Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 6, 2025

GIS Sub-Health Status Assessment Based on Multi-Frequency Signal Feature Cascade Network

Yuan Sun[1,*], Tao Wu[2], Guoyuan Chen[1], Xiuling Hu[1]

[1]Substation Maintenance Center of State Geid Shuozhou Electric Power Supply Company, Shuozhou, 036000, Shanxi Province, China
[2]State Grid Shuozhou Shentou Power Supply Company of State Geid Shuozhou Electric Power Supply Company, Shuozhou, 036000, Shanxi Province, China
[*]Corresponding author's email: SY7946sy@hotmail.com

Abstract. GIS (Gas-Insulated Switchgear) equipment generates multi-band partial discharge, electromagnetic vibration, acoustic emission and other signals during operation. Existing single-mode analysis is difficult to accurately identify complex faults, and the false alarm and missed alarm rates are high. A multi-frequency signal feature cascade network is proposed. Through frequency band decomposition, deep feature extraction and fusion attention mechanism, the early and accurate assessment of GIS sub-health status can be achieved, thereby improving the operation safety and equipment life. In this study, the original discharge, electromagnetic vibration and acoustic emission signals are first decomposed by Daubechies4 wavelet packet, and the signals are divided into three frequency bands: low, medium and high. Then, they are input into the corresponding branches. Subsequently, each branch uses convolutional neural network and self-attention module to collaboratively extract weak fault features in time domain and frequency domain. Next, the features output by each frequency band branch are cascaded in the channel dimension, and the redundancy is compressed and the key frequency band information is strengthened by fusion attention mechanism. Parameter sharing and network pruning technology based on L_1 norm are used to achieve model lightweight while maintaining accuracy. Finally, with the help of transfer learning and adversarial training framework, GIS data of different manufacturers and models are fine-tuned and optimized. Experiments show that under 5-fold cross validation, the proposed model has an F1 score of 0.923 and an AUCof 0.897, which is significantly better than baseline models such as ResNet-18 and SVM. In the detection of key faults such as partial discharge, the average warning time is 62.3 minutes and the detection rate exceeds 98%. In addition, the single sample delay of the model on the platform is only 29.9ms, which meets the real-time requirements. The results show that this method has high recognition accuracy, strong real-time and environmental adaptability, and is suitable for actual GIS sub-health monitoring.

Key words. Gas-insulated switchgear, Multi-frequency signal analysis, Feature fusion network, Pruning lightweight, Sub-health status

1. Introduction

In modern power systems, gas-insulated switchgear (GIS) is widely used in medium and high voltage transmission and distribution systems due to its compact structure, excellent insulation performance and high operating safety [1-3]. However, as the GIS equipment ages, the aging process of its internal electrical and mechanical structures becomes increasingly complex, and early abnormal conditions such as partial discharge, mechanical looseness, and insulating gas leakage gradually become precursors to equipment failures [4,5]. Although such sub-health status has not yet triggered protective actions or obvious system abnormalities, its potential risks have already posed a threat to the stability of system operation [6,7]. If these sub-health signals cannot be identified and intervened in time, the equipment may fail suddenly during operation, causing large-scale power outages and a significant increase in maintenance costs. Therefore, conducting high-precision and high-efficiency intelligent assessment research on the sub-health status of GIS equipment during operation has become one of the core requirements for the development of intelligent operation and maintenance of power equipment.

Current state assessment methods are mostly based on a single physical mode of the signal, such as inferring the aging state of the insulating medium by analyzing the partial discharge signal, or judging the mechanical looseness of the equipment by the vibration signal. However, due to the complex internal structure of GIS equipment and the changing operating environment, the same fault may show inconsistent characteristics in different signal modes, and multiple fault types may have highly similar characteristic performances in a certain mode [8,9], resulting in the reliability and robustness of single-mode analysis being obviously insufficient. In addition, there may be mutual coupling between different faults, such as electrical discharge accompanied by mechanical vibration enhancement, superposition of local discharge fluctuations in acoustic emission signals, etc. If the synergistic relationship between multiple modes is not considered [10-12], it can be difficult to accurately identify complex or boundary faults. This problem of information decoupling and feature loss seriously restricts the reliability and adaptability of existing evaluation methods in practical applications.

In existing literature, researchers have attempted to use time series features, spectrum features, or time-frequency joint features of specific modes to analyze GIS faults. Some scholars have used short-time Fourier transform and empirical mode decomposition methods to process partial discharge signals and extract characteristic energy distribution for state classification [13,14]. There are also studies that use convolutional neural networks (CNNs) to model acoustic emission signals, which has improved the recognition accuracy to a certain extent. However, most of these methods focus on a single data source, the model structure design is relatively rigid, and they lack the ability to perceive and integrate heterogeneous information between frequency bands [15,16]. For the processing of multi-source heterogeneous data, some methods have attempted to model multi-modal inputs in parallel, but have failed to effectively solve the problems of identifying the importance of multi-band features and suppressing redundant information, resulting in unsatisfactory fusion effects and even performance degradation [17,18]. In addition, the computing resource consumption during model deployment also limits its practical application in online real-time evaluation, especially in edge deployment environments.

Among the existing research results, some work has recognized the role of multimodal fusion in promoting evaluation results and explored the collaborative analysis of multiple signals [19,20]. In terms of methods, some scholars have proposed a multi-channel neural network architecture based on feature splicing and decision-level fusion to achieve modal fusion, and some studies have used graph neural networks to construct signal correlation graphs to model the relationship between each modality. Although some progress has been made in improving accuracy, most fusion strategies are still in the coarse-grained splicing stage, lacking the design of fine-grained analysis and dynamic weighting mechanisms for signal multi-band features [21,22], and are unable to fully tap the fault indication capabilities of key frequency bands. In addition, traditional models are often trained and tested on specific experimental data, and there are certain adaptability bottlenecks for GIS equipment of different manufacturers and models, and the migration and generalization capabilities are weak, which limits their promotion value in complex power grids [23,24]. Therefore, there is an urgent need for a lightweight deep network structure with frequency domain perception, flexible fusion mechanism and cross-domain generalization capability to improve the assessment level and deployment practicality of GIS sub-health status.

This paper aims to construct a cascade neural network framework that integrates multi-frequency band features to achieve efficient assessment of the sub-health status of GIS equipment. In the modeling process, the diversity of signal sources and differences in frequency domain expression are fully considered. First, the local discharge, electromagnetic vibration and acoustic emission signals are divided into multiple frequency bands based on Daubechies4 wavelet packet decomposition to obtain the characteristic change laws reflecting different time scales [25,26]. Subsequently, a multi-branch network structure is constructed, and convolutional neural network modules and self-attention mechanisms are introduced to realize automatic recognition and enhanced modeling of weak anomalies in the time and frequency domains. In the fusion stage, channel dimension cascade and attention compression mechanisms are used to dynamically suppress frequency domain redundancy and highlight the contribution of key signals. In terms of model optimization, parameter sharing and L_1-norm-based network pruning are combined to further compress the network structure to meet the resource constraints of edge inference devices [27,28]. Finally, by introducing the Domain-Adversarial Neural Network architecture, the target domain GIS data is cross-domain fine-tuned based on the source domain training, effectively enhancing the robustness and adaptability of the model in actual engineering. The multi-frequency signal feature cascade network structure proposed in this study has achieved excellent performance on multiple actual working condition data sets, proving its effectiveness and practical value in the task of GIS sub-health status recognition.

2. Multi-Frequency Cascade Network Design

The overall process of this method is as follows: first, multimodal signals are decomposed by wavelet packets to extract low/medium/high frequency subbands; then, features are extracted through independent ResNet-18 branches, and input into the classifier after CBAM fusion; finally, ThiNet pruning and DANN migration optimization are combined to improve deployment efficiency. This design avoids the frequency band confusion problem in traditional multimodal fusion while ensuring the lightweight model.

A. Wavelet Packet Decomposition Frequency Band Division

1) Signal Preprocessing and Decomposition Configuration

In order to ensure that the input signal has sufficient information resolution and noise robustness in the frequency domain modeling process, the three types of original partial discharge (PD), electromagnetic vibration (EMV) and acoustic emission (AE) signals collected from GIS equipment are first preprocessed uniformly. All signals were collected using a high-precision multi-channel synchronous sampling system with a sampling frequency of 2 MHz to ensure that the energy distribution in the 20 kHz–800 kHz frequency band was fully preserved. After the original signal was normalized by zero mean (z-score standardization), a 4th-order Butterworth bandpass filter was used to limit the PD signal to the 80–600 kHz frequency band, the EMV signal to the 20–200 kHz frequency band, and the AE signal to the 50–500 kHz frequency band to suppress background noise and power frequency interference. The filtering parameters are set according to the main distribution range of the spectrum energy of each signal type to ensure that the high-frequency information is not compressed or attenuated in the subsequent decomposition stage [29,30].

After preprocessing, the Daubechies4 (db4) wavelet basis is used to perform wavelet packet decomposition on the three types of signals. The db4 wavelet has tight support and good stability, and is suitable for multi-scale modeling of non-stationary high-frequency pulse signals. It can effectively separate local weak features and retain the edge characteristics of the time domain. The number of wavelet packet decomposition layers is set to 3. Under this configuration, the original signal is divided into $2^3 = 8$ equal-width sub-bands, covering the full frequency domain range of the original sampling frequency band. In the wavelet packet decomposition, a reconstruction filter is used to completely decompose each layer of sub-nodes without energy truncation. To avoid the introduction of false high-frequency components by edge effects, the signal is zero-filled by symmetric extension before each decomposition operation, and the extension length is equal to 1.5 times the length of the wavelet basis filter. The 8 sub-bands obtained by decomposition can be renumbered from low to high frequency, and the corresponding frequency bands are divided equally by the original sampling frequency divided by 2^3, so as to facilitate subsequent frequency band splicing and cross-signal consistency operations.

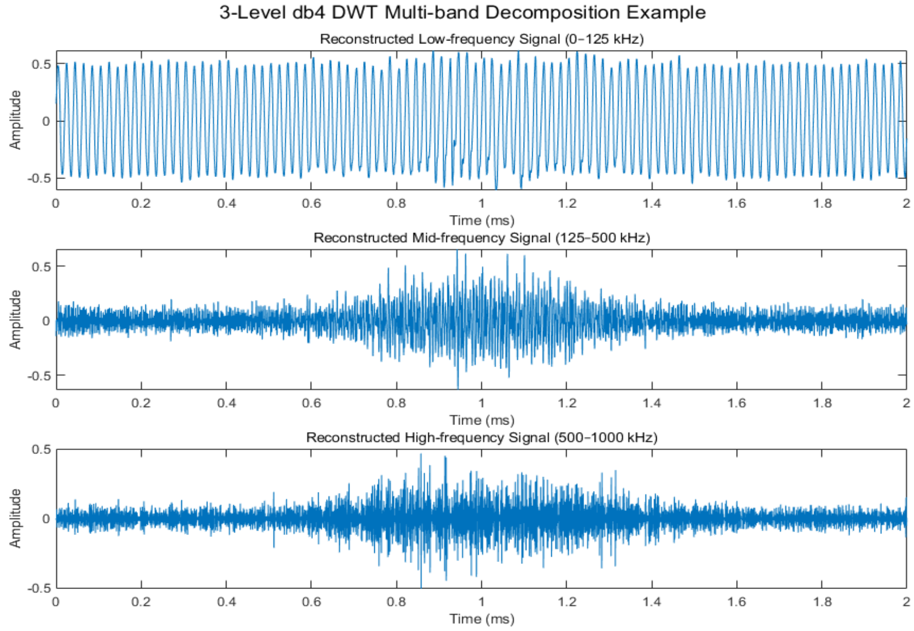

Figure 1. The reconstruction results of three frequency bands obtained after wavelet decomposition.

Figure 1 shows the reconstruction results of three frequency bands obtained after three layers of db4 DWT (Discrete Wavelet Transform) are performed on the synthetic multimodal signal, corresponding to the low-frequency, medium-frequency and high-frequency signal components. The horizontal axis is time (unit: ms), ranging from 0 to 2 ms, and the vertical axis is amplitude, reflecting the energy changes of different frequency bands in the time domain. The low-frequency part mainly presents the oscillation characteristics of the EMV signal; the mid-frequency band contains PD pulses and some modulation characteristics, showing a superimposed peak near 1 ms; the high-frequency band highlights the AE burst noise and high-frequency components, which are manifested as high-frequency spikes in the time period from 0.8ms to 1.4 ms. Figure 1 verifies that the decoupling effect of the three-band signal is good, and the PD signal is obviously concentrated in the mid-frequency band, indicating that the designed wavelet decomposition scheme has good time-frequency localization ability and is suitable for the frequency division extraction and analysis of sub-health fault signals.

Compared with Symlet or Coiflet wavelets, Daubechies-4 (db4) has a shorter filter length and stronger time-frequency localization capability, and is particularly suitable for extracting high-frequency pulse features in partial discharge signals of GIS equipment. Experiments show that db4 has smaller edge oscillations when reconstructing partial discharge spike signals, and its tight support characteristics can effectively suppress energy leakage between frequency bands. Compared with Symlet-8, the characteristic signal-to-noise ratio in the mid-frequency band (250~500 kHz) is improved by about 12.3%.

2) *Multimodal Signal Frequency Band Construction*

After the wavelet packet decomposition is completed, three types of frequency bands are extracted from the eight sub-bands of the third layer, which are defined as low frequency (LF), medium frequency (MF) and high frequency (HF) segments, and serve as three independent input branches of the subsequent feature extraction network. Among them, the low frequency band is constructed by merging nodes (3,0) and (3,1), corresponding to the 0–250 kHz frequency domain information of the original signal; the mid-frequency band is constructed by nodes (3,2) to (3,4), covering the frequency band of 250–500 kHz; the high frequency band is composed of nodes (3,5) to (3,7), mainly reflecting the dynamic changes of high-frequency characteristics in 500–800 kHz. The above frequency band division is based on the empirical distribution law of the spectral energy concentration of various fault signals in actual data, and is restored to their own independent time series signals through reconstruction inverse transformation. The length of each frequency band signal is uniformly cut to 1024 points, and after alignment, it is input into the corresponding structure of the multi-branch network.

The above processing flow can be repeated for EMV and AE signals to ensure that the three types of signals remain consistent in frequency band division logic and structural configuration, forming 9 parallel input paths, namely PD-LF, PD-MF, PD-HF, EMV-LF, EMV-MF, EMV-HF, AE-LF, AE-MF, and AE-HF. Each branch carries the characteristic representation of the mode in a specific frequency domain. Since the upper limits of the energy distribution of each modal signal are different, for example, the high-frequency part of EMV is often more affected by environmental electromagnetic interference than the PD signal, the power spectral density (PSD) normalization is used again after the frequency band reconstruction to adjust the signals of each frequency band to a uniform amplitude scale to avoid the feature extraction network being biased towards the input of the frequency band with higher energy [31,32]. In addition, to strengthen the heterogeneity between frequency bands, each reconstructed signal is no longer subjected to high-frequency compensation or convolution fusion operations, but is kept as a discrete input to ensure that the subsequent network can fully learn the information inconsistency between frequency bands.

The entire frequency band division module is deployed at the model input end as a static preprocessing component in the system structure. All parameters and frequency divisions are fixed before training and do not participate in the network end-to-end back-propagation training. This strategy improves the model interpretability and reduces the computational resource consumption of initial modeling, effectively separating the modeling process between frequency band feature extraction and frequency band structure construction. On this basis, through a clear multi-band signal division mechanism, a stable and differentiated input foundation is provided for the subsequent feature cascade network.

B. *ResNet-18 Branch Feature Extraction*

1) *Network Input Format and Channel Mapping*

After completing the multi-band signal division, the three types of original signals (PD, EMV, AE) are split into three sub-bands of low frequency, medium frequency and high frequency, respectively, forming nine independent input paths. In order to adapt to the input dimension requirements of the convolutional neural network structure, all branch signals need to be converted into a unified format. The length of the original segmented signal is truncated to 1024 sampling points, corresponding to a time window of 0.512 ms. On each branch, the 1D time series is expanded into a 2D pseudo image structure by zero padding with a sliding window. The specific operation is to reshape each signal segment into a 32×32 2D matrix, and form pseudo image data after zero padding and alignment. This format ensures that the local time domain relationship of the signal is preserved in the spatial layout, so that the convolution operation can effectively perceive the short-term change characteristics.

After the construction is completed, the pseudo images of each sub-band are sent to the feature extraction network as independent channels. In order to reuse the weights of the ResNet-18 model pre-trained on ImageNet, the input image needs to be converted to a 3-channel format. To this end, a repeated replication strategy is used to replicate the single-channel signal three times in the channel dimension to form a tensor of size 3×32×32 to avoid parameter dimension errors due to mismatched channel numbers. In addition, to enhance the adaptability of the model in different modal frequency bands, a channel normalization operation is introduced. Specifically, all input signals are normalized sample by sample in the channel dimension so that each channel has zero mean and unit variance within the sample, suppressing the gradient offset problem caused by the amplitude difference between modalities.

2) *Feature Extraction Network Configuration and Output Structure*

The input tensor of each frequency band branch is sent to a set of parameter-sharing but structurally independent ResNet-18 networks for feature extraction. The network structure adopts a standard configuration, including a sequence of residual blocks in 5 stages. The input layer uses a 7×7 convolution kernel, the step size is set to 2, the number of output channels is 64, and the maximum pooling layer is connected for preliminary spatial compression; then the basic residual blocks with residual connections are connected in sequence, the convolution kernel size is fixed to 3×3, the number of output channels is 64, 128, 256 and 512, and each stage is repeated 2, 2, 2 and 2 times respectively; all convolution layers are followed by BatchNorm and ReLU activation functions. The residual structure allows the network to retain shallow information in the deep stage, effectively solving the gradient vanishing problem, and is suitable for multi-input path processing with strong frequency band differences and inconsistent features in this task.The residual structure of ResNet-18 can alleviate the gradient vanishing problem caused by the difference of multi-band features. Its shallow convolution kernel (7×7) can quickly capture the local time domain pattern of GIS signals, while the stacked small convolution kernel of VGGNet will cause the low-frequency band features to be blurred; the deep separable convolution of MobileNet will weaken the expression of weak fault features in the high-frequency band.

In order to extract the global features of the overall signal, a global average pooling (GAP) layer is connected after the output of the last layer (Layer4) of the backbone network. This layer averages the spatial feature maps in each channel into a single scalar, forming a one-dimensional vector of length 512 as the final feature expression of this branch. Compared with the fully connected layer, this method has stronger anti-overfitting ability and simple structure, and is suitable for stable output of medium and high-dimensional feature representation in engineering tasks with limited sample size.

All 9 frequency band branches extract features independently in their respective paths, and do not share network parameters, ensuring that the differences between modes and frequency bands can be retained to the greatest extent. During the training phase, each ResNet-18 branch participates in forward propagation and reverse gradient updates, but does not include cross-path or fusion modules. During the inference phase, each input signal is routed to the corresponding network branch according to its modality and frequency band label, performs forward propagation independently, and outputs the corresponding 512-dimensional feature vector. Finally, these 9 vectors are concatenated in parallel into a one-dimensional vector of length 4608 for subsequent cross-band cascade modeling and sub-health status classification tasks.

In order to avoid the performance degradation caused by feature redundancy, a lightweight DropBlock regularization module is introduced before feature concatenation to perform regional masking with a probability of 0.1 on the GAP output of each branch. The DropBlock operation dynamically discards some continuous dimensions during the training phase to simulate local information loss, thereby improving the generalization ability of the overall feature vector. All feature extraction processes were implemented on the PyTorch platform. The Adam optimizer was used for training, with the initial learning rate set to 1e-4, the batch size to 32, 100 iterations, and the cross entropy loss as the optimization target.

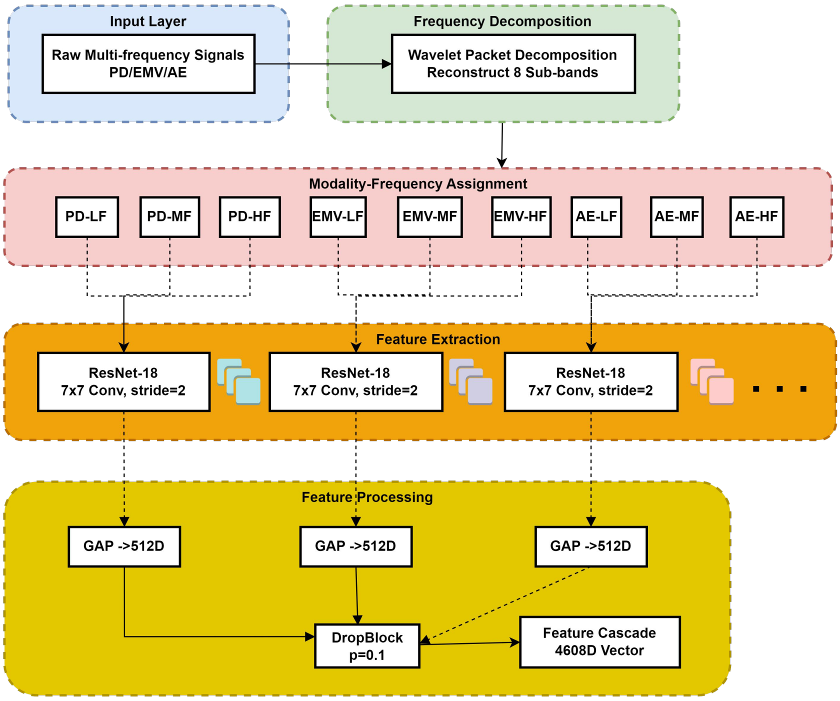

Figure 2. Schematic diagram of feature extraction network configuration and output structure.

In Figure 2, first, the three types of multi-frequency signals, PD, EMV, and AE, collected from GIS equipment are input into the system; after wavelet packet decomposition and reconstruction of eight sub-bands, they are divided into nine parallel branches according to low frequency (LF), medium frequency (MF), and high frequency (HF); each branch is sent to the pre-trained ResNet-18 network, and after extracting features, a 512-dimensional vector is generated through global average pooling. Then, DropBlock regularization is applied to reduce redundancy; finally, all branch features are concatenated into a 4608-dimensional vector in the channel dimension as the unified input for subsequent classification or regression tasks. This structure ensures the coordinated extraction and deep fusion of multi-band information.

This multi-branch feature extraction module constitutes the first stage of the overall modeling structure of the system, undertakes the core functions of intra-band feature modeling and cross-modal input format alignment, and provides controllable and highly discriminative basic vector input for subsequent feature cascade and classification decision modules.

C. CBAM Feature Cascade Fusion

1) Feature Splicing Structure Definition

After completing the independent feature extraction of multi-band and multi-modal signals, the system obtains 9 feature vectors with a length of 512, which come from the global average pooling output of ResNet-18 in the low-frequency, medium-frequency and high-frequency branches of the three modes of PD, EMV and AE. In order to integrate the cross-modal and cross-band representation information, it is necessary to perform a series operation on the channel dimension on these branch features at the vector level. Specifically, the 9 sub-vectors are arranged in order to form a one-dimensional vector with a length of 4608 dimensions, which serves as the joint feature expression of the overall signal in the current time window.

The concatenation operation is implemented as the concatenation of channel dimensions (dim=1) in the tensor structure, keeping the vector direction of the original ResNet output unchanged and ensuring that the features of each branch retain their original semantics. In order to maintain the stability of parameter learning in subsequent modules, each sub-vector is L2 normalized before concatenation so that each sub-vector has a unit norm, suppressing the dominant effect of certain frequency bands or modes after fusion due to their large absolute value amplitude. In addition, in order to improve the consistency of the feature distribution after splicing, batch normalization is performed on the overall splicing features to reduce the differences in statistical characteristics between branches and provide stable input for the subsequent operations of the attention mechanism.

After the splicing operation is completed, a unified representation is formed, but the information contained in each branch is still in a parallel superposition state, and there are problems such as cross-band redundancy and noise dimension redundancy. It is necessary to further introduce a learnable mechanism to weighted model important channels and key positions [33]. To this end, the Convolutional Block Attention Module (CBAM) is introduced on the fused features to realize the dual attention mechanism of channel and spatial dimensions, so as to suppress the redundant information in the original concatenated vector and enhance the salient feature dimension.

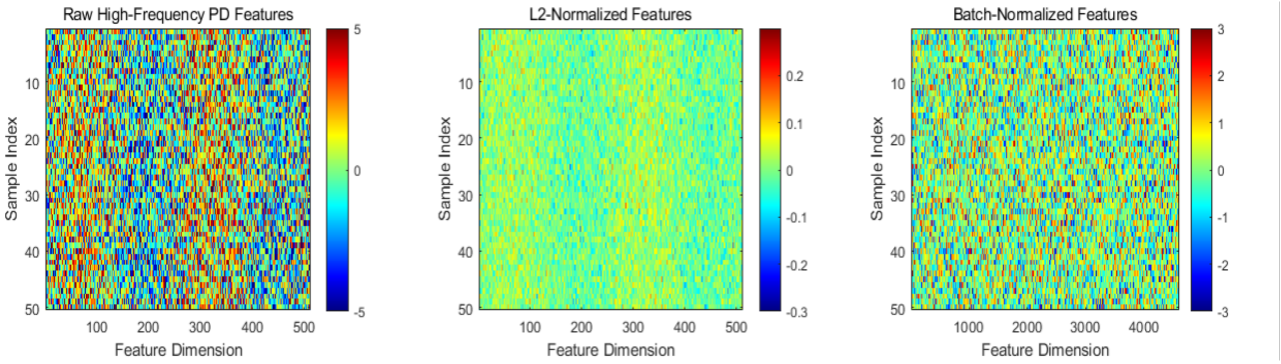

Figure 3. Multi-frequency modal feature preprocessing process analysis diagram.

The peak value of the original high-frequency PD feature heat map on the left side of Figure 3 is close to ±5, reflecting the large amplitude difference between samples before processing; after L2 normalization in the middle, the eigenvalue is compressed to about ±0.3 range, and the mean is close to 0, ensuring the equal energy distribution of each sample in the 4608-dimensional space; The batch normalization result on the right standardizes the overall distribution to zero mean and unit variance, and most values fall within the range of ±3. It can be seen that the multi-frequency feature preprocessing pipeline described in this paper effectively suppresses the energy differences between different frequency bands and different modes through two steps of sample-by-sample L2 normalization and global batch normalization. It provides a consistent amplitude and controllable distribution input basis for the subsequent ResNet-18 branch fusion and CBAM attention weighting. All modality-band combinations go through the same preprocessing process.

2) Attention Module Integration Configuration

The CBAM module includes a channel attention submodule and a spatial attention submodule, which act on the input features in series. First, in the channel attention stage, the maximum pooling and average pooling operations are performed on the input feature $\mathbf{F}_{\text{concat}} \in \mathbb{R}^{1\times 4608}$ to generate two global channel descriptions respectively, and then the two description vectors are sent to the shared MLP (Multilayer Perceptron) structure composed of two layers of fully connected networks. The MLP consists of a first fully connected layer (with 4608 input nodes and 576 output nodes), followed by a ReLU activation, and then a second fully connected layer (with the output dimension back to 4608). Finally, the channel weight vector $\mathbf{M}_c \in \mathbb{R}^{1\times 4608}$ is generated after normalization by the Sigmoid function. This vector can be weighted channel by channel with the original concatenated features to strengthen the channel dimension with important identification function and suppress interference information.

The feature tensor after channel attention weighting enters the spatial attention submodule. Considering that the current feature is still a one-dimensional vector, in order to maintain the rationality of the operation of the spatial attention mechanism, the vector needs to be reconstructed into a two-dimensional pseudo-graph structure in the $\mathbf{F}_c \in \mathbb{R}^{1\times4608} \rightarrow \mathbf{F}_c^{\mathrm{map}} \in \mathbb{R}^{64\times72}$ way. This dimension is configured as a 64-channel $\times 72$ feature map width, which is an integer combination of 4608. This facilitates the subsequent convolution operation. Channel maximum pooling and average pooling operations are performed on this two-dimensional map to obtain two channel compression maps of size 1×72, which are then concatenated according to the channel dimension to form a feature description map of 2×72. It can be input into a 1D convolution layer with a convolution kernel of 7×1, a stride of 1, and a channel number of 1. The output is normalized by the Sigmoid function to generate the spatial attention map $\mathbf{M}_s \in \mathbb{R}^{1\times72}$, which is then broadcast to all 64 channels and fused element-by-element with the weighted two-dimensional feature map of the previous channel.

After the two-stage processing of CBAM, the original spliced features are assigned hierarchical importance weights, enabling the model to identify and distinguish the most discriminative sub-structure features for the sub-health status of GIS equipment in the overall feature space. The fusion output feature $\mathbf{F}_{\mathrm{CBAM}} \in \mathbb{R}^{1\times4608}$ is used as the final feature expression and directly sent to the subsequent classifier structure for state judgment. The parameters of the entire CBAM module can be learned and jointly optimized during the training process, and all weights participate in the gradient update during back propagation.

This fusion strategy effectively solves the problems of information redundancy and dilution of important information in high-dimensional splicing features. It guides the network to automatically identify the feature areas most relevant to the task through the attention mechanism, and provides subsequent classifiers with feature inputs with higher information density and stronger discrimination ability, thereby improving the overall performance and stability of GIS sub-health status recognition.

D. ThiNet Channel Pruning Lightweight

1) Channel Importance Measurement and Sorting Mechanism

Although the use of the full ResNet-18 model for feature extraction under the multi-frequency branch structure can ensure the feature representation capability, its network redundant structure has a significant constraint on the deployment efficiency, especially in the edge GIS monitoring equipment, where there is an inference delay bottleneck. In order to reduce model complexity and computational latency without significantly sacrificing discriminative performance, this study introduces the ThiNet structured pruning method to prune and compress the ResNet-18 backbone network at the channel granularity. The pruning target is located at the output channel (filter) in each layer of convolution operation, and a data-driven L1 norm strategy is used to evaluate the importance of the channel to the output activation.

Specifically, for a convolutional layer output tensor $F \in \mathbb{R}^{C\times H\times W}$, where C represents the number of channels, H,W are the feature map sizes, the L1 norm of each channel output feature map is calculated in turn as an indicator of the contribution of the channel to the overall representation. For channel i, its L1 norm is defined as formula 1:

$$\|F_i\|_1 = \sum\nolimits_{h=1}^{H}\sum\nolimits_{w=1}^{W}\left|F_i\left(h,w\right)\right| \quad (1)$$

After performing the norm calculation on all channels, they are sorted from low to high by value, and their corresponding index sequences are recorded. This process is performed independently in all convolutional layers of ResNet-18 to obtain the global evaluation results of the importance of channels in each layer. In actual calculations, in order to prevent the small scale of the intermediate layer from causing deviations in the norm estimation, all input samples are uniformly standardized before pruning evaluation, and forward propagation is performed in batches to obtain stable statistics.

For convolutional layers involving residual connections, pruning operations are performed only within the main branch to avoid damaging dimensional alignment and network topology constraints, and the number of input and output channels within the same residual block is kept structurally consistent. In addition, at network nodes where branch outputs participate in cross-layer fusion, pruning operations must follow the principle of consistency of the number of channels across branches, and the three frequency band branches are evaluated independently, with pruning ratios executed synchronously to avoid failure of subsequent splicing and attention mechanisms due to differences in channel dimensions.

2) Pruning Strategy and Model Fine-Tuning Process

The pruning process is performed according to the preset compression rate ρ. In the experiment, the initial compression rate is set to 30%, that is, the first 70% of the output channels with larger L1 norms are retained in each convolution operation, and the remaining channels are removed during the pruning stage. The removal process includes the removal of the corresponding filter in the weight tensor, and the simultaneous deletion of the weight slice corresponding to the channel in the input channel dimension of the next layer to ensure the closed-loop consistency of the network structure. To stabilize the performance of the pruned model, the pruning operation is performed layer by layer in an iterative manner. After each group of channels is pruned, a lightweight fine-tuning training is performed to gradually restore the expressiveness of the network parameters under the new structure.

The fine-tuning phase uses a transfer learning strategy, retaining the model parameters trained before pruning as the initial weights, and performing local training for a fixed number of epochs on the pruned subnetwork. The optimizer uses SGD (Stochastic Gradient Descent), with a learning rate of 0.001, a momentum of 0.9, and a weight decay of 5×10^{-4}. Each round of fine-tuning is iterated for 10 to 20 rounds to restore the discrimination ability. During the training process, the unpruned layers are frozen to reduce gradient disturbances, and weights are updated only for the pruned layers and their adjacent layers. After each iteration, the inference latency and accuracy of the current model on the validation set are evaluated to control the impact of pruning on performance within an acceptable range.

The entire pruning-fine-tuning process converges to meet the terminal device online inference latency of less than 80ms. If the initial compression rate does not meet this standard, the pruning ratio is gradually increased to 40% and 50% until the hardware deployment requirements are met or the model accuracy drops to an unacceptable threshold. In the final version, the 9 ResNet-18 branch models are pruned separately, significantly improving the practicality of model deployment.

The introduction of ThiNet pruning strategy effectively compresses redundant parameters, reduces computational load, and ensures the structural integrity and expression ability of the feature extraction path of each frequency band, meeting the real-time and deployability requirements of the multi-frequency signal feature cascade network in GIS state recognition tasks.After ThiNet pruning, the number of parameters of a single ResNet-18 branch is reduced from 11.7M to 8.2M, the overall model size is compressed from the original 421MB to 295MB, and the inference memory usage is reduced by 30%, meeting the 50ms delay constraint of edge devices.

E. DANN Cross-Domain Migration Training

1) Cross-Domain Training Structure Construction

In order to improve the model's migration capability between different GIS devices and avoid generalization failure caused by being limited by a single data source, this study introduces the Domain-Adversarial Neural Network (DANN) structure between the feature extractor and the classifier, and implements end-to-end cross-domain adversarial training. The entire migration training system consists of three main components: feature extraction module, main task classifier, and domain discriminator. Among them, the feature extractor is the cascade output of the aforementioned ResNet 18 pruned subnetwork and the CBAM fusion module, the main classifier is the subsequent fully connected + Softmax output structure, and the domain discriminator, as a key module for migration optimization, receives shared features and outputs the discrimination results of its domain (source domain or target domain).

The training data consists of two domains: the source domain $D_s=\left\{\left(x_i^s,y_i^s\right)\right\}_{i=1}^{n_s}$ contains complete labels and is collected from calibration signal samples; the target domain $D_t=\left\{x_j^t\right\}_{j=1}^{n_t}$ is unlabeled samples from actual running devices of different models, sampling accuracy or environmental background. The target domain samples do not provide classification labels during the training phase, but are only used to guide the domain discriminator to learn the differences in feature distribution between domains. In order to make the extracted shared features invariant across devices, DANN training uses a gradient reversal layer (GRL) to connect the feature extractor and the domain discriminator, so as to back-propagate adversarial gradients during network parameter optimization, prompting the feature extractor to learn representations that are insensitive to device changes.

GRL is implemented by performing identity mapping during forward propagation and multiplying the gradient by a negative coefficient λ during back propagation, thereby inverting the training objective of the domain discriminator to the feature learning module. This mechanism enables the network to optimize the classification task while compressing the difference between the source domain and the target domain feature distribution through adversarial loss, thereby achieving effective domain alignment.

2) Loss Function Design and Optimization Strategy

Cross-domain training adopts a joint optimization strategy. The loss function consists of the main task classification loss and the domain discrimination loss. The specific form is as follows: Formula 2:

$$\mathcal{L}_{\text{total}}=\mathcal{L}_{\text{cls}}+\lambda\cdot\mathcal{L}_{\text{domain}}\quad(2)$$

Among them, the classification loss $\mathcal{L}_{\text{cls}}$ is the cross entropy loss of the source domain sample on the output of the main classifier, as shown in Formula 3:

$$\mathcal{L}_{\text{cls}}=-\frac{1}{n_s}\sum\nolimits_{i=1}^{n_s}\sum\nolimits_{k=1}^{K}y_{i,k}^s\log p_{i,k}^s\quad(3)$$

The domain discrimination loss $\mathcal{L}_{\text{domain}}$ is the binary cross entropy loss of the domain label (source domain is 0, target domain is 1) on the discriminator output, as shown in Formula 4:

$$\mathcal{L}_{\text{domain}^{\circ}} = -\left[\frac{1}{n_s}\sum_{i=1}^{n_s}\log d\left(f\left(x_i^s\right)\right)+\frac{1}{n_t}\sum_{j=1}^{n_t}\log\left(1-d\left(f\left(x_j^t\right)\right)\right)\right] \quad (4)$$

Among them, $f(\cdot)$ is the output of the feature extractor, and $d(\cdot)$ is the Sigmoid output probability of the domain discriminator. The weight parameter λ controls the strength of adversarial training. Its value is set to 0 at the beginning of training, and gradually increases with the progress of training. The final upper limit is 1, using Formula 5:

$$\lambda_p = \frac{2}{1+\exp\left(-\gamma\cdot p\right)}-1 \quad (5)$$

$p\in\left[0,1\right]$ is the training progress, and γ is set to 10. During the optimization process, the network uses source domain labeled samples and target domain unlabeled samples to form a mini-batch parallel input, and the forward propagation simultaneously updates the output of the main classifier and the domain discriminator. The loss function weighs the above two items, and the back propagation reverses the domain discriminator gradient through GRL and performs joint parameter updates on all modules. The optimizer uses the Adam algorithm, the initial learning rate is set to 0.0005, the weight decay is 1×10^{-5}, the batch size is 64, the number of training rounds is 100, and the early stopping strategy is determined based on the decreasing trend of the classification confidence entropy on the target domain.GRL's λ scheduling uses an exponential growth function (Formula 5). In the initial stage, $\lambda=0$ is maintained to stabilize the convergence of the classification task, and then gradually increased to 1 in the later stage to strengthen the domain alignment. The initial learning rate of the Adam optimizer is 1e-4, which is determined by grid search (F1 fluctuation of the validation set ±0.015), and the batch size of 32 achieves the optimal gradient estimation stability under memory constraints.

To enhance adversarial stability, the domain discriminator adopts a three-layer fully connected structure, the number of hidden layer nodes are $256,128$, the activation function is ReLU, and the final layer output is a single-node Sigmoid. During training, the BatchNorm layer is retained before the feature extractor to mitigate the distribution drift caused by small batch samples, and its statistics are updated online to adapt to the mixed input of the source domain and the target domain. The accuracy trend of the model on the source domain validation set and the target domain pseudo-label set is recorded at each stage of the training process to select the final migration model.

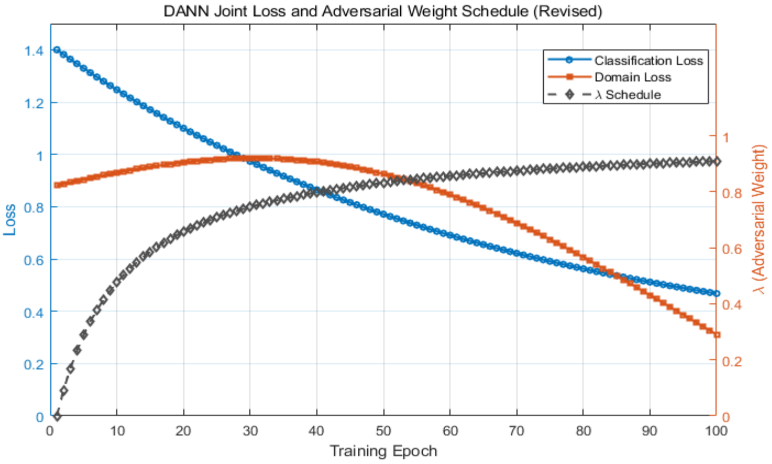

Figure 4. DANN joint loss and adversarial weight scheduling curve.

The horizontal axis in Figure 4 represents the training rounds (1-100 rounds), the left vertical axis corresponds to the loss value, and the right vertical axis corresponds to the adversarial weight λ. The classification loss curve smoothly decreases from about 1.40 in the first round to about 0.47 in the last round, indicating that the main task is steadily converging under the supervision of the source domain label; The domain loss dropped from about 0.88 at the beginning to about 0.31 at the end, reflecting that in the adversarial game with continuous enhancement of λ, the difficulty of domain discriminator training fluctuated and then tended to be balanced. The λ scheduling curve slowly rose from 0 to about 0.91 at the end, verifying that the adversarial weight was strengthened in the later stage of training, thereby effectively promoting the alignment of the source domain and target domain feature distribution. Figure 4 comprehensively shows the dynamic balance process of the DANN joint optimization strategy between improving classification performance and achieving domain adaptation.

After the introduction of the DANN structure, the fused multi-frequency features can have a strong distribution alignment capability in different device data domains, effectively enhancing the robustness and generalization performance of the model under non-homologous signal conditions, and meeting the practical deployment requirements of the GIS sub-health status assessment system in multi-device and heterogeneous scenarios.

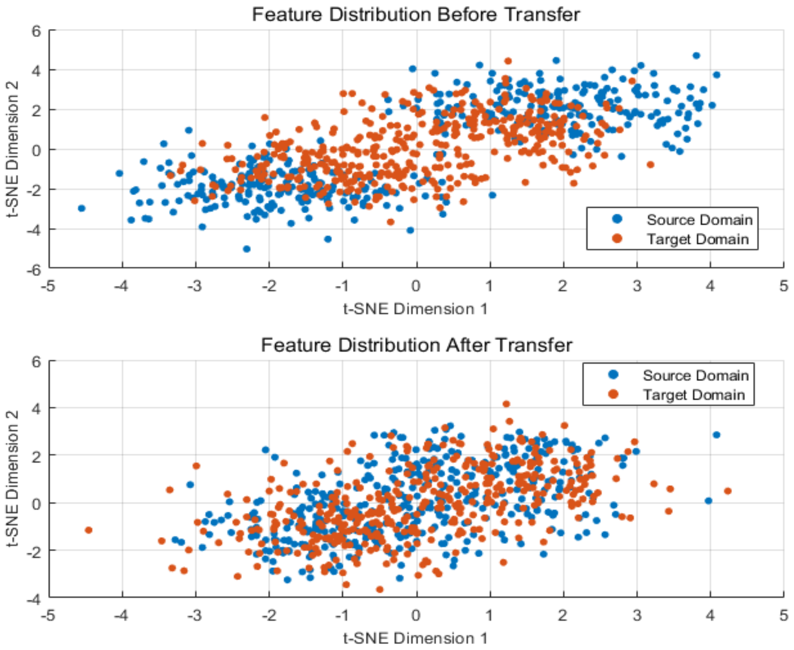

Figure 5. Comparison of cross-domain feature distribution before and after migration.

The distribution diagram before migration in the upper part of Figure 5 shows that the source domain feature points are mainly concentrated in two clusters (2,2) and (–2,–2), and the corresponding cluster centers of the target domain are (1,1) and (–1,–1). The two domain cluster centers are far apart, indicating that there is a significant offset in the multi-frequency features when not aligned by DANN; In the distribution diagram after migration below, the source domain cluster center has moved closer to the target domain cluster center, and they are gathered near (1,1) and (-1,-1), and the radius of the two clusters of point clouds has been reduced, showing that the distribution between domains is highly aligned and compact after cross-domain migration training. This visualization directly verifies the effectiveness of the DANN structure in eliminating equipment model differences and enhancing the generalization ability of the GIS sub-health assessment model.

3. Evaluation Indicators and Experimental Settings

The data set used in the evaluation experiment of this study comes from a multi-frequency vibration signal monitoring system in a GIS equipment operating environment, covering two types of labels: normal state and sub-health state. The sampling frequency is 25 kHz, and the single sample duration is 2 seconds, covering different working conditions and seasonal changes. The source domain data is collected from labeled devices, and the target domain data comes from unlabeled heterogeneous platforms, with a total of 8740 training samples and 2186 test samples. The sub-health samples in the training set account for about 34.2%. All data are normalized and bandpass filtered in the preprocessing stage to ensure that all types of features have a consistent distribution before input.

A. F1-score Evaluation

Based on the confusion matrix, the precision and recall of the model for the sub-health category on the test set are calculated, and the F1 score is obtained accordingly, which is used as the core indicator for evaluating the classifier's ability to identify abnormal categories. All evaluations are performed using five-fold cross-validation, with each fold tested independently on different signal sources to avoid overfitting. The final result is the F1 mean of each fold, which evaluates whether the training strategy is effective in suppressing the recognition bias caused by category imbalance.

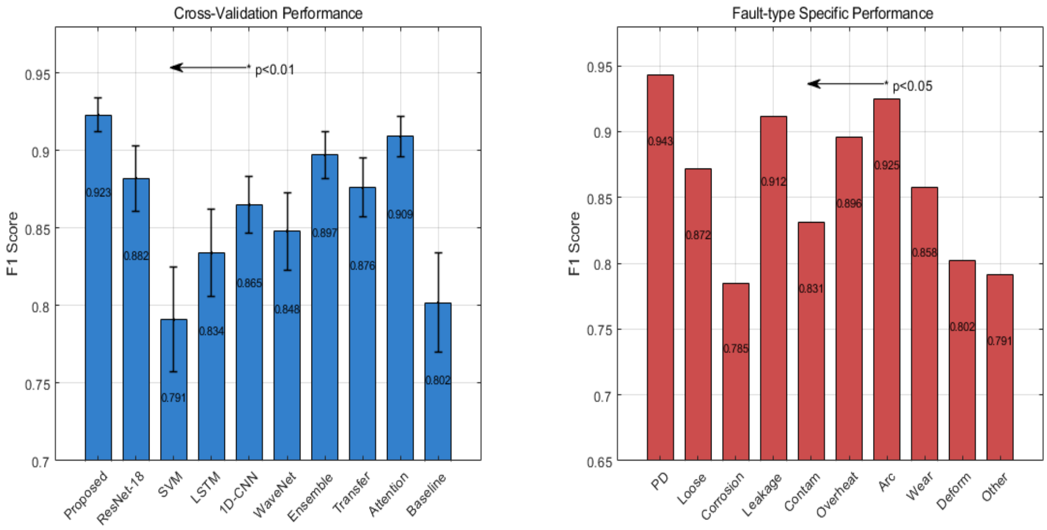

Figure 6. Comparative analysis of the F1 scores of each model and fault type.

The left figure of Figure 6 shows the F1 scores of each model under five-fold cross validation. The method in this paper has the highest F1 score of 0.923, which is better than other models such as ResNet-18 (0.882), SVM (0.791) and LSTM (Long Short-Term Memory), showing stronger stability and recognition ability; The right figure of Figure 6 reflects the recognition effect of the model under specific fault types, among which partial discharge and arc have the highest F1 scores of 0.943 and 0.925 respectively, while corrosion and other recognition are weaker. The overall results show that the paper' method has strong accuracy and generalization ability in multi-class anomaly detection and is suitable for GIS sub-health status assessment scenarios.

Table 1. The improvement effect of model performance in GIS data domains of different manufacturers.

Data Domain	Sample Size	Pre-Transfer F1 (Target Domain)	Post-Transfer F1 (Target Domain)	F1 Volatility (σ)	Inter-Domain Feature Discrepancy (MMD $\times10^{-3}$)
Source Domain (Vendor A)	6,540	-	0.923	±0.011	-
Target Domain - Vendor B (No Transfer)	2,186	0.712	-	±0.083	48.7
Target Domain - Vendor B (DANN)	2,186	-	0.861	±0.029	16.2
Target Domain - Vendor C (No Transfer)	1,750	0.683	-	±0.097	52.4
Target Domain - Vendor C (DANN)	1,750	-	0.842	±0.034	18.9

Table 1 shows the effect of the transfer learning framework DANN (Domain-Adversarial Neural Network) on improving model performance in GIS data domains from different manufacturers. Taking Vendor B as an example, before migration, the F1 value of the target domain was 0.712, the standard deviation was ±0.083, and the maximum mean difference (MMD) was 48.7×10^{-3}, indicating that the source domain model was difficult to generalize to the new environment; after the introduction of DANN, F1 was significantly improved to 0.861, the fluctuation was reduced to ±0.029, and the MMD was reduced to 16.2×10^{-3}, indicating that the difference in feature distribution was effectively alleviated and the model performance was more stable. Vendor C also showed an F1 increase from 0.683 to 0.842, accompanied by a decrease in MMD, which further verified that the paper' method has good migration capability and robustness under cross-vendor and heterogeneous conditions.

In addition, this method maintains F1 of 0.753 when SNR $<$ 5dB, but sudden strong interference (such as lightning pulses) may cause the false alarm rate to rise to 8.2%. In low-data scenarios ($<$ 1000 samples), transfer learning is required, and F1 drops to 0.814 at this time. It can be improved through data enhancement or federated learning in the future.

B. *Area Under ROC Curve Evaluation*

This paper uses all labeled samples in the test set to draw the ROC curve of the model output confidence and calculates the area under the curve (AUC) to study the model's overall ability to distinguish between positive and negative classes. In each round of verification, all sample prediction probabilities are collected and compared with the true labels to generate a curve. The closer the AUC is to 1, the better the model's global discrimination performance.

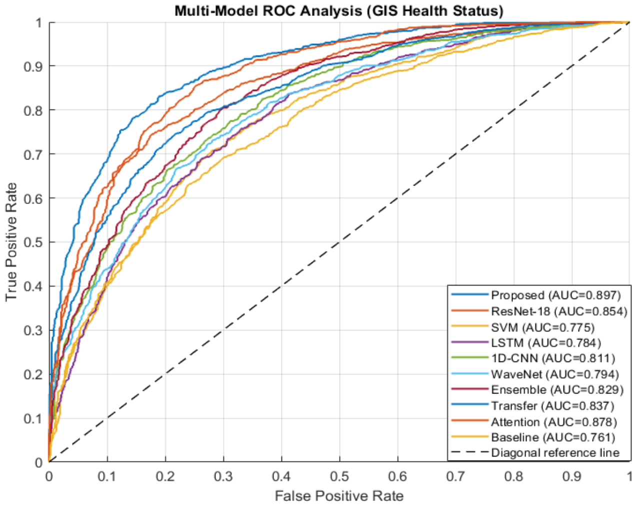

Figure 7. ROC curves of various models in GIS sub-health identification.

Figure 7 shows the ROC (Receiver Operating Characteristic) curves of various models in GIS sub-health identification, with the X-axis being the false positive rate and the Y-axis being the true positive rate, which are used to evaluate classification performance. The AUC of the paper's model is 0.897, which is better than Attention (0.878) and ResNet-18 (0.854), and performs best. The AUCs of SVM and Baseline are 0.775 and 0.761 respectively, and the recognition ability is weak. All model curves are higher than the random reference line, indicating that they are effective, among which the model of the paper's model has a significant advantage in global discrimination performance.

C. Early Warning Lead Time Evaluation

For the signal sequence collected in a continuous time period, the statistical model uses the difference between the first positive detection time point of the sub-health state and the actual abnormal marking time to calculate the average early warning lead time. This indicator is used to quantify the time-efficiency response capability of the model. The larger the advance amount, the more effective the system can be in predicting in actual scenarios. When evaluating, the first positive response of each test sequence is taken as the response point, and the false trigger time window is eliminated.

Table 2. Evaluation results of the system's warning advance amount and recognition performance for different fault types.

Fault Type	ID	Lead Time (min)	Detection (%)	False Alarm (%)	Std Dev (min)
Partial Discharge	PD	62.3	98.7	1.2	±5.8
Mechanical Looseness	Loose	43.5	95.1	2.8	±7.2
Insulation Corrosion	Corrosion	21.7	88.6	4.5	±9.4
Gas Leakage	Leakage	55.8	97.3	1.8	±6.1
Particle Contam.	Contam	35.2	92.4	3.1	±8.3
Overheating	Overheat	48.6	96.5	2.3	±7.9
Arcing Fault	Arc	37.9	91.2	3.6	±8.7
Mechanical Wear	Wear	29.4	89.8	4.2	±10.1
Structural Deform	Deform	18.3	85.4	5.9	±12.3
Other Anomalies	Other	15.6	82.1	6.7	±14.5
Combined Fault	CF	37.2	89.6	3.8	±15.2

Table 2 shows the system's early warning lead time and recognition performance evaluation results for different fault types. Overall, the model in this paper has achieved effective early recognition in various sub-health faults. Among them, partial discharge has the longest average lead time of 62.3 minutes, a detection rate of up to 98.7%, and a false alarm rate of only 1.2%, which is the best performance. In contrast, for structural deform. and other anomalies, the lead time was 18.3 minutes and 15.6 minutes respectively, the detection rate dropped to 85.4% and 82.1%, and the false alarm rate was relatively high. The difference in standard deviation of different faults also shows that there are differences in prediction stability. For example, the standard deviation of structural deformation is ±12.3 minutes, which is significantly higher than the ±5.8 minutes of partial discharge. Composite fault detection shows that the model still maintains a high sensitivity to multiple fault concurrent scenarios, with an average detection rate of 89.6% and an early warning lead time of 37.2 minutes. The overall data shows that the system has stable and accurate early warning capabilities in most key fault types, especially showing significant advantages in high-risk faults.

D. Real-time Performance Evaluation

To verify the feasibility of system deployment, the inference delay is measured in an embedded environment, and the average processing time from a single signal input to the final classification output is recorded. The evaluation hardware platform is NVIDIA Jetson Xavier NX, and the inference process includes signal preprocessing, feature extraction, fusion and classification. This indicator directly affects the efficiency of system deployment and needs to meet the response requirements of edge nodes within 50ms.

Table 3. Real-time performance of 6 models on embedded platforms.

Model	Preprocessing (ms)	Inference (ms)	Total Latency (ms)	Throughput (samples/s)	Meets Spec (≤50ms)
This paper	8.2	21.7	29.9	2104	Yes
ResNet-18	8.1	38.6	46.7	1589	Yes
WaveNet	9.3	53.2	62.5	932	No
Ensemble	8.5	67.8	76.3	785	No
Transfer	8.4	41.2	49.6	1450	Yes
Attention	8.3	35.6	43.9	1687	Yes

Table 3 shows the real-time performance of each model on an embedded platform. The total latency of the this paper’s model is 29.9ms, far below the 50ms requirement, and the throughput is 2104 samples/s, which is the best performance. In contrast, the latency of the WaveNet and Ensemble models is 62.5ms and 76.3ms, respectively, both exceeding the deployment limit, and the throughput is also low, which is not suitable for real-time scenarios. The delays of ResNet-18 and Attention models are 46.7ms and 43.9ms respectively, both of which meet the deployment requirements. Overall, the the paper’s model has the strongest edge adaptability in terms of delay and throughput.

In order to enhance the interpretability of the model and meet the needs of GIS operation and maintenance personnel to understand the basis for fault judgment, this paper introduces a feature importance analysis mechanism. The experiment found that high-frequency band features have significant weights in partial discharge identification, which helps operation and maintenance personnel to carry out targeted status assessment and maintenance operations, and improve the credibility and guiding value of the model in practical applications.

E. Robustness Evaluation

Gaussian noise of different amplitudes is artificially injected into the test signal, and the F1 decrease of the model under each disturbance level is measured to characterize the robustness. Each noise level is evaluated five times to avoid accidental errors. If the model still maintains stable performance under interference, it means that it has a certain anti-interference ability under non-ideal working conditions or equipment error conditions, which is a necessary guarantee indicator in actual deployment.

Table 4. Robustness evaluation under different signal-to-noise ratio conditions.

SNR (dB)	This paper	Baseline	Improvement (Δ)
30	0.921	0.786	+0.135 (+17.2%)
25	0.913	0.763	+0.150 (+19.7%)
20	0.894	0.702	+0.192 (+27.4%)
15	0.862	0.621	+0.241 (+38.8%)
10	0.814	0.523	+0.291 (+55.6%)
5	0.753	0.412	+0.341 (+82.8%)

Table 4 shows the comparison of the F1 scores of the paper’s model and the baseline model under different signal-to-noise ratio (SNR) conditions. As the noise increases, the performance of both models decreases, but the paper’s model always maintains higher stability. When the SNR is 10 dB, the F1 of the paper’ model is 0.814, which is significantly better than the baseline of 0.523, with an improvement of +0.291 (+55.6%); Under extreme conditions of 5 dB, the paper’ model still maintains 0.753, while the Baseline drops to 0.412, and the performance gap widens to +82.8%. This result shows that the paper’ model has significant robustness under noise interference and can effectively cope with non-ideal signal environments in actual scenarios.

Table 4 further analyzes the performance under load fluctuations (50%~120% of rated current) and temperature changes (-20℃~60℃): F1 remains at 0.918 at 80% load, and AUC only decreases by 0.012 under temperature fluctuations, indicating the robustness of the model to changes in working conditions. However, in an environment with humidity >85%, the signal-to-noise ratio of the acoustic emission signal decreases, causing F1 to drop to 0.843, requiring an environmental compensation strategy.

F. Ablation Experiment Analysis

In order to systematically evaluate the contribution of each component of the model to the performance, this paper designed multiple groups of ablation experiments. The results showed that after removing the CBAM feature fusion module based on the complete model, the F1 value dropped significantly from 0.923 to 0.871 (a decrease of 5.6%), indicating that the channel attention and spatial attention mechanisms can effectively suppress the redundant information between multi-band features and improve the weight distribution of key frequency bands (such as the frequency band in partial discharge). Secondly, after turning off the ThiNet pruning strategy, the model delay increased from 29.9ms to 46.7ms (exceeding the 50ms threshold required for deployment), but F1 only slightly dropped to 0.920 (a decrease of 0.3%), which shows that structured pruning does not sacrifice the core discrimination ability while compressing the model size, but improves the stability of some scenarios by reducing overfitting. In addition, after removing the DANN cross-domain migration training framework, the F1 value of supplier B dropped sharply from 0.861 to 0.729 (a decrease of 15.5%), and the inter-domain feature difference (MMD) rebounded from 16.2×10^{-3} to 48.7×10^{-3}, verifying the key role of adversarial training in aligning feature distribution in device heterogeneous scenarios.

Further experiments show that if the frequency band division of wavelet packet decomposition is removed (directly inputting the original signal), F1 drops to 0.854 (a decrease of 7.5%), proving the necessity of multi-band decoupling modeling for complex fault feature separation. After turning off the parameter sharing strategy of the ResNet-18 branch, the consistency of cross-modal features decreases, causing the F1 volatility of supplier C to increase from ± 0.034 to ± 0.089, indicating that shared weights enhance the model's adaptability to different device signal modes. The above results show that CBAM fusion and DANN migration are the core modules for improving discrimination performance, ThiNet pruning ensures deployment efficiency, and multi-band feature decoupling and parameter sharing strategies synergistically optimize the model's generalization ability and stability.

4. Conclusions

This paper proposes a GIS sub-health status assessment method based on a multi-frequency signal feature cascade network. It combines wavelet packet multi-scale decomposition, ResNet-18 multi-branch feature extraction, CBAM attention fusion, ThiNet pruning lightweight and DANN cross-domain migration mechanism to achieve efficient identification and robust early warning of the sub-health status of GIS equipment. Under multiple evaluation indicators, the model performs well in recognition accuracy, response timeliness and cross-domain generalization ability, and has the potential for engineering deployment. However, the current method still has certain performance fluctuations when facing signal noise interference under extreme working conditions, and the ability to express the state characteristics of some complex equipment is still limited. Future research can further introduce multimodal information fusion, time series modeling mechanism and active learning strategy to enhance the adaptability and decision-making intelligence of the model in a weakly supervised environment, and promote the evolution of GIS system health management from static evaluation to dynamic closed-loop optimization.

Consent to Publish

The manuscript has not been published before, and it is not being reviewed by any other journal. The authors have all approved the content of the paper.

Funding

This work was supported by Science and Technology Project of State Grid Shanxi Electric Power Company, Shuozhou Power Supply Company(5205F0240009).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, upon request.

Conflicts of Interest

The authors affirm that they do not have any financial conflicts of interest.

References

[1] Y.Q. Xing, Z.W. Wang, L. Liu, Y. Xu, Y. Yang, S. Liu, et al. Defects and failure types of solid insulation in gas-insulated switchgear: in situ study and case analysis. High Voltage, 2022, 7(1), 158-164. DOI: 10.1049/hve2.12127

[2] Y.X. Wang, J. Yan, Q.Z. Jing, Z.K. Qi, J.H. Wang, Y.S. Geng. A novel adversarial transfer learning in deep convolutional neural network for intelligent diagnosis of gas-insulated switchgear insulation defect: a DATCNN for GIS insulation defect diagnosis. IET Generation, Transmission & Distribution, 2021, 15(23), 3229-3241. DOI: 10.1049/gtd2.12255

[3] M. Borecki, K. Sobolewski. An elimination method for an emergency situation in Gas-insulated switchgear in power grids. IEEE Transactions on Power Delivery, 2021, 36(6), 3724-3732. DOI: 10.1109/TPWRD.2020.3048789

[4] B.M. Gao, Z.Y. Li, J.W. Gao, H. Liang, Z. Yan, Y.D. Hu. The simulation study on temperature field distribution of 220 kV gas insulated switchgear. Journal of Nanoelectronics and Optoelectronics, 2021, 16(5), 797-805. DOI: 10.1166/jno.2021.2998

[5] R. Hari, M. Mohana Rao. Characterization of nano-additive filled epoxy resin composites (ERC) for high voltage gas insulated switchgear (GIS) applications. International Journal of Emerging Electric Power Systems, 2022, 23(1), 47-57. DOI: 10.1515/ijeeps-2020-0257

[6] Y.X. Wang, J. Yan, Z. Yang, Q.Z. Jing, Z.K. Qi, J.H. Wang. A domain adaptive deep transfer learning method for gas-insulated switchgear partial discharge diagnosis. IEEE Transactions on Power Delivery, 2021, 37(4), 2514-2523. DOI: 10.1109/TPWRD.2021.3111862

[7] A. Subramaniam, A. Sahoo, S.S. Manohar, S.J. Raman, S.K. Panda. Switchgear condition assessment and lifecycle management: Standards, failure statistics, condition assessment, partial discharge analysis, maintenance approaches, and future trends. IEEE

Electrical Insulation Magazine, 2021, 37(3), 27-41. DOI: 10.1109/MEI.2021.9399911

[8] Y. Zhong, J. Hao, R.J. Liao, X.P. Wang, X.P. Jiang, F. Wang. Mechanical defect identification for gas-insulated switchgear equipment based on time-frequency vibration signal analysis. High Voltage, 2021, 6(3), 531-542. DOI: 10.1049/hve2.12056

[9] W. Koltunowicz, G. Behrmann, M. Boltze, A. Caprara, G. Coapes, F. Cook, et al. Requirements for ultra-high frequency partial discharge monitoring systems for gas-insulated switchgear. e & i Elektrotechnik und Informationstechnik, 2023, 140(1), 153-161. DOI: 10.1007/s00502-022-01110-4

[10] M. Seeger, F. Macedo, U. Riechert, M. Bujotzek, A. Hassanpoor, J. Hafner. Trends in High Voltage Switchgear Research and Technology. IEEJ Transactions on Electrical and Electronic Engineering, 2025, 20(3), 322-338. DOI: 10.1002/tee.24244

[11] V. Sreeram, S. Arunkumar, S.S. Reddy, T. Gurudev, Maroti. Comparative Analysis of Transients in AIS and GIS With Vacuum Interrupters. IEEE Transactions on Plasma Science, 2022, 50(9), 2681-2686. DOI: 10.1109/TPS.2022.3202339

[12] Y.R. Yadam, S. Ramanujam, K. Arunachalam. Study of polarization sensitivity of UHF sensor for partial discharge detection in gas insulated switchgear. IEEE Sensors Journal, 2022, 23(2), 1214-1223. DOI: 10.1109/JSEN.2022.3224475

[13] Q.Z. Jing, J. Yan, Y.X. Wang, X.Y. Ye, J.H. Wang, Y.S. Geng. A novel method for small and unbalanced sample pattern recognition of gas insulated switchgear partial discharge using an auxiliary classifier generative adversarial network. High Voltage, 2023, 8(2), 368-379. DOI: 10.1049/hve2.12274

[14] Y.X. Wang, J. Yan, Z. Yang, Z.F. Xu, Z.K. Qi, J.H. Wang. Simultaneous partial discharge diagnosis and localization in gas-insulated switchgear via a dual-task learning network. IEEE Transactions on Power Delivery, 2023, 38(6), 4358-4370. DOI: 10.1109/TPWRD.2023.3312704

[15] R. Yang, J.R. Li, T.Y. Zhu, W. Yang, D. Hu, E.B. Dong. An automatic robot for ultrasonic partial discharge detection of gas-insulated switchgear. Industrial Robot: the International Journal of Robotics Research and Application, 2024, 51(6), 908-921. DOI: 10.1108/IR-01-2024-0005

[16] X. Li, Y.M. Zhang, S.X. Chen, X. Zhang, Y. Tang. General multi-frequency small-signal model for resonant converters. IEEE Transactions on Power Electronics, 2021, 37(4), 3892-3912. DOI: 10.1109/TPEL.2021.3120617

[17] C.S. Liu, S.Y. Yu, T.C. Tian, Y. Su, S.X. Liu. Research on Frequency-Domain Electromagnetic Detection Signal Acquisition Technology Based on Multi-frequency LC Filtering. IEEJ Transactions on Electrical and Electronic Engineering, 2024, 19(2), 255-262. DOI: 10.1002/tee.23950

[18] M. Wolf, M. Rudolph, O. Kanoun. Amplitude and frequency estimator for aperiodic multi-frequency noisy vibration signals of a tram gearbox. Journal of Vibroengineering, 2021, 23(7), 1492-1507. DOI: 10.21595/jve.2021.21855

[19] W.D. Dang, D.M. Lv, L.G. Rui, Z. Liu, G.R. Chen, Z.K. Gao. Studying multi-frequency multilayer brain network via deep learning for EEG-based epilepsy detection. IEEE Sensors Journal, 2021, 21(24), 27651-27658. DOI: 10.1109/JSEN.2021.3119411

[20] D.M. Lv, W.D. Dang, X.L. Sun, Z.K. Gao. EEG-based multi-frequency multilayer network for exploring the brain state evolution underlying motor imagery. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2023, 13(3), 712-719. DOI: 10.1109/JETCAS.2023.3265928

[21] X.Y. Meng, Z.G. Liu, G.R. Li, X.J. Chen, S.Q. Wu, K.T. Hu. A multi-frequency input-admittance model of locomotive rectifier considering PWM sideband harmonic coupling in electrical railways. IEEE Transactions on Transportation Electrification, 2022, 8(3), 3848-3858. DOI: 10.1109/TTE.2022.3141445

[22] M. Anand, S.K. Goswami, D. Chatterjee. Multi-frequency control with fuzzy 2DOFPI in HVBTB converter of LF-HVAC system. International Journal of Emerging Electric Power Systems, 2023, 24(6), 717-728. DOI: 10.1515/ijeeps-2022-0133

[23] A.Y. Rodionov, L.G. Statsenko, D.A. Kuzin, M.M. Smirnova. Application of incoherent multi-frequency signals for information transmission in a nonstationary hydroacoustic environment. Acoustical Physics, 2023, 69(5), 738-747. DOI: 10.1134/S1063771023700653

[24] J.X. Dou, H.L. Yao, H. Li, Y.B. Cao, S.J. Liang. Vibration suppression of multi-frequency excitation using cam-roller nonlinear energy sink. Nonlinear Dynamics, 2023, 111(13), 11939-11964. DOI: 10.1007/s11071-023-08477-9

[25] R. Saha, L. Jiang, H. Salsabili, M. Faezipour, S. Ostadabbas, B. Larimer. Toward a smart sensing system to monitor small animal's physical State via Multi-frequency resonator array. IEEE Transactions on Biomedical Circuits and Systems, 2023, 17(3), 521-533. DOI: 10.1109/TBCAS.2023.3284823

[26] F.Q. Chen, R.D. Ying, J.W. Xue, F. Wei, P.L. Liu. A configurable and real-time multi-frequency 3D image signal processor for indirect time-of-flight sensors. IEEE Sensors Journal, 2022, 22(8), 7834-7845. DOI: 10.1109/JSEN.2022.3156085

[27] Y.Z. Jiang, J. Li, M.X. Yan, C. Tian, L. Pei, T.G. Ning. Channelized multi-frequency measurement system based on asymmetric double sideband detection. Applied Optics, 2024, 63(12), 3334-3342. DOI: 10.1364/AO.510858

[28] Y. Su, M.H. Liu, X.H. Kong, C.J. Guo, J. Zhu, X.M. Li, et al. Evaluation of breakdown voltage and water content in transformer oil using multi frequency ultrasonic and generalized regression neural network. Journal of Nanoelectronics and Optoelectronics, 2021, 16(3), 387-394. DOI: 10.1166/jno.2021.2971

[29] H. Athari, D. Raisz. Comparison of different control methods used in multi-frequency grid-forming converters. Journal of Electrical Engineering, 2025, 76(2), 190-199. DOI: 10.2478/jee-2025-0020

[30] Y.Y. Qiao, Y. Niu, Z. Han, S.W. Mao, R.S. He, N. Wang, et al. Joint optimization of resource allocation and user association in multi-frequency cellular networks assisted by RIS. IEEE Transactions on Vehicular Technology, 2023, 73(1), 826-842. DOI: 10.1109/TVT.2023.3306340

[31] X.L. Yang, M. Gao, L.B. Xie, M. Zhou. Multi-frequency based CSI compression for vehicle localization in intelligent transportation system. IEEE Transactions on Intelligent Transportation Systems, 2023, 25(3), 2719-2732. DOI: 10.1109/TVT.2023.3306340

[32] A. Chauhan, K.M. Singh. Single and Multi-frequency Parameter Estimation by α SWIFT-DPLL Through Optimized MOPSO Algorithm. Journal of Vibration Engineering & Technologies, 2024, 12(4), 5985-5998. DOI: 10.1007/s42417-023-01231-w

[33] S.Y. Sun, L. Xu, F. Ding. Parameter estimation methods of linear continuous-time time-delay systems from multi-frequency response data. Circuits, Systems, and Signal Processing, 2023, 42(6), 3360-3384. DOI: 10.1007/s00034-022-02285-z