Journal of Computers Vol. 36 No. 3, June 2025, pp. 41-54
https://doi.org/10.63367/199115992025063603004

Transformer Anomaly Detection Based on Multi-Source Information Fusion of Insulation Oil

Bing-Yu Mo[1*], Shang-Hu Zhou[1], Meng-Long Han[1], Yan-Jiao He[2], Jian-Wu Li[2], and Ya-Ping Yang[1]

[1] State Grid Qinghai Electric Power Corporation Electric Power Research Institute, Qinghai 810000, China
Kaileyu1990@163.com, 378438626@qq.com, 18797326850@163.com, 2444521655@qq.com

[2] State Grid Qinghai Electric Power Corporation, Qinghai 810008, China
360151064@qq.com, ljwu0010@qq.com

Received 28 March 2024; Revised 26 April 2024; Accepted 26 April 2024

Abstract. With the rapid construction and development of smart grids, accurately identifying abnormalities in transformers is of great practical significance for the safe operation of power systems. However, traditional transformer anomaly detection methods make it difficult to effectively detect anomalies inside the transformer from a single source of transformer data. In this paper, we propose a transformer anomaly detection method based on multi-source information fusion of insulating oil, which uses an anomaly detection and diagnosis model with a deep Transform network structure. Specifically, our method contains four modules. The Transformer module is used to receive oil and gas decomposition data in the transformer and uses an attention-based sequence encoder to quickly perform inference knowledge to obtain broader transformer operation time trend data. The Gate Recurrent Unit (GRU) module accepts transformer temperature and bushing temperature data and uses a gating mechanism and a status update mechanism to capture features in the sequence. The Convolutional Neural Networks (CNN) module receives the oil chromatography information at the time of operation of the transformer and uses convolution operations to extract the spatial features in the oil chromatography data. Multilayer Perceptron (MLP) module takes the inputs of the above three modules as input and uses fully connected layers and *softmax* layers to accurately classify the data. We conducted extensive experiments on real data sets, and the experimental results show that the Accuracy of our proposed method reaches 98.87%, and the F1 reaches 99.40%, surpassing all baseline classifiers, and the F1 exceeds the state-of-the-art method by 3.77%. Furthermore, we explore the impact of different models and features on detection results.

Keywords: transformer, insulating oil, abnormality detection

1 Introduction

The power transformer is one of the critical components in the power grid, responsible for power transmission and transformation. Any failure in a power transformer can result in significant economic losses for the power grid [1-2]. To mitigate the adverse consequences of transformer faults, it is imperative to gain a comprehensive understanding of the internal insulation condition of the transformer. This necessitates the implementation of online monitoring and the ability to diagnose abnormalities in the transformer. Early detection of any irregularities within the transformer can significantly reduce the maintenance costs, which holds immense importance in prolonging the transformer's service life and enhancing economic benefits.

Nonetheless, traditional diagnostic methods [3-4] for identifying abnormal conditions in transformers fall short of accurately capturing the intricate characteristics of complex systems. In recent years, the extensive adoption of advanced sensor monitoring technology in the operation and maintenance of power transformers has led to an exponential increase in the volume of monitoring data. This influx of data serves as a crucial foundation for conducting comprehensive status assessments and predicting equipment performance.

For instance, Rong et al. [5] conducted an analysis of transformer fault types by monitoring the content of oil decomposition gases in the transformer. To address these limitations, Qin et al. [6] treat data as points in a high-dimensional space and learn Mahalanobis matrices based on the training set. Then use WLOF to assign the degree of outliers to each point. Metric learning can be regularized to reduce the correlation between different

* Corresponding Author

components of a vector. Even with a large number of duplicate data points, the WLOF method can accurately detect outliers. The experimental results on actual datasets show that even if the different components of the vector are not similar in scale and have strong correlation, this method can still detect outliers well. For instance, Fei et al. [7] incorporated transformer oil temperature data into a decision tree model to extract clusters representing the transformer's operating conditions. Subsequently, a dataset with anomaly labels was fed into a random decision forest model to construct an abnormal oil temperature classifier and generate anomaly detection rules. While methods based on machine learning and deep learning, including LSTM, address issues like gradient explosion, gradient vanishing, and inadequate long-term memory found in traditional Recurrent Neural Network (RNN), their predictive performance on time series data still falls short of meeting the demands of real-world distribution network environments. Nevertheless, current transformer anomaly detection techniques primarily rely on data sources such as transformer oil decomposition gas content, oil chromatography, and red and ultraviolet light. These methods often do not effectively integrate and harness the wealth of multi-source information available in the extensive electrical monitoring data collected through transformer online monitoring systems.

A promising solution to address this challenge is the fusion of multi-source data, including both normal and abnormal data generated during transformer operation. Subsequently, various deep learning technologies are applied to process this diverse data, extracting deep relationships between various monitoring results obtained from the multi-source information. Finally, the pre-processed multi-source information is fed into a multi-layer neural network. The representational capabilities of the multi-layer network are harnessed to train a transformer anomaly detection method based on the fusion of multi-source information.

In this paper, we present a transformer anomaly detection method founded on the fusion of multi-source information derived from insulating oil. Our approach employs an anomaly detection and diagnostic model featuring a deep Transformer network structure. Specifically, our method comprises four distinct modules: the Transformer module is designed to receive oil and gas decomposition data from the transformer. It leverages an attention-based sequence encoder to swiftly infer knowledge, thereby obtaining a more comprehensive understanding of the transformer's operational trends over time. The GRU module processes transformer temperature and bushing temperature data. It employs both a gating mechanism and a status update mechanism to effectively capture key features within the sequence. The CNN module is responsible for handling oil chromatography data collected during the transformer's operation. This module employs convolution operations to extract spatial features from the oil chromatography data. The MLP module serves as the final layer in the pipeline, taking inputs from the three preceding modules. It utilizes fully connected layers and *softmax* layers to precisely classify the data. Through this comprehensive approach, we aim to enhance transformer anomaly detection and contribute to the field of power system reliability. Specifically, the main contributions of this paper can be briefly summarized as follows:

- We propose a transformer anomaly detection framework that leverages the fusion of multi-source information from insulating oil. This method harnesses the power of deep learning techniques to efficiently identify anomalies in transformers.
- To make effective use of the various monitoring data generated by the transformer, we have developed four distinct modules to process and integrate this information. The aim is to enhance the accuracy of transformer anomaly detection.
- We comprehensively evaluate our proposed method on a real transformer dataset. Experimental results show that our proposed method outperforms baseline classifiers and current state-of-the-art detection methods. Additionally, we conduct ablation experiments to illustrate the detection capabilities of each module.

2 Related Work

Abnormal detection of transformers is one of the core goals of power equipment safety monitoring. Transformer anomaly detection is to collect and analyze the status characteristic information of the insulating oil in the transformer to obtain the operating status of the transformer and monitor its health trend to determine whether there are abnormalities in the operation of the transformer. Existing transformer abnormality detection mainly includes online monitoring of dissolved gases in oil, online monitoring of transformer bushings, partial discharge, online monitoring of bushing insulation, etc. For example, Wang et al. [8] proposed a robust model based data processing and analysis method for non random errors in oil chromatography data. Firstly, based on certain criteria, identify and eliminate large errors or abnormal data that may cause distortion in the measurement results. In order to avoid or suppress the influence of coarse errors, an improved least squares method was proposed to handle the coarse errors in experimental data, improving the fitting accuracy. The balance criterion method was also proposed in the data analysis process of the model to comprehensively analyze the data.

Aiming at the problems that current anomaly identification methods are greatly affected by data fluctuations and missing data and require long operation times. Singh et al. [9] conducted a fault analysis on the distribution transformer of the Indian state-owned power company Punjab State Power Company Limited to identify the cause of its high failure rate. A detailed fault analysis was conducted on 348 distribution transformers in a certain urban area from 2010 to 2015. The first part conducts failure analysis according to IEEE standards and summarizes the analysis results. The second part uses Failure Mode, Effects, and Criticality Analysis (FMECA) to analyze the fault modes, causes, and impacts of distribution transformers. Furthermore, based on severity classification, occurrence probability, and discovery probability, conduct criticality analysis, assign risk priority numbers, and identify the most critical components of the transformer. FMECA determines preventive measures to be taken to reduce the risk of future transformers by eliminating the causes of faults, reducing their severity and probability of occurrence. However, selecting incomplete types of gases in oil will lead to inaccurate classification results. The abnormal data can be quickly identified based on Hotelling-T2 and Squared Prediction Error statistics, and the working status of the transformer can be monitored online by combining the data characteristics.

Different from the above studies, Yang et al. [10] conducted insulation diagnosis on oil leakage sleeves based on the frequency domain dielectric response of 0.1Hz~1kHz, extracted characteristic values for defect judgment, and compared them with normal oil sleeves. The characteristic values include the average deviation of dielectric loss within the frequency range of 0.1Hz to 1kHz and the actual capacitance ratio. The experimental results and actual disassembly inspection results indicate that the gaskets and chips of the two leaking oil sleeves are all damaged, filled with discharge and wax, indicating irreversible insulation degradation. Compared with ordinary insulation sleeves, their dielectric loss increment is greater than 70%, and their capacitance change is greater than 5%, providing a reference for on-site testing and diagnosis of moisture defects in oil paper insulation sleeves.

Li et al. [11] conducted research and summary on abnormal heating faults in high-voltage terminal blocks, neutral point terminal blocks, grounding bars, post insulator bases, and DC filter reactor terminal blocks of reactors. Based on relevant testing data and standard regulations, the causes of heating in different parts mentioned above were analyzed, and targeted treatment plans were provided. The research results indicate that due to the influence of the leakage magnetic field and equipment installation process of the dry hollow reactor, abnormal heating is generated in the metal components of the column base, neutral point, grounding bar, and circulating current inside the down lead. Established corresponding fault types, providing a basis for the types of abnormal heating faults in dry air reactors. Yu et al. [12] proposed a power transformer anomaly diagnosis method based on the Attention CNN LSTM hybrid model. Collect the sound signals emitted by actual power transformers in normal, overload, and discharge states, preprocess the sound signals, obtain the MFCC features of the sound signals, and then group the sound feature vectors through first-order and second-order differences, and enter the Attention CNN LSTM hybrid model for training.

With the advancement of machine learning, researchers are increasingly inclined to utilize machine learning algorithms for anomaly detection. This method derives the abnormal temperature threshold for each cluster through hypothesis testing and is used to mark anomalies in the working condition data set. Finally, the data set with anomaly labels is input into the random decision forest model to build an abnormal oil temperature classifier and generate anomaly detection rules.

However, detection methods based on thresholds make it difficult to accurately identify transformer fault types. Lin et al. [13] proposed a data cleaning strategy based on association rule analysis and neural networks. Firstly, establish a mathematical model using association rules to measure the degree of correlation between state monitoring variables. Then, the k-means clustering algorithm is used to detect missing and abnormal values in the sequence, and a cleaning process and rule considering sequence correlation are proposed to distinguish between sensor data anomalies and device status anomalies. Using an improved wavelet neural network model to predict missing data and correct error data for washable data points. Improved the cleaning efficiency and accuracy of the network. In response to the noise data in oil and gas data, Before applying density based spatial clustering (DBSCAN) with noisy applications to detect clusters, Bo et al. [14] used an ensemble method based on MLE to filter or label bad data, which changes over time within each time window. In the third step, develop a decision logic diagram using additional data and system information to classify observed anomalies into different types of faults, faults, events, or degradation. The method was validated using an IEEE testing system established through the interface between real-time digital simulators (RTDS) and industrial hardware PMU devices. Early detection of equipment failures will improve system reliability and reduce maintenance costs. Victoria et al. [15] described a technique for anomaly detection under the conditions experienced by working transformers, such as load, seasonal weather, and network configuration. The purpose is to model the aging but normal behavior of a given transformer while reducing the likelihood of being incorrectly detected as abnormal. This article introduces how this technology is applied to two transmission transformers in the UK, and provides a case study of 12 months of data, detailing the abnormal situations detected during this period.

With the excellent performance of deep learning in the field of detection, many researchers use deep learning-based methods to detect transformer anomalies. For example, Mitiche et al. [16] proposed a deep learning-based method for real-time detection of anomalies in current amplitude and phase angle at three casing taps. The method consists of a Long Short-Term Memory Autoencoder (LSTMAE) network that learns the normal current and phase measurements of the bushing and detects any point where these measurements change based on a mean absolute error (MAE) metric evaluation. Dai et al. [17] proposed a transformer fault diagnosis method based on deep belief networks. By analyzing the relationship between dissolved gases in transformer oil and fault types, the non coding ratio of gases was determined as a characterization parameter for the DBN model. DBN uses multi-layer and multi-dimensional mapping to extract more detailed differences in fault types. During this process, pre train the diagnostic parameters. The backpropagation algorithm adjusts samples based on their labels and optimizes parameters. To verify the effectiveness of the proposed method, a diagnostic DBN model was constructed and tested using various oil color spectrum datasets collected from State Grid Corporation of China and previous publications.

In substations, when the winding of a transformer overheats, either a short-term solution to reduce the load or a long-term plan to install another transformer box is necessary. In order to decide whether to deploy either of these two strategies, it should be possible to accurately predict transformer temperature. He et al. [18] explored the possibility of using artificial neural networks to predict the top oil temperature of transformers. Explored methods for predicting transformer oil temperature using static neural networks, time processing networks, and recurrent networks. Abduldeleev et al. [19] introduced a continuous control method and system for vacuum circuit breaker technology based on vacuum chamber temperature measurement and wireless channel data transmission. Detailed introduction of the system's functional block diagram and hardware structure. The hardware setup based on contact temperature sensors has been demonstrated. The system is deployed on the voltage circuit breaker of the 110 kV closed switchgear in the arc furnace workshop of a certain metal factory. The author reviewed the system interface and provided a temperature trend of the modules included in the circuit breaker pole. Liu et al. [20] proposed the idea of applying Time Series Generative Adversarial Networks (Time GANs) in scheduling automation systems to identify, diagnose, and predict anomalies in massive data streams. Generate a large number of reliable datasets using Time GAN to train fault diagnosis models. Using unsupervised learning methods such as density based noisy application spatial clustering (DBSCAN) and K-means to detect potential unknown anomalies in the power grid. Finally, choose some supervised learning methods to compare with unsupervised learning methods. The experimental results show that the algorithm has a higher recognition rate for known anomalies than other benchmark algorithms, and can discover new unknown anomalies.

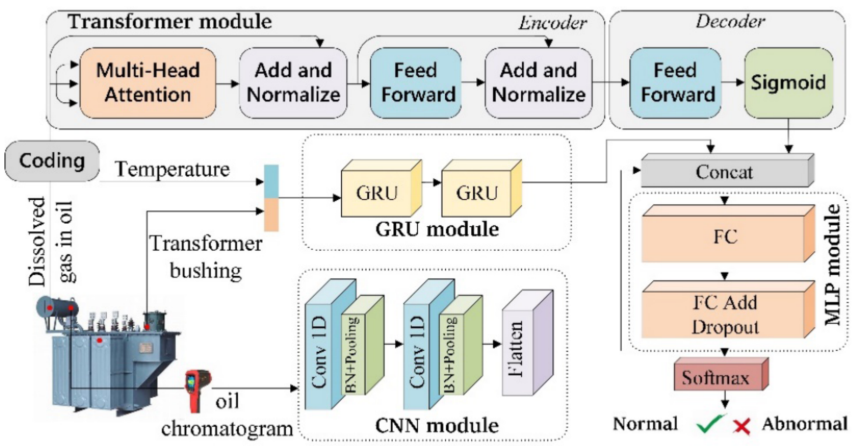

Fig. 1. The overall framework of transformer anomaly detection based on multi-source information fusion of insulating oil

3 Proposed Detection Architecture

In this section, we first introduce the overall architecture that we propose based on insulating oil transformer anomaly detection. Secondly, we introduced the principle of gas production in the transformer, the corresponding relationship between dissolved gas and internal faults of the transformer, and we introduced the coding method. Finally, we introduce the transformer anomaly detection process based on deep learning.

3.1 Overall Architecture

The transformer anomaly detection framework based on multi-source information fusion of insulating oil is shown in the Fig. 1, which mainly includes encoding, Transformer Module, GRU Module, CNN Module and MLP Module. In general, we first encoded the gas generated during the operation of the insulation transformer to prepare for subsequent model training. Second, to make full use of the data generated by the transformer, we input different feature information into different deep-learning models. Finally, the output results of the Transformer, GRU, and CNN modules are concatenated together, and then input into the MLP module, and the classification results are output through the *softmax* function. More details on each module are provided below.

1) **Transformer fault coding method.** This method primarily encodes the gas content generated by the insulating oil during the operation of the transformer, with the main technique employed being the three-ratio method. By utilizing this encoding method, it becomes possible to effectively differentiate the gas content produced during transformer failures from the gas content generated during normal operation. This approach holds promise for enhancing the accuracy and efficiency of transformer anomaly detection.
2) **Transformer Module.** In this module, the primary input consists of the encoded oil-decomposed gas data. The module employs the self-attention mechanism to extract interdependencies between various gases, providing a means to delineate the relationships used for distinguishing the oil-decomposed gas under different states of the transformer. This self-attention mechanism aids in capturing critical information to differentiate the transformer's operating conditions effectively.
3) **GRU Module.** In this module, the primary inputs are the temperature data from the transformer and the online bushing monitoring temperature. It leverages the gating mechanism and the state transition characteristics of the GRU to effectively capture the long-term dependencies within the sequence. This enables the model to understand and represent the relationships and patterns in the temperature data over an extended period, facilitating more accurate analysis and prediction.
4) **CNN Module.** In this module, the primary input is the oil chromatogram data obtained during the operation of the transformer. It utilizes the convolution and pooling operations of CNN to extract characteristic information from the oil chromatogram, which is then used to differentiate between the faulty and normal conditions of the transformer. This approach leverages the power of CNN to automatically learn and capture relevant features within the chromatogram data, contributing to more accurate fault detection in the transformer.
5) **MLP Module.** In this module, the primary inputs are the outputs of the Transformer Module, GRU Module, and CNN Module. It employs a MLP to integrate the multi-source information and combines the function to accurately classify the data. This approach allows for the effective fusion of information from different sources, enhancing the overall classification capabilities of the system.

3.2 Principle of Gas Production in Insulating Oil Transformer

Transformer insulating oil is a specialized mineral oil derived from the processing and refining of petroleum. It is extensively utilized in transformers due to its excellent insulating properties. The primary constituents of transformer insulating oil encompass the following components: **a)** Alkanes, characterized by the molecular formula C_nH_{2n+2}. **b)** Cyclic hydrocarbon saturated hydrocarbons, identified by the molecular formula C_nH_{2n}. **c)** Aromatic unsaturated hydrocarbons, denoted by the molecular formula C_nH_{2n-2}. **d)** Cellulose, represented by the molecular formula $(C_nH_{10}O_2)n$. It is evident that the principal constituents of insulating oil encompass various chemical groups connected by $C-C$ bonds.

The underlying principle of partial discharge in transformer faults typically involves ion reactions that result in the release of a certain level of energy. This energy leads to the cracking and rupture of $C-H$ bonds, accumulating hydrogen gas. However, the occurrence of $C-C$ single bond rupture requires the release of a higher level of energy, leading to the formation of hydrocarbon gases in various bond types within a relatively shorter timespan. It's worth noting that the energy requirement increases with the number of bonds involved.

Hence, it can be deduced that equipment typically remains in a continuous state of heating primarily because transformer equipment emits corresponding energy during its operation. This heating condition induces alterations in the properties of various insulation materials. During the gradual process of property change, it gradually accumulates low-molecular-weight hydrocarbons and gases. Following the occurrence of an internal fault in the transformer, a significant amount of electrothermal energy is typically generated. This, in turn, leads to the cracking and rupture of $C{-}C$ and $C{-}H$ bonds, simultaneously generating a certain amount of hydrogen molecules. Alongside the formation of unstable carbon-hydrogen compounds and free radicals, these substances may undergo chemical recombination, resulting in the assembly of new hydrogen gas and low-molecular-weight hydrocarbon gases. In cases of relatively mild faults, transformer oil is capable of dissolving these gases. However, in instances of higher fault severity, the increased energy levels produced shorten the periodicity of these gases, causing them to rapidly expand in scale and, subsequently, generate acetylene gas. During the fault occurrence process, when the temperature of solid insulation materials becomes excessively high, a significant amount of CO and CO_2 is produced. Consequently, it is evident that detecting and analyzing the gases produced in transformer oil can help determine the type of fault occurring in the transformer. The gases produced by various faults of insulating oil transformers are shown in Table 1.

Table 1. Gases from oil-immersed variable pressure faults

Fault type	Main gas components	Main gas composition
The oil temperature in the transformer is high	CH_4, C_2H_2	H_2, C_2H_6
The oil and paper in the transformer have high temperatures	CH_4, C_2H_4, CO, CO_2	H_2, C_2H_6
Partial discharge occurs in oil-paper insulation	H_2, CH_4, CO	H_2, C_2H_6, CO_2
Spark discharge occurs in transformer oil	H_2, C_2H_2	-
Arcing in transformer oil	H_2, C_2H_2	CH_4, C_2H_4, C_2H_6
Arcing in oil and paper	H_2, C_2H_2, CO, CO_2	CH_4, C_2H_4, C_2H_6

3.3 Correspondence between Dissolved Gases in Oil and Internal Faults

During the process of detecting oil substances in transformer equipment, it is of paramount importance to identify the characteristic gases that play a fundamental role in determining faults in the detection results. These crucial characteristic gases encompass carbon dioxide, carbon monoxide, ethylene, ethane, and other substances. When evaluating the transformer's condition, it is imperative to actively measure the real-time characteristics of these gases and employ them as a basis for ascertaining the transformer's fault status.

Table 2. Standards for gas content parameters in transformer oil

Type	Gas type	Gas content	
Transformers and Reactors	Total hydrocarbons	≥ 330KV	≤ 220KV
	Acetylene	150	150
	H	1	5
	CO	150	150
	CO_2	-	-

The release of energy following ion reactions causes the cracking and rupture of $C{-}H$ bonds, lead to the accumulation of hydrogen gas. Therefore, generating a significant quantity of hydrocarbon gas necessitates a more extensive cleavage of $C{-}H$ bonds, which demands a higher energy level to attain. Lower energy levels can only yield lower molecular weight alkanes. In an environment with a temperature of 50°C, molecules will produce C_2H_4 containing a double bond, whereas in an environment with temperatures ranging from 800°C to 1200°C, the molecules will generate C_2H_2, which contains a triple bond. As the ambient temperature gradually decreases, the double-bonded C_2H_4 and triple-bonded C_2H_2 produced by molecules will be swiftly suppressed and accumulated into new compounds, resulting in a larger quantity of acetylene. Smaller amounts of acetylene will continue to be generated in environments with temperatures below 800°C.

During the actual detection process of the transformer's condition, if the gas content parameter in the transformer's internal oil exceeds the specified standard, it becomes imperative to initiate key maintenance procedures. The specific standard parameters for gas content in the transformer's internal oil are outlined in Table 2. The types and examples of insulating oil transformer faults are shown in Table 3.

Table 3. Types and examples of oil-immersed transformer faults

Fault type	Fault example
Partial discharge	Partial discharges in gas-filled cavities caused by incomplete immersion, high moisture content of the paper, supersaturation of oil, or cavities.
Low energy	a. Spark discharges or arcs can be induced by varying potentials or floating potentials resulting from connection failures. These discharges may transpire in shielding rings, adjacent wire coils, or conductors within windings, as well as in open welds or closed loops of iron cores. b. The high voltage between the clamps, the bushing and the box wall, the coil and the discharge at the ground end. c. The edge connections of insulators of various types of materials and the along-surface discharge of winding pads. Oil breakdown, cut-off current of the selector switch.
High energy	a. Local high energy or flashover caused by short circuit, creeping discharge or arcing. b. Short circuit between low voltage and ground, between joints, between coils, between bushings and boxes, between copper bars and boxes, and between windings and iron cores. c. A discharge between two adjacent conductors surrounding the main magnetic flux. Discharge between the insulating screws of the iron core and the metal rings that fix the iron core.
t<300°C	a. Transformer operating beyond rated power. b. The oil flow in the winding is blocked. c. Stray magnetic flux in the choke clamp.
300°C<t<700°C	a. Poor contact at bolted joints, sliding contact surfaces, contact surfaces in selected open coffins, and connections between bushing leads and cables. b. Circulation currents between the clamps and bolts at the electrical choke, between the clamps and the core laminations, circulation currents in the ground wire, and poor solder joints in the magnetic shielding and circulation currents in the clamps. c. Wear of the insulation between adjacent parallel conductors in a winding.
t> 700°C	a. Large circulation on the tank and core b. The magnetic field on the tank wall is too high to compensate, forming a certain current c. Short circuit between core laminations

3.4 Transformer Fault Coding Method

The three-ratio method is highly effective in characterizing faults in transformers and reactors. This method is based on analyzing the properties of gases produced when the insulating material in transformer equipment fails. Its objective is to establish the relationship between various parameters and environmental factors to determine potential fault types in the transformer equipment. Table 4 presents the encoding parameters for three distinct values of five specific gases, combined with two types of similar gases such as solubility and diffusion coefficient. By evaluating the calculated ratio results and referring to the corresponding codes in Table 4, effective matching with the fault types in Table 5 can be achieved. This method is capable of eliminating the influence of oil in the transformer detection process and effectively detecting fault types in oil-immersed transformer equipment.

Table 4. Coding parameters

Gas ratio range	Coding		
	C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6
$\rho < 0.1$	0	1	0
$0.1 \leq \rho < 1$	1	0	0
$1 \leq \rho < 3$	1	2	1
$\rho \geq 3$	2	2	2

Table 5. Correspondence between transformer fault and coding

Coding combination			Fault type
C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6	
0	0	1	Low temperature overheating T<150°C
-	2	0	Low temperature overheating 150°C<T<300°C
-	2	1	Medium temperature overheating300°C<T<700°C
-	0,1,2	2	High temperature and overheating T>700°C
-	1	0	Partial discharge
2	0,1	0,1,2	low energy discharge
-	2	0,1,2	Low energy discharge and overheating
1	0,1	0,1,2	Arc discharge
-	2	0,1,2	Arc discharge and overheating

3.5 Proposed Detection Method based on Multi-source Information Fusion of Insulating Oil

Transformers, CNNs, attention mechanisms, and GRUs are popular deep learning models extensively applied in various natural language and visual processing tasks. The self-attention mechanism within the Transformer model enables dynamic focusing on different parts of the input sequence, enhancing the capture of information in different transformer oil degassing processes. Additionally, the self-attention mechanism in the Transformer model effectively captures long sequence features' dependencies, offering a better characterization of the relationship between various oil and gas elements. GRU contributes to gradient propagation efficiency in long sequences by allowing the model to skip less critical information through update gates. CNN, with its local feature extraction capabilities and weight parameter sharing, excels in extracting spatial structure information from transformer oil chromatographic features while reducing model training parameters. Our proposed detection method incorporates four distinct deep neural network structures. In this section, we will delve into the Transformer-based anomaly detection model that leverages the fusion of multi-source information from insulating oil.

1) **Transformer Module:** This module contains two parts: an encoder and a decoder. In the encoder, a sequence containing oil hydrolysis gas is taken as input, $X = \{x_1, x_2, \ldots, x_n\}$, where x_i represents the content of the i-th gas and n represents the features type. We define scaled-dot product attention of three matrices Q (query), K(key) and V (value):

$$Attention(Q, K, V) = soft\max\left(\frac{QK^T}{\sqrt{n}}\right)V \tag{1}$$

where, the *softmax* function allows the input matrix to be compressed to simplify reasoning in downstream neural network operations. V Unlike traditional attention operations, scaled dot product attention scales the weights through the square root term to reduce the variance of the weights, thereby promoting stable training. For the input matrices Q, K, and V, we exploit the multi-head self-attention mechanism by making it utilize h feed-forward layers to obtain Q_i, K_i and V_i, $i \in \{1, \ldots, h\}$. Then applying scaled-dot product attention as:

$$MultiHeadAtt(Q, K, V) = Concats(H_1, \ldots, H_h) \tag{2}$$

where, H_i = *Attention* (Q_i, K_i, V_i). Multi-head attention allows the model to jointly attend to information from different representation subspaces of different hydrolysis gases. Therefore, the following operations are performed in the encoder:

$$I_1^1 = LayerNorm(I_1 + MultiHeadAtt(I_1, I_1, I_1)) \tag{3}$$

$$I_1^2 = LayerNorm(I_1^1 + FeedForward(I_1^1)) \tag{4}$$

2) **GRU Module:** This module consists of two layers of GRU. The GRU model includes two gating mechanisms, namely the Reset Gate and the Update Gate, as well as an updated hidden state. We take the temperature of the transformer and the online monitoring temperature of the transformer bushing as inputs to this module. Assuming that *t* represents the time step, the calculation formula of the reset gate is as follows:

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \quad \textbf{(5)}$$

where, r_t represents the output of the reset gate, W_r represents the weight matrix, h_{t-1} represents the hidden layer state of the previous time period, x_t represents the input of the current state, [·] represents the connection operation, σ represents tanh function. The update gate is calculated as follows:

$$z_t = \tanh\left(W_Z \cdot \left[h_{t-1}, x_t\right]\right) \quad \textbf{(6)}$$

where, z_t represents the output of the update gate. Next, calculate the candidate hidden state $\widetilde{h_t}$ and the new hidden layer state B for the current time step h_t:

$$\widetilde{h_t} = \tanh\left(W \cdot \left[r_t \odot h_{t-1}, x_t\right]\right) \quad \textbf{(7)}$$

where, *W* is the weight matrix connecting the parts of the reset gate and the input, and $\odot$ represents element-wise multiplication. $\widetilde{h_t}$ is a candidate new hidden state that takes into account past hidden states and the current input. The calculation for updating the hidden state is as follows:

$$h_t = \left(1 - z_t\right) \odot h_{t-1} + z_t \odot \widetilde{h_t} \quad \textbf{(8)}$$

Therefore, the following operations were performed in the GRU Module:

$$g_1 = GRU_1\left(\tilde{x}\right) \quad \textbf{(9)}$$

$$g_2 = GRU_2\left(g_1\right) \quad \textbf{(10)}$$

where, GRU_i , $i \in$\{1, 2\}$ represents the *i*-th GRU layer. g_2 represents the final output of GRU Module. In this way, GRU can effectively capture long-term dependencies in input sequences through the gating mechanism and status update mechanism, and better learn relevant features when processing transformer temperature and bushing temperature data sequences.

3) **CNN Module:** When it comes to the CNN module, it takes the oil chromatogram x of the operating transformer as input to the module. The following operations are performed in this module:

$$h_1 = ConvBlock_1(x) \quad \textbf{(11)}$$

$$h_2 = ConvBlock_2(h_1) \quad \textbf{(12)}$$

$$\mu = Flatten(h_2) \quad \textbf{(13)}$$

where, $ConvBlock_i$ represents the *i*-th convolution block in the CNN Module, h_1 and h_2 represent the hidden vectors, and μ represents the output of the CNN Module.

4) **MLP Module:** After passing through three modules, the output of the Transformer Module, GRU Module and CNN Module is sent to the MLP Module, which can be described as:

$$c = Concate\left(I_1^2, g_2, \mu\right) \quad \textbf{(14)}$$

$$h_1' = FC_1\left(c\right) \quad \textbf{(15)}$$

$$h_2' = FC_2\left(h_1'\right) \quad \textbf{(16)}$$

$$\mathcal{O} = Dropout\left(h_2'\right) \quad \textbf{(17)}$$

where, *Concate* represents the connection operation, *c* represents the result of the connection, and FC_1 and FC_2 represent two fully connected layers respectively. *Dropout* represents the dropout layer to avoid overfitting of the classification model during the training process.

Finally, the classification results are obtained through the *softmax* layer. Since the gradient of the cross-entropy loss function to the last layer weight is no longer related to the derivative of the activation function, but is only proportional to the difference between the output value and the true value, the convergence speed is faster. In addition, the backpropagation of the cross-entropy loss function is multiplicative, thus speeding up the update of the weight matrix and reducing model training overhead. Therefore, during model training, we use cross-entropy as the loss function, defined as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=0}^{N-1} y_i \cdot \log\left(p_i\right) + \left(1 - y_i\right) \cdot \log\left(1 - p_i\right) \quad \textbf{(18)}$$

where, N represents the batch size, y_i represents the label of the i-th real sample, and p_i represents the model prediction label.

4 Proposed Detection Architecture

We evaluate the performance of our proposed method through extensive experiments. We compared baseline methods including support vector machine (SVM), random forest (RF), K-nearest neighbor (KNN), and naive Bayes (NB). Secondly, our experimental results are compared with state-of-the-art works, including [18], [19] and [20]. Finally, we conduct ablation experiments on the dataset to evaluate the overall performance of our detection model.

4.1 Dataset

We conducted an evaluation of our proposed method using a real dataset comprising transformer anomalies. The dataset was sourced from monitoring data collected before and after the shutdown of a 750kV reactor in 2022, which occurred due to excessive gas levels triggering a heavy gas trip at a specific time during operation. This dataset includes information on insulating oil gas content, temperature, casing temperature, and oil chromatogram data. The gas content in the insulating oil and descriptions of the current transformer are outlined in the table. The transformer status data is categorized into six groups: Normal, Low energy discharge, High energy discharge, Partial discharge, Medium and low-temperature superheating, and High-temperature superheating, each representing different abnormal conditions. These classifications were determined primarily through manual observation of changes in acetylene, hydrogen, total hydrocarbon contents, and the three-ratio chromatographic data analysis method. The dataset comprises a total of 823,769 records, including 719,103 normal data points and 104,666 abnormal data points.

4.2 Experimental Environment and Evaluation Metrics

1) Experimental Environment: Our detection model is implemented using TensorFlow2.0 and Keras. The ex-

perimental platform is Windows 10, and the model is trained and tested on the CPU Intel i9-13900k and GPU NVIDIA GeForce RTX 4090 (24G).

2) Metrics: We use the commonly used evaluation indicators including *Accuracy*, *Precise*, *Recall*, and *F*1 (weighted average of precision and recall rate) to evaluate the model's performance. where, *F*1 is used to evaluate the comprehensive performance of the model, and the calculation formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$Precise = \frac{TP}{TP + FP} \tag{20}$$

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

$$F1 = \frac{2 \times Precise \times Recall}{Precise + Recall} \tag{22}$$

where True Positive (TP) is the number of abnormal correctly classified as abnormal; False positives (FP) are the number of abnormal misclassified as normal; True Negative (TN) is the number of normal software correctly classified as normal; False Negative (FN) is the amount of abnormal misclassified as normal.

4.3 Evaluation Result

1) Compare with baseline classifier.

In this experiment, we evaluate the performance of our proposed detection method. We first compare with baseline classifiers, and the experimental results are shown in Table 6. In order to choose the optimal hyperparameters for each baseline classifier, we conduct experiments on different hyperparameters. Specifically, when the hyperparameters C = 1 of SVM, Kernel = 'poly', gamma = 0.1, the number of decision trees of RF is set to 100, NB alpha = 1, and neighbors = 5 in KNN, the base classifier achieves the highest detection accuracy. Finally, we compare the experimental results on the test set with our proposed method.

Table 6. Compare with baseline classifier

Methods	Accuracy (%)	Precise (%)	Recall (%)	F1
SVM	84.14	80.13	72.62	0.7619
RF	85.72	76.22	72.92	0.7453
KNN	82.21	84.21	66.67	0.7442
NB	80.19	79.18	80.19	0.7968
Our Method	**98.87**	**99.86**	**98.95**	**0.9940**

It can be seen from the experimental results that our proposed detection method achieves an F1 of 0.9940 on the real data set. Our model achieves higher classification accuracy compared to baseline classifiers. The main reason is that different deep learning models extract different high-level representations of features from multi-source features, reflecting the real situation of the transformer from multiple angles.

2) Compared with the state-of-the-art methods.

We further demonstrate the detection performance of our proposed method. We compare three state-of-the-art detection methods, [18], [19] and [20]. Specifically, we reproduce the detection models of [18], [19] and [20]. The experimental results are shown in Table 7. Experimental results show that our proposed detection method is more competitive compared with three state-of-the-art methods. The F1 are 10.16% higher than the [18] method, 7.69% higher than the [19] method and 4.38% higher than the [20] method. On the other hand, it shows that it is effective for us to extract multi-source information of the transformer and perform information fusion.

Table 7. Compare with the state-of-the-art methods

Methods	Accuracy (%)	Precise (%)	Recall (%)	F1
[18]	94.31	95.22	83.96	0.8924
[19]	93.06	92.88	90.56	0.9171
[20]	97.62	96.84	93.26	0.9502
Our method	**98.87**	**99.86**	**98.95**	**0.9940**

3) Ablation experiment.

To explore the influence of different combinations of deep learning models on the detection results, we conducted a classifier ablation experiment. Specifically, we input features into Transformer, GRU, CNN, and MLP, respectively, as well as different combinations of classification models. The experimental results are shown in the Table 8. The experimental results show that the detection performance of a single classifier is weak, F1 is 0.7122, 0.6652, 0.8013, and 0.7252, respectively. Compared with a single classifier, the combined model achieves a higher detection rate, which shows the effectiveness of our proposed method.

Table 8. Method ablation experiment

Methods	Accuracy (%)	Precise (%)	Recall (%)	F1
Transformer	70.85	72.50	69.98	0.7122
GRU	65.22	69.82	63.51	0.6652
CNN	80.47	82.77	77.66	0.8013
MLP	69.50	74.23	70.89	0.7252
Transformer+GRU	92.66	93.64	90.01	0.9179
Transformer+CNN	91.54	89.88	87.39	0.8862
Transformer+MLP	88.69	89.30	86.22	0.8773
GRU+CNN	93.80	94.56	92.08	0.9330
GRU+MLP	90.58	91.55	88.91	0.9021
CNN+MLP	89.64	90.12	86.45	0.8825
Transformer+GRU+CNN	96.21	95.63	93.63	0.9462
Transformer+GRU+MLP	95.87	96.81	94.80	0.9579
GRU+CNN+MLP	94.69	93.44	90.97	0.9219
Our method	**98.87**	**99.86**	**98.95**	**0.9940**

In addition, we also conducted ablation experiments on the features. We explore the effects of different features on classification results. f_1 stands for Dissolved gas in oil, f_2 stands for Temperature, f_3 stands for Transformer bushing temperature, and f_4 stands for oil chromatogram. The experimental results are shown in the Table 9. The experimental results show that the classification ability of a single feature is relatively weak. It is worth noting that oil and gas characteristics play a crucial role in detecting transformer anomalies. On the other hand, it also shows that the full use of the features information generated by the transformer is very important for the accurate detection of transformer anomalies.

Table 9. Feature ablation experiment

Features	Accuracy (%)	Precise (%)	Recall (%)	F1
f_1	67.96	70.05	66.85	0.6841
f_2	57.44	58.95	54.77	0.5678
f_3	52.63	57.64	55.82	0.5672
f_4	62.88	64.93	61.97	0.6342
f_1+f_2	82.56	84.76	83.66	0.8421
f_1+f_3	80.09	82.53	81.59	0.8206
f_1+f_4	79.04	77.68	75.86	0.7676
f_2+f_3	77.63	78.92	76.21	0.7754
f_2+f_4	75.07	76.83	74.11	0.7545
f_3+f_4	78.11	79.38	77.68	0.7852
$f_1+f_2+f_3$	95.67	96.25	94.89	0.9557
$f_1+f_2+f_4$	94.88	95.42	93.04	0.9421

$f_2+f_3+f_4$	92.70	93.09	92.88	0.9298
All features	**98.87**	**99.86**	**98.95**	**0.9940**

5 Conclusions

In this paper, we introduce a transformer-based anomaly detection method that leverages the fusion of multi-source information derived from insulating oil. Our methodology begins by gathering various features from the transformer during its operation, including data on gases generated by the insulating oil, transformer temperature, bushing monitoring temperature, and oil chromatography. To ensure the precise detection of anomalies in the transformer, we employ encoding techniques for the different gases generated. Subsequently, we feed these distinct features into different deep learning models, encompassing Transformer, GRU, CNN, and MLP. We then combine the model outputs and derive classification results after passing them through the MLP and *softmax* layers. Finally, we subject our proposed method to extensive experiments using real-world datasets. When compared to the baseline classifier, our method exhibits a 19.72% higher F1 score. Furthermore, when compared to state-of-the-art methods, our approach achieves a 3.77% higher F1 score. Additionally, we conduct ablation experiments to investigate the influence of different models and features on detection accuracy.

While our proposed method demonstrates competitiveness, it is essential to acknowledge its limitations and shortcomings. The presence of noise features within the dataset can impact the accuracy of the detection model, and irrelevant features may lead to excessive consumption of computing resources. In our future work, we plan to introduce denoising methods aimed at reducing the model's training time and enhancing its detection accuracy. Regarding the detection models, we intend to implement multi-time series information fusion to bolster the model's robustness. Additionally, we will assess the effectiveness of our proposed method across a broader range of datasets.

References

[1] S. Yu, W.-M. Tan, C.-M. Zhang, C. Tang, L.-H. Cai, D. Hu, Power transformers fault diagnosis based on a meta-learning approach to kernel extreme learning machine with opposition-based learning sparrow search algorithm, Journal of Intelligent & Fuzzy Systems 44(1)(2023)455-466. http://doi.org/10.3233/JIFS-211862.

[2] J.-B. Thomas, S.-G. Chaudhari, K.-V. Shihabudheen, N.-K. Verma, CNN-Based Transformer Model for Fault Detection in Power System Networks, IEEE Transaction on Instrumentation and Measurement 72(2023) 1-10. http://doi.org/10.1109/TIM.2023.3238059.

[3] A.J. Christina, M.A. Salam, Q.M. Rahman, F.-S. Wen, S.-P. Ang, W. Voon, Causes of Transformer Failures and Diagnostic Methods—A Review, Renewable and Sustainable Energy Reviews 82(Part 1)(2018) 1442–1456. https://doi.org/10.1016/j.rser.2017.05.165.

[4] L.-V. Badicu, U. Broniecki, W. Koltunowicz, J. Subocz, M. Zenker, A. Mrozik, Detection of bushing insulation defects by diagnostic monitoring, in: Proc. 2016 International Conference on Condition Monitoring and Diagnosis, 2016. https://doi.org/10.1109/CMD.2016.7757764.

[5] Z. Yang, W.H. Tang, A. Shintemirov, Q.H. Wu, Association rule mining-based dissolved gas analysis for fault diagnosis of power transformers, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 39(6) (2009) 597-610. http://doi.org/10.1109/TSMCC.2009.2021989.

[6] J.-F. Qin, Y. Yang, C. Gu, Z.-J. Hong, H.-Y. Du, Outlier Detection for Transformer's Oil Chromatographic Data Based on Metric Learning and the Weighted Local Outlier Factor, in: Proc. 2019 6th International Conference on Systems and Informatics (ICSAI), 2019. http://doi.org/10.1109/ICSAI48974.2019.9010155.

[7] F. Xiao, X.-W. Leng, K. Ye, Y.-L. Hu, X.-L. Li, L.-C. Zhu, The Research of Anomaly Detection Method for Transformer Oil Temperature Based on Hybrid Model of Non-Supervised Learning and Decision Forests, in: Proc. 2018 2nd International Conference on Power and Energy Engineering (ICPEE 2018), 2018. http://doi.org/10.1088/1755-1315/192/1/012020.

[8] J. Wang, M.-Z. Zhu, Z.-L. Gu, X.-L. Wang, The oil chromatographic data processing and analysis method based on the robustness model, in: Proc. 2022 7th Asia Conference on Power and Electrical Engineering (ACPEE), 2022. https://doi.org/10.1109/ACPEE53904.2022.9783808.

[9] J. Singh, S. Singh, A. Singh, Distribution transformer failure modes, effects and criticality analysis (FMECA), Engineering Failure Analysis 99(2019) 180-191. https://doi.org/10.1016/j.engfailanal.2019.02.014.

[10] D.-Q. Yang, W.-N. Zhang, J. Hao, Z.-G. Du, Diagnosis of Moisture Defects in Insulation of 252kV Main Transformer Bushing, in: Proc. 2022 5th International Conference on Energy, Electrical and Power Engineering (CEEPE), 2022.

http://dx.doi.org/10.1109/CEEPE55110.2022.9783341.
[11] Y.-X. Li, Q. Li, Z.-X. Liu, Q.-S. Chen, Diagnosis and analysis of abnormal heating fault for 35kV dry air core reactor, in: Proc. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021. http://dx.doi.org/10.1109/IAEAC50856.2021.9390877.
[12] D. Yu, W. Zhang, H. Wang, Research on Transformer Voiceprint Anomaly Detection Based on Data-Driven, Energies 16(5)(2023) 2151. https://doi.org/10.3390/en16052151.
[13] J. Lin, G.-H. Sheng, Y.-J. Yan, Q.-Q. Zhang, X.-C. Jiang, Online monitoring data cleaning of transformer considering time series correlation, in: Proc. 2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), 2018. http://doi.org/10.1109/TDC.2018.8440521.
[14] B. Cui, A.K. Srivastava, P. Banerjee, Synchrophasor-based condition monitoring of instrument transformers using clustering approach, IEEE Transactions on Smart Grid 11(3)(2020) 2688-2698. http://doi.org/10.1109/TSG.2019.2960043
[15] V.M. Catterson, S.E. Rudd, S.D.J. McArthur, G. Moss, On-line transformer condition monitoring through diagnostics and anomaly detection, in: Proc. 15th International Conference on Intelligent System Applications to Power Systems, 2009. https://doi.org/10.1109/ISAP.2009.5352860.
[16] I. Mitiche, T. McGrail, P. Boreham, A. Nesbitt, G. Morison, Data-Driven Anomaly Detection in High-Voltage Transformer Bushings with LSTM Auto-Encoder, Sensors 21(21)(2021) 7426-7432. https://doi.org/10.3390/s21217426.
[17] J.-J. Dai, H. Song, G.-H. Sheng, X.-C. Jiang, Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network, IEEE Transactions on Dielectrics and Electrical Insulation 24(5)(2017) 2828-2835. https://doi.org/10.1109/TDEI.2017.006727.
[18] Q. He, J. Si, D.J. Tylavsky, Prediction of top-oil temperature for transformers using neural networks, IEEE Transactions on Power Delivery 15(4)(2000) 1205-1211. http://doi.org/10.1109/61.891504.
[19] I.R. Abdulveleev, A.S. Karandaev, E.A. Khramshina, I.V. Liubimov, S.A. Evdokimov, V.R. Gasiyarov, Construction smart system for monitoring technical condition of furnace transformer vacuum circuit breaker, in: Proc. 2023 International Russian Smart Industry Conference (SmartIndustryCon), 2023. https://doi.org/10.1109/SmartIndustryCon57312.2023.10110723.
[20] W.-J. Liu, P.-F. Lei, D. Xu, X.-R. Zhu, Anomaly Recognition, Diagnosis and Prediction of Massive Data Flow Based on Time-GAN and DBSCAN for Power Dispatching Automation System, Processes 11(9)(2023) 2782. https://doi.org/10.3390/pr11092782.