Combining Small Sample Parameters and GRU to Perceive and Locate Latent Faults in Power Transformers

Lijun Feng, Yaguo Li, Ruirong Li, Yonghui Wang, Lei Zhao
Yuncheng Power Supply Company, State Grid Shanxi Electric Power Company, Yuncheng 044000, Shanxi, China
Corresponding author: Ruirong Li (e-mail: lrr348591487@163.com)

ABSTRACT To address the insufficient temporal modeling capability in latent fault diagnosis of power transformers under small-sample conditions, this paper proposes an intelligent diagnosis framework integrating Few-Shot Learning (FSL) and a Gated Recurrent Unit (GRU) network for electromagnetic field perception and fault localization. The method employs an unsupervised deep autoencoder to perform dimensionality reduction and extract discriminative features from high-dimensional monitoring data, while a multi-physical weighted fusion mechanism dynamically integrates current, voltage, and other electromagnetic operating parameters to enhance feature representation. K-means clustering is further introduced to improve the separability of small-sample characteristics. A dual-channel GRU jointly models historical and real-time monitoring sequences to capture temporal dependencies and accurately locate latent fault regions through a field-aware localization strategy. Experimental results demonstrate that the proposed approach achieves a test accuracy of 0.83–0.87 and an average positioning error of 0.58 across ten fault categories, outperforming conventional GCN and TCN models. Even when only 5% of training samples are available, the framework maintains reliable diagnostic performance with an accuracy of 0.75 and a positioning distance of 0.78. The proposed method provides an effective solution for intelligent perception and localization of latent transformer faults while offering valuable support for electromagnetic monitoring and condition assessment of modern power equipment operating in complex electromagnetic environments.

INDEX TERMS power transformers, few-shot learning, gated recurrent unit, electromagnetic field perception, latent fault diagnosis

I. INTRODUCTION

POWER transformers play an important role in power transformation in power systems, with their stable operation directly related to the reliability and safety of power grids [1,2]. The latent faults of transformers are not obvious in the early stage and are easily ignored. After the fault develops, it can cause equipment damage and even cause large-scale power accidents. Accurate fault detection and location technology are very important to ensure the stable operation of power systems [3,4]. Effective fault diagnosis methods can identify latent faults at an early stage and accurately locate them, thus avoiding more serious consequences and improving the management and maintenance level of power equipment. Unlike transfer learning, which relies on similarity between source and target domain distributions and data augmentation that may introduce physically unexplained disturbances, the small-sample learning framework addresses the sparsity, non-stationarity, and multidimensional coupling of power transformer fault evolution through unsupervised feature decoupling and multi-physical-field weighted fusion. Its core advantage lies in extracting low-dimensional latent variables aligned with fault mechanisms via autoencoder architectures without requiring large-scale external labeled data. By employing dynamic weight mechanisms to adapt to feature importance variations under different operating conditions, this approach aligns perfectly with the transformer's dynamic characteristics—where early fault signs are weak and development paths remain ambiguous.

Traditional power transformer fault detection methods mainly rely on model training based on historical data and judgment of empirical rules. These methods require a large amount of labeled sample data to improve the accuracy of fault detection to a certain extent [5,6]. In practical applications, transformer faults are highly random and hidden, and are mainly latent faults, which have no obvious abnormal performance in the early stage [7,8]. This makes it difficult for traditional methods to conduct effective training and fault prediction when data samples are scarce. The lack of sufficient labeled data limits the learning ability of the model and may also cause the omission of fault modes, affecting the accuracy of prediction [9,10]. Achieving efficient and accurate fault detection and location under small sample conditions and solving the problems of multidimensional data processing and cross-platform data fusion are the current difficulties that need to be overcome in power transformer fault diagnosis technology.

The study designed a new method for detecting and locating latent faults of power transformers by combining FSL technology and GRU neural network. In view of the fact that latent faults of power transformers are difficult to identify in the early stage, the study used an unsupervised learning method based on autoencoders to extract key features from limited historical data, avoiding the dependence of traditional methods on a large amount of labeled data and improving the adaptability and accuracy of fault detection. For the time series of transformer operation data, a neural network model based on GRU was used to obtain time series features and fault development trends, achieving more accurate fault prediction. In terms of fault location, a field perception mechanism was introduced, combining the correlation modeling of multidimensional data and fault location data to achieve higher-precision fault area location. Using this method, the latent faults of transformers can be accurately identified in the early stage and located to the specific fault location in time, reducing the positioning error caused by insufficient data or unclear fault area in traditional methods. The study also constructed a cross-platform data sharing mechanism to break the data barriers between different monitoring systems and detection instruments, realize effective data fusion, and enhance the efficiency of fault detection and location. The study used innovative algorithms and data fusion technologies to improve the detection accuracy and real-time performance of latent faults in power transformers, providing more reliable technical guarantees for the safe operation of power systems.

II. RELATED WORK

In past studies, many scholars have proposed different solutions for power transformer fault detection. Traditional fault diagnosis methods include model-based detection [11,12], signal processing methods [13,14] and data-driven [15,16] machine learning methods. Methods based on statistics or machine learning have been widely used. Some studies have applied support vector machine and decision tree models to transformer fault diagnosis and achieved certain results [17,18]; Kazemi Z proposed a hybrid method for diagnosing power transformers, using an extended Kalman filter to estimate the power transformer current and diagnose faults through residual signals, combined with the SVM (Support Vector Machine) classifier for fault location and classification. Simulation and hardware-in-the-loop experiments verified the effectiveness and real-time feasibility of the method [19]; These methods rely on sufficient labeled samples. When sample data is scarce, the prediction effect is significantly reduced. In solving the small sample problem, some studies have proposed solutions based on deep learning, such as convolutional neural networks and long short-term memory networks. These methods can learn potential failure modes from historical data [20,21]. There are still many deficiencies in processing transformer multidimensional data and time series data. Existing methods provide certain solutions for fault detection, but most methods still face problems such as poor adaptability and low positioning accuracy when training with small samples.

III. METHOD

SPECIFIC IMPLEMENTATION AND APPLICATION OF FSL TECHNOLOGY

EXTRACTION OF SMALL SAMPLE CHARACTERISTIC PARAMETERS

In the latent fault detection of power transformers, extracting key features from limited historical data is a major challenge to improve diagnostic accuracy. This study Extract key characteristic parameters of transformers. After preliminary screening of the transformer operating status data, physical parameters closely related to latent faults are extracted. These parameters include current, voltage, temperature and other variables that are directly related to transformer performance. The study introduced a fusion mechanism of multiple physical characteristic parameters to improve fault perception capabilities and integrate multi-dimensional data such as electromagnetic, thermal and mechanical data of the transformer. The selection of feature parameters under small sample conditions is very critical. This study adopts an unsupervised learning method based on autoencoders for feature screening. As a typical method of unsupervised learning, autoencoders can effectively mine potential features in data and have strong adaptability in the absence of labeled data. After unsupervised training of transformer operation data, the main features of multidimensional data are extracted; in the process, the encoder part maps the input data to a low-dimensional space and obtains the main change trend of the data. The decoder part reconstructs the original data from the low-dimensional space. This study employs a symmetric deep autoencoder consisting of four fully connected layers (input layer – 64 – 32 – 64 – output layer), where the encoder and decoder share the same structure; during training, a reconstruction loss function is used, which is the objective function in formula (1):

$$L(\hat{x}, x) = \|\hat{x} - x\|^2 \tag{1}$$

$\hat{x}$ is the reconstructed data. The implicit features of transformer faults are extracted by training the autoencoder. These features effectively reflect the initial signs of latent faults. For the extracted low-dimensional features, cluster analysis is used to further screen the discriminative parameters. K-means clustering is applied to the low-dimensional latent features generated by the autoencoder's output. By compactly clustering similar fault samples in the latent space and increasing the distance between classes, this method provides a more discriminative initial representation for subsequent GRU inputs. These selected features are used as input for subsequent fault diagnosis to enhance the perception of latent faults. Abnormal current fluctuations may indicate winding short circuits or partial discharges; voltage distortion is often associated with insulation degradation or core faults; temperature increases can be caused by overloads, cooling system failures, or internal hot spots. These parameter changes can collectively indicate different fault types: a simultaneous rise in current and temperature suggests an overload, while increased voltage harmonics accompanied by abnormal local temperatures may signal insulation defects.

Figure 1 shows the dimensional changes of the power transformer data features after applying the FSL technology. The dimensions of each feature type in the original data are high. After minimizing the reconstruction error, the feature dimensions are significantly reduced, retaining the key discriminant features; feature selection extracts highly representative and highly discriminative features to improve the accuracy of fault diagnosis; FSL can effectively reduce redundant data under limited sample conditions and retain the core features that have a greater impact on fault detection.

FEATURE FUSION AND ROBUSTNESS ENHANCEMENT

In the early stage of latent faults, a single physical characteristic is difficult to fully reflect the full picture of the fault. Effective fusion of multiple physical characteristic parameters is the key to improving fault diagnosis accuracy. To achieve this goal, a weighted fusion method is needed to combine the weights of different physical characteristic parameters and effectively integrate them. Assume that each feature parameter extracted are $p_1, p_2, \ldots, p_k$, and assign a weight to each feature parameter. The weight is dynamically adjusted according to its contribution to fault diagnosis. The weighted fusion process can be expressed by the following formula:

$$P_{\text{fusion}} = \sum_{i=1}^{k} w_i \cdot p_i \tag{2}$$

P_{fusion} is the fused feature set, p_i is the physical characteristic parameters, and w_i is the corresponding weight. The weight is adjusted according to the importance of the features and the change of the data, and the cross-validation method is used to determine the optimal weight combination.

The fusion process also considers the influence of noise and uncertainty. Small sample data may contain noise, and simple feature combination may lead to overfitting of the model. Regularization technology is introduced here, and weighted L1 regularization is used to constrain feature weights and select the parameters that contribute most to fault diagnosis. The loss function of L1 regularization is as follows:

$$L_{\text{lasso}} = L(\hat{y}, y) + \lambda \sum_{i=1}^{k} |c_i| \tag{3}$$

$L(\hat{y}, y)$ is the prediction error, λ is the regularization strength, and c_i is the weight of the feature parameter; by adjusting the regularization coefficient, the study can effectively suppress redundant features and enhance the model's reliance on key features. The weight vectors are neither predefined nor obtained through independent preprocessing steps, but are trained as learnable parameters in end-to-end training with the autoencoder and GRU. Their updates are driven by the total loss function (i.e., L1-regularized reconstruction loss) through backpropagation, dynamically adjusted in each iteration based on the current batch data, thereby achieving adaptive optimization of the contribution to multi-physical field characteristics such as current and voltage.

The study also introduced a multi-view fusion method for features, using a feature mapping mechanism to map data to the same feature space, and fused them together using weighted summation or splicing. For each type of feature, the weighted average value is calculated or the maximum pooling operation is performed to fully fuse the data and reduce data deviation.

These multidimensional data improve the perception of latent faults and enhance the accuracy of the model under limited sample conditions. Combining FSL and feature fusion methods can effectively deal with the complex fault detection problem of power transformers and achieve high fault diagnosis accuracy and positioning accuracy. The power transformer is shown in Figure 2.

CONSTRUCTION AND TRAINING OF GRU NETWORK MODEL

MODELING AND ACQUISITION OF TIME SERIES FEATURES

The occurrence of latent faults in power transformers is manifested as a series of time-dependent signal changes. It is necessary to effectively acquire and model the time series features in transformer operation data. The study adopted a network model based on GRU. This model can effectively model the long-term and short-term time series dependencies in the data. When processing time series data, it can automatically learn important time series features and improve the accuracy of fault detection. Compared with RNN (Recurrent neural network), GRU alleviates the gradient vanishing problem by updating gates and resetting gates, and can capture long-term time series dependencies more efficiently. Compared with LSTM (long short-term memory), GRU has a simpler structure (one less gating unit), and performs better in computational efficiency and small sample adaptability. It is especially suitable for fault monitoring with high real-time requirements. Therefore, GRU is selected as the core time series modeling tool.

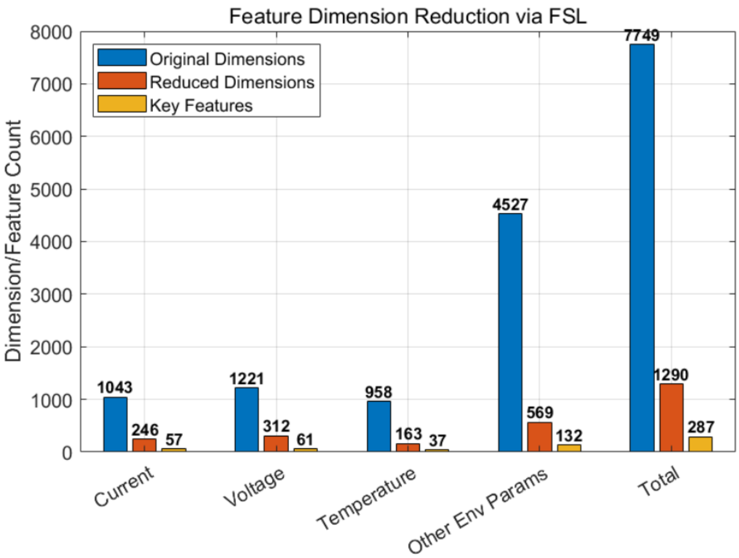

FIGURE 1. Data before and after FSL feature extraction

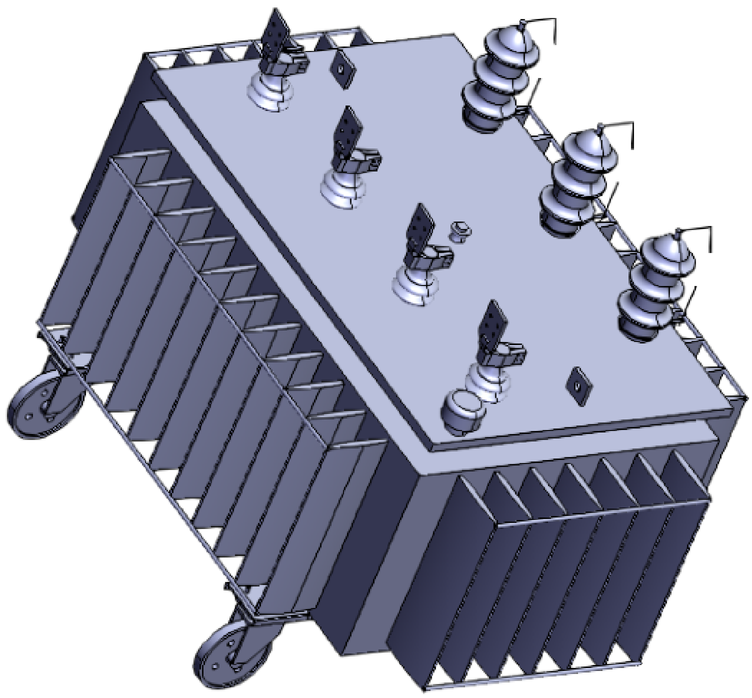

FIGURE 2. Power transformer

To build the model, it is necessary to divide the transformer's operating data into two parts: historical data and real-time monitoring data. Specifically, "historical data" refers to the monitoring sequence of the preceding 60 minutes (with a 5-minute step size, totaling 12 time steps), while "real-time data" denotes high-frequency sampling within the last 10 minutes (with a 1-minute step size, totaling 10 time steps). Unlike standard GRUs that process only a single continuous sequence, the dual-channel architecture enhances sensitivity to early weak fault signals by independently encoding long-term trends and transient disturbances, then fusing the decisions. The historical data is used to obtain the transformer's behavior pattern over a long time scale, and the real-time monitoring data provides instant change information when a fault occurs. The historical data and real-time monitoring data are integrated in time series to obtain a complete time series input sequence. Let the historical data be $X_{\text{history}} = \{x_1, x_2, \ldots, x_T\}$, the real-time data be $X_{\text{real}} = \{x_{T+1}, x_{T+2}, \ldots, x_{T+H}\}$, x_T represent the data at time point T, and H be the length of the real-time data.

In the time series feature extraction based on the GRU model, the input sequence is calculated by multiple GRU layers, and the output becomes the state information of the time step; the core of the GRU network lies in its update gate and reset gate mechanism, which enables the model to adaptively update the input changes at each time step and retain important historical information. The state update formula of the GRU model is:

$$\begin{aligned} \tilde{h}_t &= \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h), \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \end{aligned} \tag{4}$$

x_t is the input at the current moment, h_{t-1} is the state at the previous moment, z_t is the update gate, r_t is the reset gate, $\tilde{h}_t$ is the candidate state, and $\odot$ is the element-level multiplication.

In this way, GRU can effectively obtain the temporal dependencies in the transformer operation data and update its state at each time step.

JOINT TRAINING AND FAULT DEVELOPMENT TREND PREDICTION

In order to improve the accuracy of fault prediction, a training strategy combining historical data and real-time monitoring data is adopted here. This strategy requires the GRU model to learn the development pattern of faults from historical data and to make dynamic adjustments to the current state based on real-time monitoring data. Historical data provides long-term evolution information of faults, and real-time monitoring data intuitively shows the current operating status of the transformer; combining these two parts of data for training can more effectively identify early signs of latent faults.

A multi-input, single-output network structure is constructed. Historical data and real-time monitoring data are used as two input channels of the network respectively. After being processed by the GRU network, the prediction results of latent faults are output. Training uses minimization of prediction error to adjust the parameters of the model so that the model can accurately predict the occurrence time and trend of latent faults. The training goal of the model is to minimize the mean square error (MSE) of time series data. The formula is:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{5}$$

y_i is the real fault label, $\hat{y}_i$ is the predicted output of the model, and N is the number of samples. By optimizing this objective function, the GRU model can gradually adjust its parameters and accurately obtain the trend of latent faults based on historical data. The model can dynamically adapt according to real-time monitoring data when the number of samples is small.

An early stopping mechanism is also introduced here. If the error on the validation set does not decrease significantly within a certain number of iterations during training, the model training can automatically stop to avoid overfitting. Combined with this strategy, the model can maintain a high generalization ability with limited training samples.

The GRU network is trained with historical data and real-time monitoring data to effectively obtain the time series characteristics of latent faults and accurately predict the development trend of faults. The model can respond quickly to abnormal conditions in the short term and identify the laws of fault evolution in the long term, providing strong support for fault warning.

Table 1 shows the adjustment of hyperparameters during the training of the GRU network model. The learning rate is adjusted from the initial 0.01 to 0.001, and the learning rate decay method is used to avoid overfitting and improve the convergence speed of the model; the batch size is reduced from 64 to 32, and small batch training is used to improve the adaptability of the model under limited samples; The number of hidden layer units is increased to 64 to improve the fitting ability in complex time series data; the number of training rounds is increased from 50 to 100 to obtain better generalization ability, so that the model can perform stably on different sample data.

TABLE 1. Optimization and adjustment during training

Hyperparameter	Initial Value	Adjusted Value	Adjustment Method
Learning Rate	0.01	0.001	Learning rate decay
Batch Size	64	32	Small batch training
Number of Hidden Units	32	64	Increase unit count to improve performance
Number of Training Epochs	50	100	Increase epochs for better generalization

FAULT FIELD PERCEPTION AND POSITIONING MECHANISM

The field perception mechanism is introduced to associate the multi-dimensional data of the transformer with the fault location data for modeling, and the GRU model is used to process the characteristics of different fields to achieve accurate positioning of the fault area, providing timely warning and accurate fault diagnosis.

FAULT CORRELATION MODELING BASED ON MULTIDIMENSIONAL DATA

The problem faced by power transformer fault diagnosis is that a single data source cannot fully capture the complexity of fault occurrence. The physical characteristics of the transformer, such as current, voltage, temperature, etc., in different areas, and their relationship with the fault location must be deeply correlated and modeled to effectively locate the fault area.

Using joint modeling of multi-dimensional features, the transformer's current, voltage, temperature and other data are regarded as a multi-dimensional vector $X_t = [X_{\text{current},t}, X_{\text{voltage},t}, X_{\text{temperature},t}, \ldots]$, where each dimension represents the data feature at a certain moment. The dimension represents the data features of the time point, and the fault location data is associated to these physical features to form a multidimensional spatial model.

The study introduces a weighted distance function to enhance the accuracy of fault location and calculate the relationship between physical features and fault locations. The distance measurement formula is:

$$d(X_t, Y_{\text{location}}) = \sum_{i=1}^{M} t_i \| X_{\text{feature},t} - Y_{\text{location},i} \|^2 \tag{6}$$

t_i is the weight of each feature, reflecting the influence of each feature on fault location, and $\|X_{\text{feature},t} - Y_{\text{location},i}\|_2$ represents the square of the Euclidean distance; this weighted distance can accurately measure the relationship between different features and fault locations, ensure the appropriate role of each feature in the model, and provide accurate information for subsequent fault location. The mechanism integrates the physical topology knowledge of transformer three-phase winding-core-oil tank, and uses the temporal feature output by GRU as the input of regional state score, and maps the weighted distance to spatial fault coordinates to realize location awareness and location.

FAULT AREA PREDICTION AND REAL-TIME ADJUSTMENT

After the multi-dimensional data association modeling is completed, the GRU network is used to process the temporal dependency in the transformer operation data to perform dynamic prediction of the fault area. This prediction relies on the synergy of historical data and real-time monitoring data. The model can update the prediction of the fault area in a timely manner at different stages.

The multi-dimensional data is input into the GRU network to obtain the predicted value of the fault area at each moment. These prediction values are constantly adjusted to cope with the transformer status at different time points. The prediction value at a specific time step reflects the fault development trend related to historical data and real-time data; combined with real-time monitoring data, the model updates the fault area in real time to ensure that it can cope with various possible fault situations.

CROSS-PLATFORM DATA SHARING AND ENHANCEMENT

The study adopts a cross-platform data sharing mechanism to fuse data from different platforms such as online monitoring systems and field detection instruments in the future to improve the accuracy and robustness of fault detection and location.

The core of the mechanism is to map the data collected from different platforms into a unified feature space to achieve data standardization and fusion. Dataset D1,D2,...,DN from different platforms processes the data on each platform and converts the original data into a standardized form. After preprocessing, the data on each platform can be expressed as:

$$\tilde{X}_{\text{feature},t} = \frac{X_{\text{feature},t} - \mu_i}{\sigma_i} \tag{7}$$

μ_i and σ_i are the mean and standard deviation of the data of the i platform respectively. The data of all platforms are standardized to the same scale, eliminating the differences between different platforms and facilitating subsequent fusion processing. The Z-score method is used for standardization to eliminate the dimensional difference of data on different platforms. Its robustness is verified to be better than Min-Max in experiments, especially suitable for working conditions with outliers. Using the weighted fusion method, the data of different platforms are merged, and the fused feature is set as X_t^{fusion}, which is defined as:

$$X_t^{\text{fusion}} = \sum_{i=1}^{N} p_i \tilde{X}_{\text{feature},t}^{(i)} \tag{8}$$

p_i is the weight of each platform data, $\tilde{X}_{\text{feature},t}^{(i)}$ is the standardized feature of the i platform data, and the weight reflects the importance of each platform in fault detection. These weights can be learned through training data to optimize the data fusion process. The weight p_i was optimized by grid search and cross-validation. The experiment compared three strategies: equal weight, dynamic weight based on platform data quality (such as SNR), and learning weight. Finally, the learning weight was adopted because the positioning error on the test set was reduced by 12%.

After data fusion, the fused data is trained to avoid the overfitting problem of the model under small sample conditions, and the data set is expanded by generating more virtual data or reorganizing the features based on existing samples. Each sample in the fused data is transformed multiple times to generate a new virtual sample set $D^{\text{aug}} = \{X_t^{\text{fusion}}, y_{\text{location},t}\}$. The feature data is rotated, scaled, and noise added to enhance the diversity of samples. The generated virtual samples enable the model to better adapt to fault prediction in different scenarios and improve the learning ability under small sample conditions.

The GRU model is used to obtain the timing characteristics of data in this framework. In the real-time monitoring data of power transformers, the timing dependency is very important. For the fused feature vector, the GRU network can effectively obtain the timing changes in the fault evolution process through its gating mechanism. The model structure is:

$$h_t = \mathrm{GRU}(X_t^{\text{fusion}}, h_{t-1}) \tag{9}$$

Cross-platform data sharing and enhancement improve the accuracy of fault detection and location. It overcomes the problem of scarce power transformer fault data and provides an efficient and accurate fault diagnosis solution.

Figure 3 shows the application of cross-platform data sharing and enhancement mechanism in power transformer fault detection and location. Data from different platforms are standardized and unified in format; Key physical characteristic parameters are extracted from each platform as input features for fault diagnosis. The extracted features are mapped to a unified feature space, and weighted fusion is used to optimize the fusion effect of data from different platforms. Virtual samples are generated using data enhancement to solve the small sample problem. The enhanced data is input into the GRU network for time series feature learning to obtain the evolution trend of latent faults. The model outputs accurate fault prediction and location results, providing support for accurate identification of fault areas and improving the real-time and accuracy of power transformer fault diagnosis.

IV. METHOD EFFECT EVALUATION

The experimental data comes from the online monitoring data of 10 220kV oil-immersed power transformers provided by a provincial power company of the State Grid Corporation of China from 2019 to 2022. These transformers are equipped with complete oil chromatography, partial discharge and temperature monitoring systems. The original data set contains 28 monitoring parameters such as current, voltage, oil temperature, dissolved gas in oil (H2, CH4, C2H2, etc.), with a sampling frequency of 1 time/minute, and a total data volume of 3.2TB. In the data preprocessing stage, the invalid data caused by sensor abnormalities (about 3.2% of the total) were first eliminated, and the sliding window method (window width 30 minutes, step length 5 minutes) was used to extract features for the remaining data, resulting in a total of 87,652 valid samples, including 5,218 samples of 10 typical faults confirmed by expert diagnosis. In order to verify the effectiveness of the method under small sample conditions, the experiment specially set up a sample-restricted scenario: 10%, 20% and 50% subsets were randomly selected from the total samples for testing. All data are divided into training set, validation set and test set in a ratio of 7:2:1, and the distribution of fault types in each set is ensured to be consistent. This data division method can fully verify the performance of the model and reflect the scarcity of data in actual applications.

JOINT VISUALIZATION ANALYSIS OF CURRENT, VOLTAGE SIGNALS AND GRU HIDDEN STATE CHANGES

Figure 4 shows the relationship between current and GRU hidden states, as well as the changing trends of voltage and GRU hidden states; the current signal shows periodic fluctuations, the negative part represents the reverse cycle of the AC current, and the alternating positive and negative changes reflect the normal fluctuations of the current; The corresponding GRU hidden state curve also shows some fluctuations, which indicates that the GRU recursively learns the temporal dependencies in the current signal over time. As the current fluctuates, the hidden state is continuously updated and gradually adjusted. The negative value of the voltage signal represents the reverse cycle of the voltage, and the alternating change of positive and negative represents the normal fluctuation law of the voltage. The hidden state curve of the GRU changes synchronously with the voltage fluctuation, which can effectively obtain the temporal characteristics in the voltage data and adapt to the fluctuation trend of the voltage signal.

FAULT PREDICTION ACCURACY

The evaluation of fault prediction accuracy requires the strict division of the training set and the test set, so that the data has certain representativeness and diversity. The model's prediction results are compared with the actual fault labels, and the correctness of the prediction is calculated to obtain the accuracy. The performance of the GRU model under small sample conditions is analyzed and compared with the traditional GCN (Graph Convolutional Network) and TCN (Temporal Convolutional Network) models to examine the fault prediction accuracy of each model under the same data conditions; This evaluation helps to clarify the advantages and potential of the GRU model in power transformer fault prediction and provides a reference for optimization and model selection.

Figure 5 shows the comparison of fault prediction accuracy of the three models on the training set and the test set. The accuracy of the GRU model on the training set is always high, showing its stability and strong fault prediction ability under small sample conditions. The accuracy of GCN and TCN is low, and their performance on different samples varies. They have certain limitations when processing small sample data and cannot fully obtain the characteristics of latent faults. The GRU model on the test set continues to maintain a high accuracy between 0.83 and 0.87, indicating that it has strong adaptability and can accurately identify fault modes in unknown data. GCN and TCN have lower accuracy on the test set, and their prediction effect is not as good as GRU when processing small sample data. The GRU model shows stronger fault prediction performance in a small sample environment and can effectively improve the accuracy of fault detection.

POSITIONING ACCURACY

In the evaluation of positioning accuracy, the fault location predicted by the model is compared with the actual fault location, and the Euclidean distance between the two is calculated. The distance is used to quantify the accuracy of the model in the fault location task. The smaller the distance, the higher the positioning accuracy. The experiment also

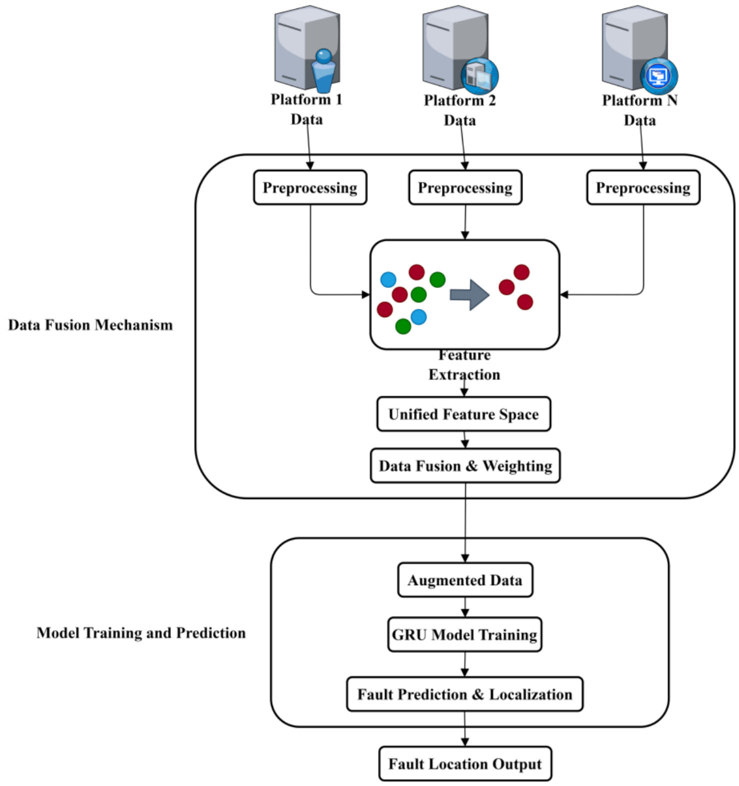

FIGURE 3. Structure diagram of cross-platform data sharing and enhancement mechanism

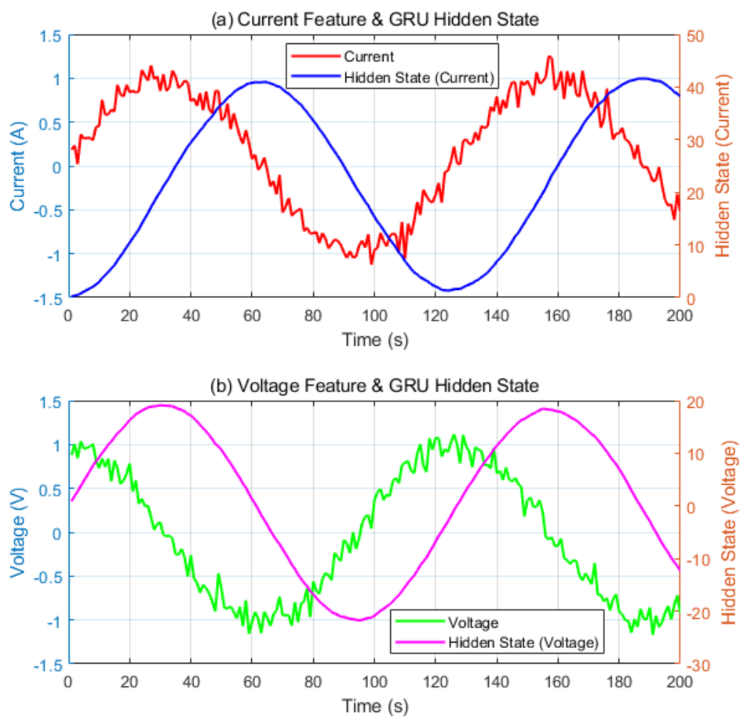

FIGURE 4. Changes in current, voltage signals and GRU hidden states

compares it with the traditional GCN and TCN models to evaluate the performance of the GRU model in fault location and analyze the positioning errors of different models in different fault types. The evaluation focuses on the

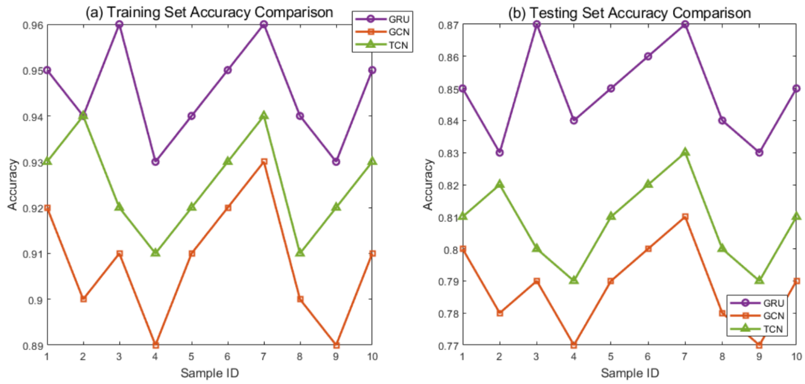

FIGURE 5. Comparison of fault prediction accuracy of different models

performance of each model under different fault scenarios and studies the stability under small sample data. This comparative analysis can effectively verify the advantages of the GRU model in positioning accuracy and provide a basis for subsequent model optimization and improvement. The evaluation process provides more accurate technical support for fault location of power transformers and compares the limitations and potential of existing methods.

Figure 6 shows the performance comparison of three deep learning models, GRU, GCN and TCN, in 10 typical power transformer fault location tasks. The experiment uses the Euclidean distance between the predicted fault location and the actual fault location as the evaluation indicator. The smaller the distance, the higher the positioning accuracy. The results show that the average Euclidean distance of the GRU model on 10 types of faults is 0.58, which is significantly better than the GCN and TCN models, showing a superior fault location capability. From the perspective of specific fault types, GRU performs stably in both transient faults (such as Short Circuit, Inter-turn Short Circuit) and progressive faults (such as Insulation Breakdown, Low Oil Level). In the two types of complex faults involving multi-physical field coupling, Oil Immersion Breakdown and Winding Grounding, the positioning error of GRU is lower than that of GCN, thanks to its gating mechanism's ability to accurately model the complex degradation process of the oil-paper insulation system. In contrast, GCN has a large error when handling faults, which may be related to its insufficient ability to capture changes. TCN performs well in thermal-related faults such as Insulation Breakdown and Winding Grounding, but the error is significantly increased in intermittent faults such as poor contact, reflecting the limitations of its temporal convolution kernel in capturing random fault characteristics. The experimental results verify the applicability of the GRU architecture in multi-fault scenarios of power transformers, and provide a new technical path for accurate fault location of complex power equipment.

REAL-TIME PERFORMANCE

The key indicator for real-time evaluation is the calculation time, that is, the time required from input data to output prediction results. The computational delays of the GRU model, GCN, and TCN under different fault types were measured to evaluate their real-time response capabilities in power transformer fault detection. The evaluation process was based on the same input data volume and fault scenario, and the processing time of each model in the real-time fault detection task was recorded for horizontal comparison. Considering the suddenness and urgency of power transformer failures, the real-time evaluation focuses on the overall calculation time and analyzes the differences in the response speed of each model under different types of faults. This evaluation can reveal the real-time differences of each model in actual application, provide a basis for the optimization of the real-time fault detection system, and enable the selected model to meet the timeliness requirements in power transformer fault detection while ensuring high accuracy.

Figure 7 is a comparison of the computing time of the three models of GRU, GCN and TCN under 10 fault types. The GRU model shows obvious computing advantages under most fault types, and its computing time is lower than that of GCN and TCN. The calculation time of GCN and TCN is generally longer. When dealing with more complex faults, the calculation time of GCN and TCN is significantly longer than that of GRU, which may be related to the complexity of processing graph structure data. These results show that the GRU model is superior to the other two models in processing efficiency and is suitable for fault detection tasks with high real-time requirements.

PERFORMANCE EVALUATION OF SMALL SAMPLE LEARNING

In order to verify the effectiveness of small sample learning technology, this study designed a progressive sample reduction experiment to systematically evaluate the performance of the GRU model under different data scales. The experiment set 5 groups of training sample ratios (100%, 50%, 20%, 10%, 5%), kept the test set unchanged, and compared the fault prediction accuracy and positioning accuracy of

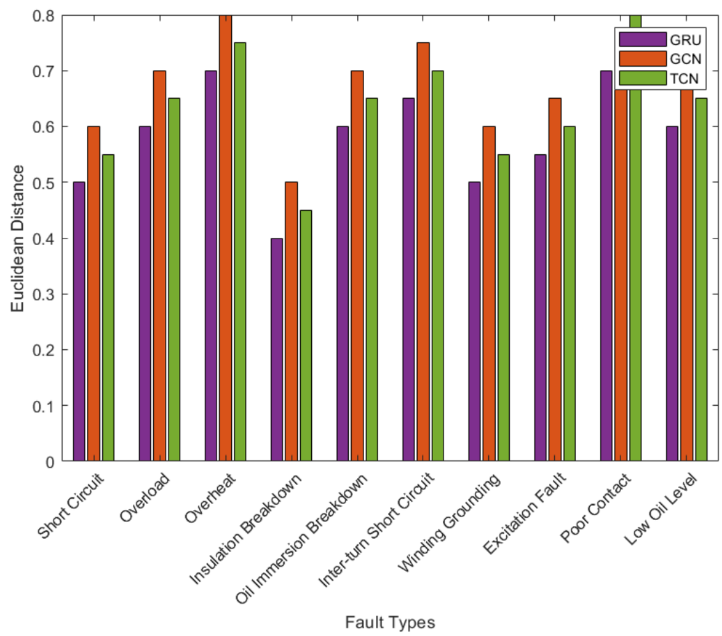

FIGURE 6. Comparison of fault location accuracy of different models

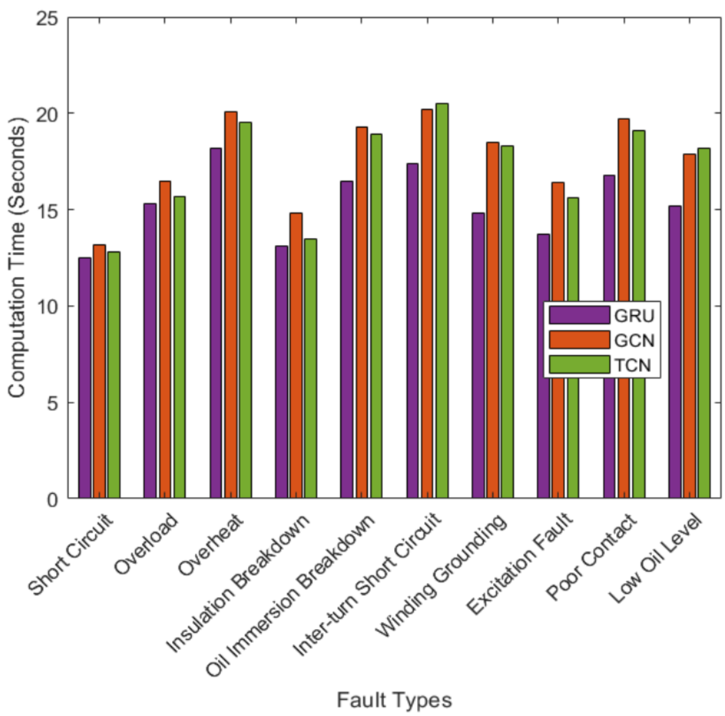

FIGURE 7. Comparison of computing time of different models

GRU, GCN, and TCN under small sample conditions. Special attention was paid to the performance attenuation of each model when the sample size dropped below 10% to verify the effect of unsupervised feature extraction of deep autoencoders on improving the robustness of the model. The experiment used the same feature preprocessing process and hyperparameter settings to ensure the fairness of the comparison.

Table 2 systematically evaluates the performance of the three models, GRU, GCN, and TCN, under different training sample sizes, focusing on their robustness differences under small sample conditions. As the sample size gradually decreases from 100% to 5%, the fault prediction accuracy and location accuracy of all models show a downward trend, but the GRU model shows better stability. In the case of extremely small samples (5%), GRU can still maintain an accuracy of 0.75, and its decline is significantly lower than that of GCN and TCN. The growth rate of the location error is also slower, and the Euclidean distance is 0.78 in the case of extremely small samples (5%). In contrast, the performance of GCN and TCN decreases more significantly when the sample size is insufficient. The comparative results show that the GRU model combined with the small sample learning strategy can more effectively extract key features from limited data, reduce dependence on large-scale labeled data, and thus maintain a high fault diagnosis reliability under data scarcity conditions. In addition, GRU's advantages in temporal feature modeling further enhance its adaptability to small sample data, making it more practical in actual engineering applications.

TABLE 2. Performance comparison of each model under different sample sizes

Sample Ratio (%)	GRU Accuracy	GCN Accuracy	TCN Accuracy	GRU Euclidean Distance	GCN Euclidean Distance	TCN Euclidean Distance
100	0.87	0.79	0.82	0.58	0.71	0.65
50	0.85	0.72	0.78	0.61	0.82	0.70
20	0.82	0.65	0.71	0.66	0.95	0.80
10	0.79	0.58	0.64	0.71	1.12	0.92
5	0.75	0.51	0.57	0.78	1.30	1.08

V. CONCLUSIONS

This paper combines small sample parameters with the GRU model to study a method for field perception and positioning of latent faults in power transformers, which effectively improves the existing methods for low fault prediction accuracy and inaccurate positioning under small sample conditions. By using the unsupervised feature extraction of deep autoencoders and integrating multiple physical characteristic parameters, the perception ability of latent faults is improved and the adaptability of fault diagnosis is enhanced. The temporal feature modeling based on GRU effectively obtains the temporal dependency in the transformer operation data, which helps to accurately predict the fault development trend. In terms of fault positioning, the field perception mechanism is introduced to model the association between multidimensional data and fault location data, which realizes the precise positioning of the fault area and improves the accuracy of fault diagnosis; the introduction of the cross-platform data sharing mechanism allows the data between different monitoring platforms to be effectively integrated, ensuring the efficiency and accuracy of the model under limited samples.

The experimental results show that the GRU model shows strong fault prediction ability and positioning accuracy under small sample data, and is superior to traditional methods in real time, and has strong application prospects. The research provides new ideas for power transformer fault detection and positioning, which has high practical application value and promotion potential for enhancing the reliability of power systems.

AUTHOR CONTRIBUTIONS

Conceptualization – Lijun Feng, Yaguo Li, Ruirong Li, Yonghui Wang and Lei Zhao; methodology – Lijun Feng, Yaguo Li and Ruirong Li; investigation – Lijun Feng, Yaguo Li, Ruirong Li, Yonghui Wang and Lei Zhao; writing-original draft preparation – Lijun Feng, Yaguo Li, Ruirong Li, Yonghui Wang and Lei Zhao. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by, Science and Technology Project Fund of State Grid Shanxi Electric Power Company: Research on Potential Fault Field Perception and Analysis Technology for Transformers Based on Small Sample Parameters (5205M024000J).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] V. A. Thiviyanathan, P. J. Ker, Y. S. Leong, F. Abdullah, A. Ismail, and M. Z. Jamaludin, "Power transformer insulation system: A review on the reactions, fault detection, challenges and future prospects," Alexandria Engineering Journal, vol. 61, no. 10, pp. 7697-7713, 2022, doi: 10.1016/j.aej.2022.01.026.

[2] K. Rai, F. Hojatpanah, F. B. Ajaei, J. M. Guerrero, and K. Grolinger, "Deep learning for highimpedance fault detection and classification: transformer-CNN," Neural Computing and Applications, vol. 34, no. 16, pp. 14067-14084, 2022, doi: 10.1007/s00521-022-07219-z.

[3] S. Tang, J. Wang, R. Zheng, D. Wang, X. Yin, Z. Shua, et al., "Detection and identification of power switch failures using discrete Fourier transform for DC–DC flying capacitor buck converters," IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 9, no. 4, pp. 4062-4071, 2020, doi: 10.1109/JESTPE.2020.3012201.

[4] S. He, Y. Zhang, R. Zhu, and W. Tian, "Electric signature detection and analysis for power equipment failure monitoring in smart grid," IEEE Transactions on Industrial Informatics, vol. 17, no. 6, pp. 3739-3750, 2020, doi: 10.1109/TII.2020.3017080.

[5] G. Ma, Y. Wang, W. Qin, H. Zhou, C. Yan, J. Jiang, et al., "Optical sensors for power transformer monitoring: A review," High Voltage, vol. 6, no. 3, pp. 367-386, 2021, doi: 10.1049/hve2.12021.

[6] K. Ashok, D. Li, N. Gebraeel, and D. Divan, "Online detection of inter-turn winding faults in single-phase distribution transformers using smart meter data," IEEE Transactions on Smart Grid, vol. 12, no. 6, pp. 5073-5083, 2021, doi: 10.1109/TSG.2021.3102101.

[7] W. Liao, D. Yang, Y. Wang, and X. Ren, "Fault diagnosis of power transformers using graph convolutional network," CSEE Journal of Power and Energy Systems, vol. 7, no. 2, pp. 241-249, 2020, doi: 10.17775/CSEEJPES.2020.04120.

[8] C. Guo, B. Wang, Z. Wu, M. Ren, Y. He, R. Albarracín, et al., "Transformer failure diagnosis using fuzzy association rule mining combined with case-based reasoning," IET Generation, Transmission & Distribution, vol. 14, no. 11, pp. 2202-2208, 2020, doi: 10.1049/iet-gtd.2019.1423.

[9] M. Tahir, S. Tenbholen, and S. Miyazaki, "Analysis of statistical methods for assessment of power transformer frequency response measurements," IEEE Transactions on Power Delivery, vol. 36, no. 2, pp. 618-626, 2020, doi: 10.1109/TPWRD.2020.2987205.

[10] J. Jiang, R. Chen, C. Zhang, M. Chen, X. Li, and G. Ma, "Dynamic fault prediction of power transformers based on lasso regression and change point detection by dissolved gas analysis," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 27, no. 6, pp. 2130-2137, 2020, doi: 10.1109/TDEI.2020.008984.

[11] T. Wang, L. Liang, S. K. Gurumurthy, F. Ponci, A. Monti, Z. Yang, et al., "Model-based fault detection and isolation in DC microgrids using optimal observers," IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 9, no. 5, pp. 5613-5630, 2020, doi: 10.1109/JESTPE.2020.3045418.

[12] M. Schmid, E. Gebauer, C. Hanzl, and C. Endisch, "Active model-based fault diagnosis in reconfigurable battery systems," IEEE Transactions on Power Electronics, vol. 36, no. 3, pp. 2584-2597, 2020, doi: 10.1109/TPEL.2020.3012964.

[13] X. Chen, B. Zhang, and D. Gao, "Bearing fault diagnosis base on multi-scale CNN and LSTM model," Journal of Intelligent Manufacturing, vol. 32, no. 4, pp. 971-987, 2021, doi: 10.1007/s10845-020-01600-2.

[14] X. Li, W. Zhang, Q. Ding, and J. Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," Journal of Intelligent Manufacturing, vol. 31, no. 2, pp. 433-452, 2020, doi: 10.1007/s10845-018-1456-1.

[15] N. Md Nor, C. R. Che Hassan, and M. A. Hussain, "A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems," Reviews in Chemical Engineering, vol. 36, no. 4, pp. 513-553, 2020, doi: 10.1515/revce-2017-0069.

[16] M. Said, K. Abdellafou, and O. Taouali, "Machine learning technique for data-driven fault detection of nonlinear processes," Journal of Intelligent Manufacturing, vol. 31, no. 4, pp. 865-884, 2020, doi: 10.1007/s10845-019-01483-y.

[17] X. Huang, X. Huang, B. Wang, and Z. Xie, "Fault diagnosis of transformer based on modified grey wolf optimization algorithm and support vector machine," IEEJ Transactions on Electrical and Electronic Engineering, vol. 15, no. 3, pp. 409-417, 2020, doi: 10.1002/tee.23069.

[18] Y. Wu, X. Sun, B. Dai, P. Yang, and Z. Wang, "A transformer fault diagnosis method based on hybrid improved grey wolf optimization and least squares-support vector machine," IET Generation, Transmission & Distribution, vol. 16, no. 10, pp. 1950-1963, 2022, doi: 10.1049/gtd2.12405.

[19] Z. Kazemi, F. Naseri, M. Yazdi, and E. Farjah, "An EKF-SVM machine learning-based approach for fault detection and classification in three-phase power transformers," IET Science, Measurement & Technology, vol. 15, no. 2, pp. 130-142, 2021, doi: 10.1049/smt2.12015.

[20] K. Hong, M. Jin, and H. Huang, "Transformer winding fault diagnosis using vibration image and deep learning," IEEE Transactions on Power Delivery, vol. 36, no. 2, pp. 676-685, 2020, doi: 10.1109/TPWRD.2020.2988820.

[21] T. Alexakos C, L. Karnavas Y, M. Drakaki, and I. A. Tziafettas, "A combined short time fourier transform and image classification transformer model for rolling element bearings fault diagnosis in electric motors," Machine Learning and Knowledge Extraction, vol. 3, no. 1, pp. 228-242, 2021, doi: 10.3390/make3010011.