CIT. Journal of Computing and Information Technology, Vol. 30, No. 4, December 2022, 223–238
doi: 10.20532/cit.2022.1005691

Construction of Transformer Fault Diagnosis and Prediction Model Based on Deep Learning

Xiaomeng Li

Department of Mechanical and Electrical Engineering, Weihai Ocean Vocational College, Rongchen, China

The current intelligent diagnosis and prediction methods for transformer faults are prone to low diagnostic accuracy and insufficient trend prediction ability when the fault categories are imbalanced. Therefore, a fault diagnosis and prediction model for transformers was constructed using a deep learning framework. The fault diagnosis model was constructed using a focus loss stacked sparse noise reduction autoencoder on the deep learning framework. The prediction model was constructed by fusing long and short term memory networks on the basis of tree structure Parzen optimization, and the two models were validated. The results obtained through validation of the diagnostic model indicate that, when the actual hidden layer is set to 3 and the quantity of neurons is 58, the model accuracy during training and testing reaches 97.5% and 92.5%, respectively. After adding 0.001 times the Gaussian white noise, the model accuracy was significantly lifted, so this study set the Gaussian noise coefficient to 0.001. In the comparison with baseline models, the actual classification ability of the research model samples is strong, significantly improving the fault diagnosis ability. In the validation of the prediction model, the three error index values of the research model in the single prediction step of CH4 concentration were 0.0699, 0.0540, and 0.8481%, respectively, and proved to be were lower than in the case of the baseline model. The three error values in the two-step prediction are 0.0194, 0.0161, and 0.6535%, which are also lower than in case of the baseline model. Overall, the diagnosis and prediction model proposed in this paper can provide real-time future numerical predictions of dissolved gas analysis and monitoring data in transformer oil. Furthermore, the research outlines the future development trend of monitoring and measurement through application of tensor flow deep learning framework in transformer fault diagnosis. The attained prediction results are innovative, and could well complete the purpose of actual transformer fault diagnosis and early warning.

ACM CCS (2012) Classification: Computing methodologies → Machine learning → Machine learning algorithms

Keywords: deep learning, transformer, fault, prediction model, TensorFlow

1. Introduction

The increasing demand for energy has come up with higher technical requirements for the safety, reliability, and stability of smart grids, as well as higher requirements for the quality of power equipment [1]. Transformers are important equipment in the actual operation of the power grid, located at the center of the grid. Therefore, the actual operating environment is relatively complex, often affected by various harsh operating conditions, and once a fault occurs, it will cause large-scale power outages [2]. In severe cases, safety accidents such as explosions and fires may occur. So, ensuring the safe and stable operation of transformers not only helps to ensure the safety of personnel in the distribution area, but also helps to avoid unnecessary maintenance costs and high equipment replacement costs [3]. On the basis of this, research on transformer fault diagnosis (TFD) and prediction technology has become particularly important. Wang *et al.* proposed a Gaussian process multi classification method on the ground of intelligent fault classification algorithms for fault diagnosis of oil immersed power transformers. This effectively improves

the effectiveness and stability of TFD while achieving multi class recognition pairs [4]. Zhou *et al.* proposed an oil immersed TFD model on the basis of the probability neural network optimized by the moth flame optimization algorithm for power TFD, which effectively improves the accuracy [5]. Kukker *et al.* used an intelligent genetic algorithm to adjust the fuzzy classifier for TFD in response to relevant issues in high-precision TFD. This improves classification accuracy while ensuring the healthy operation of the transformer [6]. From this, it can be concluded that these methods are difficult to adopt in the face of more complex operating environments, and therefore cannot guarantee safety in the event of a failure.

For oil-immersed transformers, the Dissolved Gas Analysis (DGA) method is usually used to diagnose and detect the actual fault type of the transformer [7]. The concentration of relevant gases dissolved in oil can effectively reflect the actual state of the current transformer, and monitoring the content of relevant gases can make judgments on the fault status and type of the transformer [8]. Ardi *et al.* addressed the issue of incomplete datasets when using DGA for TFD and utilized the Tertius algorithm to process the missing values in the dataset. This approach effectively improves the accuracy of monitoring and diagnosing power transformer faults [9]. Menezes *et al.* proposed an inductive decision tree method on the ground of computational intelligence to address the related issues of traditional DGA analysis in TFD. This approach improves the accuracy of the diagnostic model on the basis of effectively extracting more information [10]. Zheng *et al.* conducted in-depth analysis on the actual generation pathway of DGA main products in transformer thermal fault analysis using palm oil as raw material for its insulation liquid. This provides theoretical guidance for transformer thermal fault diagnosis [11]. From it, it can be seen that some of these methods lack the concept of quantity and rely on expert experience, while others lack coding and do not involve all coding combinations. The diagnostic accuracy and convergence speed of other models make it difficult to meet the current practical needs.

Overall, traditional methods have the problem of incomplete coding and large judgment boundaries, resulting in low diagnostic accuracy. Many diagnostic methods that integrate intelligent algorithms also have problems with low diagnostic accuracy and insufficient trend prediction ability when the fault categories are imbalanced. The reason is that the actual transformer fault diagnosis is a multi classification problem. Faced with such problems, existing intelligent algorithms have cumbersome parameter settings and the process of constructing classifiers is cumbersome. Therefore, the method of constructing a TFD and prediction model (TFD-PM) using the TensorFlow deep learning (DL) framework in this study is innovative.

2. Analysis of TFD-PM on the Ground of DL

The traditional TFD and prediction methods, based on the characteristics of dissolved gases in oil, have significant limitations when dealing with imbalanced sample data. Thus, this section mainly constructs a new TFD-PM on the ground of the TensorFlow DL framework.

2.1. TFD Method on the Ground of Sparse Noise Reduction Autoencoder

In response to the problems of low diagnostic accuracy and insufficient trend prediction ability of current intelligent diagnosis and prediction methods for transformer faults when the fault categories are imbalanced, this paper constructed a TFD-PM using the TensorFlow DL framework. The commonly used DL frameworks currently have their own shortcomings. TensorFlow is an open-source framework developed by Google that utilizes the idea of data flow graphs for large-scale distributed numerical computation. "Tensor" represents multidimensional vectors, and "flow" represents calculations using data flow graphs. Compared to other DL frameworks, TensorFlow itself has a high degree of flexibility and consistency. Meanwhile, it greatly simplifies the actual model training process due to its internal automatic differentiation mechanism. Therefore, this study uses it as the basic framework for subsequent models [12–14].

Among them, the TFD method mainly uses focal loss stacked sparse noise reduction auto-

encoder (FLS-SNRA) to diagnose transformer faults. It utilizes an incomplete autoencoder learning to encode the data distribution, making it smaller than the input. If an encoder is too large, it cannot be used to obtain any useful information. Regular autoencoders can enable the model to learn other features by utilizing a loss function. When the actual amount of hidden nodes outperforms the actual number of nodes in the input layer of the autoencoder, it is possible to artificially limit the hidden layer [15]. This study adds a penalty factor to the loss function to restrict the actual sparsity of the sparse autoencoder (SAE) on the bridge, thus forming the SAE. Even if the actual capacity of the diagnostic model is large enough to learn many meaningless identity functions, it can still obtain some practical details of the data distribution. The expression of the relevant cost function of SAE is shown in equation (1).

$$\begin{cases} H_{sparse}(M, c) = H(M, c) + \alpha \sum_{i=1}^{n} K\left(\gamma \| \hat{\gamma}_i\right) \\ H(M, c) = \frac{1}{W} \sum_{j=1}^{W} \left(\hat{q}_j - q_j\right)^2 \\ K\left(\gamma \| \hat{\gamma}_i\right) = \gamma \log \frac{\gamma}{\hat{\gamma}_i} + (1-\gamma) \log \frac{1-\gamma}{1-\hat{\gamma}_i} \end{cases} \quad (1)$$

In equation (1), $H(M, c)$ represents the MSE; α represents the sparse penalty coefficient, usually set to 0.3; n represents the quantity of hidden layer neurons (HLN); $K\left(\gamma \| \hat{\gamma}_i\right)$ represents Kullback Leibler divergence (KL); γ represents the sparsity parameter, usually set to 0.05 or 0.1, while $\hat{\gamma}_i$ represents the average activation level of HLN i for all training data; W represents the total number of HLN; q represents the actual output value; and $\hat{q}$ represents the predicted output value [16]. SAE is essentially a shallow learning model, so to obtain deeper input data and more informative features, this study uses a stack structure to stack the layers and construct a stacked coefficient autoencoder. On this basis, considering the actual impact of noise, this study added noise to the self encoder to form a Stacked Sparse Denoising Auto Encoder (SSDAE). The schematic diagram of the structure of the stacked type sparse noise reduction autoencoder is shown in Figure 1. From Figure 1, it can be seen that the autoencoder network is divided into two stages, namely unsupervised pre training and supervised fine tuning. The former utilizes unlabeled sample data and the first line of equation (1), and adopts a layer by layer greedy training strategy and a backpropagation algorithm to sequentially train the network parameters of each layer of SSDAE. The latter removes the decoding layer of SSDAE and adds a Softmax classification layer. Based on the cross entropy loss function, the backpropagation algorithm is used to optimize the network parameters of each layer. Generally speaking, the way to increase noise usually involves adding small random fluctuations and randomly assigning a partial component to zero. The former is generally Gaussian white noise (GWN), as expressed in Equation (2).

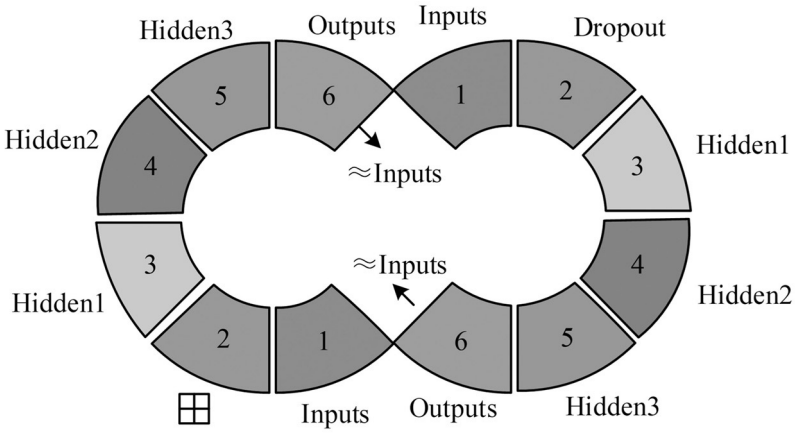


Figure 1. Schematic diagram of the structure of the stacked type sparse noise reduction autoencoder.

$$\tilde{p} = p + \varphi\psi \tag{2}$$

In equation (2), $\tilde{p}$ represents the data after adding noise; p represents raw data; φ represents the coefficient; ψ represents a random number that follows a normal distribution. SSDAE includes unsupervised pre-training (UPT) and supervised fine-tuning (SFT) in actual training, which removes the decoding layer of SSDAE and adds a maximum soft value (Softmax) classification layer [17]. On the basis of the cross-entropy loss function (CELF), backpropagation algorithm is utilized to perfect the actual parameters of each layer of the network. The expression of the CELF is shown in equation (3).

$$O = \sum_{i'}^{N} q_{i'} \log \hat{q}_{i'} \tag{3}$$

In equation (3), O represents the value of the CELF; q represents the true label; $\hat{q}$ represents the probability of predicting labels. However, when using the CELF in transformer fault data samples, it will shift the diagnostic model towards categories with more samples. Therefore, this study uses the focal loss function (FLF), which is expressed in equation (4).

$$F = -\beta\left(1-\hat{q}\right)^{\kappa} \log\left(\hat{q}\right) \tag{4}$$

In equation (4), F represents the FLF value; β is the balance parameter value; and κ represents the focusing parameter. This study uses the FLF to replace the second line equation in equation (1), reducing the loss weight corresponding to categories with more samples and increasing the loss weight corresponding to categories with fewer samples. This allows the model to focus more on categories with fewer samples, and can rise the model's precision in TFD. Therefore, the TFD process using FLS-SNRA is shown in Figure 1.

Figure 2 shows that the process mainly includes five processes: *i.e.* data preprocessing, model parameter setting (MPS), UPT, SFT, and outputting classification results. Among them, data preprocessing mainly involves adding Gaussian white noise to the input vector after data cleaning, which includes outlier handling and missing value filling. The model parameter setting mainly involves setting SSDAE parameters. Finally, the diagnostic results will be output after completing the SSDAE model training. Unsupervised pre training mainly involves reconstructing the initial values of the underlying SAE parameters of the input pre training network. Supervised fine-tuning mainly involves discarding the decoding layer of the former and adding a softmax classifier. After training, the parameters of each SAE layer are adjusted. In order to accelerate the training speed, the adaptive learning rate adjustment coefficient is set to 0.5 and the maximum number of iterations is 1000. Due to the fact that the transformer used in the experiment is an oil-immersed transformer, five special gases, namely hydrogen

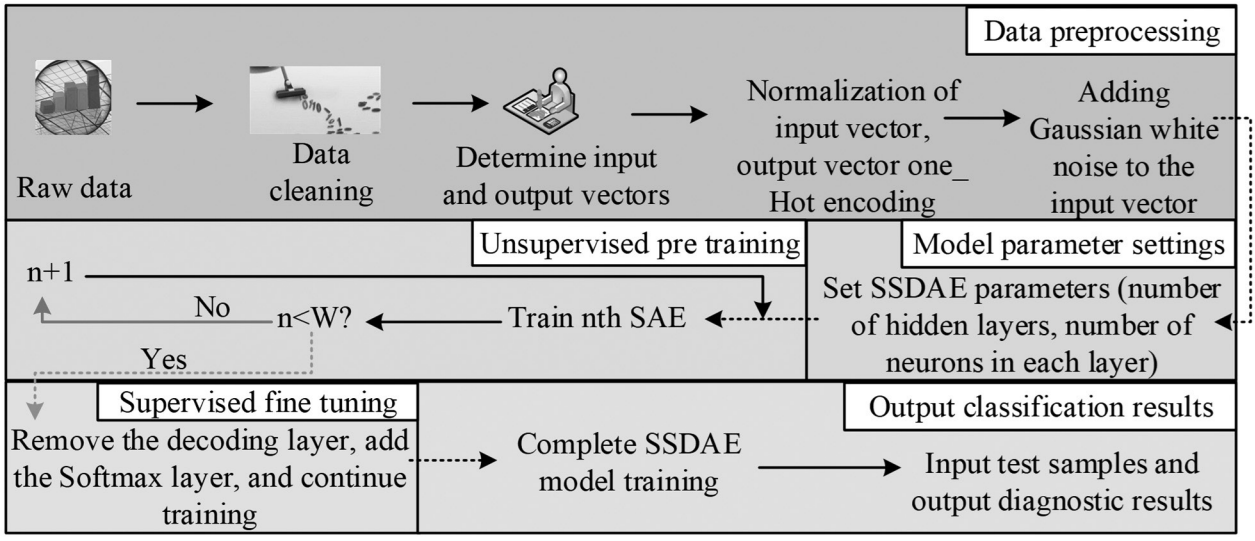


Figure 2. TFD process using FLS-SNRA.

(H2), methane (CH4), ethane (C2H6), ethylene (C2H4), and acetylene (C2H2), are usually used to determine the type of fault in the transformer. Therefore, the basic architecture of the TFD model is obtained by applying Figure 1 to actual experiments shown in Figure 2.

Figure 3 shows that the TFD model using a stacked sparse denoising autoencoder is firstly an UPT method that reconstructs the input SAE parameters. Secondly, there is a SFT that removes the decoding layer of the first step, and then adds a softmax classifier to each SAE layer to adjust the parameter through training.

2.2. Construction of Transformer Fault Prediction Model on the basis of TPE Algorithm

On the basis of TFD, to achieve early warning of transformer faults, this study proposes a Long Short-Term Memory (LSTM) oil dissolved gas concentration prediction method based on the Tree Structured Parzen Estimator (TPE). This model predicts the characteristic gas concentration values of transformers for events in the future with aim of achieving fault diagnosis. In DL algorithms, Bayesian optimization algorithms use the approximation of the posterior distribution of unknown objective functions to perform hyperparametric optimization, which can significantly lift the efficiency of search algorithms [18]. Therefore, this study comprehensively considers the use of Bayesian optimization form for parameter adjustment of TPE to optimize the hyperparameters of LSTM networks. Among them, the actual criterion for TPE selection is expected improvement, which is defined as the expectation that the value will be less than a certain threshold, and the relevant expression is shown in equation (5).

$$E_{q^*}(p') = \int_{-\infty}^{\infty} \max(q^* - q', 0)\, x(q' \mid p')\, dq' \quad (5)$$

In equation (5), E represents the actual standard for TPE selection; q^* represents the threshold; p' represents the input value; q' represents the output value; x represents the defined value of the total probability density, and the two probability densities below it are expressed as shown in equation (6).

$$x(q' \mid p') = \begin{cases} h(p'), & q' < q^* \\ s(p'), & q' \geq q^* \end{cases} \quad (6)$$

In equation (6), $h(p')$ and $s(p')$ both represent probability density. The TPE selection method is to select a value at a certain quantile δ of $\{q^{i*}\}$ as the threshold q^*, that is, $x(q < q^*) = \delta$. Due to the inability to directly obtain $x(q' \mid p')$, corresponding changes were made using the Bayesian formula, as expressed in equation (7).

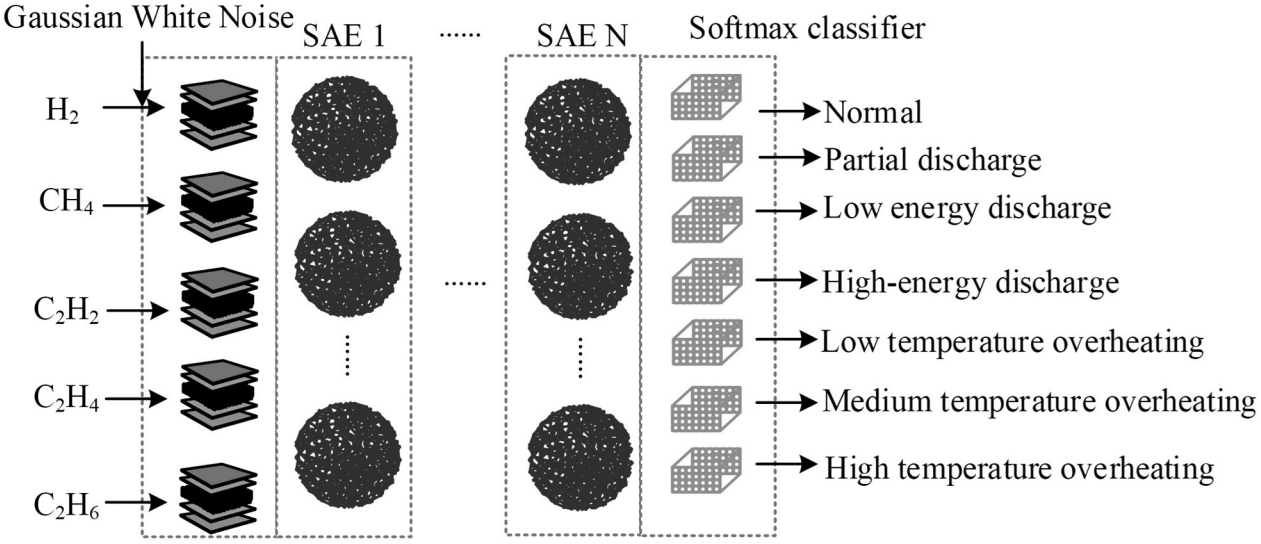


Figure 3. Basic architecture of TFD model using FLS-SNRA.

$$\begin{aligned} E_{q^*}(p') &= \int_{-\infty}^{\infty} \max\left(q^* - q', 0\right) x\left(q' \mid p'\right) dq' \\ &= \int_{-\infty}^{q^*} \max\left(q^* - q', 0\right) x\left(q' \mid p'\right) dq' \\ &= \int_{-\infty}^{q^*} \max\left(q^* - q', 0\right) \frac{x\left(p' \mid q'\right) x(q)}{x(p')} dq' \end{aligned} \tag{7}$$

According to $x\left(q < q^*\right) = \delta$ and

$$\begin{aligned} x(p') &= \int x\left(p' \mid q'\right) x(p') dq' \\ &= \delta h(p') + (1-\delta) s(p'), \end{aligned}$$

the corresponding transformation is made, and its expression is shown in equation (8).

$$\begin{aligned} E_{q^*}(p') &= \int_{-\infty}^{q^*} \max\left(q^* - q', 0\right) \frac{x\left(p' \mid q'\right) x(q')}{x(p')} dq' \\ &= \frac{\int_{-\infty}^{q^*} \max\left(q^* - q', 0\right) x\left(p' \mid q'\right) dq'}{x(p')} \\ &= \frac{h(p')^* \int_{-\infty}^{q^*} \max\left(q^* - q', 0\right) x(q') dq'}{x(p')} \\ &= \frac{h(p') q^* - h(p') \int_{-\infty}^{q^*} x(q') dq'}{\delta h(p') + (1-\delta) s(p')} \end{aligned} \tag{8}$$

In equation (8), δ represents the quantile. According to equation (8), it can be seen that using $h(p')$ to calculate the probability is higher, while using $s(p')$ to solve for p' with a lower probability will increase the value of E. Based on TPE, this study focuses on the problem of traditional neural networks being hindered in model weight optimization due to gradient vanishing and chooses to use LSTMs to predict the trend of transformer fault characteristic gases. The schematic diagram of the internal structure of the LSTM memory unit is shown in Figure 4.

From Figure 4, it can be seen that LSTM mainly avoids the problem of gradient vanishing by adding a gating system and uses this module to control useful information and omit useless information with the aim to solve the long-term dependency problem. The most important aspect of an LSTM network is the actual state of the cells, which contains all the information learned by the network at the current point in time. For the gate control system, generally an S-type nonlinear activation function is utilized, and the internal forgetting gate calculation expression is shown in equation (9).

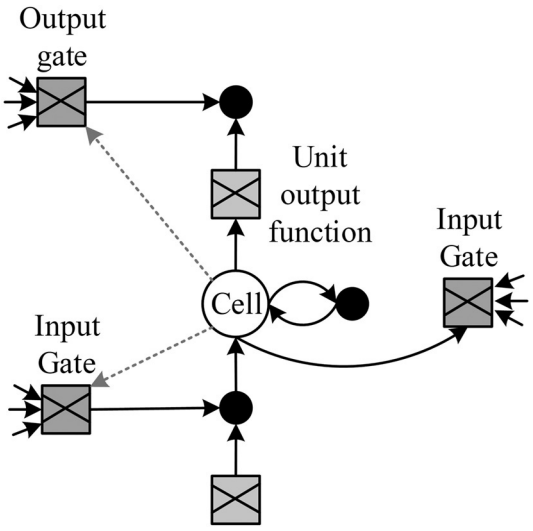


Figure 4. Schematic diagram of the internal structure of LSTM memory unit.

$$f^{(\tau)} = \rho\left(\varpi_f g^{(\tau-1)} + \xi_f p''^{(\tau)} + e_f\right) \tag{9}$$

In equation (9), $f^{(\tau)}$ represents the forgetting gate activation vector at time τ; ρ represents a vector with element values (0,1); ϖ_f and ξ_f represent the parameter matrices of forgetting gates; p'' represents the input vector; e represents the bias matrix. The expression of input gate control is shown in equation (10).

$$\begin{cases} z^{(\tau)} = \rho\left(\varpi_z g^{(\tau-1)} + \xi_z p''^{(\tau)} + e_z\right) \\ a^{(\tau)} = \tanh\left(\varpi_a g^{(\tau-1)} + \xi_a p''^{(\tau)} + e_a\right) \end{cases} \tag{10}$$

In equation (10), z represents the activation vector of the input gate; a represents a new candidate value vector; ϖ_z, ξ_z, ϖ_a, and ξ_a represent the parameter matrix of the input gate. It updates the cell state determined by the input gate and forgetting gate in LSTM on the basis of equations (9) and (10), and its expression is shown in equation (11).

$$B^{(\tau)} = f^{(\tau)} * B^{(\tau-1)} + z^{(\tau)} * a^{(\tau)} \tag{11}$$

In equation (11), B represents the cell state. In addition, the expression of output gate control is shown in equation (12).

$$v^{(\tau)} = \rho\left(\varpi_v g^{(\tau-1)} + \xi_v p''^{(\tau)} + e_v\right) \tag{12}$$

In equation (12), v represents the activation vector of the output gate; ϖ_v and ξ_v represent the parameter matrix of the output gate. Similarly, the expression after updating the cell state is shown in equation (13).

$$g^{(\tau)} = v^{(\tau)} * \tanh\left(B^{(\tau)}\right) \tag{13}$$

In equation (13), $g^{(\tau)}$ represents the output vector. Therefore, the actual predicted output is expressed as shown in equation (14).

$$\hat{q}^{(\tau)} = \rho\left(\theta g^{(\tau)} + b\right) \tag{14}$$

In equation (14), $\hat{q}^{(\tau)}$ represents the predicted output value; θ represents the weight matrix; b represents the bias vector. On the basis of this, the process of predicting the dissolved gas concentration in LSTM oil using TPE algorithm is shown in Figure 5.

Figure 5 exhibits the standardizing raw data process of LSTM and divides it into training and testing sets. Secondly, the training set is introduced into the LSTM, and a dropout layer is introduced to avoid overfitting of the model. Finally, the concentration prediction results of LSTM are output on the entire connected layer [19]. In the set parameter search space, the TPE algorithm is used to continuously execute the entire process until the model reaches its optimal performance. Among them, standardizing the raw data of LSTM essentially converts various dissolved gas contents into relative contents within the range of [0, 1]. Therefore, the expression of the standardization method is shown in equation (15).

$$u_l' = \frac{u_l - u_{l\max}}{u_{l\max} - u_{l\min}} \tag{15}$$

In equation (15), u_l' represents the concentration data of the l-th gas after standardization, while u_l represents the original concentration data; $u_{l\min}$ and $u_{l\max}$ represent the minimum and maximum concentrations of gas type l.

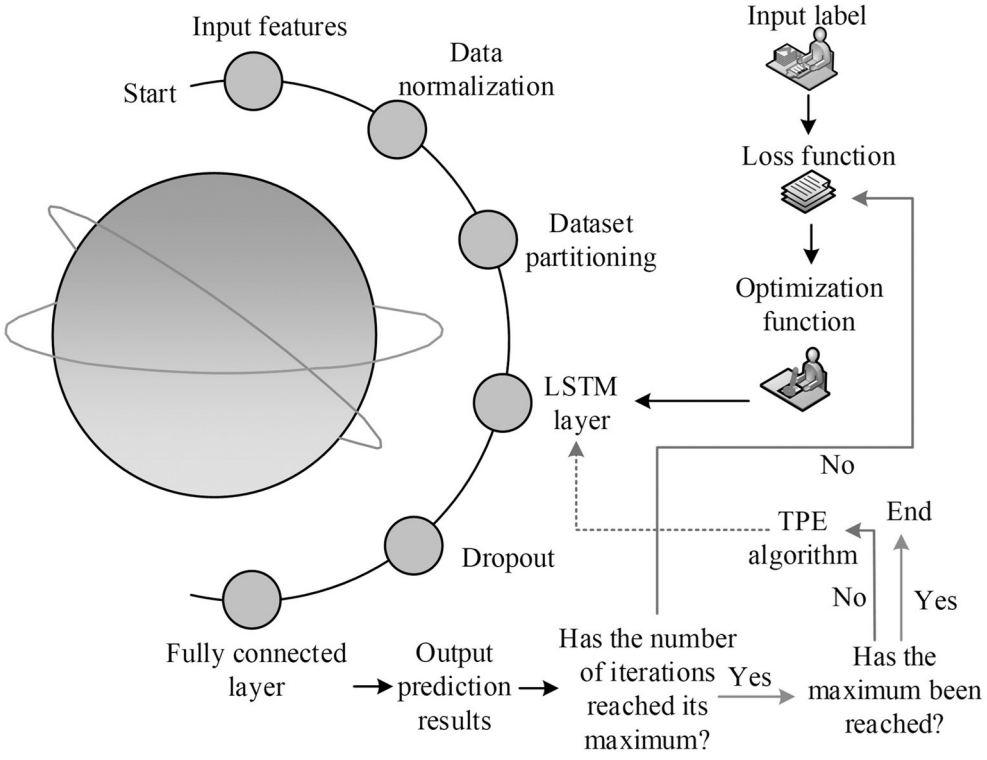


Figure 5. Process flow of the LSTM oil-dissolved gas concentration prediction model using TPE algorithm.

3. Performance Verification of TFD-PM

The validation of TFD-PMs can not only prove their performance but also lay a numerical theoretical foundation for practical applications. Therefore, this study mainly verifies the two research models in practice.

3.1. Performance Verification of TFD Model

To verify the application performance of the constructed TFD-PM, two models were experimentally validated in this study. In the analysis of transformer fault diagnosis models, the research takes the DGA fault data of a certain power supply company and 370 sets of transformer fault diagnosis datasets publicly available in China as research data. Data collection is mainly conducted online, supplemented by offline collection. Among them, there are a total of 1518 sets of samples, and the typical sample data content of different states is shown in Figure 6.

In Figure 6, A~G represent normal, partial discharge, low energy discharge (LED), high-energy discharge, low-temperature overheating, medium temperature overheating, and high-temperature overheating (HTO) states. Figure 6 shows that the highest concentrations of C2H6 are 16.55μL/L, 79.30μL/L, and 260.00μL/L under normal, LED, and HTO conditions, respectively. The highest concentration of CH4 is 20.00μL/L under partial discharge conditions. In practice, there are missing and duplicate values in this type of data, so research is conducted to handle outliers and fill in the missing values [20]. Among them, the nuclear density distribution of the original characteristic gas content and the density distribution after preprocessing are shown in Figure 7.

Figure 7 shows that there is only a small amount of data on the right side and the values are much larger than those on the left, so the data on the right is processed. The loss value obtained by using the random forest filling method is about 10, which is much lower than for other methods, that is, using this method for missing value data processing. Therefore, after detecting outliers and filling in missing values, the data is more concentrated, roughly distributed between -250 and 500, indicating that the data preprocessing effect is good. Therefore, the obtained characteristic gas concentration values are used as inputs, fault type codes are used as outputs, and the relevant dataset is modeled and analyzed using Python 3.8 and TensorFlow DL frameworks. The coding of transformer operation status and statistical content of fault samples are shown in Table 1.

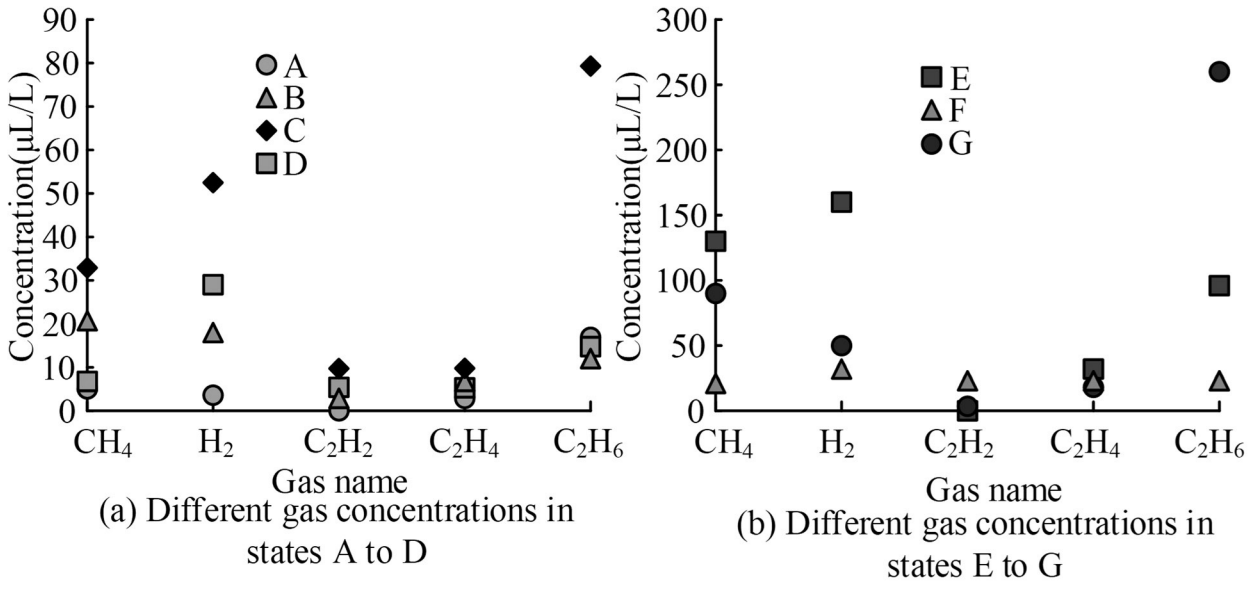


Figure 6. Typical sample data content in different states.

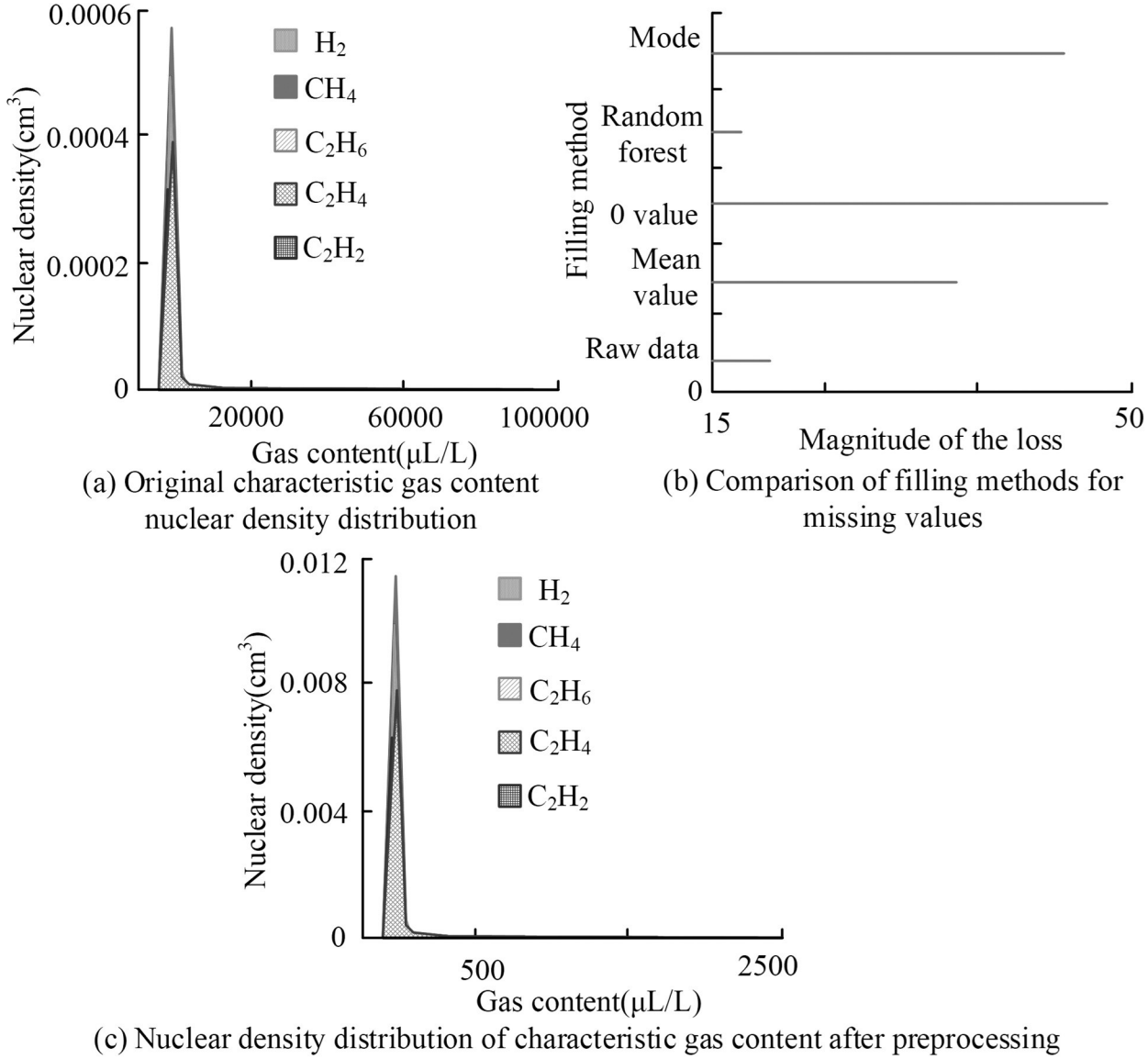


Figure 7. The nuclear density distribution of the original characteristic gas content and the density distribution after preprocessing.

Table 1. Encoding of transformer operation status and statistical content of fault samples.

Transformer operation status code							
	A	**B**	**C**	**D**	**E**	**F**	**G**
Status code	(1, 0, 0, 0, 0, 0, 0)	(0, 1, 0, 0, 0, 0, 0)	(0, 0, 1, 0, 0, 0, 0)	(0, 0, 0, 1, 0, 0, 0)	(0, 0, 0, 0, 1, 0, 0)	(0, 0, 0, 0, 0, 1, 0)	(0, 0, 0, 0, 0, 0, 1)
Fault sample statistics							
	A	**B**	**C**	**D**	**E**	**F**	**G**
H	275	34	73	109	72	51	85
I	71	11	21	30	21	15	24

In Table 1, H and I represent the numbers of training and test samples. Table 1 shows that when the transformer is in a certain state, it is directly labeled as "1", and the other states are all marked as "0". On this basis, due to the impact of HLN and GWN on network performance and model precision (the study set the hidden layers to 3), the study analyzed both effects. Meanwhile, to clarify the SSDAE diagnostic performance, this study introduced the three ratio method (1), SVM (2), decision tree (3), random forest (4), BPNN (5), stacked SAE+CELF (6), and SSDAE+CELF (7) to compare with the SSDAE+FLF (8) model studied. We perform an ablation experiment. Thus, the results of the three experiments are shown in Figure 8.

Figure 8 shows that in the HLN experiment, when the real hidden layer is 3 and the neurons is 58, the accuracy of model during training and testing reaches the maximum value, specifically reaching 97.5% and 92.5%, respectively. In addition, in the GWN experiment, the accuracy has greatly risen after adding 0.001 times the GWN. However, the model accuracy significantly decreased after exceeding the level of 0.001 times the GWN. Therefore, this study set the Gaussian noise coefficient to 0.001. Meanwhile, in the model comparison under this setting, the correct number of samples for the research model was as high as 167, and the fault diagnosis accuracy was as high as 93.30%. This result indicates the superiority of the constructed diagnostic model. To further verify its performance, the confusion matrix results of different diagnostic methods were analyzed. The confusion matrix results of the 8 diagnostic methods are shown in Figure 9.

Figure 9 shows that method 1 has weaker actual classification ability for samples of the same type, such as incorrect sample sizes of 5, 3, and 5 in states 5, 6, and 7. However, methods 2, 3, 4, and 5 have improved performance compared to method 1, but are still lower than

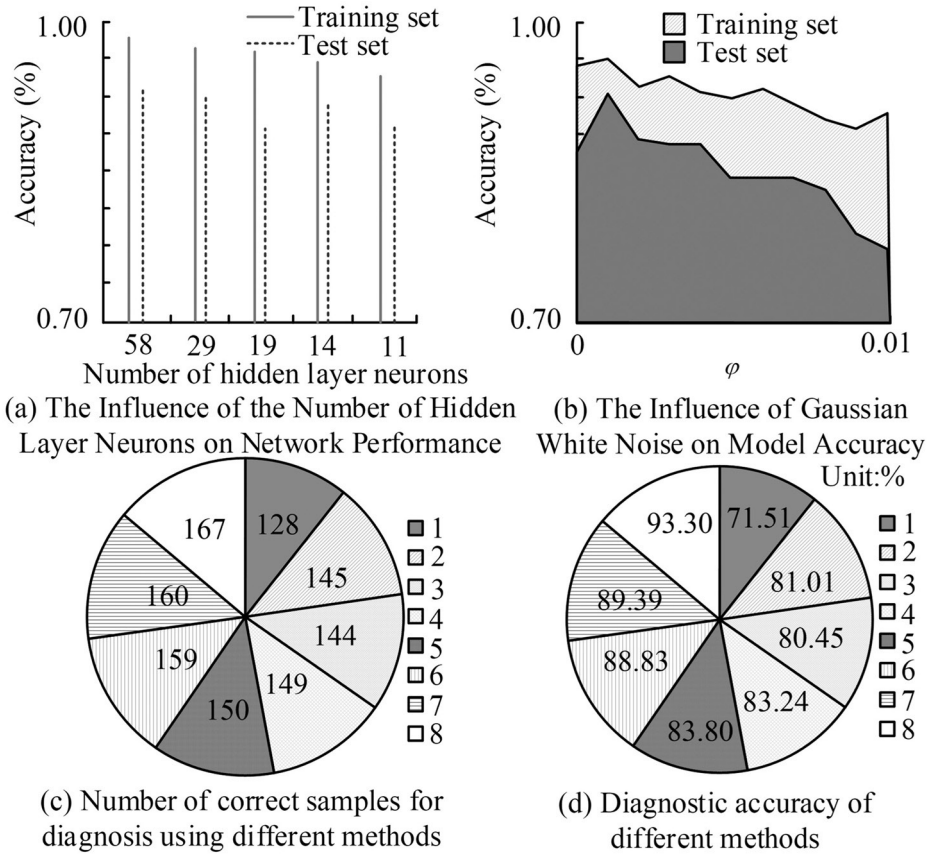


Figure 8. Three experimental results on the SSDAE model.

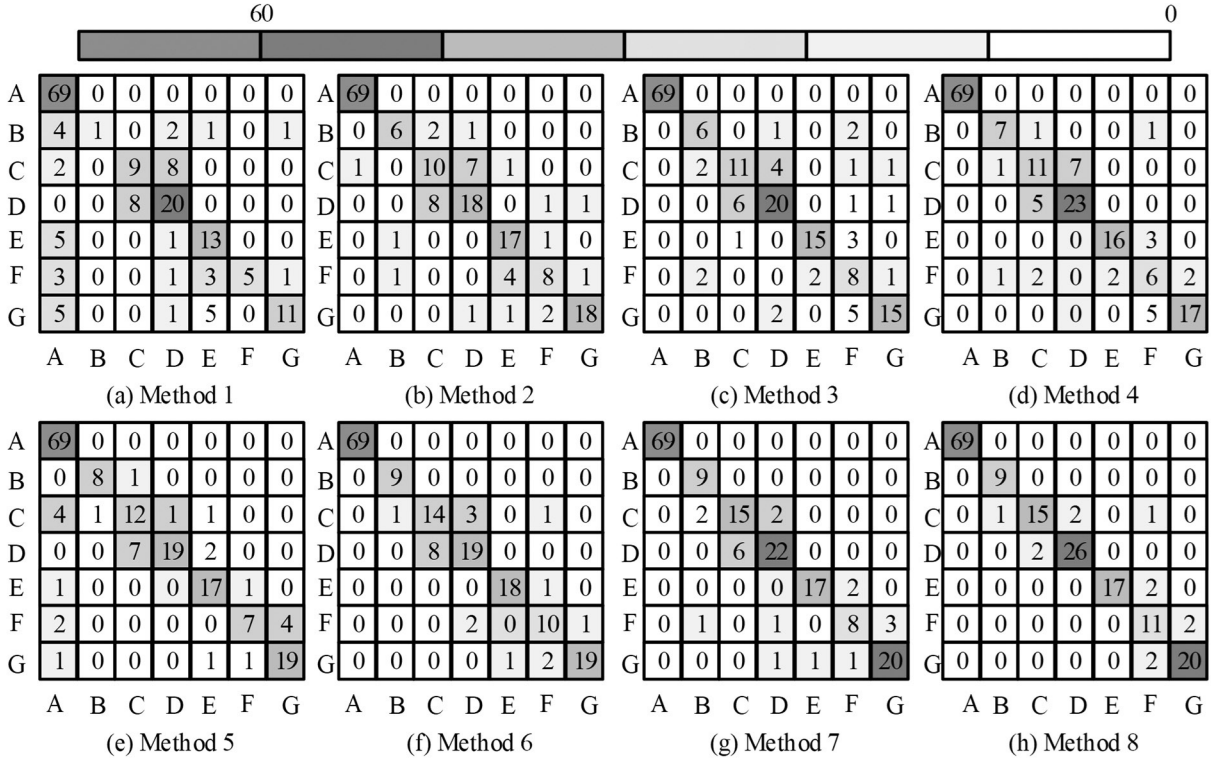


Figure 9. Confusion Matrix Results of 8 TFD Methods.

the research model. Comparing Figures f and g, it can be seen that adding GWN to the original data can effectively improve the diagnostic performance, but the increase is not significant. However, after comparing Figures g and h, it can be found that using the FLF instead of the CELF can effectively improve the diagnostic accuracy of the model. Overall, this study proposes a model of SSDAE+focal loss with high diagnostic effectiveness and practicality. The three ratio method has fewer fault codes, and some faults are difficult to diagnose using the three ratio method. Moreover, the judgment criteria of the three ratio method are too absolute, resulting in lower accuracy of the three ratio method. Adding Gaussian white noise to the input vector of the stacked sparse autoencoder can avoid overfitting of the model, increase its generalization ability, and thus the accuracy of the SSDAE model is higher than that of SSAE. Imbalanced training samples can easily lead to model bias, which in turn affects the model's diagnostic performance in the term of faults. In the fault diagnosis of transformer, sample imbalance cannot be avoided. Therefore, the study abandons the cross-entropy loss function of traditional classification models and adds the focal loss function, which has a significantly better diagnostic ability for small samples when compared to other models.

3.2. Performance Verification of Transformer Fault Prediction Model

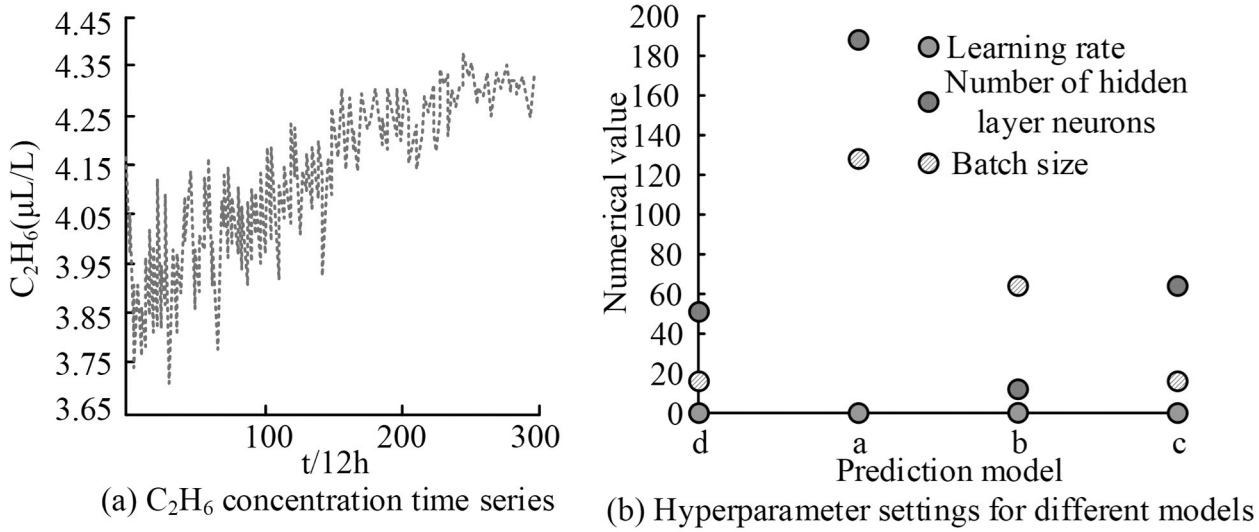
On the basis of verifying the performance of diagnostic models, we further validate the performance of predictive models. In the experiment, the actual data collected from the 500kV transformer oil related online monitoring device of a certain power supply company was used as the experimental data, and the dataset was modeled and analyzed using the Ubuntu 20.04 operating system, Python 3.8 programming language, and TensorFlow DL framework [21]. The experimental data was collected from November 26, 2022 to April 20, 2023, with a sampling frequency of 12 hours. Regarding the dataset split, the data before March 22, 2023 was the training set sample, and the subsequent data was the test set sample. C2H6 was used as an example for

predictive analysis. In addition, this study introduced artificial neural networks (a), recurrent neural networks (b), and gated recurrent units (c) to compare with the studied LSTM (d). On the basis of this, the time series of C2H6 concentration and the hyperparameter content of the four prediction models are shown in Figure 10.

Figure 10 shows that the overall concentration of C2H6 has a fluctuating growth, maintaining values between 3.7μL/L and 4.4μL/L over 300 cycles. In addition, the learning rate of the optimized model hyperparameters is set to 0.01, the HLNs are 51, and the batch size is 16. The three parameters of the comparative model are maintained between 0.01~0.1, 12~188, and 16~128, respectively. On the basis of this, the single step prediction and error comparison results of the four models are shown in Figure 10.

Figure 11 shows that in the one-step prediction results for C2H6 concentration, the LSTM prediction results are closer to the actual results, with an error of no more than 0.003. In the error comparison results, the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) of LSTM were 0.0053, 0.0048, and 0.7659%, respectively, which were lower than for the control model. Overall, the LSTM prediction model can effectively capture the time series correlation of historical monitoring data of transformer characteristic gases and can accurately learn the trend of changes in characteristic gas concentration. Similarly, the results of predicting other gases using the same method are shown in Figure 12.

In Figure 12, carbon monoxide (CO) was added to the original five characteristic gases. Figure 12 shows that in the prediction of H2 characteristic gas, the three error index values of LSTM are 0.0562, 0.0465, and 6.8565%, respectively. The three error indicators in CH4 prediction have values of 0.0699, 0.0540, and 0.8481%, respectively. The predicted values in C2H4 are 0.0257, 0.0217, and 3.0752%, respectively. The values in CO are 0.6700, 0.5311, and 1.1320%, respectively. The values of CO2 are 3.3622, 2.4893, and 1.0202%, respectively. Overall, the error index values of the LSTM model in predicting other characteristic gases are lower than those of the comparison model, further indicating that it can accurately learn the time series correlation of the actual historical monitoring data of each characteristic gas in the transformer, and has effectiveness in predicting the concentration values of each characteristic gas. This can further predict the degree of transformer malfunction for the purposes of early warning. On the basis of a single step prediction, a 2–4 step extended prediction was conducted on the C2H6 concentration to predict its future 1–2 weather volume concentration, to further verify the effectiveness of LSTM. The results are shown in Figure 13.



(a) C_2H_6 concentration time series (b) Hyperparameter settings for different models

Figure 10. C2H6 concentration time series and hyperparameter content of four prediction models.

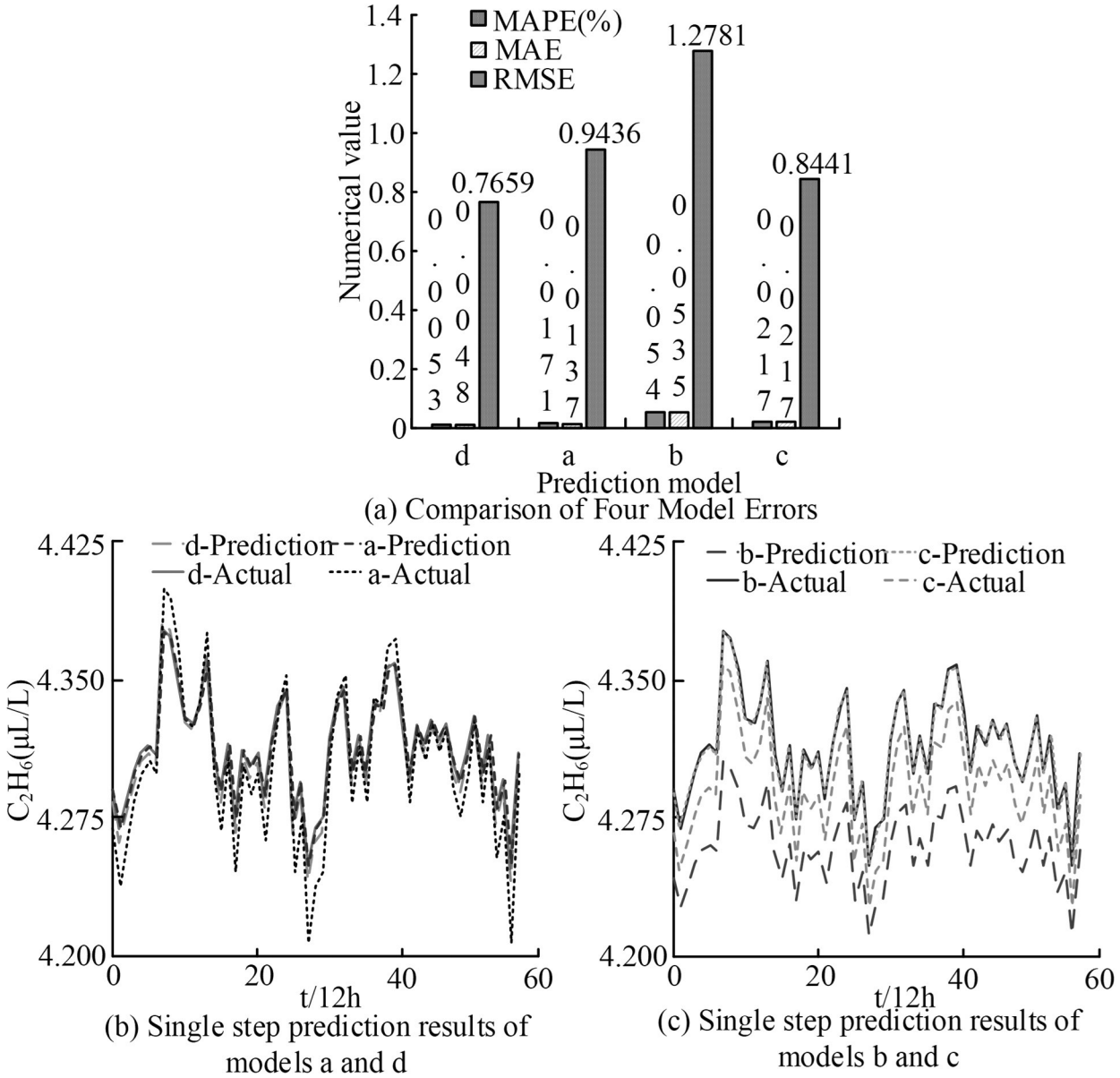


Figure 11. Single step prediction and error comparison results of four models.

Figure 13 shows that the three error values of LSTM in two-step prediction are 0.0194, 0.0161, and 0.6535%, respectively. The error values in the three-step prediction are 0.0359, 0.0307, and 0.7757%, respectively. The error values in the 4-step prediction are 0.0502, 0.0469, and 1.1365%, respectively. Overall, on the basis of the implementation data of transformer monitoring, the LSTM dissolved gas concentration prediction model using TPE algorithm can be applied to the proposed TFD method for fault diagnosis and early warning. Meanwhile, by combining the two models, real-time future numerical predictions can be made on the monitoring data of transformer DGA, and the related future development trends of monitoring and measurement can be monitored to achieve the purpose of corresponding diagnosis and early warning. Overall, by combining FLS-SNRA with LSTM+TPE, real-time future numerical prediction of transformer DGA monitoring data can be carried out, and the future trend of measurements can be monitored to achieve the corresponding diagnosis and early warning of transformer future faults. Compared to current methods, the presented research has higher effectiveness and practicality.

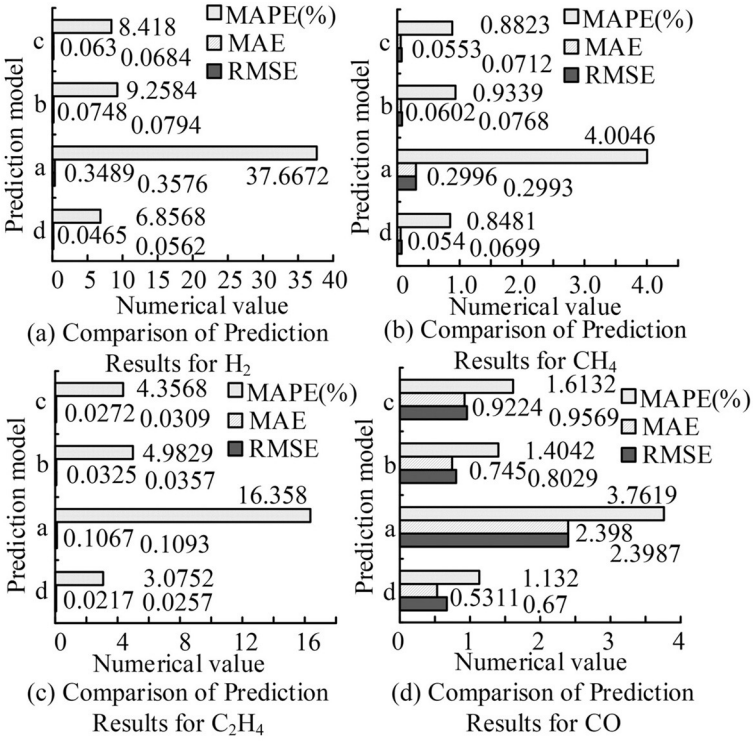


Figure 12. Prediction results of other characteristic gas concentrations.

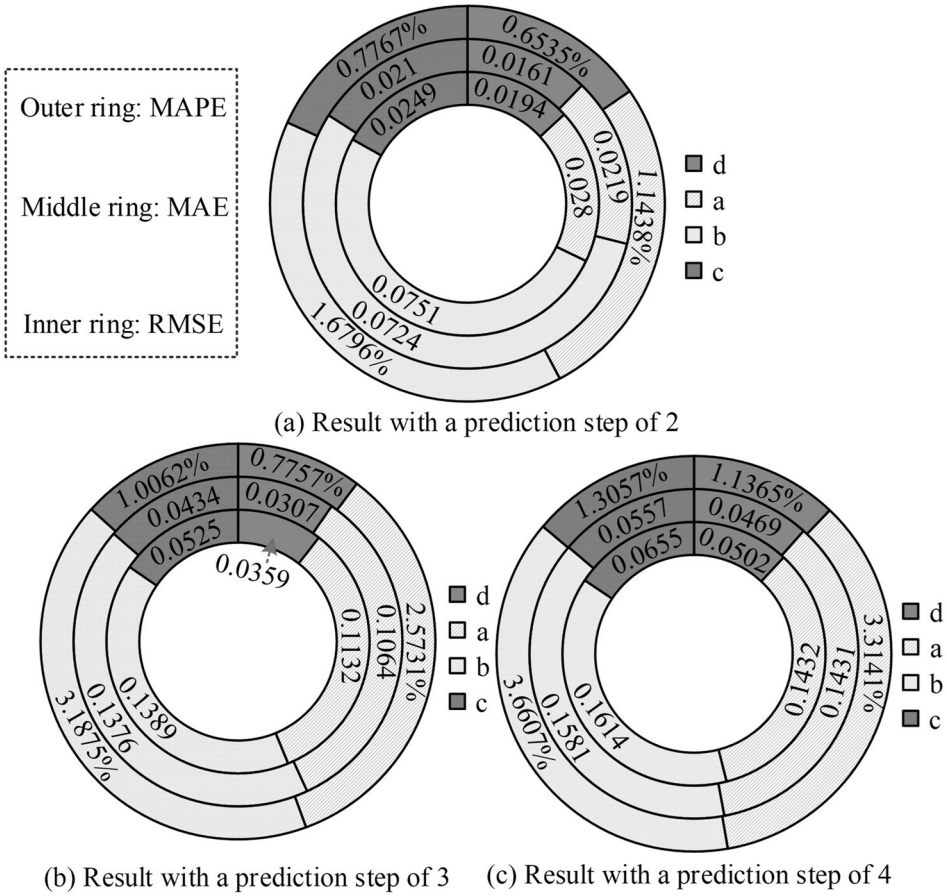


Figure 13. Comparison of multi-step prediction results of different models.

4. Conclusion

In response to the problems of low diagnostic accuracy and insufficient trend prediction ability of current intelligent diagnosis and prediction methods for transformer faults in cases when the fault categories are imbalanced, this paper proposes a TFD-PM approach built using the TensorFlow DL framework, and verified through experiments. The data above indicated that in the validation of the diagnostic model, when the actual quantity of HLN is set to 58, the accuracy of model training and testing reaches the maximum values, reaching 97.5% and 92.5%, respectively; After adding 0.001 times GWN, the accuracy was significantly perfected, so both parameters were set to 58 and 0.001. In the comparison of models on the ground of this, the fault diagnosis accuracy is as high as 93.30%, and the performance indicated by the confusion matrix is higher than that of the comparison model. In the validation of the prediction model, the one-step prediction results for C2H6 concentration showed that the predicted results of the research model were closer to the actual results, with an error of no more than 0.003, making our approach perform better than the comparison model. In H2 characteristic gas prediction, the three error index values of LSTM are 0.0562, 0.0465, and 6.8565, respectively, which are lower than for the comparison model. Other gases show the same results. Overall, the diagnostic and predictive models proposed in the study have high superiority and effectiveness and can facilitate diagnosis and early warning of transformer DGA faults. The proposed approach effectively solves the problem of incomplete coding and large judgment boundaries in traditional methods, which lead to low diagnostic accuracy, particularly in practical transformer fault diagnosis. However, the models constructed in the study belong to a class of static models, and it is necessary to optimize them for the dynamic environment to achieve dynamic diagnosis of transformer faults in the future. At the same time, only accuracy indicators were selected for research and analysis, and more indicators can be added for verification in the future.

References

[1] K. Jain and A. Saxena, "Simulation on Supplier Side Bidding Strategy at Day-ahead Electricity Market Using Ant Lion Optimizer", *Journal of Computational and Cognitive Engineering*, vol. 2, no. 1, pp. 17–27, 2023. http://dx.doi.org/10.47852/bonviewJCCE2202160

[2] Y. Zhang *et al.*, "Early Warning of Incipient Faults for Power Transformer Based on DGA Using a Two-Stage Feature Extraction Technique", *IEEE Transactions on Power Delivery*, vol. 37, no. 3, pp. 2040–2049, 2022. http://dx.doi.org/10.1109/TPWRD.2021.3103455

[3] Y. Wu *et al.*, "A Transformer Fault Diagnosis Method Based on Hybrid Improved Grey Wolf Optimization and Least Squares-support Vector Machine", *IET generation, transmission & distribution*, vol. 16, no. 10, pp. 1950–1963, 2022. http://dx.doi.org/10.1049/gtd2.12405

[4] L. Wang *et al.*, "Gaussian Process Multi-Class Classification for Transformer Fault Diagnosis Using Dissolved Gas Analysis", I*EEE Transactions on Dielectrics and Electrical Insulation*, vol. 28, no. 5, pp. 1703–1712, 2021. http://dx.doi.org/10.1109/TDEI.2021.009470

[5] Y. Zhou *et al.*, "Novel Probabilistic Neural Network Models Combined with Dissolved Gas Analysis for Fault Diagnosis of Oil-immersed Power Transformers", *ACS omega*, vol. 6, no. 28, pp. 18084–18098, 2021. http://dx.doi.org/10.1021/acsomega.1c01878

[6] A. Kukker *et al.*, "An Intelligent Genetic Fuzzy Classifier for Transformer Faults", *IETE Journal of Research*, vol. 68, no. 4, pp. 2922–2933, 2022. http://dx.doi.org/10.1080/03772063.2020.1732844

[7] H. Hu *et al.*, "A Novel Method for Transformer Fault Diagnosis Based on Refined Deep Residual Shrinkage Network", *IET Electric Power Applications*, vol. 16, no. 2, pp. 206–223, 2022. http://dx.doi.org/10.1049/elp2.12147

[8] J. Münchmeyer *et al.*, "The Transformer Earthquake Alerting Model: A New Versatile Approach to Earthquake Early Warning", *Geophysical Journal International*, vol. 225, no. 1, pp. 646–656, 2020. http://dx.doi.org/10.1093/gji/ggaa609

[9] N. Ardi *et al.*, "Predicting Missing Value Data on IEC TC10 Datasets for Dissolved Gas Analysis using Tertius Algorithm", *Journal of Applied Informatics and Computing*, vol. 7, no. 1, pp. 44–50, 2023. http://dx.doi.org/10.30871/jaic.v7i1.5361

[10] A. G. C. Menezes *et al.*, "Induction of Decision Trees to Diagnose Incipient Faults in Power Transformers", *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 29, no. 1, pp. 279–286, 2022. http://dx.doi.org/10.1109/TDEI.2022.3148453

[11] C. Zhu, "An Adaptive Agent Decision Model Based on Deep Reinforcement Learning and Autonomous Learning", *Journal of Logistics, Informatics and Service Science*, vol. 10, no. 3, pp. 107–118, 2023. http://dx.doi.org/10.33168/JLISS.2023.0309

[12] H. Zheng *et al.*, "Investigation on Micr-mechanism of Palm Oil as Natural Ester Insulating Oil for Overheating Thermal Fault Analysis of Transformers", *High Voltage*, vol. 7, no. 4, pp. 812–824, 2022. http://dx.doi.org/10.1049/hve2.12182

[13] A. Bernardić, "Power System Control and Protection Models Based on Artificial Intelligence–A Tensorflow Approach", *B&H Electrical Engineering*, vol. 16, no. 1, pp. 27–33, 2022. http://dx.doi.org/10.2478/bhee-2022-0004

[14] S. F. Stefenon *et al.*, "Classification of Distribution Power Grid Structures Using Inception v3 Deep Neural Network", *Electrical Engineering*, vol. 104, no. 6, pp. 4557–4569, 2022. http://dx.doi.org/10.1007/s00202-022-01641-1

[15] Z. Wan *et al.*, "KFIML: Kubernetes-Based Fog Computing IoT Platform for Online Machine Learning", *IEEE Internet of Things Journal*, vol. 9, no. 19, pp. 19463–19476, 2022. http://dx.doi.org/10.1109/JIOT.2022.3168085

[16] J. X. Chong *et al.*, "Deepfakes Detection using Computer Vision and Deep Learning Approaches", *Journal of System and Management Sciences*, vol. 12, no. 5, pp. 21–35, 2022. http://dx.doi.org/10.33168/JSMS.2022.0502

[17] X. Cui *et al.*, "Two-Step Electricity Theft Detection Strategy Considering Economic Return Based on Convolutional Autoencoder and Improved Regression Algorithm", *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 2346–2359, 2022. http://dx.doi.org/10.1109/TPWRS.2021.3114307

[18] V. Kosana *et al.*, "Hybrid Convolutional BI-LSTM Autoencoder Framework for Short-term Wind Speed Prediction", *Neural Computing and Applications*, vol. 34, no. 15, pp. 12653–12662, 2022. http://dx.doi.org/10.1007/s00521-022-07125-4

[19] A. Takiddin *et al.*, "Deep Autoencoder-Based Anomaly Detection of Electricity Theft Cyberattacks in Smart Grids", *IEEE Systems Journal*, vol. 16, no. 3, pp. 4106–4117, 2022. http://dx.doi.org/10.1109/JSYST.2021.3136683

[20] D. Paul *et al.*, "Bayesian Optimization-Based Gradient Boosting Method of Fault Detection in Oil-Immersed Transformer and Reactors", *IEEE Transactions on Industry Applications*, vol. 58, no. 2, pp. 1910–1919, 2022. http://dx.doi.org/10.1109/TIA.2021.3134140

[21] L. L. Sin *et al.*, "Comparing Machine Learning and Deep Learning Based Approaches to Detect Customer Sentiment from Product Reviews", *Journal of System and Management Sciences*, vol. 13, no. 2, pp. 101–110, 2023. http://dx.doi.org/10.33168/JSMS.2023.0207

Received: August 2023
Revised: September 2023
Accepted: October 2023

Contact addresses:
Xiaomeng Li
Department of Mechanical and Electrical Engineering
Weihai Ocean Vocational College
Rongchen
China
e-mail: lixiaomeng@whovc.edu.cn

Xiaomeng Li is a graduate student with an MSc degree. He graduated from the Department of Information Engineering of City College of Xi'an Jiaotong University in the area of electrical engineering and automation. Furthermore, he graduated from the School of Electrical Engineering of North China Electric Power University in the area of electromagnetic field and microwave technology. Currently he is employed at the Department of Mechanical and Electrical Engineering of Weihai Ocean Vocational College as a full-time teacher of motor control technology and a lecturer with a professional title. His research interests include transformer model construction.