Artificial Intelligence-Based Degradation Assessment And Remaining Useful Life Prediction For 10 KV Tulip Contacts Via Thermal–Electrical–Mechanical Feature Fusion

Bo Yuan[1], Xielin Shen[1], Hua Ye[1*], Lei Wang[2,3], and Yiming Liu[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: yehua0608@126.com*

Received Aug. 7, 2026; accepted Sep. 7, 2026

To address the coupled degradation of tulip contacts in 10 kV withdrawable switchgear under the combined effects of insertion–withdrawal wear, spring relaxation, increasing contact resistance, and localized overheating, this paper proposes a physics-informed thermal–electrical–mechanical feature fusion network (PI-TEMFNet). The method extracts multi-rate features, including temperature rise, contact resistance, contact force, and vibration, and employs reliability-aware masked cross-modal attention to fuse asynchronous and missing information. Contact-resistance relationships, the heat-balance equation, and a monotonic degradation constraint are embedded into the joint prediction of degradation stage, health index, and probabilistic remaining useful life (RUL). Accelerated degradation tests on 12 GC5 specimens and leave-one-specimen-out cross-validation show that the model achieves an Accuracy of 96.4%, a Macro-F1 of 96.1%, and an AUC of 0.988, with a health-index RMSE of 0.041. The MAE, RMSE, and MAPE for RUL prediction are 21.6×10^3 cycles, 29.4×10^3 cycles, and 5.8%, respectively, while the coverage probability of the 90% prediction interval is 91.3%. Under random missing rates of 20% and 40%, the MAPE remains 6.3% and 7.2%, respectively, demonstrating high accuracy and strong robustness.

Keywords: 10 kV switchgear; tulip contact; thermal–electrical–mechanical feature fusion; physics-informed neural network; degradation assessment; health index; remaining useful life; probabilistic prediction

© The Author('s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC 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_36.030

1. Introduction

10 kV metal-enclosed switchgear is essential equipment for power distribution, control, and fault isolation in distribution systems, and its operational reliability directly affects the safety and continuity of the power supply. As a separable connection component in the primary conductive circuit of a withdrawable circuit breaker, the tulip contact relies on contact fingers and retaining springs to provide stable contact force, thereby maintaining low contact resistance and satisfactory current-carrying capability. However, frequent insertion and withdrawal, mechanical vibration, assembly deviation, surface wear, and spring stress relaxation during long-term operation can reduce the contact force and cause uneven force distribution, ultimately degrading the contact performance.

According to rough-surface contact theory, current is mainly transmitted through a limited number of microscopic conductive asperities, and the real contact area is jointly affected by the normal load, material properties, and surface roughness [1–4]. When the contact force

of a tulip contact decreases or the surface roughness increases, the contact resistance rises accordingly and generates more intense localized Joule heating under high current. Sustained temperature rise further accelerates annealing and thermal deformation of the contact-finger material as well as spring stress relaxation, thereby forming a thermal–electrical–mechanical coupled degradation process of "mechanical deterioration–resistance increase–localized overheating–further mechanical degradation."

Condition monitoring of switchgear has gradually evolved from single-threshold judgment toward intelligent sensing and data-driven diagnosis. Machine-learning and deep-learning methods have demonstrated considerable potential in fault identification, thermal anomaly detection, and predictive maintenance [5–10]. However, the hot spot of a tulip contact is located inside an enclosed contact box. A single temperature or loop-resistance measurement is not only unable to locate local defects accurately but is also easily affected by load current, ambient temperature, and heat-dissipation conditions. Therefore, fusing multi-source information such as temperature rise, contact resistance, contact force, vibration, and surface condition is an important means of distinguishing operating-condition variations from genuine degradation and improving early deterioration detection.

This study integrates multidimensional thermal, electrical, and mechanical features to assess tulip-contact degradation and predict service life. Joint multiphysical representation is used to mitigate operating-condition interference and quantify contact health; RUL prediction supports the transition of distribution switchgear toward predictive maintenance; and electrical-contact physics is introduced to constrain the intelligent model, thereby improving model interpretability and adaptability to varying operating conditions.

Artificial-intelligence research on power equipment has gradually shifted from single-parameter rule-based judgment to multi-parameter feature extraction and intelligent classification. In representative studies, Xu and Lu optimized the parameters of a probabilistic neural network using an improved grey wolf optimizer and diagnosed transformer faults from dissolved-gas data [11]. Lv et al. combined a multi-kernel graph convolutional network with a dual-channel classifier to exploit topological associations and complementary information among multiple dissolved gases, thereby improving the identification of complex fault modes [12]. Wen et al. proposed a deep multimodal imaging method for detecting abnormal temperature rise in electrical equipment; different imaging information was fused through variational inference and modality-weight allocation, demonstrating that dynamically assigning modality contributions is more effective than fixed feature concatenation for extracting complementary information [13]. Although these studies did not directly address tulip contacts, their concepts of intelligent optimization, graph-structured modeling, and dynamic multimodal weighting provide methodological references for heterogeneous thermal–electrical–mechanical feature fusion.

In addition to supervised fault classification, multimodal unsupervised learning has received increasing attention. Cordoni et al. jointly fed power signals and thermal images into a deep autoencoder to learn the normal operating distribution of equipment without requiring large numbers of fault labels and then identified abnormal states using reconstruction error [14]. Tang and Jian diagnosed thermal faults in complex electrical equipment through infrared-image recognition and verified the effectiveness of deep visual features for identifying localized overheating regions [15]. These methods can automatically extract latent features from high-dimensional monitoring data and reduce dependence on manually defined thresholds and empirical rules. Nevertheless, their outputs are generally fault classes or anomaly scores and cannot readily answer predictive-maintenance questions such as "to what extent has the equipment degraded?" and "how long remains before failure?"

Remaining useful life prediction is a key component of equipment prognostics and health management. Lei et al. systematically reviewed the machinery health-prognostics process from data acquisition and health-indicator construction to degradation modeling and RUL prediction, emphasizing that a health indicator with monotonicity, trendability, and predictability is the key link between raw sensor data and lifetime models [16]. Zhao et al. summarized deep-learning applications in machine health monitoring and noted that convolutional networks, recurrent networks, and autoencoders can automatically learn hierarchical degradation features from complex signals [17]. Li et al. employed a deep convolutional neural network for RUL estimation, using local receptive fields and weight sharing to extract degradation patterns from multisensor data and providing a representative end-to-end lifetime-prediction framework [18].

Conventional deterministic RUL models generally output only a point estimate and therefore cannot characterize uncertainty caused by specimen differences, operating-condition disturbances, and model errors. Pan et al. combined Bayesian deep learning with a stochastic process for probabilistic RUL prediction in the absence of com-

plete lifetime labels and expressed lifetime-estimation uncertainty through a predictive distribution [19]. Wu et al. reviewed recent deep-learning-based RUL methods and identified multi-source information fusion, cross-condition generalization, interpretable prediction, and uncertainty quantification as important research directions [20]. For safety-critical 10 kV switchgear, a single-point RUL estimate may lead to overly aggressive or conservative maintenance decisions, whereas jointly reporting the predictive mean, confidence interval, and failure probability is more consistent with risk-oriented asset management.

Multisensor fusion and joint learning of health state and lifetime constitute another recent trend in RUL research. Wang et al. proposed a dynamically parameterized bilinear factorization framework to fuse multimodal degradation information under varying operating conditions and alleviate the influence of distribution shifts on lifetime prediction [21]. Zhou and Wang developed an adaptive multiscale feature-fusion and mixture-of-experts multitask model that simultaneously assesses equipment health state and predicts RUL, allowing discrete state labels and continuous lifetime information to constrain each other [22]. They subsequently proposed a multi-perspective spatiotemporal graph attention network that represents spatial associations among sensors through a graph structure and learns long-range dependencies in degradation processes through temporal attention [23]. These studies demonstrate that multitask learning and dynamic fusion can improve the completeness of degradation representations for complex equipment. However, most existing studies use public datasets for aero-engines, bearings, and similar equipment, where the inputs are generally synchronous, continuous, and dimensionally preprocessed multichannel sensor sequences. Such data differ markedly from the practical structure of tulip-contact data, in which thermal, electrical, and mechanical features have different sampling rates and inconsistent availability.

Physics-informed neural networks provide a new technical pathway for integrating mechanistic models with data-driven approaches. Raissi et al. transformed governing equations, initial conditions, and boundary conditions into differentiable loss terms, enabling a neural network to fit observations while satisfying physical laws [24]. Compared with black-box models that depend entirely on statistical relationships in samples, physics-informed methods can exploit prior knowledge to narrow the hypothesis space and may offer advantages under data scarcity, incomplete observations, or extrapolative prediction. For tulip contacts, the contact force–real contact area–contact resistance relationship, the Joule-heating relationship, and the heat-conduction balance can be formulated as physical constraints. Such constraints prevent physically implausible outputs, such as a predicted decrease in temperature rise when the contact force falls or an increase in the health index as cumulative damage grows.

In recent years, physics-informed learning has expanded from forward and inverse solutions of partial differential equations to degradation prediction for industrial equipment. Jiang et al. proposed a spatiotemporal-attention-based hidden physics-informed network for equipment RUL prediction that learns latent evolution laws while extracting long-term dependencies from degradation sequences [25]. Kim et al. developed an interpretable physics-informed deep-learning framework that collaboratively estimates equipment RUL through a physical degradation model and a neural network, thereby improving mechanistic interpretability [26]. Ko et al. applied a physics-informed network to long-term prognostics of proton-exchange-membrane fuel cells and constrained cross-timescale performance-degradation trajectories using electrochemical mechanisms [27]. E et al. developed a joint degradation-trajectory and RUL-prediction method for supercapacitors in which feature learning is constrained by device degradation equations [28]. Li et al. proposed a SiMBA-augmented physics-informed neural network that combines state-space temporal modeling, frequency-domain feature mixing, and physical residuals for RUL prediction of complex industrial equipment [29]. Another study used a physics-informed network to jointly estimate the state of health, short-term degradation path, and RUL of lithium-ion batteries, reflecting the trend toward integrated continuous health representation, future-trajectory prediction, and RUL estimation [30].

Despite these advances, three scenario-specific gaps remain for 10 kV tulip-contact prognostics. First, most multimodal RUL frameworks assume synchronous and relatively dense sensor sequences, whereas contact force, surface roughness, and wear are sparsely and asynchronously observed in practical tulip-contact monitoring. Second, direct concatenation or long-interval interpolation may confuse data absence with physical degradation and may distort the actual thermal–electrical–mechanical evolution trajectory. Third, existing physics-informed prognostic models rarely impose the contact-force–contact-resistance–Joule-heating coupling together with missing-data-aware multimodal fusion in a unified model. These limitations motivate PI-TEMFNet, which combines observation masks and measurement recency with reliability-aware cross-modal fusion, thermal–electrical–mechanical physical consistency, and joint degradation-stage, health-index, and

probabilistic RUL prediction.

2. Materials and methods

2.1. Structure and Coupled Degradation Mechanism of Tulip Contacts

This study considers the tulip contact used in 10 kV withdrawable switchgear. The component consists of multiple copper contact fingers, a fixed base, and retaining springs. Radial force applied by the springs provides the mechanical connection and current-transfer path between the contact fingers and the moving and stationary conductors. Although the number of contact fingers and structural parameters vary among tulip-contact models, all are elastic multi-contact systems with parallel current paths.

During long-term operation, frequent insertion and withdrawal, mechanical vibration, assembly deviation, contact-surface wear, and spring stress relaxation can reduce the contact force or cause uneven force distribution among the contact fingers. Meanwhile, increased surface roughness and mating deviation reduce the real contact area, leading to greater contact resistance and localized current density. Under high current, Joule heat generated by the contact resistance raises the contact temperature and further causes thermal deformation of the contact fingers, material softening, and deterioration of spring performance, forming a cumulative degradation process with thermal–electrical–mechanical feedback. Previous studies have shown that contact force, surface roughness, and mating angle significantly affect the contact resistance and temperature-rise characteristics of tulip contacts. The coupled degradation chain can be summarized as follows:

$$\{F_c, R_\alpha, \alpha\} \rightarrow A_r \rightarrow R_c \rightarrow P_J \rightarrow \Delta T \rightarrow \Delta F_c \tag{1}$$

where F_c is the contact force; R_α is the contact-surface roughness; α is the mating deviation or mating angle; A_r is the real contact area; R_c is the contact resistance; P_J is the Joule loss; ΔT is the contact temperature rise; and ΔF_c is the loss of contact force caused by thermal effects and cumulative wear.

$$R_c(n) = K[0.102F_c(n)]^{-m}\Phi(R_\alpha(n), \alpha(n)) \tag{2}$$

where n is the number of operating or insertion–withdrawal cycles; K and m are the material coefficient and contact-form coefficient, respectively; $\Phi(\cdot)$ is a correction function for surface roughness and mating deviation and increases with R_α and $|\alpha|$. This expression indicates that reduced contact force, increased surface roughness, and contact offset all increase the contact resistance. Existing multiphysics studies of tulip contacts have also confirmed that reduced contact force significantly increases both contact resistance and hot-spot temperature rise.

When higher-order radiation terms are neglected and a lumped thermal-parameter representation is adopted, the contact temperature rise can be approximated as follows:

$$C_{\text{th}}\frac{d\Delta T}{dt} = I^2R_c - hA_s\Delta T \tag{3}$$

where C_{th} is the equivalent thermal capacitance, I is the load current, h is the equivalent heat-transfer coefficient, and A_s is the heat-dissipation area. Eq. (3) is primarily used to describe the overall effects of load current, contact resistance, and heat-dissipation conditions on temperature rise, whereas the complete temperature field is verified through experiments and multiphysics simulation. The evolution of contact force during cyclic operation can be expressed as follows:

$$F_c(n+1) = F_c(n) - \Delta F_w(n) - \Delta F_r(T(n), \Delta t) - \Delta F_a(\alpha(n)) \tag{4}$$

where ΔF_w denotes the force loss caused by frictional wear and vibration impact, ΔF_r denotes spring stress relaxation caused jointly by temperature and exposure time, and ΔF_a denotes the equivalent force loss caused by assembly deviation. Eqs. (2)–(4) show that tulip-contact degradation exhibits pronounced accumulation and state-memory characteristics; therefore, temperature or resistance at a single instant is insufficient to characterize the health state completely.

2.2. Multi-Source Thermal–Electrical–Mechanical Degradation Features

Temperature monitoring and loop-resistance measurement are commonly used for switchgear condition assessment. However, temperature is easily affected by load current and environmental conditions, whereas loop resistance generally contains the combined contributions of multiple connection points and therefore reflects the state of the tulip contact only indirectly. Accordingly, this study constructs a multi-source degradation-feature system from the three physical dimensions of thermal, electrical, and mechanical behavior, while load current, ambient temperature, and operating-cycle count are used as auxiliary operating-condition inputs. The feature categories and their physical implications are listed in 1.

Table 1 shows that the three categories of state features are clearly complementary. Thermal features directly reflect localized overheating and abnormal temperature distribution but are susceptible to load and environmental conditions. Electrical features describe the conductive capability

Table 1. Multi-source Degradation Feature System of Thermal-Electrical-Mechanical Characteristics

Feature Category	Representative Features	Main Physical Implication
Thermal features	Maximum temperature, maximum temperature rise, temperature rise rate, thermal time constant, temperature difference among contact fingers, temperature non-uniformity coefficient	Reflect local heat accumulation, heat dissipation capacity, and spatial distribution of hotspots.
Electrical features	Contact resistance, voltage drop, resistance growth rate, equivalent Joule loss, three-phase unbalance degree	Reflect conductive performance, current constriction effect, and electrical contact degradation.
Mechanical features	Average contact force, contact-force non-uniformity coefficient, insertion–withdrawal force, spring stiffness, vibration characteristics, surface roughness, mating deviation	Reflect spring relaxation, contact wear, and assembly status.
Operating condition features	Load current, ambient temperature, energization time, and cycle number	Enable operating condition compensation and decoupling of degradation factors.

of the contact interface and the growth of resistance but cannot independently identify specific causes such as insufficient contact force, surface wear, or assembly deviation. Mechanical features have direct degradation-mechanism implications and can characterize spring relaxation, uneven force distribution, and contact-surface deterioration, but they are generally sparsely sampled and difficult to acquire online. Jointly modeling thermal, electrical, and mechanical features together with operating-condition variables therefore helps distinguish external operating-condition variations from intrinsic tulip-contact degradation.

To reduce the influence of differences in initial specimen parameters and external operating conditions, the following four derived indicators are selected:

$$\begin{cases} \Delta T_{\max}(n) = \max_i T_i(n) - T_{\text{amb}}(n) \\ \eta_T(n) = \dfrac{\text{std}\{\Delta T_i(n)\}}{\text{mean}\{\Delta T_i(n)\} + \varepsilon} \\ \gamma_R(n) = \dfrac{R_c(n) - R_c(0)}{R_c(0)} \\ \eta_F(n) = \dfrac{\text{std}\{F_i(n)\}}{\text{mean}\{F_i(n)\} + \varepsilon} \end{cases} \tag{5}$$

where T_i and F_i are the temperature and contact force of the ith contact finger or measurement point, respectively; T_{amb} is the ambient temperature; and ε is a small constant that prevents division by zero. $\Delta T_{\max}$ characterizes the overall thermal risk, η_T is the temperature non-uniformity coefficient for localized hot-spot and thermal-dispersion characterization, γ_R describes the relative degradation of electrical-contact performance, and η_F is the contact-force non-uniformity coefficient for uneven force distribution and local loss of contact.

Among these indicators, the average contact force and contact resistance mainly reflect the overall degradation trend, whereas the temperature non-uniformity coefficient η_T and contact-force non-uniformity coefficient η_F are more suitable for detecting contact offset, local relaxation, and partial contact-finger failure. Joint use of absolute state quantities, relative growth quantities, and the two non-uniformity indicators improves the identifiability of early degradation features.

2.3. Multi-Rate Data Organization

Different modalities have substantially different sampling frequencies. Temperature, current, voltage, and vibration signals can be acquired continuously or once per operating cycle and therefore constitute dense time-series data. Contact resistance and insertion–withdrawal force are generally measured at fixed cycle intervals, whereas contact force, spring stiffness, surface roughness, and wear are usually available only during outage-inspection periods and are therefore sparse features. Existing circuit-breaker lifetime studies organize continuous features, such as arcing energy, separately from sparse features, such as contact resistance, roughness, and material loss, thereby retaining both high-frequency operating information and low-frequency physical degradation information.

The multi-source input and observation state at the nth cycle are defined as follows:

$$\mathbf{X}_n = [\mathbf{X}_n^T, \mathbf{X}_n^E, \mathbf{X}_n^M, \mathbf{U}_n],$$
$$m_{n,j} = \begin{cases} 1, & x_{n,j} \text{ is actually measured in the } n\text{-th cycle} \\ 0, & x_{n,j} \text{ is missing} \end{cases} \tag{6}$$

where $\mathbf{X}_n^T$, $\mathbf{X}_n^E$, and $\mathbf{X}_n^M$ are the thermal, electrical, and mechanical features, respectively; $\mathbf{U}_n$ denotes the operating-condition features; and $m_{n,j}$ is the observation mask of the jth feature.

For sparse mechanical features, ordinary linear interpolation is not used to construct an artificial continuous degradation trajectory. Instead, the model receives the latest valid observation, an observation mask, and the elapsed time since the previous measurement. This organization allows the subsequent model to distinguish measured values from missing values and dynamically adjust the contribution of each modality according to feature recency. Multi-source feature fusion and sparse-information modeling have been shown to improve the stability of circuit-breaker lifetime prediction when features are incomplete.

2.4. Definitions of Health State and Remaining Useful Life

The end of life of a tulip contact should be jointly determined by its thermal, electrical, and mechanical states rather than by a single temperature threshold. The end-of-life cycle count is defined as follows:

$$N_{\mathrm{EOL}} = \inf\{n : \Delta T_{\max}(n) \geq \Delta T_{\lim} \vee R_c(n) \geq R_{\lim} \vee \bar{F}_c(n) \leq F_{\lim} \vee G(n) = 0\} \tag{7}$$

where $\Delta T_{\lim}$, $R_{\lim}$, and $F_{\lim}$ are the failure thresholds for temperature rise, contact resistance, and average contact force, respectively; G(n)=0 indicates a functional failure such as contact-finger melting, abnormal separation, or inability to carry current stably. The specific thresholds are determined according to the equipment model, relevant test standards, and accelerated degradation test results. At the nth operating cycle, the RUL and continuous health index are defined as follows:

$$\mathrm{RUL}_n = N_{\mathrm{EOL}} - n, \quad \mathrm{HI}_n \in [0,1], \quad \mathrm{HI}_{n+1} \leq \mathrm{HI}_n \tag{8}$$

where HI_n=1 represents the initial healthy state and HI_n=0 represents the end of life. According to the health-index level, degradation rate, and key physical parameters, the tulip contact can be further divided into four stages: healthy, early degradation, accelerated degradation, and severe degradation. The research tasks therefore include degradation-stage identification, continuous health-index estimation, and RUL prediction. These three tasks share a fused thermal–electrical–mechanical representation, and a unified artificial-intelligence prediction model is established in the next section through physical constraints and multitask learning.

To ensure reproducibility of the stage labels, the stage boundaries in the current dataset are determined from change points in the training-set health-index curves and are fixed across all test folds: HI $>$ 0.93 denotes the healthy stage; 0.75 $<$ HI $\leq$ 0.93 denotes the early-degradation stage; 0.35 $<$ HI $\leq$ 0.75 denotes the accelerated-degradation stage; and HI $\leq$ 0.35 denotes the severe-degradation stage. The thresholds are determined exclusively from the training set to prevent test-information leakage.

2.5. Physics-Informed Thermal–Electrical–Mechanical Feature Fusion and Joint Prediction Method

2.5.1. *Overall Framework of PI-TEMFNet*

To address asynchronous sampling of multi-source tulip-contact features, sparse mechanical observations, and insufficient physical consistency in purely data-driven models, this paper proposes a physics-informed thermal–electrical–mechanical feature fusion network (PI-TEMFNet). The model comprises four components: modality-specific temporal encoding, reliability-aware cross-modal fusion, physical-consistency constraints, and multitask prediction. The overall architecture is shown in Fig. 1.

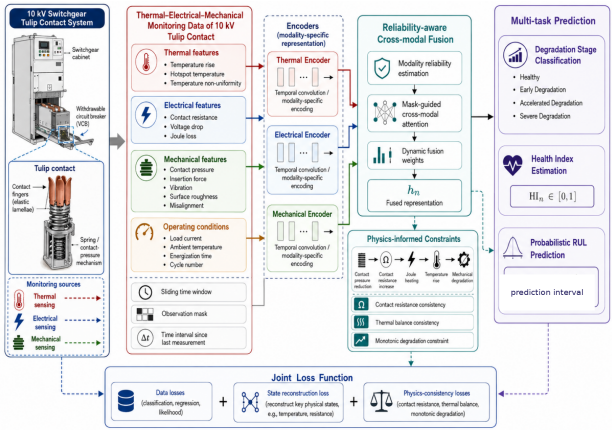

Fig. 1. Overall architecture of the proposed PI-TEMFNet

First, according to the feature system defined in Table 1 of Section 2, the thermal, electrical, and mechanical features are separately fed into independent temporal encoders to extract degradation representations for the different physical modalities. Observation masks and elapsed times since the latest measurements are then used

to estimate the validity of each modality, and masked cross-modal attention enables interaction among heterogeneous information sources. The fused features are used for degradation-stage identification, continuous health-index estimation, and probabilistic RUL prediction. Meanwhile, the contact-resistance relationship, heat-balance relationship, and degradation-monotonicity constraint are introduced to ensure that model predictions comply with the thermal–electrical–mechanical coupling laws of tulip contacts.

Existing circuit-breaker lifetime studies have shown that combining parameter features with explicit engineering meaning and deep temporal features, while using attention to emphasize critical degradation information, can improve lifetime-prediction stability. In addition, separately modeling dense and sparse degradation information enhances predictive capability when some features are missing.

For the nth operating cycle, the input window of modality q is defined as follows:

$$\mathcal{S}_n^q = \{\mathbf{X}_{n-L+1:n}^q, \mathbf{M}_{n-L+1:n}^q, \Delta\boldsymbol{\tau}_{n-L+1:n}^q, \mathbf{U}_{n-L+1:n}\}, \quad q \in \{T, E, M\} \tag{9}$$

where T, E, and M denote the thermal, electrical, and mechanical modalities, respectively; L is the time-window length; X_q is the feature sequence; M_q is the observation mask; $\Delta\tau_q$ is the elapsed time since the most recent valid measurement of each feature; and U comprises operating-condition variables such as load current, ambient temperature, energization time, and cycle count.

2.5.2. *Modality Encoding and Reliability-Aware Fusion*

Different modalities vary on different temporal scales. Thermal features mainly describe temperature-rise inertia and hot-spot evolution, electrical features characterize changes in contact resistance and Joule loss, and mechanical features represent spring relaxation, uneven force distribution, and surface wear. To preserve the independent physical meaning of each modality, temporal convolutional encoders with non-shared parameters are adopted:

$$\mathbf{H}_n^q = \mathcal{E}_q(\mathcal{S}_n^q) \in \mathbb{R}^{L\times d}, \quad q \in \{T, E, M\} \tag{10}$$

where $\mathcal{E}_q(\cdot)$ is the modality encoder and d is the latent-feature dimension. The encoder employs dilated causal convolutions to extract local fluctuations and long-term degradation dependencies. The detailed reliability-aware cross-modal fusion process is shown in Fig. 2.

The reliability-aware cross-modal fusion process comprises three successive stages: modality reliability estimation, mask-guided cross-modal attention, and dynamic feature fusion. The model first combines encoded features,

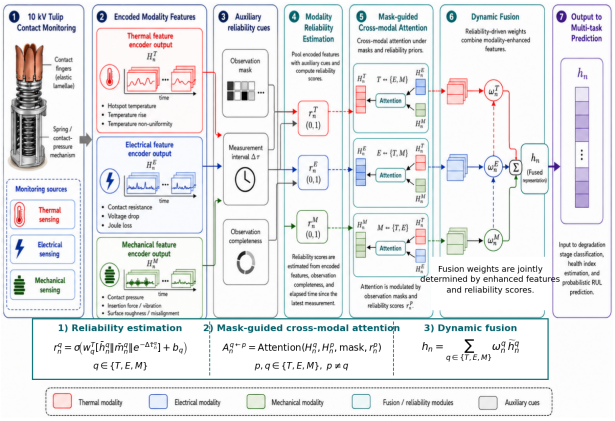

Fig. 2. Reliability-aware cross-modal feature fusion mechanism

observation masks, measurement intervals, and observation completeness to calculate a reliability score for each modality. The reliability scores and observation masks then regulate cross-modal attention, allowing the thermal, electrical, and mechanical modalities to selectively absorb complementary information from other modalities while retaining their own characteristics. Finally, the enhanced features are weighted and aggregated according to dynamic modality weights to obtain the fused representation $\mathbf{h}_n$ for subsequent multitask prediction.

Because sampling completeness and reliability vary over time among modalities, modality reliability is calculated from the latent features, observation ratio, and measurement recency:

$$r_n^q = \sigma\left(\mathbf{w}_q^T[\bar{\mathbf{h}}_n^q \parallel \bar{\mathbf{m}}_n^q \parallel e^{-\Delta\bar{\boldsymbol{\tau}}_n^q}] + b_q\right) \tag{11}$$

where $\bar{\mathbf{h}}_n^q$ is the temporal aggregation of the encoded modality features; $\bar{\mathbf{m}}_n^q$ is the average observation-completeness ratio within the window; $\Delta\bar{\boldsymbol{\tau}}_n^q$ is the average measurement interval; $\parallel$ denotes vector concatenation; and $\sigma(\cdot)$ is the Sigmoid function. When the missing proportion of a modality increases or the elapsed time since its latest measurement becomes longer, its reliability score decreases. In Fig. 2, Part 4 corresponds to Eq. (11), and its outputs r_n^T, r_n^E, and r_n^M represent the dynamic reliability of the thermal, electrical, and mechanical modalities, respectively.

To learn the nonlinear coupling relationships among thermal, electrical, and mechanical features, masked cross-modal attention is employed. The information acquired by modality q from modality p is expressed as follows:

$$\mathbf{A}_n^{q\leftarrow p} = \text{softmax}\left(\frac{\mathbf{Q}_n^q(\mathbf{K}_n^p)^T}{\sqrt{d}} + \log(r_n^p+\varepsilon) + \mathbf{M}_n^p\right)\mathbf{V}_n^p, \quad p \neq q \tag{12}$$

where $\mathbf{Q}_n^q = \mathbf{H}_n^q\mathbf{W}_Q^q$, $\mathbf{K}_n^p = \mathbf{H}_n^p\mathbf{W}_K^p$, and $\mathbf{V}_n^p = \mathbf{H}_n^p\mathbf{W}_V^p$, $\mathbf{\Gamma}_n^p$ is the attention mask generated from the observation mask. Valid observation positions are assigned 0, whereas missing positions are assigned a large negative value, causing the corresponding attention weights to approach 0. After cross-modal information interaction, the fused feature is:

$$\begin{aligned}
\bar{\mathbf{h}}_n^q &= \text{Pool}\left(\mathbf{H}_n^q + \sum_{p\neq q}\mathbf{A}_n^{q\leftarrow p}\right),\\
\omega_n^q &= \frac{\exp(g_q(\bar{\mathbf{h}}_n^q, r_n^q))}{\sum_{p\in\{T,E,M\}}\exp(g_p(\bar{\mathbf{h}}_n^p, r_n^p))},\\
\mathbf{h}_n &= \sum_{q\in\{T,E,M\}}\omega_n^q\bar{\mathbf{h}}_n^q
\end{aligned} \tag{13}$$

where $\text{Pool}(\cdot)$ is the temporal aggregation operation, $g_q(\cdot)$ is the modality-weight network, and ω_n^q is the dynamic fusion weight. This method prevents information imbalance caused by fixed weighting and direct feature concatenation and dynamically adjusts the contribution of each modality according to the degradation stage and data quality.

2.5.3. *Physics-Constrained Multitask Joint Prediction*

The fused feature is simultaneously fed into the degradation-stage, health-index, and RUL prediction branches, whose outputs are:

$$\begin{cases}
\mathbf{p}_n = \text{softmax}(\mathbf{W}_c\mathbf{h}_n + \mathbf{b}_c)\\
\widehat{\text{HI}}_n = \sigma(\mathbf{W}_h\mathbf{h}_n + b_h)\\
\mu_n = \text{Softplus}(\mathbf{W}_\mu\mathbf{h}_n + b_\mu)\\
\sigma_n = \text{Softplus}(\mathbf{W}_\sigma\mathbf{h}_n + b_\sigma) + \varepsilon\\
\widehat{\text{RUL}}_n \sim \mathcal{N}(\mu_n, \sigma_n^2)
\end{cases} \tag{14}$$

where $\mathbf{p}_n$ is the predicted probability of the four states—healthy, early degradation, accelerated degradation, and severe degradation; $\widehat{\text{HI}}_n \in [0,1]$ is the health index; and μ_n and σ_n are the predictive mean and standard deviation of RUL, respectively. The probabilistic output provides not only an RUL estimate but also a characterization of predictive uncertainty. Existing dual-channel lifetime-prediction studies use predictive variance to describe the credibility of different information channels and employ inverse-variance fusion to improve prediction stability. The loss function of the data-driven component is defined as follows:

$$\begin{aligned}
\mathcal{L}_{\text{data}} &= \lambda_c\mathcal{L}_{\text{cls}} + \lambda_h\mathcal{L}_{\text{HI}} + \lambda_r\mathcal{L}_{\text{NLL}} + \lambda_s\mathcal{L}_{\text{state}},\\
\mathcal{L}_{\text{NLL}} &= \frac{1}{2N}\sum_{n=1}^{N}\left[\log\sigma_n^2 + \frac{(\text{RUL}_n-\mu_n)^2}{\sigma_n^2}\right]
\end{aligned} \tag{15}$$

where $\mathcal{L}_{\text{cls}}$ is the weighted cross-entropy loss, $\mathcal{L}_{\text{HI}}$ is the health-index Huber loss, $\mathcal{L}_{\text{NLL}}$ is the Gaussian negative log-likelihood loss for RUL, and $\mathcal{L}_{\text{state}}$ is the auxiliary state-reconstruction loss. To ensure that the fused features retain explicit physical-state information, an auxiliary state decoder is introduced:

$$[\hat{R}_{c,n}, \Delta\hat{T}_n] = \mathcal{D}_{\text{phy}}(\mathbf{h}_n) \tag{16}$$

where $\hat{R}_{c,n}$ and $\Delta\hat{T}_n$ are the reconstructed contact resistance and hot-spot temperature rise, respectively. The auxiliary states are used only as physical constraints during training; when the corresponding states are observed, a masked Huber loss is used to calculate $\mathcal{L}_{\text{state}}$.

In Section 2, Eq. (2) is used to calculate the theoretical contact resistance $R_{c,n}^{\text{phy}}$ from the contact force, surface roughness, and mating deviation. The contact-resistance consistency loss and heat-balance consistency loss are defined as follows:

$$\begin{aligned}
\mathcal{L}_R &= \frac{1}{|\Omega_R|}\sum_{n\in\Omega_R}(\log\hat{R}_{c,n} - \log R_{c,n}^{\text{phy}})^2\\
\mathcal{L}_T &= \frac{1}{|\Omega_T|}\sum_{n\in\Omega_T}\left(C_{\text{th}}\frac{\Delta\hat{T}_n - \Delta\hat{T}_{n-1}}{\Delta t} - I_n^2\hat{R}_{c,n} + hA_s\Delta\hat{T}_n\right)^2
\end{aligned} \tag{17}$$

where Ω_R is the set of samples for which contact force, roughness, and contact resistance are all available, and Ω_T is the set of samples for which continuous temperature-rise and current data are available. The logarithmic contact-resistance error reduces the influence of different numerical scales on model training.

By converting heat generation, transfer, and boundary laws into training losses, a physics-informed network can reduce its dependence on massive observational datasets and suppress predictions that violate physical laws. For tulip contacts, contact force and roughness first alter the contact resistance; the resulting contact resistance produces Joule heat; and temperature rise and thermal stress then feed back into the mechanical state. This physical process has been verified through multiphysics analysis. Considering that tulip contacts exhibit generally irreversible degradation in the absence of maintenance, the following monotonicity constraint is further defined:

$$\mathcal{L}_{\text{mono}} = \frac{1}{N-1}\sum_{n=1}^{N-1}\Big[\max(0,\widehat{\text{HI}}_{n+1}-\widehat{\text{HI}}_n)^2 + \max(0,\mu_{n+1}-\mu_n)^2\Big] \tag{18}$$

This constraint permits local fluctuations in the monitoring data but suppresses physically unreasonable long-term increases in the health index and RUL as the cumulative operating-cycle count rises. The total loss function of PI-TEMFNet is:

$$\mathrm{L} = \lambda_c \mathrm{L}_{\text{cls}} + \lambda_h \mathrm{L}_{\text{HI}} + \lambda_r \mathrm{L}_{\text{NLL}} + \lambda_s \mathrm{L}_{\text{state}} + \lambda_{\text{phy}}\left(\lambda_R \mathrm{L}_R + \lambda_T \mathrm{L}_T + \lambda_M \mathrm{L}_{\text{mono}}\right) \tag{19}$$

where λ_c, λ_h, λ_r, and λ_s are the weights of the degradation-stage loss, health-index regression loss, probabilistic RUL prediction loss, and auxiliary state-reconstruction loss, respectively; λ_R, λ_T, and λ_M are the weights of the contact-resistance consistency, heat-balance consistency, and degradation-monotonicity constraints, respectively; and λ_{phy} is the overall scaling coefficient of the physical-constraint terms. By minimizing Eq. (19), the model jointly optimizes degradation-stage identification, continuous health-state representation, probabilistic RUL prediction, reconstruction of key physical states, and thermal–electrical–mechanical consistency.

2.6. Accelerated Degradation Test Platform and Data Acquisition

To verify the effectiveness of PI-TEMFNet for degradation assessment and RUL prediction of 10 kV tulip contacts, a thermal–electrical–mechanical coupled accelerated degradation test platform was constructed. The test objects were GC5-series tulip contacts rated at 630 A, each comprising 12 contact-finger groups and four retaining springs. The platform mainly consisted of a power-frequency high-current generator, an ambient-temperature control chamber, a contact-resistance tester, contact-force and insertion–withdrawal force measurement devices, a vibration-acquisition module, and a multichannel temperature-recording system. The platform structure and sensor arrangement are shown in Fig. 3.

For thermal-feature acquisition, K-type thermocouples were installed in non-load-bearing regions near the contact interface of each of the 12 contact-finger groups, while ambient temperature was recorded simultaneously. After the specimen was energized, the steady-state temperature rise was recorded when the fluctuation at every measurement point remained within ±1 K for 1 h. Previous tulip-contact experiments have used multichannel thermocouples to monitor the distributed temperature rise of individual contact fingers, enabling effective identification of temperature nonuniformity caused by mating deviation and local loss of contact. The stress conditions of the three test groups are summarized in

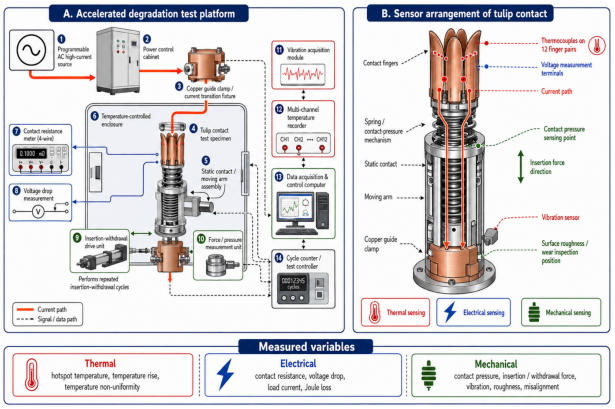

Fig. 3. Accelerated degradation test platform and sensor arrangement for the 10 kV tulip contact

According to Eq. (7), a specimen was considered to have reached the end of life when the maximum temperature rise reached 65 K, the contact resistance reached 2.5 times its initial value, the average contact force decreased to 60% of its initial value, or a functional failure occurred, such as an inability to carry current stably or abnormal separation of contact fingers. The 65 K criterion was adopted as a conservative laboratory end-of-life threshold; it is stricter than the 75 K temperature-rise limit specified for silver- or nickel-coated contacts in oxidizing gas in GB/T 11022-2020 under the corresponding continuous-current test conditions [31]. The 2.5R0 resistance criterion and 0.60F0 contact-force criterion are operational thresholds used in the accelerated-degradation protocol to represent pronounced electrical and mechanical deterioration rather than universal industry limits. Because the ground-truth end-of-life cycle is determined by the earliest threshold crossing, tightening any threshold moves the labeled end of life earlier and reduces absolute RUL labels, whereas relaxing it has the opposite effect. The same label definition is applied to all comparison models, so relative model comparisons are expected to be less sensitive than absolute RUL values; nevertheless, deployment on another contact design requires threshold recalibration according to the applicable standard, manufacturer specification, and field condition.

The electrical features included load current, contact voltage drop, contact resistance, and equivalent Joule loss. Contact resistance was measured using the four-wire method to reduce the effects of lead and fixture re-

Table 2. Accelerated Degradation Test Conditions for Tulip Contacts

Test group	Number of specimens	Load current	Ambient temp. (°C)	Initial contact force (N)	Mating angle (°)	Main acceleration factors
A	4	0.8–1.0 I_r	20–25	180 ± 5	0–0.5	Mechanical cycling
B	4	1.0–1.1 I_r	30–35	160 ± 5	0.5–1.0	Thermal stress / force decay
C	4	1.1–1.2 I_r	35–40	140 ± 5	1.0–1.5	Coupled acceleration

Note: $I_r = 630$ A.

sistance. The mechanical features included average contact force, contact-force non-uniformity coefficient, insertion–withdrawal force, and vibration during insertion and withdrawal. The vibration sampling frequency was set to 20 kHz, while surface roughness and contact-finger wear were measured at specified shutdown-inspection points. Each equivalent degradation cycle consisted of 20 insertion–withdrawal operations and one continuous energized thermal cycle. To obtain different degradation rates and failure modes, 12 specimens were divided into normal-stress, medium-stress, and high-coupled-stress groups.

Measurement traceability of experimental instruments is ensured by periodic calibration. The K-type thermocouple has a measuring range of $-20\,^{\circ}\mathrm{C} \sim 200\,^{\circ}\mathrm{C}$ and a measurement accuracy of $\pm 0.5\,\mathrm{K}$, which is calibrated according to national metrology certificates before tests. The multichannel temperature recorder supports 16-channel synchronous sampling with sampling resolution of 0.1 K. A four-wire micro-ohmmeter is adopted for contact-resistance measurement, with a measurement range of $0.1\ \mu\Omega \sim 200\,\mathrm{m}\Omega$ and basic accuracy $\pm 0.5\%$ reading. The contact-force / insertion-withdrawal-force sensor covers 0–500 N with resolution 0.1 N. The vibration acquisition module achieves a sampling rate up to 20 kHz. All sensing devices are calibrated before each group of accelerated degradation experiments to guarantee the validity of measured data.

2.7. Dataset Construction and Experimental Settings

Temperature, current, voltage-drop, and vibration signals were collected in every degradation cycle; contact resistance and insertion–withdrawal force were measured every 10 cycles; and contact force, surface roughness, and wear were measured every 50 cycles. After outlier removal, operating-condition normalization, and feature extraction, each time step contained 14 thermal features, 8 electrical features, 16 mechanical features, and 4 operating-condition variables. For sparsely observed features, the latest valid measurement was retained together with an observation mask and elapsed-time information; long-interval linear interpolation was not used.

Existing multi-source lifetime studies have demonstrated that prediction accuracy depends strongly on the completeness of degradation information and that missing degradation features significantly increase prediction errors. In addition to complete-modality tests, random-missing, single-modality-missing, and sparse-sampling experiments were therefore designed to evaluate robustness under incomplete field-monitoring conditions. Leave-one-specimen-out cross-validation was used to prevent sample leakage: in each round, 1 specimen was used as the test set, 1 specimen was used for model selection, and the remaining specimens were used for training; the final results are averages over 12 rounds. The input-sequence window length was L = 20, and the latent-feature dimension was 64. Each of the three modality encoders contained 3 layers of dilated temporal convolution with a kernel size of 3 and dilation rates of 1, 2, and 4. The model was optimized using AdamW, with an initial learning rate of 1×10^{-3}, a batch size of 64, and a maximum of 200 epochs. Training followed a staged warm-up strategy: during epochs 1–30, the overall physical-loss scale was set to $\lambda_{\mathrm{phy}} = 0$ so that the modality encoders and multitask prediction heads first learned stable data representations. From epoch 31 onward, the physical-loss contribution was progressively increased toward its target level, while λ_R, λ_T, and λ_M were evaluated only for samples with the corresponding valid physical observations. This staged schedule prevents large early physical residuals from dominating the data losses before the latent representation has stabilized. Hyperparameter selection was performed using the validation specimen in each fold; the held-out test specimen was not used for tuning.

The target loss weight coefficients after hyper-parameter tuning via grid-search on the validation specimen of each fold are set as: λ_c=0.4, λ_h=0.3, λ_r=0.5, λ_s=0.2, λ_R=0.25, λ_T=0.25, λ_M=0.20, and the overall physical-loss scaling coefficient λ_{phy}=0.8. Grid search was conducted within predefined parameter intervals on the validation fold, and the combination yielding the best overall performance (Macro-F1, HI RMSE and RUL MAPE) was selected for final model training.

The baseline models included SVR, GPR, LSTM, GRU-MSA, TCN, Transformer, and TEMFNet without physical constraints. Here, GRU-MSA is an attention-based lifetime-prediction algorithm that fuses parameter and temporal features and has been applied to the mechanical-life prediction of circuit breakers. Degradation-stage identification was evaluated using Accuracy (%), Macro-F1 (%), and AUC; health-index estimation used HI RMSE and R^2; and remaining useful life (RUL) prediction used RUL MAE, RUL RMSE, and RUL MAPE (%). Probabilistic prediction additionally used the 90% prediction-interval coverage probability (PICP90, %), normalized mean prediction-interval width (MPIW), and negative log-likelihood (NLL). In Table 4, RUL MAE and RUL RMSE are reported in units of 10^3 equivalent degradation cycles. The ideal objective was for PICP90 to approach 90% while maintaining a low MPIW.

3. Results and discussion

3.1. Evolution of Degradation Features and Health-State Assessment

Fig. 4 presents representative full-life degradation trajectories for a specimen subjected to high coupled stress. As the equivalent cycle count increased, the average contact force gradually decreased from 181.6 N to 105.2 N, the contact resistance increased from 5.9 $\mu\Omega$ to 13.8 $\mu\Omega$, and the maximum temperature rise increased from 22.4 K to 66.8 K. Meanwhile, the contact-force non-uniformity coefficient and temperature non-uniformity coefficient increased markedly during the middle and late lifetime stages, indicating that some contact fingers experienced local loss of contact before deterioration became evident in the overall mean state.

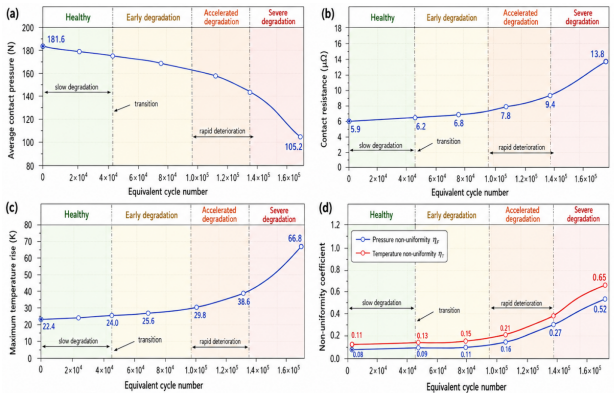

Fig. 4. Full-life degradation trajectories of thermal, electrical, and mechanical features for a representative specimen

During the early degradation stage, the contact force decreased slowly, while the temperature rise and contact resistance changed only slightly. After entering the accelerated-degradation stage, the reduction in contact force decreased the real contact area, accelerated resistance growth, and produced localized hot spots. Multiphysics studies have likewise shown that contact force, surface roughness, and mating angle jointly affect contact resistance and temperature rise, while local contact-finger failure increases the current density in the remaining effective fingers. The degradation-stage classification and health-index prediction results of the different models are summarized in Table 3.

Table 3 shows that PI-TEMFNet achieves an Accuracy of 96.4%, a Macro-F1 of 96.1%, and an AUC of 0.988, outperforming all other models. Its health-index RMSE is 0.041, representing reductions of 34.9% and 32.8% relative to TCN and Transformer, respectively.

Figure 5(a) presents the confusion matrix normalized by the true class. In the current consistency-check dataset, the healthy, early-degradation, accelerated-degradation, and severe-degradation classes contain 193, 203, 493, and 446 samples, respectively. The resulting Accuracy and Macro-F1 are % and 96.1%, consistent with Table 3. The model identifies the healthy and severe-degradation states reliably, while most misclassifications occur between the early- and accelerated-degradation stages. These two stages transition continuously around the turning point, causing some overlap among boundary samples. Figure 5(b) shows that the health index predicted by PI-TEMFNet closely tracks the actual degradation trend and accurately captures the turning point at which the degradation rate increases during the middle and late lifetime stages.

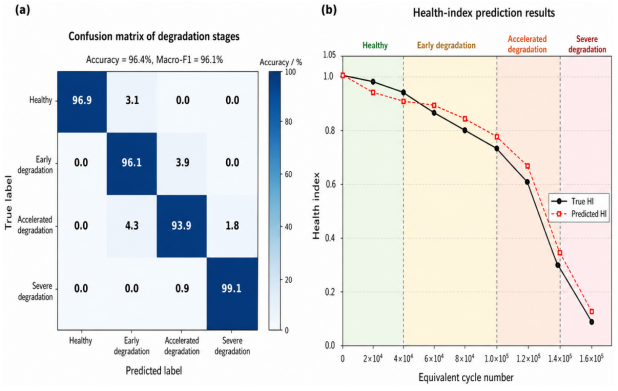

Fig. 5. Confusion matrix of degradation stages and health-index prediction results of PI-TEMFNet

Compared with TEMFNet, the physical constraints make the health-index trajectory smoother and preserve an overall monotonic decrease, reducing abnormal rebounds

Table 3. Degradation Assessment Results of Different Models

Model	Accuracy (%)	Macro-F1 (%)	AUC	HI RMSE	R^2
SVR	84.9	83.1	0.910	0.105	0.842
GPR	87.6	86.2	0.931	0.089	0.881
LSTM	90.8	89.6	0.950	0.078	0.907
GRU-MSA	92.3	91.6	0.963	0.069	0.928
TCN	93.2	92.5	0.970	0.063	0.941
Transformer	93.7	93.1	0.973	0.061	0.945
TEMFNet	95.0	94.6	0.981	0.050	0.963
PI-TEMFNet	96.4	96.1	0.988	0.041	0.975

caused by local measurement fluctuations. This result indicates that thermal–electrical–mechanical physical priors improve the long-term consistency of the health representation without eliminating genuine local fluctuations.

3.2. Remaining Useful Life Prediction and Uncertainty Analysis

The RUL prediction performance of the different models is presented in Table 4.

Table 4 shows that PI-TEMFNet achieves an MAE of 21.6×10^3 cycles, an RMSE of 29.4×10^3 cycles, and a MAPE of 5.8%, significantly outperforming the single-modality temporal models and TEMFNet without physical constraints. Compared with GRU-MSA, PI-TEMFNet reduces MAPE by 43.1%; compared with TEMFNet, it reduces MAPE by 21.6%.

The 90% prediction-interval coverage probability of PI-TEMFNet is 91.3%, close to the nominal coverage level of 90%. Its MPIW is lower than those of GPR and TEMFNet, indicating that the higher PICP is not obtained by excessively widening the prediction interval. The lower NLL further demonstrates that the output probability distribution has favorable accuracy and calibration. To analyze predictive performance at different monitoring stages, 30%, 50%, 70%, and 90% of the full-life data were used as the observed sequence. The MAPE values of the models under different observed-life ratios are listed in Table 5.

The performance values reported in Tables 3 and 4 are averages over the 12 leave-one-specimen-out test folds. All quantitative metrics reported in Table 3 and Table 4 correspond to the mean value over 12-fold leave-one-specimen-out cross-validation. For PI-TEMFNet, key metrics are given as mean ± standard deviation: Accuracy 96.4±1.1%, Macro-F1 96.1±1.3%, HI RMSE 0.041±0.006, RUL MAPE 5.8±0.7%. Corresponding key results of the representative baseline TEMFNet are Accuracy 95.0±1.4%, Macro-F1 94.6±1.5%, HI RMSE 0.050±0.008, RUL MAPE 7.4±0.9%. Fold-level raw metric records are available in supplementary materials.

As shown in Table 5, the prediction errors of all models decrease as more observations become available, but PI-TEMFNet maintains the lowest error even in the early lifetime stage. This is mainly because average contact force, the contact-force non-uniformity coefficient, and resistance growth rate provide degradation priors before pronounced overheating occurs, while the physical constraints further restrict unreasonable extrapolation under limited early-stage samples. Figure 6 shows the RUL prediction curves and 90% prediction intervals of a representative test specimen under different observed-life ratios. The prediction intervals are wider in the early lifetime stage, reflecting uncertainty caused by limited degradation information. As thermal, electrical, and mechanical features accumulate, the predictive mean gradually approaches the true RUL and the interval narrows accordingly.

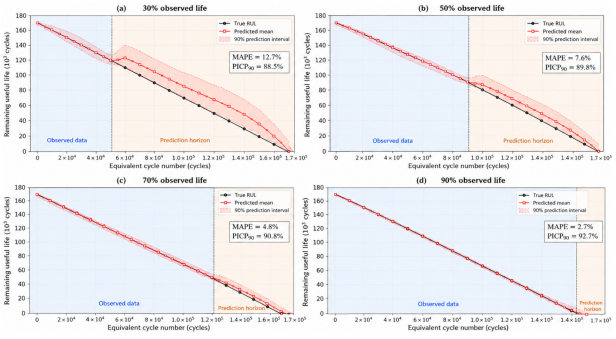

Fig. 6. Probabilistic RUL prediction results under different observed-life ratios

3.3. Ablation Experiments and Robustness Analysis

To quantify the contribution of each component, ablation experiments were conducted using the same data partitions and training settings. The results are summarized in Table 6

To quantify the contribution of each component, ablation experiments were performed under identical data partitions and training settings, as shown in Table 6. Removing any physical modality reduces model performance, with re-

Table 4. RUL Prediction Results of Different Models

Model	RUL MAE (10^3 cycles)	RUL RMSE (10^3 cycles)	RUL MAPE (%)	PICP90 (%)	MPIW	NLL
SVR	56.3	72.4	15.6	—	—	—
GPR	48.5	63.1	13.4	88.1	0.302	2.14
LSTM	42.7	55.9	11.7	—	—	—
GRU-MSA	37.1	49.0	10.2	—	—	—
TCN	34.5	45.6	9.4	—	—	—
Transformer	33.2	44.0	9.1	—	—	—
TEMFNet	27.0	35.9	7.4	86.8	0.238	1.94
PI-TEMFNet	21.6	29.4	5.8	91.3	0.216	1.59

Table 5. RUL Prediction MAPE under Different Observed-Life Ratios

Model	30% observed life (%)	50% observed life (%)	70% observed life (%)	90% observed life (%)
SVR	28.4	20.1	13.5	8.1
GPR	25.1	17.5	11.2	6.6
LSTM	22.3	14.9	9.4	5.7
TCN	19.6	12.6	8.0	4.9
TEMFNet	15.2	9.6	6.3	3.8
PI-TEMFNet	12.7	7.6	4.8	2.7

moval of the mechanical modality causing the largest deterioration. Mechanical features directly reflect contact-force decay, uneven force distribution, and surface wear and therefore provide important evidence for early degradation identification. Removing the thermal modality weakens the identification of severe degradation and near-failure states, while removing the electrical modality impairs the intermediate representation of the transition from mechanical change to thermal fault. After reliability estimation is removed, RUL MAPE increases from 5.8% to 7.8%, indicating that fixed weights cannot adapt to different sampling completeness levels and degradation stages. After the physical constraints are removed, the mean error increases by only 1.5 percentage points, but PICP decreases from 91.3% to 86.8%, demonstrating that physical constraints improve not only point-prediction accuracy but also probabilistic-interval calibration.

Random packet loss, long-term absence of a single modality, and noise interference were further simulated. The results are presented in Table 7.

Under random missing rates of 20% and 40%, the RUL MAPE of PI-TEMFNet increases by only 0.5 and 1.4 percentage points, respectively, indicating that the observation mask, measurement interval, and reliability weights effectively reduce the influence of incomplete data. By contrast, when linear interpolation is directly applied without an observation mask, MAPE rises to 12.1%, showing that continuous interpolation of sparse mechanical features can produce unrealistic degradation trajectories. When the input-signal SNR is set to 30, 20, and 10 dB, the RUL MAPE values are 6.1%, 6.8%, and 8.2%, respectively. Performance gradually deteriorates as noise increases, but under strong 10 dB noise it remains superior to TCN and Transformer with complete data, indicating that multimodal complementarity and physical constraints suppress the influence of single-sensor noise on the final prediction.

4. Conclusions

(1) PI-TEMFNet jointly assesses the degradation stage, health index, and probabilistic remaining useful life of 10 kV tulip contacts through modality-specific encoding, reliability-aware cross-modal fusion, and physical-consistency constraints.

(2) The model achieved an Accuracy of 96.4%, a Macro-F1 of 96.1%, an AUC of 0.988, and an HI RMSE of 0.041. The RUL MAE, RUL RMSE, and RUL MAPE were 21.6×10^3 cycles, 29.4×10^3 cycles, and 5.8%, respectively, while $PICP_{90}$, MPIW, and NLL were 91.3%, 0.216, and 1.59.

(3) Ablation and incomplete-monitoring tests confirmed the contribution of the thermal–electrical–mechanical fusion framework. Removing the mechanical modality, reliability estimation, and physical constraints increased MAPE to 10.5%, 7.8%, and 7.3%, respectively; MAPE remained 6.3% and 7.2% under 20% and 40% random missing data, and was 8.2% under 10 dB noise.

(4) Future work will proceed progressively in three stages. First, long-term field samples will be expanded across

Table 6. Ablation Results of PI-TEMFNet

Model Configuration	Macro-F1 (%)	HI RMSE	RUL MAPE (%)	PICP90 (%)
Remove thermal modality	92.0	0.066	9.6	87.5
Remove electrical modality	92.7	0.060	8.7	88.3
Remove mechanical modality	91.1	0.072	10.5	86.9
Remove reliability estimation	93.6	0.055	7.8	88.8
Remove cross-modal attention	94.0	0.053	7.4	89.0
Remove physical constraints	94.7	0.050	7.3	86.8
Remove multi-task joint learning	94.2	0.052	7.1	89.6
Full PI-TEMFNet	96.1	0.041	5.8	91.3

Table 7. Model Robustness under Incomplete Monitoring Conditions

Test Condition	Macro-F1 (%)	HI RMSE	RUL MAPE (%)
Complete data	96.1	0.041	5.8
20% random missing	95.3	0.045	6.3
40% random missing	93.9	0.052	7.2
Thermal modality completely missing	92.0	0.066	9.6
Electrical modality completely missing	92.7	0.060	8.7
Mechanical modality completely missing	91.1	0.072	10.5
Linear interpolation without mask	89.8	0.081	12.1

manufacturers, contact models, sites, and operating conditions to establish cross-model generalization. Second, after a broader field dataset is available, digital-twin calibration and online incremental learning will be introduced to reduce model–equipment mismatch and update degradation trajectories during service. Third, on the basis of validated cross-model robustness and calibrated online prediction, lightweight edge deployment and maintenance-decision support will be implemented and evaluated in operational switchgear.

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] J. A. Greenwood and J. B. P. Williamson, (1966) *“Contact of nominally flat surfaces”* **Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences *295***(1442): 300–319. DOI: 10.1098/rspa.1966.0242.

[2] M. G. Cooper, B. B. Mikic, and M. M. Yovanovich, (1969) *“Thermal contact conductance”* **International Journal of Heat and Mass Transfer *12***(3): 279–300. DOI: 10.1016/0017-9310(69)90011-8.

[3] M. M. Yovanovich, (2005) *“Four decades of research on thermal contact, gap, and joint resistance in microelectronics”* **IEEE Transactions on Components and Packaging Technologies *28***(2): 182–206. DOI: 10.1109/TCAPT.2005.848483.

[4] M. Bahrami, J. R. Culham, M. M. Yovanovich, and G. E. Schneider, (2006) *“Review of thermal joint resistance models for nonconforming rough surfaces”* **Applied Mechanics Reviews *59***(1): 1–12. DOI: 10.1115/1.2110231.

[5] M. W. Hoffmann, S. Wildermuth, R. Gitzel, A. Boyaci, J. Gebhardt, H. Kaul, I. Amihai, B. Forg, M. Suriyah, T. Leibfried, V. Stich, J. Hicking, M. Bremer, L. Kaminski, D. Beverungen, P. zur Heiden, and T. Tornede, (2020) *“Integration of novel sensors and machine learning for predictive maintenance in medium voltage switchgear to enable the energy and mobility revolutions”* **Sensors *20***(7): 2099. DOI: 10.3390/s20072099.

[6] Y. A. M. Alsumaidaee, C. T. Yaw, S. P. Koh, S. K. Tiong, C. P. Chen, and K. Ali, (2022) *“Review of medium-voltage switchgear fault detection in a condition-based monitoring system by using deep learning”* **Energies *15***(18): 6762. DOI: 10.3390/en15186762.

[7] Y. A. M. Alsumaidaee, J. K. S. Paw, C. T. Yaw, S. K. Tiong, C. P. Chen, T. F. Yusaf, F. Benedict, K. Kadirgama, T. C. Hong, and A. N. Abdalla, (2023) *"Fault detection for medium voltage switchgear using a deep learning hybrid 1D-CNN-LSTM model"* **IEEE Access *11***: 97574–97589. DOI: 10.1109/ACCESS.2023.3294093.

[8] K. A. A. Mahmoud, M. M. Badr, N. A. Elmalhy, R. A. Hamdy, S. Ahmed, and A. A. Mordi, (2024) *"Transfer learning by fine-tuning pre-trained convolutional neural network architectures for switchgear fault detection using thermal imaging"* **Alexandria Engineering Journal *103***: 327–342. DOI: 10.1016/j.aej.2024.05.102.

[9] M. Matin, A. Dehghanian, A. H. Zeinaddini, and H. Darijani, (2025) *"Interpretable machine learning modeling of temperature rise in a medium voltage switchgear using multiphysics CFD analysis"* **Case Studies in Thermal Engineering *65***: 105585. DOI: 10.1016/j.csite.2024.105585.

[10] C. Tang, J. Zhang, Y. Yang, K. Gao, B. Chen, A. Hussain, and S. Fan, (2026) *"Infrared anomaly detection of switchgear equipment based on lightweight deep neural networks"* **Electric Power Systems Research *256***: 112910. DOI: 10.1016/j.epsr.2026.112910.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, (2018) *"Machinery health prognostics: A systematic review from data acquisition to RUL prediction"* **Mechanical Systems and Signal Processing *104***: 799–834. DOI: 10.1016/j.ymssp.2017.11.016.

[17] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, (2019) *"Deep learning and its applications to machine health monitoring"* **Mechanical Systems and Signal Processing *115***: 213–237. DOI: 10.1016/j.ymssp.2018.05.050.

[18] X. Li, Q. Ding, and J.-Q. Sun, (2018) *"Remaining useful life estimation in prognostics using deep convolution neural networks"* **Reliability Engineering & System Safety *172***: 1–11. DOI: 10.1016/j.ress.2017.11.021.

[19] J. Pan, B. Sun, Z. Wu, Z. Yi, Q. Feng, Y. Ren, and Z. Wang, (2024) *"Probabilistic remaining useful life prediction without lifetime labels: A Bayesian deep learning and stochastic process fusion method"* **Reliability Engineering & System Safety *250***: 110313. DOI: 10.1016/j.ress.2024.110313.

[20] F. Wu, Q. Wu, Y. Tan, and X. Xu, (2024) *"Remaining useful life prediction based on deep learning: A survey"* **Sensors *24***(11): 3454. DOI: 10.3390/s24113454.

[21] Y. Wang, Y. Lei, N. Li, X. Gao, X. Liu, Q. Ren, and J. Jiang, (2024) *"A multimodal dynamic parameterized bilinear factorized framework for remaining useful life prediction under variational data"* **Reliability Engineering & System Safety *245***: 110025. DOI: 10.1016/j.ress.2024.110025.

[22] L. Zhou and H. Wang, (2024) *"An adaptive multi-scale feature fusion and adaptive mixture-of-experts multi-task model for industrial equipment health status assessment and remaining useful life prediction"* **Reliability Engineering & System Safety *248***: 110190. DOI: 10.1016/j.ress.2024.110190.

[23] L. Zhou and H. Wang, (2024) *"MST-GAT: A multi-perspective spatial-temporal graph attention network for multi-sensor equipment remaining useful life prediction"* **Information Fusion *110***: 102462. DOI: 10.1016/j.inffus.2024.102462.

[24] M. Raissi, P. Perdikaris, and G. E. Karniadakis, (2019) *"Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations"* **Journal of Computational Physics *378***: 686–707. DOI: 10.1016/j.jcp.2018.10.045.

[25] F. Jiang, X. Hou, and M. Xia, (2025) *"Spatio-temporal attention-based hidden physics-informed neural network for remaining useful life prediction"* **Advanced Engineering Informatics *63***: 102958. DOI: 10.1016/j.aei.2024.102958.

[26] M. Kim, S. Yoo, S. Son, S. Y. Chang, and K. Y. Oh, (2025) *"Physics-informed deep learning framework for explainable remaining useful life prediction"* **Engineering Applications of Artificial Intelligence *143***: 110072. DOI: 10.1016/j.engappai.2025.110072.

[27] T. Ko, D. Kim, J. Park, J. Park, and S. H. Lee, (2025) *"Physics-informed neural network for long-term prognostics of proton exchange membrane fuel cells"* **Applied Energy *382***: 125318. DOI: 10.1016/j.apenergy.2025.125318.

[28] L. E, J. Wang, R. Yang, C. Wang, H. Li, and R. Xiong, (2025) *"A physics-informed neural network-based method for predicting degradation trajectories and remaining useful life of supercapacitors"* **Green Energy and Intelligent Transportation *4***(3): 100291. DOI: 10.1016/j.geits.2025.100291.

[29] M. Li, J. Qin, H. Fan, and T. Ke, (2025) *"SiMBA-augmented physics-informed neural networks for industrial remaining useful life prediction"* **Machines *13***(6): 452. DOI: 10.3390/machines13060452.

[30] Y. Li, M. He, Y. Ren, B. Gao, and H. Qi, (2025) *"Physics-informed neural network for co-estimation of state of health, remaining useful life, and short-term degradation path in lithium-ion batteries"* **Applied Energy *398***: 126427. DOI: 10.1016/j.apenergy.2025.126427.

[31] National Standard of the People's Republic of China. *GB/T 11022-2020: Common specifications for high-voltage alternating-current switchgear and controlgear standards*. Standards Press of China. Published 14 December 2020; implemented 1 July 2021. 2020.