# Spatio-Temporal Adaptive Graph Convolutional Networks for Real-Time Power Equipment Monitoring and Fault Detection

Liu Yi*, Zhao Tianle, Zhu Haotian
School of Architectural Intelligence, Jiangsu Vocational Institute of Architectural Technology, Xuzhou, 221008, China
E-mail:Rockliujsviat@126.com

**Keywords:** power equipment monitoring, fault warning, big data analysis, multi-scale feature extraction, spatio-temporal graph convolutional network, parallel computin

**Received:** July 23, 2025

*Traditional methods for monitoring the operational status of power equipment often suffer from limited feature representation and inaccurate dynamic modeling, resulting in low early-fault detection accuracy and elevated false-alarm rates. To address these limitations, this paper proposes a comprehensive, big-data-driven monitoring and fault warning framework, optimized end-to-end from data preprocessing to online alert generation. The system incorporates multi-scale sliding-window feature extraction, weighted trend-deviation quantification, and redundant-feature compression to enhance the representation of degradation signals. A spatio-temporal adaptive graph convolutional network (STAGCN) is employed to jointly capture equipment topology and temporal dependencies. In the warning module, an anomaly discrimination strategy based on confidence-score reconstruction is deployed on a hierarchical, parallel big-data platform, enabling sub-second alerts for critical faults (e.g., line short circuits, insulation degradation). Experimental evaluations reveal that, following feature compression, overall detection accuracy increases from 89.0% to 92.8%, while the false-alarm rate decreases from 4.8% to 3.5%. For typical line short-circuit faults, the proposed system achieves a warning accuracy of 96.5%, with response times under 3.1 seconds under full-load conditions. These results demonstrate that the proposed framework delivers high real-time performance, accuracy, and reliability in complex grid environments.*

*Povzetek: Predlagan sistem za zgodnje opozarjanje na napake v elektroenergetski opremi, ki izboljša natančnost (89,0 % → 92,8 %), zniža lažne alarme (4,8 % → 3,5 %) in zazna kratke stike z 96,5 % v ~3,1 s.*

## 1 Introduction

As a critical unit in ensuring energy transmission and distribution, the operational health of power equipment directly impacts the overall security and stability of the power grid. With the rapid expansion of smart grids [1–2], distributed generation [3–4], and the integration of new loads, the operating environment of power equipment [5–6] has become increasingly multi-source, variable, and complex. Traditional approaches based on periodic inspections and static threshold judgments can no longer meet the requirements for real-time and precise equipment condition control. Consequently, power equipment condition monitoring [7–8] and fault warning [9–10] have become essential means of enhancing power system resilience, holding both high engineering application value and significant theoretical importance.

However, several key technical challenges persist in practical applications. In terms of fault identification [11–12], existing methods typically rely on fixed time-window processing or simple statistical feature extraction, which fail to capture subtle and gradually evolving degradation characteristics during equipment operation, making early fault signals difficult to detect. For dynamic modeling, current time-series analysis methods often neglect spatial structural associations when addressing equipment status changes in complex environments, resulting in the inability to model potential coupling effects between different devices and thus reducing overall fault detection sensitivity. Regarding anomaly detection [13–14] and early warning mechanisms [15–16], most traditional approaches are based on static rules or fixed inference models, which are slow to respond to minor abnormal changes in equipment operation [17–18], leading to high false-alarm and missed-alarm rates and impairing operation and maintenance (O&M) decision-making. These limitations reveal significant shortcomings in feature representation, spatiotemporal modeling accuracy, and anomaly detection sensitivity, hindering the advancement of intelligent, refined O&M for large-scale, dynamically evolving power systems.

In recent years, driven by diverse emerging technologies, power equipment condition monitoring has advanced rapidly. Long [19] applied Internet of Things (IoT) technology for real-time monitoring and analysis of power equipment condition data, exploring approaches to enhance grid reliability and O&M efficiency. To address signal interference and feature extraction challenges in real-time monitoring [20–21], several studies have explored strategies integrating advanced signal processing

with intelligent optimization algorithms. Tong [22] et al. collected operational signals from power equipment, applying singular value decomposition and particle filtering to monitor condition changes. They further optimized the pheromone configuration and update strategy in the ant colony algorithm to construct a fault-location evaluation function, achieving timely monitoring and over 90% fault-location accuracy.

Considering the increasing computational and responsiveness demands of condition monitoring under complex operating conditions, researchers have applied multi-level collaborative computing and deep learning to enhance system performance. Zhang [23] et al. combined edge computing and cloud computing, integrating SURF (Speeded-Up Robust Features) detection with a CBAM (Convolutional Block Attention Module)-enhanced Yolov5 model to achieve precise monitoring and analysis in smart substations. Experiments showed recall and accuracy rates of 91.21% and 90.54%, respectively, for insulator fault detection. These studies form the methodological basis for this work in terms of feature extraction, signal processing, and edge–cloud collaborative computing: Long's IoT framework informs the multi-source online acquisition and transmission strategy; Tong et al.'s decomposition and optimization algorithms inspire the multi-scale dynamic feature extraction approach; and Zhang et al.'s edge–cloud and attention mechanisms guide the design for rapid state modeling in high-load environments. Nevertheless, despite achieving high accuracy in specific condition-monitoring and fault-location scenarios [24–25], existing research remains insufficient in multi-scale condition perception, unified data processing, and the development of full-link early warning mechanisms for complex grid environments.

Similarly, fault warning technology has attracted increasing attention as a vital tool for maintaining grid stability. Zhang [26] et al. investigated power equipment failure prediction and lifecycle management using big data analytics, proposing and validating prediction strategies that enhance operational stability and safety. As fault warning research deepens [27–28], multi-source data fusion and spatial information technologies have emerged as key enablers for performance improvement. Kang [29] et al. introduced a smart grid fault warning method combining multi-sensor fusion with GIS-based positioning, verifying its effectiveness in improving both accuracy and timeliness. With growing grid complexity, efficient transmission-layer fault detection [30] has become essential for enhancing system resilience. Mu [31] et al. developed a power-flow-based transmission network security situation warning framework and applied oversampling for sample balancing. Tests on a modified IEEE 36-node system demonstrated approximately 90% cascading-fault warning accuracy with prediction times under one second, significantly improving security awareness and reducing anomaly identification latency.

To further advance early fault detection and reduce false alarms in complex monitoring scenarios, recent studies have explored advanced classification and temporal modeling techniques. For example, Dahr and Farjami [32, 33] proposed a hard voting ensemble integrating logistic regression, support vector machines, and random forests, achieving notable improvements in anomaly detection for network intrusion scenarios. In real-time spatiotemporal analysis, Xu [34] developed a YOLO-based framework integrating temporal context and network analysis, demonstrating the potential of combining spatial feature extraction with temporal dependency modeling in high-speed detection tasks.

These related works lay the technical foundation for this study in big data analytics, multi-sensor fusion, and real-time alarm mechanisms. Zhang et al.'s big data–driven lifecycle management strategy supports the multi-stage state modeling and prediction framework proposed herein; Kang et al.'s multi-source fusion and GIS positioning approach informs the fusion strategy for alarm location and timeliness optimization; and Mu et al.'s rapid transmission-network warning framework validates this work's confidence reconstruction and rapid decision-making approach. Although existing studies have achieved progress in improving the real-time performance, accuracy, and adaptability of warning models, gaps remain in large-scale grid-wide validation and dynamic responsiveness under complex fault scenarios.

To address these challenges, this paper proposes a big data–driven power equipment condition monitoring and fault warning system. In the data preprocessing stage, a multi-scale sliding-window method captures both local variations and global evolution patterns in operational time-series data across multiple time scales, mitigating degradation-feature loss from single-scale processing. In feature construction, a weighted trend-deviation index is introduced, applying differentiated weights to variation magnitude and trend to amplify subtle anomalies in the feature space. In feature optimization, redundant high-dimensional features are compressed via correlation analysis to eliminate noise and enhance feature compactness and effectiveness.

In the state modeling stage, a spatiotemporal adaptive graph convolutional network (STAGCN) is proposed, which dynamically updates both spatial adjacency and temporal dependencies to jointly model equipment topological relationships and operational evolution patterns, improving complex-system state representation accuracy. For anomaly detection, a confidence-score reconstruction strategy models the normal operating-state distribution and computes real-time deviations to generate quantitative anomaly scores, enabling early detection of potential faults.

The entire system adopts a hierarchical deployment architecture combined with a parallel big-data processing framework to support real-time ingestion, processing, and monitoring of massive equipment datasets, ensuring efficiency and reliability in large-scale scenarios. By systematically addressing the challenges of small-degradation feature capture, inaccurate dynamic-state modeling, and large errors in early fault identification, this work advances intelligent O&M methods for power equipment under big data environments. Compared with existing approaches that combine graph convolution and time-series modeling, the proposed STAGCN introduces

a learnable adaptive adjacency matrix and a dynamic temporal dependency update mechanism, enabling joint adaptive modeling of spatial topology and temporal evolution. This design, unlike fixed-topology or static time-window approaches, maintains high-precision state representation and weak-fault detection performance even under heterogeneous network structures and long-term dependencies. In large-scale concurrent data-stream scenarios, the integration of multi-scale feature processing with a parallel computing framework enables superior accuracy and lower latency compared with representative existing methods.

# 2 Method design and system construction

In the process of constructing the power equipment status monitoring and fault warning system, in order to ensure the method's systematicness and the implementation efficiency, it is necessary to clearly plan the overall design architecture in layers. Each functional module is organically organized according to the data flow and processing logic, forming a closely connected collaborative system at the level of data preprocessing, feature construction, feature optimization, status modeling, fault warning, and system architecture. Figure 1 intuitively shows the overall framework of the method design and system construction in this paper.

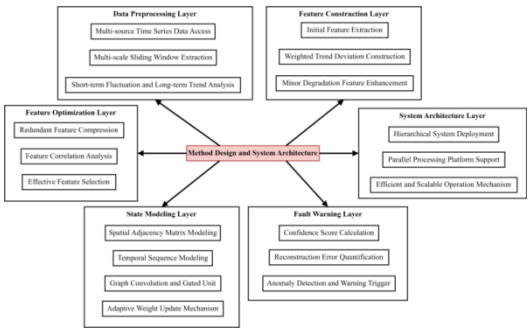


Figure 1: System overall framework diagram

Figure 1 shows the complete process of the system from raw data access to final warning output, and each layer is linked through clear data and information transmission. The data preprocessing layer focuses on the dynamic feature extraction of the original multi-source time series data; the feature construction layer is committed to strengthening the small degradation features through weighted trend deviation; the feature optimization layer improves the effectiveness of the feature set through redundant compression and screening. The state modeling layer combines spatial topology and time evolution behavior for unified modeling; the fault warning layer realizes anomaly discrimination and warning generation based on the confidence score reconstruction strategy; the system architecture layer uses hierarchical deployment and parallel processing technology to ensure the efficiency and scalability of the overall system. The overall design is closely centered on the goals of equipment status monitoring and fault warning, emphasizing the systematic and targeted nature of the method in each link.

## 2.1 Data preprocessing layer

The raw data collected in the power equipment status monitoring system usually comes from multi-source sensors, showing high-frequency, continuous, and heterogeneous time series characteristics. The data types cover physical quantities such as voltage, current, temperature, and vibration acceleration. The original time series data has obvious short-term fluctuations and long-term trend superposition characteristics under actual working conditions. Directly using the original data for modeling leads to insufficient feature expression, so it is necessary to design a reasonable dynamic feature extraction mechanism to process the data.

For the original multi-source time series data, this paper adopts a multi-scale sliding window dynamic feature extraction mechanism. It is assumed that the original single-channel time series is $X = \{x_1, x_2, \ldots, x_T\}$, where T is the total number of sampling points. The sliding window length is set to L; the step size is set to S; the scale factor is set to α. The window length L is set from the sampling frequency and dynamic response of the equipment, with the step size S = L/4 for moderate overlap. The scale factor α is in [0.5, 2.0] to represent different time scales. The final scale factor α values of 0.5, 1.0, and 2.0 were determined through grid search on a dataset, optimizing for maximum early fault detection accuracy while maintaining a false alarm rate below 4%. At each scale α, the original data is scaled to generate a new time series sequence $X^{(\alpha)} = \{x_1^{(\alpha)}, x_2^{(\alpha)}, \ldots, x_{T/\alpha}^{(\alpha)}\}$. Among them, the local average value $x_t^{(\alpha)}$ corresponding to the scale factor is calculated by Formula (1), which is used to smooth short-term high-frequency noise and highlight the long-term trend characteristics:

$$x_t^{(\alpha)} = \frac{1}{\alpha}\sum_{i=0}^{\alpha-1} x_{t+i} \quad (1)$$

At each scale, $X^{(\alpha)}$ is divided into multiple subsequences according to the sliding window method, and each subsequence is defined by Formula (2):

$$W_k^{(\alpha)} = \{x_{(k-1)S+1}^{(\alpha)}, x_{(k-1)S+2}^{(\alpha)}, \ldots, x_{(k-1)S+L}^{(\alpha)}\} \quad (2)$$

Among them, k is the window number, and $W_k^{(\alpha)}$ is the input of a single feature extraction.

To extract short-term volatility information and long-term trend characteristics, six statistical features are calculated in each window $W_k^{(\alpha)}$: mean $\mu_k^{(\alpha)}$, standard deviation $\sigma_k^{(\alpha)}$, maximum value, minimum value, kurtosis $\kappa_k^{(\alpha)}$, and skewness $\gamma_k^{(\alpha)}$. The mean $\mu_k^{(\alpha)}$ is defined by Formula (3):

$$\mu_k^{(\alpha)} = \frac{1}{L}\sum_{i=1}^{L} x_{(k-1)S+i}^{(\alpha)} \quad (3)$$

The standard deviation $\sigma_k^{(\alpha)}$ is defined by Formula (4):

$$\sigma_k^{(\alpha)} = \sqrt{\frac{1}{L}\sum_{i=1}^{L}\left(x_{(k-1)S+i}^{(\alpha)} - \mu_k^{(\alpha)}\right)^2} \quad (4)$$

The kurtosis $\kappa_k^{(\alpha)}$ and skewness $\gamma_k^{(\alpha)}$ are calculated based on the fourth-order moment and the third-order central moment standardization, respectively, to capture abnormal spikes and skew characteristics, as shown in Formula (5) and Formula (6):

$$\kappa_k^{(\alpha)} = \frac{\frac{1}{L}\sum_{i=1}^{L}\left(x_{(k-1)S+i}^{(\alpha)} - \mu_k^{(\alpha)}\right)^4}{\left(\sigma_k^{(\alpha)}\right)^4} \quad (5)$$

$$\gamma_k^{(\alpha)} = \frac{\frac{1}{L}\sum_{i=1}^{L}\left(x_{(k-1)S+i}^{(\alpha)} - \mu_k^{(\alpha)}\right)^3}{\left(\sigma_k^{(\alpha)}\right)^3} \quad (6)$$

The multi-scale sliding window extracts feature sets at different scales and different statistics. For unified processing, the features of all scales are concatenated to form the final preliminary feature matrix $F = [F^{(0.5)}, F^{(1.0)}, F^{(2.0)}]$, where $F^{(\alpha)}$ represents the feature set extracted from all windows at scale α. All-scale features are concatenated to form the preliminary feature matrix, supporting subsequent construction and modeling by retaining local fluctuations and global trends.

## 2.2 Feature construction layer

After the preliminary extraction of multi-scale dynamic features, in order to enhance the characterization capability of the minor degradation behavior of power equipment, a feature construction method based on weighted trend deviation quantification is designed. It is assumed that the processed feature sequence is $\{s_k\}_{k=1}^{N}$, where N is the sequence length, and $s_k$ is the feature value extracted at the k-th moment. For this sequence, the local weighted mean $\nu_k$ is calculated by Formula (7):

$$\nu_k = \frac{\sum_{j=k-L}^{k}\theta_j s_j}{\sum_{j=k-L}^{k}\theta_j} \quad (7)$$

Among them, L is the length of the local weighted window, and $\theta_j$ is the weighting factor corresponding to time j, which is set according to the exponential decay law shown in Formula (8):

$$\theta_j = e^{-\alpha(k-j)} \quad (8)$$

Here, α is the weight decay coefficient, which controls the degree of influence of historical features on the current local mean. The local weighted mean $\nu_k$ is used to describe the evolution trend of the feature value in the short term. The trend deviation $\epsilon_k$ is further defined by Formula (9) to represent the degree of deviation of the current feature value relative to the local trend:

$$\epsilon_k = s_k - \nu_k \quad (9)$$

To enrich the feature representation, a variety of statistics are extracted from the $\{\epsilon_k\}_{k=1}^{N}$ sequence, including weighted deviation mean, weighted deviation standard deviation, skewness, and kurtosis, and the calculation formulas are as follows:

$$M(\epsilon) = \frac{1}{N}\sum_{k=1}^{N}\epsilon_k \quad (10)$$

$$S(\epsilon) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\epsilon_k - M(\epsilon))^2} \quad (11)$$

$$Sk(\epsilon) = \frac{\frac{1}{N}\sum_{k=1}^{N}(\epsilon_k - M(\epsilon))^3}{\left(\sqrt{\frac{1}{N}\sum_{k=1}^{N}(\epsilon_k - M(\epsilon))^2}\right)^3} \quad (12)$$

$$Ku(\epsilon) = \frac{\frac{1}{N}\sum_{k=1}^{N}(\epsilon_k - M(\epsilon))^4}{\left(\sqrt{\frac{1}{N}\sum_{k=1}^{N}(\epsilon_k - M(\epsilon))^2}\right)^4} \quad (13)$$

The weighted deviation mean reflects overall trend deviation, the standard deviation measures local fluctuation intensity, the skewness indicates distribution asymmetry, and the kurtosis describes peak characteristics. Based on trend deviation quantification, these statistics are extracted across multiple time windows to capture both short-term disturbances and long-term evolution, forming feature vectors that support subsequent feature optimization and state modeling.

## 2.3 Feature optimization layer

After the initial feature construction is completed, to improve the modeling efficiency and identification accuracy, the high-dimensional feature set needs to be optimized. In the feature optimization layer, this paper adopts a redundant feature compression algorithm, combined with feature correlation evaluation and validity screening mechanism, to remove information redundancy and enhance the compactness of the feature set. It is assumed that the original feature set is $F = \{f_1, f_2, \ldots, f_n\}$, where each feature vector is $f_i \in \mathbb{R}^T$, and T represents the time series length. The Pearson correlation coefficient $\rho(f_i, f_j)$ shown in Formula (14) is used to measure the linear correlation between any two features:

$$\rho(f_i, f_j) = \frac{\sum_{t=1}^{T}(f_i^t - \bar{f}_i)(f_j^t - \bar{f}_j)}{\sqrt{\sum_{t=1}^{T}(f_i^t - \bar{f}_i)^2}\cdot\sqrt{\sum_{t=1}^{T}(f_j^t - \bar{f}_j)^2}} \quad (14)$$

Among them, $\bar{f}_i$ and $\bar{f}_j$ are the time means of features $f_i$ and $f_j$, respectively. When $|\rho(f_i, f_j)| > \theta$, it is considered that there is a high degree of redundancy between the two. Here, θ is the threshold, which is determined through multiple groups of experiments to ensure the balance between compression strength and information retention. The threshold θ was set to 0.92 based on iterative trials, where feature pairs exceeding this correlation were removed to prevent collinearity. The number of retained features after compression was fixed at 65 to ensure stability across different datasets. On this basis, the maximum mutual information method is used to evaluate the information contribution between each feature and the target state label. The mutual information calculation method is as shown in Formula (15):

$$I(f_i; y) = \sum_{f_i \in \mathcal{F}}\sum_{y \in \square} p(f_i, y)\log\frac{p(f_i, y)}{p(f_i)\cdot p(y)} \quad (15)$$

Among them, $p(f_i, y)$ represents the joint distribution probability, and $p(f_i)$ and $p(y)$ are the marginal probabilities of the feature and the label, respectively. The first k features with the largest mutual information value are retained to construct the final optimized feature subset $F^* = \{f_1^*, f_2^*, \ldots, f_k^*\}$. To ensure the uniformity of feature scale, all retained features are normalized using the Z-score normalization method shown in Formula (16):

$$f_i^{*\prime} = \frac{f_i^* - \mu_i}{\sigma_i} \quad (16)$$

Among them, $\mu_i$ and $\sigma_i$ represent the mean and standard deviation of the optimized feature $f_i^*$, respectively. Through the above redundant compression and effectiveness screening process, the optimized feature

set is significantly compressed in dimension; the redundant interference information is effectively eliminated; the key information with discriminative ability for state identification is retained, providing a higher quality input basis for the subsequent modeling layer.

## 2.4 State modeling layer

The state modeling layer realizes the joint modeling of the power equipment state by constructing STAGCN. The network structure uses the structural topology information between equipment to construct the adjacency relationship graph in the spatial dimension, and applies the gating mechanism for dynamic modeling in the time dimension in combination with the evolution trend of the sequence data. It is assumed that the network composed of power equipment is a graph $G = (V, E, A)$. Among them, V represents the node set; E represents the edge set; $A \in \mathbb{R}^{N\times N}$ is the static adjacency matrix; N is the number of nodes.

This paper applies a weight update mechanism to dynamically adjust the adjacency matrix, and constructs a learnable adaptive adjacency matrix $\widehat{A} \in \mathbb{R}^{N\times N}$ through Formula (17):

$$\widehat{A}_{ij} = \frac{\exp\left(\mathrm{ReLU}\left(z_i^T W_1 + z_j^T W_2 + b_1\right)\right)}{\sum_{k=1}^{N} \exp\left(\mathrm{ReLU}\left(z_i^T W_1 + z_k^T W_2 + b_1\right)\right)} \quad (17)$$

Among them, $z_i$ and $z_j$ are the embedded representations of nodes i and j; $W_1$ and $W_2$ are trainable weight matrices; $b_1$ is the bias term; ReLU represents the activation function. Exponential normalization ensures that the matrix is normalized by row, meeting the requirements of graph convolution.

The graph convolution operation is based on the adaptive adjacency matrix to aggregate node features and define the l-th layer graph convolution operation as shown in Formula (18):

$$H^{(l+1)} = \sigma\left(\widehat{A} H^{(l)} W^{(l)}\right) \quad (18)$$

Among them, $H^{(l)} \in \mathbb{R}^{N\times d_l}$ represents the node feature representation of the l-th layer; $W^{(l)} \in \mathbb{R}^{d_l \times d_{l+1}}$ is the weight matrix; $\sigma(\cdot)$ is the activation function.

The temporal modeling module applies the gated recurrent unit (GRU) to model the temporal features. Assuming that the temporal feature of each node is $X_t \in \mathbb{R}^{N\times d}$, the state update process of GRU is as shown in Formulas (19) to (21):

$$r_t = \sigma(W_r X_t + U_r H_{t-1} + b_r) \quad (19)$$

$$z_t = \sigma(W_z X_t + U_z H_{t-1} + b_z) \quad (20)$$

$$H_t = (1 - z_t) \odot H_{t-1} + z_t \odot \tanh\left(W_h X_t + U_h(r_t \odot H_{t-1}) + b_h\right) \quad (21)$$

Among them, $r_t$ and $z_t$ are the reset gate and update gate, respectively; $\odot$ represents the Hadamard product; $W_*$ and $U_*$ are parameter matrices; $b_*$ is the bias term; $H_t$ is the hidden state at the current moment. The spatial graph convolution and the temporal GRU are fused in the network through inter-layer alternation to achieve unified modeling of the power equipment state in the dimensions of topological structure and time evolution. The final state representation uses the aggregation result of multiple layers of spatiotemporal features as the output, and the input fault warning layer is used to discriminate the model and generate warning signals. This modeling method takes into account the complex dependencies between nodes and the temporal dynamics of the node's own status, effectively supporting high-precision identification of subtle degradation trends and abnormal evolution behaviors of equipment.

## 2.5 Fault warning layer

The fault warning layer applies an anomaly discrimination strategy based on confidence score reconstruction, aiming to quantify the degree of state deviation and generate early warning signals. This strategy is based on the normal distribution characteristics of the equipment state, and measures the anomaly of the input state by jointly measuring the reconstruction error and confidence score. First, the input state feature vector $x \in \mathbb{R}^d$ is nonlinearly mapped and input into the multi-layer autoencoder network to obtain its reconstructed output $\hat{x}$. The reconstruction error is defined as $e = x - \hat{x}$, and its Euclidean norm is calculated by Formula (22):

$$E(x) = \| x - \hat{x} \|_2^2 \quad (22)$$

Among them, $E(x)$ represents the reconstruction error of the input state, which is used to characterize the degree of deviation between the current state and the normal state distribution. To improve the ability to distinguish weak anomalies, the confidence score function $\mathcal{C}(x)$ is applied and constructed according to the distribution density of the normal state of the input sample in the training set. The confidence score is defined by the kernel density estimation method shown in Formula (23):

$$\mathcal{C}(x) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (23)$$

Among them, $x_i$ is the i-th normal state sample in the training set; n is the number of normal samples; σ is the kernel width parameter. The confidence score reflects the degree of belonging of the current state in the normal sample distribution. The lower the value, the greater the difference from the normal state. The comprehensive discriminant function is defined by Formula (24):

$$\mathcal{A}(x) = \alpha \cdot E(x) - \beta \cdot \log \mathcal{C}(x) \quad (24)$$

Among them, α and β are empirically determined adjustment coefficients that control the weights of reconstruction error and confidence score in the final discriminant value. The abnormal discrimination threshold θ is set according to the statistical distribution of reconstruction error and confidence score in the training set, and the discrimination logic is shown in Formula (25):

$$y = \begin{cases} 1, & \text{if } \mathcal{A}(x) > \theta \\ 0, & \text{otherwise} \end{cases} \quad (25)$$

Among them, y=1 denotes an alarm-triggering abnormal state and y=0 a normal state. By combining state reconstruction deviation and distribution attribution, the strategy enhances early fault sensitivity while maintaining a low false alarm rate. Historical alarms and real-time data are jointly processed for timestamping and severity classification, then pushed via the message bus to support rapid front-end decision-making. Integrated with the layered architecture, it ensures low-latency processing under large-scale data streams.

## 2.6 System architecture layer

The system adopts a collaborative architecture integrating edge and centralized computing to form a distributed data processing and unified state management platform. Field sensors collect raw data, transmitted via communication modules to edge nodes, where a lightweight unit standardizes data, calibrates time, detects anomalies, and sends results to the core cloud cluster equipped with CUDA-enabled embedded GPUs for asynchronous batch processing. The cloud cluster comprises task scheduling, GNN modeling, early warning, and result caching nodes, enabling non-blocking data flow, dynamic graph modeling with temporal reasoning, anomaly scoring, and high-performance in-memory storage for visualization. ZeroMQ ensures low-latency communication, while CPU/GPU coordination with non-blocking multitasking maximizes throughput. Containerized deployment with Kubernetes supports elastic scaling and load-adaptive scheduling, and ELK-based logging enables real-time monitoring and fault tracing, ensuring reliable operation and rapid fault warning under large-scale, high-concurrency conditions.

# 3 Experimental setup and implementation

## 3.1 Experimental platform and system environment

To ensure the reproducibility of system construction and the fairness of experimental evaluation, this paper follows the principles of high performance, strong stability, and wide adaptability in the configuration of the experimental platform and the selection of the software environment. Taking into account the actual needs of big data processing capabilities, deep learning model training efficiency, and multi-module concurrent operation, a set of experimental environments with universality and engineering application value is constructed. The platform resources used include hardware equipment, operating system, program language environment, and core framework dependencies. The specific configuration is shown in Table 1.

Table 1: Experimental platform and system environment parameter settings

| Category | Name/Model | Size/Version | Usage Description |
|---|---|---|---|
| GPU | NVIDIA GeForce RTX 3060 | 12 GB | Model training and graph convolution |
| CPU | AMD Ryzen 7 5800X | 8 cores, 16 threads, 3.8GHz | Data processing and system computation |
| Operating System | Windows 10 Pro | 64-bit, 21H2 | System deployment and resource management |
| Programming Language | Python | 3.7 | Feature processing, model building, integration |
| Deep Learning Framework | PyTorch | 1.10.2 | STAGCN implementation and training control |
| Data Processing Platform | Apache Spark | 3.1.2 | Distributed preprocessing and feature extraction |

This configuration has completed heterogeneous computing resource support to meet the parallel computing requirements of deep graph neural networks for graph convolution and time modeling. The operating system version is stable and compatible with mainstream dependency libraries and middleware environments. Python and its dependent framework versions are consistent with the STAGCN model implementation requirements. The big data processing platform has good scalability and task scheduling capabilities, providing efficient and reliable operation support for the multi-level processing flow proposed in this paper.

## 3.2 Dataset source and preprocessing method

The original data comes from the status monitoring equipment deployed at the actual substation and transmission line site, covering key facilities such as main transformers, circuit breakers, capacitor compensation equipment, busbars, and transmission line monitoring terminals. Various types of equipment continuously collect multi-dimensional time series information such as temperature, current, voltage, and power. The sampling frequency is mainly in seconds, and the daily data volume is maintained at a high-frequency and high-dimensional level, forming a huge original dataset. The collection period covers 180 consecutive days, and the data storage structure is organized on the HDFS platform according to column encoding, building a unified and accessible distributed data foundation. Table 2 shows the number of monitoring dimensions, number of daily samples, sampling frequency, and cumulative data volume corresponding to the collection time of various types of equipment, which helps to quantify the structural characteristics and scale of the data foundation.

Table 2: Structural statistical information of the original monitoring data

| Equipment Type | Number of Monitoring Dimensions | Number of Daily Samples (records) | Sampling Frequency (Hz) | Cumulative Data Volume (records) |
|---|---|---|---|---|
| Main Transformer | 8 | 86,400 | 1 | 15,552,000 |
| Circuit Breaker | 6 | 64,800 | 0.75 | 11,664,000 |
| Capacitor Compensation | 7 | 86,400 | 1 | 15,552,000 |
| Busbar | 5 | 57,600 | 0.67 | 10,368,000 |
| Transmission Line Monitoring Terminal | 9 | 86,400 | 1 | 15,552,000 |

Table 2 shows the data structure formed by various key equipment in the power system during the state monitoring process, highlighting the differences in sampling frequency, data dimension, and monitoring cycle between different equipment types, and reflecting the high-density characteristics of the overall time series data of the system. The data entries are all derived from the real sampling process. By standardizing the sampling frequency and acquisition time, the consistency and comparability of the data structure are ensured, providing structural support for the parameter setting and method selection of the subsequent data preprocessing process.

The data preprocessing process first performs data cleaning operations to remove physical illegal values and logical conflict values. The physical value constraint domain is defined as $x_t \in [x_{min}, x_{max}]$. Among them, $x_t$ represents the monitoring value at any time, and $x_{min}$ and $x_{max}$ are the theoretical lower and upper limits of the corresponding variables of the sensor, respectively. Data exceeding this interval are directly set as missing items. The logical conflict judgment criterion adopts the multi-variable linkage consistency rule. For example, the power factor between voltage, current, and power is reversed to verify whether the constraint relationship is established. If not, it is marked as abnormal data.

The missing data is filled by backfilling the mean of the stable segment in the adjacent time period, as shown in Formula (26):

$$x'_t = \frac{1}{N}\sum_{i=1}^{N} x_{t-i} \quad (26)$$

Among them, $x'_t$ represents the value after filling at the t-th moment, and $x_{t-i}$ represents the observation value in the window of N stable moments in history. The stable moment is determined based on the sliding standard deviation threshold criterion. When the standard deviation $\sigma_{t-w:t+w} < \delta$, the observation value in the interval is considered to be in a stable state. The window width is set to $w = 10$, and the threshold is $\delta = 0.02 \times \bar{x}$, where $\bar{x}$ is the global mean of the current variable. The data standardization process uses the Z-Score normalization method to ensure that multi-source heterogeneous variables enter the subsequent feature extraction process under the same scale. This method ensures that each variable is processed in a scale space with a mean of 0 and a standard deviation of 1, thereby eliminating the impact of dimensional interference on the modeling process. To ensure data alignment and consistency, the time synchronization alignment process based on the nearest matching method is performed to address the timestamp offset problem of sensors of different equipment. It is assumed that the observation timestamps of equipment i and j are $\{t_i^k\}$ and $\{t_j^l\}$, respectively, and the matching principle is to find $\min|t_i^k - t_j^l| \le \epsilon$, where $\epsilon$ is the maximum synchronization error tolerance, set to 1 second. The data points that are not successfully matched are interpolated and removed. After completing the dataset cleaning, completion, standardization, and synchronization, a time series data tensor with a unified structure is finally formed, providing a data basis for subsequent feature extraction and state modeling.

## 3.3 Experimental parameter configuration

To ensure the reproducibility of the model training and evaluation process and the consistency of parameter tuning, this section systematically organizes the key experimental configurations. The experiment uses multiple sets of parameter combinations to examine the impact of different configurations on fault identification performance, training efficiency, and system robustness. Table 3 summarizes the core parameter settings of each experiment to support subsequent model performance comparison and result analysis.

Table 3: Experimental parameter configuration and combination settings

| Parameter Configuration | Configuration 1 | Configuration 2 | Configuration 3 | Configuration 4 |
|---|---|---|---|---|
| Learning Rate | 0.001 | 0.0005 | 0.0001 | 0.0001 |
| Number of Epochs | 100 | 150 | 200 | 200 |
| Batch Size | 64 | 128 | 64 | 32 |
| Optimizer | Adam | AdamW | RMSprop | SGD |
| Sliding Window Length | 50 | 100 | 150 | 100 |
| Multi-scale Parameters | [1,3,5] | [1,2,4,6] | [2,4,8] | [1,5,10] |
| GCN Layers | 2 | 3 | 3 | 4 |

| | | | | |
|---|---|---|---|---|
| Temporal Step Size | 1 | 1 | 2 | 2 |
| Weight Initialization | Xavier | Kaiming | Xavier | Normal |

Table 3 shows the core parameter configuration of the model under different training strategies, covering learning rate, number of training rounds, batch settings, optimizer type, time series feature processing parameters, and graph neural network structure settings, forming a multi-dimensional experimental variable combination. The parameter settings provide a basic basis for the subsequent model performance differences, and also reflect the adaptability and stability of the system under complex configuration space.

This experiment designs four groups of model training schemes to evaluate the actual operation effects of different parameter configurations in the power equipment status monitoring and fault warning system, which differ in feature processing granularity, graph network depth, optimizer type, and time series modeling step size. Parameter combination 1 uses a medium sliding window, a lower number of graph convolution layers and an Adam optimizer, emphasizing the balance between model response speed and generalization ability. Parameter combination 2 applies the AdamW optimizer with stronger weight regularization while maintaining high step size modeling capabilities, aiming to improve long-term prediction stability. Parameter combination 3 configures a deeper graph convolution structure and uses the RMSprop optimizer, aiming to capture more complex state evolution patterns. Parameter combination 4 reduces the batch size and adopts the SGD optimization strategy to try to enhance the feature convergence performance under the local optimal update path. The system operation results corresponding to different schemes are summarized in Table 4.

Table 4: System performance under different parameter combinations

| Parameter Configuration | Fault Detection Accuracy (%) | False Alarm Rate (%) | Real-time Processing Rate (%) |
|---|---|---|---|
| Configuration 1 | 94.7 | 3.5 | 96.2 |
| Configuration 2 | 93.2 | 4.1 | 95.5 |
| Configuration 3 | 91.5 | 5.2 | 93.8 |
| Configuration 4 | 92.4 | 4.6 | 94.1 |

From the data results, parameter combination 1 performs best in all three indicators. Its high accuracy is mainly due to the balanced capture ability of the mesoscale sliding window for key feature fluctuations, and the shallow graph convolution structure reduces the risk of overfitting, forming an efficient and stable state modeling effect. From the overall performance, parameter combination 1 forms a good balance in accuracy, stability, and real-time response capability, and is the optimal configuration under the current setting.

# 4 Results

## 4.1 Multi-scale feature extraction effect

In the empirical phase of the multi-scale feature extraction method, this paper takes the current fluctuation data of power equipment as the object, designs and constructs a time series signal with local mutation, slow-changing trend, and background noise superposition characteristics, and simulates the dynamic change characteristics of equipment data under real operating conditions. By setting the scale parameters of the sliding window to 3, 7, and 15, respectively, the sliding mean features at different levels are extracted to observe its ability to capture the key change patterns in the signal. The original signal and the mesoscale feature extraction results are compared synchronously to verify the role of sliding window processing in suppressing disturbances and enhancing trend identification. Figure 2 shows the comparative relationship between the original signal and the mesoscale sliding mean, as well as the response differences of feature changes at three scales.

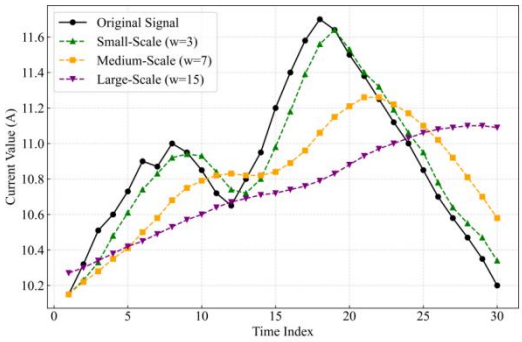


Figure 2: Comparison of multi-scale sliding mean feature extraction effects

As can be seen from Figure 2, the original current signal shows obvious nonlinear fluctuation characteristics from the 1st moment to the 30th moment, and is at a lower level from the 1st to the 5th moment. The current value rises from 10.15 A to 10.73 A, peaks at 11.70 A at the 18th moment, and then gradually decreases to 10.20 A at the 30th moment. After small-scale (w=3) sliding processing, the sequence still retains most of the spike information, with values of 10.94 A and 11.64 A at the 9th and 19th moments. The reason for this phenomenon is that the window width is small, resulting in local disturbances not being fully smoothed, thus maintaining the high-frequency details in the original signal. The medium-scale (w=7) curve changes gently between the 1st moment and the 22nd moment, and the current peak is significantly

lower than the peak of the original signal, indicating that some disturbances are suppressed, while the rising and falling trends are still retained.

The large-scale (w=15) smooth curve is the most stable overall. From the 1st to the 30th moment, the current value is always controlled between 10.27A and 11.10 A, and no mutation points appear in the original signal. The fundamental reason is that the average of the wide window for multi-time data causes short-term changes to be smoothed and filtered, thus presenting a response structure dominated by low-frequency trend features. It can be seen that the sliding window scale determines the degree of manifestation of short-term disturbances and long-term trends in the signal.

## 4.2 Feature optimization to improve modeling performance

In large-scale power equipment status monitoring tasks, the original sensor data has high dimensions and a large proportion of redundant features. Direct modeling often faces problems such as heavy computational burden, slow convergence speed, and unstable model discrimination ability. To improve the model's learning efficiency and expression accuracy in complex feature space, the system applies a redundant feature compression mechanism to eliminate invalid variables and enhance the modeling weight of core feature signals. This section compares and analyzes the modeling performance before and after compression, and tracks the performance evolution trend of the two types of models in the iteration process through the dual indicators of training accuracy and false alarm rate. The results are shown in Figure 3.

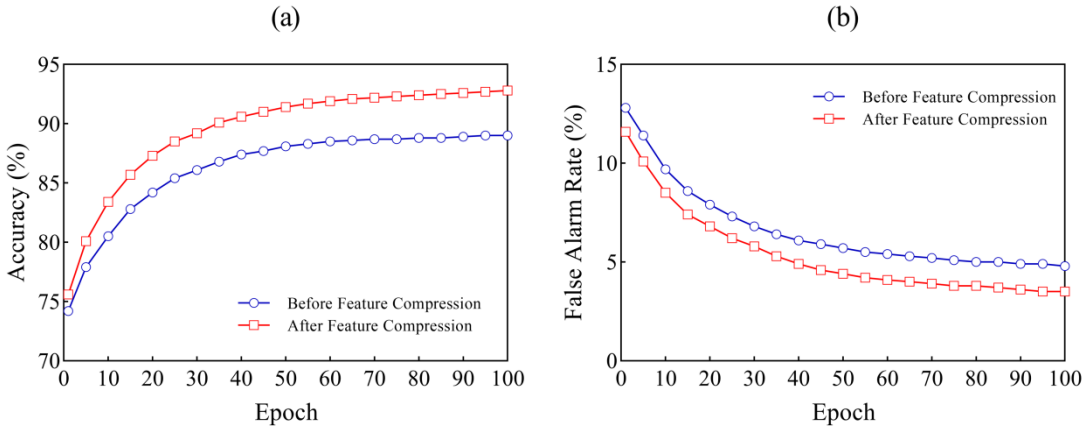


Figure 3: Effect of redundant feature compression on model accuracy and false alarm rate
Figure 3 (a): Accuracy
Figure 3 (b): False alarm rate

Figure 3 shows that after feature optimization, the model accuracy shows a faster upward trend in the early stage of training. The accuracy in the first 15 rounds increases from 75.6% to 85.7%, which is 2.9 percentage points higher than the performance of the unoptimized model in the same round. This difference is mainly due to the reduction of interference items brought by the highly redundant feature set. The optimization process improves the effective density of feature expression, allowing the model to capture feature structures that are highly relevant to the target state earlier. In the later stage of training, the accuracy of the compressed model tends to stabilize at 92.8%, showing stronger convergence than the final accuracy of 89.0% of the unoptimized model. In terms of false alarm rate, the optimized model always maintains a low level, dropping to 4.1% after the 60th round and finally controlled at 3.5%, while the false alarm rate of the unoptimized model during the same period remains at around 4.8%. The difference comes from the reduced disturbance of the redundant factors in the high-dimensional features on the abnormal discrimination boundary. Overall, the feature compression strategy effectively improves the model's generalization ability and the sensitivity to weak fault signals.

4.3 System Fault Warning Effect

To comprehensively evaluate the performance of the fault warning system under different working conditions, this paper constructs experimental scenarios with fault type and operating load as variables to verify the stability and sensitivity of the system under multiple working conditions. Five typical fault types cover insulation aging, poor contact, mechanical looseness, line short circuit, and vibration anomaly, representing the main degradation modes and sudden anomalies that may occur in power equipment during long-term operation. The operating load is set to five levels of 20%, 40%, 60%, 80%, and 100%, simulating the real-time working status of the equipment under different power outputs to observe the system's response to changes in operating pressure. The experiment records the warning accuracy of each fault type and the average response time under each load level, and the results are shown in Figure 4.

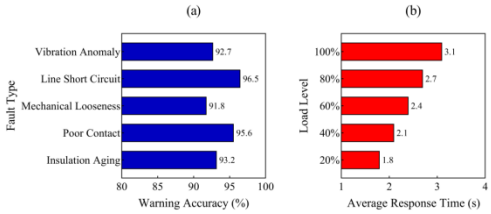


Figure 4: Comparison of the accuracy and response time of power equipment fault warning

Figure 4 (a). Warning accuracy under different fault types

Figure 4 (b). Warning response time under different operating loads

Figure 4 shows that in the fault type dimension, the system's warning accuracy for line short circuits reaches 96.5%, significantly higher than the 91.8% for mechanical looseness. This difference is due to the fact that line short circuits are more drastic in terms of data fluctuation amplitude, frequency characteristics, and energy mutation, which makes them have higher discrimination in the model, while mechanical looseness is easily misjudged as normal fluctuations due to weak interference characteristics and slow degradation process, affecting model judgment. Poor contact and insulation aging reach 95.6% and 93.2%, respectively, mainly driven by their boundary characteristics in the mid-frequency vibration spectrum. The accuracy of vibration anomalies is 92.7%, which is related to their strong noise interference and high degree of feature overlap. Under load conditions, the warning response time increases linearly with the increase in load level, reaching 3.1 seconds at 100% load and only 1.8 seconds at 20% load. The increase in system processing pressure leads to a decrease in the parallel efficiency of feature extraction and modeling modules, which is the main reason for the extension of high-load response time. The results show that the system maintains strong robustness in typical fault identification and high-load real-time warning, and has a basis for promotion and application.

4.4 Verification of Overall System Real-time and Scalability

To fully verify the operating effect of the system based on big data analysis proposed in this paper in a multi-source data stream environment, the experiment designs a system performance comparison experiment covering different data scales, number of equipment, and number of nodes. By quantifying the differences between this system and the traditional centralized architecture in real-time processing capabilities and resource scheduling efficiency from multiple angles, the processing stability and scalability of the system under high concurrency conditions are deeply evaluated. The experimental results are shown in Figure 5.

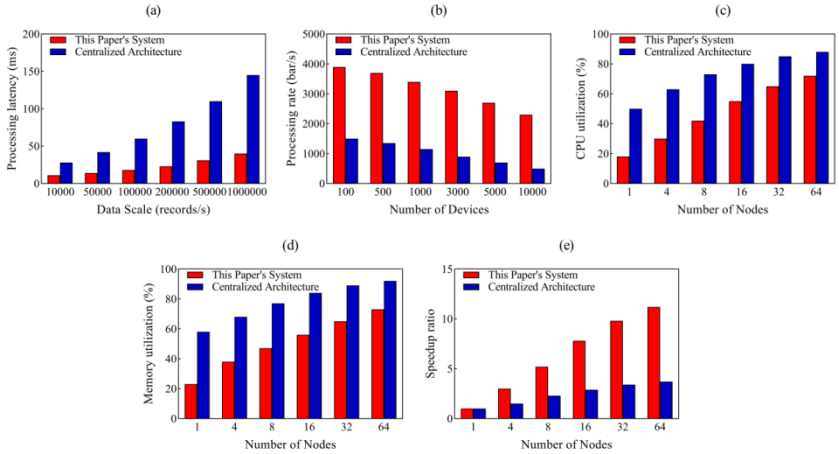


Figure 5: Comparison of processing performance and resource utilization between this paper's system and centralized architecture under different scales and configurations

Figure 5 (a): Comparison of processing latency under different data scales

Figure 5 (b): Comparison of processing rate under different number of equipment

Figure 5 (c): Comparison of CPU utilization under different number of nodes

Figure 5 (d): Comparison of memory utilization under different number of nodes

Figure 5 (e): Comparison of parallel speedup ratio under different number of nodes

Figure 5 (a) shows that as the data input scale increases from 10,000/second to 1 million/second, the processing latency of this system increases from 11ms to 40ms, while the centralized architecture latency increases from 28ms to 145ms, mainly because this system adopts a hierarchical parallel mechanism to effectively alleviate the problem of single-point load concentration. In Figure 5 (b), when the number of equipment is expanded to 10,000, the processing rate of this system reaches 2,300/second, while the centralized architecture is only 500/second, which is due to the adaptive scheduling mechanism of this system for node load balancing. Figure 5 (c) and Figure 5 (d) show the CPU and memory utilization, respectively. When the number of nodes increases from 4 to 64, the CPU utilization of this system increases from 30% to 72%, and the memory utilization increases from 38% to 73%. In contrast, the CPU utilization of the centralized architecture quickly soars to 88%, and the memory utilization reaches 92%. The centralized resource bottleneck effect is obvious, due to the lack of dynamic

migration strategy for cross-node resources. Figure 5 (e) shows the trend of parallel speedup ratio changing with the number of nodes. When the number of nodes is 64, this system reaches 11.2 times, while the centralized architecture is only 3.7 times, which is attributed to the combined effect of data flow segmentation and communication load optimization strategy. The overall results show that the system proposed in this paper has significant advantages over the traditional architecture in terms of multi-dimensional performance indicators, especially in the case of large-scale data and high node parallelism, showing excellent processing stability and resource adaptation capabilities.

Conclusions

Based on big data analysis, this paper proposes a power equipment condition monitoring and fault warning system that integrates multi-scale sliding-window dynamic feature extraction, weighted trend-deviation quantitative feature construction, redundant-feature compression optimization, and STAGCN-based state modeling. The system enhances early fault identification accuracy through an anomaly discrimination mechanism based on confidence-score reconstruction. Experimental results show that, after redundant-feature compression, the model's accuracy improves from 89.0% to 92.8%, while the false-alarm rate decreases from 4.8% to 3.5%, thereby significantly improving modeling efficiency and identification accuracy. In fault warning experiments, the proposed method achieves a line short-circuit identification accuracy of 96.5%, with a warning response time of 3.1 seconds under 100% load conditions, demonstrating strong real-time performance and high fault-detection capability. Although the system performs well under typical load and fault scenarios, certain limitations remain in detecting weak anomalies under conditions of multiple fault superposition and extreme load fluctuations. Future work will explore the incorporation of multimodal perception and dynamic anomaly-threshold mechanisms to further enhance early fault detection and response capabilities in complex environments, thereby improving the system's adaptability and stability in ultra-large-scale power system deployments.

# Acknowledgements

This work was supported in part by the Industry University Research Cooperation Project of Jiangsu Province under Grant BY20230961, and in part by the Research Project of Jiangsu Vocational Institute of Architectural Technology under Grant JYA324-21.

# References

[1] Omitaomu, Olufemi A., and Haoran Niu. "Artificial intelligence techniques in smart grid: A survey." Smart Cities 4.2 (2021): 548-568. https://doi.org/10.3390/smartcities4020029

[2] Li, Yuanzheng, Chaofan Yu, Mohammad Shahidehpour, Tao Yang, Zhigang Zeng, and Tianyou Chai. "Deep reinforcement learning for smart grid operations: Algorithms, applications, and prospects." Proceedings of the IEEE 111.9 (2023): 1055-1096. https://doi.org/10.1109/jproc.2023.3303358

[3] Ufa, R. A., Y.Y. Malkova, V.E. Rudnik, M.V. Andreev, and V.A. Borisov. "A review on distributed generation impacts on electric power system." International journal of hydrogen energy 47.47 (2022): 20347-20361. https://doi.org/10.1016/j.ijhydene.2022.04.142

[4] Liu, Keyan, Wanxing Sheng, Zhao Li, Fang Liu, Qianyi Liu, and Yucong Huangi, et al. "An energy optimal schedule method for distribution network considering the access of distributed generation and energy storage." IET Generation, Transmission & Distribution 17.13 (2023): 2996-3015. https://doi.org/10.1049/gtd2.12855

[5] Xia, Changjie, Ming Ren, Bing Wang, Ming Dong, Guanghao Xu, and Jiacheng Xie, et al. "Infrared thermography-based diagnostics on power equipment: state-of-the-art." High Voltage 6.3 (2021): 387-407. https://doi.org/10.1049/hve2.12023

[6] Gehao, S. H. E. N. G., QIANG Yong, LUO Lingen, SONG Hui, LIU Yadong, and JIANG Xiuchen. "Key technologies and development trends of digital power equipment for new type power system." High Voltage Engineering 49.5 (2023): 1765-1778. https://doi.org/10.1049/hve2.12027

[7] Liu, Zhengguang, Qinyue Tan, Yubo Zhou, and Hengshan Xu. "Syncretic application of IBAS-BP algorithm for monitoring equipment online in power system." IEEE access 9 (2021): 21769-21776. https://doi.org/10.1109/access.2021.3055247

[8] Zhou, Nan, Yan Xu, Sungin Cho, and Cheng Tian Wee. "A systematic review for switchgear asset management in power grids: condition monitoring, health assessment, and maintenance strategy." IEEE Transactions on Power Delivery 38.5 (2023): 3296-3311. https://doi.org/10.1109/tpwrd.2023.3272883

[9] Yang, Hanrui, Huanhuan Lv, Jingwei Zhang, Weiliang Huang, Shibo Xu and Shengxi Jiao. "Cable current-carrying status analysis and early fault warning method based on temperature information." Measurement Science and Technology 34.6 (2023): 065012. https://doi.org/10.1088/1361-6501/acbd22

[10] Yuming, Huang, Luo Jiaohong, Ma Zhenguo, Tang Bing, Zhang Keqi, and Zhang Jianyong. "On combined PSO-SVM models in fault prediction of relay protection equipment." Circuits, Systems, and Signal Processing 42.2 (2023): 875-891. https://doi.org/10.1007/s00034-022-02056-w

[11] Tian, Xiuxia, Can Li, and Bo Zhao. "A novel classification model SA-MPCNN for power equipment defect text." Transactions on Asian and Low-Resource Language Information Processing 20.6 (2021): 1-21. https://doi.org/10.1145/3464380

[12] Liu, Ti, Jia Feng, Dahong Fu, and Wenhan Chen. "Application of the optimised yolov3 algorithm in substation power equipment defect identification and detection [J]." Mechatronic Systems and Control 51.4 (2023): 235-243. https://doi.org/10.2316/j.2023.201-0386

[13] Li, Shimiao, Amritanshu Pandey, Bryan Hooi, Christos Faloutsos, and Larry Pileggi. "Dynamic graph-based anomaly detection in the electrical grid." IEEE

Transactions on Power Systems 37.5 (2021): 3408-3422. https://doi.org/10.1109/tpwrs.2021.3132852

[14] Dhiman, Harsh S., Dipankar Deb, S. M. Muyeen, and Innocent Kamwa. "Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines." IEEE Transactions on Energy Conversion 36.4 (2021): 3462-3469. https://doi.org/10.1109/tec.2021.3075897

[15] Lei, Tianxiang, Fangcheng Lv, Jiaomin Liu, and Jiahao Feng. "Research on electrical equipment monitoring and early warning system based on Internet of things technology." Mathematical Problems in Engineering 2022.1 (2022): 6255277. https://doi.org/10.1155/2022/6255277

[16] Wang, Shaoke, Zhaoyan Zhang, Peiguang Wang, and Yaru Tian. "Failure warning of gearbox for wind turbine based on 3σ-median criterion and NSET." Energy Reports 7 (2021): 1182-1197. https://doi.org/10.1016/j.egyr.2021.09.146

[17] Bo, Qi, ZHANG Peng, ZHANG Shuqi, ZHAO Linjie, WANG Hongbin, and HUANG Meng, et al. "Application status and development prospects of digital twin technology in condition assessment of power transmission and transformation equipment." High Voltage Engineering 47.5 (2021): 1522-1538.

[18] Zhang, Qin, and Yutang Liu. "Reliability evaluation of Markov cyber–physical system oriented to cognition of equipment operating status." Computer communications 181 (2022): 80-89. https://doi.org/10.1016/j.comcom.2021.10.004

[19] Long, Li. "Research on status information monitoring of power equipment based on Internet of Things." Energy Reports 8 (2022): 281-286. https://doi.org/10.1016/j.egyr.2022.01.018

[20] Ju, Xin, Ruixin Gou, Yanli Xiao, Zheng Wang and Shangke Liu. "The use of edge computing-based internet of things big data in the design of power intelligent management and control platform." International journal of grid and utility computing 13.1 (2022): 76-86. https://doi.org/10.1504/ijguc.2022.121426

[21]Jiang, Xianglong. "Automatic monitoring system of power equipment based on Internet of Things technology." International Journal of Emerging Electric Power Systems 23.6 (2022): 807-818. https://doi.org/10.1515/ijeeps-2022-0144

[22] Tong, Wei, and Qi-ping Huang. "A method for power equipment condition monitoring and fault location based on improved ant colony algorithm." International Journal of Energy Technology and Policy 19.3-4 (2024): 363-376. https://doi.org/10.1504/ijetp.2024.141388

[23] Zhang, Lishuo, Zhuxing Ma, Hao Gu, Zizhong Xin, and Pengcheng Han. "Condition Monitoring and Analysis Method of Smart Substation Equipment Based on Deep Learning in Power Internet of Things." International Journal of Information Technologies and Systems Approach (IJITSA) 16.3 (2023): 1-16. https://doi.org/10.4018/ijitsa.324519

[24] Afrasiabi, Shahabodin, Mousa Afrasiabi, Mohammad Amin Jarrahi, and Mohammad Mohammadi. "Fault location and faulty line selection in transmission networks: Application of improved gated recurrent unit." IEEE Systems Journal 16.3 (2022): 5056-5066. https://doi.org/10.1109/jsyst.2022.3172406

[25] Minullin, R. G., I. G. Akhmetova, V. A. Kasimov, and A. A. Piunov. "Location monitoring with determining the location of damage and the current performance of overhead power lines." Power Technology and Engineering 57.1 (2023): 145-152. https://doi.org/10.1007/s10749-023-01635-4

[26] Zhang, Kang, and Joan Lazaro. "Failure Prediction and Life Cycle Management of Power Equipment Based on Big Data Analysis." Journal of Computer, Signal, and System Research 2.1 (2025): 70-79. https://doi.org/10.71222/4ptjr081

[27] Meng, Fanqi, Shuaisong Yang, Jingdong Wang, Lei Xia, and Han Liu. "Creating knowledge graph of electric power equipment faults based on BERT–BiLSTM–CRF model." Journal of Electrical Engineering & Technology 17.4 (2022): 2507-2516. https://doi.org/10.1007/s42835-022-01032-3

[28] Shcherbatov, Ivan, Evgeny Lisin, Andrey Rogalev, Grigory Tsurikov, Marek Dvorak, and Wadim Strielkowski. "Power equipment defects prediction based on the joint solution of classification and regression problems using machine learning methods." Electronics 10.24 (2021): 3145. https://doi.org/10.3390/electronics10243145

[29] Kang, Zhihui, Yanjie Zhang, and Yuhong Du. "Application of multi-sensor information fusion technology in fault early warning of smart grid equipment." Energy Informatics 7.1 (2024): 123. https://doi.org/10.1186/s42162-024-00433-0

[30] Liu, Xiaoming, Jun Liu, Yu Zhao, Tao Ding, Xinglei Liu, and Jiacheng Liu. "A Bayesian deep learning-based probabilistic risk assessment and early-warning model for power systems considering meteorological conditions." IEEE Transactions on Industrial Informatics 20.2 (2023): 1516-1527. https://doi.org/10.1109/tii.2023.3278873

[31] Mu, Zeyu, Peidong Xu, Ke Zhang, Tianlu Gao, and Jun Zhang. "Cascading fault early warning and location method of transmission networks based on wide area time-series power system state." IEEE Journal of Radio Frequency Identification 7 (2022): 6-11. https://doi.org/10.1109/jrfid.2022.3208359

[32] Dahr J M, Farjami Y. A Hard Voting Ensemble Model of the Logistic Regression, Support Vector Machine and Random Forest for Network Intrusion Detection[J]. Informatica, 2025, 49(27). https://doi.org/10.31449/inf.v49i27.8573

[33] Dahr J M, Farjami Y. A Hard Voting Ensemble Model of the Logistic Regression, Support Vector Machine and Random Forest for Network Intrusion Detection[J]. Informatica, 2025, 49(27). https://doi.org/10.31449/inf.v49i27.8573

[34] Xu M. YOLO-Based Framework with Temporal Context and Network Analysis for Real-Time Basketball Video Understanding[J]. Informatica, 2025, 49(27). https://doi.org/10.31449/inf.v49i27.8657