EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

Safety risk point identification and localization in power equipment based on electromagnetic signals and a multiscale attention convolutional neural network

Siwu Yu[1]*, Fan Song[2], Xing Liu[2], Yang Mei[2], Jiangang Liu[1], Yuanyuan Zhao[1], Siqi Guo[1] and Yumin He[1]

[1]Electric Power Research Institute of Guizhou Power Grid Co., Ltd, Guiyang 550002, China
[2]Guizhou Power Grid Co., Ltd, Guiyang 550002, China

Abstract

INTRODUCTION: Safety risk points in power equipment often produce weak and non-stationary electromagnetic responses. These responses are difficult to identify when partial discharge, contact anomalies, insulation deterioration, and shielding defects occur under complex operating conditions.
OBJECTIVES: To improve risk recognition and spatial localization, a multichannel electromagnetic signal identification method is proposed.
METHODS: Sliding windows are used to construct signal samples. Outlier correction, normalization, and wavelet threshold denoising are then performed to suppress background interference while preserving transient pulse structures. Time-domain statistics, frequency-domain energy characteristics, and short-time Fourier transform (STFT) spectra are extracted to describe waveform fluctuation, band energy migration, and time-frequency coupling. A multiscale convolutional neural network (CNN) with an attention mechanism is designed to enhance key frequency bands and sensitive measurement channels. Spatial correlation weights are further introduced to estimate the location of risk points.
RESULTS: Experimental results show that the proposed method achieves a recognition accuracy of 95.28%, an F1-score of 94.58%, and an area under the receiver operating characteristic curve (AUC) of 0.981. The mean localization error is 8.46 cm.
CONCLUSION: Compared with conventional methods, the proposed method improves the discrimination of weak electromagnetic disturbances and provides a practical solution for safety risk identification in power equipment.

Keywords: power equipment; electromagnetic signals; risk point identification; multiscale convolution; attention fusion

Received on 03 August 2026, accepted on 08 September 2026, published on 22 September 2026

doi: 10.4108/ew.14304
*Corresponding author. Email: swyu2012@163.com

1. Introduction

Deep learning based on electromagnetic (EM) signal analysis provides an effective method for identifying safety risks in power equipment. During partial discharge (PD), insulation deterioration, contact anomalies, and shielding defects, electromagnetic responses usually present non-stationary variation, weak abrupt changes, and multi-frequency coupling. Traditional threshold-based discrimination and single-frequency analysis are therefore insufficient for joint risk classification and spatial localization. Hussain et al. reviewed PD diagnostic techniques for high-voltage equipment and showed that diagnosis increasingly depends on coordinated signal acquisition, feature extraction, and intelligent classification, while multichannel electromagnetic information remains insufficiently exploited [1]. Kumar et al. examined PD detection, feature extraction, artificial-intelligence classification, and optimization methods, indicating that handcrafted characteristics retain physical interpretability but remain sensitive to noise and feature-selection quality [2]. Rauscher et al. introduced deep learning and data augmentation for PD detection in electrical machines, improving diagnosis under limited samples, although the task remained centered on event detection and classification [3]. Carvalho et al. identified PD sources from signal-conditioning features, but temporal location, spectral migration, and multiscale disturbance structures were not jointly modeled [4]. Khodaveisi et al. applied machine-learning methods to PD localization in transformer tanks and demonstrated that measurement responses contain spatial information associated with discharge positions, but classification and localization cues were not represented jointly [5]. Almehdhar and Prochazka combined time-frequency analysis with deep learning for multiple PD source classification, improving pattern separability, while adaptive weighting of sensitive channels and frequency regions remained limited [6]. Li et al. strengthened time-frequency representation for high-voltage cable PD recognition, but the output remained classification-oriented [7]. Das et al. developed a deep-learning framework for identifying and locating single and multiple PD events, demonstrating the feasibility of joint diagnosis and localization [8]. Compared with these studies, the originality of the present method lies in three connected aspects. Parallel multiscale convolutions preserve transient pulses, mesoscale fluctuations, and longer-duration band migration in a common feature space rather than relying on a single receptive field. Attention weights are learned for sensitive measurement channels and discriminative frequency regions, so informative responses contribute adaptively to risk representation. More importantly, spatial localization is coupled with recognition rather than treated as independent post-processing: attention-derived channel responses, class probabilities, target-band electromagnetic intensity, and sensor coordinates are combined into spatial correlation weights for position estimation. Classification and localization losses are optimized jointly, allowing both tasks to constrain the same parameter-learning process. This coordinated treatment of multiscale features, channel-frequency attention, and classification-guided spatial correlation distinguishes the method from classification-oriented or standalone localization approaches.

2. Technical foundation for electromagnetic signal risk identification in power equipment

Abnormal operating states of power equipment alter the local electromagnetic field through changes in discharge activity, conductive contact, insulation properties, and shielding continuity. Partial discharge produces short-duration broadband pulses because of rapid charge movement in localized insulation defects, whereas unstable contacts can generate intermittent electromagnetic transients accompanied by repeated amplitude fluctuations. Progressive insulation deterioration tends to redistribute spectral energy over a longer duration, while shielding defects weaken electromagnetic confinement and increase broadband leakage responses. These mechanisms cause the externally measured electromagnetic signals to exhibit non-stationary variation, weak abrupt changes, and multi-frequency coupling, so risk identification depends on converting physical disturbances into measurable signal characteristics [9]. The observed response also varies among sensing channels. For the same risk source, propagation distance, structural obstruction, shielding condition, and sensor orientation affect signal attenuation and frequency response, producing differences in amplitude, energy distribution, and local time-frequency patterns between measurement points. Such inter-channel gradients contain spatial information and provide the physical basis for subsequent risk-point localization. Raw measurements, however, contain radio-frequency background interference, operating noise, electromagnetic crosstalk, and occasional sampling spikes. Preprocessing is therefore required to suppress non-risk disturbances without removing transient structures associated with actual abnormalities. A single feature domain is insufficient because different risks occupy different temporal and spectral scales [10]. Time-domain statistics characterize pulse intensity, waveform dispersion, and abrupt changes; frequency-domain energy features describe abnormal band concentration and energy migration; and short-time Fourier analysis preserves the correspondence between the occurrence time of a disturbance and its frequency composition. Their complementary representation is particularly important when partial discharge and contact anomalies share similar transient responses or when early insulation deterioration produces only weak spectral changes. In addition, the duration and spatial sensitivity of abnormal responses are not fixed. Short pulses require a relatively small receptive field, sustained fluctuations require broader contextual information, and different sensing channels may contribute unequally because of the spatial attenuation of electromagnetic propagation. These characteristics provide the technical basis for using multiscale convolution to capture disturbances with different durations, attention weighting to emphasize informative frequency regions and sensitive channels, and spatial correlation to transform inter-channel response differences into location information. Section III therefore constructs the identification and localization model according to this correspondence between electromagnetic disturbance mechanisms, multi-domain signal characteristics, and spatial measurement responses.

3. Design of a model for identifying safety risk points in power equipment based on electromagnetic signals

3.1. Design of the overall framework for safety risk point identification

According to the features of electromagnetic signals — nonstationary, weak, and multi-channel coupling — a model for identifying safety hazard points is designed in accordance with the calculation flow of "signal acquisition — sample construction — noise suppression — time-frequency characterization — feature fusion — hazard classification — spatial localization". [11] Figure 1 shows a system that uses multichannel EM sensor signals as input. On the left is the original EM sequence from various measuring points on the device. In the center section, the modules are connected in sequence, which are used to correct the anomaly sampling point, de-noising the wavelet threshold, the time domain statistics, the frequency domain energy characteristic, and the time frequency spectrum construction. On the left, there is a multiscale convolutional network, an attention fusion module, and a spatial correlation localization module. To avoid truncating partial-discharge pulses due to excessively short signal segments or diluting risk disturbances with background operational signals due to excessively long segments, the sample window length is determined as

$$L = \lceil \alpha f_s \tau_{\min} \rceil \quad (1)$$

where $L \in N^{+}$ is the number of sampling points in one sample window, f_s is the electromagnetic-signal sampling frequency in Hz, $\tau_{\min}$ is the minimum observable duration of an abnormal pulse in seconds, α is the dimensionless coverage factor and is set to $\alpha \geq 5$, and $\lceil\ \rceil$ denotes the ceiling operator. In the whole frame, the preprocessor is used to reduce the RF background noise and EMI, and the time frequency feature extraction module is used to keep short duration pulses, high frequency energy migration, and local spectrum peak variations. The multi-scale convolution module is used to capture risk patterns across different time scales, and the attention module is used to highlight key frequency bands and sensitive measurement points, and the spatial correlation location module estimates the distribution region of the risk points according to the difference of the characteristic response between the channels, so as to form a complete calculation process from the signal characteristic to the identification of the safety hazard point [12].

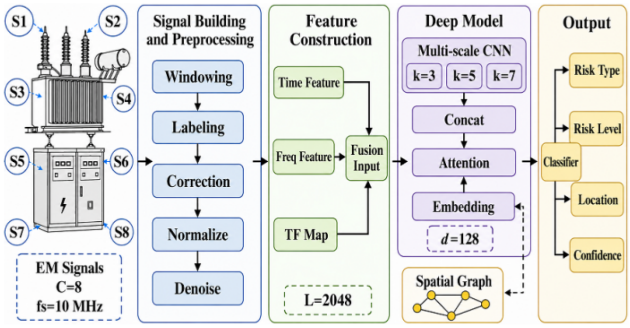

Figure 1. Overall framework for safety risk point identification

3.2. Electromagnetic signal acquisition and sample construction

3.2.1. Multi-channel electromagnetic signal acquisition structure

The multichannel electromagnetic acquisition scheme is designed to deal with the problem of uneven space distribution of danger points, fast decay of abnormal electromagnetic interference, and the sensitivity of single point signal to local obstruction and background disturbance. The collection system deploys electromagnetic sensing channels around the transformer sleeve, the switchgear bus connection area, the cable terminal, the insulation support, and the grounding shield, so as to simultaneously capture the high frequency pulses and broadband interference caused by the partial discharge, the loose contacts, the insulation degradation, and the shielding defects in the multiple measuring points [13]. Each channel uses a uniform time stamp and a common source trigger mechanism to ensure that the difference in amplitude, time of arrival, and frequency band response of the same anomaly at different measuring points can be compared, thus providing a basis for later spatial correlation and localization. In engineering implementations, some degree of redundancy is usually incorporated to avoid aliasing of high frequency risk components in the digitization process [14]. The acquired multichannel signal can be expressed as:

$$X = [x_1(t), x_2(t), \ldots, x_C(t)]^T, \quad t = 1, 2, \ldots, T \quad (2)$$

where $X \in R^{C \times T}$ is the multichannel electromagnetic-signal matrix, C is the number of sensing channels, T is the number of temporal sampling points in a continuous acquisition sequence, $x_c(t)$ is the electromagnetic amplitude measured by channel c at sampling instant t, $c = 1, \ldots, C$, and the superscript T denotes matrix transpose. This configuration can not only support the construction of the subsequent sliding window but also provide the basis for noise reduction, time-frequency generation, multiscale feature fusion and confidence estimation.

3.2.2. Signal segmentation and sample labeling

It is necessary to transform the sequential sequence into a window which is suitable for the input of the model, so that the short term disturbances, such as partial discharge, contact anomaly, and insulation degradation, can be characterized in a fixed scale. Sample segmentation uses a sliding window method, in which the length of the window is limited by the principle of abnormal pulse coverage, and the step size is set on the basis of the overlapping information of neighboring samples. This approach preserves the continuous evolution of the risk interference and avoids the training bias due to the large number of redundant samples [15]. For the multi-channel signal matrix X , the i -th sample can be expressed as:

$$D_i = \left(S_i, y_i, p_i\right), \quad S_i = X_{[:,iH+1:iH+L]} \tag{3}$$

where D_i is the i-th labeled sample, $S_i \in R^{C\times L}$ is its multichannel signal segment, H is the sliding step measured in sampling points, i is the window index, $y_i \in \{0,1,\ldots,K-1\}$ is the risk-class label, K is the number of risk classes, and $p_i \in R^{d_p}$ is the reference coordinate of the corresponding risk point, with d_p =2or 3 according to the adopted equipment coordinate system. Class labels are determined from equipment operating records, partial-discharge calibration tests, contact-anomaly simulations, calibrated insulation-defect samples, shielding-defect configurations, and manual verification, and are divided into normal operation, partial discharge, contact anomaly, insulation deterioration, and shielding defect. For abnormal samples C1–C4, the ground-truth localization target is obtained from the physically configured abnormal source or defect position determined before signal acquisition. The geometric center of each abnormal source is recorded in the local Cartesian coordinate system of the equipment as $p_n^{gt} = [x_n^{gt}, y_n^{gt}, z_n^{gt}]^T$. The same structural datum and axis definitions are used for sensor and defect coordinates in all acquisition batches, maintaining a one-to-one correspondence among the class label, channel response, and physical risk position. Normal samples C0 contain no physical risk point and therefore contribute only to the classification loss rather than the localization loss. When multiple abnormal responses occur in one window, the primary class is determined from the temporal overlap ratio and risk priority, and the coordinate of the corresponding primary abnormal source is used as the localization target.

3.3. Electromagnetic signal preprocessing and noise suppression

The acquired EM signals contain several non-risk-related components acquired by multi-channel acquisition and sliding window segmentation, for example, the random spike, the transient drift of the sensor, the RF background noise, the electromagnetic cross-talk, and the operating noise of the device. High frequency energy transitions resulting from PD pulses, contact anomalies, and insulation degradation are easily masked by noise if they are directly inputted into the time-frequency feature extraction module. Its function is to convert the EM signals from the various measuring points, the sampling lot, and the working load into the model input with the same amplitude scale, the relatively stable waveform structure, and the key interference is completely preserved [15]. This phase consists of three steps: correcting the anomalous data point, normalizing the amplitude, and wavelet-based noise suppression. In this paper, a new method is proposed, which is to correct the occasional spike in the time domain, and to eliminate the sensitivity difference of the channel. Furthermore, the method of de-noising is adopted to attenuate the background noise in the multi-scale decomposition space, and to preserve the short term pulse structure, thus providing a stable data base for constructing time domain, frequency domain, and time frequency spectrum characteristics.

3.3.1. Outlier correction and normalization

Outlier correction deals with isolated spikes due to unstable sensor contact, transient electromagnetic crosstalk, and sampling circuit jitter, with emphasis on controlling the effect of non-risk-related fluctuations on the subsequent computation of kurtosis, peak and band energy. For the signal $S_c(t)$ in the window of the th channel (c), a local median constraint is applied for discrimination. When a sample deviates from the neighborhood median by more than a set threshold, the original value is replaced with the neighborhood median, thereby preserving the true continuous pulse structure while attenuating single-point noise spikes:

$$\tilde{S}_c(t) = \begin{cases} med\left(S_c(t-r:t+r)\right), & \left|S_c(t) - med(S_c(t-r:t+r))\right| > \lambda\sigma_c \\ S_c(t) & \left|S_c(t) - med(S_c(t-r:t+r))\right| \le \lambda\sigma_c \end{cases} \tag{4}$$

where $\tilde{S}_c(t)$ denotes the corrected sample value, med denotes the local median, r denotes the neighborhood radius, λ denotes the anomaly detection coefficient, and σ_c denotes the standard deviation of the signal in the channel window. The signal with corrected outliers is subjected to Z-score normalization to unify the amplitude scales of different sensor channels into the same distribution space:

$$\hat{S}_c(t) = \frac{\tilde{S}_c(t) - \mu_c}{\sigma_c + \varepsilon} \tag{5}$$

where $\hat{S}_c(t)$ represents the normalized signal value, μ_c represents the mean of the signal in the c th channel window, and ε is a stability term to prevent the denominator from becoming zero. This process ensures that the multichannel electromagnetic signals are kept at a similar scale prior to wavelet de-noising and time-frequency feature extraction, so as to avoid the dominance of a single high amplitude channel.

3.3.2. Wavelet threshold-based electromagnetic noise suppression

The risk disturbance of electromagnetic signals is usually presented as a short duration, a local energy spike, or a band shift, while the background noise usually exhibits a random or wide-band weak disturbance. Therefore, the de-noising method based on wavelet threshold can be used to preserve the local nonstationary structure [16]. The normalized window signal $\hat{S}_c(t)$ is subjected to discrete wavelet decomposition, which decomposes the signal into low-frequency approximation coefficients and multiple layers of high-frequency detail coefficients; the low-frequency component preserves changes in the device's operational state, while the high-frequency component attenuates random noise based on a threshold function. The threshold value is determined jointly by the noise standard deviation and the window length:

$$\theta_j = \sigma_j \sqrt{2 \ln L} \quad (6)$$

In the equation, θ_j represents the threshold for the detail coefficients of the j th layer, σ_j represents the estimated standard deviation of the noise in that layer, and L represents the window length. The threshold function uses a soft threshold to compress the detail coefficients $d_{j,k}$:

$$d'_{j,k} = sign(d_{j,k}) \max(|d_{j,k}| - \theta_j, 0) \quad (7)$$

where $d'_{j,k}$ denotes the k th detail coefficient of the j th layer after thresholding. Soft thresholding reduces waveform oscillations caused by hard clipping, enabling the denoised electromagnetic signal to suppress random interference while preserving the spike edges generated by partial discharges and contact anomalies [17]. Therefore, the signal obtained by wavelet reconstruction can be used as a common input for time domain statistics, frequency domain energy characteristics, and time-frequency characteristics.

3.4. Extraction of time-frequency features from electromagnetic signals

After correcting the abnormal sampling point, normalization, and the de-noising of the wavelet, the EM signal satisfies the requirement of stability. However, the security risks of the facility are not limited to the increase in amplitude; their electromagnetic responses frequently include short duration pulses, band energy shifts, local spectral peaks, and cross-channel responses. [18] To make sure that the subsequent multi-scale feature fusion model receives inputs with physical meaning and computational specificity, this section constructs features for the windowed signal at three levels: time-domain, frequency-domain, and time-frequency. Time domain characteristics are used to describe the impact strength and waveform discontinuity of electromagnetic disturbance, so that they can be used to represent the transient anomaly, such as partial discharge, loose contact, etc.; the frequency domain characteristic is used to describe the aggregation and migration of the abnormal energy in different frequency bands, reflecting the state variations such as insulation degradation and shielding defects [19]; the time-frequency spectrum maintains the coupling relation between the temporal location and the frequency component, which allows the formation of a learning local structure in a two-dimensional feature space--the frequency, frequency band division, and spectrum resolution of all characteristic types correspond to the above-mentioned sampling frequency, abnormal pulse duration, and channel synchronization requirements, so that the characteristic scale is not disconnected from the capture structure. Then, the multi-source feature will be used as input to the multiscale convolution and attention fusion module in section 3.5.

3.4.1. Time-domain statistical feature extraction

The time domain statistics are used to describe the amplitude distribution, waveform fluctuation, and transient impact severity of electromagnetic signals in the sampling window, and they are used as the first level of feature input to the hazard point identification model to acquire basic interference information [20]. The preprocessed multi-channel windowed signals $\hat{S}_c(t)$ have eliminated some isolated spikes and amplitude-scale discrepancies; however, risk states such as partial discharges, loose contacts, and insulation degradation still manifest in the time series as pulse amplification, increased fluctuations, peak concentration, or sudden energy changes.To preserve these local responses associated with equipment abnormal states, this section extracts statistical measures such as mean, variance, root mean square, kurtosis, skewness, peak factor, and impulse factor from each channel window to form a time-domain feature vector:

$$f_c^{time} = [\mu_c, \sigma_c^2, x_{rms,c}, K_c, Sk_c, C_{f,c}, I_{f,c}] \quad (8)$$

where f_c^{time} denotes the time-domain statistical feature vector for the th channel (c), μ_c denotes the window mean, σ_c^2 denotes the variance, $x_{rms,c}$ denotes the root mean square (RMS), K_c denotes the kurtosis, Sk_c denotes the skewness, $C_{f,c}$ denotes the peak factor, and $I_{f,c}$ denotes the impulse factor. Kurtosis and peak factors are more sensitive to short-duration, high-amplitude pulses, and can be used to capture transient surges due to partial discharge and contact anomalies; variance and skewness describe discrete changes in signals around the steady state operating level, and are suitable for characterizing sustained waveform shifts due to insulation degradation or shielding defects [21]. Time domain characteristics of each channel are concatenated in the order of measuring points to form a multi-channel time domain characteristic matrix, which allows the simultaneous acquisition of amplitude statistics and spatial measuring points of difference in frequency domain and multiscale fusion models.

3.4.2. Extraction of frequency-domain energy features

Energy characteristics in the frequency domain are used to describe the energy distribution of electromagnetic signals in a wide range of frequencies, and to overcome the limitations of time domain statistics in terms of spectral migration, local spectral peaks, and high frequency anomalies [22]. For the preprocessed window signals of the first c channels $\hat{S}_c(t)$, the Fast Fourier Transform (FFT) is used to obtain the spectral amplitude sequence. According to the features of abnormal electromagnetic interference, the spectrum is classified as low frequency operation background, middle frequency structure fluctuation, and high frequency pulse disturbance band. Therefore, the frequency domain analysis focuses on the extraction of the band energy, the energy ratio, the spectral centroid, and the spectral entropy. The energy of the b th frequency band can be expressed as:

$$E_{c,b} = \sum_{f \in B_b} \left| F_c(f) \right|^2 \tag{9}$$

where $E_{c,b}$ represents the energy of the c th channel within the b th frequency band, B_b denotes the range of the b th frequency band, and $F_c(f)$ denotes the complex spectral amplitude of the windowed signal at frequency f after a fast Fourier transform. To mitigate the impact of absolute amplitude differences across channels on frequency-domain analysis, the band energy is further converted into a normalized energy ratio:

$$\rho_{c,b} = \frac{E_{c,b}}{\sum_{b=1}^{B} E_{c,b} + \varepsilon} \tag{10}$$

where $\rho_{c,b}$ represents the energy proportion of the th frequency band (b), B denotes the number of frequency bands, and ε is a numerical stability term. The frequency domain characteristic vector, which is made up of the energy ratio of each frequency band and the spectral entropy, is used to reflect the degree of clustering and the dispersion of risk disturbance in spectrum space. Combined with time domain statistics, it forms a multi-source feature input, which provides the base of frequency domain for the construction of time frequency spectrum and attention based fusion.

3.4.3. Construction of time-frequency spectrum features

It is possible to identify and capture by the model of short-duration high-frequency pulses, local spectral peak enhancement, and energy migration processes caused by the abnormal state of partial discharge, loose contact, and insulation degradation in a two-dimensional characteristic space [23] .Temporal and frequency domain metrics extracted in sections 3.4.1 and 3.4.2, respectively, are capable of describing amplitude fluctuations and variation in band energy, but they are difficult to maintain specific temporal locations in which anomalous components occur. Therefore, this section employs the short-time Fourier transform (STFT) to construct the time-frequency energy matrix of the windowed signal. For the preprocessed signal $\hat{S}_c(t)$ of the c th channel, its time-frequency representation can be written as:

$$T_c(\tau, f) = \left| \sum_{t=1}^{L} \hat{S}_c(t) w(t-\tau) e^{-j2\pi ft} \right|^2 \tag{11}$$

where $T_c(\tau, f)$ represents the time-frequency energy of the c th channel at time τ and frequency f , $w(t-\tau)$ denotes the sliding analysis window function, L denotes the window length, and j denotes the imaginary unit. The length of the analysis window is determined by the duration and frequency resolution of the abnormal pulse. An excessively short window reduces the ability to recognize low frequency energy, whereas an excessively long window will smooth out short-duration discharge pulses; thus, a local window function that matches the sample window is used to keep the time and frequency resolution balanced [24]. To reduce the impact of amplitude differences between channels on the spectrum intensity, the time-frequency energy matrix is normalized as follows:

$$\overline{T}_c(\tau, f) = \frac{T_c(\tau, f) - \min(T_c)}{\max(T_c) - \min(T_c) + \varepsilon} \tag{12}$$

where $\overline{T}_c(\tau, f)$ represents the normalized time-frequency spectrum value, and ε is a numerical stability term. Then, a multi-channel time-frequency feature tensor is obtained by stacking the normalized spectra of every channel in the order of measurement, which enables the multiscale convolution module to extract the risk disturbance patterns from the local structures. Combined with time domain statistics and frequency domain energy characteristics, these patterns can be used to identify the hazard and determine the space position.

3.5. Design of the multiscale feature fusion recognition model

The time domain statistic characteristic, frequency domain energy characteristic and multi-channel time frequency spectra jointly describe the electromagnetic response (EMR) data. However, because these characteristics are different in time scale, frequency response, and spatial measuring point sensitivity, a unified fusion model is needed to extract and discriminate between risk patterns. Fig. 2 shows a computational workflow that is centered on "Feature Input — Multiscale Convolution-Attention Fusion — Feature Linking — Risk Classification." In this process, the time-frequency spectrum is used as a two-dimensional convolutional input to capture PD pulses and HF peaks, while the time domain and the frequency domain are used as auxiliary statistical vectors to supplement the information on amplitude peaks, energy ratios, and spectral distributions [25]. The multiscale convolution branch extracts short term fluctuations, mesoscale oscillations, and risk evolution characteristics over a longer period of time. The attention fusion module assigns weights according to channel response strength, frequency band contribution, and local spectrum activation level, so as to reduce the interference of background operation signals on classification boundaries. Each convolutional kernel is set according to the duration of abnormal pulses, window length, and time frequency spectrum resolution. The fusion architecture combines the statistical, spectral, and topological characteristics of electromagnetic signals into a unified risk representation space, which can be used to classify risk points and locate them in space.

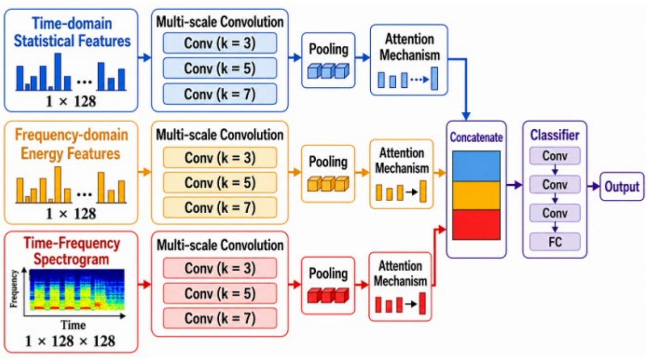

Figure 2. Structural diagram of the multiscale electromagnetic feature fusion recognition model

Multiscale convolutional feature extraction module

The multiscale convolution feature extraction module is used to extract the risk disturbance patterns from the multi-channel time-frequency spectrum, and to represent the short duration high frequency pulses produced by the partial discharge, the local energy fluctuation due to the contact anomaly, and the frequency band shift corresponding to the degradation of the insulation in the uniform characteristic space. The input features are normalized time-frequency spectrum tensors, which are associated with the channel, time slice, and frequency bin. The small-size kernel focuses on the extraction of short term peaks and local spectral peaks; the medium-scale kernel captures the energy diffusion in successive time slices; and the larger kernel describes the change in frequency band structure within longer windows. The feature extraction process for the r th-scale convolution branch can be expressed as:

$$Z_r = \phi\left(W_r * \overline{T} + b_r\right) \quad (13)$$

where Z_r denotes the feature map output by the r th convolution branch, $\overline{T}$ denotes the multi-channel time-frequency spectrum input tensor, W_r denotes the convolution kernel parameters for the r th scale, b_r denotes the bias term, $*$ denotes the convolution operation, and ϕ denotes the nonlinear activation function. Based on the duration of anomalous pulses, spectral resolution, and sampling window length, the convolution kernel scale is determined to ensure that the local interference is not smoothed by too large a receptive field, and the migration relation is not lost because of too small a receptive field. Then, a multi-scale feature tensor is generated from each scale, which provides the attention fusion module with the local impact, frequency band clustering, and the difference in response between measurement points.

3.5.1. Attention feature fusion module

The attention feature fusion module is applied to the deep features produced by the multiscale convolution branches to adaptively emphasize informative electromagnetic responses. Its basic channel-recalibration procedure follows the squeeze-and-excitation principle of SE-Net. Specifically, global average pooling corresponds to the squeeze operation and compresses each multiscale feature channel into a global response descriptor. The subsequent two-layer nonlinear mapping performs excitation to estimate the relative importance of different feature channels. The generated attention weights are then used to recalibrate the concatenated multiscale features:

$$\alpha = softmax\left(W_2 \delta\left(W_1 GAP\left(Z\right) + b_1\right) + b_2\right) \quad (14)$$

where α denotes the attention weight vector, Z denotes the multi-scale convolutional feature tensor, GAP denotes global average pooling, W_1 and W_2 denote weight mapping matrices, b_1 and b_2 denote bias terms, and δ denotes the nonlinear activation function. The fused features are obtained by re-scaling the multi-scale features using the attention weights:

$$F_{att} = \alpha \odot Z \quad (15)$$

where F_{att} denotes the attention-enhanced fused feature, and $\odot$ denotes the channel-wise, weighted operation. In this module, the weights are generated from the feature response strength, not the fixed ratio, which reduces the disturbance caused by the background noise and the low contribution measurement points. At the same time, it improves the representation of the high frequency spectrum peak from the partial discharge, the impact structure from the contact anomaly, and the frequency band migration from the insulation degradation, so as to provide a more discriminating risk feature for the risk point classification output module.

3.5.2. Risk point classification output module

The Risk Point Classification Output Module maps the Attention Enhanced Fused Features to Power Equipment Safety Risk Classes and Risk Levels, and provides Category Constraints for Spatial Correlation Localization. Since the fused feature F_{att} already incorporates time-domain impulse, frequency-domain energy migration, multi-channel time-frequency spectral responses, and key measurement point weight information, the classification layer compresses the high-dimensional features via fully connected mapping and uses the Softmax function to generate probability distributions for each risk category:

$$\hat{y}_k = \frac{\exp\left(w_k^T F_{att} + b_k\right)}{\sum_{q=1}^{K} \exp\left(w_q^T F_{att} + b_q\right)} \quad (16)$$

where $\hat{y}_k$ represents the predicted probability that a sample belongs to risk class k , K denotes the number of risk classes, and w_k and b_k represent the classification weights and bias terms corresponding to class k , respectively. The hazard categories can be defined on the basis of normal operation, partial discharge, contact abnormal, insulation deterioration, and shielding defects, and the hazard level is coded together on the basis of the predicted probability, the severity of the anomaly category, and the intensity of the electromagnetic response. During the model training phase, cross-entropy loss is used to constrain the classification boundary:

$$L_{cls} = -\sum_{k=1}^{K} y_k \log(\hat{y}_k) \quad (17)$$

where y_k denotes the one-hot encoded label of the true class, and L_{cls} represents the classification loss. This output module converts multiscale fused features into interpretable risk types and reliability results, allowing for the identification of transient hazards (e.g. partial discharge), structural hazards (e.g. contact anomalies) and progressive hazards (e.g. isolation degradation) in a uniform classification area.

3.6. Risk point localization method based on spatial correlation

Risk Point Classification Output Module provides risk types and confidence levels related to EMF. Nevertheless, the safe operation and maintenance of the electrical equipment also requires the identification of possible locations in the installation's structure in which such anomalies are likely to occur. Thus, the location module integrates the multiple measurement point response difference, the spatial sensor layout, and the hazard category prior knowledge into its computation. Figure 3 shows the process involving the device measuring point, the channel specific response characteristics, the spatial proximity relation, the risk weight calculation, and the risk zone output. Different measuring points on the outside of the device are used as spatial nodes, each of which includes attention fused features and the electromagnetic response strength of the respective channel. The relative weights of the nodes are determined according to the physical distance between the measuring points, the orientation of the installation and the similarity of the signal. For the c-th measurement point, the normalized spatial correlation weight is defined as

$$\omega_c = \frac{\exp[\lambda a_c + (1-\lambda)\hat{y}_{k^*} r_c]}{\sum_{m=1}^{C} \exp[\lambda a_m + (1-\lambda)\hat{y}_{k^*} r_m]} \quad (18)$$

where $\omega_c \in (0,1)$ denotes the normalized contribution of sensing channel cc to risk-point localization and satisfies $\sum_{c=1}^{C} \omega_c = 1$; a_c is the attention-derived response score of channel cc; $r_c \in [0,1]$ is the normalized electromagnetic-response intensity of that channel within the target frequency band or time-frequency region; $k^* = \arg\max_k \hat{y}_k$ denotes the predicted risk class; and $\hat{y}_{k^*}$ is its classification confidence. The coefficient $\lambda \in [0,1]$ controls the relative contribution of learned attention response and measured electromagnetic intensity. It is fixed at $\lambda = 0.5$ for all experiments, assigning equal contributions to the two normalized information sources. This setting avoids a priori preference for either the learned feature response or the physical signal intensity when no additional evidence supports asymmetric weighting, while preserving the joint use of feature-level and signal-level spatial information. The estimated risk-point coordinate is subsequently obtained as

$$\hat{p} = \sum_{c=1}^{C} \omega_c p_c \quad (19)$$

where $p_c \in R^{d_p}$ is the known coordinate of sensing channel c , and $\hat{p} \in R^{d_p}$ is the estimated risk-point coordinate.

The spatial location of the risk point is estimated by weighting the coordinates of each measurement point:

$$\hat{\tilde{p}}_n = \sum_{c=1}^{C} \beta_{n,c} \tilde{p}_c \quad (20)$$

where $\hat{\tilde{p}}_n$ denotes the normalized predicted position of the nn-th abnormal sample, $\tilde{p}_c$ is the normalized spatial coordinate of the cc-th electromagnetic measurement point, and $\beta_{n,c}$ is the spatial correlation weight obtained from Eq. (17), satisfying $\sum_{c=1}^{C} \beta_{n,c} = 1$ after normalization. Sensor coordinates and ground-truth risk coordinates are transformed using the same local equipment coordinate system and identical spatial bounds. This treatment prevents differences in the physical ranges of individual coordinate axes from dominating the localization objective and ensures that spatial weighting and ground-truth supervision are represented on the same scale.

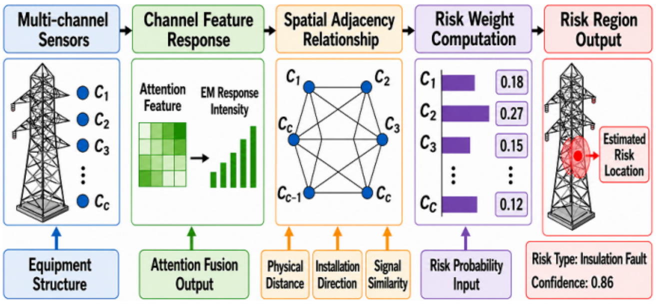

Figure 3. Schematic diagram of multi-channel electromagnetic signal spatial correlation localization

3.7. Model training process and parameter settings

Model training jointly constrains risk-category discrimination and abnormal-source localization. For abnormal samples C1–C4 with valid position labels, the ground-truth coordinate p_n^{gt} is obtained from the physically calibrated source or deliberately configured defect position recorded before signal acquisition. Normal samples C0 have no physical risk-point coordinate and participate only in classification training. Because the physical dimensions and sensor spacing may differ among coordinate directions, both sensor coordinates and ground-truth risk coordinates are transformed by Min-Max normalization using fixed boundaries of the monitored equipment region. For coordinate dimension $d \in \{x, y, z\}$, the normalization is defined as

$$\tilde{p}_{n,d}^{gt} = \frac{p_{n,d}^{gt} - p_d^{\min}}{p_d^{\max} - p_d^{\min}}, \quad \tilde{p}_{c,d} = \frac{p_{c,d} - p_d^{\min}}{p_d^{\max} - p_d^{\min}} \qquad (21)$$

where $p_d^{\min}$ and $p_d^{\max}$ denote the fixed physical boundaries of the monitored region along coordinate dimension dd. These bounds are determined from equipment geometry and sensor coverage and remain unchanged for the training, validation, and test sets. After obtaining the normalized predicted position $\hat{\tilde{p}}_n$ from Eq. (18), the localization loss is defined as the mean squared Euclidean distance between the predicted and ground-truth coordinates:

$$L_{loc} = \frac{1}{\sum_{n=1}^{B} m_n} \sum_{n=1}^{B} m_n \left\| \hat{\tilde{p}}_n - \tilde{p}_n^{gt} \right\|_2^2 = \frac{1}{\sum_{n=1}^{B} m_n} \sum_{n=1}^{B} m_n \sum_{d \in \{x,y,z\}} \left(\hat{\tilde{p}}_{n,d} - \tilde{p}_{n,d}^{gt} \right)^2 \qquad (22)$$

where B denotes the batch size and m_n is a localization-validity mask. Specifically, $m_n = 1$ for abnormal samples C1–C4 and $m_n = 0$ for normal samples C0. The mask prevents samples without physical risk locations from producing invalid localization gradients. During experimental evaluation, the normalized prediction is transformed back into the physical equipment coordinate system as

$$\hat{p}_{n,d} = \hat{\tilde{p}}_{n,d} \left(p_d^{\max} - p_d^{\min} \right) + p_d^{\min} \quad (23)$$

and the Euclidean distance between the predicted and ground-truth positions is reported in centimeters. The joint optimization objective is therefore expressed as

$$L = L_{cls} + \eta_{loc} L_{loc} + \rho \|\Theta\|_2^2 \quad (24)$$

where L_{cls} is the cross-entropy loss for the five risk states, $\eta = 0.4$ is the localization-loss weight, $\lambda = 1 \times 10^{-4}$ is the regularization coefficient, and Θ denotes the set of trainable model parameters. Coordinate normalization prevents the physical range of an individual axis from dominating localization optimization, whereas inverse transformation is applied only for physical position output and centimeter-level localization-error evaluation.

The key parameters of the training phase are given in Table 1, in which window length, sliding step size, convolution kernel size, and batch size meet the requirements of abnormal pulse coverage, sample continuity, multi-scale disturbance capture, and gradient stability, etc. In training, when the validation set loss is not reduced for a number of successive epochs, an early stopping mechanism is activated; the learning rate is decreased in response to the variation of the validation loss, ensuring that the model maintains a stable optimization under restricted electromagnetic sample conditions. The above training process ensures that risk classification, key channel response, and space position estimation are constrained in the same parameter update process, which provides a uniform implementation basis for further experiments, including model performance comparison, risk type analysis, localization error validation, and ablation research.

Table 1. Model training parameter settings

Parameter Name	Parameter Setting	Basis for Setting
Sample Window Length L	2048 points	Covers short-duration electromagnetic pulses while preserving local frequency band

		structure
Step size H	512 points	Maintains continuity between adjacent windows, reducing the probability of missing high-risk segments
Multi-scale convolution kernels	3×3, 5×5, 7×7	Corresponding to short-term spikes, mesoscale fluctuations, and frequency-band migration features
Batch size	32	Balances gradient stability and memory usage
Initial learning rate	0.001	Ensures convergence speed and parameter update stability in the early stages of training
Optimizer	Adam	Suitable for adaptive gradient updates under non-stationary signal characteristics
Number of training iterations	100	Ensures sufficient update of multi-scale convolutional and attention weights
Dropout Rate	0.3	Reduces the risk of overfitting in the feature fusion layer
Positioning loss weight η	0.4	Maintain classification as the primary objective while optimizing position estimation in a coordinated manner
Regularization coefficient λ	1×10^{-4}	Constraint on model parameter complexity

4. Experimental design and results analysis

4.1. Experimental platform and dataset construction

The experimental platform is constructed for electromagnetic response acquisition around typical safety-risk regions of power equipment, including the switchgear bus connection, cable terminal, insulation support, and shielding connection areas. The acquisition end employs an 8-channel electromagnetic sensor array and a synchronous module with a sampling frequency of 2 MHz. Model training and inference are performed on a workstation equipped with an Intel Core i7-12700 processor, 32 GB RAM, and an NVIDIA GeForce RTX 3060 GPU with 12 GB memory. The model is implemented in PyTorch, and all computational-efficiency measurements are obtained on the same hardware configuration with a batch size of 1 during inference. The eight channels are arranged to provide paired spatial coverage of the principal structural regions, so that each potential risk area can be observed from more than one measurement position. This configuration reduces the dependence on a single channel when electromagnetic propagation is affected by structural obstruction, shielding, or local interference, while preserving sufficient inter-channel response differences for spatial correlation localization. The choice of eight channels is further examined through channel-number experiments in Section 4.6 rather than being treated as a fixed empirical setting. All channels share the same trigger clock, allowing response intensity and arrival characteristics at different measurement points to remain temporally comparable. The continuous signals are segmented using a window length of 2048 points and a step size of 512 points. To prevent information leakage between adjacent windows, the data are divided by acquisition batch into training, validation, and test sets at a ratio of 7:1:2. Table 2 summarizes the sample composition, label sources, and principal electromagnetic characteristics of each risk state.

Table 2. Composition of the electromagnetic signal dataset

Risk State	Label ID	Number of Samples	Number of Acquisition Channels	Label Source	Primary Electromagnetic Characteristics
Normal Operation	C0	1200	8	Steady-State Operation Records	Stable amplitude, concentrated energy in the frequency band
Partial Discharge	C1	1150	8	Partial Discharge Calibration Test	Short-duration high-frequency pulse enhancement
Contact Anomaly	C2	1080	8	Contact resistance anomaly simulation	Local surges and periodic fluctuations
Insulation Deterioration	C3	1020	8	Insulation Defect Sample Calibration	Band-specific energy migration
Shielding Defects	C4	980	8	Shielding structure defect configuration	Broadband disturbance enhancement
Total	—	5430	8	Multi-source Annotation Fusion	Covering Classification and Localization Tasks

4.2. Performance comparison of different recognition models

In order to evaluate the classification performance under both conventional and recent recognition frameworks, SVM, Random Forest, 1D-CNN, LSTM, CNN-LSTM, and Transformer are retained as baseline methods, while Improved CBAM-ResNet, TFMT, and STFT-CNN, representing recent models from 2023 to 2025, are additionally included. Since the published methods were originally evaluated using different signal sources and datasets, their reported metrics are not directly compared. Instead, the three recent architectures are reproduced under the same electromagnetic dataset, 7:1:2 data partition, preprocessing procedure, and evaluation criteria used in this study. The test results are presented in Table 3. Conventional 1D-CNN and LSTM achieve accuracies of 91.37% and 90.68%, respectively, whereas Transformer reaches 93.16%. Recent feature-enhanced models further improve recognition: Improved CBAM-ResNet, TFMT, and STFT-CNN achieve accuracies of 93.48%, 94.16%, and 94.57%, respectively. Their results confirm the effectiveness of attention enhancement and time-frequency representation for weak electromagnetic disturbances. The proposed model reaches an accuracy of 95.28%, an F1-score of 94.58%, and an AUC of 0.981. Compared with TFMT and STFT-CNN, the improvement is not derived solely from deeper feature extraction; multiscale convolution preserves disturbances with different durations, while attention fusion reallocates contributions among sensitive measurement channels and discriminative frequency regions. The resulting representation therefore provides more stable discrimination when weak transient pulses, band migration, and multichannel response differences coexist.

Table 3. Comparison of performance among different recognition models

Recognition Model	Accuracy /%	Precision /%	Recall/ %	F1-score/ %	AUC
SVM	86.42	85.91	84.76	85.33	0.902
Random Forest	88.15	87.62	86.94	87.28	0.918
1D-CNN	91.37	90.84	90.21	90.52	0.946
LSTM	90.68	89.95	89.42	89.68	0.939
CNN-LSTM	92.54	92.11	91.48	91.79	0.957
Transformer	93.16	92.73	92.05	92.39	0.963
Improved CBAM-ResNet	93.48	93.06	92.71	92.88	0.968
TFMT	94.16	93.73	93.38	93.55	0.973
STFT-CNN	94.57	94.18	93.82	94.00	0.976
Model in this paper	95.28	94.86	94.31	94.58	0.981

4.3. Analysis of recognition results for different risk types

In order to further analyze the model's discrimination ability across different types of security risk, a class of statistical analysis was carried out on the test set recognition results. Figure 4 shows the classification distribution of the five types of samples: normal operation, partial discharge, contact abnormal, insulation deterioration, and shielding defect. The values along the main diagonal in Figure 4 are 96.85%, 95.10%, 92.96%, 93.64%, 93.02%, and 93.02% respectively. This shows that the model exhibits a high level of recognition stability for both normal and PD samples. In particular, PD samples typically show concentrated short-duration high-frequency energy in the time-frequency spectrum; after multiscale convolution, they form distinct local active areas, which are highly distinguishable from normal operating samples. The main cause of this phenomenon is that the transient EMP induced by the loose contact is similar to the PD pulse in time domain and high frequency energy. Also, there is some confusion between insulation degradation and shielding defects, as both can lead to energy diffusion across the band and an increase in broadband interference. This is especially the case at the beginning of the degradation, when the electromagnetic response is weak, which makes it difficult to differentiate them on the basis of a single band property, after the attention fusion module applies the weighting to the high-contribution and the sensitive measuring points, it reduces the effect of the background noise and the low response channel, and concentrates the confusion between the different hazard classes on the samples with the same electromagnetic mechanism. This shows that the model's identification results are consistent with the electromagnetic response characteristics of equipment anomalies.

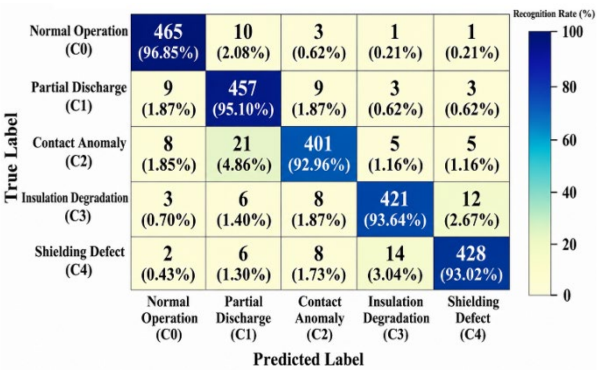

Figure 4. Confusion matrix for different risk types

4.4. Validation of risk point localization performance

Based on the above multichannel electromagnetic acquisition and classification results, the risk-point localization experiment estimates the spatial position of abnormal sources within the equipment. The switchgear connection, cable terminal, insulation support, shielding connection, and other risk-sensitive structures are divided into predefined localization zones according to equipment geometry and sensor coverage. For each sample, the zone containing the calibrated risk point is taken as the ground-truth zone, and candidate zones are ranked according to their localization scores. Top-1 localization is considered correct when the ground-truth zone has the highest localization score, whereas Top-3 localization is considered correct when the ground-truth zone is included among the three highest-scoring candidate zones. Accordingly,

$$Top-k=\frac{1}{N}\sum_{i=1}^{N}\mathrm{I}\left(z_i^{*}\in R_i^{(k)}\right)\times 100\% \quad (25)$$

where N is the number of localization test samples, z_i^* is the ground-truth localization zone of sample i, $R_i^{(k)}$ denotes the set of the kk highest-ranked candidate zones, and I is the indicator function. Thus, Top-1 and Top-3 accuracy evaluate correct zone identification rather than applying a fixed Euclidean-distance threshold. Continuous coordinate accuracy is evaluated separately by the localization error

$$e_i=\left\|\hat{p}_i-p_i\right\|_2 \quad (26)$$

where $\hat{p}_i$ and p_i are the estimated and reference coordinates, respectively. This definition is adopted because equipment inspection and maintenance are organized by structural risk zones, while the Euclidean error independently measures coordinate-level localization precision. Table 4 compares the localization methods according to these criteria. The proposed spatial correlation method achieves an average localization error of 8.46 cm, a Top-1 accuracy of 91.27%, and a Top-3 accuracy of 96.85%.

To further evaluate the localization stability across different risk categories, the localization errors of individual test samples were calculated according to the Euclidean distance between the estimated and reference risk-point coordinates. The category-specific results are summarized in Table 5. Normal operation is not included in the localization error statistics because no abnormal risk point exists. Partial discharge achieves the lowest localization error of 7.92 ± 2.31 cm, followed by contact anomaly with 8.24 ± 2.47 cm. The errors for insulation deterioration and shielding defects are 8.71 ± 2.65 cm and 9.08 ± 2.72 cm, respectively. The overall mean localization error remains 8.46 cm. The relatively limited variation among the four abnormal categories indicates that the spatial correlation localization method maintains stable performance under different electromagnetic response patterns.

Table 4. Comparison of risk point localization results

Localization Method	Average Localization Error/cm	Top-1 Accuracy/%	Top-3 Accuracy (%)	Average Confidence
Maximum Amplitude Channel Localization	18.64	78.35	89.72	0.741
Time-of-Arrival Differential Positioning	15.28	81.46	91.38	0.768
Weighted Center of Mass Positioning	12.73	85.92	93.64	0.804
Spatial Correlation Localization in This Paper	8.46	91.27	96.85	0.873

Table 5. Localization errors for different risk states

Risk State	Localization Error/cm (Mean ± Std)
Normal Operation	—
Partial Discharge	7.92 ± 2.31
Contact Anomaly	8.24 ± 2.47
Insulation Deterioration	8.71 ± 2.65
Shielding Defects	9.08 ± 2.72
Overall	8.46 ± 2.54

4.5. Ablation studies and module contribution analysis

In order to validate the contribution of each module to identify and locate risk points, we designed an ablation experiment. Then, we removed or replaced the multi-scale convolutional, attention, frequency-domain energy, time-frequency, and spatial correlation modules. Table 6 shows the performance metrics for the absence of each module, clearly illustrating the contribution of the modules to the overall performance of the model. It is shown that the removal of the multiscale convolution branch results in inadequate capture of the short duration, high frequency pulse characteristics, which reduces the F1 score by about 2.7%; the removal of the attention module weakens the weighting effect of the key and sensitive channels, resulting in a decrease of approximately 1.8% in precision; the removal of the frequency domain energy characteristics or the time-frequency spectral inputs leads to an inadequate representation of the local energy migration and short-duration pulse patterns, resulting in a reduction in the F1-score of about 2 - 3%; the removal of the spatial correlation module has little effect on the classification performance, but the average localization error has increased to 16.72cm, suggesting that the spatial weight plays a key role in the accuracy of localization. In this paper, we show that multiscale convolutional and attention modules can improve the performance of classification. The spatial correlation module is mainly used to improve the precision of the location of risk points, which is a reliable base for defining the hazard maintenance area in engineering.

Table 6. Comparison of Ablation Experiment Results

Model Variants	Accuracy/%	F1-score/%	Average Localization Error/cm	Module Description
Full Model	95.28	94.58	8.46	Keep All Modules
Remove wavelet denoising	92.81	92.06	11.92	Background noise not sufficiently suppressed
Removal of frequency-domain energy features	93.47	92.88	10.84	Insufficient representation of frequency band shifts
Removal of time-frequency spectrum input	91.96	91.12	13.35	Missing short-term high-frequency features
Remove multiscale convolution	92.63	91.85	12.46	Insufficient capture of perturbations of different durations
Removal of attention modules	93.18	92.37	11.08	Decreased weighting of key frequency bands and sensitive channels
Removal of Spatial Correlation Localization	95.01	94.21	16.72	Minimal impact on classification, increased localization error

4.6. Influence of the number of acquisition channels

To examine whether the 8-channel configuration is necessary and to evaluate the adaptability of the method to different sensing densities, additional experiments are conducted using 4, 6, 8, 10, and 12 electromagnetic channels. The sampling frequency, window segmentation, dataset division, model parameters, and training procedure remain unchanged, and only the number and spatial distribution of sensing channels are adjusted. For reduced-channel configurations, measurement points are selected according to uniform coverage of the principal equipment regions; additional channels are placed between adjacent risk-sensitive regions for the 10- and 12-channel configurations. Table 7 shows that increasing the number of channels from 4 to 8 improves the accuracy from 91.64% to 95.28% and reduces the average localization error from 14.72 cm to 8.46 cm. The improvement is mainly associated with stronger spatial redundancy and more complete inter-channel response gradients. Increasing the number further to 10 and 12 produces only limited gains: the 12-channel configuration reaches an accuracy of 95.51% and a localization error of 7.98 cm, while the inference time increases from 7.54 ms for 8 channels to 10.68 ms. For the default 8-channel configuration, the complete model contains 1.86 M trainable parameters and requires 3.42 GFLOPs for one forward pass. The inference time is measured with a batch size of 1 after 100 warm-up runs and averaged over 1000 forward passes on the RTX 3060 GPU. The resulting mean inference time is 7.54 ms per sample. These results indicate that the 8-channel configuration maintains moderate computational complexity while providing high recognition and localization performance.

Table 7. Recognition and localization performance for different numbers of acquisition channels

Number of Channels	Configuration Level	Accuracy/%	F1-score/%	Average Localization Error/cm	Top-1 Localization Accuracy/%	Inference Time/ms
4	Low	91.64	90.83	14.72	82.46	4.91
6	Medium-low	93.87	93.15	10.63	87.92	6.18
8	Medium	95.28	94.58	8.46	91.27	7.54
10	Medium-high	95.46	94.77	8.12	91.74	9.06
12	High	95.51	94.82	7.98	91.92	10.68

4.7. Limitations and Future Work

The dataset contains 5,430 samples obtained from controlled laboratory measurements, calibration tests, and simulated abnormal configurations. Therefore, the reported results should be interpreted as validation under controlled experimental conditions and cannot be assumed to represent the full variability of field operating environments. Dataset bias may arise from the limited equipment configurations, controlled interference patterns, and predefined risk categories. Sensor deployment is another source of variability because channel number, sensor position, orientation, propagation distance, and shielding structure can alter the measured electromagnetic response and spatial gradients. These factors may reduce generalization when the model is applied to unseen equipment structures, operating loads, noise levels, or defect severities. Future work will incorporate long-term field monitoring data from multiple devices and operating environments, together with cross-device transfer learning, domain adaptation, and adaptive sensor-topology calibration, to evaluate and improve generalization under practical operating conditions.

5. Conclusions

The abnormal electromagnetic response of power equipment is characterized by nonstationary, abrupt change, multi-frequency coupling, and remarkable variation between measuring points. Single domain or frequency domain analysis alone is not enough to identify risk categories at the same time and locate the risk points. The model integrates multichannel electromagnetic preprocessing, time-domain, frequency-domain, and STFT representations with multiscale convolution, attention fusion, and spatial correlation localization. Unlike classification-oriented or standalone localization approaches, multiscale disturbance representation, adaptive channel-frequency weighting, and classification-guided spatial estimation are coupled within a unified optimization process, enabling risk type and location to be inferred from the same electromagnetic evidence. Experimental results show that the model is highly stable in identifying partial discharge, contact anomaly, insulation deterioration, and shielding defects. The results show that the combination of multi-source feature fusion and spatial localization can improve the applicability of the system. Challenges remain at this stage, including limited sample sources, device-dependent sensor deployment, and inadequate cross-device operational state generalization validation. Future work could be extended to include long term on-line monitoring data, to optimize light weight reasoning and edge deployment mechanisms, and to introduce cross-device transfer learning and adaptive sensor topology updating to enhance the capability of real time recognition and localization in complicated operating and maintenance environments. Experimental results show that the model is highly stable in identifying partial discharge, contact anomaly, insulation deterioration, and shielding defects. The results show that the combination of multi-source feature fusion and spatial localization can improve the applicability of the system.

Acknowledgement

This work is financially supported by the Science and technology project of China Southern Power Grid Co., Ltd. (No. GZKJXM 060000KC23110067).

References

[1] Hussain G A, Hassan W, Mahmood F, et al. Review on partial discharge diagnostic techniques for high voltage equipment in power systems[J]. IEEE Access, 2023, 11: 51382-51394.

[2] Kumar H, Shafiq M, Kauhaniemi K, et al. A review on the classification of partial discharges in medium-voltage cables: detection, feature extraction, artificial intelligence-based classification, and optimization techniques [J]. Energies, 2024, 17(5): 1142.

[3] Rauscher A, Kaiser J, Devaraju M, et al. Deep learning and data augmentation for partial discharge detection in electrical machines[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108074.

[4] Carvalho I F, da Costa E G, Nobrega L A M M, et al. Identification of partial discharge sources by feature extraction from a signal conditioning system [J]. Sensors, 2024, 24(7): 2226.

[5] Khodaveisi F, Karami H, Karimpour M Z, et al. Partial discharge localization in power transformer tanks using machine learning methods [J]. Scientific Reports, 2024, 14(1): 11785.

[6] Almehdhar A, Prochazka R. Classification of multiple partial discharge sources using time-frequency analysis and deep learning [J]. Applied Sciences, 2025, 15(10): 5455.

[7] Li A, Wei G, Li S, et al. Pattern recognition of partial discharge in high-voltage cables using TFMT model [J]. IEEE Transactions on Power Delivery, 2024, 39(6): 3326-3337.

[8] Das R, Das A K, Chatterjee S, et al. A novel deep-learning framework to identify and locate single and multiple partial discharge events [J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2023, 30(6): 2633-2641.

[9] Govindarajan S, Morales A, Ardila-Rey J A, et al. A review on partial discharge diagnosis in cables: Theory, techniques, and trends [J]. Measurement, 2023, 216: 112882.

[10] Tag A, Refaat S S, Kameli S M, et al. Machine learning applications for online partial discharge detection, classification, and localization in power transformers: A review[C]//2024 4th International Conference on Smart Grid and Renewable Energy (SGRE). IEEE, 2024: 1-6.

[11] Thobejane L T, Thango B A. Partial discharge source classification in power transformers: A systematic literature review [J]. Applied Sciences, 2024, 14(14): 6097.

[12] Sahoo R, Karmakar S. Comparative analysis of machine learning and deep learning techniques on classification of artificially created partial discharge signal [J]. Measurement, 2024, 235: 114947.

[13] Hassan W, Hussain G A, Wahid A, et al. Optimum feature selection for classification of PD signals produced by multiple insulation defects in electric motor s[J]. Scientific Reports, 2024, 14(1): 23446.

[14] Ortego J, Garnacho F, Álvarez F, et al. Locating Insulation

Defects in HV Substations Using HFCT Sensors and AI Diagnostic Tools[J]. Sensors, 2024, 24(16): 5312.
[15] Dalila R A M, Turkben A K. Artificial intelligence based partial discharge detection using CNN and KNN to increase the quality of electrical insulation [J]. Discover Computing, 2025, 28(1): 101.
[16] Kaziz S, Said M H, Imburgia A, et al. Radiometric partial discharge detection: A review [J]. Energies, 2023, 16(4): 1978.
[17] Jung H, Kim Y T, Lee S K, et al. Study on deep-learning model for phase resolved partial discharge pattern classification based on convolutional neural network algorithm [J]. Journal of Electrical Engineering & Technology, 2025, 20(1): 873-878.
[18] Aldosari O, Aldowsari M A, Batiyah S M, et al. Image-based partial discharge identification in high voltage cables using hybrid deep network [J]. Ieee Access, 2023, 11: 50325-50333.
[19] Liu K, Jiao S, Nie G, et al. On image transformation for partial discharge source identification in vehicle cable terminals of high-speed trains [J]. High Voltage, 2024, 9(5): 1090-1100.
[20] Hu D, Chen Z, Yang W, et al. A GIS Partial Discharge Pattern Recognition Method Based on Improved CBAM-ResNet [J]. Journal of Electrical and Computer Engineering, 2023, 2023(1): 9948438.
[21] Zhang C, Chen M, Zhang Y, et al. Partial discharge pattern recognition algorithm of overhead covered conductors based on feature optimization and bidirectional LSTM-GRU [J]. IET Generation, Transmission & Distribution, 2024, 18(4): 680-693.
[22] Beura C P, Wolters J, Tenbohlen S. Application of pathfinding algorithms in partial discharge localization in power transformers [J]. Sensors, 2024, 24(2): 685.
[23] Monzón-Verona J M, González-Domínguez P, García-Alonso S. Characterization of partial discharges in dielectric oils using high-resolution CMOS image sensor and convolutional neural networks [J]. Sensors, 2024, 24(4): 1317.
[24] Klein L, Seidl D, Fulneček J, et al. Antenna contactless partial discharges detection in covered conductors using ensemble stacking neural networks[J]. Expert Systems with Applications, 2023, 213: 118910.
[25] Saad M H, Hashima S, Omar A I, et al. Deep learning approach for cable partial discharge pattern identification [J]. Electrical Engineering, 2025, 107(2): 1525-1540.