Research Article

Dissolved Gas Analysis of Insulating Oil in Electric Power Transformers: A Case Study Using SDAE-LSTM

Zhao Luo ¹, Zhiyuan Zhang,¹ Xu Yan,² Jinghui Qin,¹ Zhendong Zhu,¹ Hao Wang,¹ and Zeyong Gao¹

¹Faculty of Electric Power Engineering, Kunming University of Science and Technology, Kunming 650500, China

²Jiangsu Frontier Electric Technology Co., Ltd., Nanjing 211102, China

Correspondence should be addressed to Zhao Luo; waiting.1986@live.com

Received 8 February 2020; Revised 9 September 2020; Accepted 24 October 2020; Published 11 November 2020

Academic Editor: Ricardo Aguilar-Lopez

Copyright © 2020 Zhao Luo 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.

Dissolved gas analysis (DGA) is the most important tool for fault diagnosis in electric power transformers. To improve accuracy of diagnosis, this paper proposed a new model (SDAE-LSTM) to identify the dissolved gases in the insulating oil of power transformers and perform parameter analysis. The performance evaluation is attained by the case studies in terms of recognition accuracy, precision ratio, and recall ratio. Experiment results show that the SDAE-LSTM model performs better than other models under different input conditions. As evidenced from the analyses, the proposed model achieves considerable results of recognition accuracy (95.86%), precision ratio (95.79%), and recall ratio (97.51%). It can be confirmed that the SDAE-LSTM model using the dissolved gas in the power transformer for fault diagnosis and analysis has great research prospect.

1. Introduction

Power transformers are considered as the core of the electric power systems, and its running state determines whether the power network is controllable or not. Power transformers tend to decompose and produce some gases during operation. These gases are dissolved in insulating oil due to stress (electrical stress, mechanical stress, and thermal stress, respectively) [1–3]. Generally, there are many kinds of gases dissolved in insulating oil, such as hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), carbon monoxide (CO), and carbon dioxide (CO_2) [4]. In fact, the dissolved concentrations of these gases are closely associated with the running state of the power transformers. It is common that Dissolved Gas Analysis (DGA) is considered as a potential way to diagnose the essential fault both in interior and exterior of the power transformer and predict how the power transformer will turn out [5–8]. Therefore, the result of DGA directly provides enough effective information to diagnose the running states of the power transformers.

Over the past several decades, two different types of approaches for obtaining dissolved gas analysis results have been proposed. The first type belongs to mathematical approaches, which includes International Electrochemical Commission (IEC) ratio, Doerneburg ratios, David triangle method, and Rogers ratios. However, these approaches have some limitations, such as incompletely encoding and the absolution of fault boundary distinction. These approaches are often difficult to exactly describe the running state and trend of power transformers due to its limitations, even though they are simple to operate. Meanwhile, the mathematical approaches cannot be able to provide an interpretation for faults or possible trends [9].

The second type of the approaches is applying the machine learning method that stems from artificial intelligence to overcome these issues. These approaches applied machine learning method to diagnose the running state of the power transformers and revealed the trends of the power transformers. Compared with mathematical approaches, machine learning-based approaches performed better diagnostic accuracy due to its excellent capabilities of

classification and predication of machine learning. Therefore, these are potential means of fault diagnosis in power transformers.

Typical machine learning methods had been implemented for DGA, including artificial neural networks (ANNs), support vector machine (SVM), relevance vector machine (RVM), and fuzzy theory [10, 11]. To some extent, these approaches improve the accuracy of fault diagnosis, whereas they are usually some specific weaknesses, too. For instance, the diagnostic accuracy of an ANN is closely related to the choice of lots of parameters. Failure to find the most appropriate parameters for the model has a vast influence on the diagnostic accuracy. Another problem of the ANN is overfitting. The diagnostic accuracy of SVM depends on the choice of the kernel function and corresponding parameters (e.g., cost parameter, slack variables, and the margin of the hyper plane), and that is not an easy thing. An RVM model reduces the computation of the kernel function, so it costs more training time than SVM.

In recent years, deep learning known as deep neural network (DNN) has attracted considerable attentions in machine learning field. On the one hand, recurrent neural network (RNN) is a type of ANNs, which especially contain a loop structure inside. This loop structure makes the input get more complete information at each point, which not only contains the current information but also includes the previous information of the hidden layer. The hidden layers reveal a function of memory block, so that preserves historical information and constantly renews the information with the updated input. Although the RNN model uses historical information to predict the current state, it improves the limitations of remote dependence in the field of time series. But, in practical applications, it will bring problems such as vanishing gradient and explosive gradient. To solve the issue, Hochreiter proposed the long short-term memory (LSTM), a special RNN method. LSTM is an RNN architecture published by Sepp Hochreiter and Jürgen Schmidhuber. Like most RNNs, a LSTM network is universal in the sense that, given enough network units, it can compute anything a conventional computer can compute, provided it has the proper weight matrix, which may be viewed as its program. Unlike traditional RNNs, an LSTM network is well-suited to learn from experience to classify, process, and predict time series when there are very long time lags of unknown size between important events [12, 13]. This is one of the main reasons why LSTM outperforms alternative RNNs and Hidden Markov Models and other sequence learning methods in numerous applications. At present, the LSTM has been widely used in some fields, like machine translation, automatic speech recognition, and natural language processing. Meanwhile, autoencoder (AE), denoising autoencoder (DAE), and stacking denoising autoencoder (SDAE) stem from DNNs have been applied to machine health monitoring and fault diagnosis. Especially, DAE and SDAE can learn feature representations from input data by denoising, such as reconstructing input from artificial corruption [14]. Inspired by that, we try to consider using SDAE-LSTM in fault diagnosis of power transformers.

SDAE adopts unsupervised learning method, which can learn the internal characteristics of fault data more deeply.

And SDAE contains noise reduction factor, which can avoid noise interference. In order to enhance the ability of model to extract fault data features, an improved SDAE-LSTM transformer fault diagnosis method is proposed. LSTM overcomes the “gradient” existing in traditional neural network in information processing. Therefore, SDAE-LSTM transformer fault diagnosis method can not only effectively enhance the model’s ability to extract fault features but also can track the variation of gas concentration in oil with time better, improve the diagnostic accuracy, and provide a strong guarantee for the safe and stable operation of power transformers.

In this study, a LSTM-based model is developed to diagnose the essential fault by the DGA. As the traditional fault diagnosis method of transformer difficult to deal with the fault feature information effectively, using Sade to fully mine the fault feature information of transformer is helpful to improve the accuracy of transformer fault diagnosis. Due to the interference and influence of monitoring device, ambient temperature, personnel operation, and other aspects of dissolved gas content in transformer oil, it is necessary to extract features from the original data. Feature extraction can reduce the impact of data on the model performance and improve the training speed and diagnostic accuracy of the model. The processed data are used as the input of LSTM. Then, the LSTM is performed for further analysis. To testify the proposed model, its diagnostic accuracy is obtained from case studies. Furthermore, three other approaches, namely, BPANN, SVM, and RF, respectively, are compared with the proposed model.

The rest of this study is organized as follows. Sections 2.1 and 2.2 briefly demonstrates the needed methods including SDAE and LSTM. Section 2.3 describes the proposed model. The experiment results of case study are presented in Section 3. Section 4 gives conclusions and extensive discussions.

2. Proposed Model

2.1. Stacked Denoising Autoencoder. For a given data, DNN stems from ANN is a classifier consisting of a set of models attempts to learn high-level representations with very deep neural networks, which is usually deeper than three layers. There are three parts of a typical DNN: an input layer, an output layer, and a lot of hidden layers stacked between the former two layers. The network is firstly layerwise initialized by a type of unsupervised training and then it turns into a supervised way. A large number of layers consisting of a finite number of nodes are introduced to realize highly nonlinear functions in DNNs, which can grasp the presentation of statistical regularity from the data. And the representation features have many virtues for classification as a common way in deep learning [15]. Figure 1 presents the difference between typical machine learning and deep learning. Where, SVM (support vector machine), RVM (relevance vector machine), RF (random forest), and PCA (principal component analysis) are few of the most widely used data dimensionality reduction algorithms. The main idea of PCA is to map the n -dimensional features to the k -dimension, which is a new orthogonal feature, also known as

FIGURE 1: Comparison between regular machining learning and deep learning.

the principal component. It is a k -dimensional feature reconstructed from the original n -dimensional features.

An AE is a three-layer network, which consist of an encoder and a decoder. On the one hand, the encoder acts on the input data from a high-dimensional space to a low-dimensional space. On the other hand, the decoder reconstructs the input data from the corresponding codes. On the basis of the training samples $x_i \in Z$, the encoder transforms the input vector x into a hidden representation $h_i \in Z$ through a nonlinear mapping as in the following equation:

$$h = \alpha(W_1 x + b_1), \quad (1)$$

where $\alpha(x) = 1/[1 + \exp(-x)]$ is a nonlinear activation function, which is used for nonlinear deterministic mapping. Then, decoder maps hidden representation back to the original representation in a similar way in the following equation:

$$g = \alpha(W_2 x + b_2). \quad (2)$$

The training process of an AE is conducted to optimize the parameter set, $\theta = \{W_1, b_1, W_2, b_2\}$, that is evaluated by minimizing the reconstruction error between g and x . Generally, the measurement of the average reconstruction error is mean square error (MSE). The calculation method of MSE is shown in

$$\begin{aligned} J_{\text{MSE}}(\theta) &= \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|z_i - x_i\|^2 \right) \\ &= \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|f_{W,b}(x_i) - x_i\|^2 \right). \end{aligned} \quad (3)$$

An AE attempts to learn an identity map for the approximation between output and input in the training process. However, the retaining information of the original input does not guarantee to separate effectively the useful information from

the noise in the original input. In the other words, it is not helpful to extract the useful features from the original input.

Vincent first proposed a SDAE model to escape an equivalent representation of the original input. A SDAE is a stack of single-layer DAE, which takes a corrupted version of data as input to reconstruct or denoise original input. Figure 2 presents the construction of the SDAE [16]. Above all, a DAE randomly breaks the original input based on q_D , so that the original input turns to $\tilde{x}$. To make y become the hidden representation, take the encoding function f that acts on $\tilde{x}$. And, the decoding function g finally reconstructs y to z . Here, q_D is one of the kinds of functions, such as Gaussian noise, Masking noise, and Salt-and-pepper noise. Likewise, the encoding and decoding function both are active function (i.e., sigmoid function, tanh function, and ReLU function). Generally, the reconstruction error is minimized to acquire the appropriate parameters of DAE. The well-trained encoding function acts on the original input to generate the feature representation, which is served as the input at the back [17]. The SDAE could be built quite mechanically by repeating the certain straightforward processes over and over again as shown in Figure 2. By adding noise in training data, the AE will reconstruct the original input data, so that the DAE not only overcomes the interference of the noise but also captures more effective features and improves the generalization ability of the model.

In general, the training process of SDAE can be divided into two stages: the unsupervised pretraining and the supervised finetune. Unsupervised pretraining is conducted by training the parameters of initialization network, so that each layer is trained as a DAE by minimizing the error in reconstructing its input. After that, the network goes through a supervised finetune by adding a logistic regression layer on top of network. As a result, the generalization performance of the SDAE is improved dramatically.

FIGURE 2: The structure of SDAE.

2.2. LSTM. LSTM is an excellent variant model of RNN, which inherits the characteristics of most RNN models, and at the same time solves the Vanishing Gradient problem caused by the gradual reduction in the gradient back propagation process. When it comes to information processing tasks, LSTM is very suitable for processing problems that are highly related to time series. LSTM has a very good performance when dealing with time-series data classification problems. Considering that LSTM has solved the “gradient dissipation” problem of traditional neural networks in terms of information processing, it works better than RNN when dealing with some problems that require long-term storage. Therefore, this article will use the LSTM model for data processing.

LSTM is a collection of basic unit, which is the memory block consisting of memory cell and three gates of control memory cell state, namely, forget gate, input gate, and output gate, respectively. The forget gate plays an important role on removing some unnecessary information, and the input gate determines the effect of input on the state of memory units. Likewise, the output gate selects and sends out the output. Figure 3 presents the structure of the memory block. Given the input vector x_t at the time t and the output h_{t-1} at the time $t-1$, we define the W , U , and b as weight matrix and bias vector. The memory block refreshes the state and decides the output.

Firstly, the forget gate removes the historical and futile information, as can be shown in

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f). \quad (4)$$

Then, on the basis of the input and historical information, the input gate refreshes the state. It is given by

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\ \tilde{C}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c), \\ C_t &= f_{t-1} * C_{t-1} + i_t * \tilde{C}_t. \end{aligned} \quad (5)$$

FIGURE 3: The structure of memory block.

Finally, the output gate gives the current information as follows:

$$\begin{aligned} o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o), \\ h_t &= o_t * \tanh(C_t), \end{aligned} \quad (6)$$

where σ represents the logistic sigmoid function and C_t represents the state of the memory cell at the time t . Meanwhile, f_t , i_t , and o_t represent the output state of forget gate, input gate, and output gate at the time t , respectively.

LSTM improves the traditional memory block of RNN model (single tanh layer or single sigmoid layer), that overcome the limitation of long-term preservation for information [18]. By the design of gate structure and memory cell, LSTM can effectively update and transmit the key information in time series. Compared with the traditional RNN model, LSTM performs better on information selection and learning characteristics of time series, which can

FIGURE 4: The internal loop structure of the LSTM.

apply the long-distance information to the current prediction. The internal loop structure of the LSTM neural network is shown in Figure 4 [19]. As can be seen from it, the input of each time step is a vector, and the state of the hidden layer gradually moves backward according to the time series.

As a special structure, it is necessary to input a sequence of feature vectors, which consists of continuous feature vector from M time steps. In that, it is necessary to construct the input sequence of LSTM model before training. Given the input vector x_t at the t time step, the constructed sequence is $\{x_{t-M+1}, x_{t-M+2}, \dots, x_t\}$. More specifically, the first sequence is $x_1, x_2, \dots, x_M$ and the second sequence is $x_2, x_3, \dots, x_{M+1}$. Likewise, the other sequences are acquired.

2.3. SDAE-LSTM Model. In this section, we apply a SDAE-LSTM model consisting of deep neural network and LSTM to diagnose the faults of the power transform.

The framework of SDAE-LSTM model is shown in Figure 5. As seen in Figure 5, the first step is to extract the features from the original data vector by SDAE. Then, the last hidden layer of the SADE is regarded as the input of the LSTM, which is an effective representation of the original data.

According to the common faults in daily monitoring, this paper establishes the following data tags of power transformer status: classified into normal pattern (F1), discharge with high energy pattern (F2), discharge with low energy pattern (F3), partial discharge pattern (F4), high-temperature thermal pattern (F5), middle-temperature thermal pattern (F6), and low-temperature thermal pattern (F7).

2.4. Input Vector. Based on a large number of researches on transformer, the gases selected usually contains five hydrogenous gases, such as hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), and ethane (C_2H_6). However, it is not uncommon that some limitations are revealed for diagnosing faults on the basis of these five gases. On the one hand, the values of these gases largely are

dispersed that reflect enormous discrepancy among the orders of magnitudes. On the other hand, the DGA data may have great uncertainty for the same type of fault due to the diversity. In that, various ratio method are proposed to solve these drawbacks, such as composition ratio ($H_2\%$, $CH_4\%$, $C_2H_2\%$, $C_2H_4\%$, $C_2H_6\%$), IEC three-ratio codes (CH_4/H_2 , C_2H_2/C_2H_4 , C_2H_4/C_2H_6), Rogers four-ratio codes (CH_4/H_2 , C_2H_2/C_2H_4 , C_2H_4/C_2H_6 , C_2H_6/CH_4), and noncode ratio (CH_4/H_2 , C_2H_2/C_2H_4 , C_2H_4/C_2H_6 , $C_2H_2/(C_1 + C_2)$, $H_2/(H_2 + C_1 + C_2)$, $C_2H_4/(C_1 + C_2)$, $CH_4/(C_1 + C_2)$, $C_2H_6/(C_1 + C_2)$, $(CH_4 + C_2H_4)/(C_1 + C_2)$).

In this study, we have proposed a SDAE-LSTM model for transformer incipient fault diagnosis. As a typical unsupervised deep learning network, SDAE can dig the deep-level features of unlabeled data. Without adding any labels to the input data, training can derive the intrinsic feature of the input data, so that the input intrinsic can extract high-dimensional features; tap the potential value within the data. Through the monitoring of relevant data during transformer failure, it can be found that there are large numerical differences between different parameters. If the original gas concentration data are directly placed in the model for training operations, the weight setting may be affected due to the large difference in data. Due to different orders of magnitude, a larger value will affect the training results, resulting in distortion of the training results of the entire network. In order to avoid the influence of different original data orders on training weights, it is necessary to preprocess the gas data in transformer oil. Because the data range is relatively concentrated, this article uses a standardized method. A normalization method has been developed to normalize the input vector as follows:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (7)$$

where x presents the original vector, x^* means standardized vector, and $x_{\max}$ and $x_{\min}$ represent the maximum and minimum separately.

After the LSTM training, the output results are the final classification of the original input features, and all the

FIGURE 5: The framework of the SDAE-LSTM transformer fault diagnosis system.

experimental results are obtained by 10-fold cross-validation.

3. Case Studies

In this section, case experiments confirmed that the proposed model has good classification performance. All experiments are coded in PYTHON 3.7 and implemented on Intel Core i7-6200CPU, 2.30 GHz computer with an 8 GB memory. The SDAE and LSTM are established by TensorFlow of Python.

In this case, we collect the practical data consisting of normal and abnormal state from China national grid and China southern power grid, which is called the original dataset. The dataset was composed of 1723 samples and randomly divided into a training set (1378 samples, 80 percentage) and a testing set (345 samples, 20 percentage). On basis of that, we have conducted some experiments to validate the efficiency of the proposed model in this section.

3.1. Evaluation of Fault Diagnosis Accuracy. The fault diagnosis accuracy is quantified by recognition accuracy (RA), which has been widely used in machine learning models. The criterion means the proportion of samples recognized correctly in the whole samples recognized, which represents the overall performance of the SDAE-LSTM model:

$$RA = \frac{a_f}{a_f + c_f} \times 100\%, \quad (8)$$

where a_f means the number of samples recognized correctly and c_f gives the number of the whole samples recognized.

The fault diagnosis accuracy also can be evaluated by precision ratio (PR), which gives the information on the

percentage of particular samples recognized correctly in the entire same samples. This criterion reflects the performance on recognizing correctly the particular patterns:

$$PR = \frac{b_{f_1}}{b_f} \times 100\%, \quad (9)$$

where b_f means the number of samples, which come from the same pattern. b_{f_1} gives the information about the number of samples recognized correctly in the whole same samples.

In addition, the recall ratio (RR) as an effective index is utilized for measuring the performance of the proposed model in fault diagnosing. It demonstrates the percentage of a particular pattern recognized in the whole number of the same patterns. Moreover, recall ratio and prediction ration are regarded as two interconstrain measures:

$$RR = \frac{b_{f_1}}{d_f} \times 100\%, \quad (10)$$

where d_f means the whole number of samples.

In this study, RA not only is an indicator in measuring, but beyond that, PR and RR are effective indexes for evaluating the fault diagnosis. In that, performance of the proposed model can be quantified objectively and overall.

3.2. Comparison with Different Inputs. The RAs, PRs, and RRs are attained from the same model (LSTM model) with different inputs in this section, classified into DGA, composition ratio, IEC three ratio codes, Rogers four-ratio codes, noncode ratio, and SDAE. That is to say, the experiments are conducted to verify the importance of SDAE for LSTM

TABLE 1: RAs' comparison for different inputs.

Patterns	Inputs					
	DGA (%)	Composition ratio (%)	IEC three-ratio codes (%)	Rogers four-ratio codes (%)	Noncode ratio (%)	SDAE (%)
F1	93.13	96.62	93.37	97.04	97.12	98.34
F2	89.74	96.67	90.94	95.59	95.58	97.32
F3	87.54	94.01	91.83	92.97	93.35	94.57
F4	89.96	91.25	89.44	93.21	93.28	95.24
F5	89.68	90.43	88.09	94.80	92.39	94.21
F6	92.80	90.53	95.20	92.77	92.45	98.15
F7	85.34	94.91	94.65	90.56	92.93	93.16
Average	89.74	93.49	91.93	93.85	93.87	95.86

model in Section 4.2. Table 1 gives the information about RAs, after inputting the different vectors to the same model.

As is evident from the Table 1, the model gains the best recognition accuracy when it comes to SDAE, which attained 95.86% on average alone. Four recognition accuracies in seven patterns are more than 95%, especially two of them are more than 98% dramatically. Moreover, the model performs well in terms of recognition accuracy on average, which are moderately worse than SDAE-LSTM model, 93.49% in composition ratio, 93.85% in Rogers four ratio codes, and 93.87% noncode ratio. Every RA is no less than 90%, whilst no more than two patterns' recognition accuracies surpass to 95%. In a stark contrast, negative phenomena are found in results of the model, which are inputted by DGA. The average recognition accuracy is merely to 89.74%, and most of patterns are recognized to less than 90%. After all, it is not uncommon due to its originality in DGA. In that, the SDAE attaches its importance on the recognition ability of the model.

Table 2 presents the precision ratios for different inputs by the proposed model. A thorough analysis from the table is that the precision ratios of normal pattern are acquired to more than 95% for all different inputs. For other patterns, the model has poor performances for the former five inputs, where one of patterns acquires the precision ratio less than 90% at least. Especially, most of precision ratios are less than 90% when it comes to DGA as the input of the proposed model. Interestingly, the most outstanding performance is shown for SDAE input in terms of the average precision ratio, which is to considerable 95.79%. The next one is slightly less by 2% around on average, 94.84% in noncode ratio, 92.59% in Rogers four-ratio codes, 92.67% in IEC three-ratio codes, and 93.46% in composition ratio, respectively. There is no doubt it is not a good result for DGA by the model, which is merely to 89.83% on average. As a result, the proposed model reflects the most potential performance for SDAE when it comes to precision ratio.

Table 3 illustrates the recall ratios for different inputs. As shown in Table 3, the LSTM model for SDAE reflects better recall ratios than other inputs, which attained 97.51% on average alone, and even recall ratio and prediction ration are regarded as two interconstrain measures. More specifically, every pattern's recall ratio almost is more than 95%, which are considerable. On the contrary, there are not great generalization capacity in the terms of the average

recall ratio via the model, after inputting the former five inputs separately, 91.92% in noncode ratio, 89.58% in Rogers four-ratio codes, 83.44% in IEC three-ratio codes, 76.88% in composition ratio, and 82.41% in DGA. Most recall ratios of seven patterns are less than 80% for the former five inputs. Considering that, according to the RRs' comparison for different input, the model is more sensitive to recognize seven patterns of electric power transformers for SDAE.

3.3. Comparison with Other Methods. Backpropagation neural network (BPANN), support vector machine (SVM), and random forest (RF) are the most traditional machine learning methods to recognize patterns in this interest. In this section, four comparative experiments are conducted to illustrate the potential merits of the SDAE-LSTM model which compare the RAs, PRs, and RRs with these three common methods. The parameters of these three models are shown in Table 4.

Table 5 gives the information about recognition accuracies of different models for dissolved gas analysis process of insulating oil in electric power transformers. As shown from it, the SDAE-LSTM model achieves the best average recognition accuracy to 95.86%. The recognition accuracies of patterns surpass 95% except three of them. In particular, the recognition accuracies of normal discharge with high energy and middle temperature thermal pattern are more than 97%. The next one is RF model, which is moderately less by 3.6% on average RA. Interestingly, two of seven perform well, 95.89% in normal pattern, and 95.13% in partial discharge pattern. The rest of them have middle positions, even the RA of low temperature thermal pattern is merely to 88.54%. The BPANN and SVM have the worst average RA, which are 82.90% and 86.98%, respectively. It is obvious that the RAs of patterns are less than 90% except only one pattern, sometimes even less than 80%. Therefore, the proposed model performs better than other three common approaches on the recognition accuracies.

Table 6 represents detailed results of different models in terms of precision ratio. As is the most evident from it, the SDAE-LSTM model reflects the largest average precision ratio, which attained 95.79%. More specifically, the precision ratios are more than 95% considerably except the precision ratios of partial discharge pattern (93.86%) and low

TABLE 2: PRs' comparison for different inputs.

Patterns	Inputs					
	DGA (%)	Composition ratio (%)	IEC three-ratio codes (%)	Rogers four-ratio codes (%)	Noncode ratio (%)	SDAE (%)
F1	97.49	96.11	95.91	95.65	97.21	98.52
F2	87.69	93.87	89.75	89.34	94.46	95.53
F3	87.37	91.96	91.64	92.49	93.91	95.19
F4	88.53	89.54	92.31	93.56	97.05	93.86
F5	86.42	94.25	94.97	90.73	89.57	95.85
F6	92.98	96.08	92.85	94.77	96.73	97.62
F7	88.34	92.40	91.26	91.58	94.94	93.97
Average	89.83	93.46	92.67	92.59	94.84	95.79

TABLE 3: RRs' comparison for different inputs.

Patterns	Inputs					
	DGA (%)	Composition ratio (%)	IEC three-ratio codes (%)	Rogers four-ratio codes (%)	Noncode ratio (%)	SDAE (%)
F1	94.10	89.93	90.48	93.24	94.37	94.48
F2	90.11	90.01	90.35	93.13	95.61	97.44
F3	56.14	65.79	56.91	73.58	90.76	97.79
F4	92.51	97.11	82.38	91.14	88.57	99.73
F5	77.95	85.74	82.15	90.79	94.11	97.66
F6	87.36	89.24	93.40	92.79	90.91	95.89
F7	78.75	79.55	88.40	92.42	89.15	99.61
Average	82.41	76.88	83.44	89.58	91.92	97.51

TABLE 4: RAs' comparison for different methods.

Models	Parameters	Values
BPANN	Hidden unit number	16
	Activation function	Tansig
	Training method	Trainngd
	Learning rate	0.05
	Training iterations	0.001
	Objective value	1000
SVM	Kernel function	RBF
	RBF sigma σ	8.1493
	Box constraint C	1
RF	The number of trees (ntree)	500
	The capacity of feature space (mtry)	4

temperature thermal pattern (93.97%). Moreover, the average precision ratio of RF model is attained to 91.04%, which is less than 2.93%. Most of the seven patterns' precision ratios are more than 90%. In addition, the SVM model gets a worse average precision ratio (85.83%), since four patterns' precision ratios are less than 85%. In particular, the SVM model gains the smallest value in discharge with low patterns (70%) among twenty-eight data when it comes to precision ratio. Furthermore, the average precision ratio of BPANN model is the smallest among the figures of other three models, which is merely to 81.34%. As a consequence, these experiment results reveal that the SDAE-LSTM model performs best among four models in terms of precision ratio.

Recall ratios' comparison of different models is shown in the Table 7. To compare rationally the SDAE-LSTM model with other approaches, recall ratios are conducted as well due to interconstrain relationship between precision ratio and recall ratio. As can be seen in Table 7, the SDAE-

LSTM model reflects better recall ratio performance than the other three traditional machine learning models, while the single recall rate is greater than 94.4%, and the average recall rate is reaching 97.51%. The following one is the RF model, which is 91.64% in average when it comes to recall ratio. Interestingly, although one of patterns is merely 87.50% (discharge with low energy pattern), the rest are more than 90%. Comparing with RF and SDAE-LSTM models, the average recall ratios are in stark contrast both in BPANN model and SVM model, 78.57% and 84.24%, respectively. It should be clear that sharp declines are found in these fourteen datums, such as discharge with low energy patterns in BPANN model (46.88%) and SVM model (65.63%). Arising from analyses represented, performance of the SDAE-LSTM model is the best when it comes to recall ratio.

In conclusion, merits of the SDAE-LSTM model are revealed in three aspects in comparison with other three

TABLE 5: RAs' comparison of different models.

Patterns	Models			
	BPANN	SVM	RF	SDAE-LSTM
F1	86.41	91.77	95.89	98.34
F2	82.84	85.53	91.12	97.32
F3	79.23	84.29	92.67	94.57
F4	80.17	88.69	95.13	95.24
F5	82.33	86.49	90.28	94.21
F6	89.45	86.16	91.55	98.15
F7	79.86	85.78	88.54	93.16
Average	82.90	86.96	92.17	95.86

TABLE 6: PRs' comparison of different models.

Patterns	Models			
	BPANN	SVM	RF	SDAE-LSTM
F1	86.11	85.37	91.89	98.52
F2	83.00	87.91	92.39	95.53
F3	78.95	70.00	87.50	95.19
F4	80.00	84.21	85.71	93.86
F5	83.61	84.62	91.67	95.85
F6	85.71	93.67	96.15	97.62
F7	72.00	95.00	92.00	93.97
Average	81.34	85.83	91.04	95.79

TABLE 7: RRs' comparison of different models.

Patterns	Models			
	BPANN	SVM	RF	SDAE-LSTM
F1	86.11	97.22	94.44	94.48
F2	92.22	88.89	94.44	97.44
F3	46.88	65.63	87.50	97.79
F4	80.00	80.00	90.00	99.73
F5	85.00	91.67	91.67	97.66
F6	87.80	90.24	91.46	95.89
F7	72.00	76.00	92.00	99.61
Average	78.57	84.24	91.64	97.51

traditional machine learning models (BPANN, SVM, and RF models), classified into recognition accuracy, precision ratio, and recall ratio. To be more specific, the recognition accuracies of seven patterns all reach high level, especially in normal and middle temperature thermal patterns. Similarly, higher precision ratios and recall ratios are found in the SDAE-LSTM model than in the other three models. To put it in a nutshell, the SDAE-LSTM model is more effective model for pattern recognition in the dissolved gas analysis process of insulating oil in electric power transformers.

4. Conclusions

This study proposes a new model based on LSTM to analyze the dissolved gas of insulating oil in electric power transformers. Two parts' experiments are conducted to evaluate the proposed model. As the comparison experiments for different inputs (DGA, composition ratio, IEC three-ratio codes, Rogers four-ratio codes, noncode ratio, and SDAE) show, the SDAE-LSTM model is superior on recognition

accuracy, precision ratio, and recall ratio [20, 21]. Next, another comparison experiment result demonstrated that the proposed model achieves better performance than three traditional machine learning models (BPANN, SVM, and RF models). When it comes to recognition accuracy, the SDAE-LSTM attains 95.86%, which is obviously higher than BPANN (82.90%), SVM (86.96%), and RF (92.17%). Similarly, the SDAE-LSTM model gains better precision ratio (95.79%) than BPANN (81.34%), SVM (85.83%), and RF (91.04%). Obviously, a larger advantage is found in terms of recall ratio for the proposed model than of BPANN (78.57%), SVM (84.24%), and RF (91.64%) as well, which is 97.51% alone. That means the SDAE-LSTM model plays more promising role for recognizing patterns in dissolved gas analysis of insulating oil in electric power transformers.

Considering that the SDAE-LSTM model is a promising model for analyzing dissolved gas in the insulating oil of power transformers by identifying patterns, we can further study its application in online monitoring of power transformers. In addition, the comparisons with different capacities of samples are also interesting.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (grant no. 51907084), Yunnan Provincial Talents Training Program (grant no. KKS201704027), and Scientific Research Foundation of Yunnan Provincial Department of Education (grant no. 2018JS032).

References

- [1] N. Lelekakis, D. Martin, W. Guo, and J. Wijaya, "Comparison of dissolved gas-in-oil analysis methods using a dissolved gas-in-oil standard," *IEEE Electrical Insulation Magazine*, vol. 27, no. 5, pp. 29–35, 2011.
- [2] J. I. Aizpurua, V. M. Catterson, B. G. Stewart et al., "Power transformer dissolved gas analysis through Bayesian networks and hypothesis testing," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 25, no. 2, pp. 494–506, 2018.
- [3] 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.
- [4] J. Jiang, Z. Wang, X. Han et al., "Multi-gas detection in power transformer oil based on tunable diode laser absorption spectrum," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 26, no. 1, pp. 153–161, 2019.
- [5] 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.

- [6] 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.
- [7] S. Eeckhoudt, S. Autru, and L. Lerat, "Stray gassing of transformer insulating oils: impact of materials, oxygen content, additives, incubation time and temperature, and its relationship to oxidation stability," *IEEE Electrical Insulation Magazine*, vol. 33, no. 6, pp. 27–32, 2017.
- [8] J. Li, Q. Zhang, K. Wang, J. Wang, T. Zhou, and Y. Zhang, "Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 2, pp. 1198–1206, 2016.
- [9] J. Jiang, R. Chen, M. Chen, W. Wang, and C. Zhang, "Dynamic fault prediction of power transformers based on hidden markov model of dissolved gases analysis," *IEEE Transactions on Power Delivery*, vol. 34, no. 4, pp. 1393–1400, 2019.
- [10] S. S. M. Ghoneim, I. B. M. Taha, and N. I. Elkalashy, "Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 3, pp. 1838–1845, 2016.
- [11] 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 arctangent transform," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 1, pp. 586–595, 2016.
- [12] J. Maumela, F. Nelwamondo, and T. Marwala, "Condition monitoring of transformer bushings using rough sets, principal component analysis and granular computation as pre-processors," in *Proceedings of the 2013 International Conference on System Science and Engineering (ICSSE)*, Budapest, Hungary, July 2013.
- [13] L. Kong, L. Luan, K. Zhou, C. Chen, J. Chen, and Z. Wang, "Running state prediction and evaluation of power transformers," in *Proceedings of the 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*, Toyonaka, Japan, July 2019.
- [14] J. Dai, H. Song, G. Sheng, and X. Jiang, "Cleaning method for status monitoring data of power equipment based on stacked denoising autoencoders," *IEEE Access*, vol. 5, pp. 22863–22870, 2017.
- [15] B. Tan, J. Yang, Y. Tang, S. Jiang, P. Xie, and W. Yuan, "A deep imbalanced learning framework for transient stability assessment of power system," *IEEE Access*, vol. 7, pp. 81759–81769, 2019.
- [16] E. K. Wang, X. Zhang, and L. Pan, "Automatic classification of CAD ECG signals with SDAE and bidirectional long short-term network," *IEEE Access*, vol. 7, pp. 182873–182880, 2019.
- [17] W. Mao, J. Chen, X. Liang, and X. Zhang, "A new online detection approach for rolling bearing incipient fault via self-adaptive deep feature matching," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 2, pp. 443–456, 2020.
- [18] H. Shi, L. Guo, S. Tan, and X. Bai, "Rolling bearing initial fault detection using long short-term memory recurrent network," *IEEE Access*, vol. 7, pp. 171559–171569, 2019.
- [19] J. Dai, H. Song, G. Sheng, and X. Jiang, "LSTM networks for the trend prediction of gases dissolved in power transformer insulation oil," in *Proceedings of the 2018 12th International Conference on the Properties and Applications of Dielectric Materials (ICPADM)*, Xi'an, China, May 2018.
- [20] I. B. M. Taha, S. S. M. Ghoneim, and H. G. Zaini, "Improvement of rogers four ratios and IEC code methods for transformer fault diagnosis based on dissolved gas analysis," in *Proceedings of the 2015 North American Power Symposium (NAPS)*, Charlotte, NC, USA, October 2015.
- [21] X. F. Wang, Z. D. Wang, Q. Liu, G. Wilson, D. Walker, and P. W. R. Smith, "Dissolved gas analysis (DGA) of mineral oil under thermal faults with tube heating method," in *Proceedings of the 2017 IEEE 19th International Conference on Dielectric Liquids (ICDL)*, Aberdeen City, UK, June 2017.