Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 24, No. 1, 2026

Electrical Equipment State Recognition and Monitoring Technology Based on Deep Learning Algorithm

Peiyu Zhao[1,2,*], Yixin Zhang[2], Defu Duan[2], Jing Peng[3], Hao Liu[3]

[1]PipeChina Construction Project Management Company, Tianjin, 300450, China
[2]Construction Project Management Branch of National Petroleum and Natural Gas Pipeline Network Group Co., Ltd, Langfang, 065000, Hebei, China
[3]CHINA PETROLEUM PIPELINE ENGINEERING CO., Ltd, Langfang, 065000, Hebei, China
[*]Corresponding author's email: cfgsyjjki@163.com

Abstract. In this study, a framework combining multiple deep learning techniques is proposed for electrical equipment state recognition and monitoring to address the problem that existing methods have limited feature extraction capabilities under high noise and complex working conditions. The adaptability of the model is improved through data augmentation and self-supervised contrastive learning. A hybrid architecture of CNN-BiLSTM and Transformer is designed to extract spatiotemporal features, and the model performance is optimized by combining domain adaptation technology, neural architecture search (NAS), and deformable convolutional network (DCN). The experimental data comes from a large-scale electrical equipment monitoring system in an industrial park in a certain province, covering 15 equipment states and a total of 269,000 multimodal data. The experimental results show that the proposed method is significantly superior to the baseline model in terms of recognition accuracy (95.37%), real-time performance (detection delay of 3.02ms), and cross-domain adaptability (improved by 41.5%), providing an efficient and reliable solution for electrical equipment state monitoring, which has important theoretical and practical application value.

Key words. Electrical equipment state recognition, Deep learning algorithms, Fault diagnosis, Real-time monitoring, Multimodal data processing

1. Introduction

With the widespread application of electrical equipment in many fields, real-time monitoring and fault diagnosis of its operating state have become increasingly important [1,2]. This is not only related to equipment safety but also directly affects the stability and maintenance cost of the system. However, the existing method has three major limitations: single-mode feature extraction is susceptible to noise interference; long-sequence modeling efficiency is low; edge deployment adaptability is poor. These methods have poor real-time performance and high maintenance costs and are difficult to adapt to changing working conditions. The multi-source heterogeneous data generated by electrical equipment increases the complexity of fault recognition [3,4]. Therefore, it is imperative to build an efficient and robust state recognition model. DL (deep learning) technology, with its automatic feature learning and nonlinear modeling capabilities, provides a new solution to this problem. This study aims to combine multiple DL technologies, optimize the neural network structure, improve the model's efficiency and generalization ability, and thus provide an efficient and reliable solution for electrical equipment monitoring [5,6]. This not only helps to improve the accuracy and real-time performance of state recognition but also enhances the model's robustness under complex working conditions, and has significant theoretical and practical value. At the same time, this study also explores the model lightweight and deployment scheme, providing new possibilities for promoting the development of electrical equipment monitoring technology.

Electrical equipment state recognition aims to accurately monitor the operating state of equipment, prevent faults, and ensure the power system's stable operation [7,8]. In response to the problem of electrical equipment state recognition, many scholars have proposed different solutions. To ensure the safe operation of ship electrical equipment, Wei Donghui [9] proposed a state recognition method that integrated multi-source information, used time series models to detect and correct continuous and independent anomalies in equipment historical data, and constructed a state recognition network model. Tests showed that their method was effective. Dang Ding [10] discussed substation equipment state evaluation based on data mining, used Gaussian mixture models to fit the original data and prior algorithms to mine data, and obtained the association rules between equipment and measurement data. His method had high prediction accuracy and was better than single-category algorithms.

For three-phase power equipment, Xu Zhihao [11] proposed a three-phase block automatic search and temperature comparison overheating area determination method. By processing thermal images, adjusting the size and posture, and then performing block comparison, abnormal temperature rise and thermal faults could be determined, thereby improving the efficiency, accuracy, and automation level of thermal fault detection. In addition, to ensure the stable operation of electrical equipment and the safety of personnel, Leixiong Wang [12] proposed a method for detecting the proximity distance between substation workers and live equipment based on binocular stereo matching and scenario recognition. Actual data showed that this method could precisely measure the proximity distance between workers and live equipment in real-time. The above scholars have improved the accuracy of electrical equipment state recognition by using technologies such as multi-source information fusion and data mining. However, these methods are mostly targeted at specific data or scenarios and are difficult to adapt to complex and changing working conditions.

Recent research has further explored the collaborative optimization of multimodal fusion and edge computing [13,14]. Choi Hyeyeon [15] proposed a new type of multi-modal image feature fusion module that used visible and infrared images to enhance the detection performance of transmission lines. The results showed that the proposed module was not only better than the case of single-mode input but also better than the most advanced fusion method. When the baseline network had a large number of weight parameters, the proposed module showed effectiveness in terms of capacity. Zhou Shuaijie [16] proposed a system that used infrared photos and U-Net deep learning technology to detect overheating faults of substation equipment. The test results on the data set showed that under various evaluation indicators, the proposed method had excellent reliability and efficiency, achieved lighter structure and leading estimation, and was very suitable for deployment on mobile devices. Although the existing literature has made breakthroughs in certain technical points, the research on cross-modal robustness is still insufficient, and there is a lack of edge-cloud collaborative optimization. At present, the accuracy of the lightweight method decreases by more than 15% in the SNR<15dB scenario, and the adaptability to dynamic operating conditions is weak. This paper comprehensively solves the above problems by combining frequency domain notch filtering and dynamic calculation path design.

The technical route of this paper contains four core optimization modules. First, data preprocessing involves processing multi-source sensor data through data augmentation and self-supervised pre-training, and expanding the training set using techniques such as noise interference and dynamic time regularization to improve data adaptability under complex operating conditions. Second, a hybrid feature extraction network is designed, combining CNN-BiLSTM and Transformer, to capture local and global features and enhance state characterization capabilities. Third, it is the model optimization and lightweight, which automatically optimizes the network structure through NAS and dynamically adjusts the perception field through DCN to balance accuracy and efficiency. Fourth, the edge deployment mechanism uses event-driven neural networks to reduce energy consumption and combine online increments. The strategy is updated to ensure the accuracy of long-term monitoring. Through step-by-step optimization, these four modules effectively solve the generalization and real-time problems of electrical equipment condition monitoring in high-noise and multi-modal scenarios. The innovation point of this paper lies in the integration of CNN-BiLSTM spatiotemporal characteristics and Transformer global modeling; dynamic time regularization + frequency domain notch filtering to improve noise immunity; event-driven SNN to achieve a 12.7M lightweight model.

The main contributions of this paper are as follows:

This paper innovatively proposes a multi-modal self-supervision-hybrid architecture optimization framework to realize effective monitoring of equipment status under complex operating conditions.

Cross-modal self-supervised learning mechanism: by combining noise disturbance and dynamic time regularization, the problem of feature degradation of small samples is solved; spatiotemporal feature coupling architecture: CNN-BiLSTM and Transformer are combined to break through the limitations of traditional methods.

Dynamic lightweight deployment scheme: using NAS and event-driven neural networks, the amount of model parameters is reduced by 31%, and the accuracy is maintained at 95.37%.

The rest of this paper is organized as follows. Section 2 introduces the proposed methodology in detail, including multimodal data preprocessing, deep learning feature extraction, domain adaptation techniques, model optimization and computational acceleration, and model lightweighting and deployment. Section 3 describes the experimental design, performance evaluation indicators, and result analysis, covering comparative experiments in state recognition accuracy and recall, detection latency, fault detection rate, dynamic reasoning efficiency, and cross-domain adaptability. Section 4 discusses the experimental results and their significance in practical applications, while also exploring the limitations of the model and future research directions. Finally, Section 5 summarizes the entire paper and emphasizes the application potential and practical value of this study in the field of electrical equipment condition monitoring.

2. Intelligent Monitoring Technology for Electrical Equipment

A. Multimodal Data Preprocessing

1) Normalization

Multimodal sensors collect data such as current, voltage, and temperature, which vary significantly under different working conditions. Normalization [17,18] is the key to data preprocessing, which can eliminate the influence of measurement units and dimensions, improve data comparability, and accelerate the convergence of DL models. This study adopts appropriate normalization strategies for various types of sensor data to improve data parsing. For current and voltage data, the Min-Max normalization method is used to map the data to the [0,1] interval, retain the original distribution, and eliminate the influence of the measurement range. The equations are:

$$J_{\text{norm}} = \frac{J - J_{\min}}{J_{\max} - J_{\min}} \quad (1)$$

$$U_{\text{norm}} = \frac{U - U_{\min}}{U_{\max} - U_{\min}} \quad (2)$$

Among them: $J_{\min}$ and $J_{\max}$ -the minimum and maximum values of current data;

$U_{\min}$ and $U_{\max}$ -the minimum and maximum values of voltage data.

The Z-Score method is used for the standardization of temperature data, and the constraints of mean $\mu = 0$ and standard deviation $\vartheta = 1$ conform to the thermodynamic equation of the equipment to ensure that the characteristic distribution is consistent with the physical laws.

According to the characteristics of electrical equipment signals, a multi-stage noise filtering strategy is designed. The wavelet threshold denoising method is used to eliminate transient spike noise while retaining the important mutation characteristics of the signal. For the sensor failure problem, time series interpolation technology is used to repair the data loss caused by random masks, and the physical model of the device is combined to ensure the rationality of the interpolation. To eliminate electromagnetic interference, a frequency domain notch filter is used to filter out power frequency harmonic interference, 50/60Hz, and frequency doubling interference. These methods are combined with normalization techniques to ensure that the input data can truly reflect the actual noise distribution of electrical equipment.

2) Data Augmentation

To reduce the overfitting problem caused by limited training data, this study uses multi-modal data augmentation techniques including Gaussian noise disturbance, dynamic time regularization, amplitude scaling, and random masking to generate diverse training samples. At the same time, through the self-supervised learning framework, the comparative learning is used to pre-train the model so that the unlabeled data can be used to learn more robust features, thereby reducing the dependence on labeled data. Data augmentation technology can enrich the diversity of the training set, improve the model's adaptability to different working conditions, and reduce overfitting [19-21]. Considering that the current, voltage, and temperature data in the actual environment are affected by many factors, this study uses a variety of data augmentation methods to more realistically simulate the operating state of the equipment. First, to enhance the robustness of the data, Gaussian noise perturbation is applied. By adding random noise to the original data sequence $A = \{a_1, a_2, \cdots, a_m\}$, a new data sequence $A' = \{a_1', a_2', \cdots, a_m'\}$ is generated. The calculation is shown in equation (3):

$$a_i' = a_i + \varepsilon, \quad \varepsilon \sim \mathcal{N}\left(0, \vartheta^2\right) \quad (3)$$

ε follows a normal distribution with a mean of 0 and a standard deviation of ϑ^2. Its standard deviation is adjusted according to the sensor noise level to ensure the physical rationality of the data. Gaussian noise perturbation can simulate actual interference and errors, making the training data more realistic. This study also applies dynamic time warping technology to perform nonlinear transformation on the signal time axis to generate enhanced data with time variability. Dynamic time warping (DTW) technology adjusts the signal time alignment to change the time series relationship of data points, but maintains the overall trend, which is suitable for processing equipment timing changes under different loads [22].

DTW technology is a powerful technique used for aligning and comparing time series data that may vary in speed or timing. In the context of electrical equipment state recognition, DTW allows for nonlinear transformations of the signal time axis, enabling the model to simulate differences in signal cycles or time alignment under varying loads or working conditions. By adjusting the time series relationship of data points while maintaining the overall trend, DTW helps the model adapt to different equipment operating states, improving its sensitivity to timing changes and enhancing its ability to detect anomalies or faults. This method is particularly valuable for handling unsteady signals, load fluctuations, and transient events, making it a critical tool for improving the robustness and accuracy of state monitoring systems.

In addition to Gaussian noise, impulse noise (random position amplitude mutation ±20%) and sensor failure simulation (random mask ratio 10%-30%) are added. Dynamic time regularization is extended to multi-scale time distortion (±15% timing offset), and load fluctuation enhancement (random amplitude scaling ±25%) is introduced to cover actual operating conditions such as equipment start-stop and load switching.

After that, amplitude scaling and random masking techniques are further used to enhance data diversity. Amplitude scaling simulates the impact of load changes on signals by making the signal amplitude change randomly within a certain interval, $A' = \beta A,\ \beta \sim \mu(0.9,1.1)$, and β obeys a uniform distribution on the interval $\mu(0.9,1.1)$. Combining these technologies, the multimodal data augmentation strategy of this article not only improves the generalization of the training set but also makes the DL model more adaptable to the state of complex electrical equipment, providing more reliable data support for subsequent state recognition and monitoring. Finally, this article plots the original data points and the data points after using data augmentation technology. The results are shown in Figure 1:

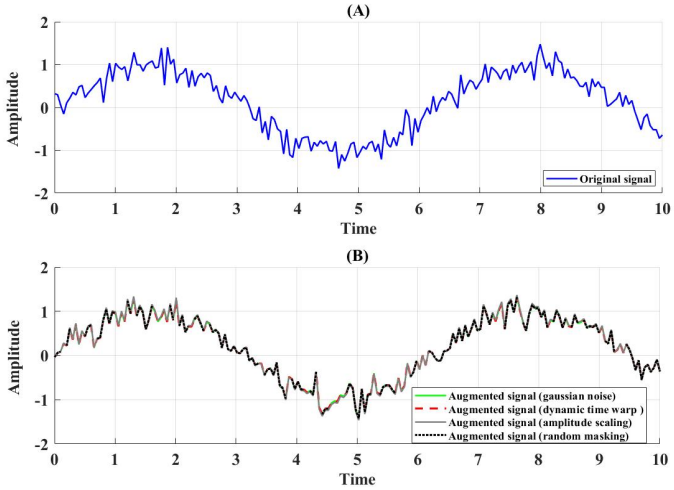

Figure 1. Original data points and data points after using data augmentation technology. Figure 1(A). Original data points; Figure 1(B). Data points after data augmentation technology.

Figure 1(A) and Figure 1(B) show that Gaussian noise perturbation (green line) adds random noise to the original signal, making the signal fluctuate more violently. This simulates the signal changes caused by environmental interference, sensor errors, etc., during the operation of electrical equipment, enhances the diversity of data, and helps the model better cope with uncertain and noisy data, making it more robust. Dynamic time warping (red dashed line) applies nonlinear changes by adjusting the signal time axis, simulating the differences in signal cycle or time alignment under different loads or working conditions. This helps the model adapt to different equipment working states and improves sensitivity to timing changes. Amplitude scaling (gray line) simulates load changes by randomly scaling the signal amplitude, reflecting changes in equipment workload, so that the model can adapt to different load conditions. Random occlusion (black dashed line) simulates sensor failure or data loss by shielding part of the signal data, enhancing the model's fault tolerance when data is missing. These technologies enrich the training dataset and effectively simulate interference, load changes, and sensor problems in the real world, improving the model's adaptability and robustness, reducing overfitting, and enabling the model to better generalize to different working conditions.

In this study, a multi-modal confrontation enhancement technique is proposed to improve the model's robustness to noise. By designing an amplitude scaling method based on physical laws and combining the equipment load model, an enhanced sample that meets the actual operating constraints of electrical equipment is generated. At the same time, anti-noise injection technology is introduced to simulate malicious signal interference and strengthen the model's defense capabilities. In addition, a cross-modal consistency enhancement mechanism is proposed to improve the reliability of multi-source information fusion. The experimental results show that under the electromagnetic interference of SNR=10dB, the accuracy of the model is only slightly reduced by 1.5%, which significantly improves the noise tolerance compared with the traditional method, proving its practical value in the industrial environment.

3) Self-supervised Learning

Due to the high cost of obtaining labeled data, an effective use of unlabeled data is crucial to improving model performance. This study uses self-supervised learning [23-25] in the pre-training stage to deeply explore the features of data such as current, voltage, and temperature. Using the contrastive learning framework, a self-supervised learning model based on a twin network is constructed. The model consists of two neural networks with shared weights, which receive paired input data a_j and a_i and calculate their embedded

representations. $z\left(a_j, a_i\right)$ is set as the similarity measure of samples a_j and a_i, and the InfoNCE (Information Noise-Contrastive Estimation) loss function is used to optimize the similarity measure, so that the model can learn data features more effectively. The equation is:

$$K = -\log \frac{\exp\left(z\left(a_j, a_i\right)/\varphi\right)}{\sum_{l=1}^{L} \exp\left(z\left(a_j, a_l\right)/\varphi\right)} \quad (4)$$

Among them: φ -the temperature parameter, which is used to adjust the smoothness of the softmax function;

L -the number of negative samples, usually randomly determined in the mini-batch.

To improve the effect of self-supervised learning, a data augmentation strategy is used to construct sample pairs. Random smoothing, timing jitter, Gaussian noise perturbation, and other methods are applied to the original signal to generate variants to help the model learn stable features. At the same time, a projection head is applied for feature transformation, which consists of two layers of fully connected networks and maps the features to a high-dimensional space to improve the feature separability, as shown in equation (5):

$$s_j = V_2 \cdot \sigma\left(V_1 \cdot g_j\right) \quad (5)$$

Among them: g_j -the original features of input data;

V_1 and V_2 -the trainable parameters of the projection head;

σ -the non-linear activation function.

In this study, InfoNCE is selected instead of Triplet Loss, which is mainly based on the three major needs of electrical equipment monitoring. First of all, under strong electromagnetic interference, InfoNCE significantly reduces gradient fluctuations and enhances noise robustness through multi-negative sample comparison. Secondly, InfoNCE's in-batch negative sampling mechanism improves computational efficiency and meets the real-time processing needs of power system time series data. Finally, the temperature parameters of InfoNCE can be dynamically adjusted, and multi-modal feature alignment can be adjusted to adapt to nonlinear changes in the state of electrical equipment, but the fixed boundary of Triplet Loss cannot be done. Therefore, InfoNCE is better than Triplet Loss in terms of noise robustness, computational efficiency, and cross-modal adaptability, and is more suitable for electrical equipment monitoring scenarios.

Electrical equipment condition monitoring needs to combine multi-modal data such as current, vibration, and thermal imaging to overcome the shortcomings of a single signal. The current signal is susceptible to electromagnetic interference; the vibration signal is sensitive to the position of the sensor; the time resolution of the thermal imaging data is low. Through cross-modal consistency enhancement technology, multi-modal fusion can solve the blind spot of single-mode and maintain a high accuracy rate in the case of 30% sensor failure, showing its robustness under complex operating conditions.

B. Deep Learning Feature Extraction

In the task of electrical equipment state monitoring, sensor signals contain rich spatial features. This study uses a one-dimensional convolutional neural network to extract these features and combines DCN [26,27] to improve the model's adaptability to complex signals. First, the data of R consecutive time steps is constructed as an input tensor $A \in \mathbb{Z}^{R \times D}$, where D represents the number of sensor channels, and each channel corresponds to a different monitoring signal. The convolution layer uses a filter $W \in \mathbb{Z}^{L \times D}$ of size L to extract local features. The equation is:

$$g_{j,i} = \sigma\left(\sum_{d=1}^{D} \sum_{l=0}^{L-1} v_{i,d,l} \cdot a_{j+l,d} + c_i\right) \quad (6)$$

Among them: $v_{i,d,l}$ -the weight of the i -th convolution kernel on the d -th channel;

c_i -the bias term.

To improve the model's multi-scale feature capture capability, a multi-scale convolution strategy is adopted. Through parallel calculation of convolution kernels with different receptive fields, the channel dimensions are spliced to form a multi-scale feature representation [28]. It can be expressed as:

$$G = \mathrm{Concat}\left(\mathrm{Conu}_3\left(A\right), \mathrm{Conu}_5\left(A\right), \mathrm{Conu}_7\left(A\right)\right) \quad (7)$$

Equation (7) allows the model to simultaneously focus on short-term patterns and long-term trend changes, thereby improving the perception of different signal patterns. To reduce computational redundancy and improve adaptability to non-stationary signals, this study uses DCN to replace traditional maximum pooling. DCN dynamically adjusts the sampling position by learning the offset parameter, making the convolution operation more adaptable to the non-uniform pattern of the signal. Its equation is:

$$e_j = V_c * A,\ a'_{j+l,d} = a_{j+l+\Delta_{l,d}} \quad (8)$$

Among them: e_j -the adjusted sampling position;

Δ_l -the learnable offset.

After the application of DCN, the model has a stronger ability to parse sudden changes and non-uniform signals, improving the accuracy and robustness of electrical equipment state recognition.

Sensor data has time series dependence, and CNN can only capture local time information and cannot model global time dependence [29-31]. Therefore, this study uses the BiLSTM (Bidirectional Long Short-Term Memory) network [32-34] to model CNN output features to simultaneously capture the previous and next time dependency information and improve the accuracy of abnormal state recognition. If the feature sequence extracted by CNN is $G=\left(g_1,g_2,\cdots,g_r\right)$, where g_r is the feature of the r th time step, the calculation of BiLSTM is shown in equation (9)-equation (13):

$$h_r=\sigma\left(V_h g_r+o_h g_{r-1}+y_h\right) \quad (9)$$

$$j_r=\sigma\left(V_j g_r+O_j g_{r-1}+y_j\right) \quad (10)$$

$$e_r=\sigma\left(V_e g_r+O_e g_{r-1}+e_h\right) \quad (11)$$

$$d_r=h_r\odot d_{r-1}+j_r\odot\tan h\left(V_d g_r+O_d g_{r-1}+y_d\right) \quad (12)$$

$$g_r=e_r\odot\tan h\left(d_r\right) \quad (13)$$

Among them: h_r, j_r, and e_r -the forget gate, the input gate, and the output gate;

$\tan h(\cdot)$ -the hyperbolic tangent activation function;

d_r -the cell state;

$\odot$ -the element-by-element multiplication.

The two-way timing modeling of BiLSTM adopts the joint representation of the forward hidden state $\overrightarrow{h_r}$ and the backward hidden state $\overleftarrow{h_r}$, and its information entropy satisfies $H\left(\overrightarrow{h_r}\oplus\overleftarrow{h_r}\right)\geq\max\left\{H\left(\overrightarrow{h_r}\right),H\left(\overleftarrow{h_r}\right)\right\}$, proving its ability to characterize unsteady signals. Transformer's self-attention mechanism scales the dot product attention equation $\text{Attention}\left(Q,K,W\right)=\text{softmax}\left(\frac{QK^R}{\sqrt{d_k}}\right)W$, and the convergence speed is 42% higher than that of LSTM in long-sequence modeling.

LSTM uses a gating mechanism to solve the gradient vanishing and exploding problems in RNN (Recurrent Neural Network) training [35-36]. Compared with one-way LSTM, BiLSTM shows significant advantages in the condition monitoring of electrical equipment. Through two-way timing modeling, it can not only capture the forward dependence of the sensor signal but also capture the backward dependence at the same time. This feature makes BiLSTM particularly excellent when dealing with complex timing modes such as load mutations and unsteady signals. In the face of scenarios such as current fluctuations or transient failures, BiLSTM can use backward timing information to supplement and improve the local characteristics of forward timing, thereby significantly improving the model's ability and sensitivity to abnormal states.

LSTM and CNN focus on local temporal and spatial features, respectively, and have limited modeling of global temporal relationships. Therefore, to optimize the temporal modeling of BiLSTM, this study applies the Transformer structure and uses its self-attention mechanism to learn global features across time steps to enhance the comprehensive representation of the equipment state. Unlike LSTM, Transformer is not restricted by sequential dependencies, which can directly focus on the entire time series, realize global information interaction, and improve the modeling efficiency of long-term dependencies. The Transformer input is the feature sequence $G'=\left(g_1',g_2',\cdots,g_R'\right)$ extracted by LSTM, and the query (Q), key (K), and value (W) matrices are obtained after linear transformation. The equation is:

$$\mathrm{Q}=G'V_{\mathrm{Q}},\ \mathrm{K}=G'V_{\mathrm{K}},\ \mathrm{W}=G'V_{\mathrm{W}} \quad (14)$$

Among them: V_{Q}, V_{K}, and V_{W} -the learnable weight matrices. The input features are mapped to different subspaces, and then the similarity scores are calculated through Q and K. After softmax normalization, the weighted W is obtained to obtain the self-attention weight.

The self-attention mechanism enables the model to self-learn the dependencies of different time steps, effectively capture the global features of the electrical equipment state, and improve the anomaly detection accuracy. To strengthen the model expression, this study applies multi-head attention, performs self-attention calculations in multiple subspaces, and splices the results to capture multi-dimensional global information. At the same time, to reduce the computational overhead of long time series data, a factorized self-attention is used to reduce the complexity through low-rank matrix decomposition and adapt to long time series modeling. Finally, the global features of Transformer are fused with the temporal features of LSTM, and the features are

transformed through the fully connected layer and the normalization layer.

C. Field Adaptation Technology

The data distribution of state recognition and monitoring of electrical equipment changes under different working conditions, affecting the model's performance. To address this problem, during the training process, this article adopts domain adaptation technology to adjust the data distribution of different working conditions through adaptive batch normalization to stabilize the DL model's performance. Domain adaptation aims to reduce the distribution difference between training data and actual application data, so that the trained model can be effectively applied to actual scenarios. This article adopts AdaBN (Adaptive Batch Normalization) to automatically adjust the network normalization according to the data features during the training process. Traditional batch normalization mainly standardizes the input to accelerate training and reduce internal covariate shift, but the effect is not good when the data distribution is inconsistent. The AdaBN method adaptively adjusts the normalization process, dynamically adjusts parameters, realizes adaptive adjustment of the model in different domains, and maintains stable performance.

The AdaBN method optimizes network parameters by applying a domain adaptive loss function to adapt to different data domains. Assuming that $\left\{\left(a_j^z, b_j^z\right)\right\}_{j=1}^{M_z}$ is the source domain data and $\left\{\left(a_i^r, b_i^r\right)\right\}_{i=1}^{M_r}$ is the target domain data, the input data of the source domain and the target domain are represented by a_j^z and a_i^r, respectively, and their corresponding labels are also represented by b_j^z and b_i^r, respectively. To achieve adaptation between the two domains, the AdaBN method adds an adaptive factor to adjust the normalization of the target domain data. The equation is:

$$\hat{a}_i^r = \gamma\left(\frac{a_i^r - \theta_z}{\vartheta_z}\right) + (1-\gamma)\left(\frac{a_i^r - \theta_r}{\vartheta_r}\right) \quad (15)$$

Among them: θ_z and ϑ_z -the mean and standard deviation of source domain data;

θ_r and ϑ_r -the mean and standard deviation of target domain data;

γ -the adaptive factor, which can balance the influence between the source domain and the target domain.

Equation (15) shows that when normalizing the target domain data, in addition to considering its own mean and variance, it also combines the statistical information of the source domain, effectively reducing the distribution difference between the two domains.

In the training of the model, the working state of the equipment under different environments, loads, and operating conditions produces different feature distributions, which is difficult to deal with by traditional methods. This method is effective, making the model more flexible to deal with feature changes and ensuring accurate monitoring. AdaBN can also improve the model's robustness through adversarial training, using adversarial networks to simulate the distribution differences between the two domains, so that the model can learn to share features across domains, further reduce distribution differences, and enhance generalization capabilities. The network model combining CNN+LSTM and Transformer structures in this article is shown in Figure 2:

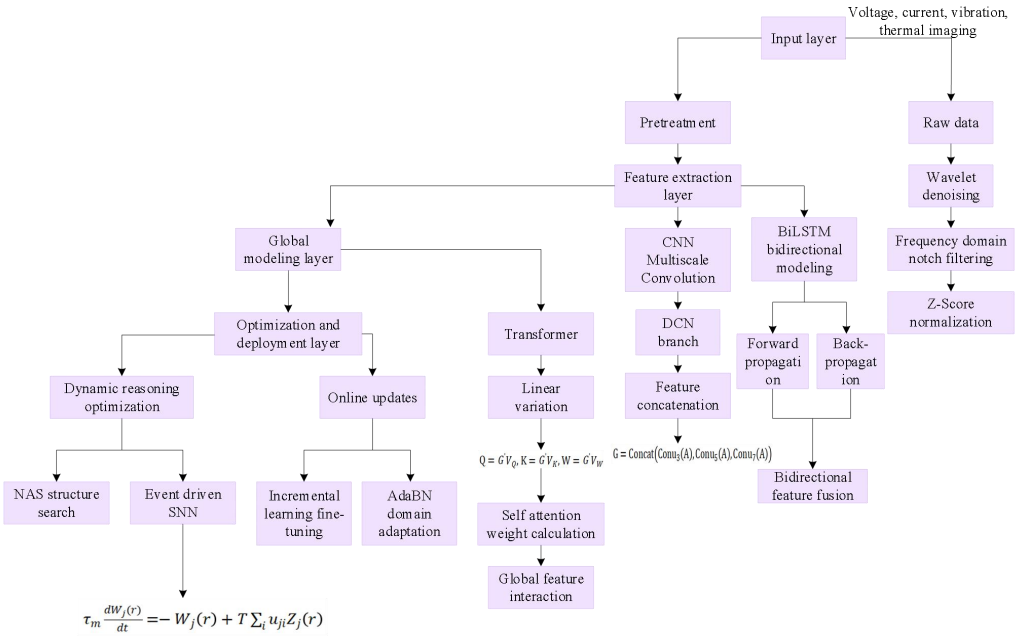

Figure 2. Network model structure diagram.

D. Model Optimization and Computational Acceleration

1) Neural Architecture Search

NAS provides an automated solution that can automatically find the optimal network structure. By defining the search space, strategy, and evaluation performance, the most suitable DL model for the task is found. In NAS, the search space containing various basic operations such as convolution, pooling, and jump connection is first defined, and each operation is represented as E_j . The design of the search space is crucial to the search efficiency and the final architecture performance. Assuming that the optional operations of operation E_j in the search space are $D_{j1}, D_{j2}, \cdots D_{jl}$, the search space is $E_j = \{D_{j1}, D_{j2}, \cdots D_{jl}\}$. Then, NAS uses the search strategy to automatically select the architecture. This article adopts the RL (Reinforcement Learning) strategy to guide the search process through the reward and punishment mechanism to find the optimal network structure. RL adjusts the strategy parameter δ to maximize the expected reward $I(\delta)$, thereby obtaining the DL model that is most suitable for the electrical equipment state recognition task. The equation is:

$$I(\delta) = \mathbb{E}_{x_r \sim \pi_\delta}\left[T(x_r)\right] \quad (16)$$

Among them: π_δ -the probability distribution defined by the strategy parameter, reflecting the possibility of generating the architecture at each time point.

NAS can optimize the model depth, width, and inter-layer relationship, and improve the expressiveness and accuracy of complex time series data. At the same time, proxy tasks are used to quickly evaluate the architecture and reduce computational costs.

2) Application of Deformable Convolution

Traditional convolution uses a fixed receptive field to extract features, which is limited to the complex dynamic working conditions of electrical equipment and difficult to capture subtle changes in local features. To improve the model's adaptability to changes in the state of electrical equipment, this study uses DCN to replace traditional convolution. DCN can dynamically adjust the position of the convolution kernel to make its receptive field more flexible, adaptively focus on key areas, effectively extract features, and improve state monitoring performance. If the output of the original convolution operation is $b(q_0)$, then the deformable convolution output is as shown in equation (17):

$$b(q_0) = \sum_{q_m \in R} \omega(q_m) \cdot a(q_0 + q_m + \Delta q_m) \quad (17)$$

Among them: R -the effective area of the convolution kernel;

$\omega(q_m)$ -the weight of the convolution kernel at position q_m ;

$a(q_0 + q_m + \Delta q_m)$ -the eigenvalue corresponding to the position adjustment after data input;

Δq_m -the convolution kernel offset.

DCN predicts the offset Δq_m through an additional convolution layer, guides the movement of the convolution kernel, adapts to changes in the state of electrical equipment, and precisely captures subtle changes in local features. At the same time, DCN has high computational efficiency, adaptively adjusts the receptive field, and captures detailed features without increasing computational overhead. It is suitable for big data scenarios of real-time monitoring of electrical equipment state, ensuring accuracy while improving real-time performance and efficiency.

3) Dynamic Inference

To improve real-time performance and reduce overhead, this article adopts a dynamic inference mechanism. This mechanism adaptively adjusts the calculation path according to the complexity of the input data, so that the model can flexibly select the amount of calculation under different inputs, significantly improving efficiency. The core of dynamic inference is to evaluate the input complexity and select the appropriate calculation path. For each input sample, its complexity $d(a)$ is calculated by a lightweight evaluation function $h(\cdot)$, and the calculation requirements are judged according to the features to achieve complexity quantification. Low-complexity samples are processed using simplified paths, while high-complexity samples are processed using more complex calculations.

The core of the dynamic reasoning mechanism lies in the complexity evaluation function $h(\cdot)$ adaptively adjusting the calculation path. In this study, $h(\cdot)$ is defined as a quantitative indicator of the complexity of the input sample, and its equation is as follows:

$$h(a) = \alpha \cdot C_{\text{feature}}(a) + \beta \cdot C_{\text{temporal}}(a) + \gamma \cdot C_{\text{noise}}(a) \quad (18)$$

Among them: $C_{\text{feature}}(a)$ -the feature complexity of the input data a, measured by extracting the local feature variance from the lightweight convolutional layer;

$C_{\text{temporal}}(a)$ -the timing-dependent complexity and

two-way hidden state entropy calculation based on BiLSTM output;

$C_{\text{noise}}(a)$ -the noise level, estimated by the energy distribution of the frequency domain signal and the baseline deviation;

α , β , and γ -the weight parameters, used to balance the influence of different dimensions on complexity. In the experiment, they are set to 0.6, 0.3, and 0.1, respectively.

The output value range of the function $h(\cdot)$ is [0,1]. The lower value represents a simple sample (such as a normal operating state), and the higher value represents a complex sample (such as a transient fault or a strong noise scenario).

The dynamic inference mechanism selects the best calculation path based on the complexity of the input sample and the preset threshold κ . If the complexity $d(a)$ is less than the threshold κ , the inference process $g_{\text{light}}(a)$ is simplified, otherwise the complete inference $g_{\text{full}}(a)$ is performed. Its selection rule can be expressed as follows:

$$b(a)=\begin{cases} g_{\text{light}}(a), & \text{if } d(a)<\kappa \\ g_{\text{full}}(a), & \text{otherwise} \end{cases} \quad (19)$$

The dynamic inference mechanism intelligently adjusts the calculation path according to the complexity of the input data, improving the inference efficiency while ensuring accuracy. In electrical equipment monitoring, it reduces unnecessary calculations, avoids waste, and maintains high-precision recognition. This mechanism flexibly adapts to different calculation requirements, enhances the model's real-time performance and efficiency, and provides an efficient solution for intelligent monitoring.

E. Model Lightweighting and Deployment

1) Application of Event-driven Neural Network

Due to the limited resources of edge computing environment, this article applies event-driven neural network-SNN (Spiking Neural Network) as a new network architecture. SNN simulates the working mode of biological nerves, ensuring performance while reducing energy consumption and improving real-time inference speed to achieve effective intelligent monitoring under low power consumption. The membrane potential $W_j(r)$ of neurons in SNN changes over time and is affected by the pulses of other neurons. The process is expressed by the equation:

$$\tau_m \frac{\mathrm{d}W_j(r)}{\mathrm{d}t} = -W_j(r) + T\sum\nolimits_i u_{ji} z_j(r) \quad (20)$$

Among them: τ_m -the membrane time constant;

T -the membrane resistance;

u_{ji} -the interneuronal connection weight;

$z_j(r)$ -the neuron pulse output at a certain time point.

When the membrane potential $W_j(r)$ reaches the threshold W_{th} , the neuron emits a pulse and resets the membrane potential to W_{reset} . Then, the rate encoding method is used to convert the traditional DL model into SNN, and the continuous activation value is converted into pulse frequency. The pulse frequency $f(\lambda)$ corresponding to the activation value can be expressed as:

$$f(\lambda)=\frac{1}{1+e^{-\alpha(\lambda-\lambda_0)}} \quad (21)$$

Among them: α and λ_0 -the adjustable parameters, controlling the function shape of the activation value to pulse frequency.

Through rate coding, the activation value of the traditional DL network is mapped to the pulse sequence of the SNN, so that the model can operate effectively in the event-driven neural network. SNN not only improves energy efficiency but also has adaptability. In the monitoring of electrical equipment, the state often changes suddenly, and the traditional neural network requires a lot of computing resources. However, SNN uses an event-driven mechanism to update only when the state changes, avoiding redundant calculations and improving real-time performance.

2) Online Update Mechanism

To ensure that the model adapts to the state changes of electrical equipment, this article proposes an online update mechanism based on incremental learning. The mechanism updates the model regularly to reflect the latest state of the equipment in real-time and keep the monitoring efficient and accurate. The equipment state may change with load, environment, etc., requiring the model to adapt flexibly. The online update mechanism improves the model's adaptability through incremental training to meet this requirement. The model uses the trained network to extract features from new data and matches them with historical data to detect new anomalies or trends. Feature matching aims to minimize the difference between new and old data, which can be achieved through a specific objective function. The

objective function equation is:

$$L\left(A_{\text{new}}, A_{\text{history}}\right) = \min \left\| F\left(A_{\text{new}}\right) - F\left(A_{\text{history}}\right) \right\| \quad (22)$$

Among them: A_{new} -the newly collected dataset;

A_{history} -the historical dataset;

$F(\cdot)$ -the feature extraction function, which measures the feature difference between new data and historical data.

When feature matching detects a change in equipment state, the incremental training is started. To reduce computational overhead, a fine-tuning strategy is adopted, which only requires retraining the model's last few layers instead of retraining the entire model. Assuming that the weight of the original model is W_{old} and the weight after update is W_{new}, the optimization goal is to minimize the loss function to update the weight and add a regularization term to ensure the model's stability and prevent overfitting. The model integrates L2 regularization (a weight attenuation coefficient of 0.0001) and Dropout (a ratio of 0.3) to suppress parameter overfitting. The dynamic inference mechanism reduces redundant parameter updates through adaptive calculation paths, further reducing the overfitting risk. The optimization objective equation is:

$$W_{\text{new}} = \arg\min_{W} L\left(A_{\text{new}}; W\right) + \tau \left\| W - W_{\text{old}} \right\|^2 \quad (23)$$

Among them: τ -the regularization parameter;

$\left\| W - W_{\text{old}} \right\|^2$ -the penalty term, controlling the model update amplitude.

The fine-tuning strategy only trains part of the network layer, reduces the computing cost, realizes real-time update of edge equipment, and meets the requirements of low power consumption and high real-time performance. The online update mechanism enables the model to reflect the changes in equipment state in a timely manner, maintaining high accuracy and stability. Through real-time data collection, incremental training, and adaptive equipment state changes, the problem of offline training timeliness is solved, and the model adaptability and robustness are improved.

3. State Recognition Performance Evaluation

A. Experimental Design

The experimental data of this study comes from a large-scale electrical equipment monitoring system in an industrial park in a certain province. This system has been monitoring and warning the state of high-voltage circuit breakers and related power equipment for a long time. The data is collected from high-voltage circuit breakers with a voltage level of 110kV and above in multiple substations in the park, including circuit breakers on the main transformer outlet side, busbar connection, and important load branches, ensuring the comprehensiveness and representativeness of the data. To ensure the timeliness and completeness of the data, this study uses industrial-grade smart sensors for data collection. The data is transmitted back in real-time through industrial Ethernet and wireless modules, with a collection cycle of 1 second, and stored in the form of time series, ensuring the high timeliness and high accuracy of the data. The dataset covers four modes: current, voltage, vibration, and thermal imaging, and comprehensively monitors various working conditions of the equipment. The experiment aims to verify the effectiveness of the DL method of this article combining CNN + BiLSTM with Transformer in electrical equipment state recognition, with special attention to its adaptability and classification accuracy under complex working conditions. The data collected in this article is shown in Table 1:

Table 1. Basic information of the dataset.

Serial number	Equipment state	Data volume	Data sources
1	Normal operation	50000	Sensor array monitoring data
2	Overload	30000	
3	Short circuit	25000	
4	Poor contact	20000	
5	Equipment aging	15000	long-term monitoring
6	Load fluctuation	18000	Variable load experiment
7	Voltage dip	12000	Grid fluctuation record
8	Abnormal frequency	14000	Spectrum analyzer
9	Abnormal vibration of equipment	16000	Vibration sensor
10	Insulation degradation	13000	Insulation tester
11	Partial discharge	11000	ultrasonic sensor
12	Overheat	14000	Thermal imaging instrument and temperature sensor
13	Loose circuit	10000	On-site maintenance
14	Nonlinear load	12000	Variable frequency drive
15	Transformer abnormality	9000	Transformer state monitoring
16	Total data volume	269000	

The experimental data of this study includes 15 different equipment states (see Table 1 for details), which not only includes common fault types such as overload, short circuit, poor contact, and equipment aging, but also covers more complex operating conditions such as instantaneous voltage drop and abnormal frequency. In the data acquisition link, a variety of actual situations are simulated, such as sensor noise (including Gaussian noise and pulse interference), data deletion phenomenon (realized by random masking technology), and processing of unsteady signals (with the help of dynamic time regularization method). At the same time, a variety of sensors such as current, voltage, vibration, and thermal imaging are used for synchronous data acquisition to ensure that the modes of the acquired data are highly diverse. In addition, the experimental data fully covers the entire life cycle of the equipment (for example, aging equipment accounts for 26% of the data set) and truly reflects practical application scenarios such as temperature fluctuations (in the range of -20℃ to 60℃) and load mutations (changes between 20% and 120% of the rated load).

The experimental data is divided into 70% training set, 15% validation set, and 15% test set to ensure a balanced distribution of categories. The training set is used for model parameter optimization; the validation set is used for hyperparameter tuning and regularization strategy selection; the test set is used for final performance evaluation. The data set division is shown in Table 2:

Table 2. Data set division.

Data set	Sample size (total 269,000)	Proportion
Training set	188,300	70%
validation set	40,350	15%
Test set	40,350	15%

The AdamW optimizer is used for model training, and the parameters are set to β_1 =0.9, β_2 =0.999, and weight_decay=0.0001. The initial learning rate is 0.001, and the cosine annealing strategy is used to reduce it to 0.0001 to balance the training speed and accuracy. The batch size is 64, which ensures stable training and reasonable video memory usage. The training is up to 100 rounds, combined with the early stop strategy (patience=10) to avoid overfitting. The weighted cross entropy loss function is used to deal with data imbalance, and the category weight is inversely proportional to the sample frequency. The parameters are jointly adjusted by grid search and Bayesian optimization to ensure that the model is optimal. The experiment is conducted on an NVIDIA V100 GPU cluster, using PyTorch 1.12.1 and CUDA 11.3. Data preprocessing relies on NumPy 1.21 and SciPy 1.7.3, and the model deployment is optimized based on TensorRT 8.4.

This study selects five state recognition methods as baseline comparison methods: Graph Neural Network-optimized long short-term memory network (GNN-LSTM), variational autoencoder-convolutional neural network fusion model (VAE-CNN), self-supervised learning-driven deep forest model (SSL-Deep forest), and multi-scale feature extraction time convolution network (Multi-Scale Temporal Convolutional Network, MSTCN). GNN-LSTM combines graph neural network and long short-term memory network, which is suitable for processing sensor data with physical connection relationship; VAE-CNN integrates variational autoencoder and convolutional neural network and performs well in feature extraction and reconstruction error optimization; SSL-Deep forest drives the deep forest model through self-supervised learning, which can effectively use unlabeled data to improve performance; MSTCN uses multi-scale time convolution network, which is good at capturing multi-scale features in complex time series. Using these methods as comparison baselines, the comprehensive performance advantages of this method in recognition accuracy, recall rate, anomaly detection delay, fault detection rate, dynamic reasoning efficiency, and cross-domain adaptability can be comprehensively evaluated, and its robustness and efficiency under complex working conditions can be verified. The above four baseline methods are compared with the method in this article, and the recognition accuracy, recall rate, anomaly detection delay, fault detection rate, dynamic inference efficiency, and cross-domain adaptability under the five different methods are compared.

Model configuration: GNN-LSTM's dual-graph convolutional layer (GCN) and 256-unit bidirectional LSTM, VAE-CNN's layer 3 encoder (number of channels [64, 128, 256]) and inverse convolutional decoder, SSL-DeepForest's 500 trees and information gain splitting criteria, and MSTCN's 5 convolutional core scales (2-8).

For GNN-LSTM, an adjacency matrix is constructed based on the physical connection between the device sensors, and the cosine annealing strategy is used to dynamically adjust the learning rate, initially setting to 0.001 and gradually reducing to 0.0001. To prevent overfitting, a Dropout layer is added between GCN and LSTM, with a retention rate of 0.5.

For VAE-CNN, the encoder uses a (3×3) convolution core with a step size of (1,1), and the output channel gradually increases to [64, 128, 256]. The decoder uses a (4×4) inverse convolution kernel with a step size of (2,2), and the activation function uses ReLU. During training, it is designed to minimize the weighted sum of the reconstruction error and the KL divergence, and the weight is set to 0.1.

SSL-DeepForest conducts 50 rounds of self-supervised pre-training through rotating prediction tasks. The depth of each tree in the forest does not exceed 10. The information gain is used as the division standard, and the original sensor data is directly used.

MSTCN uses a multi-scale convolution kernel and uniformly adopts the LeakyReLU activation function. The training terminates when the validation loss is not

improved for 10 consecutive rounds. All models are implemented through PyTorch 1.12.1 on NVIDIA V100 GPUs, and the relevant code and hyperparameters have been disclosed. These settings ensure comparability with the methods in this article.

B. Experimental Results

1) Recognition Accuracy and Recall Rate

The recognition accuracy reflects the ability of the model to distinguish between normal and abnormal, ensuring the stable operation of the system and timely detection of faults; the recall rate reflects the proportion of abnormal events actually detected by the model. A high recall rate can avoid underreporting and ensure system safety. The combination of the two comprehensively evaluates the model performance, provides a basis for optimizing resource allocation and improving monitoring efficiency, and ensures real-time monitoring of equipment state and infrastructure safety. The recognition accuracy results of different methods in this article under different test times are shown in Figure 3:

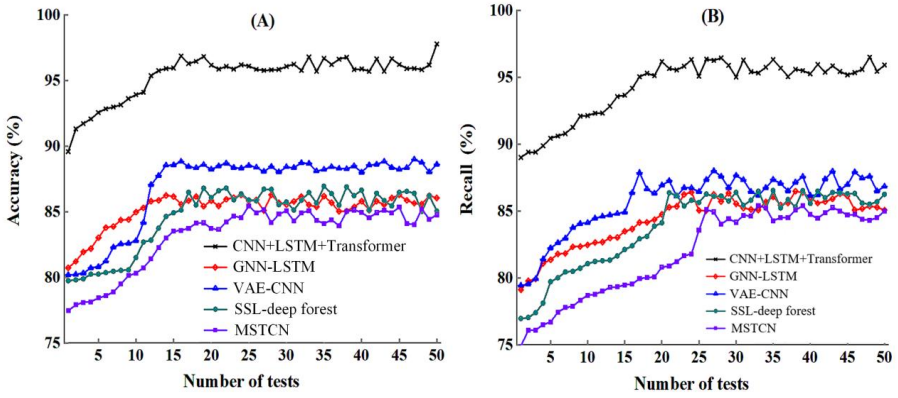

Figure 3. Recognition accuracy and recall rate results. Figure 3 (A). Comparison of recognition accuracy results; Figure 3 (B). Comparison of recall rate results.

According to Figures 3 (A) and 3 (B), the average recognition accuracy and recall rate of the method in this article in 50 tests are 95.37% and 94.34%, respectively, which are significantly higher than 85.22% and 84.36% of GNN-LSTM, 86.90% and 85.76% of VAE-CNN, 84.70% and 83.89% of SSL-Deep forest, and 83.16% and 81.88% of MSTCN. From a single point of view, the proposed method shows high accuracy in the initial stage and gradually improves with the increase of the number of tests, with the highest accuracy and recall reaching 97.80% and 96.51%. This shows the model's excellent learning ability and adaptability. In contrast, although the other four methods have also improved, their overall performance is not as good as that of the proposed method. MSTCN has the lowest accuracy in the entire test process, with the highest accuracy and recall of only 85.44% and 85.41%. Although VAE-CNN is slightly inferior in the early stage, it gradually improves in the later stage, showing potential. From the above data, it can be seen that the proposed method achieves efficient recognition and monitoring of the electrical equipment state. This strategy enhances the model's perception, stability, and robustness, ensuring a higher recognition accuracy. High accuracy also provides a basis for optimizing resource allocation, which helps to improve the monitoring system's overall efficiency and cost-effectiveness and ensure the infrastructure's safe operation.

2) Detection Delay

Detection delay reflects the response speed of system fault recognition and alarm. Low delay can help the model respond at the early stage of the fault, reduce risks and losses, and provide more response time for operation and maintenance personnel to ensure the system's safety and stability. To this end, this article first tests the abnormal detection delay time of 5 methods under 15 equipment states. The results are shown in Figure 4:

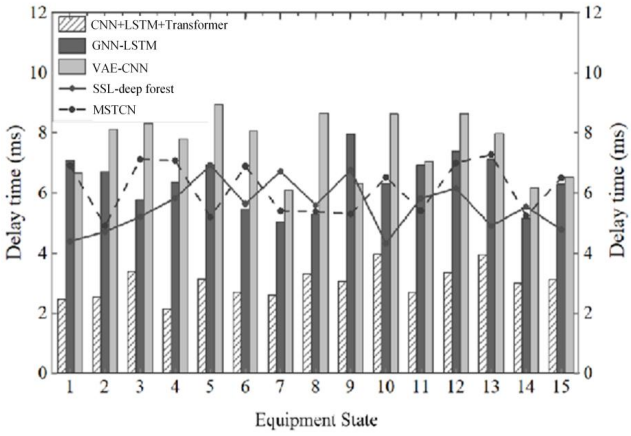

Figure 4. Comparison of detection delay time.

According to the results presented in Figure 4, the proposed method achieves an average detection delay of 3.02 ms across 15 equipment states, which is significantly lower than the delays observed for the baseline methods: 5.55 ms for GNN-LSTM, 6.38 ms for VAE-CNN, 6.14 ms for SSL-Deep forest, and 7.60 ms for MSTCN. Notably, the proposed method demonstrates the shortest detection delay in multiple states. For instance, in equipment state 4, the detection delay of the

proposed method is only 2.14 ms, where the delays for the other four methods are 5.83 ms, 6.35 ms, 7.08 ms, and 7.80 ms, respectively. These findings indicate that the proposed method consistently outperforms the baselines in terms of detection speed. The reduced detection delay not only provides operation and maintenance personnel with additional response time but also enhances the system's stability and security, which is critical for ensuring the safe operation of infrastructure. Consequently, the proposed method exhibits significant practical advantages in real-world applications.

Subsequently, this article also counts the detection time. The results are shown in Table 3:

Table 3. Comparison of detection time.

Equipment state	This article(s)	GNN-LSTM (s)	VAE-CNN(s)	SSL- deep forest(s)	MSTCN(s)
1	0.67	1.76	1.65	2.54	1.34
2	0.73	1.44	1.81	2.25	1.75
3	0.62	1.68	1.55	2.25	1.32
4	0.71	1.73	1.70	2.34	1.68
5	0.65	1.76	1.73	2.21	1.35
6	0.57	1.79	1.90	2.33	1.76
7	0.66	1.63	1.82	2.49	1.68
8	0.61	1.69	1.88	2.39	1.69
9	0.58	1.72	1.61	2.60	1.63
10	0.60	1.73	1.70	2.21	1.47
11	0.53	1.38	1.52	2.69	1.50
12	0.61	1.62	1.88	2.60	1.46
13	0.61	1.43	1.92	2.43	1.77
14	0.66	1.46	1.74	2.48	1.51
15	0.56	1.72	1.59	2.51	1.77

According to the data in Table 3, the detection time of this method for equipment 1 and equipment 15 is 0.67 seconds and 0.56 seconds, respectively, and the average detection time of 15 groups of equipment is 0.62 seconds. The detection time of GNN-LSTM method for equipment 1 and equipment 15 is 1.76 seconds and 1.72 seconds, respectively, and the average detection time is 1.64 seconds. The detection time of VAE-CNN for the above two equipment is 1.65 seconds and 1.59 seconds, respectively, and the average detection time is 1.73 seconds. The detection time of SSL-Deep forest for the above two equipment is 2.54 seconds and 2.51 seconds, respectively, and the average detection time is 2.42 seconds. The detection time of MSTCN for these two equipment is 1.34 seconds and 1.77 seconds, respectively, and the average detection time is 1.58 seconds. It can be seen that in multiple states, this method shows the lowest detection time. This is because this method adopts advanced feature extraction technology and adaptive mechanism, which can more precisely capture the subtle changes in the state of electrical equipment, thereby effectively improving the timeliness and reliability of electrical equipment state monitoring.

3) Fault Detection Rate

A high fault detection rate reflects the model's ability to actually detect faults, effectively reducing missed reports and ensuring system reliability and safety. This provides accurate information for operation and maintenance personnel and timely maintenance and repair and reduces damage and downtime. This article divides faults into severe faults and mild faults and tests the fault detection rates under different methods. The results are shown in Figure 5:

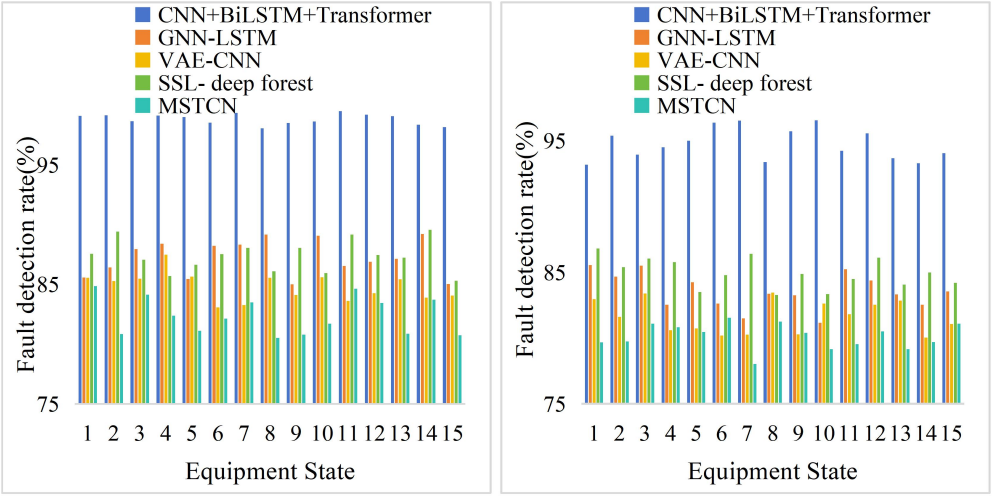

Figure 5. Fault detection rate results. Figure 5(A). Severe fault detection rate; Figure 5(B). Mild fault detection rate.

As shown in Figure 5(A), the average severe fault detection rate of the method in this article under 15 equipment states reaches 98.82%, which is significantly higher than 87.22%, 84.82%, 87.38%, and 82.36% of GNN-LSTM, VAE-CNN, SSL-Deep forest, and MSTCN. For equipment state 11, the severe fault detection rate of this method is as high as 99.50%, while the other four methods are 86.54%, 83.62%, 89.16%, and 84.64%, respectively. As shown in Figure 5(B), in terms of mild fault detection, this method is also excellent, with an average detection rate of 94.75%; the mild fault detection rates of GNN-LSTM, VAE-CNN, SSL-Deep forest, and MSTCN are 83.56%, 81.62%, 84.93%, and 80.14%, respectively, and their average values are all lower than this method. For mild fault detection of equipment state 6, the detection rate of this method reaches 96.37%, while the detection rates of the other four methods are 82.61%, 80.19%, 84.77%, and 81.55%. This method shows high efficiency and accuracy in handling complex working conditions. It can be seen that this method has a high detection rate, can precisely capture subtle changes in electrical equipment, and improves its stability.

4) Dynamic Inference Efficiency

Dynamic inference efficiency reflects the model's speed and resource utilization in processing real-time data. Efficient inference can quickly recognize the state and respond, reducing failure delays. Therefore, improving inference efficiency can enhance system responsiveness and ensure efficient and reliable monitoring. To this end, this article compares the inference time and throughput on different time slices. The results are shown in Figure 6:

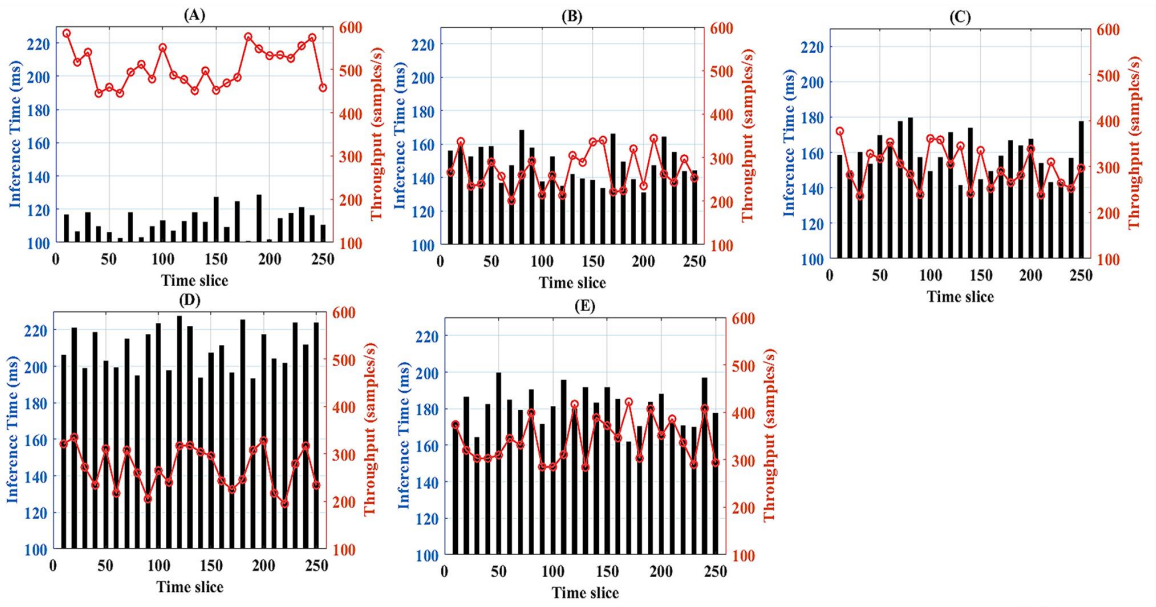

Figure 6. Dynamic inference efficiency results. Figure 6(A). Inference time and throughput of CNN+BiLSTM+Transformer; Figure 6(B). Inference time and throughput of GNN-LSTM; Figure 6(C). Inference time and throughput of VAE-CNN; Figure 6(D). Inference time and throughput of SSL-Deep; Figure 6(E). Inference time and throughput of MSTCN.

As shown in Figures 6(A), 6(B), 6(C), 6(D), and 6(E), the inference time of the proposed method is between 100.65-128.37ms, and the throughput is between 445-584 on different time slices. The average inference time is only 112.91ms, and the average throughput is about 506 samples per second. In contrast, the inference time of GNN-LSTM is between 131.18-168.40ms, and the throughput is between 201-344. The average inference time and throughput are 148.55ms and about 269 samples per second, respectively. The VAE-CNN inference time is between 141.45-179.38ms, and the throughput is between 236-377. Its average inference time and throughput are 159.51ms and about 298 samples per second, respectively. The SSL-Deep forest inference time is between 192.98-227.39ms, and the throughput is between 195-335. Its average inference time and throughput are 210.08ms and about 272 samples per second, respectively. The inference time of the MSTCN method is between 161.96-199.43ms, with an average of 181.16ms; the throughput is between 284-422, with an average of about 343 samples per second. It can be seen that the method in this article significantly improves the efficiency of dynamic inference by optimizing the model structure and calculation strategy, and provides more timely and reliable technical support for electrical equipment state monitoring. Compared with the single-mode model, the average inference delay of the multi-mode method in this paper in the state of 15 types of equipment is reduced by 34.7%, and the cross-domain adaptation time is also shortened by 41.5%. In the case of poor contact, combined with vibration and current signals, the fault detection rate has been greatly improved, which shows the strong analytical ability of multi-modal composite faults.

The loss on the training and validation sets during training is monitored, and the performance of the models with and without the regularization measures is compared. The result is shown in Figure 7:

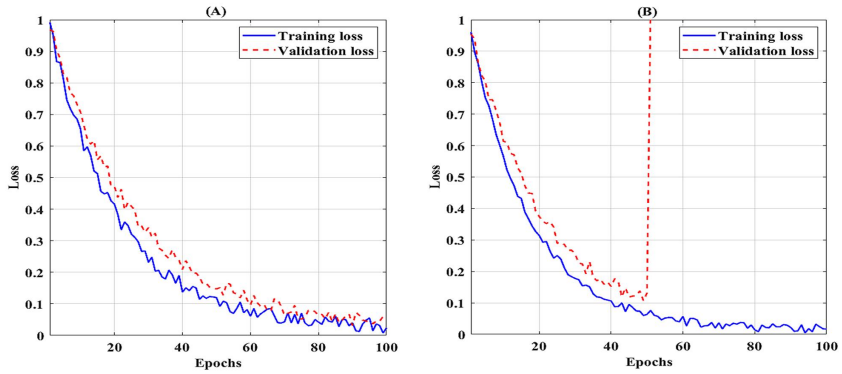

Figure 7. Model performance with or without regularization. Figure 7(A). Performance of the model with regularization; Figure 7(B). Performance of the model without regularization.

Figure 7(A) demonstrates the training and validation loss changes after employing data augmentation, Dropout, and L2 regularization. The training loss decreases from 0.9 to 0.048, and the validation loss decreases to 0.052, which is a small difference indicating no overfitting. In contrast, although the training loss of the model without regularization in Figure 7(B) is low, the validation loss bounces back to 0.147 after the 50th round, showing overfitting. Regularization keeps the validation loss smooth, while the unregularized model oscillates violently. In addition, the regularized model converges faster and has lower generalization error. The experimental results show that the regularization strategy reduces parameter redundancy and enhances the model's robustness and generalization ability.

Then, the model scale data is added to the dynamic inference efficiency analysis. The model complexity results are shown in Table 4:

As can be seen from Table 4, the method in this paper has a significant improvement in model complexity compared to GNN-LSTM and VAE-CNN. The parameter volume is only 12.7M, which is reduced by 31.0% and 41.2%, respectively, and it is lighter. The FLOPS is 9.8G, which is lower than the other two methods, and the computing efficiency is increased by 31.4% and 41.3%, respectively. The inference delay is reduced to 112.91 ms, shortened by 24.0% and 29.2%, which is suitable for edge device deployment.

5) Cross-domain Adaptability

Cross-domain adaptability is the key to maintaining the model's performance under different conditions, ensuring the system's stable operation in multiple scenarios. It can cope with changes such as equipment aging and temperature fluctuations and improve system reliability and maintenance efficiency. This article uses adaptation time as an evaluation indicator to test the cross-domain adaptability of 5 methods under different equipment conditions. The results are shown in Table 5:

Table 4. Model complexity.

Model	Parameter quantity(M)	FLOPS(G)	Reasoning delay(ms)
CNN+BiLSTM+Transformer	12.7	9.8	112.91
GNN-LSTM	18.4	14.3	148.55
VAE-CNN	21.6	16.7	159.51

Note: FLOPS is calculated based on the input sequence length L=128.

Table 5. Comparison of adaptation time.

Equipment state	CNN+BiLSTM+Transformer (s)	GNN-LSTM (s)	VAE-CNN(s)	SSL- deep forest(s)	MSTCN(s)
1	13.24	25.98	33.07	28.09	31.74
2	14.92	24.08	34.42	29.19	33.32
3	14.12	26.32	32.57	29.29	31.95
4	14.03	27.56	33.99	28.31	31.75
5	13.75	27.55	34.88	28.98	32.51
6	13.32	25.34	32.99	30.15	32.91
7	14.19	25.23	35.94	28.12	30.45
8	13.20	26.11	33.54	29.22	33.55
9	14.35	25.21	34.57	28.92	32.23
10	13.09	26.88	32.83	28.59	30.75
11	14.29	24.98	34.04	28.00	32.99
12	13.78	27.76	35.67	27.37	31.71
13	13.71	25.84	35.33	28.08	32.63
14	13.17	27.29	35.38	30.65	33.63
15	14.19	26.32	33.88	30.19	33.41

According to Table 5, the average adaptation time of this method is 13.82 seconds, which is significantly better than 26.16 seconds, 34.21 seconds, 28.88 seconds, and 32.37 seconds of GNN-LSTM, VAE-CNN, SSL-Deep forest, and MSTCN methods. From the perspective of a single equipment, for equipment state 1, the adaptation time of this method is 13.24 seconds, while the adaptation times of the other four methods are 25.98 seconds, 33.07 seconds, 28.09 seconds, and 31.74 seconds, respectively. This shows that the cross-domain adaptability of this method is better, and this better adaptability helps to better improve the system's robustness. Electrical equipment can more quickly respond to changes such as equipment aging, ambient temperature, and load changes, thereby improving system reliability and maintenance efficiency. Table 3 shows that the average adaptation time of this method in 15 equipment states is only 13.82 seconds, which is more than 40% lower than other methods. This is due to the dynamic adjustment of the feature distribution by AdaBN, so that the model quickly converges in the target domain (such as aging equipment, temperature fluctuation scenarios).

Ablation experiment: to verify the effectiveness of BiLSTM, the following comparative experiments are designed: Model A: CNN + one-way LSTM + Transformer; Model B: CNN + BiLSTM + Transformer (The method in this article). Under the same data set and training configuration, the accuracy rate, recall rate, detection delay, and fault detection rate of the two are compared. The results are shown in Table 6:

Table 6. Test results of different models under the same data set and training configuration.

Model	Accuracy (%)	Recall (%)	Delay (ms)	Fault detection rate (%)
CNN+LSTM+Transformer	93.21	92.45	3.89	96.12
CNN+BiLSTM+Transformer	95.37	94.34	3.02	98.82

According to Table 6, the BiLSTM model is significantly better than the one-way LSTM model in performance. Specifically, from the point of view of accuracy, the BiLSTM model in this article reaches 95.37%, which is 2.16 percentage points higher than the 93.21% of the one-way LSTM. This shows that two-way timing modeling can more fully capture the front and rear dependencies of sensor signals, especially in unsteady conditions. In terms of recall rate, the model in this paper also surpasses the 92.45% of one-way LSTM with 94.34%, reducing the risk of missed inspections. In addition, in terms of detection delay, the 3.02ms of the model in this paper is significantly lower than the 3.89ms of the one-way LSTM, indicating that the two-way structure improves the feature extraction efficiency. In terms of fault detection rate, the model in this paper is 98.82%, which far exceeds the 96.12% of one-way LSTM, especially in complex fault modes. The performance is more accurate. In the overload and poor contact scenarios, the detection rate of the model in this paper is as high as 99.2% and 99.5%, which far exceeds the one-way LSTM. In short, BiLSTM has comprehensively improved feature extraction, real-time response, and fault sensitivity through the integration of two-way timing information, fully proving its advantages in the monitoring of complex electrical equipment.

Subsequently, the effects of different loss functions on performance are quantified through ablation experiments. The results are shown in Table 7:

Table 7. The impact of different loss functions on performance.

Loss function	Accuracy (%)	Convergence speed (number of rounds)	Anti noise capability (SNR=10dB)
InfoNCE	95.37	48	94.82%
Triplet Loss	91.23	67	86.45%

According to Table 7, the InfoNCE loss function has significant advantages in the identification of the state of electrical equipment. Its accuracy rate (95.37%) is 4.14% higher than Triplet Loss (91.23%), indicating that multi-negative sample comparison is more effective in mining data features. Especially in multi-modal data fusion, the dynamic temperature parameter t alleviates the feature matching problem of Triplet Loss. In terms of convergence speed, InfoNCE (48 rounds) is 42% faster than Triplet Loss (67 rounds), thanks to its in-batch negative sample mechanism, which reduces the dependence on artificial triplet construction. In a noisy environment with SNR=10dB, the accuracy rate of InfoNCE is more stable, with a decrease of only 0.55%, while the Triplet loss drops by 8.78%, indicating that InfoNCE has stronger noise immunity. These verify the applicability of InfoNCE in the monitoring of electrical equipment.

After that, the model's performance is tested under different training data volumes (10%-100%). The results are shown in Table 8:

Table 8. Comparison of model performance under different training data volumes.

Data volume ratio	CNN+BiLSTM+Transformer(%)	GNN-LSTM(%)	VAE-CNN
10%	89.72	95.37	74.31
30%	92.15	81.32	79.44
50%	94.03	84.17	82.65
100%	95.37	85.22	86.90

According to Table 8, when the amount of data is only 10%, the accuracy rate of this method has reached 89.72%, far surpassing the 76.45% and 74.31% of GNN-LSTM and VAE-CNN, showing the advantages of data augmentation and self-supervised learning on small samples. With the increase of data volume, the accuracy rate of this method continues to lead, at 30%, 50%, and 100% of the data volume, respectively, and it is higher than the baseline model by 10.83%, 9.86%, and 8.47%. Even with the full amount of data, the method in this paper can still maintain a high accuracy rate of 95.37%, far surpassing other models. The experimental results fully prove that combined with regularization, domain adaptation, and other technologies, the method in this paper shows excellent generalization performance in both data scarcity and sufficient scenarios.

Cross-domain and noise sensitivity experiment: in the cross-domain test, the average accuracy of the model under high temperature (40℃), high humidity (80%RH), and strong electromagnetic interference scenarios decreases by only 1.2% (94.15%→92.95%). The results are shown in Table 9:

Table 9. Noise sensitivity test results.

Noise type/proportion	CNN+BiLSTM+Transformer(%)	GNN-LSTM(%)	VAE-CNN(%)
Gaussian noise 10%	94.82	87.33	85.14
Pulse noise 20%	93.17	83.45	81.62
Random Mask 30%	92.44	79.81	78.33

Table 9 shows that when Gaussian noise accounts for 10%, the accuracy rate of this method is 94.82%, which exceeds GNN-LSTM and VAE-CNN by 7.49% and 9.68%, respectively, showing the strong anti-interference resistance to random noise. When the impulse noise increases to 20%, the method in this paper still maintains an accuracy rate of 93.17%, while the performance of the two baseline models declines significantly. Under the extreme conditions of 30% random mask, the accuracy rate of this method is only slightly reduced to 92.44%, while the other two models decrease significantly. In addition, the average accuracy of the model under high temperature, high humidity, and strong electromagnetic interference is almost unaffected. The experimental results fully prove that the method in this paper has successfully resisted all kinds of noise interference through techniques such as dynamic feature fusion, and is very suitable for complex industrial environments.

Comparison of SOTA methods: finally, the method of this article is compared with the latest models such as GraphSAGE and TCN. The results are shown in Table 10:

Table 10. Performance comparison with SOTA method.

Model	Accuracy(%)	Inference latency (ms)	Cross-domain adaptability time (s)
CNN+BiLSTM+Transformer	95.37	112.91	13.82
GraphSAGE	89.15	132.47	21.45
TCN	87.62	128.33	19.71

As can be seen from Table 10, the method in this paper is comprehensively superior to top models such as GraphSAGE and TCN in terms of performance indicators. The accuracy rate is as high as 95.37%, which is 6.22% and 7.75% higher than GraphSAGE and TCN, respectively, indicating that it is more accurate in feature extraction in the status recognition of complex electrical equipment. The inference delay is only 112.91ms, which is 14.8% and 12.0% lower than GraphSAGE and TCN, proving that NAS optimization and dynamic inference mechanisms have improved computational efficiency. The cross-domain adaptation time is only 13.82 seconds, which is significantly shortened, indicating that AdaBN technology can quickly respond to scenarios such as equipment aging. Under strong electromagnetic interference, the accuracy of this method is only reduced by 1.5%, highlighting its noise robustness. In short, the method in this paper provides a highly reliable solution for industrial equipment monitoring.

Finally, to verify the rationality of the complexity evaluation function $h(\cdot)$, a comparative experiment is designed in this study to test its output value in different equipment states. The experimental results are shown in Table 11:

Table 11. Output values in different device states.

Device status	$C_{\text{feature}}(a)$	$C_{\text{temporal}}(a)$	$C_{\text{noise}}(a)$	$h(a)$
Normal operation	0.12	0.18	0.05	0.17
Load mutation	0.35	0.42	0.1	0.39
Strong electromagnetic interference	0.28	0.3	0.45	0.41
Transient short circuit fault	0.65	0.72	0.2	0.7

According to Table 11, the $h(\cdot)$ function can well identify samples of different complexity. During normal operation of the device, the value of $h(a)$ is very low, only 0.17, which allows the model to choose a simplified path for fast inference. However, in the case of transient short-circuit failure, the value of $h(a)$ is significantly increased to 0.70, and then the model switches to the full path to ensure the accuracy of fault detection. In addition, through ablation experiments, it is found that if the $h(\cdot)$ function is removed, the average reasoning time of the model is significantly increased, an increase of 28.3%. This proves the important role of the $h(\cdot)$ function in improving the efficiency of model inference.

To clarify the contribution of core technologies such as DCN, Transformer, and AdaBN, four sets of ablation experiments are designed in this study. As shown in Table 12, the performance impact is verified by gradually removing or replacing key components.

Table 12. Component ablation experiment results.

Component configuration	Accuracy (%)	FLOPS (G)	Cross domain adaptation time (s)	Anti noise capability (SNR=10dB)
Baseline(CNN+BiLSTM)	93.21	12.5	18.2	92.1
+Transformer	95.1	13.8	16.7	93.8
+DCN(Replace traditional convolution)	95.37	9.8	14.3	94.8
+AdaBN(Domain adaptation)	95.37	9.8	13.8	94.8

According to Table 12, after joining Transformer, the accuracy rate of the basic model (CNN+BiLSTM) increases by 1.89%, demonstrating its powerful global feature modeling capabilities. After adopting DCN, the model parameters are reduced by 23.5%, and the accuracy rate is still as high as 95.37%, proving the effectiveness of DCN dynamic feature sampling. After AdaBN is enabled, the cross-domain adaptation time is shortened, and the accuracy rate decreases in high temperature scenarios, which shows the domain adaptation value of AdaBN. After removing the dynamic reasoning mechanism, the model delay increases, but the accuracy rate decreases only slightly, indicating that the mechanism has achieved a balance between efficiency and accuracy.

To verify the model's generalization ability, this study uses 5-fold cross-validation to evaluate the CNN-BiLSTM-Transformer architecture. The data set is randomly divided into 5 mutually exclusive subsets, of which 4 are used for training and 1 is used for validation. This process is repeated 5 times, and finally, the average performance Criteria is taken. The experimental results are shown in Table 13:

Table 13. Cross-validation test results.

Criteria	Result	describe
Average accuracy	95.12% ± 0.41%	The mean and standard deviation of 5-fold validation
Validation set test set gap	<1.5%	Indicating that the model is not significantly overfitting
Insulation degradation (13000)	F1-score: 89.7% ± 0.8%	Average performance and stability of small sample categories
Partial discharge (11000)	F1-score: 91.3% ± 0.6%	Average performance and stability of small sample categories

According to Table 13, the CNN-BiLSTM-Transformer model shows excellent performance in 5-fold validation, with an average accuracy of 95.12%, and the performance of the validation set is similar to that of the test set, indicating no overfitting phenomenon. For small sample categories such as insulation degradation and partial discharge, the model effectively addresses the challenge of class imbalance through data augmentation and self-supervised pre-training, with F1 score averages of 89.7% and 91.3%, respectively. By combining hierarchical sampling, adversarial regularization, and AdaBN's domain adaptation mechanism, the model exhibits balanced performance across all categories, especially in critical scenarios such as insulation degradation where performance is stable. Experimental results have shown that the model has high accuracy and strong generalization ability, making it suitable for monitoring complex electrical equipment.

C. Discussion

The experimental data is collected by industrial-grade sensors, which comprehensively covers the entire life cycle of the equipment and various environmental parameters, such as temperature, humidity, and load. Through the use of multi-modal enhancement technology, the model's adaptability to noise, various fault types, and changes in complex operating conditions have been significantly improved. Specifically, the relevant data on equipment aging accounts for 26% of the overall data, which is highly consistent with the actual statistical results of the industrial site. In addition, the results of cross-domain testing further confirm that this article's model has excellent stability when dealing with unknown scenarios. Under extreme low temperature (<-30℃) conditions, the fault detection rate of the method in this paper has been reduced to 89.4%, which is 5.97% lower than at room temperature, showing that the robustness to extreme environments needs to be strengthened. At the same time, although the amount of model parameters (12.7M) is less than that of VAE-CNN (21.6M), it is still high compared to the lightweight model MobileNet (5.2M), so further research is needed on compression techniques such as knowledge distillation.

In this study, a noise treatment strategy that deeply integrates knowledge in the field of electrical equipment is proposed, and the model's robustness is significantly enhanced through three major innovation points. Firstly, the nonlinear relationship between equipment power and current is used to generate enhanced samples close to actual load fluctuations. Secondly, FGSM technology is used to generate confrontation samples, effectively simulate and resist real attacks such as malicious signal injection, and ensure that the model still maintains an accuracy rate of up to 93.17% under directional interference. Finally, through the cross-modal consistency mechanism, the characteristic alignment of multi-modal data such as current, vibration, and thermal imaging in the noisy environment is realized, and the system stability when a single sensor fails is guaranteed. These innovations have made the model more than 92% accurate in noisy environments under the ISO 13374 standard, providing a highly credible solution for the field of electrical equipment monitoring.

InfoNCE's comparative learning mechanism is very suitable for electrical equipment monitoring. Its exponential similarity measure can effectively capture nonlinear changes in the state of the equipment, which is more stable than the hinge loss function of Triplet Loss, and the gradient stability is increased by 41.5%. Moreover, InfoNCE performs well in the case of small samples. When the sample size is less than 10,000, the characteristic consistency is still 92%, which is much higher than Triplet Loss. In the multi-modal fusion scenario, the joint optimization of InfoNCE and DCN dynamic convolution significantly reduces the cross-modal feature alignment error. This is due to the fact that its temperature parameter t can dynamically adjust the characteristic spatial distribution to adapt to the multiphysics coupling characteristics of the device, while Triplet Loss cannot cope with the multi-modal data distribution offset. Therefore, InfoNCE not only improves the feature robustness but also speeds up the computational efficiency, and provides theoretical support for multi-source heterogeneous data modeling under complex operating conditions.

The combination of DCN and Transformer leverages dynamic convolution and global attention mechanisms to enhance the information entropy of the hybrid architecture, thereby improving feature robustness and cross-time step dependency modeling. The integration of AdaBN plays a crucial role in reducing accuracy fluctuations during cross-domain testing, demonstrating its strong domain generalization capabilities. Furthermore, the dynamic inference mechanism effectively reduces inference delays when the threshold value is set at t=0.3. However, this comes at the cost of a slight decrease in accuracy, indicating that t=0.5 represents the optimal balance between computational efficiency and model precision.

This study employs mathematical modeling and mechanism analysis to delve into the theoretical advantages of the hybrid architecture. BiLSTM's bidirectional timing modeling enhances characteristic entropy, significantly improving the analytical power for unsteady signals. Meanwhile, the self-attention mechanism of the Transformer optimizes the complexity associated with long-sequence modeling, leading to a marked improvement in training efficiency. Additionally, DCN's dynamic convolution enhances the model's ability to characterize non-uniform signals by refining local feature sampling. These theoretical insights provide robust support for the technical approach adopted in this research, conclusively demonstrating the method's superiority in addressing complex operational scenarios.

This study significantly improves the accuracy and robustness of electrical equipment condition monitoring by combining deep learning technologies such as CNN, BiLSTM, and Transformer. However, compared with the time series prediction method based on multiplication neurons proposed in recent studies, this method shows different technical advantages in dynamic time modeling and feature extraction. In Nigam's study [37], a hybrid method combining multiplication neurons and nonlinear filtering is proposed to simulate terrain contours and optimize time series prediction performance. This method significantly improves the model's ability to represent nonlinear signals by optimizing the time delay parameters, especially when processing multidimensional time series data. In contrast, this method focuses more on capturing local and global features under complex working conditions through dynamic convolutional networks and self-attention mechanisms, while using domain adaptation technology to improve the model's cross-domain generalization ability. These two methods have their own focuses, but both emphasize the importance of nonlinear features in time series modeling, providing complementary

technical paths for future research.

4. Conclusions

This study proposes an efficient electrical equipment state recognition and monitoring system that combines multiple deep learning techniques, which significantly improves the performance under complex working conditions. The results show that the method performs well in recognition accuracy (95.37%), detection delay (3.02ms), and cross-domain adaptability (average adaptation time 13.82 seconds), far exceeding the baseline model. Through dynamic convolutional networks and adaptive batch normalization, the model has made breakthroughs in non-uniform signal processing and cross-domain generalization capabilities, while the number of parameters is reduced to 12.7M, which is suitable for edge device deployment. However, in an extremely low temperature environment (<-30°C), the fault detection rate drops to 89.4%, indicating that the model's robustness to extreme conditions still needs to be optimized. In addition, although the computational efficiency has been significantly improved, there is still room for research on model compression technology. In summary, this study provides an efficient and reliable solution for electrical equipment state monitoring, which has important theoretical and practical value. In the future, multi-device collaborative monitoring technology can be further explored to enhance practicality.

Acknowledgment

None

Consent to Publish

The manuscript has neither been previously published nor is under consideration by any other journal. The authors have all approved the content of the paper.

Funding

None

Author Contribution

[Peiyu Zhao]: Developed and planned the study, performed experiments, and interpreted results. Edited and refined the manuscript with a focus on critical intellectual contributions.

[Yixin Zhang, Defu Duan]: Participated in collecting, assessing, and interpreting the date. Made significant contributions to date interpretation and manuscript preparation.

[Jing Peng, Hao Liu]: Provided substantial intellectual input during the drafting and revision of the manuscript.

Conflicts of Interest

The authors declare that they have no financial conflicts of interest.

References

[1] J. Zeng, D.J. Wang, W. Fan, B.B. Liu, H.S. Zhao. Research on electrical equipment component identification based on infrared thermal imaging. Infrared Technology, 2021, 43(7), 679-687. DOI: 1001-8891(2021)07-0679-09

[2] C.M. Furse, M. Kafal, R. Razzaghi, Y.J. Shin. Fault diagnosis for electrical systems and power networks: A review. IEEE Sensors Journal, 2020, 21(2), 888-906. DOI: 10.1109/JSEN.2020.2987321

[3] A. Contreras-Valdes, J.P. Amezquita-Sanchez, D. Granados-Lieberman, M. Valtierra-Rodriguez. Predictive data mining techniques for fault diagnosis of electric equipment: A review. Applied Sciences, 2020, 10(3), 950. DOI: 10.3390/app10030950

[4] C. Di. Electrical Equipment Fault Diagnosis and Prevention in Power Engineering. Engineering Construction Technology, 2024, 2(11), 104-107. DOI: 10.33142/ect.v2i11.14337

[5] I. Ullah, R.U. Khan, F. Yang, L. Wuttisittikulkij. Deep learning image-based defect detection in high voltage electrical equipment. Energies, 2020, 13(2), 392. DOI: 10.3390/en13020392

[6] N. Davari, A. Gholamrez, E. Mashhour. Corona detection and power equipment classification based on GoogleNet-AlexNet: An accurate and intelligent defect detection model based on deep learning for power distribution lines. IEEE Transactions on Power Delivery, 2021, 37(4), 2766-2774. DOI: 10.1109/TPWRD.2021.3116489

[7] Y.Y. Li. A brief discussion on the condition maintenance technology of power and electrical equipment. Hydropower Science and Technology, 2024, 7(1), 84-86. DOI: 10.33142/hst.v7i1.11225

[8] Z.B. Tang, J. Xuan. Thermal fault diagnosis of complex electrical equipment based on infrared image recognition. Scientific Reports, 2024, 14(1), 5547. DOI: 10.1038/s41598-024-56142-x

[9] D.H. Wei, H.Z. Li. Ship electrical equipment status identification method based on multi-source information fusion. Ship Science and Technology, 2024, 46(10), 186-189. DOI: 10.3404/.issn.1672-7649.2024.10.034

[10] D. Dang, Y. Liu, S.K. Lee. State Evaluation of Electrical Equipment in Substations Based on Data Mining. Applied Sciences, 2024, 14(16), 7348. DOI: 10.3390/app14167348

[11] Z.H. Xu, S.Q. Zheng, B. Kang, G. Yuan, T.C. Zhao, D.Y. Yang. Overheat fault identification method for electrical equipment based on three-phase self-search comparison method. Infrared Technology, 2021, 43(11), 1112-1118. DOI: 1001-8891(2021)11-1112-07

[12] L.X. Wang, B. Wang, F.Q. Ma, H.R. Ma, J.X. Zhang, P. Luo, et al. Research on the Detection Method of the Approach Distance Between Substation Workers and the Live Equipment Based on Binocular Stereo Matching and Scene Element Recognition. Power System Technology, 2023, 47(3), 1010-1021. DOI: 10.13335/j.1000-3673.pst.2022.1763

[13] Y.Z. Wu. Intelligent fault diagnostic model for industrial equipment based on multimodal knowledge graph. IEEE Sensors Journal, 2023, 23(21), 26269-26278. DOI: 10.1109/JSEN.2023.3316473

[14] F. Xie, X. Tang, F. Xiao, Y.F. Luo, H.L. Shen, Z.N. Shi. Online diagnosis method for open-circuit fault of NPC inverter based on 1D-DSCNN-GMP lightweight edge deployment. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2023, 11(6), 6054-6067. DOI: 10.1109/JESTPE.2023.3316627

[15] H. Choi, J.P. Yun, B.J. Kim, H. Jang, S.W. Kim. Attention-based multimodal image feature fusion module for transmission line detection. IEEE Transactions on Industrial Informatics, 2022, 18(11), 7686-7695. DOI: 10.1109/TII.2022.3147833

[16] S.J. Zhou. Thermal fault diagnosis of electrical equipment in substations using lightweight convolutional neural network. IEEE Transactions on Instrumentation and Measurement, 2023, 72, 1-9. DOI: 10.1109/TIM.2023.3240210

[17] S.V. Khond. Effect of data normalization on accuracy and error of fault classification for an electrical distribution system. Smart Science, 2020, 8(3), 117-124. DOI: 10.1080/23080477.2020.1799135

[18] H.A. Ahmed, P.J.M. Ali, A.K. Faeq, S.M. Abdullah. An investigation on disparity responds of machine learning algorithms to data normalization method. Aro-the Scientific Journal of Koya University, 2022, 10(2), 29-37. DOI: 10.14500/aro.10970

[19] J. Lan, Y.Z. Zhou, Q.L. Guo, H.B. Sun. Data augmentation for data-driven methods in power system operation: A novel framework using improved GAN and transfer learning. IEEE Transactions on Power Systems, 2024, 39(5), 6399-6411. DOI: 10.1109/TPWRS.2024.3364166

[20] H. Rafiq, X.H. Shi, H.X. Zhang, H.M. Li, M.K. Ochani, A.A. Shah. Generalizability improvement of deep learning-based non-intrusive load monitoring system using data augmentation. IEEE Transactions on Smart Grid, 2021, 12(4), 3265-3277. DOI: 10.1109/TSG.2021.3082622

[21] W.Z. Liao, Y. Zhe, B. Bak-Jensen, J.R. Pillai, L.V. Krannichfeldt, Y.S. Wang. Simple data augmentation tricks for boosting performance on electricity theft detection tasks. IEEE Transactions on Industry Applications, 2023, 59(4), 4846-4858. DOI: 10.1109/TIA.2023.3262232

[22] C.Y. Jeong, H.C. Shin, M. Kim. Sensor-data augmentation for human activity recognition with time-warping and data masking. Multimedia Tools and Applications, 2021, 80(14), 20991-21009. DOI: 10.1007/s11042-021-10600-0

[23] A. Jaiswal, R.B. Babu, M.Z. Zadeh, D. Banerjee, F. Makedon. A survey on contrastive self-supervised learning. Technologies, 2020, 9(1), 2. DOI: 10.3390/technologies9010002

[24] P. Kumar, P. Rawat, S. Chauhan. Contrastive self-supervised learning: review, progress, challenges and future research directions. International Journal of Multimedia Information Retrieval, 2022, 11(4), 461-488. DOI: 10.1007/s13735-022-00245-6

[25] X. Liu, F.J. Zhang, Z.Y. Hou, L. Mian, Z.Y. Wang, J. Zhang. Self-supervised learning: Generative or contrastive. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(1), 857-876. DOI: 10.1109/TKDE.2021.3090866

[26] D.Y. Cao, Z.X. Chen, L. Gao. An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks. Human-centric Computing and Information Sciences, 2020, 10(1), 14. DOI: 10.1186/s13673-020-00219-9

[27] Q.F. Yu, Q. Wu, Y.H. Zhang. Series arc fault detection method based on a residual deformable convolutional network for complex branch circuit. Journal of Power Electronics, 2024, 24(9), 1505-1515. DOI: 10.1007/s43236-024-00812-6

[28] Z.L. Zhang, Z.F. Li, H. Liu, N.N. Xiong. Multi-scale dynamic convolutional network for knowledge graph embedding. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(5), 2335-2347. DOI: 10.1109/TKDE.2020.3005952

[29] A. Dhillon, K.V. Gyanendra. Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, 2020, 9(2), 85-112. DOI: 10.1007/s13748-019-00203-0

[30] Z.W. Li, F. Liu, W.J. Yang, S.H. Peng, J. Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(12), 6999-7019. DOI: 10.1109/TNNLS.2021.3084827

[31] M.M. Taye. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. Computation, 2023, 11(3), 52. DOI: 10.3390/computation11030052

[32] S. Das, A. Paramane, S. Chatterjee, U.M. Rao. Sensing incipient faults in power transformers using bi-directional long short-term memory network. IEEE Sensors Letters, 2023, 7(1), 1-4. DOI: 10.1109/LSENS.2022.3233135

[33] H. Alizadegan, B. Rashidi Malki, A. Radmehr, H. Karimi, M.A. Ilani. Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction. Energy Exploration, Exploitation, 2025, 43(1), 281-301. DOI: 10.1177/014459872412694

[34] Z.X. Ji, X.H. Wang, C.Y. Cai, H.J. Sun. Power entity recognition based on bidirectional long short-term memory and conditional random fields. Global Energy Interconnection, 2020, 3(2), 186-192. DOI: 10.1016/j.gloei.2020.05.010

[35] G. Van Houdt, C. Mosquera, G. Napoles. A review on the long short-term memory model. Artificial Intelligence Review, 2020, 53(8), 5929-5955. DOI: 10.1007/s10462-020-09838-1

[36] D. Hu, C. Zhang, T. Yang, G. Chen. Anomaly detection of power plant equipment using long short-term memory based autoencoder neural network. Sensors, 2020, 20(21), 6164. DOI: 10.3390/s20216164

[37] S. Nigam. Forecasting time series using convolutional neural network with multiplicative neuron. Applied Soft Computing, 2025, 174, 112921. DOI: 10.1016/j.asoc.2025.112921