

Flashover Risk Prediction And Predictive Maintenance Of Contact Boxes In 10 KV Switchgear Based On Artificial-Intelligence-Driven Multi-Sensor Fusion

Xielin Shen¹, Yipeng Wang¹, Xiaomei Pan¹, Bo Yuan¹, Lei Wang^{2,3}, and Chaofei Gao^{2,3*}

¹ Quanzhou Power Supply Company, State Grid Fujian Electric Power Co., Ltd., Quanzhou, 362064, China

² College of Automation (College of Artificial Intelligence), Beijing Information Science & Technology University, Beijing, 100192, China

³ Beijing Key Laboratory of Low-Carbon and Flexible Urban Power Supply Technology and Equipment, Beijing, 100192, China

* Corresponding author. E-mail: gaocf@bistu.edu.cn

Received: Aug. 01, 2026; Accepted: Aug. 25, 2026

To address the difficulty of uniformly characterizing the coupled evolution of partial discharge, thermal accumulation, surface conduction, and environmental stress in contact boxes of 10 kV switchgear, as well as the limited capability of conventional methods for early flashover prediction, a physics-consistent hierarchical sensor-temporal fusion network (PC-HSTFNet) is proposed. The method integrates TEV/UHF/HFCT, ultrasonic, infrared and point-temperature, leakage-current, temperature-humidity, and operating-condition information. Modality-specific encoding, dynamic weighting of homogeneous sensors, reliability-biased cross-modal attention, and causal temporal modeling are combined to identify degradation stages, predict multi-horizon flashover probabilities, and estimate the remaining safe operating time. A dataset containing 28,640 multi-sensor time windows was constructed from 18 contact-box specimens and 72 controlled test runs. PC-HSTFNet achieved a Macro-F1 of 0.931, an S4 recall of 0.927, and a 5-min risk AUPRC of 0.902. Its Brier score and ECE were 0.052 and 0.016, respectively, while the MAE of the remaining safe operating time was 6.8 min. Compared with the Multimodal Transformer, Macro-F1 and AUPRC increased by 2.6 and 3.3 percentage points, respectively. Under a representative contamination-condensation condition, the model provided an average early-warning lead time of 18.9 min, with false-alarm and missed-alarm rates of 3.4% and 1.6%. These results demonstrate that the proposed method can support reliable flashover warning and predictive maintenance of switchgear contact boxes.

Keywords: 10 kV switchgear; contact box; flashover risk prediction; multi-sensor fusion; PC-HSTFNet; predictive maintenance

© The Author(s). This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202701_36.026

1. Introduction

The 10 kV metal-clad switchgear is a key component of distribution systems, performing power distribution, control, and protection. The contact box provides mechanical support for the fixed contact as well as phase-to-phase and phase-to-ground insulation, and is continuously subjected to power-frequency electric fields, load-induced heat-

ing, mechanical stress, and variations in temperature and humidity. Manufacturing defects, installation deviations, poor contact, surface contamination, and condensation can distort the local electric field and reduce insulation resistance, thereby inducing partial discharge, increased leakage current, local temperature rise, and surface carbonization. In severe cases, a conductive discharge path may form and

cause surface flashover [1].

Partial-discharge signals in switchgear are weak, non-stationary, and highly susceptible to field noise. Although adaptive denoising and intelligent feature extraction improve signal distinguishability [2], a single sensing modality cannot fully characterize the insulation degradation of the contact box. TEV and UHF signals are sensitive to discharge pulses, ultrasonic signals exhibit strong immunity to electromagnetic interference, and infrared temperature, temperature-humidity, and load data reflect thermal, environmental, and operating stresses, respectively. Existing studies have improved insulation-defect recognition through semi-supervised learning, partial-discharge image fusion, and highly sensitive acoustic detection [3–5]; however, most methods still focus on current fault-type identification and do not jointly analyze the temporal evolution of multiple physical quantities.

With advances in artificial intelligence, edge computing, and digital twins, power-equipment maintenance is shifting from periodic inspection and post-fault repair toward continuous condition monitoring and predictive maintenance. Studies on digital-twin-assisted partial-discharge analysis and virtual-physical integration in substations have shown that the coordination of real-time sensing data, intelligent models, and virtual equipment replicas can support condition prediction and maintenance decisions [6, 7]. Nevertheless, the flashover risk associated with the same partial-discharge pattern can vary substantially with humidity, contamination level, load, and insulation-aging stage. Instantaneous discharge amplitude or discrete fault labels alone are therefore insufficient to determine future flashover probability and safe operating time.

Accordingly, this study integrates partial-discharge, acoustic, thermal, environmental, and operating-condition information to establish a multi-sensor temporal risk-prediction method for contact boxes in 10 kV switchgear. The method extends static fault identification to dynamic insulation-degradation analysis and provides a basis for flashover early warning, remaining-safe-operating-time assessment, and maintenance-timing optimization, thereby reducing unexpected outage risk and unnecessary maintenance.

In addition to partial discharge, contact overheating and local temperature rise are important indicators of abnormal switchgear contact systems. Tang and Jian diagnosed thermal faults in complex electrical equipment through infrared-image denoising, object detection, semantic segmentation, and temperature-difference analysis, demonstrating that deep vision models can extract abnormal hot spots and temperature-distribution features from thermal

images [8]. To address distribution mismatch among multimodal data under varying operating conditions, Zhang et al. processed vibration signals and infrared images using one-dimensional and two-dimensional convolutional networks, respectively, and improved cross-condition diagnosis through domain alignment and cross-modal consistency learning [9]. Although these methods mainly target general electrical or mechanical equipment, their multimodal feature extraction and cross-domain fusion strategies can be transferred to the joint analysis of TEV, ultrasonic, infrared, and environmental data from contact boxes.

Intelligent optimization has also been used to improve parameter selection in power-equipment fault models. Xu and Lu optimized a probabilistic neural network using an improved grey wolf optimizer, reducing the likelihood of convergence to local optima and improving the stability of transformer fault diagnosis [10]. This study indicates that metaheuristic algorithms can optimize key hyperparameters of classifiers or fusion models; however, its target remained static fault categories rather than multi-sensor temporal risk prediction.

To support synchronous monitoring and rapid analysis of large numbers of switchgear units, Jia et al. developed a cloud-edge collaborative partial-discharge recognition architecture. UHF and broadband pulse-current sensors were used to acquire discharge signals, feature extraction and dimensionality reduction were conducted at the edge, and a deep belief network was trained in the cloud [11]. Although this architecture improves model training and multi-device collaborative analysis, its output is primarily the discharge-defect category rather than future flashover probability.

The objective of multi-sensor information fusion is to reduce the uncertainty of individual sensors by exploiting complementary observations. Lin et al. systematically reviewed data-level, feature-level, and decision-level fusion methods for equipment fault diagnosis and identified sensor heterogeneity, time synchronization, data-quality differences, missing modalities, and cross-condition generalization as major challenges [12]. To avoid separately modeling partial-discharge type and severity, Li et al. proposed a gradient-balanced selective mixture-of-experts network that performs phase-level fusion of acoustic-emission and UHF PRPS data and jointly identifies discharge patterns and severity [13]. This finding indicates that multi-sensor fusion can characterize not only defect types but also risk levels, providing a methodological basis for contact-box flashover prediction.

Optical-electromagnetic fusion is an important direction in multimodal partial-discharge detection. Li et al.

designed a combined UHF and light-guide-rod detection device and fused optical and UHF PRPS patterns using guided filtering, thereby exploiting the complementary advantages of the two detection methods [14]. The effectiveness of multi-sensor fusion also depends on sensor condition. Cao et al. identified sensitivity degradation in GIS UHF sensors by cross-comparing S-parameters, showing that sensor aging or failure can directly affect the reliability of online monitoring data [15]. A multi-sensor risk model should therefore dynamically evaluate signal-to-noise ratio, drift, missing-data rate, and inter-channel consistency rather than assuming that all channels remain reliable. Class imbalance is another practical challenge because normal and mildly abnormal samples greatly outnumber critical-fault samples. Lv et al. combined a multi-kernel graph convolutional network with dual-channel classifiers and improved minority-class recognition through sample-level reweighting [16], which is relevant to contact-box risk prediction with scarce flashover samples.

Laboratory-simulated and field-operating data differ considerably in noise level, discharge density, equipment structure, and sensor installation. Liu et al. proposed a domain-adversarial ConvNeXt model that reduces PRPD distribution shift across test conditions through domain discrimination and adversarial training, thereby improving generalization on multi-condition GIS data [17]. This result indicates that high accuracy obtained by randomly splitting data from a single test platform may overestimate engineering performance; cross-device, cross-condition, or laboratory-to-field validation is therefore necessary.

Regarding the utilization of raw information, Pan et al. constructed a ConvNeXt model based on GIS partial-discharge mechanisms and full-wave UHF signals, avoiding information loss caused by the use of only a small number of manually designed discrete features [18]. He et al. combined time-frequency energy grayscale maps with an improved variational Bayesian autoencoder to model feature distributions of noisy signals while recognizing discharge patterns [19]. Jia et al. fused PRPD statistical features measured by HFCT with UHF frequency-domain features and performed multi-source GIS insulation-defect diagnosis using PCA and a ResNet-MLP model [20]. These studies show that jointly modeling complete waveforms, time-frequency images, and multi-source features yields a more comprehensive representation of insulation condition than individual handcrafted features.

For temperature-rise detection, Wen et al. proposed a deep multimodal imaging temperature-measurement method that aggregates infrared and visible-light imaging information through variational inference and dynamic

weighting, enabling the model to emphasize higher-quality modalities [21]. This further demonstrates that learnable multimodal weighting is more appropriate than fixed concatenation for condition monitoring of complex electrical equipment.

Early multi-source partial-discharge studies focused on the joint discrimination of different detection signals. Li et al. applied multi-information fusion to GIS partial-discharge recognition and verified the complementarity of heterogeneous sensor data [22]. Hussein et al. developed robust feature extraction and classification for acoustic partial-discharge signals under strong noise [23]. Zhang et al. monitored partial discharge in metal-enclosed switchgear using distributed non-contact sensors, showing that multi-point deployment improves the observability of internal discharge activity [24]. These studies established foundations for multi-sensor data acquisition and feature fusion.

With the development of edge intelligence and deep learning, research has shifted toward lightweight models and automatic feature extraction. Wang et al. proposed a lightweight convolutional neural network that balances GIS partial-discharge recognition accuracy and computational efficiency in a ubiquitous power-IoT environment [25], and subsequently used MixNet to improve multi-scale feature extraction from discharge patterns [26]. Fechet et al. developed a switchgear-maintenance application using in-cabinet microclimate sensors and introduced temperature and humidity into partial-discharge monitoring, confirming the importance of environmental information for interpreting insulation-risk variations [27].

Subsequently, partial-discharge diagnosis expanded from a single deep representation toward the fusion of statistical, image, and condition features. Yin et al. fused PRPD statistical and image features and applied evidence theory for comprehensive GIS partial-discharge decision-making [28]. Wu et al. assessed defects and health conditions of GIS epoxy insulators using multi-source information [29]. Cordoni et al. fed electrical-power signals and thermal images into a deep autoencoder for multimodal unsupervised fault detection [30]. Although these studies advanced equipment diagnosis from single-fault classification toward comprehensive health assessment, they did not establish a complete chain of degradation-stage identification, flashover-probability prediction, remaining-safe-time estimation, and maintenance strategy for contact boxes.

In summary, existing studies have five main limitations. First, most methods output current partial-discharge types or health grades but do not predict future-window flashover probability or remaining safe operating time. Sec-

ond, partial discharge, thermal condition, temperature-humidity, load, and operating voltage have not been modeled within a unified multi-physical framework. Third, fixed weights and direct feature concatenation are vulnerable to sensor noise, drift, and missing modalities. Fourth, laboratory-to-field distribution shift, scarce critical-flashover samples, and class imbalance remain significant. Fifth, model outputs are not directly linked to maintenance timing, maintenance level, and maintenance cost.

Based on these limitations, the contact box of 10 kV switchgear is regarded as a dynamic degradation system jointly driven by electrical, thermal, environmental, and insulation-aging stresses. The objective is extended from current partial-discharge-type recognition to multi-sensor temporal degradation modeling, future flashover-probability prediction, remaining-safe-operating-time estimation, and maintenance decision-making, thereby forming a closed-loop framework of condition sensing, reliability-aware fusion, risk prediction, uncertainty assessment, and predictive maintenance.

The main contributions of this study are as follows:

1. A synchronous sensing system comprising five modalities—electromagnetic, acoustic, thermal, leakage current, and environmental/operating conditions—is constructed, and three prediction tasks are consistently defined: degradation-stage identification, multi-horizon flashover-risk prediction, and remaining-safe-operating-time estimation.
2. A physics-consistent hierarchical sensor-temporal fusion network, PC-HSTFNet, is proposed. Modality-specific encoding, dynamic weighting of homogeneous sensors, and reliability-biased cross-modal attention reduce the effects of noise, channel drift, and missing modalities on risk representation.
3. A multi-task prediction method integrating degradation-stage identification, discrete-time survival analysis, and probability calibration is established. Trend-consistency and stage-order constraints improve the physical plausibility of flashover probability and remaining-safe-operating-time estimates.
4. A graded maintenance strategy is constructed from multi-horizon risk, remaining safe operating time, and prediction uncertainty. Ablation, noise-disturbance, modality-missing, and cross-switchgear tests are conducted to verify reliability and engineering applicability.

2. Materials and methods

2.1. Flashover Evolution Mechanism

Contact boxes in 10 kV switchgear are generally molded from epoxy-resin insulating materials and are continuously exposed to power-frequency electric fields, load heating, mechanical stress, and environmental temperature-humidity variations. Internal bubbles, interface debonding, surface cracks, fixed-contact eccentricity, and sharp metal edges may cause local electric-field concentration, while contamination and condensation reduce insulation resistance and increase leakage current, ultimately inducing partial discharge and surface tracking [1, 2]. The surface current density of the contact box can be expressed as:

$$\mathbf{J}_s(t) = \sigma_s [H(t), T_s(t), c_p(t)] \mathbf{E}_s(t) \quad (1)$$

where the first two terms denote the surface current density and tangential electric-field strength, respectively; the third term is the equivalent surface conductivity; and H , T_s , and c_p denote the in-cabinet relative humidity, contact-box surface temperature, and contamination level, respectively. Increasing humidity or contamination increases the equivalent surface conductivity and consequently increases leakage current and local Joule heating. Resistive losses caused by poor contact, surface leakage losses, and partial-discharge energy jointly determine the thermal state of the contact box. A simplified thermal-balance relationship is:

$$C_{th} \frac{dT_s(t)}{dt} = I^2(t)R_c(t) + P_{leak}(t) + P_{PD}(t) - \frac{T_s(t) - T_a(t)}{R_{th}} \quad (2)$$

where C_{th} and R_{th} are the equivalent thermal capacitance and thermal resistance, respectively; I is the load current; R_c is the contact resistance; P_{leak} and P_{PD} are the equivalent thermal powers generated by leakage current and partial discharge, respectively; and T_a is the ambient temperature inside the cabinet.

Sustained partial discharge erodes, roughens, and carbonizes the insulating material, thereby further increasing the local electric field and surface conductivity. Under humid and contaminated conditions, nonuniform leakage-current distribution may form local dry bands, and voltage concentration across the dry-band edges can initiate intermittent arcing. As discharge energy and carbonized paths accumulate, the insulation state evolves from initial defects to intermittent partial discharge, sustained discharge and temperature rise, dry-band arcing and surface carbonization, and finally surface flashover. The corresponding electromagnetic, acoustic, thermal, and environmental responses are illustrated in Fig. 1.

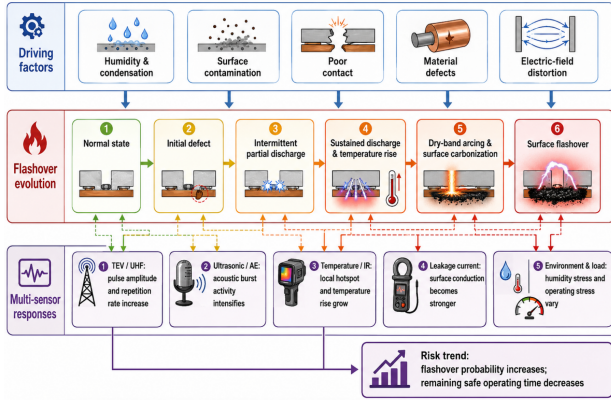


Fig. 1. Flashover evolution mechanism and multi-sensor responses of the contact box in 10 kV switchgear.

2.2. Multi-Sensor Condition Observation and Data Representation

The true insulation degradation state of the contact box cannot be directly obtained using a single sensor. Let the latent degradation state and operating condition at time t be z_t and u_t , respectively. The observation of the m -th sensing modality is expressed as:

$$\mathbf{x}_t^{(m)} = h_m(\mathbf{z}_t, \mathbf{u}_t) + \varepsilon_t^{(m)} \quad (3)$$

where $h_m(\cdot)$ is the mapping from the degradation state to the m -th sensing signal, and $\varepsilon_t^{(m)}$ denotes measurement noise and environmental interference. Different sensors observe different physical responses of the same degradation process; their complementary information is summarized in Table 1.

Existing studies on multi-sensor partial-discharge recognition and equipment-condition assessment have shown that complementary detection information improves diagnostic completeness and interference immunity [12–15, 22, 24, 29]. However, modalities differ significantly in sampling frequency, dimension, noise level, and data reliability, and direct concatenation can introduce redundant or low-quality information. Therefore, multi-sensor data are organized using sliding time windows, and sensor-validity indicators are retained. The model input at time t is expressed as:

$$\mathbf{X}_t = \left\{ \mathbf{X}_{t-L+1:t}^{(m)}, \mathbf{r}_{t-L+1:t}^{(m)} \right\}_{m=1}^M \quad (4)$$

where L is the time-window length and M is the number of sensing modalities. The validity mask indicates missing data, sensor anomalies, or inadequate signal quality. High-frequency electromagnetic and acoustic data characterize rapid discharge changes; infrared and point-temperature data characterize thermal accumulation; leakage current

characterizes surface conduction; and temperature, humidity, condensation, load current, and operating voltage describe slowly accumulating environmental and operating stresses. Together, these variables form a multi-timescale risk representation.

2.3. Definition of Flashover-Risk Prediction Tasks

Unlike conventional diagnosis that only determines the current discharge type, this study formulates contact-box analysis as three related tasks: degradation-stage identification, multi-horizon flashover-probability prediction, and remaining-safe-operating-time estimation.

First, according to partial discharge, leakage current, temperature rise, and surface condition, the degradation process is divided into normal state S_0 , early anomaly S_1 , developing degradation S_2 , severe degradation S_3 , and critical flashover S_4 . These stage labels reflect the current insulation-health level but do not independently describe future risk.

Second, let T_f denote the time of actual flashover or the critical time at which shutdown is required. For a prediction horizon Δ , the flashover-risk label at time t is defined as:

$$y_t^{(\Delta)} = \begin{cases} 1, & 0 < T_f - t \leq \Delta \\ 0, & T_f - t > \Delta \end{cases} \quad (5)$$

By setting multiple horizons Δ_1 , Δ_2 , and Δ_3 , the model outputs both short-term emergency risk and medium-term maintenance risk. Short horizons support emergency shutdown or load transfer, whereas longer horizons support live inspection and planned outage scheduling.

Finally, the remaining safe operating time of the contact box at time t is defined as:

$$\tau_t = T_f - t \quad (6)$$

Samples that do not experience flashover before the end of observation are treated as right-censored data, preventing incompletely observed degradation processes from being incorrectly labeled as long-term safe. The model outputs the degradation stage, flashover probabilities at different horizons, and the remaining safe operating time.

This subsection defines the contact-box flashover problem from the perspectives of physical mechanism, sensor observation, and prediction objectives. Based on these definitions, Section 2.4 develops a reliability-aware multi-sensor temporal fusion model for dynamic sensor weighting, cross-modal feature interaction, and joint prediction of flashover risk and remaining safe operating time.

Table 1. Multi-sensor Observation Information for Flashover Risk of Switchgear Contact Boxes

Sensing Modality	Key Observation Parameters	Risk Characterization Function
Transient Earth Voltage (TEV), Ultra-high Frequency (UHF) or High Frequency Current Transformer (HFCT)	Pulse amplitude, Repetition rate, Phase-resolved distribution and frequency spectrum	Characterize partial discharge intensity and discharge pattern
Ultrasonic or Acoustic Emission (AE)	Impulse waveform, Frequency band energy and pulse density	Supplement electro-acoustic information and mitigate electromagnetic interference
Infrared Thermography or Point-type Temperature Sensor	Surface temperature, Hot spot location and temperature rise rate	Characterize contact overheating and local thermal accumulation
Leakage Current Sensor	Leakage current amplitude, RMS value, growth rate and harmonic content	Characterize surface conduction and tracking development
Environment and Operating-condition Sensors	Ambient temperature, relative humidity, condensation duration, load current and operating voltage	Characterize environmental and operating stresses

2.4. Reliability-Aware Multi-Sensor Temporal Flashover-Risk Prediction Method

To address inconsistent multi-sensor signal quality, redundant cross-modal information, and the limitation of conventional models to current-fault identification, a physics-consistent hierarchical sensor-temporal fusion network (PC-HSTFNet) is proposed. The network sequentially performs modality-specific feature extraction, reliability fusion of homogeneous sensors, heterogeneous-modal interaction, temporal degradation modeling, and multi-task risk prediction, and jointly outputs the contact-box degradation stage, flashover probabilities at different horizons, and remaining safe operating time.

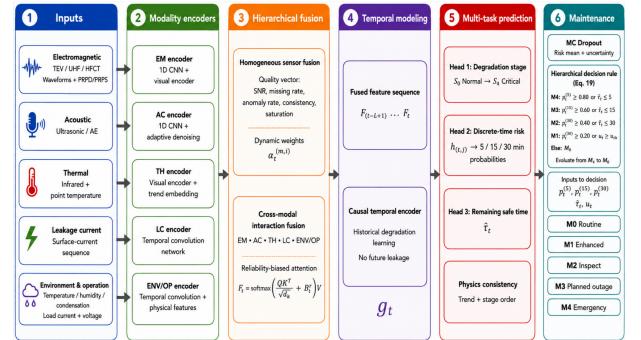
2.4.1. Overall Model Framework

The model input is the multi-sensor time-window sample X_t defined in Section 2.3, including the electromagnetic modality (TEV/UHF/HFCT waveforms and PRPD/PRPS patterns), acoustic modality (ultrasonic/AE signals), thermal modality (infrared images and point temperatures), leakage-current modality, and environmental/operating-condition modality (temperature-humidity, condensation, load current, and operating voltage). PC-HSTFNet comprises modality-specific encoders, a reliability-aware fusion module, a cross-modal interaction module, a causal temporal encoder, and multi-task prediction heads, as shown in Fig. 2.

For the i -th sensor channel of the m -th modality, the unified feature representation is:

$$\mathbf{H}_t^{(m,i)} = \mathcal{E}_m \left(\mathbf{X}_t^{(m,i)}, \mathbf{r}_t^{(m,i)} \right) \quad (7)$$

where $\mathcal{E}_m(\cdot)$ is the modality-specific encoder, $\mathbf{r}_t^{(m,i)}$ is the data-validity mask, and $\mathbf{H}_t^{(m,i)} \in \mathbb{R}^{L_m \times d}$ is the encoded feature. L_m and d are the feature-sequence length and unified embedding dimension, respectively.

**Fig. 2.** Overall architecture of the proposed PC-HSTFNet.

2.4.2. Modality-Specific Feature Extraction and Adaptive Denoising

High-Frequency Discharge and Acoustic Waveform Encoding

TEV, UHF, HFCT, and ultrasonic signals are nonstationary, pulse-sparse, and affected by complex noise. Parallel one-dimensional convolutions with different dilation rates are used to extract local impulses, discharge clusters,

and long-range dependencies, followed by a learnable soft-thresholding module. Let the convolutional feature of the m -th modality be $\mathbf{F}_t^{(m,i)}$, its denoised representation can be expressed as:

$$\tilde{\mathbf{F}}_t^{(m,i)} = \text{sign}(\mathbf{F}_t^{(m,i)}) \odot \max(|\mathbf{F}_t^{(m,i)}| - \theta_t^{(m,i)}, 0) \quad (8)$$

where $\theta_t^{(m,i)}$ is a nonnegative threshold adaptively generated from the channel features, and $\odot$ denotes element-wise multiplication. The module suppresses low-amplitude noise and redundant components while preserving discriminative positive and negative discharge pulses. The combination of multi-scale convolution, attention, and soft thresholding improves feature representation for noisy partial-discharge signals.

Encoding of Images and Low-Frequency Operating Sequences

PRPD/PRPS patterns and infrared images are processed using a lightweight hierarchical visual encoder to extract local textures, phase distributions, hot-spot regions, and global structural features. PRPD/PRPS features are assigned to the electromagnetic modality, whereas infrared-image and point-temperature features are assigned to the thermal modality. Leakage-current sequences are encoded by an independent temporal convolutional network. Temperature-humidity, condensation, load current, and operating voltage jointly form the environmental/operating-condition modality to extract change rates, cumulative effects, and long-term trends. Handcrafted features, including discharge-pulse count, kurtosis, band energy, temperature-rise rate, and humidity duration, are mapped by a multilayer perceptron and added to the deep features of the corresponding modality, thereby combining automatic feature learning with physical interpretability.

2.4.3. Reliability-Aware Hierarchical Multi-Sensor Fusion

To reduce the influence of low-quality sensor channels and redundant modalities on risk representation, a hierarchical fusion module consisting of dynamic weighting of homogeneous sensors and heterogeneous cross-modal interaction is designed, as detailed in Fig. 3.

Dynamic Weighting of Homogeneous Sensors

Sensor points within the same modality may have different information quality because of installation position, propagation path, noise level, or sensor drift. Channel weights are dynamically calculated from feature content and sensor-quality vectors. The reliability score and normalized weight of the i -th sensor are:

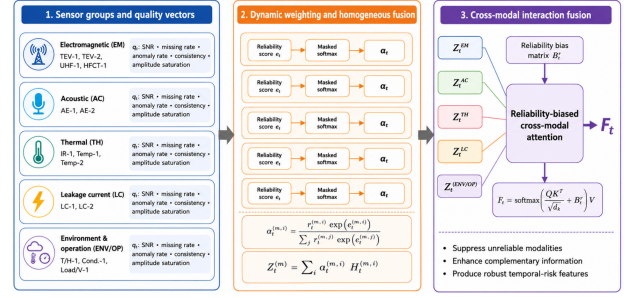


Fig. 3. Detailed structure of the reliability-aware hierarchical multi-sensor fusion module.

$$e_t^{(m,i)} = \mathbf{w}_m^T \left[\text{Pool}(\mathbf{H}_t^{(m,i)}) \parallel \mathbf{q}_t^{(m,i)} \right] + b_m, \quad (9)$$

$$\alpha_t^{(m,i)} = \frac{r_t^{(m,i)} \exp(e_t^{(m,i)})}{\sum_{j=1}^{N_m} r_t^{(m,j)} \exp(e_t^{(m,j)}) + \varepsilon}$$

where $\mathbf{q}_t^{(m,i)}$ consists of quality indicators including signal-to-noise ratio, missing rate, anomaly rate, inter-channel consistency, and amplitude saturation; N_m is the number of sensors in the m -th modality; and $r_t^{(m,i)} \in \{0, 1\}$ is the validity indicator.

The fused homogeneous-sensor feature is expressed as:

$$\mathbf{Z}_t^{(m)} = \sum_{i=1}^{N_m} \alpha_t^{(m,i)} \mathbf{H}_t^{(m,i)} \quad (10)$$

When a channel is missing or its quality decreases, its weight is automatically reduced, whereas channels with high signal-to-noise ratios and strong agreement with other sensor points receive larger weights. Thus, fusion weights vary dynamically with each sample and operating condition rather than remaining fixed after training. Adaptive weighting based on measurement error and channel quality reduces the deviation between the fused representation and the true state.

Heterogeneous Cross-Modal Interaction Fusion

After homogeneous fusion, the modality features are mapped to a common dimension and arranged as a modality-token matrix:

$$\mathbf{Z}_t = [\mathbf{Z}_t^{(1)}; \mathbf{Z}_t^{(2)}; \dots; \mathbf{Z}_t^{(M)}] \quad (11)$$

Reliability-biased cross-modal attention is then used to learn complementary relationships among the electromagnetic, acoustic, thermal, leakage-current, and environmental/operating-condition modalities:

$$\mathbf{F}_t = \text{softmax} \left(\frac{\mathbf{Q}_t \mathbf{K}_t^T}{\sqrt{d_k}} + \mathbf{B}_t^r \right) \mathbf{V}_t \quad (12)$$

where $\mathbf{Q}_t$, $\mathbf{K}_t$, and $\mathbf{V}_t$ are linearly mapped from $\mathbf{Z}_t$; d_k is the key-vector dimension; and $\mathbf{B}_t^r$ is the attention-bias matrix constructed from modality reliability. The mechanism adapts modality contributions to the degradation stage. Electromagnetic and acoustic modalities dominate in the early-defect stage before a clear temperature rise occurs. As sustained discharge produces hot spots and stronger surface conduction, the contributions of the thermal and leakage-current modalities increase, while the environmental/operating-condition modality provides an external-stress explanation. Compared with direct feature concatenation, hierarchical fusion retains complementary information while reducing redundant features and interference from low-quality modalities.

2.4.4. Physics-Consistent Multi-Task Risk Prediction

Cross-modal features are fed chronologically into a causal temporal encoder to learn the evolution from partial discharge and thermal accumulation to critical flashover. A causal mask ensures that prediction at time t depends only on current and historical information. The encoder output is connected to a degradation-stage identification head and a discrete-time survival-prediction head.

Discrete-Time Flashover Risk

The future is divided into J consecutive intervals, and the model outputs the conditional flashover hazard in the j -th interval:

$$\hat{h}_{t,j} = \text{sigmoid}(\mathbf{w}_{h,j}^T \mathbf{g}_t + b_{h,j}), \quad j = 1, \dots, J \quad (13)$$

where $\mathbf{g}_t$ is the temporal-encoder output, and $\hat{h}_{t,j}$ is the probability of flashover in the j -th interval conditional on no previous flashover. The corresponding survival probability and cumulative flashover probability within the future horizon Δ_k are:

$$\hat{S}_{t,j} = \prod_{\ell=1}^j (1 - \hat{h}_{t,\ell}), \quad \hat{p}_t^{(\Delta_k)} = 1 - \hat{S}_{t,J_k} \quad (14)$$

where J_k is the number of intervals corresponding to Δ_k . Because the survival probability decreases naturally over time, Eq. (14) ensures that the cumulative risk over a longer horizon is not lower than that over a shorter horizon. The expected remaining safe operating time is:

$$\hat{\tau}_t = \delta \sum_{j=0}^J \hat{S}_{t,j} \quad (15)$$

where δ is the duration of one discrete interval. The discrete survival model can simultaneously use complete-flashover and right-censored samples, preventing equipment that has not flashed over by the end of observation from being incorrectly labeled as long-term safe.

Physics-Consistency Constraint

Let D_t be a composite degradation indicator constructed from standardized partial-discharge activity, leakage current, temperature rise, and humidity stress, and let $\hat{R}_t$ be the composite flashover risk output by the model. When D_t increases over consecutive windows while the predicted risk decreases, the prediction conflicts with insulation-degradation behavior. A joint temporal-trend and stage-order constraint is therefore constructed:

$$\begin{aligned} L_{\text{phy}} = & \frac{1}{N_T} \sum_t \max [0, -\Delta \hat{R}_t \Delta D_t] \\ & + \frac{1}{|\mathcal{P}|} \sum_{(a,b) \in \mathcal{P}} \max [0, \gamma - (\hat{R}_a - \hat{R}_b)] \end{aligned} \quad (16)$$

where Ω is the set of sample pairs with a known degradation order, sample a has a higher actual degradation degree than sample b , and γ is the minimum risk margin. For reproducibility, each constituent of the composite degradation indicator D_t is normalized to $[0, 1]$ using statistics estimated from the training subset only. The normalized partial-discharge activity, leakage-current level, temperature-rise term, and humidity-stress term are combined as $D_t = 0.35D_{PD} + 0.25D_{LC} + 0.25D_T + 0.15D_H$, respectively, and the risk-margin parameter is set to $\gamma = 0.10$. The normalization parameters and component weights are fixed after training-data preprocessing and are reused unchanged for validation and test samples, thereby avoiding information leakage. The first term of Eq. (16) enforces consistency between multi-physical degradation trends and risk changes, whereas the second preserves the known degradation-stage ordering.

2.4.5. Joint Loss and Predictive Uncertainty

The model is trained in a multi-task manner with the total loss:

$$L = \lambda_1 L_{\text{stage}} + \lambda_2 L_{\text{surv}} + \lambda_3 L_{\text{phy}} + \lambda_4 L_{\text{cal}} \quad (17)$$

where L_{stage} denotes the class-weighted classification loss for the five degradation stages, L_{surv} is the discrete-time survival negative log-likelihood incorporating right-censored samples, L_{phy} is the physics-consistency loss defined in Eq. (16), and L_{cal} is the probability-calibration loss. The coefficients λ_1 – λ_4 control the relative contributions of the four objectives and are set to $\lambda_1 = 1.00$, $\lambda_2 = 1.00$, $\lambda_3 = 0.20$, and $\lambda_4 = 0.10$, respectively. Class-dependent weights are incorporated only into L_{stage} to alleviate degradation-stage imbalance, particularly the relatively scarce S_4 critical-flashover samples. The loss weights are selected on the validation set by jointly considering Macro-F1, AUPRC,

Table 2. Risk Levels and Predictive Maintenance Measures for Switchgear Contact Boxes

Maintenance Level	Predicted State	Recommended Measures
M0	No M1–M4 criterion is satisfied	Maintain routine online monitoring
M1	30-min risk ≥ 0.20 or normalized uncertainty ≥ 0.15 , with no higher-level criterion satisfied	Increase sampling frequency and verify sensor operating status
M2	30-min flashover risk ≥ 0.40 or predicted RST ≤ 30 min	Conduct live inspection, dehumidification and insulation surface cleaning
M3	15-min flashover risk ≥ 0.60 or predicted RST ≤ 15 min	Arrange planned power outage for contact box inspection or replacement
M4	5-min flashover risk ≥ 0.80 or predicted RST ≤ 5 min	Implement emergency shutdown or load transfer

ECE, and remaining-safe-operating-time MAE, without reference to test-set performance. The risk-margin parameter γ in Eq. (16) is independent of the loss weighting coefficients in Eq. (17). A one-factor-at-a-time sensitivity analysis therefore varies γ and each λ separately while keeping the remaining parameters fixed. The tested ranges are $\gamma = 0.05\text{--}0.20$, $\lambda_1 = 0.75\text{--}1.25$, $\lambda_2 = 0.75\text{--}1.25$, $\lambda_3 = 0.15\text{--}0.25$, and $\lambda_4 = 0.075\text{--}0.125$. Across these retained ranges, Macro-F1 varies from 0.927 to 0.932, AUPRC from 0.896 to 0.903, ECE from 0.015 to 0.022, and RST MAE from 6.7 to 7.2 min; the nominal settings above provide the best balanced validation performance.

To prevent overconfident risk estimates for unknown noise, sensor anomalies, or out-of-distribution operating conditions, Monte Carlo Dropout is applied during inference with $K = 20$ stochastic forward passes. A dropout probability of 0.20 is retained during the stochastic inference passes. The same setting is used for all comparison tests, and the predictive mean and epistemic uncertainty are calculated from the 20 stochastic outputs as defined in Eq. (18):

$$\bar{p}_t^{(\Delta)} = \frac{1}{K} \sum_{k=1}^K \hat{p}_{t,k}^{(\Delta)}, \quad u_t^{(\Delta)} = \frac{1}{K} \sum_{k=1}^K [\hat{p}_{t,k}^{(\Delta)} - \bar{p}_t^{(\Delta)}]^2 \quad (18)$$

High-risk and low-uncertainty samples directly trigger a maintenance warning; high-risk and high-uncertainty samples are referred for manual review; and low-risk samples with high uncertainty trigger increased sampling and sensor-condition inspection.

2.4.6. Risk-Driven Predictive Maintenance

The maintenance level is determined hierarchically from the multi-horizon flashover probabilities, predicted remain-

ing safe operating time, and predictive uncertainty according to the following explicit decision rule:

$$M_t = \begin{cases} M_4, & p_t^{(5)} \geq 0.80 \text{ or } \hat{\tau}_t \leq 5 \\ M_3, & p_t^{(15)} \geq 0.60 \text{ or } \hat{\tau}_t \leq 15 \\ M_2, & p_t^{(30)} \geq 0.40 \text{ or } \hat{\tau}_t \leq 30 \\ M_1, & p_t^{(30)} \geq 0.20 \text{ or } u_t \geq u_{th} \\ M_0, & \text{otherwise.} \end{cases} \quad (19)$$

where $p_t^{(5)}$, $p_t^{(15)}$, and $p_t^{(30)}$ denote the cumulative flashover probabilities within the next 5, 15, and 30 min, respectively; $\hat{\tau}_t$ denotes the predicted remaining safe operating time (min); and u_t denotes the normalized epistemic uncertainty. The uncertainty threshold is set to $u_{th} = 0.15$. Eq. (19) is evaluated from M4 to M0 so that the most conservative applicable action has priority. Specifically, M4 is assigned when the 5-min risk reaches 0.80 or the predicted remaining safe operating time is no greater than 5 min; M3 is assigned when the 15-min risk reaches 0.60 or the remaining safe operating time is no greater than 15 min; M2 is assigned when the 30-min risk reaches 0.40 or the remaining safe operating time is no greater than 30 min; M1 is assigned when the 30-min risk reaches 0.20 or the normalized uncertainty reaches 0.15, provided that no higher-level criterion is satisfied; otherwise, M0 is assigned. Thus, 0.20/0.40, 0.60, and 0.80 are horizon-specific risk thresholds rather than four boundaries applied to a single probability variable. All thresholds are determined using the validation set and are frozen before test-set evaluation.

As shown in Table 2, the maintenance level becomes more stringent as short-horizon flashover risk increases, the predicted remaining safe operating time decreases, or predictive uncertainty becomes high. For threshold-sensitivity

analysis, the four horizon-specific probability thresholds are perturbed by ± 0.05 around their nominal values, corresponding to 0.15–0.25 for the M1 30-min threshold, 0.35–0.45 for the M2 30-min threshold, 0.55–0.65 for the M3 15-min threshold, and 0.75–0.85 for the M4 5-min threshold; u_{th} is additionally varied from 0.10 to 0.20. Across these perturbations, the validation false-alarm rate ranges from 3.2% to 4.5%, the missed-alarm rate from 1.6% to 2.5%, and the average warning lead time from 17.2 to 19.6 min. The nominal settings of 0.20, 0.40, 0.60, 0.80, and $u_{th} = 0.15$ yield a validation false-alarm rate of 3.6%, a missed-alarm rate of 1.9%, and an average lead time of 18.5 min, and are therefore retained for the test and case-study evaluations.

2.5. Experimental Setup

2.5.1. Test Platform and Dataset

To evaluate PC-HSTFNet, an accelerated-degradation platform was constructed using a 10 kV metal-clad switchgear cabinet with replaceable contact-box specimens, controlled electrical/loading stress, temperature-humidity control, and synchronized multi-sensor acquisition, as shown in Fig. 4.

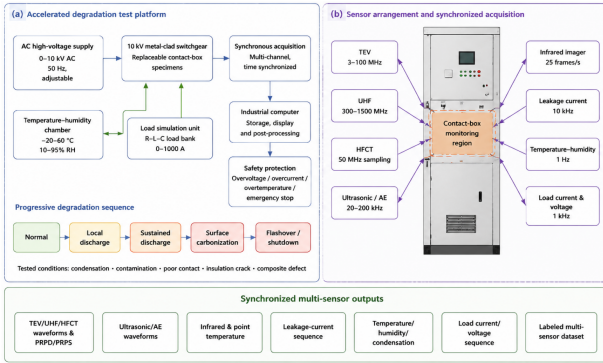


Fig. 4. Accelerated degradation test platform and sensor arrangement for the contact box in 10 kV switchgear.

The controlled tests covered condensation, contamination, poor contact, insulation cracks, and composite defects. Electrical, thermal, or environmental stress was progressively increased until the predefined non-destructive termination point or until flashover/protective shutdown in destructive tests. To avoid repetition, the sensor frequency bands, sampling rates, data formats, and intended uses are reported once in Table 3 and were kept consistent across the corresponding test runs.

Eighteen contact-box specimens were used in 72 controlled test runs, including 60 normal or single-factor runs and 12 composite-defect runs. Each specimen underwent at most one destructive run; specimens exhibiting irreversible

carbonization or flashover were immediately retired and were not reused. Sliding windows of 60 s with a step of 10 s produced 28,640 multi-sensor samples. The five stages S_0 – S_4 contained 6,020, 7,850, 6,480, 5,030, and 3,260 samples, respectively. Data were split by complete specimen and test batch into training, validation, and test sets at a ratio of 6:2:2, ensuring that adjacent windows from the same degradation process did not appear in different subsets. Prediction horizons were set to 5, 15, and 30 min, and samples without flashover before test termination were treated as right-censored. For the cross-domain generalization test, an additional independent dataset was collected from four contact-box specimens installed in a third 10 kV switchgear model supplied by a second anonymized manufacturer. Twelve controlled runs produced 4,320 sliding-window samples. These specimens and all associated windows were completely excluded from training, validation, hyperparameter tuning, and threshold selection and were used only for the held-out cross-domain evaluation in Section 3.3. No separate long-term field-operating dataset was included in the present study.

2.5.2. Comparison Methods and Experimental Settings

The proposed method was compared with representative models, including XGBoost based on handcrafted features; 1D-CNN and TCN for temporal-signal modeling; Swin Transformer for image inputs; MAADNet for partial-discharge recognition; Early Fusion as a feature-concatenation baseline; and Multimodal Transformer as a multimodal-fusion baseline. PC-HSTFNet is the complete proposed model.

All deep-learning models used identical specimen/batch-level data partitions and the same maximum training budget. PC-HSTFNet was optimized using AdamW with an initial learning rate of 1×10^{-4} , a batch size of 64, and a maximum of 150 epochs; early stopping was activated after 20 consecutive epochs without validation-loss improvement. The 60-s input window, 10-s sliding step, and 5-, 15-, and 30-min prediction horizons defined in Section 2.5.1 were used consistently. The unified embedding dimension was $d = 128$, the cross-modal attention module used four attention heads, the causal temporal encoder contained two layers with a feed-forward dimension of 256, and dropout was set to 0.20. The final physics-consistency margin was $\gamma = 0.10$, the multi-task loss weights were $\lambda_1 = 1.00$, $\lambda_2 = 1.00$, $\lambda_3 = 0.20$, and $\lambda_4 = 0.10$, and Monte Carlo Dropout used $K = 20$ stochastic forward passes. All hyperparameters were selected using the validation set only and were then frozen for test evaluation. Reported performance values are averages over five independent runs. Degradation-stage identification was evaluated us-

Table 3. Multi-sensor Configuration and Data Formats

Sensing Modality	Effective Frequency Band / Sampling Rate	Data Format	Primary Application
Transient Earth Voltage (TEV)	3–100 MHz	Time-domain waveform, PRPD	Characterize partial discharge pulses and phase-resolved distribution
Ultra-high Frequency (UHF) Sensor	300–1500 MHz	UHF waveform, frequency spectrum, PRPS	Supplement weak and high-frequency discharge information
High Frequency Current Transformer (HFCT)	50 MHz sampling rate	High-frequency current waveform, PRPD	Characterize conducted discharge pulses and phase-resolved activity
Ultrasonic / Acoustic Emission (AE)	20–200 kHz	Acoustic waveform, frequency band energy	Characterize discharge-induced acoustic emission activity
Infrared Thermography	25 frame/s	Thermal image, hot spot sequence	Characterize local overheating and temperature rise propagation
Leakage Current Sensor	10 kHz	Current time series	Characterize surface conduction and tracking development process
Temperature and Humidity Sensor	1 Hz	Low-frequency time series	Characterize insulation dampness, condensation and environmental stress
Load Current and Voltage Sensor	1 kHz	Operating condition time series	Characterize electro-thermal stress and equipment operational status

ing Macro-F1 and critical-stage recall; risk prediction used AUROC, AUPRC, Brier score, and ECE; and remaining-safe-operating-time prediction used MAE and C-index.

3. Results and discussion

3.1. Overall Performance Comparison

The overall test-set performance of the compared methods is presented in Table 4.

As shown in Table 4, PC-HSTFNet achieved the best performance for all metrics. Compared with the Multimodal Transformer, Macro-F1 and critical-stage recall increased by 2.6 and 3.1 percentage points, respectively, and the 5-min flashover-risk AUPRC increased by 3.3 percentage points. The Brier score and ECE decreased by 21.2% and 36.0%, respectively, indicating that the output risk probabilities provide both strong discrimination and better agreement with observed flashover frequencies. Single-modality models were prone to confusion during early degradation or under composite defects. Although MAADNet effectively suppresses partial-discharge noise, it does not fully exploit thermal, environmental, and operating-condition information and therefore has limited ability to distinguish samples

with similar discharge amplitudes but different risk levels. Early Fusion introduces multimodal information but cannot suppress redundant and low-quality channels through direct concatenation. By dynamically weighting sensors and modeling cross-modal interactions, PC-HSTFNet obtains a more complete degradation representation. Confusion matrices and the ROC and PR curves for 5-min risk prediction are shown in Fig. 5.

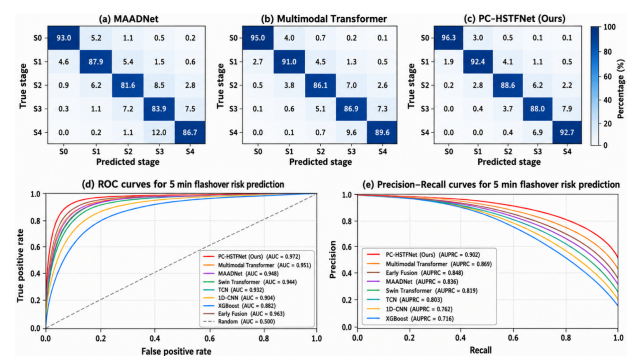


Fig. 5. Degradation-stage identification and short-term flashover-risk prediction results of selected models.

Table 4. Test-Set Performance over Five Independent Runs (Mean $\pm$ SD for Main Metrics)

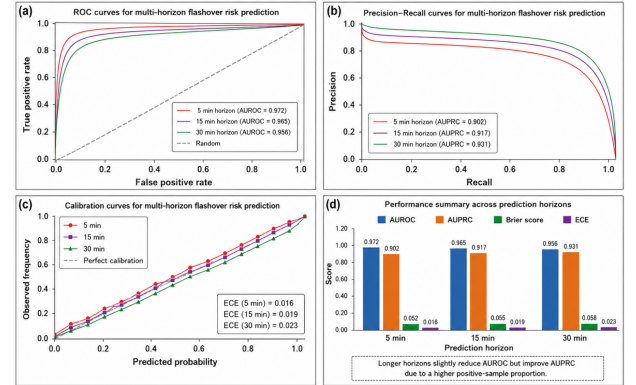
Method	Macro-F1	S4 Recall Rate	AUPRC@5min	Brier Score	ECE	RST MAE (min)	C-index
XGBoost	0.781$\pm$0.010	0.742	0.716$\pm$0.012	0.132$\pm$0.006	0.061	18.7$\pm$0.8	0.731
1D-CNN	0.824$\pm$0.008	0.791	0.762$\pm$0.010	0.108$\pm$0.005	0.047	15.2$\pm$0.7	0.774
TCN	0.851$\pm$0.007	0.828	0.803$\pm$0.009	0.094$\pm$0.004	0.041	12.8$\pm$0.6	0.801
Swin Transformer	0.864$\pm$0.007	0.841	0.819$\pm$0.008	0.088$\pm$0.004	0.036	11.7$\pm$0.5	0.816
MAADNet	0.879$\pm$0.006	0.867	0.836$\pm$0.007	0.079$\pm$0.004	0.032	10.6$\pm$0.5	0.833
Early Fusion	0.887$\pm$0.006	0.874	0.848$\pm$0.007	0.075$\pm$0.003	0.030	9.9$\pm$0.4	0.845
Multimodal Transformer	0.905$\pm$0.006	0.896	0.869$\pm$0.007	0.066$\pm$0.004	0.025	8.7$\pm$0.4	0.866
PC-HSTFNet (Proposed Model)	0.931$\pm$0.004	0.927	0.902$\pm$0.005	0.052$\pm$0.003	0.016	6.8$\pm$0.3	0.893

Table 4 reports the mean $\pm$ standard deviation of the main metrics over five independent training runs, whereas Fig. 5 presents one representative test run; therefore, minor differences exist between individual class recalls in the figure and the averages in Table 4. Using matched random seeds, PC-HSTFNet was compared with the Multimodal Transformer by a two-sided paired t-test. The differences were statistically significant for Macro-F1 ($p = 0.003$), AUPRC@5min ($p = 0.004$), Brier score ($p = 0.006$), and RST MAE ($p = 0.002$). Across the five runs, PC-HSTFNet achieved 0.931 ± 0.004 Macro-F1, 0.902 ± 0.005 AUPRC@5min, a Brier score of 0.052 ± 0.003 , and an RST MAE of 6.8 ± 0.3 min, compared with 0.905 ± 0.006 , 0.869 ± 0.007 , 0.066 ± 0.004 , and 8.7 ± 0.4 min, respectively, for the Multimodal Transformer. These results indicate that the observed performance improvements are larger than the run-to-run training variation.

As shown in Fig. 5, PC-HSTFNet produces clear class boundaries for S_0 , S_1 , and S_4 . Most errors occur between adjacent stages S_2 and S_3 because the transition from sustained partial discharge to severe carbonization has no strictly defined instantaneous boundary. The proportion of S_3 or S_4 samples misclassified as low-risk states is relatively small, satisfying the priority of reducing missed alarms for high-risk equipment. Calibration curves and risk metrics at different prediction horizons are shown in Fig. 6.

At prediction horizons of 5, 15, and 30 min, PC-HSTFNet achieved AUROC values of 0.972, 0.965, and 0.956 and AUPRC values of 0.902, 0.917, and 0.931, respectively. As the horizon increased, the proportion of positive samples that eventually experienced flashover increased, improving AUPRC; however, uncertainty over a longer interval slightly reduced AUROC. Calibration curves at all hori-

zons were close to the ideal diagonal, demonstrating that discrete-time survival modeling and the calibration loss generate reliable cumulative risk probabilities.

**Fig. 6.** Discrimination and probability calibration of multi-horizon flashover risks.

3.2. Ablation Experiments and Fusion-Weight Analysis

To quantify the contribution of each module, adaptive denoising, dynamic reliability weighting, cross-modal attention, causal temporal encoding, the physics-consistency constraint, and the calibration loss were removed individually. The results are presented in Table 5.

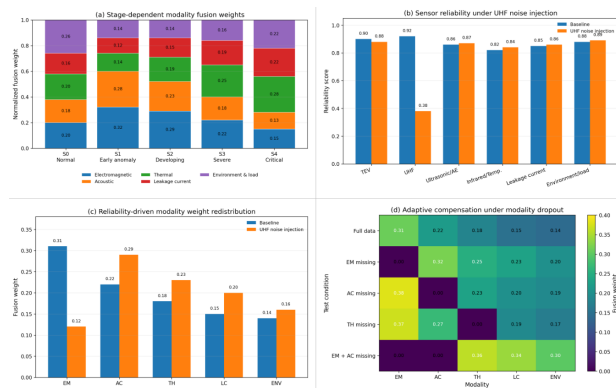
As shown in Table 5, removing causal temporal encoding caused the largest performance decrease, indicating that flashover risk depends primarily on the continuous evolution of multiple physical variables rather than instantaneous features from a single time window. Removing cross-modal attention and reliability weighting reduced AUPRC by 4.4 and 3.5 percentage points, respectively, con-

Table 5. Ablation Experiment Results of PC-HSTFNet

Model Configuration	Macro-F1	AUPRC@5min	Brier Score	ECE	RST MAE (min)
w/o Adaptive Denoising Module	0.913	0.881	0.063	0.024	8.1
w/o Reliability-based Feature Weighting	0.906	0.867	0.068	0.029	8.6
w/o Cross-modal Attention Mechanism	0.899	0.858	0.071	0.031	9.0
w/o Causal Temporal Encoding	0.885	0.842	0.078	0.037	10.2
w/o Physical Consistency Constraint	0.918	0.889	0.060	0.022	7.5
w/o Probability Calibration Loss	0.930	0.899	0.057	0.045	6.9
Full PC-HSTFNet (Proposed Model)	0.931	0.902	0.052	0.016	6.8

firming that dynamically selecting reliable sensing information and learning complementary modal relationships are essential to risk prediction.

Removing the physics-consistency constraint caused only a moderate decrease in overall classification accuracy but substantially degraded risk ranking and remaining-time prediction. Without this constraint, 12.6% of consecutive degradation windows exhibited increasing degradation indicators but decreasing predicted risk; the complete model reduced this proportion to 3.1%. Removing the calibration loss had little effect on Macro-F1, but increased ECE from 0.016 to 0.045, indicating that classification accuracy cannot replace evaluation of risk-probability reliability. Fig. 7 shows the dynamic fusion weights at different degradation stages and under sensor anomalies.

**Fig. 7.** Sensor reliability and modality-fusion weights under different operating conditions.

Electromagnetic and acoustic modalities receive higher weights during the early-anomaly stage. As discharge and thermal accumulation continue, the contributions of thermal, leakage-current, and environmental/operating-condition information increase. Under UHF noise injection, the electromagnetic fusion weight decreases from 0.31 to 0.12, while the other modalities are correspondingly up-weighted. Fig. 7(d) further quantifies the compensation

mechanism under missing modalities. With complete data, the [EM, AC, TH, LC, ENV] fusion weights are [0.31, 0.22, 0.18, 0.15, 0.14]. When EM is missing, the weights become [0.00, 0.32, 0.25, 0.23, 0.20]; when AC is missing, they become [0.38, 0.00, 0.23, 0.20, 0.19]; and when TH is missing, they become [0.37, 0.27, 0.00, 0.19, 0.17]. When both EM and AC are missing, the remaining TH, LC, and ENV weights increase to 0.36, 0.34, and 0.30, respectively. Thus, the validity mask forces unavailable modalities to zero contribution and the reliability-aware fusion module renormalizes attention over the remaining channels, explaining why the performance degradation in Table 6 is gradual rather than catastrophic.

3.3. Robustness and Cross-Condition Generalization

To examine robustness to deployment-relevant disturbances and distribution shift, tests were conducted under low signal-to-noise ratios, single- and dual-modality dropout, sensor drift, and a held-out cross-domain switchgear condition. The primary 18-specimen dataset was obtained from two 10 kV metal-clad switchgear models (Models A1 and A2) supplied by Manufacturer A (anonymized). The held-out cross-domain set used a third switchgear model (Model B1) supplied by Manufacturer B (anonymized), with different cabinet geometry and sensor-mounting layout. It contained four additional contact-box specimens, 12 controlled runs, and 4,320 sliding-window samples. None of these specimens, runs, or windows was used for model fitting, validation, hyperparameter tuning, or maintenance-threshold selection. No separate long-term field-operating dataset was included; therefore, the analysis is limited to controlled cross-device and cross-condition generalization. The results for Early Fusion, Multimodal Transformer, and PC-HSTFNet are listed in Table 6.

As shown in Table 6, all models degrade under noise and missing modalities, but PC-HSTFNet exhibits the smallest decline. With one missing modality, its AUPRC remains 0.874; with two simultaneously missing modalities, AUPRC remains 0.826, exceeding Early Fusion and Mul-

Table 6. Model robustness under disturbances and cross-domain conditions

Test Scenario	Early Fusion AUPRC	Early Fusion MAE (min)	Multimodal Transformer AUPRC	Multimodal Transformer MAE (min)	PC-HSTFNet (Proposed) AUPRC	PC-HSTFNet (Proposed) MAE (min)
Complete sensor data (baseline)	0.848	9.9	0.869	8.7	0.902	6.8
0 dB Gaussian noise	0.791	12.3	0.824	10.7	0.879	7.9
-5 dB Gaussian noise	0.733	15.6	0.782	13.1	0.851	9.2
Single modality missing	0.762	13.8	0.811	11.5	0.874	8.3
Two modalities missing	0.691	17.4	0.748	15.0	0.826	10.5
15% sensor drift	0.721	16.1	0.775	13.7	0.858	9.0
Unseen switchgear model (cross-domain)	0.714	15.8	0.782	13.2	0.846	9.4

timodal Transformer by 13.5 and 7.8 percentage points, respectively. This trend is consistent with the weight-redistribution mechanism in Fig. 7(d): increasing missing-modality severity removes more high-information channels, but the validity mask prevents invalid features from entering the fused representation and redistributes the remaining fusion mass to available thermal, leakage-current, acoustic, and environmental channels. In the held-out cross-domain switchgear test, PC-HSTFNet achieves an AUPRC of 0.846 and an RST MAE of 9.4 min. These results support robustness under the tested cross-domain condition, while the limited specimen scale and the absence of a separately quantified long-term field dataset mean that broader manufacturer-level generalization should be interpreted cautiously.

3.4. Predictive-Maintenance Case Study

A contamination-condensation composite condition was selected as a representative case. At the beginning of the test, the in-cabinet relative humidity gradually increased from 68% to 92%, followed by increases in the TEV pulse repetition rate and leakage current. During the late stage, the contact-box hot-spot temperature increased rapidly and intermittent dry-band arcing occurred. The multi-sensor conditions, predicted risks, and remaining safe operating time are shown in Fig. 8.

Actual flashover occurred 118.4 min after the start of the test. Using the fixed decision thresholds defined in Section 2.4.6, the 30-min risk reached the M2 boundary of 0.40 at 94.8 min (23.6 min before flashover), the 15-min risk reached the M3 boundary of 0.60 at 103.6 min (14.8 min before flashover), and the 5-min risk reached the M4 boundary of 0.80 at 112.2 min (6.2 min before flashover). When

the actual remaining time was 25 min, the predicted value was 26.3 min; when the actual remaining time decreased to 10 min, the prediction was 9.4 min. The maintenance-decision performance of different warning methods is presented in Table 7. The 2.1-min MAE in Fig. 8(c) is calculated over the full trajectory of this representative case, whereas the 6.8-min MAE in Table 4 is averaged over all test samples; the two values therefore have different statistical scopes.

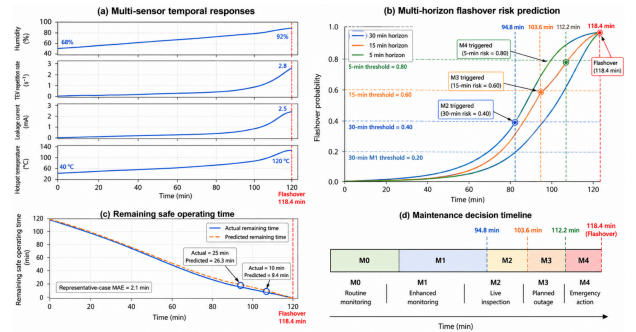


Fig. 8. Multi-sensor responses and maintenance decisions during a representative contact-box degradation process.

As shown in Table 7, the fixed-threshold method tends to misclassify transient interference as high risk and generally issues an alarm only after discharge activity has increased substantially. By jointly tracking partial discharge, temperature rise, environmental stress, and operating stress, PC-HSTFNet increases the average early-warning lead time to 18.9 min while reducing the missed-alarm rate to 1.6%. Thus, the proposed method controls false alarms while providing more decision time for live inspection, load transfer, or planned outage.

Table 7. Predictive Maintenance Decision Performance of Different Methods

Method	False Alarm Rate / %	Missed Alarm Rate / %	Average Early Warning Lead Time / min
TEV Fixed Threshold (traditional baseline)	12.7	9.5	6.8
Single-modality MAADNet	8.9	6.3	10.4
Multimodal Transformer	6.1	4.8	14.7
PC-HSTFNet (Proposed)	3.4	1.6	18.9

4. Conclusions

This study addresses heterogeneous multi-source information, changing sensor reliability, and inadequate risk prediction during the evolution from partial discharge to surface flashover in contact boxes of 10 kV switchgear. A physics-consistent hierarchical sensor-temporal fusion network, PC-HSTFNet, is proposed together with a closed-loop method from risk prediction to maintenance decision-making. The main conclusions are as follows:

- (1) Dynamic weighting of homogeneous sensors and reliability-biased cross-modal attention effectively fuse electromagnetic, acoustic, thermal, leakage-current, and environmental/operating-condition information. On the test set, Macro-F1, S4 recall, and 5-min risk AUPRC reached 0.931, 0.927, and 0.902, respectively, outperforming all comparison models.
- (2) The physics-consistency constraint and discrete-time survival modeling improve risk ranking, probability calibration, and remaining-safe-operating-time estimation. The Brier score and ECE were 0.052 and 0.016, respectively, and the MAE of the remaining safe operating time was 6.8 min. The model maintained good robustness under noise, sensor drift, and missing modalities.
- (3) The predictive-maintenance case study yielded false-alarm and missed-alarm rates of 3.4% and 1.6%, respectively, and an average early-warning lead time of 18.9 min, providing additional decision time for

live inspection, planned outage, and emergency load transfer.

- (4) The present study is limited by specimen scale, accelerated-degradation conditions, and the absence of a separately quantified long-term field-operating dataset. Future work will extend the dataset across sites, manufacturers, and long-term operating periods and investigate lightweight edge deployment and online incremental updating to further validate engineering generalization.

Acknowledgements

This work is supported by the Science and Technology Project of State Grid Fujian Electric Power Co., Ltd. (Grant No: 52133024000C), titled "Research on Key Technologies for Fault Diagnosis and Mitigation of Conductive Circuit Abnormalities in High-Current Circuit Breakers inside 10 kV Switchgear."

References

- [1] Z. Fei, Y. Li, and S. Yang, (2024) "Pattern recognition of partial discharge faults in switchgear using a back propagation neural network optimized by an improved mantis search algorithm" *Sensors* 24(10): 3174. DOI: [10.3390/s24103174](https://doi.org/10.3390/s24103174).
- [2] S. Chen, H. A. Illias, J. K. R. Wong, and N. N. Mansor, (2024) "Denoising of partial discharges in switchgear insulation material using hybrid wavelet denoising-optimization-machine learning" *Ain Shams Engineering Journal* 15(11): 103032. DOI: [10.1016/j.asej.2024.103032](https://doi.org/10.1016/j.asej.2024.103032).
- [3] J. Yang, K. Hu, J. Zhang, and J. Bao, (2024) "Semi-supervised learning for gas insulated switchgear partial discharge pattern recognition in the case of limited labeled data" *Engineering Applications of Artificial Intelligence* 137: 109193. DOI: [10.1016/j.engappai.2024.109193](https://doi.org/10.1016/j.engappai.2024.109193).
- [4] Z. Xu, H. Xu, C. Yuan, S. Chen, and Y. Chen, (2024) "Recognition of partial discharge in GIS based on image feature fusion" *AIMS Energy* 12(6): 1096–1112. DOI: [10.3934/energy.2024052](https://doi.org/10.3934/energy.2024052).
- [5] Z. Wang, Z. Liu, L. Qiao, D. Qian, Z. Chen, C. Gao, and W. Wang, (2024) "Research on partial discharge pattern recognition in GIS based on EFPI sensor" *Journal of Electrical Engineering & Technology* 19(1): 577–584. DOI: [10.1007/s42835-023-01526-8](https://doi.org/10.1007/s42835-023-01526-8).

- [6] J. Yan, Y. Wang, Y. Zhou, J. Wang, and Y. Geng, (2024) "A novel meta-learning network for partial discharge source localization in gas-insulated switchgear via digital twin" **IET Generation, Transmission & Distribution** 18(9): 1785–1794. DOI: [10.1049/gtd2.13156](https://doi.org/10.1049/gtd2.13156).
- [7] R. Loggia and A. Flamini, (2024) "Electrical safety enhanced with BIM, SCADA and digital twin integration: A case study of a MV-LV substation" **IEEE Transactions on Industry Applications** 60(2): 2725–2731. DOI: [10.1109/TIA.2023.3332311](https://doi.org/10.1109/TIA.2023.3332311).
- [8] Z. Tang and X. Jian, (2024) "Thermal fault diagnosis of complex electrical equipment based on infrared image recognition" **Scientific Reports** 14: 5547. DOI: [10.1038/s41598-024-56142-x](https://doi.org/10.1038/s41598-024-56142-x).
- [9] Y. Zhang, J. Ding, Y. Li, Z. Ren, and K. Feng, (2024) "Multi-modal data cross-domain fusion network for gearbox fault diagnosis under variable operating conditions" **Engineering Applications of Artificial Intelligence** 133: 108236. DOI: [10.1016/j.engappai.2024.108236](https://doi.org/10.1016/j.engappai.2024.108236).
- [10] Y. Xu and X. Lu, (2024) "Power transformer fault diagnosis based on hybrid intelligent algorithm" **Journal of Applied Science and Engineering** 27(1): 1859–1866. DOI: [10.6180/jase.202401_27\(1\).0002](https://doi.org/10.6180/jase.202401_27(1).0002).
- [11] Z. Jia, S. Fan, Z. Wang, S. Shao, and D. He, (2025) "Partial discharge defect recognition method of switchgear based on cloud-edge collaborative deep learning" **Scientific Reports** 15: 10956. DOI: [10.1038/s41598-024-81478-9](https://doi.org/10.1038/s41598-024-81478-9).
- [12] T. Lin, Z. Ren, L. Zhu, Y. Zhu, K. Feng, W. Ding, K. Yan, and M. Beer, (2025) "A systematic review of multi-sensor information fusion for equipment fault diagnosis" **IEEE Transactions on Instrumentation and Measurement** 74: 1–48. DOI: [10.1109/TIM.2025.3529577](https://doi.org/10.1109/TIM.2025.3529577).
- [13] J. Li, J. Tian, A. Banerjee, X. Zhai, and S. Wang, (2025) "Gradient balanced selective mixture-of-experts for gas insulated switchgear partial discharge diagnosis" **Measurement** 256: 118139. DOI: [10.1016/j.measurement.2025.118139](https://doi.org/10.1016/j.measurement.2025.118139).
- [14] Z. Li, Y. Zang, C. Wang, Y. Tang, T. Ren, and X. Jiang, (2025) "Analysis and diagnosis of optical and UHF partial discharges in GIS based on guided filtering fusion" **IEEE Transactions on Dielectrics and Electrical Insulation** 32(5): 2978–2985. DOI: [10.1109/TDEI.2025.3543147](https://doi.org/10.1109/TDEI.2025.3543147).
- [15] T. Cao, Y. Cui, H. Tan, W. Lu, F. Zhang, K. Liu, X. Chen, and L. Wang, (2025) "Research on a degradation identification method for GIS UHF partial discharge sensors based on S-parameters" **Sensors** 25(22): 6860. DOI: [10.3390/s25226860](https://doi.org/10.3390/s25226860).
- [16] X. Lv, F. Liu, M. Jiang, F. Zhang, and L. Jia, (2025) "Fault diagnosis of power transformers based on dissolved gas analysis and multi-kernel graph convolution network integrated with dual-channel classifiers" **Journal of Applied Science and Engineering** 28(10): 2217–2226. DOI: [10.6180/jase.202510_28\(10\).0014](https://doi.org/10.6180/jase.202510_28(10).0014).
- [17] Y. Q. Liu, X. C. Hua, W. D. Li, Z. K. Lai, G. J. Zhang, X. X. Chen, and X. J. Shao, (2026) "Enhancement of GIS partial discharge pattern recognition accuracy based on domain adversarial ConoNeXt" **IEEE Transactions on Dielectrics and Electrical Insulation** 33(1): 742–751. DOI: [10.1109/TDEI.2025.3579444](https://doi.org/10.1109/TDEI.2025.3579444).
- [18] S. Pan, W. Zhang, Y. Ma, Y. Su, and W. Huang, (2026) "Pattern recognition of GIS partial discharge based on UHF signal characteristics" **Electronics** 15(5): 1096. DOI: [10.3390/electronics15051096](https://doi.org/10.3390/electronics15051096).
- [19] Y. He, Y. Fang, Z. Zhang, D. Zhou, S. Chen, and S. Jing, (2026) "Partial discharge pattern recognition of GIS with time-frequency energy grayscale maps and an improved variational Bayesian autoencoder" **Energies** 19(1): 127. DOI: [10.3390/en19010127](https://doi.org/10.3390/en19010127).
- [20] B. Jia, Q. Sun, W. Guo, M. Wang, Q. Wang, and H. Zhao, (2026) "GIS partial discharge fault diagnosis based on multi-source feature fusion and ResNet-MLP" **Energies** 19(4): 1073. DOI: [10.3390/en19041073](https://doi.org/10.3390/en19041073).
- [21] J. Wen, J. Li, and Y. Zhang, (2026) "Deep multi-modal imaging temperature measurement method for detecting temperature rise in electrical equipment faults" **Journal of Applied Science and Engineering** 29(3): 603–610. DOI: [10.6180/jase.202603_29\(3\).0011](https://doi.org/10.6180/jase.202603_29(3).0011).
- [22] L. Li, J. Tang, and Y. Liu, (2015) "Partial discharge recognition in gas insulated switchgear based on multi-information fusion" **IEEE Transactions on Dielectrics and Electrical Insulation** 22(2): 1080–1087. DOI: [10.1109/TDEI.2015.7076809](https://doi.org/10.1109/TDEI.2015.7076809).
- [23] R. Hussein, K. B. Shaban, and A. H. El-Hag, (2017) "Robust feature extraction and classification of acoustic partial discharge signals corrupted with noise" **IEEE Transactions on Instrumentation and Measurement** 66(3): 405–413. DOI: [10.1109/TIM.2016.2639678](https://doi.org/10.1109/TIM.2016.2639678).

- [24] C. Zhang, M. Dong, M. Ren, W. Huang, J. Zhou, X. Gao, and R. Albarracín, (2018) “Partial discharge monitoring on metal-enclosed switchgear with distributed non-contact sensors” **Sensors** 18(2): 551. DOI: [10.3390/s18020551](https://doi.org/10.3390/s18020551).
- [25] Y. Wang, J. Yan, Z. Yang, Y. Zhao, and T. Liu, (2020) “GIS partial discharge pattern recognition via lightweight convolutional neural network in the ubiquitous power Internet of Things context” **IET Science, Measurement & Technology** 14(8): 864–871. DOI: [10.1049/iet-smt.2019.0542](https://doi.org/10.1049/iet-smt.2019.0542).
- [26] Y. Wang, J. Yan, Z. Yang, Y. Zhao, and T. Liu, (2021) “Optimizing GIS partial discharge pattern recognition in the ubiquitous power Internet of Things context: A MixNet deep learning model” **International Journal of Electrical Power & Energy Systems** 125: 106484. DOI: [10.1016/j.ijepes.2020.106484](https://doi.org/10.1016/j.ijepes.2020.106484).
- [27] R. Fechet, A. I. Petrariu, and A. Graur, (2021) “Partial discharge and Internet of Things: A switchgear cell maintenance application using microclimate sensors” **Sensors** 21(24): 8372. DOI: [10.3390/s21248372](https://doi.org/10.3390/s21248372).
- [28] K. Yin, Y. Wang, S. Liu, P. Li, Y. Xue, B. Li, and K. Dai, (2022) “GIS partial discharge pattern recognition based on multi-feature information fusion of PRPD image” **Symmetry** 14(11): 2464. DOI: [10.3390/sym14112464](https://doi.org/10.3390/sym14112464).
- [29] Y. Wu, D. Ding, Y. Wang, C. Zhou, H. Lu, and X. Zhang, (2022) “Defect recognition and condition assessment of epoxy insulators in gas insulated switchgear based on multi-information fusion” **Measurement** 190: 110701. DOI: [10.1016/j.measurement.2022.110701](https://doi.org/10.1016/j.measurement.2022.110701).
- [30] F. Cordoni, G. Bacchiega, G. Bondani, R. Radu, and R. Muradore, (2022) “A multi-modal unsupervised fault detection system based on power signals and thermal imaging via deep AutoEncoder neural network” **Engineering Applications of Artificial Intelligence** 110: 104729. DOI: [10.1016/j.engappai.2022.104729](https://doi.org/10.1016/j.engappai.2022.104729).