



OPEN Intelligent fault diagnosis and operation condition monitoring of transformer based on multi-source data fusion and mining

Jingping Cui¹✉, Wei Kuang², Kai Geng³ & Pihua Jiao⁴

Transformers are important equipment in the power system and their reliable and safe operation is an important guarantee for the high-efficiency operation of the power system. In order to achieve the prognostics and health management of the transformer, a novel intelligent fault diagnosis of the transformer based on multi-source data fusion and correlation analysis is proposed. Firstly, data fusion for multiple components of transformer dissolved gases is performed by an improved entropy weighting method. Then, the combination of bidirectional long short-term memory network, attention mechanism, and convolution neural network is employed to predict the load rate, upper oil temperature, winding temperature data, and the fusion indices of dissolved gas components in the transformer. Furthermore, Apriori correlation analysis is performed on the transformer load rate and upper oil layer, winding temperature, and fusion indices of gas components by support and confidence levels to achieve a predictive assessment of the transformer state. Finally, the validity of the algorithm is verified by applying actual data from a power system monitoring platform. The results show that in the vicinity of sample point 88, the dissolved gas, upper oil temperature, and winding temperature data are not within the normal range of intervals, and it is presumed that the arc discharge phenomenon. Furthermore, the average correct fault diagnosis rate of 100 diagnoses of the transformer fault diagnosis model proposed in this paper is 0.917, and the mean square error of the correct rate is 0.018. The proposed model can achieve the prediction of the accident early warning, to prevent further expansion of the accident.

Transformers are widely used in power systems and are the most important equipment in power supply and distribution, assuming the role of electrical energy supply in residential electricity, industrial parks, and public services^{1,2}. Irregularities such as poor contact of the transformer tap changer, short circuits of winding turn-to-turn, blockage of oil passages, and cooling system failures can lead to changes in the composition of dissolved gases, oil temperature, and winding temperature of the internal transformer^{3–5}. The traditional method of transformer governance and condition assessment is mainly based on the analysis of the problem of status quo and real-time assessment, observing the transformer operating parameters through the monitoring platform and setting alarm thresholds^{6,7}. Maintenance personnel can take appropriate measures after receiving the alarm for sudden abnormal events without warning. Inversely, for the relatively gentle changes in the working conditions, the performance of the early warning ability is insufficient^{8,9}. Therefore, how to effectively carry out the short-term state predictive assessment of transformers realize the prediction and early warning, and take preventive measures to avoid the occurrence of faults is a key issue in the construction of a strong smart grid¹⁰.

There are some conventional approaches to predict the power transformer data, such as the autoregressive integral moving average (ARIMA)^{11,12}, random walk (RW)¹³, generalized autoregressive conditional heteroscedasticity (GARCH)^{14,15} and vectorial autoregression (VAR)¹⁶. These conventional approaches have satisfactory prediction performance for linear correlation variables, but they cannot capture the nonlinear characteristics of data. Due to the limitations of conventional approaches, plentiful nonlinear artificial intelligence and deep learning methods rise in response to the proper time and conditions^{17–20}, and can be employed for data prediction based on the time series, such as artificial neural network (ANN)²¹, support vector machine (SVM)²² and recurrent neural network (RNN)²³. However, aiming at the problems of gradient

¹Ural International Institute of Rail Transit, Shandong Polytechnic, Jinan 250104, China. ²Jinan Zhongran Technology Development Co., Ltd, Jinan 250104, China. ³Shandong Huineng Electric Co., Ltd, Zibo 255022, China. ⁴School of Mechanical and Electronic Engineering, Shandong Agriculture and Engineering University, Zibo 255300, China. ✉email: cjp_sdp_edu@163.com

explosion and disappearance in RNN, the long short-term memory network (LSTM) has been developed based on RNN^{24,25}. The effective message can be extracted through the various gate structures of LSTM and gathered in historical data during the training process. In addition, in contrast to LSTM, BiLSTM consists of two layers of LSTM, which can take full advantage of both forward and reverse information²⁶. Therefore, it is especially suitable for solving data forecasting problems based on the time series. Furthermore, considering that historical data contribute differently to data at different points in time, while current neural network models (including BiLSTM) contribute equally to each point in time to be predicted, attention mechanisms have been developed²⁷. In order to further improve the prediction accuracy of the models, Attention mechanisms are usually introduced in deep learning models. Attention mechanisms can obtain valid information and important spatiotemporal features from new coding sequences. One of the drawbacks of conventional neural networks is the poor scalability due to the complete connectivity of neurons, which is overcome by convolutional neural networks (CNN). CNN enhances the efficiency of the algorithm and decreases the number of parameters. Many kinds of literature have demonstrated that CNN has the advantage of extraction and reorganization^{28,29}. Therefore, based on the advantages of extracting effective information from the Attention mechanism layer and CNN capturing the hierarchical structure, it is of great importance to study the combination model to predict the power transformer data.

With the application and development of smart grids, transformers are also gradually developing in the direction of intelligence³⁰. Transformer online monitoring can provide real-time monitoring of transformer operation data. Through the processing and analysis of big data, transformer failure can achieve early discovery and early treatment, which is conducive to solving the problem of transformer condition evaluation and prediction³¹. Therefore, judging and evaluating the transformer status based on data prediction and data mining offers a different approach to online monitoring of transformer faults. Multi-source data fusion and data extraction is an advanced data technology for data-driven insights and data correlation analysis, i.e. identifying the relationships, trends, and linkages between massive and complicated datasets³². Multiple factors hamper the modeling and analysis of their interaction and complex relationships with the creation of data sources for knowledge acquisition and, eventually, the process of decision finding. Consequently, leveraging the intellectual properties of models that can handle massive and sophisticated datasets can lead to more acceptable results. A wide range of investigations in the last two decades have taken advantage of the possibilities offered by data excavation techniques in diverse fields, such as prediction of atmospheric pollutant levels³³, prediction of the optical depth of aerosols³⁴, and mapping of subsidence susceptibility³⁵.

Apriori is a powerful approach based on information retrieval. It has been deployed for application in learning exploration and forecasting, such as wind speed³⁶, landslides³⁷ and road accidents³⁸. Research in the area of smart diagnostics for power devices typically utilizes various methods of machine learning, such as ANN, Support Vector Machines (SVM), and Random Forests (RF), as well as data exploration techniques, such as Boosted Generalized Additive Models (BGAM), which seek to determine correlations through modeling the mathematical relationships among various performance properties^{39–42}. In spite of the established performance of these techniques, they are not able to deliver connection patterns between the events and the contributory factors. Thus, the approaches are not universal, and for implementation in other areas, complicated parameters need to be reassigned and re-run. The primary strength of the Apriori approach is that, on the basis of the patterns it produces for an occurrence, it is feasible to extrapolate these patterns to similar occurrences and detect associations without having to re-run the procedure. The Apriori methodology, a powerful rule-based data exploration technique, is deployed for the first time in the current work to detect faults through the detection of association patterns by analyzing the complicated behavior of various factors in the transformer load rate and the top oil layer, winding temperature and fusion indices of the gas components. Hence, concrete decisions can be undertaken to enhance the applicability of transformer fault analysis methods by reducing handling costs, decreasing data demands, and eliminating associated problems.

How to play the role of data mining methods to support the stable operation of transformers based on the means of online monitoring of data, how to realize the intelligent diagnosis of transformer operation status by using the constantly changing operation data, as well as to realize the predictive assessment and warning of transformers are still the outstanding issues of the intelligent operation of transformers. To the best knowledge of the authors of this paper, there are no works of literature on the application of the intelligent fault diagnosis of transformers based on multi-source data fusion and data mining. Given the context discussed, the contribution of this paper is fourfold:

- (1) Intelligent fault diagnosis of the transformer based on multi-source data fusion and data mining is modeled to realize the stable operation of the transformer under different operating conditions by predictive assessment and early warning.
- (2) The components of dissolved gas, upper oil temperature, winding temperature, and load rate of the transformer are selected as state characteristic parameters, and data fusion is performed on the multiple components of dissolved gas of the transformer.
- (3) The state characteristic parameters of the transformer are predicted by CNN-BiLSTM-Attention to ensure the basis of data application and data accuracy in the predictive evaluation of the power transformer.
- (4) The correlation analyses of dissolved gas composition, upper oil temperature, and winding temperature under different load rates are achieved by multi-source data fusion and Apriori correlation analysis. The remainder of the paper is organized as follows. Section “[Transformer state characteristic parameter selection and data fusion](#)” gives the transformer state characteristic parameter selection and data fusion. Section “[Data prediction](#)” deploys the data prediction model based on CNN-BiLSTM-Attention fusion. Section “[Intelligent fault diagnosis of transformers based on correlation analysis](#)” further proposes the intelligent fault diagnosis of transformers based on correlation analysis. In “[Experimental results](#)” Section,

experimental results are investigated to demonstrate the proposed method. Section “Conclusions” draws the main conclusions.

Transformer state characteristic parameter selection and data fusion

In this paper, the condition assessment of the transformer is proposed to be related to the composition of dissolved gases in the transformer, the upper oil temperature, the winding temperature, and the load rate, where the composition of the transformer oil mainly contains H_2 , CH_4 , C_2H_2 , C_2H_4 , C_2H_6 , CO and CO_2 . The documented basis for fault diagnosis of dissolved gas in transformer oil mainly includes GB/T 7252-2001 Guidelines for Analysis and Judgement of Dissolved Gas in Transformer Oil and other related standards or guidelines. These documents provide the framework of transformer fault diagnosis based on dissolved gas analysis and the corresponding fault judgment basis. The content in the oil mainly includes H_2 , CH_4 , CO , CO_2 , C_2H_6 , C_2H_4 , C_2H_2 and other gases. Different types of faults produce different changes in gas composition and content. Therefore, The data fusion of multiple compositions of transformer dissolved gases, the upper oil temperature, and the winding temperature are used as the characteristic parameters for condition assessment in this paper.

The multiple components of the transformer dissolved gas are first normalized and the j -th component p_{ij} on the t -th time scale after normalization can be expressed as:

$$p_{tj} = v_{tj} / \sum_{j=1}^m v_{tj} \quad (1)$$

where v_{tj} denotes the j -th component on the t -th time scale before normalization.

The entropy value e_j of the j -th component can be denoted as:

$$e_j = -K \sum_{j=1}^m (p_{ij} \times \ln p_{ij}) \quad (2)$$

where $K = 1/\ln m$. If $p_{ij} = 0$, then $\lim_{p_{ij} \rightarrow 0} p_{ij} \ln p_{ij} = 0$.

Thus, the entropy weights can be expressed as follows.

$$w_j = 1 - e_j / \sum_{j=1}^m (1 - e_j) \quad (3)$$

The expected value E of the weights is obtained by adaptive optimization, which can be defined as:

$$E = \sum_{j=1}^m w_j v_{tj} \quad (4)$$

A prediction error exists when the actual value REA and the predicted expected value are not equal, then the prediction error can be shown as follows.

$$ERR = \frac{1}{2} \sum_{j=1}^m (REA_j - E_j)^2 \quad (5)$$

where ERR indicates the prediction error between the actual value and the predicted expected value.

The weights are continuously adjusted by using an error gradient descent algorithm, then the adjusted values of weights can be defined as:

$$\Delta w_j = -\eta \frac{\partial ERR}{\partial w_j} \quad (6)$$

where the negative sign indicates a gradient decrease and η is the scaling factor.

$$\begin{aligned}
 \frac{\partial ERR}{\partial w_j} &= \frac{\partial}{\partial w_j} \frac{1}{2} \sum_{j=1}^m (REA_j - E_j) \\
 &= \frac{1}{2} \frac{\partial}{\partial w_j} \sum_{j=1}^m (REA_j - E_j)^2 \\
 &= \frac{1}{2} \sum_{j=1}^m 2(REA_j - E_j) \frac{\partial}{\partial w_j} (REA_j - E_j) \\
 &= \sum_{j=1}^m (REA_j - E_j) \frac{\partial}{\partial w_j} (REA_j - w_j z_j) \\
 &= \sum_{j=1}^m [(REA_j - E_j) \times (-z_j)]
 \end{aligned}
 \tag{7}$$

Therefore, based on the adjusted values of weights, the weights can be updated by iteration:

$$w_j = w_{j-1} + \Delta w \tag{8}$$

The dissolved gas index on the t -th time scale is:

$$F_t = w_1 \times \nu_{t1} + w_2 \times \nu_{t2} + \dots + w_m \times \nu_{tm} \tag{9}$$

Data prediction

The network structure of the data prediction model based on CNN-BiLSTM-Attention fusion is shown in Fig. 1. For the input data, firstly, the longitudinal feature extraction module with convolutional layer as the core is passed, and on the basis of this, the horizontal feature extraction module with BiLSTM network as the core is passed. These two modules are cascaded back and forth to fully explore the data features from both longitudinal time points and horizontal time series perspectives. By adding the attention mechanism, the model pays more attention to the feature change pattern of the data near the moment of fault occurrence. Compared with the complex deep learning network structure, the model has a simple structure and faster speed, which can achieve timely and accurate data prediction.

CNN network

CNN is a deep learning network with a convolutional structure that extracts detailed features of data and is commonly used in the field of image classification. In the longitudinal feature extraction module, each CNN layer includes three computations: convolution, normalization, and activation function. Taking the winding temperature data as input, the formula for the convolution operation can be defined as:

$$z_j^l = \sum_{i=1}^m z_j^{l-1} * k_{ij}^l + b_j^l, j = 1, \dots, n \tag{10}$$

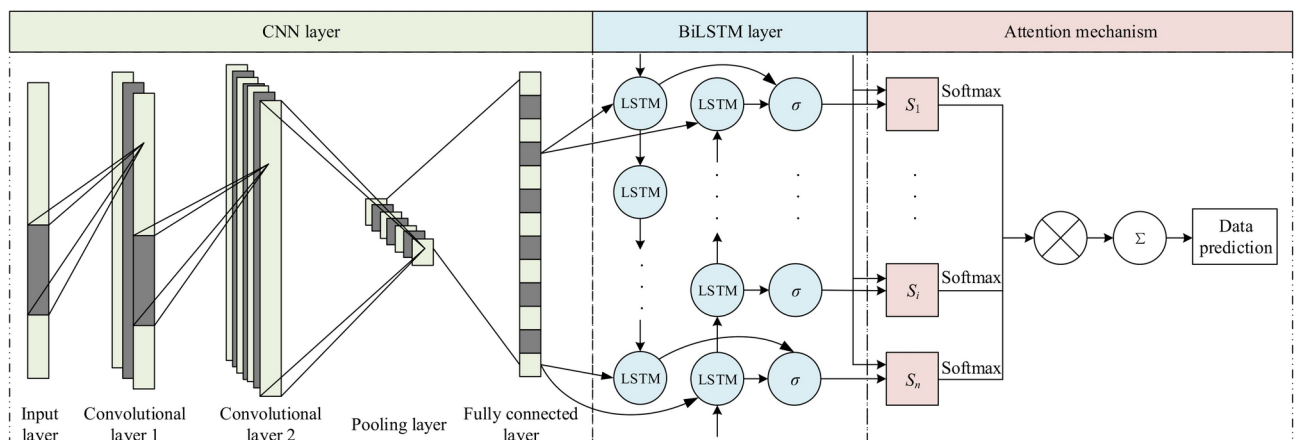


Fig. 1. Network structure of CNN, BiLSTM and Attention model.

where z_j^l represents the feature vector of the j -th feature surface of the convolutional output, z_j^{l-1} represents the winding temperature signal of the i -th input feature surface, k_{ij} represents the parameter of the j -th convolution kernel connected to the i -th input feature surface, and b_j^l represents the bias of the j -th convolution kernel.

Convolution is able to extract linear features of the data by performing a linear computation of multiplying and summing the elements of the input data with a convolution kernel, while the convolution operation moves over the input sequence by sliding the convolution kernel in a way that captures structural features of local patterns in the input data.

As the model is trained, the distribution of data will be shifted. Therefore, normalization is used to avoid the problem of gradient disappearance caused by the data falling into the saturation region of the activation function and to speed up the convergence of the model. The normalization formula is as follows:

$$z_{LN} = \frac{z - E(z)}{\sqrt{\text{Var}(z) + \epsilon}} \times \alpha + \beta \quad (11)$$

where z is the result of the convolution operation on the winding temperature, $E(z)$ is the mean, $\text{Var}(z)$ is the variance, ϵ is a small amount greater than 0 to prevent the denominator from being zero, which is generally taken as 10^{-5} , α and β are trainable parameters.

The essence of the activation function is to perform a nonlinear transformation of the input data to extract the nonlinear features of the data and increase the fitting ability of the network. The activation function used in this paper is the ReLU function:

$$x = \text{ReLU}(z_{LN}) = \begin{cases} z_{LN} & \text{if } z_{LN} \geq 0 \\ 0 & \text{if } z_{LN} < 0 \end{cases} \quad (12)$$

Compared with other activation functions, the ReLU function solves the problem of vanishing gradient on positive intervals. In addition, since it only needs to judge whether the input is greater than zero, its computational speed and convergence speed are faster, which lays the foundation for the model to be able to achieve data prediction quickly. The feature vectors are passed into the pooling layer, which calculates the average value of the data for each output channel, increasing the robustness of the model and reducing the number of parameters, which prevents model overfitting and speeds up model convergence. The next flattening layer unfolds the data of each convolutional channel in one dimension for the transition between CNN and BiLSTM.

BiLSTM network

The BiLSTM network is a variant of the LSTM network and is formed by combining the forward LSTM network and the backward LSTM network. LSTM network is an improvement of Recurrent Neural Network (RNN), which improves the short-term memory problem of RNN due to the disappearance of gradient, which causes the more distant information to have almost no effect on the current moment by adding three special gate structures and memory units. The LSTM unit structure is shown in Fig. 2.

The input set is $\{x_1, x_2, \dots, x_t\}$, where $x_t = \{x_{t,1}, x_{t,2}, \dots, x_{t,k}\}$, which denotes the k -dimensional vector data at time t . The forgetting gate f_t , the candidate state of the memory cell \tilde{c}_t , the input gate i_t , the state of the memory cell c_t , the output gate o_t and the hidden layer output value h_t can be expressed as follows.

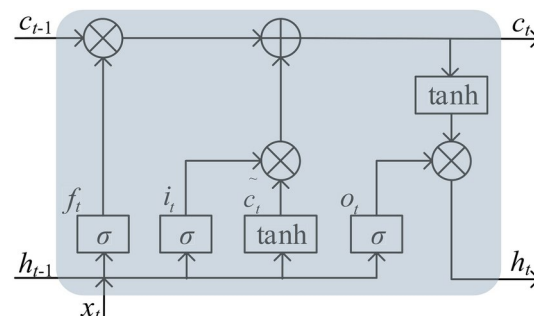


Fig. 2. LSTM unit structure.

$$\begin{cases} f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ \tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \\ i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ c_t = f_t \otimes c_{t-1} \oplus i_t \otimes \tilde{c}_t \\ h_t = o_t \otimes \tanh(c_t) \end{cases} \quad (13)$$

where W_f is the weight matrix of the oblivious gate, b_f is the bias of the oblivious gate, σ and \tanh denote the sigmoid and hyperbolic tangent activation function, W_c , W_i , b_c and b_i denote the weight matrix and bias corresponding to the candidate state \tilde{c}_t and the input gate i_t , respectively, W_o and b_o are the weight matrix and bias of the output gate O_t , \oplus and \otimes denote the add and multiply, respectively.

In addition, the activation functions of σ and \tanh can be shown as:

$$\sigma(x) = 1 / (1 + \exp(-x)) \quad (14)$$

$$\tanh(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x)) \quad (15)$$

From the above, it can be seen that the network parameters of LSTM are trained on the data in the order from front to back, which is low utilization of the data and cannot fully extract the intrinsic characteristics of the data in the time series. BiLSTM network combines forward LSTM and backward LSTM, which can simultaneously extract the forward and backward historical transverse features of the data, and further explore the intrinsic connection between the current data and the past and future data, so as to improve the utilization rate of the data and the prediction accuracy of the model. The structure of the BiLSTM network is shown in Fig. 3.

The hidden layer output value h_t of BiLSTM consists of forward vector \vec{h}_t and inverse vector \overleftarrow{h}_t , where the forward vector and inverse vector output are:

$$\vec{h}_t = \text{LSTM}(h_{t-1}, x_t, c_{t-1}), t \in [1, T] \quad (16)$$

$$\overleftarrow{h}_t = \overleftarrow{\text{LSTM}}(h_{t+1}, x_t, c_{t+1}), t \in [T, 1] \quad (17)$$

Therefore, the output of BiLSTM at time t can be expressed as:

$$y_t = \sigma(W_y \cdot [\vec{h}_t, \overleftarrow{h}_t] + b_y) \quad (18)$$

where W_y and b_y are the weight matrix and bias terms, respectively.

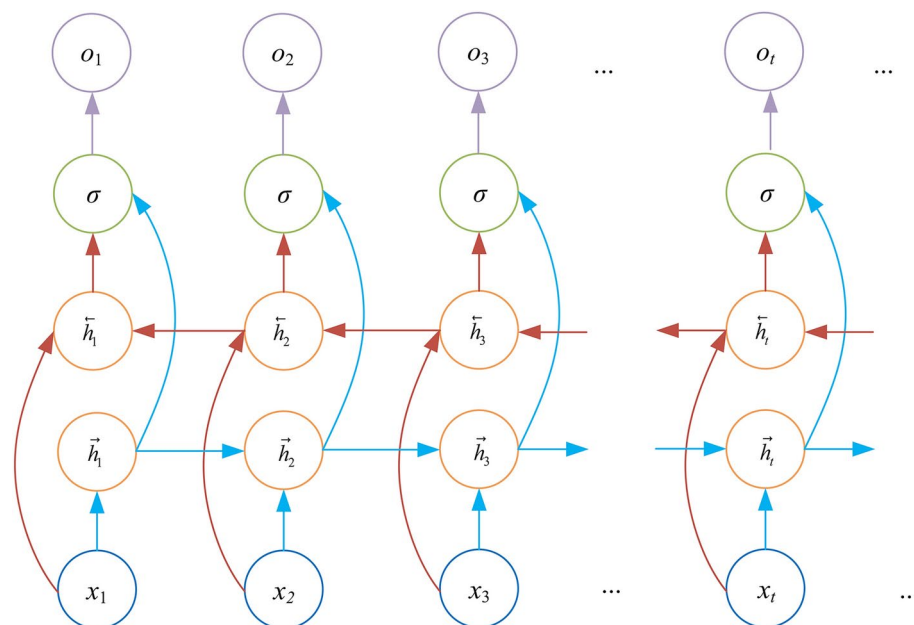


Fig. 3. BiLSTM structure.

The forward transmission layer extracts the forward history of faults in the direction of the time series, from front to back. The backward transmission layer traces the historical feature correlations of the faults from backward to forward in the reverse direction of the time series. By fusing the two features, the horizontal features of the data are obtained.

Attention mechanism

The Attention mechanism is an idea based on human visual attention that assigns different weights to different input features to enhance important features and avoid irrelevant information from influencing the final result, thus improving the performance and effectiveness of the model. The Attention mechanism is depicted in Fig. 4. Specifically, the implementation of an Attention mechanism typically involves the following steps.

Step 1 A set of query vectors $y = [y_1, y_2, \dots, y_n]$, is obtained by encoding the input sequence q .

Step 2 Subsequently, using the scoring function s , the expression can be expressed as:

$$e_{ij} = \tanh(W_1 * h_i + W_2 * h_j + b) \quad (19)$$

where h_i and h_j are the hidden layer states, e_{ij} indicates the correlation between the i -th state and the j -th state, W is the weight, b represents the offset vector.

Step 3 Then, the Softmax function is used for normalization to convert the value of each correlation into a probability weight a_i , which can be calculated as follows:

$$a_{ij} = \text{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_j \exp(e_{ij})} \quad (20)$$

where a_{ij} is the attention weight of j to i , and $\sum_j a_{ij} = 1$.

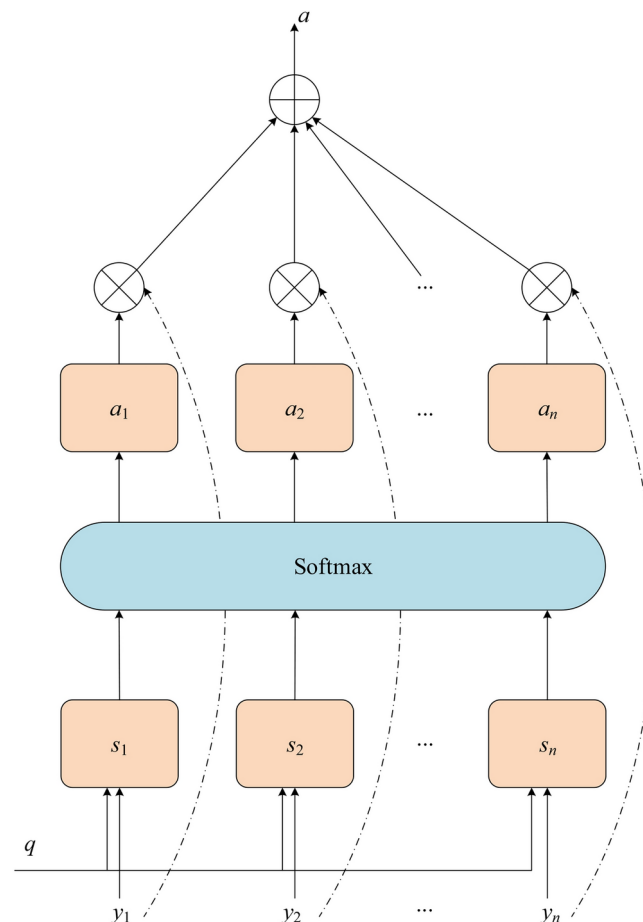


Fig. 4. Attention mechanism.

Step 4 The final Attention output value H_i is calculated from the weight coefficients a_{ij} input vector h_j , which is shown as following:

$$H_i = \sum_j a_{ij} * h_j \quad (21)$$

Intelligent fault diagnosis of transformers based on correlation analysis

Based on the load rate of the transformer, the transient stability of the transformer is identified through the multivariate data correlation analysis of the composition of dissolved gas, upper oil temperature, and winding temperature, so as to avoid the situation of leakage and misjudgement of a single data source. Through the correlation analysis of multivariate data, the causal pairs formed by the transformer load rate, the composition of dissolved gases, the temperature of the upper layer of oil, and the temperature of the windings are taken into account, and the correlation rules of support and confidence are used to determine the data ranges corresponding to the correlated data.

The data is divided into n intervals based on the maximum and minimum values of the composition of dissolved gases, upper oil temperature, winding temperature, and load rate data of the transformer, which can be expressed as:

$$\begin{cases} G_{\Delta}^A = G_{\max}^A - G_{\min}^A/n \\ O_{\Delta}^B = O_{\max}^B - O_{\min}^B/n \\ W_{\Delta}^C = W_{\max}^C - W_{\min}^C/n \\ L_{\Delta} = L_{\max} - L_{\min}/n \end{cases} \quad (22)$$

where load rate $L = (S_{\text{real}}/S_{\text{rated}}) \times 100\%$, S_{real} represents the apparent power, S_{rated} denotes rated capacity of transformer.

The intervals of the composition of dissolved gases, upper oil temperature, winding temperature, and load rate for the transformer can be shown as follows:

$$\begin{cases} [G_{\min}, G_{\min} + G_{\Delta}], [G_{\min} + G_{\Delta}, G_{\min} + 2G_{\Delta}], \dots, [G_{\min} + (n-1)G_{\Delta}, G_{\min} + nG_{\Delta}] \\ [O_{\min}, O_{\min} + O_{\Delta}], [O_{\min} + O_{\Delta}, O_{\min} + 2O_{\Delta}], \dots, [O_{\min} + (n-1)O_{\Delta}, O_{\min} + nO_{\Delta}] \\ [W_{\min}, W_{\min} + W_{\Delta}], [W_{\min} + W_{\Delta}, W_{\min} + 2W_{\Delta}], \dots, [W_{\min} + (n-1)W_{\Delta}, W_{\min} + nW_{\Delta}] \\ [L_{\min}, L_{\min} + L_{\Delta}], [L_{\min} + L_{\Delta}, L_{\min} + 2L_{\Delta}], \dots, [L_{\min} + (n-1)L_{\Delta}, L_{\min} + nL_{\Delta}] \end{cases} \quad (23)$$

In order to facilitate the presentation of the causal pairs of data, the composition of the dissolved gases, the upper oil temperature, and the winding temperature intervals may be briefly described as $\{G_1, G_2, \dots, G_n\}$, $\{O_1, O_2, \dots, O_n\}$ and $\{W_1, W_2, \dots, W_n\}$. Similarly, the interval of the load rate of the transformer is abbreviated as $\{L_1, L_2, \dots, L_n\}$.

In addition, the load rate intervals, the dissolved gases, the upper oil temperature, and the winding temperature intervals can be matched to form causal pairs of load rate and composition of dissolved gas, upper oil temperature, and winding temperature, respectively, which can be expressed as follows.

$$\begin{cases} [L_1, G_1], [L_1, G_2], \dots, [L_2, G_1], [L_2, G_2], \dots, [L_n, G_n] \\ [L_1, O_1], [L_1, O_2], \dots, [L_2, O_1], [L_2, O_2], \dots, [L_n, O_n] \\ [L_1, W_1], [L_1, W_2], \dots, [L_2, W_1], [L_2, W_2], \dots, [L_n, W_n] \end{cases} \quad (24)$$

Then, calculate the support level Sup and confidence level Con of each causal pair:

$$\begin{cases} Sup(L_a \cap G_b) = \text{count}(L_a \cap G_b)/N_G \\ Sup(L_a \cap O_b) = \text{count}(L_a \cap O_b)/N_O \\ Sup(L_a \cap W_b) = \text{count}(L_a \cap W_b)/N_W \end{cases} \quad (25)$$

$$\begin{cases} Con(L_a \cap G_b) = \text{count}(L_a \cap G_b)/\text{count}(L_a) \\ Con(L_a \cap O_b) = \text{count}(L_a \cap O_b)/\text{count}(L_a) \\ Con(L_a \cap W_b) = \text{count}(L_a \cap W_b)/\text{count}(L_a) \end{cases} \quad (26)$$

where L_a denotes the interval of a -th load rate, G_b , O_b and W_b are the interval of b -th dissolved gas upper oil temperature, and winding temperature, respectively, $\text{count}(L_a \cap G_b)$ represents the number of causal pairs belonging to both the a -th load rate interval and the b -th interval of the dissolved gas, N_G indicates the total number of causal pairs in the set of load rate and dissolved gases, $\text{count}(L_a \cap O_b)$ represents the number of causal pairs belonging to both the a -th load rate interval and the b -th interval of the upper oil temperature, N_O

indicates the total number of causal pairs in the set of load rate and the upper oil temperature, $count(L_a \cap W_b)$ represents the number of causal pairs belonging to both the a -th load rate interval and the b -th interval of the winding temperature, N_G indicates the total number of causal pairs in the set of load rate and winding temperature.

The causal pairs of the dissolved gases in the transformer, the upper oil temperature, and the winding temperature need to be greater than or equal to the threshold of minimum support and confidence at the same time, and the flowchart of correlation analysis for load rate, dissolved gas, oil temperature of upper layer and winding temperature can be shown in Fig. 5.

Experimental results

In this paper, a three-phase oil-immersed amorphous alloy distribution transformer was chosen as the object of study in the study of intelligent fault diagnosis of the transformer. The simulation corresponding to the proposed model is implemented in MATLAB platform, on a PC with Intel Core i7, 5.4GHz processor, and 32 GB of memory. The transformer is equipped with a dissolved gas analyzer in oil, and a predictive assessment of transformer condition is performed using the actual transformer load factor, the composition of dissolved gases, upper oil temperature, and winding temperature from a transformer monitoring platform in this paper. The data used are from one sampling point at 15-minute intervals from 1 May 2023 to 31 August 2023 at the substation, and the sample of data sets are shown in Fig. 6. This transformer condition predictive assessment involves multivariate feature inputs and provides a field application example reference for subsequent studies.

The algorithm of the transformer condition predictive assessment method based on correlation analysis is as follows: (1) Data fusion of multiple components of dissolved gases in the transformer is carried out by the improved entropy weight method, and the data after fusion with the dissolved gases, the upper oil temperature, winding temperature, and load rate are selected as the characteristic parameters for condition assessment. (2) Initial training and prediction based on BiLSTM neural network. Select the transformer load rate and the dissolved gas, upper oil temperature, and winding temperature data after data fusion as inputs. Divide the training set and test set, initialize the input dimension, output dimension, iterations, and activation function

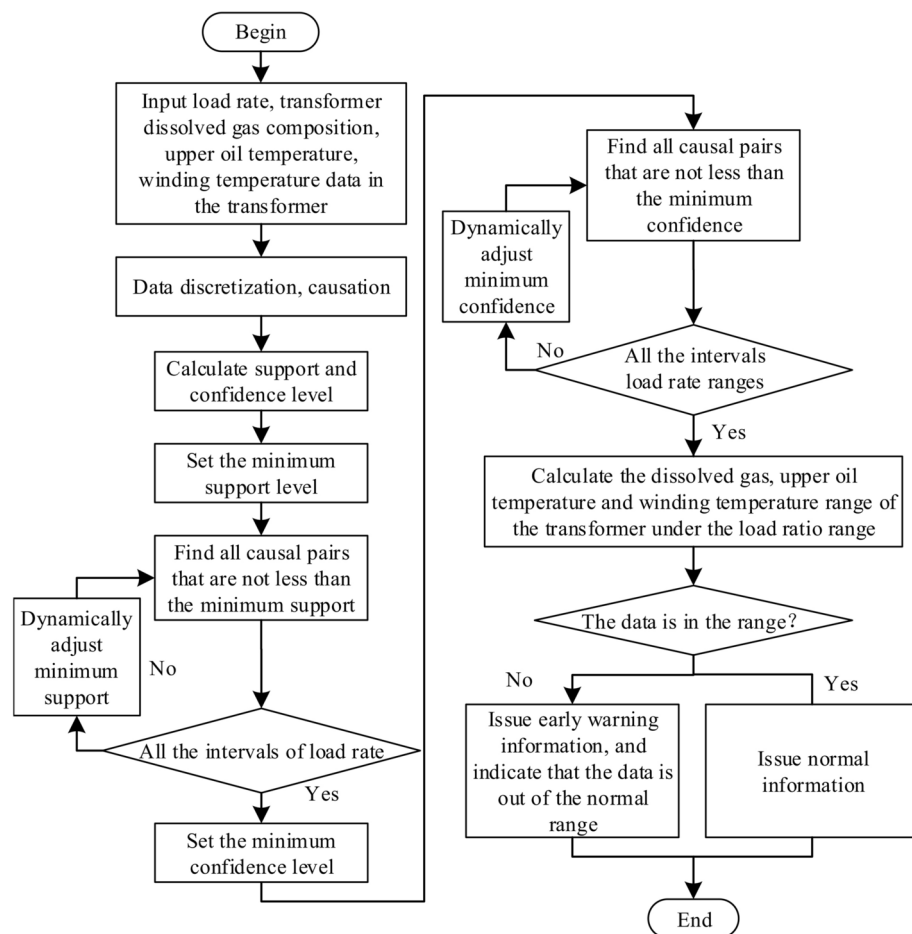


Fig. 5. The flowchart for correlation analysis of load rate, dissolved gas, oil temperature of upper layer and winding temperature.

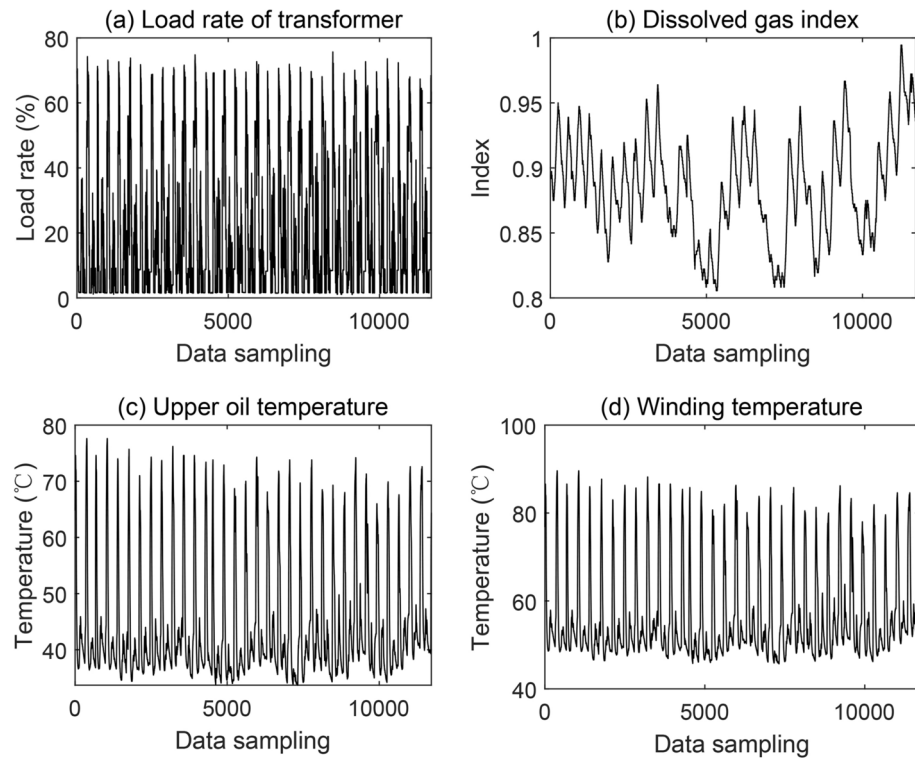


Fig. 6. The sample of data sets.

based on experience, supervise the training of the model according to the gradient descent algorithm, and get the prediction value of the data. (3) Through the correlation analysis of multivariate data, the transformer load rate, the composition of dissolved gas in the transformer, the upper oil temperature, and the winding temperature are considered to form the causal pairs, and the correlation rules of support and confidence are used to determine the data range corresponding to the correlated data. (4) If the predicted dissolved gas, upper oil temperature, and winding temperature are not within the corresponding load ratio interval, the power system monitoring platform will issue a warning signal. If the dissolved gas is not in the corresponding interval, you can check the data in Table 2 to find the corresponding fault information. (5) For the purpose of evaluating the performance of data prediction, the prediction results are evaluated using root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination R^2 . The RMSE, MAE, and R^2 can be computed by the following formulas:

$$\text{RMSE} = \sqrt{\sum_{i=1}^n (Y_{pi} - Y_{ti})^2 / n} \quad (27)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_{pi} - Y_{ti}| \quad (28)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{pi} - Y_{ti})^2 / n}{\sum_{i=1}^n (Y_{ti} - \bar{Y}_t)^2 / n} \quad (29)$$

where n is the number of samples, Y_{pi} is the i -th predicted value, Y_{ti} is the i -th actual value, and \bar{Y}_t is the average of the actual values. The closer R^2 is to 1, the better the prediction effect of the model is.

Data prediction

In this paper, the BiLSTM neural network uses the sigmoid function as the activation function of neurons. A total of 122 days of data from 1st May 2023 to 30th August 2023 were used as the training set, data from 31st August 2023 were used as the test set, and the RMSE, MAE, and R^2 are used as the evaluation index of data prediction. The threshold r and the learning efficiency η of the error gradient are 0.1 and 0.05, respectively, and the learning factors dec_r and dec_η are 0.05 and 0.8, respectively. The training set and the test set are inputted into the network after the completion of the training, and the prediction value \hat{Y}_{train} of the training set and the

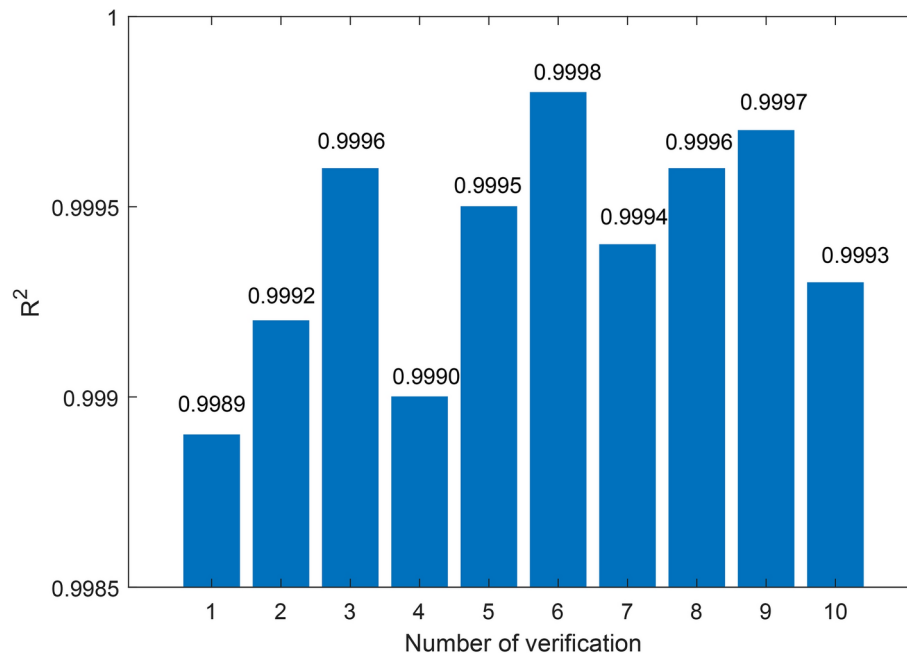


Fig. 7. The result of the ten-fold cross-validation.

| Algorithm | MAE | RMSE | R ² | Training time (s) |
|----------------------|-------|--------|----------------|-------------------|
| BP | 2.721 | 23.073 | 0.9547 | 35 |
| SVM | 5.423 | 6.259 | 0.9612 | 47 |
| LSTM | 0.717 | 0.891 | 0.9785 | 51 |
| CNN-LSTM | 0.659 | 0.823 | 0.9871 | 75 |
| CNN-LSTM-Attention | 0.553 | 0.694 | 0.9986 | 88 |
| CNN-BiLSTM-Attention | 0.474 | 0.586 | 0.9994 | 102 |

Table 1. Prediction errors of different models.

prediction value Y_a of the test set are computed to obtain the training set's predicted value and the residuals are calculated as $Y_e = Y_a - Y_{train}$.

In order to verify the reliability and superiority of the CNN-BiLSTM-Attention model used in this paper on the prediction of the battery charge state of an electric loader, the data of winding temperature is selected for prediction analysis, and experimental comparisons are also made with support vector machine (SVM), CNN-LSTM and CNN-LSTM-Attention models. Using the ten-fold cross-validation method, the dataset is divided into ten subsets, nine of which are used as the training set and the remaining one as the validation set in turn, and the training is repeated ten times. The prediction results and errors of several models are shown in Table tab. Prediction errors of different models, and the ten-fold cross-validation results of the CNN-BiLSTM-Attention model are shown in Fig. 7. It can be seen that the model used in this paper has an accuracy of up to 0.9998 and down to 0.9989 in the ten-fold cross-validation experiments, with an average accuracy of 0.9994. It shows that the model does not fall into the overfitting state under different data subsets and has some stability. In addition, the average MAE of the model proposed in this paper is 0.474% and the average RMSE is 0.586%. Compared with the CNN-LSTM model, the MAE and RMSE are improved by 28.07% and 28.80%, respectively, and compared with the CNN-LSTM-Attention model, the MAE and RMSE are improved by 14.29 and 15.56%, respectively, which are superior to some extent.

Combined with Fig. 7 and Table 1, it can be found that the CNN-BiLSTM-Attention model is significantly better than the CNN-LSTM-Attention model. This is because the BiLSTM structure can read the data information in the forward and backward direction respectively, mine the intrinsic connection between the data, fit the current data, and improve the prediction accuracy. Hence, the proposed algorithm brings together the advantages of multiple models and results in the highest accuracy, while at the same time the computational speed is relatively slow. In summary, the CNN-BiLSTM-Attention model used in this paper can predict the winding temperature data more accurately and has a certain generalization ability. Meanwhile, the CNN-BiLSTM-Attention prediction model is applied to the composition of dissolved gas, upper oil temperature, and winding temperature in this paper.

| Fault type | H ₂ | CH ₄ | C ₂ H ₂ | C ₂ H ₄ | C ₂ H ₆ | CO | CO ₂ |
|------------------------------|----------------|-----------------|-------------------------------|-------------------------------|-------------------------------|----|-----------------|
| Partial discharge | ✓ | ✓ | – | – | – | – | – |
| Arc discharge | ✓ | ✓ | ✓ | ✓ | – | – | – |
| Upper oil temperature 140 °C | – | ✓ | ✓ | ✓ | ✓ | – | – |
| Solid material | – | – | – | – | – | ✓ | ✓ |

Table 2. Fault type and corresponding gas composition.

| Constituent | Prewarning value μL/L | Constituent | Prewarning value μL/L |
|-------------------------------|-----------------------|-------------------------------|-----------------------|
| H ₂ | 25.20 | C ₂ H ₆ | 15.1 |
| CH ₄ | 11.40 | CO | 315 |
| C ₂ H ₂ | 1.55 | CO ₂ | 1850 |
| C ₂ H ₄ | 5.60 | | |

Table 3. The residual values of gas composition prediction and warning values.

| Load rates | Dissolved gas index | Upper oil temperature (°C) | Winding temperature (°C) |
|-----------------------------|---------------------|----------------------------|--------------------------|
| $0 \leq L \leq 12.610$ | [0.805, 0.837] | [33.700, 41.016] | [45.700, 53.017] |
| $12.610 \leq L \leq 25.230$ | [0.805, 0.837] | [33.700, 41.016] | [45.700, 53.017] |
| $25.230 \leq L \leq 37.840$ | [0.837, 0.868] | [41.016, 48.333] | [45.700, 60.333] |
| $37.840 \leq L \leq 50.450$ | [0.868, 0.900] | [41.016, 48.333] | [45.700, 67.651] |
| $50.450 \leq L \leq 63.070$ | [0.900, 0.963] | [41.016, 70.283] | [45.700, 67.651] |
| $63.070 \leq L \leq 75.680$ | [0.900, 0.994] | [48.333, 77.600] | [53.017, 67.651] |

Table 4. Corresponding interval for correlation analysis under the different load rates.

Intelligent fault diagnosis

Upper oil temperature and winding temperature can directly reflect the oil overheating and winding overheating in the transformer. The composition of dissolved gases in the oil is a condition assessment index reflecting the specific faults inside the transformer. The transformer is in normal operation, and the dissolved gases in the oil mainly include O₂ and N₂. When the transformer fault occurs, the composition and concentration of dissolved gases in the oil will change, and the characteristic gases of the fault may include H₂, CH₄, C₂H₂, C₂H₄, C₂H₆, CO and CO₂. Different fault types correspond to different components of the characteristic gas. The main types of transformer faults and the corresponding gas composition are shown in Table 2, and the residual values of gas composition prediction and warning values are shown in Table 3.

Through the correlation analysis of the transformer load rate and the data of transformer oil dissolved gas, upper oil temperature, and winding temperature, the correspondence between the current load rate and the data of dissolved gas, upper oil temperature, and winding temperature is obtained, and six data correspondence intervals are divided at an environment temperature of 10 °C, as shown in Table 4. Based on the data in Table 4, the curves of dissolved gas, upper oil temperature, and winding temperature data are obtained under different loads, so as to carry out a short-term predictive assessment of the transformer. The three-phase winding temperature rise intervals and predicted temperature rises are shown in Fig. 8.

As can be seen from Fig. 8, in the vicinity of sample point 88, the dissolved gas, upper oil temperature, and winding temperature data are not within the normal range of intervals and have not returned to the normal range until sample point 96, then it indicates that there may be a fault within the transformer. At the same time, through the detection of gas composition found that C₂H₂ and C₂H₄ exceeded the warning value, and it is presumed that the arc discharge phenomenon. The power monitoring platform sends out an early warning message and issues a notification for timely overhaul and maintenance. In addition, the cause of the transformer fault was determined by realizing the field test on site, and the fault type was the same as that detected by the method proposed in this paper. Through the field test to confirm the cause of the discharge and oil filtration, degassing treatment processing, to achieve the prediction of the accident early warning, to prevent further expansion of the accident.

From the analyses in Table 4 and Fig. 8, it can be seen that the composition of dissolved gases, the upper oil temperature, and the winding temperature of the transformer are affected at different load rates. In addition to this, the stable operation of the transformer is also affected by the environment temperature. The standard regulations for oil-immersed transformer oil in direct contact with the atmosphere of the top oil temperature rise shall not exceed 55 °C, and the average temperature rise of the windings shall not exceed 65 °C. The transformer will often work at 80–1000 °C, long-term in the role of higher temperatures will gradually age brittle, in the range of 80–140 °C, the transformer temperature rises 8 °C for each, and the shortening of its insulation life of

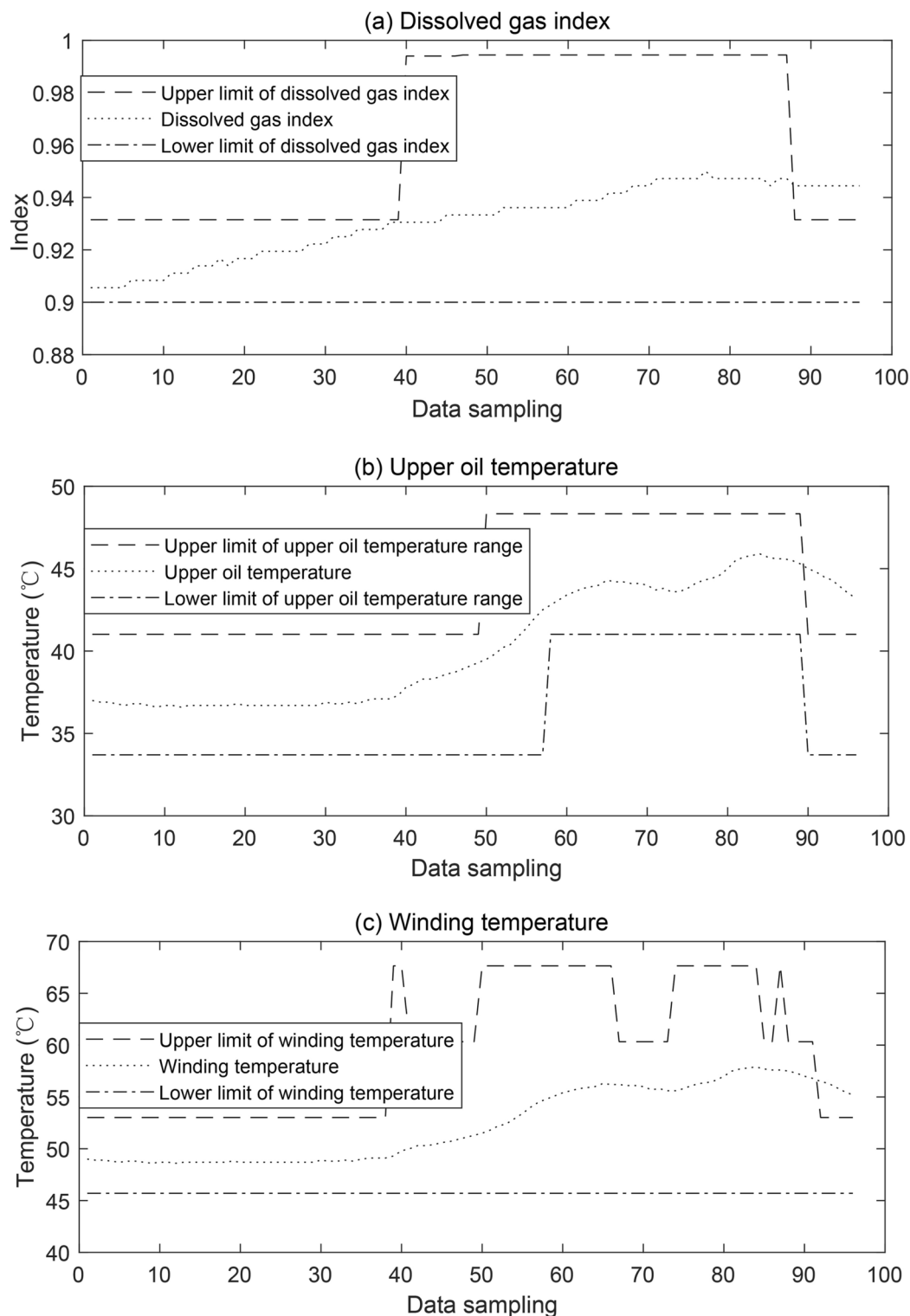


Fig. 8. Intelligent fault diagnosis of transformer based on multi-source data fusion and correlation analysis.

about half. Normal environment amorphous alloy transformer is in rated operating conditions, the temperature rise will not exceed the limit value, but due to the weather is too cold or hot, will cause the transformer to run in harsh environments, so that the oil-immersed self-cooling transformer's cooling capacity is weakened, the cooling effect is reduced. Therefore, when the amorphous alloy transformer operates in harsh environmental conditions, the temperature rise of the transformer's high and low-voltage windings should be very careful.

In order to verify the validity of the model under different operating conditions of the transformer, the amorphous alloy transformer temperature field under different environment temperatures is also studied,

| Environment temperatures (°C) | Dissolved gas index | Upper oil temperature (°C) | Winding temperature (°C) |
|-------------------------------|---------------------|----------------------------|--------------------------|
| −10 | [0.356, 0.386] | [29.856, 30.003] | [39.310, 39.635] |
| 0 | [0.508, 0.518] | [39.017, 39.255] | [39.268, 39.473] |
| 10 | [0.627, 0.634] | [48.135, 48.266] | [50.364, 50.471] |
| 20 | [0.756, 0.772] | [59.024, 59.318] | [61.087, 61.352] |
| 30 | [0.883, 0.904] | [71.267, 73.586] | [83.281, 85.636] |
| 40 | [0.991, 0.996] | [95.558, 98.215] | [101.0386, 103.659] |

Table 5. Corresponding interval for correlation analysis under the different environment temperatures.

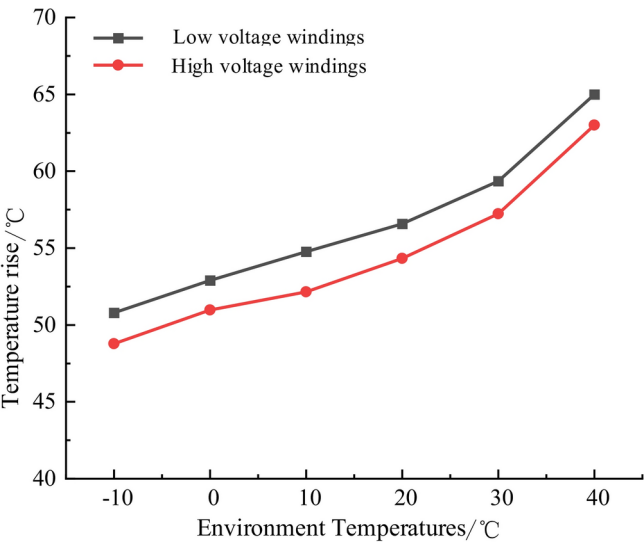


Fig. 9. The temperature rise of high and low voltage transformer winding under different environment temperatures.

respectively, to analyze the corresponding interval for correlation analysis under −10 °C, 0 °C, 10 °C, 20 °C, 30 °C, 40 °C environment temperatures at 50% load rate in Table 5.

Additionally, this section also analyzes the temperature field distribution of high and low-voltage windings under −10 °C, 0 °C, 10 °C, 20 °C, 30 °C, 40 °C environment temperatures. Figure 9 gives the temperature rise of high and low voltage transformer winding under different environment temperatures, as can be seen from the figure, the environment temperature within 35 °C, the high and low voltage winding temperature rise rises gently and does not exceed 65 °C temperature rise limit. When the environment temperature is more than 35 °C, the slope of the curve increases sharply, the high and low-voltage winding temperature rise increases markedly and has even exceeded the 65 °C limit value of the temperature rise, the normal operation of the transformer and winding insulation pose a great threat to the normal operation of the transformer, so the transformer is easy to be damaged when running in a high-temperature environment, especially when the environment temperature exceeds 35 °C.

Comparative analysis of fault diagnosis for multiple models

Five mainstream supervised learning models, namely, Linear Discriminant Analysis (LDA), K-Nearest Neighbour Algorithm (KNN), SVM, Random Forest (RF), and gradient boosting decision tree (GBDT), are selected to be trained under empirical parameters, and the test results are compared with the proposed model, and the results are shown in Table 6. As shown in the table, the diagnosis accuracy of the proposed model is the highest among the six models.

In order to demonstrate and illustrate the computational stability of the proposed method in this paper, 100 consecutive random samples of the sample set are performed, each time 20% of the samples are taken as the test set, and the remaining 80% of the samples are taken as the training set. The obtained training samples are used to train different models and the diagnosis results are counted and the results are shown in Table 7. As can be seen from the table, the average correct fault diagnosis rate of 100 diagnoses of the transformer fault diagnosis model proposed in this paper is 0.917, and the mean square error of the correct rate is 0.018. Compared with five mainstream supervised learning models, the transformer fault diagnosis model proposed in this paper has the highest correct rate, and the mean square error of the correct rate is smaller compared with the RF and GBDT models which have higher accuracy, which indicates that the method proposed in this paper is able to stably maintain the computational output with high accuracy.

| Fault type | LDA | KNN | SVM | RF | GBDT | The proposed method |
|------------------------------|-------|-------|-------|-------|-------|---------------------|
| Partial discharge | 0.692 | 0.818 | 0.753 | 0.955 | 0.923 | 0.936 |
| Arc discharge | 0.909 | 1 | 0.692 | 0.918 | 0.906 | 1 |
| Upper oil temperature 140 °C | 0.923 | 0.875 | 0.894 | 0.923 | 0.885 | 1 |
| Solid material | 0.768 | 0.471 | 0.692 | 1 | 1 | 1 |

Table 6. Accuracy rate of fault diagnosis for different models.

| Methods | Highest correct rate | Lowest correct rate | Mean correct rate | Mean variance of correct rate |
|---------------------|----------------------|---------------------|-------------------|-------------------------------|
| LDA | 0.709 | 0.518 | 0.611 | 0.038 |
| KNN | 0.845 | 0.706 | 0.741 | 0.044 |
| SVM | 0.802 | 0.718 | 0.724 | 0.035 |
| RF | 0.936 | 0.855 | 0.873 | 0.032 |
| GBDT | 0.945 | 0.818 | 0.887 | 0.026 |
| The proposed method | 0.982 | 0.891 | 0.917 | 0.018 |

Table 7. Diagnosis results of different models with repeated training.

Conclusions

In this paper, a novel intelligent fault diagnosis of transformers based on multi-source data fusion and data mining has been developed to figure out issues of faults and realize the prognostics and health management under multiple operation conditions in the power transformer. Above all, an improved entropy weighting method is employed to achieve the data fusion of various components for transformer dissolved gases. Then, the load rate, upper oil temperature, winding temperature data, and the fusion indices of dissolved gas components in the transformer are predicted by the combination of a bidirectional long short-term memory network, attention mechanism, and convolution neural network. In addition, for the purpose of the predictive assessment of the transformer state, Apriori correlation analysis based on the support and confidence levels, is performed on the transformer load rate, upper oil layer, winding temperature, and fusion indices of gas components. The specific conclusions to be drawn are the following:

- (1) Aiming at the problems of transformer operating conditions and loads, complicated parameters, and difficulty in effectively achieving the state predictive assessment, the proposed method is the method based on the data prediction and correlation analysis method to assess the health state of the transformer. Compared with past research, the proposed method can include a variety of characteristic parameters. Effective matching and correlation analysis of the characteristic parameters under different load rates of the predicted data is based on real-time assessment to improve early warning capability. The results show that in the vicinity of sample point 88, the dissolved gas, upper oil temperature, and winding temperature data are not within the normal range of intervals, and it is presumed that the arc discharge phenomenon.
- (2) By using the method of multi-source data fusion and data mining, the operating state of the transformer can be preliminarily judged by the data change of upper oil temperature, winding temperature data, and the fusion indices of dissolved gases components, which provides a simple and efficient intelligent online monitoring method for transformers that have been put into use, and also an effective method to identify a single phase fault. The experiment result shows that the transformer is easy to be damaged when running in a high-temperature environment, especially when the environment temperature exceeds 35 °C.
- (3) Compared with the method of setting thresholds, the method proposed in this paper can sense the operating situation of the equipment in advance and take corresponding measures to reduce the incidence of accidents and improve the reliability of the power supply. Compared with five learning models, i.e., LDA, KNN, SVM, RF, and GBDT, the transformer fault diagnosis model proposed has the highest correct rate, and the mean square error of the correct rate is smaller. The average correct fault diagnosis rate of 100 diagnoses of the transformer fault diagnosis model proposed in this paper is 0.917, and the mean square error of the correct rate is 0.018.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author (Jingping Cui) on reasonable request.

Received: 5 September 2024; Accepted: 24 February 2025

Published online: 04 March 2025

References

1. Deng, W. et al. Few-shot power transformers fault diagnosis based on gaussian prototype network. *Int. J. Electr. Power Energy Syst.* **160**, 110146–110159 (2024).

2. Raja, B., Venkatakrishnan, G. & Rengaraj, R. Power transformer fault diagnosis and condition monitoring using hybrid TDO-SNN technique. *Int. J. Hydrog. Energy* **68**, 1370–1381 (2024).
3. Gamel, S. A., Ghoneim, S. S. & Sultan, Y. A. Improving the accuracy of diagnostic predictions for power transformers by employing a hybrid approach combining smote and dnn. *Comput. Electr. Eng.* **117**, 109232–109245 (2024).
4. Fei, X. et al. Power system fault diagnosis with quantum computing and efficient gate decomposition. *Sci. Rep.* **14**, 16991–17012 (2024).
5. Sutikno, H. et al. Machine learning based multi-method interpretation to enhance dissolved gas analysis for power transformer fault diagnosis. *Heliyon* **10**, 25975–25989 (2024).
6. Xing, Z. & He, Y. A two-step image segmentation based on clone selection multi-object emperor penguin optimizer for fault diagnosis of power transformer. *Expert Syst. Appl.* **244**, 122940–122954 (2024).
7. Tao, W., Li, X., Liu, J. & Li, Z. Multi-scale attention network (MSAN) for track circuits fault diagnosis. *Sci. Rep.* **14**, 8886–8897 (2024).
8. Xiao, B. et al. Digital twin-driven prognostics and health management for industrial assets. *Sci. Rep.* **14**, 13443–13456 (2024).
9. Thelen, A. et al. Probabilistic machine learning for battery health diagnostics and prognostics-review and perspectives. *NPJ Mater. Sustain.* **2**, 14–25 (2024).
10. Zuñiga, J., Coria, G., Harms, Y., Valois, M. & Romero, A. A. Methodology for the optimal replacement of power transformers based on their health index. *Electr. Power Syst. Res.* **234**, 110582–110594 (2024).
11. Xiang, Y. & Zhuang, X. H. Application of Arima model in short-term prediction of international crude oil price. *Adv. Mater. Res.* **798**, 979–982 (2013).
12. Choi, T.-M., Yu, Y. & Au, K.-F. A hybrid SARIMA wavelet transform method for sales forecasting. *Decis. Support Syst.* **51**, 130–140 (2011).
13. Murat, A. & Tokat, E. Forecasting oil price movements with crack spread futures. *Energy Econ.* **31**, 85–90 (2009).
14. Hou, A. & Suardi, S. A nonparametric GARCH model of crude oil price return volatility. *Energy Econ.* **34**, 618–626 (2012).
15. Zhang, J.-L., Zhang, Y.-J. & Zhang, L. A novel hybrid method for crude oil price forecasting. *Energy Econ.* **49**, 649–659 (2015).
16. Mirmirani, S. & Cheng Li, H. A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. In *Applications of Artificial Intelligence in Finance and Economics* 203–223 (Emerald Group Publishing Limited, 2004).
17. Zhang, C. et al. Battery SOH estimation method based on gradual decreasing current, double correlation analysis and GRU. *Green Energy Intell. Transp.* **2**, 100108–100125 (2023).
18. Dhaked, D. K., Dadhich, S. & Birla, D. Power output forecasting of solar photovoltaic plant using LSTM. *Green Energy Intell. Transp.* **2**, 100113–100125 (2023).
19. Xiao, S. et al. Battery state of health prediction based on voltage intervals, BP neural network and genetic algorithm. *Int. J. Green Energy* **21**, 1743–1756 (2024).
20. Sun, C. et al. Convolutional neural network-based pattern recognition of partial discharge in high-speed electric-multiple-unit cable termination. *Sensors* **24**, 2660–2674 (2024).
21. Sheth, T. S. & Acharya, F. Optimization and evaluation of modified release solid dosage forms using artificial neural network. *Sci. Rep.* **14**, 16358–16369 (2024).
22. Jędrzejczyk, A., Firek, K., Rusek, J. & Alibrandi, U. Prediction of damage intensity to masonry residential buildings with convolutional neural network and support vector machine. *Sci. Rep.* **14**, 16256–16271 (2024).
23. Teğin, U., Dinç, N. U., Moser, C. & Psaltis, D. Reusability report: Predicting spatiotemporal nonlinear dynamics in multimode fibre optics with a recurrent neural network. *Nat. Mach. Intell.* **3**, 387–391 (2021).
24. Guo, Q., He, Z. & Wang, Z. Monthly climate prediction using deep convolutional neural network and long short-term memory. *Sci. Rep.* **14**, 17748–17752 (2024).
25. Karbasi, M. et al. Multi-step ahead forecasting of electrical conductivity in rivers by using a hybrid convolutional neural network-long short-term memory (CNN-LSTM) model enhanced by Boruta-XGBoost feature selection algorithm. *Sci. Rep.* **14**, 15051–15063 (2024).
26. Abduljabbar, R. L., Dia, H. & Tsai, P.-W. Development and evaluation of bidirectional LSTM freeway traffic forecasting models using simulation data. *Sci. Rep.* **11**, 23899–23910 (2021).
27. Kumar, M., Patel, A. K., Biswas, M. & Shitharth, S. Attention-based bidirectional-long short-term memory for abnormal human activity detection. *Sci. Rep.* **13**, 14442–14453 (2023).
28. El Said, B. Predicting the non-linear response of composite materials using deep recurrent convolutional neural networks. *Int. J. Solids Struct.* **276**, 112334–112349 (2023).
29. Sun, S., Wang, J., Xiao, Y., Peng, J. & Zhou, X. Few-shot RUL prediction for engines based on CNN-GRU model. *Sci. Rep.* **14**, 16041–16058 (2024).
30. Bjelić, M., Brković, B., Žarković, M. & Miljković, T. Machine learning for power transformer SFRA based fault detection. *Int. J. Electr. Power Energy Syst.* **156**, 109779–109791 (2024).
31. Zheng, W., Zhang, G., Zhao, C. & Zhu, Q. Multichannel consecutive data cross-extraction with 1dcnn-attention for diagnosis of power transformer. *Int. J. Electr. Power Energy Syst.* **158**, 109951–109966 (2024).
32. Hand, D. J. Data mining. In *Wiley statsref: statistics reference online* (2014).
33. Siwek, K. & Osowski, S. Data mining methods for prediction of air pollution. *Int. J. Appl. Math. Comput. Sci.* **26**, 467–478 (2016).
34. Papi, R., Argany, M., Moradipour, S. & Soleimani, M. Modeling the potential of sand and dust storm sources formation using time series of remote sensing data, fuzzy logic and artificial neural network (a case study of Euphrates basin). *Eng. J. Geosp. Inf. Technol.* **8**, 61–82 (2021).
35. Mohammady, M., Pourghasemi, H. R. & Amiri, M. Assessment of land subsidence susceptibility in semnan plain (Iran): A comparison of support vector machine and weights of evidence data mining algorithms. *Nat. Hazards* **99**, 951–971 (2019).
36. Guo, Z., Chi, D., Wu, J. & Zhang, W. A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriori algorithm. *Energy Convers. Manag.* **84**, 140–151 (2014).
37. Wu, X., Benjamin Zhan, F., Zhang, K. & Deng, Q. Application of a two-step cluster analysis and the Apriori algorithm to classify the deformation states of two typical colluvial landslides in the Three Gorges, china. *Environ. Earth Sci.* **75**, 1–16 (2016).
38. Hong, J., Tamakloe, R. & Park, D. Application of association rules mining algorithm for hazardous materials transportation crashes on expressway. *Accid. Anal. Prev.* **142**, 105497–105509 (2020).
39. Noussaiba, L. A. E. & Abdelaziz, F. Ann-based fault diagnosis of induction motor under stator inter-turn short-circuits and unbalanced supply voltage. *ISA Trans.* **145**, 373–386 (2024).
40. Wang, B., Qiu, W., Hu, X. & Wang, W. A rolling bearing fault diagnosis technique based on recurrence quantification analysis and Bayesian optimization SVM. *Appl. Soft Comput.* **156**, 111506–111521 (2024).
41. Su, X. et al. Application of DBN and GWO-SVM in analog circuit fault diagnosis. *Sci. Rep.* **11**, 7969–7981 (2021).
42. Berrisch, J., Narajewski, M. & Ziel, F. High-resolution peak demand estimation using generalized additive models and deep neural networks. *Energy AI* **13**, 100236–100242 (2023).

Acknowledgements

This work was supported in part by Shandong Province Key R&D Programme (2022TSGC2243), and Shandong Province Key R&D Programme (2023TSGC0960).

Author contributions

Jingping Cui : Data curation, Formal analysis, Investigation, Resources, Visualization, Writing—original draft, Writing—review & editing. Kuang Wei: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing. Kai Geng: Data curation, Formal analysis, Supervision, Writing—review & editing. Pihua Jiao: Data curation, Formal analysis, Writing—review & editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025