Research on Fault Location Technology of Transformer Acoustic Feature Recognition and Field Perception Data Fusion under Small Sample Parameters

Yaguo Li, Lijun Feng, Ruirong Li, Xi Cui, Jun Li
State Grid Shanxi Electric Power Company Yuncheng Power Supply Company, Yuncheng 044000, Shanxi, China
Corresponding author: Ruirong Li (e-mail: Lrr348591487@163.com)

ABSTRACT Reliable transformer fault diagnosis under limited fault samples remains a significant challenge in intelligent power systems. To address the difficulties associated with weak fault signatures, severe environmental interference, and insufficient training samples, this study investigates transformer fault location technology based on acoustic feature recognition and field perception data fusion. The generation mechanism and propagation characteristics of transformer acoustic signals are first analyzed, and an improved time–frequency feature extraction method is developed to enhance feature representation under small-sample conditions. A multi-physics data fusion framework integrating acoustic, vibration, and electrical sensing information is then established, and a dedicated attention mechanism is designed to achieve deep feature fusion across heterogeneous data sources. Finally, an enhanced deep neural network model is employed for accurate fault localization and condition identification. Experimental results demonstrate that the proposed framework effectively improves fault recognition performance and location accuracy under small-sample constraints. The study provides technical support for intelligent power equipment monitoring and offers methodological references for signal propagation analysis, sensor fusion, and electromagnetic condition monitoring systems.

INDEX TERMS transformer fault diagnosis, acoustic feature recognition, data fusion, signal propagation, electromagnetic monitoring

I. INTRODUCTION

As a core hub in the power system, the operating status of power transformers directly affects the safety of the entire grid and the continuous operation of high-precision machinery. Any fluctuation in transformer stability can disrupt the synchronized motor control and intricate structural logic of automated production systems, making reliable power conversion essential for modern textile manufacturing. With increasing voltage levels and transmission capacity, transformer structures are becoming increasingly complex, and operating environments are becoming more severe, leading to a higher probability of faults and their potential risks. A fault can not only damage equipment but also trigger widespread power outages, causing significant economic losses and social impact. Therefore, achieving early warning and precise location of transformer faults is of great practical significance for ensuring power grid safety. For a long time, transformer fault diagnosis has relied primarily on traditional methods such as oil chromatography and electrical testing. While these methods can reflect the insulation status of transformers to some extent, they generally suffer from drawbacks such as long detection cycles, inability to achieve online monitoring, and difficulty in capturing signs of sudden faults. In recent years, with the development of sensor and signal processing technologies, fault diagnosis methods based on vibration and acoustic signals have gradually become a research hotspot. Among them, acoustic fingerprint detection technology, as a non-contact detection method, has advantages such as convenient installation and maintenance, wide coverage, and the ability to reflect acoustic changes caused by mechanical and electrical faults, providing a new perspective for transformer condition monitoring.

In practical applications, transformers have a complex operating environment with severe background noise, such as fan noise, cooler noise, and electromagnetic noise, etc., which are all interwoven, and the background noise is very rich, which easily covers the acoustic signature of the fault, resulting in a low signal-to-noise ratio. Secondly, transformer faults, in particular early-stage, have weaknesses in the electrical signal characteristics of transformer faults such as weak, non-stationary and non-linear, so the feature extraction of transformer faults is extremely difficult. More critically, transformers are extremely reliable devices and have a very low failure rate, resulting in an extremely low number of fault sample data - a typical small sample problem. In data-driven diagnostic models, when the sample size is too small, it will lead to overfitting, poor generalization ability and the inability to adapt to different transformer models and operating conditions. Furthermore, single-type sensor data often only reflects one aspect of the equipment's condition, for example, vibration signal reflects the loosening and deformation of the mechanical structure mainly, and acoustic signature signal reflects the winding deformation and partial discharge more sensitive. Therefore, integrating multi-physics sensing data from both the power transformer and the high-precision machinery is key to achieving the information complementarity necessary for accurate fault diagnosis. This data fusion allows for the simulta neous monitoring of electrical stability and the mechanical vibration patterns of automated weaving systems, ensuring the seamless operation of the industrial manufacturing environment.

II. RELATED WORK

With the development of smart grid and digital operation and maintenance, transformer fault diagnosis technology is evolving from traditional model-driven methods to data-driven and artificial intelligence-integrated approaches. Kumar et al. [1] proposed a method based on fault location factors to analyze leakage inductance or short-circuit impedance measured during a fault . Baadji et al. [2] constructed a deep neural network model based on Transformer for the detection, classification and location prediction of faults on transmission lines. Kumar et al . [3] established a mathematical model by deriving the analytical relationship between fault location and measured reactance. Then, by solving the model, the specific location of the fault and the severity of the fault were determined. Fault diagnosis models based on deep learning lack interpretability. Li et al. [4] introduced variational inference into a self-attention mechanism to enable the model to learn more robust and interpretable feature representations. An et al. [5] used Transformer to extract features and introduced Gaussian mixture variational autoencoder to jointly model the features of the source and target domains . Sun et al. [6] significantly reduced the number of model parameters and computational complexity by designing a grouped self-attention mechanism and an efficient convolutional feedforward network, while maintaining a strong ability to capture local and global features in vibration signals. Lin et al. [7] deeply integrated convolutional neural networks with Transformers and achieved advanced performance on multiple bearing fault diagnosis datasets. The fusion residual attention model proposed by Zhou et al . [8] can effectively learn discriminative features from transformer vibration signals and achieve high-precision fault diagnosis. The infrared spectroscopy technology combined with pattern recognition method proposed by Darwish et al . [9] can quickly and accurately evaluate the state of transformer insulating oil and diagnose faults. Nanfak et al. [10] combined unsupervised machine learning with human expert experience to construct a two-stage hybrid diagnostic method to improve the accuracy and reliability of diagnosis. This paper aims to construct a fault location model with high interpretability and strong generalization ability under small sample conditions by fusing acoustic features and multi-physics field data, so as to improve the accuracy and reliability of transformer fault diagnosis in complex operating environments.

III. METHODS

A. TRANSFORMER ACOUSTC FEATURE EXTRACTON MECHANISM UNDER SMALL SAMPLE PARAMETERS

Under the condition of the parameters of small samples, how to extract highly discriminative features from limited acoustic fingerprint data, is the main difficulty of the localization of faults. This paper proposes a timefrequency analysis feature extraction framework based on adaptive parameter optimization. The framework preprocesses the collected raw acoustic fingerprint signal, including mean removal, filtering, and background noise suppression based on spectral subtraction, in order to improve the signal-to-noise ratio in the first place. Considering there exists a limited amount of small data samples, a data augmentation strategy is introduced in the feature extractive part[11-12] .

In the process of specific feature extraction, an improved technique of Mel frequency cepstral coefficient extraction was adopted. Traditional MFCCs are mostly based on the characteristics of human hearing. Although they can reflect envelope characteristics of acoustic signals, they are not sensitive enough to high frequency impact components unique to transformer faults. In order to overcome this, a nonlinear frequency scaling factor is introduced in this paper when computing the Mel filter bank. Based on the energy distribution characteristics of transformer fault acoustic signatures, the center frequency and bandwidth of the filter are dynamically adjusted, so as to make the feature extraction more exquisite to the fault frequency band. Simultaneously, in order to capture the dynamic change of the acoustic signature signal, the first-order difference coefficients and the second-order difference coefficients are calculated respectively based on the static MFCCs and a composite feature vector containing the dynamic information in the time domain is constructed. While this high-dimensional feature vector has a lot of rich information, it also creates redundancy and computational clutter. Therefore, a feature dimensionality reduction method based on Principal Component Analysis (PCA) is further proposed. PCA maps the original features to a new orthogonal coordinate system through linear transformation, retaining the component with the largest variance, thereby reducing the feature dimension. This method has few parameters, is easy to implement, and has good reproducibility, effectively extracting the main information of acoustic features.

In order to further explore the deep correlation between small sample data, a feature optimization module based on metric learning was designed. This module constructs a triplet loss function to force the model to narrow the distance between similar fault acoustic features in the feature space and widen the distance between different fault features, thereby enhancing the intra-class compactness and inter-class separation of features. In specific implementation, by selecting and combining anchor samples, positive samples and negative samples, the model can still learn a robust feature distribution even with a limited number of samples. Through the above mechanism, not only are acoustic features that can reflect the essence of transformer faults extracted, but also the training difficulties caused by small samples are overcome through data augmentation and metric learning [13] .

B. MULTPHYSICS SENSING DATA FUSION AND ALIGNMENT TECHNOLOGY

Transformer faults are generally accompanied by coupled changes in several physical fields, e.g. sound, vibration, temperature and electrical quantities. Monitoring a single physical quantity is easily affected by the interference of the environment and the limitations of the quantity itself, so it is not easy to demonstrate the true state of the fault. Therefore, the fusion of multi-physics sensing data plays an important role in promoting the accuracy of fault location. This paper proposes a multi-source data processing framework based on the spatiotemporal alignment and feature level fusion. This framework first solves a time synchronization issue of multi-source heterogeneous data. Since the sampling rate of acoustic sensors, vibration sensors, and electrical quantity acquisition devices are different, and the beginning time of data acquisition is different, the direct fusion will cause the feature to be misaligned. In order to solve this, the method of time delay estimation based on cross-correlation functions is adopted. Using the high frequency sampled acoustic signal as benchmark, the time delay of other sensor data with respect to the acoustic signal is calculated, and interpolation algorithm is used to ensure that all data streams are consistent with respect to the time axis, which ensures that different physical quantities at the same time indicate the exactly same equipment state event.

In the spatial dimension, considering the large size of the transformer structure, the response intensity and phase of sensors at different locations to the same fault source are significantly different. This paper establishes a three-dimensional spatial coordinate model of the transformer and maps the installation positions of all sensors to the model. By analyzing the propagation attenuation characteristics of different physical field signals, a spatial weight matrix is constructed to compensate for the signal intensity differences caused by distance and medium attenuation. On this basis, a feature fusion module based on a multi-head attention mechanism is adopted. This module inputs the acoustic feature vector, vibration feature vector and other auxiliary feature vectors as queries, keys and values into the attention layer, respectively. Through selfattention operation, the model can automatically learn the correlation weights between different physical quantity features and dynamically adjust the contribution of each source data in the fusion feature [14] . Furthermore, to achieve deep fusion of field-aware data, a tensor fusion network was introduced. This network models the feature interactions of different modalities as a high-order tensor decomposition process, generating more expressive joint representations by mining the nonlinear interaction information between modalities. During the fusion process, to address the instability of parameter estimation caused by small sample sizes, regularization constraints and DropConnect techniques were employed to prevent overfitting in the fusion layer.

C. CONSTRUCTON OF FAULT LOCATON MODEL BASED ON IMPROVED TRANSFORMER

After completing the acoustic fingerprint feature extraction and multi-physics data fusion, the core part of this paper is to build the model of fault localization, which can effectively deal with the time feature and adapt to the situation of small samples. The Transformer architecture with its powerful long sequence modelling ability and global attention mechanism shows natural advantages in deal with such problems. However, the standard Transformer model contains huge number of parameters, which is prone to the overfitting issue when trained directly on small sample datasets with less data, and its high computational complexity is not ideal for the realtime diagnosis. To overcome such problems, this paper determines a lightweight improved Transformer fault localization model. This model conducts deep optimization on the encoder structure, and replaces part of the linear mapping structure in the traditional multi-head self-attention mechanism structure with depthwise separable convolutional layers, which greatly reduces the number of parameters and calculation cost, while maintaining the ability to capture local temporal features.

The key novelty of the model is the addition of sparse attention mechanism. Considering often times fault signals show anomalies in specific frequency bands or over a certain period of time, there is a huge amount of redundancy when computing fully connected attention. The sparse attention mechanism, by limiting the queries so that they attend only to a subset of key-value pairs, for example, only to the local neighborhood and a few key nodes at a longer distance, greatly reduces computational complexity, while preserving the ability of the model to capture long-range dependency. Furthermore, this constraint on sparsity plays a regularizing role, which helps to enhance the generalization effect of the model on small samples. Regarding the encoding of the location, based on the fact that the location information of the transformer fault signals is an important factor for localization, the learnable relative location encoding is used in this paper. This allows the model to better see the relative positions of the signals over time and this is important for helping to find the specific time and location of the fault. To improve the learning ability of the model further in the case of limited sample, in this paper a prototype network learning module is designed based on the pre-training and fine-tuning paradigms. This module forms prototype centers for each type of fault in the feature space and make decisions during the classification or localization by calculating the distance between features of the test sample and each type of prototype. This metric based learning approach helps the model to adapt fast with very few samples to newer type of faults. At the output part of the model, a multi-task learning head is constructed and one task is a fault type classifier and another task is a fault location regression predictor. The regression predictor is designed to output a continuous coordinate value representing the physical location of the fault point along the winding axis. To correspond with the experimentally defined discretized evaluation criterion (dividing the winding into 10 uniform segments), the model-predicted continuous coordinate values are quantized to the nearest segment number in post-processing. This mapping from continuous physical space to discrete evaluation metrics allows for intuitive evaluation of positioning accuracy in "segments".By optimizing not only the loss for classification but also the loss for regression, the model indeed can not only determine what fault occurred, but can locate the location of the fault at the same time. The model is trained with the AdamW optimizer and the cosine annealing learning rate schedule strategy to be sure that the model ends up with a global optimum (or near immune) solution on small sample datasets.

IV. RESULTS AND DISCUSSION

A. DATASET CONSTRUCTON AND EXPERIMENT SETUP

To check the effectiveness of the proposed method, a transformer multiphysics data set with various operating conditions and types of faults were built. Experimental data were collected from the monitoring data of actual substations. The acoustic signals were obtained using a high-sensitivity microphone array sampling with a sampling rate of 44.1 km Hz to achieve the coverage of high-frequency acoustic components produced by transformer faults. Vibration signals were acquired with piezoelectric accelerometers located at different positions on the tank wall with a sampling rate of 10 kHz. Simultaneously, electrical and non-electrical parameters like top oil temperature, ambient temperature and load current were recorded. The experimental setup was designed so that five typical conditions were simulated, such as inter turn short circuit, winding deformation, multi point grounding of the core, loose insulation, and normal operation. Different severity levels were applied to each fault condition and data was gathered under no-load, half and full load conditions to simulate complex operating environments. The training set consists of only 50 samples for each fault class, whereas the test set consists of 200 samples under other conditions to be used for testing the generalization ability. Some of the evaluation metrics are accuracy, precision, recall, F1 score (classification), mean absolute error, and localization success rate (localization). Comparison models: SVM, 1D-CNN, standard transformer and LSTM. All models are trained and optimized in the same conditions. Five-fold cross-validation is a mechanism used to determine robustness. Ablation experiments are aimed at comparing the performance differences among single acoustic fingerprint, single vibration and fused features.

B. PERFORMANCE EVALUATON AND COMPARATVE ANALYSIS OF FUSION MODELS

The results of performance evaluation for each model after training and testing are presented in Table 1. The results in the data table demonstrate that the proposed fusion model has better performance than other comparative models in all parameters. Under conditions of small samples, although the traditional support vector machine model has a certain ability of generalization, its feature extraction ability is limited when dealing with complex acoustic fingerprint and vibration signals, and the accuracy can only be 78.5%, and the F1 score reaches 0.76. While the one-dimensional convolutional neural network can extract the local features, they lack the ability to capture the global information and have the issue of overfitting when dealing with a small sample case. Its accuracy is increased up to 84.2% but its F1 score is only 0.83 which shows it is not stable much. The normal modeling power of the Transformer model, which is a global model, has an accuracy rate of 88.7%, which is better than the above two. However, the high number of parameters is causing it to seriously overfit on a training set (only 50 samples) and giving relatively low recall rate on a test set. In contrast, the improved model proposed in this paper, by introducing a sparse attention mechanism and lightweight design, not only greatly reduces the number of parameters, but also effectively solves the overfitting problem. More importantly, by fusing multi-physics information such as acoustic signatures and vibrations, the model would have richer feature representations. Experimental results show that accuracy, precision, recall and F1 score of the proposed model are 94.6%, 95.1%, 94.2% and 0.946 (all of which are at the leading level) on the test set, respectively. This fully shows the success of the introduction of adaptive time-frequency analysis in the feature extraction stage and the use of a multi-head attention mechanism and tensor fusion network in the model fusion stage. In particular, the performance improvement of the fused model is extremely important compared with models using only a single acoustic signature or one vibration feature.

A deeper analysis of the data in Table 1 reveals that the single acoustic signature model performs slightly worse in identifying low-frequency vibration faults such as multi-point grounding of the iron core, while the single vibration signature model is insufficiently sensitive to high-frequency acoustic emission signals generated in the early stages of inter-turn short circuits in the winding. The fusion model proposed in this paper dynamically adjusts the weights of the two models through an attention mechanism, suc-

TABLE 1. Performance comparison of different fault diagnosis models on transformer fault datasets

Model Name	Accuracy (%)	Precision (%)	Recall rate (%)	F1 score	Parameters
Support Vector Machine (SVM)	78.5	77.2	79.8	0.760	-
One-dimensional convolutional neural network (1D-CNN)	84.2	83.5	85.0	0.830	1.2M
Long Short-Term Memory (LSTM) network	86.7	85.9	87.4	0.865	2.5M
Standard Transformer	88.7	87.5	89.9	0.886	12.8M
Single acoustic fingerprint feature model	89.3	88.5	90.2	0.893	3.5M
Single vibration characteristic model	90.1	89.8	90.5	0.901	3.5M
This article fusion model	94.6	95.1	94.2	0.946	1.68M

cessfully achieving complementary advantages, which is the key reason why it maintains high performance across various fault types. Furthermore, a comparison of parameter counts shows that while the proposed model outperforms the standard Transformer, its parameter count is only about one-third of the latter. This greatly facilitates the deployment of the model on edge computing terminals and meets the real-time and low-power requirements of the power Internet of Things.

C. FAULT LOCATON ACCURACY AND ROBUSTNESS ANALYSIS

In addition to a classification of types of faults, an assessment of the fault location capacity of the method is presented in this paper. The results of the fault location are given in Table 2, which shows the mean absolute error in the predicted location of the model as compared to the actual location with different types of faults. To denote the location error, the transformer winding was evenly divided into 10 segments along the axial direction, and named as 1-10, with the predicted location corresponding to the number of the segment. As described in Section 3.3, the model's regression predictor first outputs continuous physical coordinates, which are then quantized to the nearest segment number to align with the evaluation criteria set in the experiment. This means that the "Average Location Error (Segment)" in Table 2 is calculated based on this quantized result, reflecting the deviation between the predicted segment and the actual fault segment. As can be seen in the data presented in Table 2, for winding deformation faults, the average error is very small, being only 0.4 segments, meaning that the model is able to locate the fault within a very small range; hence, extremely high location accuracies. For minor faults like loose insulation, due to the relatively weak characteristics in sound pattern and vibration, it is a little larger in the location error, but it is also only 0.8 segments, which fully satisfies the location requirement for fault diagnosis in engineering field. Overall, the average location error of the model with all the types of faults is controlled within 0.6 segments with a high rate of location success , which verifies the model's effectiveness in locating faults.

To further confirm the robustness of the model in the complex real working environment, the anti-interference performance tests were performed. Gaussian white noise at different signal-to-noise ratios was added to the original test data and used for the simulation of different levels of environmental noise interference at the substation sites. The results of the test are available in table 3. As the signal-to-noise ratio (SNR) decreases, i.e. noise intensity increases, the accuracy of all models decreases. However, the fusion model proposed in this paper has very good anti-interference ability in the noisy environment. When the SNR decreases to 0 dB i.e., when the noise energy equals the signal energy, the model's accuracy of the traditional 1D-CNN model is extremely low at 65.4%, whereas the accuracy for the proposed model is 87.5%. This is mainly attributed to the improvement of the design of the Mel filter bank in the acoustic feature extraction stage, which can well suppress noise in non-fault frequency bands, and the improvement of the weights of effective fault features by the multi-head attention mechanism in the fusion layer.

Analysis of data in the third table indicates the benefit of multi-physics data fusion in high noise environments is further enhanced. As far as noise is concerned, there is a high probability that data of a single sensor is randomly contaminated, and a lower probability that acoustic signature, vibration data and electrical quantities will be contaminated at the same time. By mining the consistency between multimodal data with tensor fusion networks, the model can detect and remove abnormal features caused by noise to ensure high diagnostic accuracy. This result shows that the proposed small-sample fault location technique not only perform excellently under ideal conditions but also has good robustness under harsh field conditions, which is a good solution for the actual operation and maintenance process of transformers.

TABLE 2. Location error analysis for different fault types

Fault type	Sample size	Average positioning error (segment)	Location success rate (%)
Winding inter-turn short circuit	50	0.5	94.0
Winding deformation	50	0.4	96.0
Multi-point grounding of iron core	50	0.7	90.0
Loose insulation	50	0.8	88.0
Insulation damp	50	0.6	93.0
average value	250	0.6	92.2

TABLE 3. Comparison of model robustness tests under different signal-to-noise ratios

Signal-to-noise ratio	1D-CNN accuracy (%)	Standard Transformer Accuracy (%)	The accuracy rate of the model in this paper (%)
Clean data (noise-free)	84.2	88.7	94.6
20 dB	81.5	86.2	93.5
10 dB	75.3	82.1	91.2
5 dB	68.7	76.5	89.4
0 dB	65.4	72.8	87.5

V. CONCLUSION

This paper addresses key challenges in transformer fault diagnosis, such as small sample parameters and the isolation of complex acoustic signature signals within the high-decibel environment of automated production machinery. By resolving the unreliability of single information sources, the study ensures precise monitoring of power stability required for the continuous operation of high-precision equipment. It proposes a fault location technology based on the fusion of acoustic signature feature recognition and field perception data. By constructing an adaptive time-frequency analysis feature extraction mechanism, the difficulty in representing acoustic signature features under small sample conditions is effectively solved, and data augmentation and metric learning strategies are used to improve the discriminative power of the features. The proposed multiphysics spatiotemporal alignment and deep feature fusion framework successfully achieves complementarity and enhancement of heterogeneous data such as acoustic signatures and vibration, constructing a comprehensive fault feature representation. Based on this, a lightweight improved Transformer fault location model is designed. Through sparse attention mechanisms and prototype network learning, the model significantly reduces the number of parameters while significantly improving its generalization ability and location accuracy under small sample conditions. Future work will explore the model's adaptability to diverse fault types within complex field environments, particularly those involving high-precision weaving machinery, and integrate digital twin technology to achieve intelligent health management throughout the transformer's lifecycle. This holistic approach ensures a robust technical guarantee for the smart grid by synchronizing power stability with the mechanical operational demands of the modern textile manufacturing sector.

AUTHOR CONTRIBUTIONS

Conceptualization –Yaguo Li, Lijun Feng, Ruirong Li, Xi Cui and Jun Li; methodology – Yaguo Li, Ruirong Li, Xi Cui and Jun Li; investigation – Yaguo Li, Lijun Feng, Ruirong Li, Xi Cui and Jun Li; writing-original draft preparation – Yaguo Li, Lijun Feng, Ruirong Li, Xi Cui and Jun Li. 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] A. Kumar, B. R. Bhalja, and G. B. Kumbhar, "A new improved methodology to identify an interturn fault location in transformer winding based on fault location factor," IEEE Transactions on Industrial Electronics, vol. 71, no. 11, pp. 15024-15033, 2024, doi: 10.1109/TIE.2024.3357897.

[2] B. Baadji, S. Belagoune, and S. E. Boudjellal, "Transformer-based deep learning networks for fault detection, classification, and location prediction in transmission lines," Network: Computation in Neural Systems, vol. 36, no. 4, pp. 1837-1857, 2025, doi: 10.1080/0954898X.2024.2393746.

[3] A. Kumar, B. R. Bhalja, and G. B. Kumbhar, "Novel technique for location identification and estimation of extent of turn-to-turn fault in transformer winding," IEEE Transactions on Industrial Electronics, vol. 70, no. 7, pp. 7382-7392, 2022, doi: 10.1109/TIE.2022.3201309.

[4] Y. Li, Z. Zhou, C. Sun, et al., "Variational attention-based interpretable transformer network for rotary machine fault diagnosis," IEEE transactions on neural networks and learning systems, vol. 35, no. 5, pp. 6180-6193, 2022, doi: 10.1109/ TNNLS.2022.3202234.

[5] Y. An, K. Zhang, Y. Chai, et al., "Gaussian mixture variational-based transformer domain adaptation fault diagnosis method and its application in fault bearing diagnosis," IEEE Transactions on Industrial Informatics, vol. 20, no. 1, pp. 615-625, 2023, doi: 10.1109/TII.2023.3268750.

[6] W. Sun, R. Yan, R. Jin, et al., “LiteFormer: a lightweight and efficient transformer for rotating machine fault diagnosis,” IEEE Transactions on Reliability, vol. 73, no. 2, pp. 1258-1269, 2023, doi: 10.1109/TR.2023.3322860.
[7] T. Lin, Y. Zhu, Z. Ren, et al., “CCFT: The convolution and cross-fusion transformer for fault diagnosis of bearings,” IEEE/ ASME Transactions on Mechatronics, vol. 29, no. 3, pp. 2161-2172, 2023, doi: 10.1109/TMECH.2023.3312935.
[8] Y. Zhou, Y. He, Z. Xing, et al., “Vibration signal-based fusion residual attention model for power transformer fault diagnosis,” IEEE Sensors Journal, vol. 24, no. 10, pp. 17231-17242, 2024, doi: 10.1109/JSEN.2024.3382811.
[9] M. M. F. Darwish, M. H. A. Hassan, N. M. K. Abdel-Gawad, et al., “A new technique for fault diagnosis in transformer insulating oil based on infrared spectroscopy measurements,” High Voltage, vol. 9, no. 2, pp. 319-335, 2024, doi: 10.1049/hve2.12405.
[10] A. Nanfak, A. Hechifa, S. Eke, et al., “A combined technique for power transformer fault diagnosis based on k-means clustering and support vector machine,” IET Nanodielectrics, vol. 7, no. 3, pp. 175-187, 2024, doi: 10.1049/nde2.12088.
[11] H. Fang, J. An, H. Liu, et al., “A lightweight transformer with strong robustness application in portable bearing fault diagnosis,” IEEE Sensors Journal, vol. 23, no. 9, pp. 9649-9657, 2023, doi: 10.1109/JSEN.2023.3260469.
[12] J. Cen, Z. Yang, Y. Wu, et al., “A mask self-supervised learning-based transformer for bearing fault diagnosis with limited labeled samples,” IEEE Sensors Journal, vol. 23, no. 10, pp. 10359-10369, 2023, doi: 10.1109/JSEN.2023.3264853.
[13] A. Hechifa, A. Lakehal, A. Nanfak, et al., “Improved intelligent methods for power transformer fault diagnosis based on tree ensemble learning and multiple feature vector analysis,” Electrical Engineering, vol. 106, no. 3, pp. 2575-2594, 2024, doi: 10.1007/s00202-023-02084-y.
[14] L. Zhang, Z. Song, Q. Zhang, et al., “Generalized transformer in fault diagnosis of Tennessee Eastman process,” Neural Computing and Applications, vol. 34, no. 11, pp. 8575-8585, 2022, doi: 10.1007/s00521-021-06711-2.