature Information Processing Methods.

To verify the advantages of the proposed isomeric feature processing method, it needs to be compared with

| Layer          | Kernel size | Stride | Output size | Padding |
|----------------|-------------|--------|-------------|---------|
| Conv-1         | 1024 $\times$ 1 $\times$ 4 | 1 $\times$ 1 | 2 $\times$ 5 $\times$ 1024 | Yes     |
| Pooling-1      | 1 $\times$ 2 | 1 $\times$ 1 | 2 $\times$ 3 $\times$ 1024 | Yes     |
| Conv-2         | 1024 $\times$ 1 $\times$ 2 | 1 $\times$ 1 | 2 $\times$ 3 $\times$ 1024 | Yes     |
| Pooling-2      | 1 $\times$ 2 | 1 $\times$ 1 | 2 $\times$ 2 $\times$ 1024 | Yes     |
| Flatten        | 4096        | —      | 4096        | —       |
| Dense          | 150         | —      | 150         | —       |
| Softmax        | 6           | —      | 6           | —       |
Network structure optimization

- Initial network structure
- Identification of the network impact factors
- Optimization of the network parameters

Data preprocessing

- Raw data
  - Model input feature selection
  - Sample running status code
  - Sample feature preprocessing

Modeling and testing

- CNN construction
- CNN training
- The test set is input into the training CNN
- Effect of CNN fault diagnosis

Training set
Validation set

Figure 11: Diagnosis flow based on a multilayer CNN.

(a) (Convolutional layer 1) (Convolutional layer 2)

- Discharges of low energy
- Discharges of high energy
- Thermal faults > 700°C
- Thermal faults < 700°C
- Normal
- Partial discharges

Figure 12: Continued.
other feature information processing methods. The detailed workflow of other models is as follows:

F1-1DCNN: a single feature of the gas content of the transformer is taken as the input feature information. Since the input feature is a 1 × 5 feature, the 1DCNN model remains unchanged.

F2-1DCNN [39]: a single characteristic of the transformer gas ratio is input into the model. Similarly, the 1DCNN model remains unchanged.

F3-1DCNN: a mixed feature, which combines the characteristics of gas content and gas ratio, is input into the model. Compared with the proposed heterogeneous feature, the feature is one dimensional and has no heterogeneous processing. To be similar to the model in this paper, the size of the first convolution kernel L is changed to 1 × 9, the second convolution kernel is changed to 1 × 4, and the other parameters remain unchanged.

The above three network models and the proposed model are trained, and the best model is saved. The accuracies of the saved models were compared, and the results are shown in Figure 13.

From Figure 13, comparing the trained models for two kinds of single feature spatial data, we find that the 1DCNN model can better learn the classification features from the gas content ratio features to carry out classification. The performance of the network model is improved from 74.23% to 82.47% when the input feature is a mixed feature. This finding suggests that it is difficult to improve the accuracy of the 1DCNN model for transformer fault diagnosis with single feature information. The 1DCNN model can extract more information from the mixed features to distinguish the fault types and improve the accuracy. In this paper, the isomeric features are input into the model to further improve the accuracy of the model, and the accuracy of the model is improved to 86.59%, which proves the rationality of the proposed IF-CNN model.

To further verify the superiority of the IF-CNN model, the average epoch training time of each model and the test time are compared. The comparison results are shown in Figure 14.

It can be seen from Figure 14 that the performance of the first three models is very similar. The average epoch time and test time gap between the training and testing of the three models are very small. From Figures 13 and 14, the F3-DCNN model outperforms the F1-DCNN and F2-DCNN models, but the average epoch time and test time required are also substantially increased. Compared with the F3-CNN, the IF-CNN model is superior, and the time required for the model does not increase significantly. This finding shows that the proposed IF-CNN can greatly improve the diagnosis ability of the transformer convolution network.

2.12. Comparison of the Fault Diagnosis Accuracies of the Different Models. The proposed method and other machine learning models are used for fault diagnosis of power transformers and compared. Without changing the training set and validation set, the normalized transformer gas content feature (Feature 1), gas-ratio feature information (Feature 2), and mixed feature information (Feature 3) based on the transformer gas content feature...
and gas ratio are input into the traditional machine learning models, and then each model is tested on the validation set.

The machine learning models are genetic algorithm-extreme gradient boosting (GA-XGBoost) [23], particle swarm optimization-support vector machine (PSO-SVM), backpropagation neural network (BPNN), gradient boosted decision tree (GBDT), and DBN models. GA-XGBoost uses the genetic algorithm to optimize several parameters in the model. PSO-SVM uses the Gaussian radial basis function as the kernel function, and the search range of kernel function parameters $c$ and $G$ is determined by particle swarm optimization. The hidden-layer neural network structure of the BPNN is (1024-1024-512), the activation function is the ReLU function, the learning rate algorithm is Adam, the learning rate is 0.01, and the number of training cycles is 1000. The number of decision trees in the GBDT models is 100, the depth of each tree is 6, and the learning rate is 0.1. The structure of the DBN model designed in this paper is (256-256), with 50 pretraining epochs, and the pretraining learning rate is 0.05, the number of training times is 400, the learning rate is 0.1, and its activation function is also the ReLU function [30, 31]. Each model was run ten times. The highest accuracy on the validation set is recorded in Figure 15.

Figure 15 shows that the accuracy of models is higher, when each model is trained on Feature 2. This shows that classification models’ algorithms often need to design feature extractors. The final classification effect is closely related to whether the designed feature extractors can describe the classified objects well. Except for the GBDT and PSO-SVM models, the accuracies of the other models are further improved when trained on Feature 3. At the same time, it also shows that the mixed feature information enables most models to learn more fault classification features, which further improves the transformer fault diagnosis accuracy. However, sometimes, the efficiency of some models decreases because of the influence of the different data distributions of different features. From Figure 12 and previous studies, the transformer fault diagnosis accuracy can be effectively improved by establishing the isomeric feature of the transformer gas content and gas ratio and training a 1DCNN model.

To prove that this method can overcome the bad influence of different features space-isomeric feature data and the significance of the double-convolutional layer, the isomeric features of transformer samples are input into the best IF-CNN model, and the output of the first convolutional layer (conv1-out) and the second convolutional layer (conv2-out) of the model are taken as the input features. Both the input features and raw data are input into the GBDT for training and testing, and the accuracy of the validation samples is observed; the results are shown in Figure 16.

From the figure, we can observe that the accuracy of the model is greatly increased when conv1-out is used as the input, and the performance of the model is further improved when the input is conv2-out. At the same time, combined with Figure 15, the accuracy of both models is higher than the highest accuracy of the model under the single feature or mixed feature. This finding shows that the proposed method can effectively eliminate the adverse effects of the spatial data distribution of mixed features and improve the accuracy of the model. Moreover, the model has a better performance when conv2-out is used as the input than when conv1-out is used as the input. This finding shows that the double-convolutional layer can further extract the isomeric feature information, and the IF-CNN model design is reasonable. The conv2-out and raw data are input into the other models for training and testing, the dimension of conv2-out is reduced to 3-dimensions by t-SNE, and then input into the PSO-SVM model training, when the conv2-out is input into PSO-SVM. And, the accuracies of the validation samples are shown in Figure 17. The performance of the other models is improved to a certain extent, and the accuracy of the model has been improved. The results show that the proposed IF-CNN model can overcome the bad influence of different feature space data to a certain extent for improved accuracy.
3. Conclusion

To improve the transformer fault diagnosis accuracy, this paper proposes a 1DCNN transformer condition diagnosis method based on an isomeric feature. The following conclusions can be drawn from this study:

1. The characteristic information of transformer gas in different characteristic spaces is processed according to the method in this paper and then input into the model. The convolutional and pooling layers of the 1DCNN are used to process the characteristic information; the experimental result shows that this method can improve the performance of the CNN model.

2. Compared with transformer gas normalization, most models can learn better feature information from the gas content ratio data, which improves the classification effect of the models.

3. With the transformer fault sample data, the characteristic information of each gas is processed as described in this paper. The diagnostic accuracy of the transformer fault diagnosis method based on a 1DCNN is higher than that based on other machine learning methods.

4. The double-convolutional layer IF-CNN proposed in this paper can overcome the bad influence of different feature space data to a certain extent to improve the transformer fault classification accuracy.

The transformer state diagnosis model proposed in this paper provides a novel idea. The proposed method can be further extended to other power equipment and power system fault diagnosis tasks and has certain application prospects. However, there are many data closely related to transformer fault types, such as the gas production rate of different characteristic spaces of the same category, the electrical test data of transformers in different characteristic spaces of different categories, oil temperature data, and frequency response data. With the continuous improvement of data mining technology, it is possible to obtain more
accurate results for transformer fault diagnosis based on multidimensional data.

**Data Availability**

In this paper, from the recent relevant literatures and transformer fault databases, we collected DGA samples to verify the performance of the IF-1DCNN method.

**Conflicts of Interest**

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

**Acknowledgments**

This work was supported by the Science and Technology Project of the China Jilin Provincial Education Department during the 13th Five-Year Plan (Grant no. JJKH20200045KJ), China Jilin Province Science and Technology Development Plan Project (Grant no. 20190303038SF), and China Jilin Province Science and Technology Innovation Development Plan Project (Grant no. 20190302018). Thanks are due to the authors of the literature for providing the DGA sample data.

**References**

[1] C. Jettanase and A. Ngasopitakkul, "A novel probabilistic neural network-based algorithm for classifying internal fault in transformer windings," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 8, no. 2, pp. 123–131, 2013.

[2] J. Jia, F. Tao, G. Zhang, J. Shao, X. Zhang, and B. Wang, "Validity evaluation of transformer DGA online monitoring data in grid edge systems," *IEEE Access*, vol. 8, pp. 60759–60768, 2020.

[3] F. Yuan, J. Guo, Z. Xiao, B. Zeng, W. Zhu, and S. Huang, "A transformer fault diagnosis model based on chemical reaction optimization and twin support vector machine," *Energies*, vol. 12, no. 5, p. 960, 2019.

[4] M. H. Samimi, S. Tenbohlen, A. A. S. Akmal, and H. Mohseni, "Dismissing uncertainties in the FRA interpretation," *IEEE Transactions on Power Delivery*, vol. 33, no. 4, pp. 2041–2043, 2018.

[5] L. Ning, L. Jian, and Q. Chenghao, "Application of the time-varying outage model of electric transmission and transformation equipment in condition-based maintenance," in *IEEE PES Innovative Smart Grid Technologies*, pp. 1–5, IEEE, Tianjin, China, 2012.

[6] M. Gutten, D. Korenciak, M. Kucera, R. Janura, A. Glowacz, and E. Kantoch, "Frequency and time fault diagnosis methods of power transformers," *Measurement Science Review*, vol. 18, no. 4, pp. 162–167, 2018.

[7] R. Jia, Y. Xie, H. Wu, J. Dang, and K. Dong, "Power transformer partial discharge fault diagnosis based on multidimensional feature region," *Mathematical Problems in Engineering*, vol. 2016, Article ID 4835694, 11 pages, 2016.

[8] T. Hong, D. Deswal, and F. de Leon, "An online data-driven technique for the detection of transformer winding deformations," *IEEE Transactions on Power Delivery*, vol. 33, no. 2, pp. 600–609, 2018.

[9] J. Liu, Z. Zhao, C. Tang, C. Yao, C. Li, and S. Islam, "Classifying transformer winding deformation fault types and degrees using FRA based on support vector machine," *IEEE Access*, vol. 7, pp. 112494–112504, 2019.

[10] J. Faiz and M. Soleimani, "Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 25, no. 5, pp. 1798–1806, 2018.

[11] S. M. Ghoneim and I. B. M. Taha, "A new approach of DGA interpretation technique for transformer fault diagnosis," *International Journal of Electrical Power & Energy Systems*, vol. 81, pp. 265–274, 2016.

[12] L. Zhang, "Fault diagnosis of oil-immersed transformers using self-organization antibody network and immune operator," *Mathematical Problems in Engineering*, vol. 2014, Article ID 847623, 8 pages, 2014.

[13] W. H. Tang and Q. H. Wu, *Condition Monitoring and Assessment of Power Transformers Using Computational Intelligence*, Springer-Verlag, New York, NY, USA, 2011.

[14] Z. Yongli and Y. Jinliang, "Study on application of multi-kernel learning relevance vector machines in fault diagnosis of power transformers," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 33, no. 22, pp. 68–74, 2015.

[15] I. B. M. Taha, S. M. Ghoneim, and A. S. A. Duaywah, "Refining DGA methods of IEC Code and Rogers four ratios for transformer fault diagnosis," in *IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, Boston, MA, USA, 2016.

[16] Z.-X. Liu, B. Song, E.-W. Li, Y. Mao, and G.-L. Wang, "Study of "code absence" in the IEC three-ratio method of dissolved gas analysis," *IEEE Electrical Insulation Magazine*, vol. 31, no. 6, pp. 6–12, 2015.

[17] N. Bakar, A. Abu-Siada, and S. Islam, "A review of dissolved gas analysis measurement and interpretation techniques," *IEEE Electrical Insulation Magazine*, vol. 30, no. 3, pp. 39–49, 2014.

[18] C. Yan, M. Li, and W. Liu, "Transformer fault diagnosis based on BP-adaboost and PNN series connection," *Mathematical Problems in Engineering*, vol. 2019, Article ID 1019845, 10 pages, 2019.

[19] Y. Jia, L. Shi, and Y. Xin, "Transformer fault diagnosis based on wavelet neural network optimized by improved artificial fish swarm algorithm," *Journal of Henan University of Technology*, vol. 38, no. 2, pp. 103–109, 2019.

[20] M. D. Equbal, S. A. Khan, and T. Islam, "Transformer incipient fault diagnosis on the basis of energy-weighted DGA using an artificial neural network," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 26, pp. 77–88, 2018.

[21] Y. Wang and L. Zhang, "A combined fault diagnosis method for power transformer in big data environment," *Mathematical Problems in Engineering*, vol. 2017, Article ID 9670290, 6 pages, 2017.

[22] Y. Zhang, X. Li, H. Zheng et al., "A fault diagnosis model of power transformers based on dissolved gas analysis features selection and improved krill herd algorithm optimized support vector machine," *IEEE Access*, vol. 7, pp. 102803–102811, 2019.

[23] Y. Zhang, "Fault diagnosis method of oil immersed transformer based on genetic algorithm optimization of xgboost," *Electric Power Automation Equipment*, vol. 41, no. 2, pp. 200–206, 2020.

[24] S. Li, G. Wu, B. Gao, C. Hao, D. Xin, and X. Yin, "Interpretation of DGA for transformer fault diagnosis with complementary SaE-ELM and arc tangent transform," *IEEE
Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 1, pp. 586–595, 2016.

[25] E. Li, L. Wang, and B. Song, "Fault diagnosis of power transformers with membership degree," IEEE Access, vol. 7, pp. 28791–28798, 2019.

[26] S. C. Wang, C. X. Sun, and R. J. Liao, "Monitoring transformer chromatography with BPNN faults diagnosis method," Proceedings of the Chinese Society of Electrical Engineering, vol. 17, no. 5, pp. 35–38, 1997.

[27] S.-W. Fei, M.-J. Wang, Y.-B. Miao, J. Tu, and C.-L. Liu, "Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil," Energy Conversion and Management, vol. 50, no. 6, pp. 1604–1609, 2009.

[28] 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.

[29] Y. Song, D. Wu, W. Deng et al., "MPPCEDE: multi-population parallel co-evolutionary differential evolution for parameter optimization," Energy Conversion and Management, vol. 228, Article ID 113661.

[30] A. K. Mishra, S. R. Das, P. K. Ray, R. K. Mallick, A. Mohanty, and D. K. Mishra, "PSO-GWO optimized fractional order PID based hybrid shunt active power filter for power quality improvements," IEEE Access, vol. 8, pp. 74497–74512, 2020.

[31] Y. Song, D. Wu, W. Ali, X. Zhou, B. Zhang, and W. Deng, "Enhanced success history adaptive DE for parameter optimization of photovoltaic models," Complexity, vol. 2021, Article ID 6660115, 22 pages, 2021.

[32] J. Dai, H. Song, G. Sheng, and X. Jiang, "Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 24, no. 5, pp. 2828–2835, 2017.

[33] X. Shi, Y. Zhu, and X. Ning, "Transformer fault diagnosis based on deep auto-encoder network," Electric Power Automation Equipment, vol. 36, no. 4, pp. 122–126, 2016.

[34] J. Dai, H. Song, and Y. Yang, "Dissolved gas analysis of insulating oil for power transformer fault diagnosis based on ReLU-DBN," Power System Technology, vol. 42, no. 2, pp. 658–664, 2018.

[35] R. Shuangzhan, "Comparative study of deep belief network and typical neural network applied to dissolved gas analysis in oil," High Voltage Apparatus, vol. 56, no. 9, pp. 39–45, 2020.

[36] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 10, pp. 7319–7327, 2020.

[37] L. Hui, Z. Zhipan, and Z. Zhongwei, "Transformer fault diagnosis based on convolution neural network," Journal of Henan University of Technology, vol. 37, no. 185, pp. 123–128, 2018.

[38] J. Jinglong, Y. Tao, and W. Zijie, "Transformer fault diagnosis method based on convolution neural network," Electrical Measurement and Instrumentation, vol. 54, no. 13, pp. 62–67, 2017.

[39] P. Xiaodeng, "Convolution neural network diagnosis method for dissolved gases in power transformer oil," Journal of Liaoning University of Petroleum and Chemical Technology, vol. 40, no. 5, pp. 79–85, 2020.

[40] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016.

[41] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, 2012.

[42] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016.

[43] J. J. Dong, Y. Tao, and W. Zijie, "Transformer fault diagnosis method based on convolution neural network," Electrical Measurement and Instrumentation, vol. 54, no. 13, pp. 62–67, 2017.

[44] P. Xiao-deng, "Convolution neural network diagnosis method for dissolved gases in power transformer oil," Journal of Liaoning University of Petroleum and Chemical Technology, vol. 40, no. 5, pp. 79–85, 2020.

[45] T. Tieleman and G. Hinton, "Lecture 6.5-RMSprop: divide the gradient by a running average of its recent magnitude," COURSERA: Neural Networks for Machine Learning, vol. 4, no. 2, pp. 26–31, 2009.