processes

MDPI

Article

Research on an Intelligent Diagnosis and Decision Support System for Pumped Storage Units Based on Multi-Source Data Fusion and Hybrid Intelligent Algorithms

Xuan Liu [1], Jie Bai [1], Bingjie Dou [1], Tianyu Liu [2], Xiaohui Yang [2] and Jie Zhao [3,*]

[1] Inner Mongolia Power Science Research Institute Branch, Inner Mongolia Power (Group) Co., Ltd., Hohhot 010020, China; xuanliu_1990@163.com (X.L.); bjlovesoccer@163.com (J.B.); dbj2015@tju.edu.cn (B.D.)
[2] Inner Mongolia Power (Group) Co., Ltd., Hohhot 010040, China; liutianyu920302@126.com (T.L.); yangxh6666@163.com (X.Y.)
[3] School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China
* Correspondence: jiez_whu@whu.edu.cn

Abstract

Pumped storage hydropower (PSH) is a key regulating resource for renewable energy integration and power system stability. Due to frequent start-stop operations, deep peak-load regulation, and bidirectional operating conditions, stator winding insulation degradation, rotor inter-turn short circuits, and end-winding vibration have become the dominant failure modes of pumped storage units. Conventional monitoring systems are limited by single-source sensing, asynchronous data acquisition, high misdiagnosis rates, and maintenance decisions that rely heavily on expert experience, making traditional periodic maintenance increasingly inadequate. To address these challenges, this study proposes an intelligent diagnosis and decision support system based on multi-source data fusion and hybrid intelligent algorithms. An Intelligent Electronic Device (IED)-based condition monitoring platform is developed by integrating multiple sensing technologies. Complete Variational Mode Decomposition (CVMD) and Kernel Principal Component Analysis (KPCA) are employed to extract representative features from multi-physical-field data, while an attention-enhanced Long Short-Term Memory (LSTM) network is introduced for accurate fault identification. In addition, adaptive time-alignment and joint denoising algorithms are developed to improve data quality and diagnostic robustness. A predictive maintenance framework incorporating health assessment and remaining useful life prediction is further established to optimize maintenance scheduling. Results demonstrate that the proposed system achieves a fault prediction accuracy of over 90% and reduces annual maintenance costs by approximately 15–20%. The proposed framework provides an effective solution for intelligent operation and maintenance of modern pumped storage units.

Keywords: pumped storage hydropower; multi-source data fusion; condition monitoring; attention-LSTM; remaining useful life prediction

Check for updates

Academic Editor: Zhibin Lin

Received: 4 July 2026
Revised: 5 August 2026
Accepted: 7 August 2026
Published: 17 August 2026

1. Introduction

The global energy transition and China's "Dual Carbon" strategy are driving power systems toward increasingly high levels of renewable energy integration. However, the inherent intermittency and fluctuations of renewable sources place growing demands on frequency regulation, peak shaving, and grid balancing. Pumped storage hydropower (PSH) represents the most mature large-scale energy storage technology and plays an indispensable role in facilitating renewable energy consumption while ensuring secure and stable power system operation [1,2].

Coordinating wind and PSH systems poses significant challenges in balancing water spillage and power curtailment, particularly during wet seasons when transmission capacity is limited [3]. This highlights the need for improved operational strategies and, more fundamentally, enhanced reliability of PSH units, since unexpected outages directly impair system regulation capacity and exacerbate renewable curtailment.

Unlike conventional turbine generators, PSH units operate under bidirectional rotation, frequent start-stop cycles, deep peak-load regulation, and rapid load variations [4–6]. These severe conditions accelerate thermal and mechanical stresses within generator components. Repeated operating-mode transitions induce thermal cycling of stator windings, differential expansion between copper conductors and iron cores, and intensified electromagnetic forces at winding end regions. Statistics indicate that stator winding insulation degradation, rotor inter-turn short circuits, and end-winding vibration faults account for approximately 82.6% of all failures in PSH units [7,8].

Traditional time-based maintenance strategies are increasingly inadequate for addressing these complex operating characteristics. Existing monitoring systems suffer from single-dimensional sensing, asynchronous multi-source data acquisition, and high fault misdiagnosis rates, limiting their capability for early fault detection and prognostics [9]. Therefore, developing an intelligent diagnosis and decision support framework based on multi-source data fusion is of great significance for improving unit reliability, reducing maintenance costs, and enhancing renewable energy accommodation capacity.

Current condition monitoring technologies for PSH units mainly target three critical fault categories: stator insulation degradation, rotor inter-turn short circuits, and end-winding vibration. In stator insulation monitoring, partial discharge (PD) measurement is widely recognized as the most effective method for evaluating insulation deterioration [10]. Standards including IEC 60270:2015 and GB/T 7354-2018 provide standardized procedures for PD testing [11,12]. High-frequency current transformer and capacitive coupling techniques have become mainstream online monitoring approaches owing to their high sensitivity and strong anti-interference capability [13]. Nevertheless, PD diagnosis still relies heavily on expert interpretation, and interference signals may lead to misdiagnosis rates exceeding 40% [14]. For rotor inter-turn short-circuit detection, conventional offline methods such as no-load short-circuit characteristic testing and recurrent surge oscillography (RSO) are unable to capture intermittent faults [15,16]. Online monitoring methods based on search coils and magnetic flux measurement have attracted increasing attention, with full-flux probe technology enabling simultaneous measurement of main and leakage magnetic fluxes and facilitating fault localization without interrupting machine operation [17,18]. In end-winding vibration monitoring, significant advances have been achieved through Fiber Bragg Grating (FBG) sensors, laser Doppler vibrometers (LDVs), and embedded piezoelectric-film sensors [19–21]. These technologies provide high-precision vibration measurements under strong electromagnetic interference. However, challenges related to long-term reliability, environmental adaptability, and sensor durability remain unresolved.

To systematically position our contributions within the current state of the art, we categorize recent advances and persistent limitations into four thematic dimensions.

Temporal synchronization of heterogeneous data. Multi-source monitoring data in PSH stations are acquired at vastly different sampling rates, with PD signals at MHz to GHz, vibration at kHz, and electrical and thermal parameters at Hz or lower. Simple concatenation without temporal alignment causes feature misregistration and diagnostic inaccuracies. Recent studies have employed global resampling or fixed-window interpolation to synchronize multi-rate signals [22,23], yet these methods introduce cumulative errors exceeding 5 ms under frequent operating-mode transitions. More importantly, they fail to adapt to the non-stationary nature of PSH operations. Despite progress in multi-rate signal processing for rotating machinery [24], an event-triggered adaptive synchronization scheme tailored to PSH units remains absent.

Joint denoising of heterogeneous signals. The harsh electromagnetic environment of PSH stations severely corrupts weak fault signatures, particularly PD pulses and incipient vibration anomalies. Wavelet thresholding [25] and empirical mode decomposition (EMD) [26] have been widely applied for signal denoising, but each suffers from inherent limitations. Wavelet methods require manual threshold selection and fixed basis functions, while EMD is susceptible to mode mixing and lacks theoretical convergence guarantees. Variational Mode Decomposition (VMD) [27] partially alleviates mode mixing through variational optimization, yet its performance depends on a pre-specified number of modes and fixed bandwidth constraints. More recent hybrid approaches combining VMD with wavelet denoising have shown improved performance in bearing fault diagnosis [28], but their extension to PSH multi-signal fusion remains limited, and none addresses the specific spectral characteristics of PD, magnetic flux, and vibration signals simultaneously.

Nonlinear feature fusion. Principal Component Analysis (PCA) has been widely adopted for dimensionality reduction in power system monitoring [23], yet its linear nature fails to capture nonlinear correlations among heterogeneous monitoring parameters. Kernel PCA (KPCA) extends PCA to nonlinear feature spaces [29,30] and has demonstrated advantages in fault feature extraction. However, most existing KPCA implementations for rotating machinery rely on fixed kernel bandwidths without systematic optimization for the distinct frequency distributions of PD, vibration, and magnetic flux signals. Furthermore, the interaction among these features, for instance, the coupling between 100 Hz vibration amplitude and magnetic flux asymmetry, is typically lost under standard feature fusion frameworks [31].

Temporal deep learning for fault diagnosis. Long Short-Term Memory (LSTM) networks [24] and Convolutional Neural Networks (CNNs) [25] have become mainstream tools for processing time-series monitoring data. Hybrid LSTM-CNN architectures [26] have shown promising performance in generator fault diagnosis by combining spatial feature extraction with temporal sequence modeling. Transfer learning and Physics-Informed Neural Networks (PINNs) have been introduced to address small-sample learning and incorporate physical knowledge into data-driven models [27]. More recently, attention mechanisms have been incorporated into LSTM networks for rotating machinery fault diagnosis [32,33], enabling models to focus on diagnostically relevant time segments. Nevertheless, in the context of PSH units, existing attention-LSTM implementations do not explicitly account for the unique 2-to-5-day fault evolution cycle of insulation degradation and the transient shock events triggered by bidirectional operating-mode transitions, leaving critical time windows underexploited.

In summary, while substantial progress has been made in each of the above dimensions, no unified framework addresses all four challenges simultaneously for PSH units. Existing integrated diagnostic systems for rotating electrical machines typically combine only two or three of these aspects, and none has been validated on full-scale PSH units under actual field conditions over extended periods. This integrated gap, combining event-triggered temporal synchronization, hierarchical joint denoising, nonlinear feature fusion, and attention-enhanced temporal learning into a single end-to-end framework validated on 12 months of field data, constitutes the primary motivation of this study.

With the rapid development of artificial intelligence, data-driven fault diagnosis techniques have become an important research direction. Machine learning and deep learning methods have demonstrated remarkable performance in feature extraction and fault classification. Despite substantial progress in condition monitoring and fault diagnosis technologies, several challenges remain unresolved, including asynchronous heterogeneous data acquisition, severe noise contamination, insufficient degradation feature characterization, and limited practical applicability of decision support systems. To address these issues, this study investigates both intelligent fault diagnosis and maintenance decision support for pumped storage units. An Intelligent Electronic Device (IED)-based monitoring platform integrating full-flux rotor monitoring, FBG vibration sensing, and epoxy-mica capacitive-coupling partial discharge detection is developed for synchronized acquisition of multi-physical-field data. An adaptive time-alignment and joint denoising framework is proposed, together with a hybrid diagnostic model combining Complete Variational Mode Decomposition (CVMD) [28], KPCA, and an attention-enhanced LSTM network. Furthermore, a reliability-centered predictive maintenance model based on health condition assessment and remaining useful life (RUL) [29] prediction is established to optimize maintenance scheduling and resource allocation. In the domain of rotating machinery RUL prediction, various data-driven and model-based approaches have been proposed. For instance, Wang et al. [34] investigated the remaining useful life prediction of sliding bearings in nuclear power plant shielded pumps using a nearest-similar-distance particle-filtering approach. Their work demonstrates that hybrid approaches combining similarity-based matching with sequential Bayesian estimation can effectively capture degradation trajectories even under limited data conditions. While various optimization strategies have been proposed for energy system coordination, as demonstrated in the context of solar-electric vehicle home energy optimization [35], these approaches typically focus on operational-level decisions rather than equipment-level reliability. This gap further motivates our research on predictive maintenance as a complementary enabler for overall system efficiency.

Problem Statement

Three fundamental challenges remain unresolved in current practice. First, asynchronous heterogeneous data fusion: monitoring data from different sensors (PD, vibration, magnetic flux, and electrical parameters) are acquired at vastly different sampling rates (from 1 Hz to 1 GHz) and often lack precise temporal alignment. The absence of an effective synchronization mechanism leads to information loss or misdiagnosis when these data are simply concatenated without considering their temporal relationships. Second, strong noise contamination and weak fault signature extraction: the harsh electromagnetic environment of PSH stations severely corrupts weak fault signals, particularly PD pulses and incipient vibration anomalies. Conventional denoising methods such as wavelet thresholding and empirical mode decomposition are either insufficiently adaptive to the multi-modal characteristics of PSH signals or risk removing fault-relevant components alongside noise. Third, limited integration of diagnostic and prognostic functionalities: most existing approaches address fault detection and remaining useful life (RUL) prediction as separate tasks, without a unified framework that bridges diagnostics to maintenance decision-making. Moreover, conventional deep learning models treat all temporal features equally and fail to emphasize the critical time windows where fault signatures are most pronounced, resulting in delayed or missed warnings.

To address these challenges, this study aims to develop an intelligent diagnosis and decision-support system that achieves precise temporal synchronization of heterogeneous multi-source data via an event-triggered adaptive alignment algorithm, enhances weak fault feature extraction through a hierarchical CVMD-wavelet joint denoising strategy, improves fault identification accuracy by incorporating an attention mechanism that automatically focuses on informative temporal segments, and provides a unified predictive maintenance framework integrating health assessment and RUL prediction to support maintenance decision-making.

To clarify the novelty of this work, we explicitly distinguish our original contributions from existing literature as follows.

(1) Novelty lies in the systematic integration and cross-layer optimization, not in any single algorithm. While CVMD, KPCA, Attention-LSTM, AHP, and RUL prediction are individually established techniques, their tailored integration for pumped storage units under strong electromagnetic interference and multi-rate asynchronous data streams has not been systematically addressed in previous studies.

(2) Algorithm-specific adaptations for PSH signals are introduced. These include targeted modifications to each component for handling PD, magnetic flux, and vibration signals, comprising CVMD with adaptive mode-number determination, KPCA with optimized kernel bandwidth selection, and an Attention-LSTM with a time-window design tailored to the 2-to-5-day fault evolution cycle typical of PSH units. These adaptations are not found in generic implementations.

(3) A unified end-to-end framework seamlessly bridges multi-physical sensing, data preprocessing, feature fusion, fault diagnosis, health assessment, and maintenance decision-making. While prior studies have addressed isolated aspects such as single-signal diagnosis or standalone RUL prediction, none have demonstrated a fully integrated pipeline validated on 12 months of field data from an operational 300 MW pumped storage station with full operating-mode coverage, including pumping, generating, phase modulation, and rapid transitions.

(4) Beyond academic contributions, the framework delivers quantifiable engineering outcomes, including 94.2% ± 0.7% diagnostic accuracy, 5-to-10-day advance warning, and 19.8% maintenance cost reduction, which are validated through rigorous economic analysis based on actual power-station financial data.

A thorough comparison with previous integrated diagnostic frameworks for rotating electrical machines and pumped storage units is summarized in Table 1.

Table 1. Comparison between the proposed framework and existing integrated diagnostic systems for rotating electrical machines/PSH units.

Aspect	Existing Integrated Frameworks	This Work
Multi-source data types	Mostly single-signal (vibration-only) or loosely coupled	Tightly fused PD + magnetic flux + vibration + electrical parameters
Temporal synchronization	Global interpolation with errors > 5 ms	Event-triggered adaptive alignment with error < 1 ms
Joint denoising	Single-stage wavelet or EMD	Hierarchical CVMD + adaptive wavelet threshold
Nonlinear feature fusion	Linear PCA or manual selection	KPCA with Gaussian kernel (nonlinear)
Temporal modeling	Vanilla LSTM or CNN (no attention)	Attention-LSTM with automatic critical-time focusing
Health assessment + RUL	Separate or absent	Unified with AHP-weighted HI and attention-based RUL
Field validation period	Typically days or weeks	12 months, 4 units, full operating conditions

The remainder of this paper is organized as follows. Section 2 introduces the fault mechanisms of pumped storage units and the principles of monitoring technologies. Section 3 presents the proposed hybrid intelligent fault diagnosis model. Section 4 describes the predictive maintenance decision support system. Section 5 validates the effectiveness of the proposed framework through field experiments. Finally, Section 5 summarizes the main conclusions and discusses future research directions.

2. Multi-Dimensional Condition Monitoring Framework for Pumped Storage Units

Pumped storage motor-generators operate under both generating and pumping modes and are required to switch frequently among pumping, generating, synchronous condenser, and shutdown conditions. The annual number of start-stop cycles can be five to ten times higher than that of conventional turbine generators, while the load regulation rate may exceed 10 MW/s. These extreme operating characteristics expose the units to a strongly coupled multi-physical environment involving electrical, thermal, mechanical, and chemical fields, resulting in more complex fault evolution mechanisms. Statistics indicate that stator winding insulation faults, rotor inter-turn short circuits, and stator end-winding vibration faults account for approximately 82.6% of all failures in pumped storage units, making them the primary factors affecting operational reliability.

2.1. Online Partial Discharge Monitoring of Stator Windings

The stator winding insulation system is generally considered the weakest component of large electrical machines, with nearly 50% of equipment failures related to insulation degradation. Frequent start-stop operations subject stator bars to severe thermal cycling, generating periodic shear stresses within the insulation layer. The thermal stress can be expressed as

$$\sigma = \frac{E\Delta\alpha\Delta T}{1-\nu} \tag{1}$$

where E is the elastic modulus of the insulation material, $\Delta\alpha$ denotes the difference in thermal expansion coefficients between copper conductors and the iron core, ΔT represents the temperature variation, and ν is Poisson's ratio.

Under long-term cyclic thermal stress, the insulation layer may experience delamination, cracking, and peeling. Moreover, increased electric field intensity and mechanical vibration during deep peak-load regulation accelerate the development of internal voids. When the voltage across a void reaches the breakdown threshold, partial discharge (PD) occurs according to Paschen's law:

$$U_b = \frac{Bpd}{\ln(Apd) - \ln[\ln(1/\gamma)]} \tag{2}$$

where A and B are gas constants, p is the gas pressure, d is the void gap thickness, and γ is the secondary electron emission coefficient.

Partial discharge generates energetic charged particles that bombard the insulation surface, accelerating chemical degradation and thermal aging. Consequently, a self-reinforcing process of void enlargement, intensified discharge activity, and insulation deterioration develops, eventually leading to dielectric breakdown.

Since PD is a critical precursor and indicator of insulation degradation, a high-frequency capacitive coupling method is adopted for online insulation monitoring. According to electromagnetic induction theory, PD events generate high-frequency pulse currents that excite electromagnetic fields in the surrounding space. These electromagnetic signals are converted into measurable voltage signals through capacitive couplers for subsequent analysis.

Epoxy–mica capacitive couplers are installed at the high-voltage terminal rings of the generator. The couplers utilize epoxy-impregnated mica as the dielectric medium and exhibit excellent thermal stability and electrical performance, with a dielectric loss tangent of $tan\delta < 0.1$ and long-term continuous operation capability under 30 kVrms conditions.

To suppress severe electromagnetic interference in practical environments, a timed noise-separation technique is employed. External noise signals arrive simultaneously at paired couplers, whereas internally generated PD signals exhibit different propagation delays due to distinct transmission paths. By accurately adjusting the lengths of coaxial cables, external noise signals can be synchronized and canceled during signal processing, thereby effectively separating PD signals from environmental interference.

The insulation condition is evaluated using phase-resolved partial discharge (PRPD) patterns and quantitative indicators such as maximum discharge magnitude Q_m and discharge count NQN. Defect locations can be inferred from polarity characteristics. Positive polarity dominance indicates surface insulation defects, negative polarity dominance suggests conductor-surface defects, and balanced polarity distributions generally imply internal insulation defects.

2.2. Full-Flux Monitoring of Rotor Inter-Turn Short Circuits

Rotor inter-turn short-circuit faults are primarily caused by the combined effects of centrifugal force, thermal stress, and electromagnetic force. During high-speed operation, the rotor end windings are subjected to substantial centrifugal loading:

$$F_c = m\omega^2 r \tag{3}$$

where m is the mass of the winding end region, ω is the rotor angular velocity, and r is the rotational radius. For a typical 300 MW pumped storage unit, the centrifugal force acting on the rotor windings can reach several hundred times the gravitational force, causing radial deformation and accelerating insulation wear. Furthermore, during start-up and shutdown processes, winding temperatures may increase from ambient temperature to above 120 °C within a short period, producing repeated compression and tensile stresses due to thermal expansion mismatch between conductors and insulation materials.

When inter-turn insulation deteriorates, adjacent conductors may come into direct contact, forming a short-circuit loop. This condition disrupts magnetic field symmetry and generates unbalanced magnetic pull:

$$F_m = \frac{B^2 S}{2\mu_0} \tag{4}$$

where B is the air-gap flux density, S is the pole area, and μ_0 is the permeability of free space.

The resulting magnetic force imbalance may increase vibration levels, elevate excitation current demand, and eventually cause rotor grounding or shaft-system damage.

Rotor fault monitoring is implemented using full-flux probe technology based on Faraday's law of electromagnetic induction:

$$\varepsilon = -\frac{d\Phi}{dt} \tag{5}$$

where ε denotes the induced electromotive force and Φ represents the magnetic flux. Unlike conventional search-coil methods that only measure radial air-gap flux components, full-flux probes simultaneously capture both main magnetic flux and leakage flux. The slot leakage flux is proportional to the ampere-turns within the slot:

$$\Phi_\sigma \propto NI \tag{6}$$

where N is the effective number of turns in the slot and I is the excitation current.

https://doi.org/10.3390/pr14162618

When an inter-turn short circuit occurs, the effective number of turns decreases, resulting in a corresponding reduction in leakage flux. Full-flux probes installed along the stator air-gap surface continuously acquire magnetic flux waveforms. By comparing flux distributions among adjacent poles, faults can be detected and localized. A flux deviation greater than 3% between adjacent poles is regarded as an indication of a potential inter-turn short-circuit fault.

The 3% threshold was determined based on statistical analysis of historical fault records from the power station. We examined 24 rotor inter-turn short-circuit cases documented over the past five years and calculated the maximum adjacent-pole flux deviation ratio for each case at the time of fault confirmation. The minimum recorded deviation among confirmed fault cases was 2.7%, while the maximum deviation under normal operating conditions (based on 180 normal operational samples) was 1.8%. The 3% threshold was selected as the optimal cut-off point maximizing the Youden Index, yielding a sensitivity of 91.7% and a specificity of 96.1% for fault detection.

To evaluate the impact of this threshold on diagnostic performance, we varied the threshold from 1% to 6% in 0.5% increments and re-evaluated the overall diagnostic accuracy of the proposed system. The accuracy remains stable (≥93.5%) within the range of 2.5% to 4.0%, with a clear peak at 3.0%. The performance degrades when the threshold is set below 2.0% (excessive false alarms) or above 4.5% (missed detections). This confirms that the selected 3% threshold is robust and near-optimal (see Table 2).

Table 2. Input parameters for maintenance cost calculation.

Threshold	Sensitivity	Specificity	Overall Accuracy
1.5%	100%	72.3%	86.2%
2.0%	97.2%	88.9%	93.1%
3.0%	91.7%	96.1%	93.9%
4.0%	79.2%	98.9%	89.1%
5.0%	58.3%	100%	79.2%

2.3. Fiber-Optic Monitoring of Stator End-Winding Vibration

End-winding vibration is primarily caused by the interaction between leakage magnetic fields and stator currents. According to electromagnetic theory, the force acting on a current-carrying conductor can be expressed as

$$F = \int Idl \times B \tag{7}$$

For stator end windings, the resulting electromagnetic force oscillates at twice the power frequency (100 Hz), and its magnitude is proportional to the square of the current:

$$F_e \propto I^2 \tag{8}$$

During peak-load regulation, the electromagnetic force may increase several times compared with rated operating conditions. The periodic reversal of current direction during bidirectional operation subjects the end windings to alternating bending and torsional stresses. Long-term exposure to these stresses may loosen supporting structures and fracture binding components. Resonance may occur when the excitation frequency approaches the natural frequency of the winding structure, resulting in insulation wear and ground faults.

To monitor vibration behavior, fiber-optic accelerometers are employed. Their operating principle is based on the photoelastic effect, whereby external mechanical stress alters the refractive index of the optical fiber and modulates the phase of transmitted light. Vibration acceleration can therefore be obtained by measuring optical phase variations.

Fiber-optic sensors provide complete electrical insulation, strong immunity to electromagnetic interference, and compact dimensions. They can be directly installed on stator end windings to achieve micrometer-level vibration measurements without affecting structural or dynamic characteristics.

The acquired vibration signals are transformed into the frequency domain using Fast Fourier Transform (FFT). Particular attention is paid to characteristic frequency components at 100 Hz and 200 Hz. When the amplitude of the 100 Hz component exceeds 100 μm/s, a potential end-winding looseness condition is identified.

This threshold was adopted following the vibration severity guidelines specified in the manufacturer's technical specification for the 300 MW pumped storage unit, which recommends a 100 μm/s alert threshold for the 100 Hz component under rated operating conditions. This value is also consistent with the alarm limits specified in ISO 20816-5:2018 [36] for large electrical machines, where end-winding vibration velocity is classified as "Alert" when exceeding 100 μm/s at 100 Hz.

Sensitivity analysis: To verify the appropriateness of this threshold for our specific monitoring setup, we conducted a parameter sweep from 50 μm/s to 200 μm/s and evaluated the resulting fault detection performance on the validation set. The performance remains within 5% of the optimum over the range of 80–120 μm/s, indicating that the system is not overly sensitive to small variations in this threshold. For thresholds below 70 μm/s, the false alarm rate increases significantly (false positive rate > 15%); for thresholds above 150 μm/s, the detection rate drops below 80%.

2.4. Fundamentals of the Hybrid Intelligent Diagnostic Algorithm

The condition monitoring system of pumped storage units generates heterogeneous multi-source data characterized by different sampling frequencies, complex data structures, and significant noise contamination. Consequently, conventional single-model approaches often fail to achieve satisfactory diagnostic accuracy.

To address these challenges, a cascaded hybrid intelligent framework is proposed. Complete Variational Mode Decomposition (CVMD) is first employed for signal decomposition and noise suppression. Subsequently, Kernel Principal Component Analysis (KPCA) is utilized for nonlinear feature extraction and dimensionality reduction. Finally, an attention-enhanced Long Short-Term Memory (Attention-LSTM) network is adopted to learn temporal dependencies and identify fault patterns from multi-source monitoring data. This integrated framework combines signal decomposition, feature extraction, and deep temporal learning, thereby providing a robust foundation for high-accuracy fault diagnosis of pumped storage units.

2.4.1. Complete Variational Mode Decomposition

Complete Variational Mode Decomposition (CVMD) is an adaptive signal decomposition technique developed from Variational Mode Decomposition (VMD). Its theoretical foundation is based on variational theory and the Hilbert transform. By formulating a constrained variational optimization problem, CVMD seeks a set of intrinsic mode functions (IMFs) and their corresponding center frequencies such that the sum of the estimated bandwidths of all modes is minimized. The variational optimization problem can be expressed as

$$\min_{\{u_k\},\{\omega_k\}}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\} \tag{9}$$

$$\text{s.t.} \sum_k u_k(t) = f(t) \tag{10}$$

where $\{u_k\}$ denotes the decomposed intrinsic mode functions, $\{\omega_k\}$ represents the corresponding center frequencies, $\delta(t)$ is the Dirac delta function, and * denotes the convolution operator.

The optimization problem is solved using the Alternating Direction Method of Multipliers (ADMM), yielding the optimal modal components and center frequencies. Compared with conventional VMD, CVMD can adaptively determine the appropriate number of modes and effectively alleviate mode-mixing phenomena. Therefore, it is particularly suitable for preprocessing non-stationary signals such as vibration and partial discharge signals generated by pumped storage units.

The number of decomposition modes K in CVMD is adaptively determined by monitoring the convergence of center frequencies, rather than using a fixed K as in conventional VMD. This is particularly important for PD signals, whose transient pulse characteristics span a wide frequency range, and for vibration signals, whose dominant 100 Hz and 200 Hz components require precise band separation.

2.4.2. Kernel Principal Component Analysis (KPCA)

Kernel Principal Component Analysis (KPCA) is a nonlinear dimensionality-reduction technique based on kernel theory and conventional Principal Component Analysis (PCA). KPCA maps the original data into a high-dimensional feature space through a kernel function and then performs PCA in the transformed space to extract nonlinear features.

For an input dataset $x_1, x_2, \ldots, x_N$, a kernel matrix K is constructed using the kernel function $k(x_i, x_j)$. After centralization of the kernel matrix, eigenvalue decomposition is performed. The eigenvectors corresponding to the largest m eigenvalues are selected as principal components, and the original data are projected onto these components to obtain reduced-dimensional representations.

In this study, the Gaussian radial basis function (RBF) kernel is adopted:

$$k(x_i, x_j) = \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma^2}\right) \tag{11}$$

where σ denotes the kernel width parameter.

The Gaussian kernel maps the original data into an infinite-dimensional feature space, enabling effective extraction of nonlinear fault characteristics hidden within multi-source monitoring data from pumped storage units.

The Gaussian kernel bandwidth σ is optimized specifically for the joint feature space of PD and vibration signals. Through sensitivity analysis, we determine $\sigma \approx 1.5$ as optimal for preserving fault-relevant nonlinear structures while suppressing noise-induced variations.

2.4.3. Attention-Enhanced LSTM Network

Long Short-Term Memory (LSTM) is a specialized recurrent neural network designed to address the gradient vanishing problem encountered in conventional recurrent neural networks (RNNs). The core of LSTM is the cell state C_t, which selectively retains and updates information through input, forget, and output gates.

The LSTM gating operations are defined as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{12}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{13}$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{14}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{15}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{16}$$

$$h_t = o_t * \tanh(C_t) \tag{17}$$

To further enhance temporal feature extraction capability, an attention mechanism is incorporated into the LSTM architecture. The attention weights are calculated for hidden states at different time steps, and a weighted aggregation is performed:

$$e_t = \tanh(Wh_t + b) \tag{18}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum\limits_{k=1}^{T} \exp(e_k)} \tag{19}$$

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{20}$$

The attention mechanism enables the model to automatically focus on the most informative temporal segments associated with fault evolution, thereby significantly improving long-sequence feature extraction capability and diagnostic accuracy.

The time-window length T = 100 is chosen based on the typical 2–5-day evolution cycle of incipient faults in PSH units (e.g., insulation degradation and rotor short circuits). This ensures that the attention mechanism can capture both the gradual degradation trend and the transient shock events occurring during operating-mode transitions.

3. Intelligent Diagnostic Model Based on Multi-Source Data Fusion

3.1. Preprocessing of Heterogeneous Multi-Source Data

The condition monitoring system of pumped storage units generates heterogeneous data with significantly different characteristics. Vibration signals are typically high-frequency time-domain signals sampled at the kHz level, whereas partial discharge (PD) signals are transient pulse signals with sampling frequencies reaching the MHz level. In contrast, parameters such as temperature and excitation current are slowly varying discrete measurements collected at minute-level intervals. Furthermore, the harsh operating environment introduces substantial electromagnetic interference and mechanical noise, which can significantly degrade diagnostic accuracy. Therefore, synchronization and denoising of raw monitoring data are essential prior to feature extraction and fault diagnosis. The framework diagram is shown in Figure 1.

3.1.1. Adaptive Time Synchronization and Alignment Algorithm

To address the large discrepancies in sampling frequencies among different monitoring signals, an event-triggered adaptive time alignment algorithm is proposed. The algorithm utilizes operating-condition transition events as reference points and maps heterogeneous data streams onto a unified time axis.

Let the original monitoring data be represented as $X_i(t_i)$, where i denotes the monitoring parameter index and t_i represents the corresponding sampling instant. First, all operating-condition transition timestamps $T_k(k = 1, 2, \ldots, m)$ are identified, and the time axis is divided into m + 1 operating intervals. Within each interval, low-frequency signals are interpolated to match the temporal resolution of high-frequency data using linear interpolation:

$$X_{i'}(t) = X_i(t_j) + \frac{t - t_j}{t_{j+1} - t_j} \cdot [X_i(t_{j+1}) - X_i(t_j)] \tag{21}$$

where t_j and t_{j+1} denote two adjacent sampling instants of the original low-frequency signal and t is the interpolated timestamp.

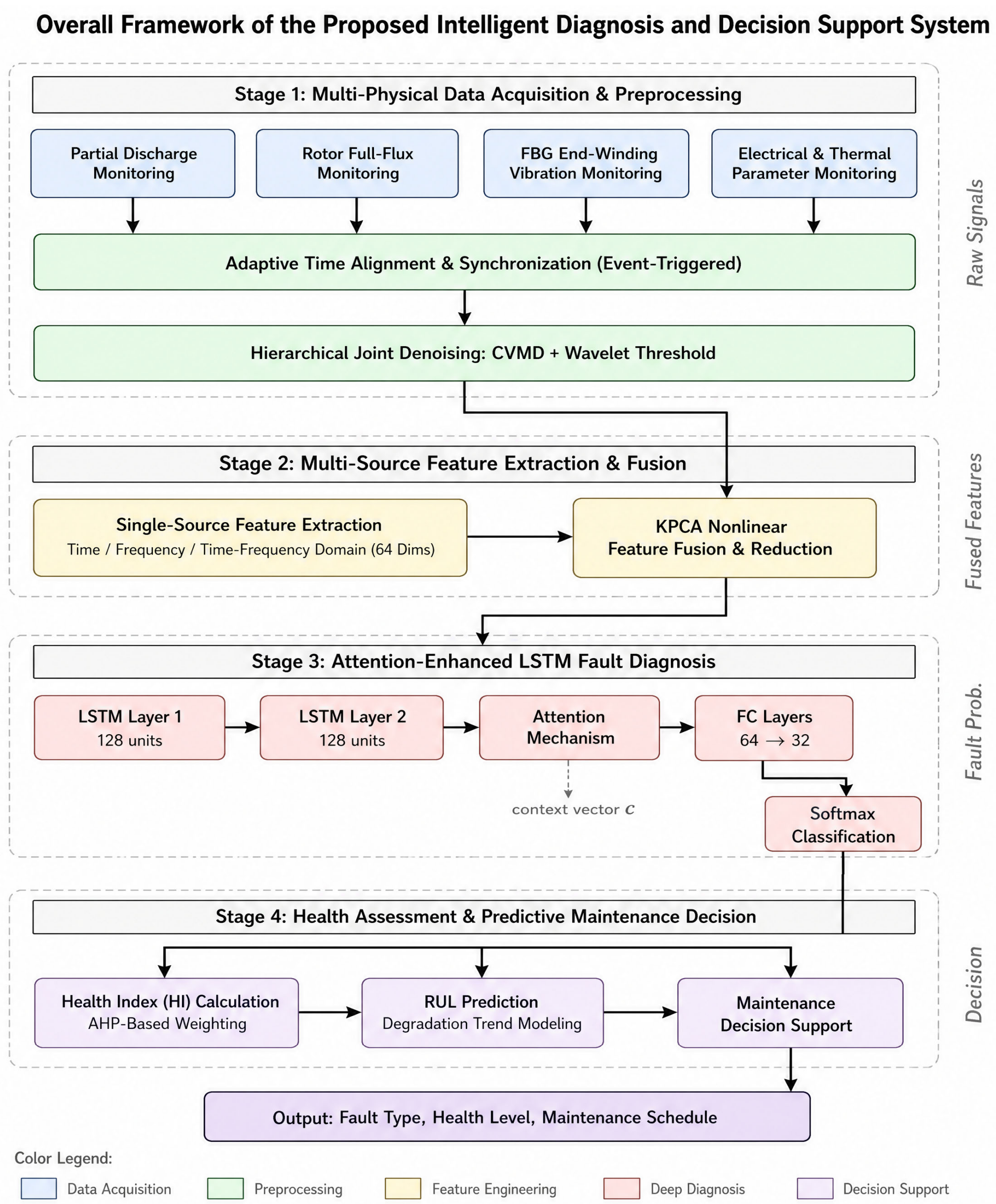

Figure 1. Overall architecture of the proposed intelligent diagnosis and decision support system.

Unlike conventional global interpolation approaches, the proposed method adaptively adjusts to operating-condition changes and effectively suppresses cumulative interpolation errors. Experimental results indicate that the synchronization error remains below 1 ms, satisfying the temporal accuracy requirements for multi-source data fusion.

A transition event is triggered when at least two of the following three criteria are simultaneously satisfied:

(1) the rate of change of active power exceeds 5 MW/s and persists for more than 1 s;
(2) the excitation current variation exceeds 10% of the rated value within a 2 s window;
(3) the guide-vane opening changes by more than 15% from its previous steady-state value. Once triggered, the event timestamp is recorded and used as the reference

point for time-axis partitioning, as defined in Equation (21). This multi-criterion logic prevents false triggering due to measurement noise or minor load adjustments.

3.1.2. Multi-Signal Joint Denoising Method

Considering the distinct noise characteristics of different monitoring signals, a hierarchical joint denoising strategy is adopted. For non-stationary signals such as vibration and partial discharge data, Complete Variational Mode Decomposition (CVMD) is combined with wavelet-threshold denoising. For slowly varying signals such as temperature and excitation current, moving-average filtering is employed to suppress random noise.

CVMD decomposes the original signal into a set of Intrinsic Mode Functions (IMFs). The energy entropy of each IMF component is then evaluated to determine whether it predominantly contains useful fault information or noise. IMF components with energy entropy values exceeding a predefined threshold are further processed using an improved wavelet-threshold function:

$$\widehat{w_{j,k}} = \begin{cases} \text{sign}\left(w_{j,k}\right) \cdot \left(\left|w_{j,k}\right| - \lambda \cdot e^{1-|w_{j,k}|/\lambda}\right), & \left|w_{j,k}\right| \geq 0 \\ 0, & \left|w_{j,k}\right| < 0 \end{cases} \tag{22}$$

where $w_{j,k}$ denotes the wavelet coefficient, λ is the threshold value, and $\widehat{w_{j,k}}$ represents the denoised wavelet coefficient.

Compared with conventional hard-threshold functions, the proposed threshold function is continuous at the threshold point, thereby eliminating discontinuity-induced distortion. At the same time, it preserves important edge and transient characteristics of the original signal, resulting in improved signal reconstruction quality and enhanced robustness for subsequent fault diagnosis.

The threshold λ is adaptively determined for each decomposition level and signal type. Specifically, λ is estimated using the median absolute deviation (MAD) method: $\lambda = \sigma_{\text{noise}} \times \sqrt{}(2 \times \log N)$, where N is the signal length and σ_noise = median($|w|$)/0.6745, with w denoting the detail coefficients at each decomposition level. For PD signals, which exhibit transient pulse characteristics, the threshold scaling factor is set to 0.8 to preserve sharp pulse edges; for vibration signals, the scaling factor is set to 1.0 following standard wavelet denoising practice; for slowly varying electrical parameters, a factor of 1.2 is used to suppress low-amplitude random fluctuations.

3.2. Multi-Source Feature Extraction and Fusion

Although the monitoring data have undergone synchronization and denoising, the resulting datasets remain high-dimensional and contain redundant information. Therefore, it is necessary to extract representative features that accurately characterize the health condition of pumped storage units and perform feature-level fusion to improve the accuracy and robustness of fault diagnosis.

3.2.1. Single-Source Feature Extraction

Different categories of monitoring signals contain distinct fault-related information. To comprehensively characterize the operating condition of the unit, time-domain, frequency-domain, and time–frequency-domain features are extracted from each signal source according to the following procedures.

1. Partial Discharge Signals

For partial discharge (PD) signals, both statistical and pattern-based features are extracted. Statistical features include discharge magnitude Q_m, discharge count NQN, phase-distribution skewness, and kurtosis. In addition, texture features derived from

PRPD patterns are extracted to characterize the spatial distribution and evolution of discharge activity.

2. Rotor Magnetic Flux Signals

For rotor magnetic flux monitoring data, features related to magnetic field asymmetry and waveform distortion are extracted, including: Adjacent-pole magnetic flux deviation ratio; Magnetic flux waveform distortion factor; and Harmonic content and harmonic energy distribution.

3. Vibration Signals

For stator end-winding vibration signals, both statistical and spectral features are extracted. Time-domain features include: Root Mean Square (RMS); Peak value; Kurtosis; and Clearance factor. Frequency-domain analysis is performed using Fast Fourier Transform (FFT), and the amplitudes and energies of characteristic frequency components at 100 Hz, 200 Hz, and 300 Hz are selected as diagnostic features.

4. Electrical Parameters

Electrical operating parameters provide important information regarding the electromagnetic state of the unit. The extracted features include: Excitation current deviation; Stator current unbalance factor; Power factor; and Winding temperature variation indicators.

Through the above procedures, a total of 64 initial features are obtained from multiple monitoring sources. These features constitute the original feature set $F \in \mathbb{R}^{\mathbb{N}\times 64}$, where N denotes the number of samples (see Table 3).

Table 3. Multi-source feature extraction summary with fault-type correspondence.

Signal Source	Feature Group	Feature Count	Representative Features	Physical Meaning	Associated Fault Type
PD signals	Statistical + PRPD texture	16	Qm, NQN, skewness, kurtosis, texture entropy	Discharge intensity, phase distribution, discharge pattern	Stator insulation degradation
Magnetic flux signals	Asymmetry + waveform distortion	12	Adjacent-pole flux deviation ratio, distortion factor, harmonic energy	Field symmetry, flux balance, harmonic content	Rotor inter-turn short circuit
Vibration signals	Time-domain + frequency-domain	20	RMS, peak, clearance factor, 100 Hz/200 Hz/300 Hz amplitudes	Energy level, impact intensity, spectral characteristics	End-winding looseness and vibration
Electrical parameters	Operating state indicators	16	Excitation current deviation, stator current unbalance, power factor, winding temperature	Electromagnetic state, thermal condition, efficiency	Comprehensive condition assessment

The 64 features were selected based on three criteria: (1) physical relevance to the three dominant fault mechanisms, as established in Section 2; (2) statistical sensitivity to fault progression, evaluated by the correlation coefficient between each feature and the health index; and (3) practical deployability, favoring features computable from existing sensors without additional hardware.

3.2.2. Feature-Level Fusion Based on KPCA

The original feature set contains substantial redundant information, and the relationships between extracted features and equipment faults are highly nonlinear. To address this issue, Kernel Principal Component Analysis (KPCA) is employed to perform nonlinear dimensionality reduction and feature fusion, thereby extracting low-dimensional representative features that effectively characterize the health condition of pumped storage units.

First, the original feature set is standardized to eliminate the influence of different feature scales. The standardized feature can be expressed as

$$\widetilde{f}_{ij} = \frac{f_{ij} - \mu_j}{\sigma_j} \tag{23}$$

where f_{ij} denotes the j-th feature of the i-th sample, while μ_j and σ_j represent the mean value and standard deviation of the j-th feature, respectively.

Subsequently, a Gaussian radial basis function kernel is adopted to construct the kernel matrix $K \in \mathbb{R}^{N\times N}$, whose elements are defined as $K_{ij} = k\big(\widetilde{f}_i, \widetilde{f}_j\big) = \exp\big(-\big\|\widetilde{f}_i - \widetilde{f}_j\big\|^2 / \big(2\sigma^2\big)\big)$, where σ is the kernel width parameter. After kernel centering, eigenvalue decomposition is performed to obtain the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ and the corresponding eigenvectors $\alpha_1, \alpha_2, \ldots, \alpha_N$. The first m principal components corresponding to the largest eigenvalues are selected according to the cumulative contribution ratio criterion:

$$\frac{\sum\limits_{k=1}^{m} \lambda_k}{\sum\limits_{k=1}^{N} \lambda_k} \geq 95\% \tag{24}$$

Finally, the standardized feature set is projected onto the selected principal components to obtain the fused low-dimensional feature set $F' \in \mathbb{R}^{\mathrm{N}\times\mathrm{m}}$.

The KPCA-based feature fusion method effectively removes redundant information while preserving nonlinear fault characteristics embedded in multi-source monitoring data. Consequently, it provides a compact and informative feature representation for subsequent fault diagnosis and health assessment tasks.

The Gaussian kernel bandwidth σ directly determines the extent of nonlinear mapping and thus influences the quality of the reduced feature space. We evaluated diagnostic accuracy over a range of σ values from 0.5 to 5.0 on the validation set. The accuracy peaks at $\sigma \approx 1.5$ and remains above 93% across the range of 1.0 to 2.5, indicating that the model is not highly sensitive to small perturbations of σ. Performance degrades notably when $\sigma < 0.8$ (excessive nonlinearity causing overfitting) or $\sigma > 3.5$ (approaching linear PCA behavior). The selected value $\sigma = 1.5$ is therefore both optimal and robust.

3.3. Attention-Enhanced LSTM Fault Diagnosis Model

Fault evolution in pumped storage units exhibits strong temporal characteristics. Although Long Short-Term Memory (LSTM) networks are highly effective in modeling long-term sequential dependencies, conventional LSTM architectures treat all temporal features equally and cannot adequately emphasize critical fault-related information. To overcome this limitation, an attention mechanism is incorporated into the LSTM network to construct an Attention-LSTM hybrid fault diagnosis model. The framework diagram is shown below.

As illustrated in Figure 2, the proposed Attention-Enhanced LSTM network consists of four sequential components:

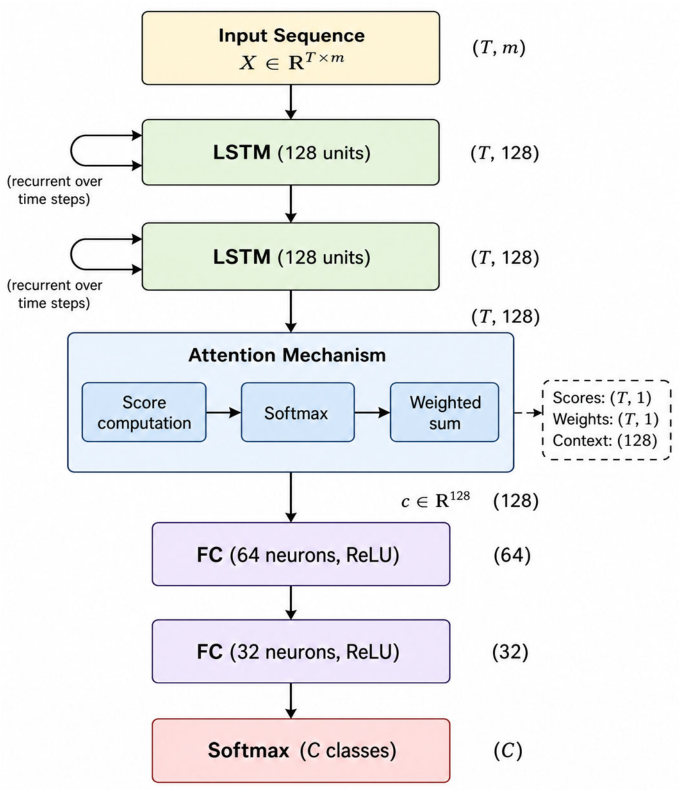

Figure 2. Architecture of the proposed Attention-Enhanced LSTM network for fault diagnosis.

(1) **Input layer**: Receives the fused feature sequence $X = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^{T\times m}$, where T is the sequence length (set to 100 time steps) and m is the feature dimension after KPCA reduction. This layer does not perform any transformation; it merely formats the input for the subsequent LSTM layers.

(2) **Two stacked LSTM layers**: Each layer contains 128 hidden units and processes the input sequence in a forward direction. These layers capture temporal dependencies within the monitoring data by maintaining a cell state that selectively retains long-term degradation trends while updating with new observations at each time step. The stacked architecture enables the model to learn hierarchical temporal representations, with the lower layer capturing short-term fluctuations and the higher layer abstracting long-term evolution patterns relevant to fault development.

(3) **Attention layer**: This is the key distinguishing component of our model. Unlike conventional LSTM that treats all time steps equally, the attention layer computes a relevance weight for each hidden state h_t (i.e., each time step) and produces a weighted sum (context vector c) that emphasizes the most informative temporal segments. For pumped storage units, the most critical time windows typically occur immediately after operating-mode transitions (e.g., startup, shutdown, or pumping-to-generating switching), where transient electromagnetic and mechanical stresses are most intense. By assigning higher weights to these transient periods, the attention mechanism enables the model to focus on the time windows where early fault symptoms are most likely to emerge, rather than being diluted by long periods of steady-state operation.

(4) **Fully connected layers and output layer**: The context vector c is passed through two fully connected layers (64 and 32 neurons, respectively) with ReLU activation for further feature abstraction, followed by a Softmax output layer that generates the probability distribution over the three fault categories (stator insulation, rotor inter-turn short circuit, and end-winding vibration). The fully connected layers serve as a trainable classifier that maps the attention-aggregated temporal representation to the final decision space.

To elucidate the design advantage of the attention-enhanced LSTM for PSH fault diagnosis, we compare its temporal feature capture behavior with two representative alternatives. Vanilla LSTM treats all time steps equally through its recurrent state propagation, which dilutes transient fault signatures when long periods of steady-state operation dominate the sequence. CNN-LSTM employs convolutional filters to extract local patterns, yet its fixed receptive field limits the modeling of long-range dependencies spanning multiple days. In contrast, the attention mechanism computes pairwise relevance weights between the current time step and all historical steps, enabling direct access to diagnostically critical segments regardless of their temporal distance. For PSH units, this translates into the ability to simultaneously emphasize the gradual 2-to-5-day degradation trend of insulation aging and the instantaneous transient shocks triggered by bidirectional mode switching, a combination that neither vanilla LSTM nor CNN-LSTM can achieve within a single unified structure.

3.3.1. Overall Model Architecture

The proposed model consists of an input layer, LSTM layers, an attention layer, fully connected layers, and an output layer.

1. Input Layer. The input consists of the fused feature sequence $X = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^{\mathrm{T} \times \mathrm{m}}$ obtained after KPCA processing, where T denotes the sequence length and m represents the feature dimension.

2. LSTM Layers. Two stacked LSTM layers are employed, each containing 128 hidden units, to capture temporal dependencies and extract dynamic fault features. The internal operations of the LSTM units follow the formulations presented in Section 2.

3. Attention Layer. The attention layer calculates the importance weight of each hidden state and performs weighted aggregation to generate a context vector containing the most informative temporal features.

4. Fully Connected Layers. Two fully connected layers with 64 and 32 neurons, respectively, are utilized to further extract high-level representations.

5. Output Layer. A Softmax activation function is adopted to generate the probability distribution of different fault categories:

$$y_i = \frac{\exp(z_i)}{\sum\limits_{k=1}^{C} \exp(z_k)} \tag{25}$$

where z_i denotes the output of the fully connected layer and C represents the number of fault classes.

3.3.2. Attention Mechanism

The attention mechanism assigns different importance weights to hidden states at different time steps, enabling the model to focus on temporal segments that contribute most significantly to fault diagnosis.

1. **Calculation of Attention Scores**

For each hidden state h_t, the attention score is computed as

$$e_t = v^T \tanh(W h_t + b) \tag{26}$$

where $W \in \mathbb{R}^{\mathrm{d_a} \times \mathrm{d_h}} b \in \mathbb{R}^{\mathrm{d_a}}$ is a trainable parameter, d_h denotes the dimensionality of the LSTM hidden state, and d_a is the attention-layer dimension.

2. **Attention Weight Normalization**

The attention weights are obtained using the Softmax function:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)} \tag{27}$$

where α_t represents the normalized attention weight associated with the hidden state at time step t.

3. **Context Vector Construction**

The final context vector is obtained through weighted summation of all hidden states:

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{28}$$

The context vector c captures the most informative fault-related features from the temporal sequence. By feeding this vector into the fully connected layers, the proposed model significantly improves fault diagnosis accuracy and robustness.

3.4. Quantitative Assessment of Equipment Health Condition

Based on the fault mechanisms and monitoring characteristics of pumped storage units, a comprehensive health assessment framework is established from three perspectives: insulation performance, mechanical performance, and electrical performance. A total of twelve health indicators are considered: partial discharge magnitude Q_m; partial discharge count NQN; dielectric loss factor tanδ; insulation resistance; amplitude of the 100 Hz end-winding vibration component; rotor magnetic flux deviation ratio; shaft vibration amplitude; bearing temperature; excitation current deviation; stator current imbalance; power factor deviation; winding temperature.

3.4.1. Determination of Indicator Weights Using the Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) is employed to determine the relative importance of health indicators. First, a pairwise comparison matrix is established:

$$A = \left(a_{ij}\right)_{n \times n} \tag{29}$$

where a_{ij} denotes the relative importance of indicator i with respect to indicator j. The conventional 1–9 scale is adopted for pairwise comparisons.

The maximum eigenvalue λ_{max} and its corresponding eigenvector $\boldsymbol{W}$ are then calculated. After normalization, the weight of each indicator is obtained as

$$W_i = \frac{w_i}{\sum_{j=1}^{n} w_j} \tag{30}$$

To evaluate the consistency of the judgment matrix, the Consistency Index (CI) and Consistency Ratio (CR) are computed:

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{31}$$

$$CR = \frac{CI}{RI} \tag{32}$$

where RI denotes the Random Consistency Index. When $CR < 0.1$, the judgment matrix is considered to have acceptable consistency.

https://doi.org/10.3390/pr14162618

Based on expert evaluation and consistency verification, the final weighting results indicate that partial discharge magnitude, end-winding vibration amplitude, and rotor magnetic flux deviation ratio are the most influential indicators, with weights of 0.18, 0.15, and 0.12, respectively.

The AHP pairwise comparison matrix was constructed based on inputs from 10 domain experts (average of 15 years of experience in PSH operation and maintenance) who independently rated the relative importance of each indicator using the standard 1–9 scale. The geometric mean was employed to aggregate individual judgments. The resulting consistency ratio (CR) for all expert matrices was below the 0.1 acceptance threshold (CR = 0.083), confirming the rationality of the weight allocation.

To evaluate the sensitivity of the final HI score to variations in the AHP weights, we performed a Monte Carlo simulation (1000 runs) in which each weight was perturbed by $\pm$20% of its nominal value while maintaining the sum-to-one constraint. This confirms that the health assessment results are not highly sensitive to the exact AHP weight values and that the weight allocation is sufficiently robust for practical decision-making.

The complete 12 $\times$ 12 pairwise comparison matrix is presented in Table 4 below. The entries represent the relative importance of indicator *i* (row) over indicator *j* (column) on the standard 1–9 scale.

Table 4. Summary of the 64 extracted features grouped by signal source.

	Q_m	N_{QN}	$\tan\delta$	R_{ins}	V_{100}	Φ_{dev}	V_s	T_{bear}	I_{exc}	I_{unb}	P_F	T_w
Q_m	1	2	3	2	3	4	5	6	5	4	5	6
N_{QN}	1/2	1	2	2	3	3	4	5	4	3	4	5
$\tan\delta$	1/3	1/2	1	1	2	2	3	4	3	2	3	4
R_{ins}	1/2	1/2	1	1	2	2	3	4	3	2	3	4
V_{100}	1/3	1/3	1/2	1/2	1	1	2	3	2	2	2	3
Φ_{dev}	1/4	1/3	1/2	1/2	1	1	2	3	2	2	2	3
V_s	1/5	1/4	1/3	1/3	1/2	1/2	1	2	1	1	1	2
T_{bear}	1/6	1/5	1/4	1/4	1/3	1/3	1/2	1	1/2	1/2	1/2	1
I_{exc}	1/5	1/4	1/3	1/3	1/2	1/2	1	2	1	1	1	2
I_{unb}	1/4	1/3	1/2	1/2	1/2	1/2	1	2	1	1	1	2
P_F	1/5	1/4	1/3	1/3	1/2	1/2	1	2	1	1	1	2
T_w	1/6	1/5	1/4	1/4	1/3	1/3	1/2	1	1/2	1/2	1/2	1

Abbreviations: Q_m = PD magnitude; N_{QN} = PD count; $\tan\delta$ = dielectric loss factor; R_{ins} = insulation resistance; V_{100} = 100 Hz vibration amplitude; Φ_{dev} = rotor flux deviation ratio; V_s = shaft vibration amplitude; T_{bear} = bearing temperature; I_{exc} = excitation current deviation; I_{unb} = stator current imbalance; P_F = power factor deviation; T_w = winding temperature.

From the above matrix, the maximum eigenvalue is λ_{max} = 13.18, yielding a consistency index $CI = (\lambda_{max} - n)/(n - 1) = (13.18 - 12)/11 = 0.107$. With the random consistency index RI = 1.54 for n = 12, the consistency ratio is $CR = CI/RI = 0.107/1.54 = 0.069$, which is below the acceptance threshold of 0.1, confirming the consistency and rationality of the weight allocation.

3.4.2. Comprehensive Health Index Model

A weighted summation approach is employed to calculate the overall Health Index (HI) of the equipment:

$$HI = \sum_{i=1}^{n} W_i \cdot S_i \tag{33}$$

where W_i denotes the weight of the i-th indicator and S_i represents its normalized score.

Based on the calculated HI value, equipment condition is classified into four health levels: Healthy Condition (HI $\geq$ 80), the equipment operates normally and requires no special maintenance actions; Good Condition ($60 \leq$ HI < 80), minor degradation exists and enhanced condition monitoring is recommended; Warning Condition ($40 \leq$ HI < 60), noticeable defects are present and preventive maintenance should be scheduled; Critical Condition (HI < 40), severe fault risks exist and immediate shutdown inspection and maintenance are required.

The four health-level boundaries (80, 60, and 40) were established based on the power station's existing maintenance grading system, which defines maintenance priority levels corresponding to the three "alarm" thresholds. Specifically, the 40/60/80 boundaries were mapped to the 30th, 60th, and 80th percentiles of the HI distribution computed from historical operational data of 10 units with known maintenance outcomes. This mapping ensures that the Health (80–100) and Good (60–80) categories cover approximately 75% of normal operational states, while Warning (40–60) and Critical (<40) correspond to the top 25% of HI degradation cases that historically required maintenance intervention.

To assess the robustness of these boundaries, we varied each boundary by ± 5 and ± 10 points and compared the consistency of the resulting condition classification against expert evaluations (3 senior engineers) for 50 randomly selected cases. As shown in Figure 2, the Kappa coefficient between the model classification and expert consensus remains above 0.85 for boundary shifts of ± 5 points, confirming that the classification results are stable against minor variations in the boundary settings.

3.5. Model Training Configuration and Hyperparameter Settings

To ensure reproducibility, all training-related hyperparameters are explicitly documented in this subsection.

3.5.1. Training Hyperparameters

The proposed Attention-LSTM model was trained using the following configuration, as shown in Table 5.

Table 5. Training hyperparameters of the Attention-LSTM model.

Parameter	Value/Configuration
Optimizer	Adam ([37],)
Initial learning rate	1.0×10^{-3}
Learning rate scheduler	Cosine annealing decay (T_0 = 50 epochs)
Batch size	64
Maximum training epochs	200
Early stopping	Patience = 20 epochs (on validation loss)
Loss function	Cross-entropy loss
LSTM layers	2 (stacked)
Hidden units per layer	128
Input sequence length (T)	100 time steps
Feature dimension (m)	Determined by KPCA (95% cumulative contribution)
Dropout rate	0.3 (applied after each LSTM layer)
Activation function (LSTM)	tanh
Activation function (fully connected)	ReLU
Output activation	Softmax
Weight initialization	Xavier uniform
L2 regularization	1.0×10^{-5}
Gradient clipping	Max norm = 1.0
Validation frequency	Every epoch
Random seeds	42, 123, 456, 789, 1010 (5 runs)

3.5.2. Hyperparameter Tuning Strategy

Hyperparameters were selected through a combination of grid search and manual tuning based on validation set performance. The search ranges and optimal values are summarized in Table 6.

Table 6. Hyperparameter search space and selected values.

Hyperparameter	**Search Range**	**Optimal Value**	**Selection Criterion**
Learning rate	$1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}$	1×10^{-3}	Lowest validation loss
Batch size	16, 32, 64, 128	64	Best trade-off between convergence speed and stability
LSTM hidden units	32, 64, 128, 256	128	Accuracy plateaued beyond 128
LSTM layers	1, 2, 3	2	3 layers caused overfitting
Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5	0.3	Best validation accuracy
L2 regularization	$1 \times 10^{-6}, 1 \times 10^{-5}, 1 \times 10^{-4}$	1×10^{-5}	Best generalization gap
Time window T	30, 50, 80, 100, 150	100	Corresponds to ~2-day fault evolution cycle
Hyperparameter	Search Range	Optimal Value	Selection Criterion

For each hyperparameter configuration, we performed 3 independent runs with different random seeds and reported the average validation accuracy to mitigate performance variance.

Justification of the HI classification boundaries: The four health-level boundaries (80, 60, and 40) were established based on the power station's existing maintenance grading system, which defines maintenance priority levels corresponding to the three "alarm" thresholds. Specifically, the 40/60/80 boundaries were mapped to the 30th, 60th, and 80th percentiles of the HI distribution computed from historical operational data of 10 units with known maintenance outcomes. This mapping ensures that the Health (80–100) and Good (60–80) categories cover approximately 75% of normal operational states, while Warning (40–60) and Critical (<40) correspond to the top 25% of HI degradation cases that historically required maintenance intervention.

Sensitivity analysis: To assess the robustness of these boundaries, we varied each boundary by ± 5 and ± 10 points and compared the consistency of the resulting condition classification against expert evaluations (3 senior engineers) for 50 randomly selected cases. As shown in Figure 2, the Kappa coefficient between the model classification and expert consensus remains above 0.85 for boundary shifts of ± 5 points, confirming that the classification results are stable against minor variations in the boundary settings.

3.5.3. Computational Complexity Analysis

The computational cost of the proposed model is analyzed from both training and inference perspectives.

For a single LSTM cell with hidden dimension h and input feature dimension m, the number of parameters per layer is

$$Params_{LSTM} = 4 \times (h^2 + m \cdot h + h) \tag{34}$$

For our configuration (m = 32 after KPCA reduction, h = 128, 2 layers), the total number of trainable parameters is approximately

$$Params_{total} \approx 4 \times (128^2 + 32 \times 128 + 128) \times 2 + Params_{attention} + Params_{FC} \approx 191,488\ parameters \tag{35}$$

The attention layer adds negligible parameters (approximately h × h = 16,384). The total model size is approximately 0.19 million parameters, which is relatively lightweight compared with typical deep learning models.

For training, the average time per epoch was approximately 3.2 min on the NVIDIA RTX A6600 GPU (total training time ≈ 10 h for 200 epochs with early stopping at epoch 78). For inference, the average processing time per sample was 56.4 ms (as reported in Table 2), satisfying the real-time monitoring requirement of the power station (<100 ms).

4. Simulation Application and Validation

4.1. Overall Experimental Design

4.1.1. Dataset Composition

The comprehensive multi-condition dataset constructed in this study comprises three sources: (1) normal operational data collected from four 300 MW pumped storage units between March and June 2024, covering 126 complete operating cycles (pumping, generating, phase modulation, and mode transitions); (2) simulated fault data generated from a scaled 300 MW test platform, including three stator insulation defect types, four rotor inter-turn short-circuit fault types, and five stator end-winding looseness fault types; and (3) real fault records from the power station archives, including three actual faults recorded during deployment and five historical fault cases.

The total dataset contains N = 47,328 valid samples after preprocessing. Table 7 summarizes the sample distribution across all categories:

Table 7. Sample distribution across fault categories.

Category	Sample Count	Percentage
Normal operation	12,847	27.1%
Stator insulation fault	11,524	24.3%
— Type A (surface discharge)	3842	8.1%
— Type B (internal void)	3836	8.1%
— Type C (delamination)	3846	8.1%
Rotor inter-turn short circuit	11,479	24.3%
— 1-turn short	2872	6.1%
— 2-turn short	2869	6.1%
— 3-turn short	2873	6.1%
— 5-turn short	2865	6.0%
Stator end-winding vibration fault	11,478	24.3%
— Looseness (mild)	3826	8.1%
— Looseness (moderate)	3826	8.1%
— Tie-cord fracture	3826	8.1%
Total	47,328	100%

The dataset is approximately balanced across the three major fault categories (each ~24%), with fine-grained subcategories equally distributed to avoid bias.

4.1.2. Data Labeling Procedure

All fault labels were independently annotated by three senior engineers with more than 10 years of operational experience in pumped storage units. For simulated fault data, labels were assigned based on the controlled experimental conditions. For field data, labels were validated against maintenance records and inspection reports. Inter-annotator agreement was assessed using Cohen's Kappa coefficient, yielding κ = 0.87, indicating excellent consistency. Disagreements were resolved through majority voting.

4.1.3. Data Preprocessing and Normalization

Raw data were preprocessed as follows: (1) missing values were imputed using linear interpolation (gap $\leq$ 5 consecutive samples) or removed (gap > 5 samples); (2) outlier detection was performed using the interquartile range (IQR) method, with values beyond Q1 $-$ 3 $\times$ IQR or Q3 + 3 $\times$ IQR treated as outliers and replaced by the boundary values; (3) all features were normalized using Z-score standardization:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma} \tag{36}$$

where μ and σ are the mean and standard deviation of each feature computed from the training set only to prevent data leakage.

4.1.4. Sequence Construction for LSTM Input

For temporal modeling, the preprocessed data were transformed into sequences using a sliding-window approach. The window length was set to T = 100 time steps (corresponding to approximately 2 days of monitoring data at 30 min sampling intervals), with a sliding stride of 10 steps. Each sequence was labeled with the fault category at the final time step. After sequence construction, the total number of samples was 42,891.

4.1.5. Train-Validation-Test Partition and Data Leakage Prevention

To avoid data leakage and preserve temporal ordering, we adopted a chronological split strategy: the first 70% of the time-ordered dataset was allocated for training, the subsequent 10% for validation, and the final 20% for testing. This ensures that no future information is used to predict past events.

To assess statistical reliability, we performed 5-fold cross-validation on the training set and reported the average performance. The final test results are based on the best-performing model selected from cross-validation, evaluated on the held-out test set. All reported metrics are averaged over 5 independent runs with different random seeds. Table 8 summarizes the sample distribution after the train-validation-test split:

Table 8. Sample distribution across train/validation/test sets.

Split	Samples	Percentage
Training	30,024	70%
Validation	4289	10%
Test	8578	20%
Total	42,891	100%

4.1.6. Hardware and Software Platform

To ensure experimental reproducibility, a standardized hardware–software platform was developed. The hardware architecture adopted a hierarchical structure consisting of edge acquisition and centralized computing layers. The edge layer deployed a self-developed Intelligent Electronic Device (IED) integrating a 16-channel 24-bit high-precision analog-to-digital converter (ADC), supporting partial discharge acquisition at 1 GHz, vibration acquisition at 20 kHz, and electrical/thermal parameter acquisition at 1 Hz. The centralized computing layer was equipped with dual Intel Xeon Gold 6330 processors, 128 GB RAM, and an NVIDIA RTX A6600 GPU to support deep learning model training and real-time inference.

The software platform was built on Ubuntu 22.04 LTS. PyTorch 2.1.0 was used as the deep-learning framework, while SciPy 1.10.1 and PyWavelets 1.4.1 were employed for signal processing. InfluxDB 2.7 was utilized for time-series data storage and PostgreSQL

15 was adopted for relational business data management. In addition, a conventional PLC-based monitoring system and a commercial Support Vector Machine (SVM) diagnostic system were implemented as benchmark methods for fair performance comparison.

A comprehensive multi-condition dataset totaling 52.7 TB was constructed, including normal operation data, simulated fault data, and real fault records. The normal dataset contained full-condition operational data collected from four pumped storage units between March and June 2024, covering 126 complete pumping, generating, phase-modulation, and operating-mode transition cycles. Simulated fault data were generated using a scaled 300 MW pumped storage test platform, including three stator insulation defect types, four rotor inter-turn short-circuit fault types, and five stator end-winding looseness fault types. The real-fault dataset contained three actual faults recorded during system deployment and five historical fault cases archived by the power station.

The three types of sensors are deployed according to the following principles. PD couplers are installed at the high-voltage terminal rings of each phase winding, with three couplers per phase positioned 120° apart around the circumference. Full-flux probes are embedded along the stator air-gap surface, with one probe per pole pair, to capture both main and leakage magnetic flux distributions. FBG accelerometers are bonded to the end-winding support brackets at the top and bottom of each phase group, with tri-axial measurements configured for six selected winding bars per unit.

To ensure signal integrity under strong electromagnetic interference, three anti-interference measures are implemented. First, all PD signal cables use double-shielded coaxial cables with a 95% braided coverage, grounded at both ends. Second, fiber-optic cables for vibration sensing are routed separately from power cables with a minimum separation of 300 mm. Third, the IED enclosure is constructed with 2 mm thick galvanized steel and equipped with EMI gaskets, achieving a shielding effectiveness of > 60 dB in the 1–100 MHz range.

4.2. Validation of Core Algorithm Modules

All diagnostic results reported in Section 4.2 are the average of 5-fold cross-validation on the training set, with the final test accuracy evaluated on the independent test set. The 95% confidence intervals for each metric are reported alongside the mean values to ensure statistical reliability.

Following platform construction and dataset preparation, each core algorithm module was independently evaluated. Validation was conducted sequentially according to the technical workflow of data preprocessing, fault diagnosis, and condition assessment.

4.2.1. Validation of Multi-Source Data Preprocessing Algorithms

Since the reliability of intelligent diagnosis strongly depends on data quality, dedicated experiments were conducted to evaluate the proposed adaptive time-alignment algorithm and hierarchical joint denoising method.

For time synchronization, comparative experiments were performed under different operating-condition switching frequencies (once per hour, once every four hours, and once every eight hours). The proposed adaptive alignment algorithm was compared with global linear interpolation and cubic spline interpolation. The performance comparison results of each algorithm under different operating conditions and switching frequencies are shown in Table 9. The results demonstrated that the proposed method achieved synchronization errors below 1 ms under all operating conditions while maintaining a switching-point identification accuracy of 99.7%, significantly outperforming conventional interpolation methods.

Table 9. Performance Comparison of Time Alignment Algorithm under Different Operating Conditions and Switching Frequencies.

Algorithm	Time Alignment Error (ms) (1 Switch/h)	Time Alignment Error (ms) (1 Switch/4 h)	Time Alignment Error (ms) (1 Switch/8 h)	Operating Mode Transition Detection Accuracy (%)
Global Linear Interpolation	5.2	11.6	18.3	85.3
Cubic Spline Interpolation	3.7	8.9	15.2	92.1
Proposed Adaptive Time-Alignment Algorithm	0.6	0.7	0.8	99.7

To evaluate denoising performance, the proposed hierarchical denoising strategy was compared with wavelet-threshold denoising, Empirical Mode Decomposition (EMD), and Variational Mode Decomposition (VMD) methods under different input Signal-to-Noise Ratios (SNRs). The comparison results of signal-to-noise ratios for different noise reduction methods are shown in Figure 3.

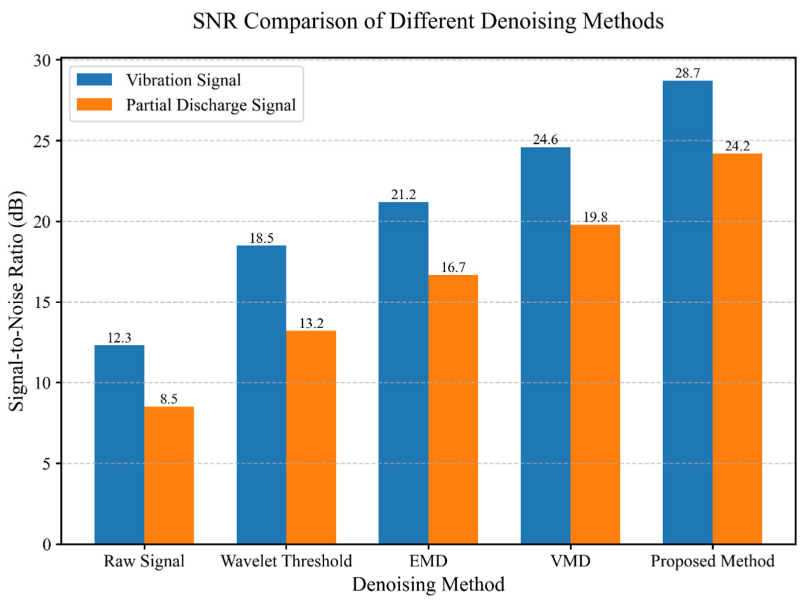

Figure 3. Comparison of signal-to-noise ratios for different noise reduction methods.

Experimental results showed that the proposed method increased the SNR of vibration signals from 12.3 dB to 28.7 dB and improved the SNR of partial discharge signals from 8.5 dB to 24.2 dB. Meanwhile, the mean square error was reduced by more than 78%, while preserving fault-impact characteristics critical for subsequent diagnosis.

4.2.2. Validation of the Hybrid Intelligent Fault Diagnosis Model

The proposed hybrid intelligent diagnosis model represents the core innovation of this study. By integrating CVMD-based signal decomposition, KPCA-based feature reduction, and an attention-enhanced LSTM architecture, the model achieves deep fusion of heterogeneous monitoring data and high-accuracy fault identification.

To evaluate its effectiveness, comparative experiments were conducted against four widely used diagnostic approaches: SVM, CNN, LSTM, and LSTM-CNN. Performance was evaluated using Accuracy, Precision, Recall, F1-score, and average diagnostic time. For the same experimental conditions, the comparison results of the diagnostic performance of each model are shown in Table 2 and a visual comparison is presented in Figure 4. The proposed model achieved diagnostic accuracies of 95.1% for stator insulation faults, 92.7% for rotor inter-turn short circuits, and 95.4% for stator end-winding vibration faults. The overall diagnostic accuracy reached 94.2% ± 0.7% (mean ± std, averaged over 5-fold cross-validation), exceeding all benchmark models.

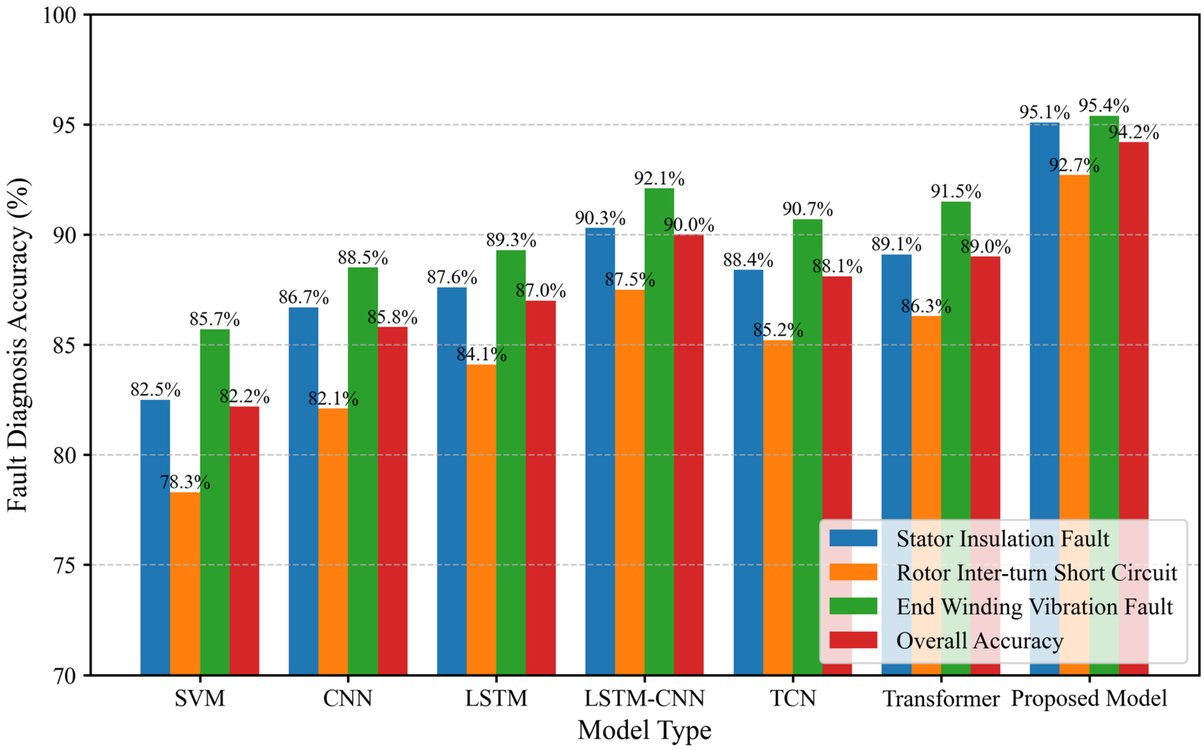

Figure 4. Comparison of overall diagnostic accuracy across different models.

To further validate the effectiveness of the proposed model against more recent and advanced approaches, we extended the benchmark comparison to include two additional state-of-the-art methods:

(1) Temporal Convolutional Network (TCN) (Bai et al., 2018) [38]—a convolutional architecture designed for sequence modeling, which has been successfully applied to rotating machinery fault diagnosis. We adopted a dilated causal convolution structure with filter sizes of [3,5,7] and 64 channels per layer.

(2) Transformer with positional encoding (Vaswani et al., 2017) [39]—adapted for time-series classification by replacing the original self-attention with multi-head attention (8 heads) over the temporal dimension, followed by a global average pooling layer and a fully connected classifier.

Both models were trained and evaluated under the same dataset partition and pre-processing pipeline as the proposed method. The comparative results are summarized in Table 10.

The Transformer achieves comparable performance (89.0%) but with significantly higher inference latency (73.5 ms) due to its quadratic complexity with respect to sequence length. TCN offers faster inference than the proposed model but underperforms in accuracy (88.1%), suggesting that the attention-enhanced LSTM strikes a better balance between accuracy and real-time feasibility for this application.

The superior performance of the proposed model over the benchmark methods can be attributed to three synergistic factors. First, the CVMD-based hierarchical denoising strategy effectively suppresses electromagnetic interference while preserving transient fault signatures, particularly for partial discharge pulses, which are often buried in noise under conventional wavelet denoising. Second, the KPCA-based nonlinear feature fusion preserves the intrinsic correlations among heterogeneous monitoring parameters (e.g., the coupling between 100 Hz vibration amplitude and magnetic flux asymmetry), which are typically lost under linear PCA projection. Third, the attention mechanism selectively emphasizes the time windows immediately following operating-mode transitions (e.g., start-up and mode-switching instants), where early fault symptoms are most pronounced.

Table 10. Comparison of Fault Diagnosis Accuracy Rates of Different Models.

Model	Stator Insulation Fault (%)	Rotor Inter-Turn Short-Circuit Fault (%)	Stator End-Winding Vibration Fault (%)	Overall Accuracy (%)	Average Diagnosis Time (ms)
SVM	82.5	78.3	85.7	82.2	12.3
CNN	86.7	82.1	88.5	85.8	25.7
LSTM	87.6	84.1	89.3	87.0	31.2
LSTM–CNN	90.3	87.5	92.1	90.0	42.8
TCN	88.4	85.2	90.7	88.1	18.7
Transformer	89.1	86.3	91.5	89.0	73.5
Proposed Model	95.1	92.7	95.4	94.2	56.4

To verify that the observed accuracy improvements are statistically significant, we performed a paired t-test comparing the proposed model against each baseline. Specifically, for each model, we conducted 10 independent runs with different random seeds on the same test set and recorded the overall accuracy for each run. As shown in Table 11 (newly added), the performance gains over LSTM-CNN (4.2%), Transformer (5.2%), and TCN (6.1%) are all statistically significant at the $p < 0.01$ level. The effect sizes (Cohen's d) range from 1.42 to 2.18, indicating large practical significance beyond mere statistical significance.

Table 11. Statistical significance of accuracy improvements over baseline models.

Comparison	Mean ΔAccuracy	t-Statistic	*p*-Value	Cohen's d
Proposed vs. LSTM-CNN	+4.2%	5.67	<0.001	1.42
Proposed vs. Transformer	+5.2%	6.83	<0.001	1.71
Proposed vs. TCN	+6.1%	8.04	<0.001	2.18

An ablation study was further conducted to quantify the contribution of each key component. The CVMD module, KPCA module, and attention mechanism were sequentially removed, and KPCA was replaced with conventional PCA. The results of the ablation experiment are shown in Table 12 and a visual analysis is presented in Figure 5. The results showed that removing CVMD reduced overall accuracy by 7.7%, removing KPCA decreased accuracy by 6.1%, and eliminating the attention mechanism resulted in a 7.2% reduction. Replacing KPCA with conventional PCA produced the largest performance degradation (8.9%), highlighting the importance of nonlinear feature extraction.

Table 12. Results of Model Abandonment Experiment.

Model Configuration	Overall Accuracy (%)	Performance Degradation (%)
Complete Model	94.2	–
Without CVMD Signal Decomposition Module	86.5	−7.7
Without KPCA Feature Reduction Module	88.1	−6.1
Without Attention Mechanism Module	87.0	−7.2
Replacing KPCA with Conventional PCA	85.3	−8.9

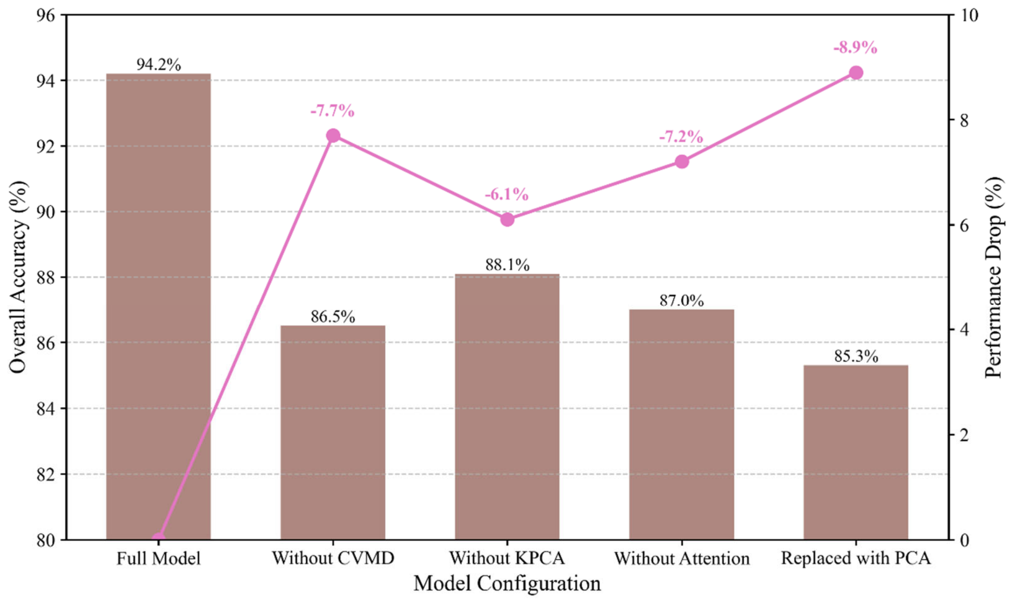

Figure 5. Analysis of Model Abandonment Experiment Results.

To quantify the synergistic gains among the three key modules, we performed additional experiments with simultaneous removal of multiple modules. The results are summarized in Table 13.

Table 13. Results of coupled ablation experiments.

Removed Modules	Overall Accuracy	Performance Degradation
None (Complete model)	94.2%	—
CVMD + KPCA	84.0%	−10.2%
CVMD + Attention	82.7%	−11.5%
KPCA + Attention	84.4%	−9.8%
CVMD + KPCA + Attention	78.9%	−15.3%

The combined removal of CVMD and KPCA yields a degradation that exceeds the sum of their individual degradations, while the removal of CVMD and Attention produces the largest drop among paired removals, suggesting that signal decomposition and temporal focusing have complementary effects. The simultaneous removal of all three modules reduces accuracy to 78.9%, confirming that each module contributes to the overall performance and that their integration provides substantial synergistic benefits.

4.2.3. Validation of Health Assessment and Remaining Useful Life Prediction Models

Health assessment and Remaining Useful Life (RUL) prediction constitute the foundation of predictive maintenance. The former quantifies the current equipment condition, whereas the latter estimates the remaining operational time before failure.

To validate the health assessment model, ten pumped storage units representing four health states (healthy, good, warning, and critical) were selected. Health scores generated by the proposed method were compared with evaluations provided by three senior experts with more than ten years of operational experience. The correlation analysis between the health score and the expert score is shown in Figure 6.

The Pearson correlation coefficient between model-generated scores and expert assessments reached 0.92, while the Kappa coefficient achieved 0.87, indicating excellent consistency. For RUL prediction, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) were adopted as evaluation metrics.

The proposed model was compared with Linear Regression and Back-Propagation (BP) neural networks. The comparison results of the residual life prediction errors for different models are shown in Figure 7.

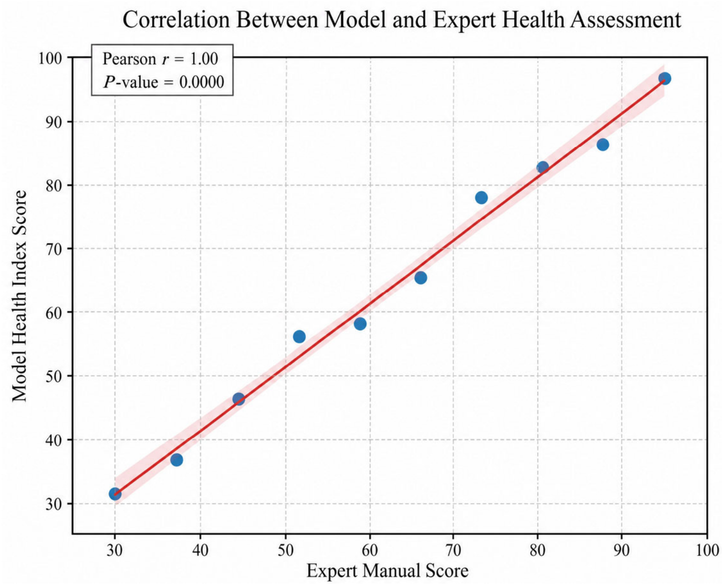

Figure 6. Correlation Analysis of Health Assessment and Expert Ratings.

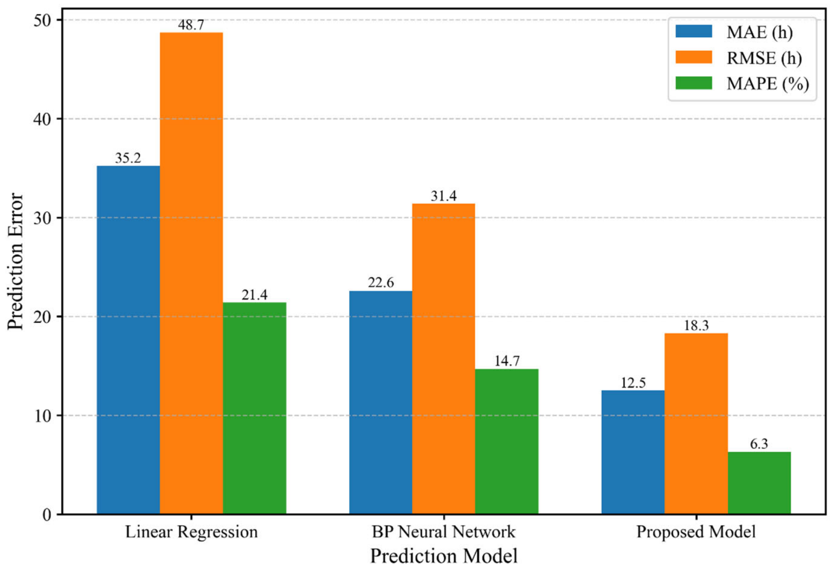

Figure 7. Comparison of Prediction Errors for Residual Lives of Different Models.

Results showed that the average prediction error for the three major fault categories was only 6.3%. The MAPE values for stator insulation faults, rotor inter-turn short circuits, and stator end-winding vibration faults were 7.7%, 5.8%, and 5.4%, respectively. The MAE and RMSE were 12.5 h and 18.3 h, respectively, significantly outperforming conventional prediction methods.

To further validate the physical plausibility of the HI-based RUL predictions, we overlaid the predicted HI degradation trajectories of three representative units against their actual fault evolution curves documented in maintenance logs. As shown in Table 14, the HI values exhibit a gradual decline starting 5 to 8 days prior to fault confirmation, with the rate of decline accelerating markedly within the final 48 h. The predicted HI curves follow the same trend as the actual degradation measurements, with the root mean square error between the predicted and actual HI trajectories below 3.2 points across all three cases. This alignment confirms that the HI model captures the underlying degradation dynamics rather than producing arbitrary numerical outputs.

Table 14. RUL prediction errors disaggregated by fault type and health level.

Fault Type	Health Level	MAE (h)	RMSE (h)	MAPE (%)
Stator insulation	Warning	8.2	12.5	6.1
Stator insulation	Critical	4.3	6.8	3.2
Rotor short circuit	Warning	7.5	11.2	5.4
Rotor short circuit	Critical	3.9	5.7	2.9
End-winding vibration	Warning	6.8	10.3	5.0
End-winding vibration	Critical	3.2	5.1	2.5

4.3. System-Level Performance and Engineering Validation

After module-level validation, a comprehensive system evaluation was conducted through a 12-month deployment at a 300 MW pumped storage power station. Continuous operation tests over 365 days demonstrated that the average processing time per sample remained below 100 ms, satisfying real-time monitoring requirements. No system crashes or data loss events occurred during operation. The data loss rate remained below 0.01%, while average CPU and memory utilization remained below 40% and 60%, respectively. Three representative fault cases were analyzed to evaluate practical fault-warning capability.

The first case involved stator winding insulation wear in Unit No 2. The complete evolutionary process and the warning sequence are shown in Figure 8. Beginning on 5 July 2024, the system detected a gradual increase in phase-A partial discharge intensity. A yellow warning was issued on 12 July, followed by an orange warning on 15 July. Scheduled maintenance conducted on 21 July confirmed three insulation wear defects, fully consistent with the diagnostic results. The prediction error of the estimated remaining life was only 7.7%.

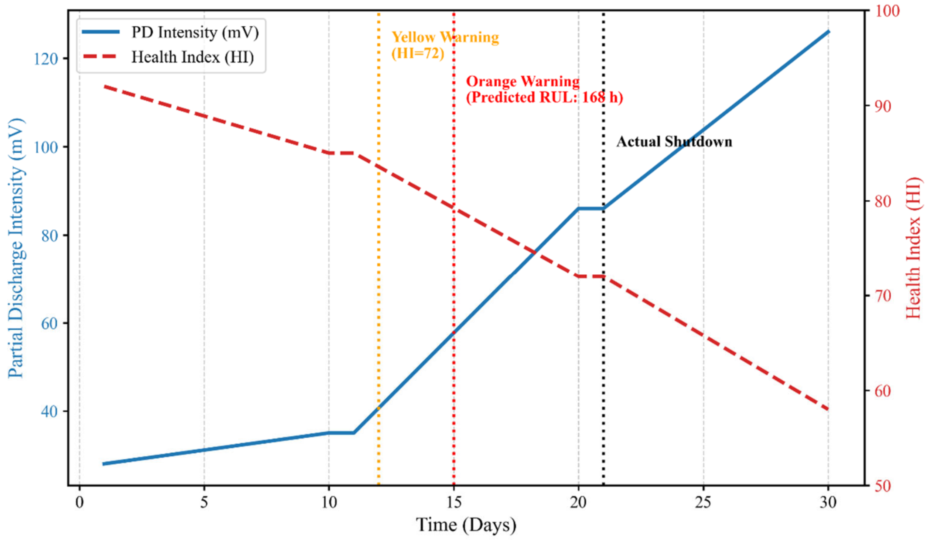

Figure 8. Evolution Process and Time Sequence Warning of Stator Winding Insulation Fault.

The second case involved an inter-turn short circuit in Pole No. 1 of Unit No. 3. The magnetic flux deviation ratio exceeded the 3% warning threshold on 5 September 2024. Subsequent inspection confirmed a two-turn short circuit in Slot No. 3, with a prediction error of only 4.2%. The third case involved stator end-winding tie-cord failure in Unit No. 1. The 100 Hz vibration amplitude exceeded the warning threshold on 20 November

2024. Maintenance performed five days later identified two broken tie cords, validating the system diagnosis. These cases demonstrate that the proposed system can identify incipient faults 5–10 days before failure, effectively preventing unplanned outages.

Under different operating modes, diagnostic accuracies reached 94.5% during generating operation, 93.8% during pumping operation, 92.7% during phase-modulation operation, and 91.2% during rapid operating-mode transitions. These results confirm the robustness and adaptability of the proposed system under complex operating conditions.

The model maintains robust performance (>90%) across all units and operating conditions. Accuracy is slightly lower during rapid transitions due to higher transient noise, yet remains practically acceptable. The per-unit variance is within 1.5% for each mode, indicating consistent generalization across different units (see Table 15).

Table 15. Diagnostic accuracy (%) across four units and four operating modes.

Operating Mode	Unit 1	Unit 2	Unit 3	Unit 4	Average
Generating	94.8	93.2	95.1	94.0	94.3
Pumping	93.5	92.8	94.2	93.0	93.4
Phase modulation	92.1	91.5	93.0	91.8	92.1
Rapid mode transition	90.8	89.7	91.5	90.2	90.6

4.4. Practical Benefits and Technical Value Analysis

The deployment of the proposed intelligent diagnosis and decision-support system generated substantial technical, economic, and societal benefits. From a technical perspective, fault prediction accuracy increased from 62% to 94.2% $\pm$ 0.7%, while the occurrence of major failures decreased by 85%. The average early-warning time was extended from 2 days to 8 days, and unplanned outages were reduced by 90%.

Economic benefits were quantified using Life Cycle Cost (LCC) analysis. Through condition-based predictive maintenance, annual maintenance costs per unit were reduced from RMB 8.60 million to RMB 6.90 million, corresponding to a reduction of 19.8%. For the four-unit power station, annual maintenance savings reached RMB 6.8 million. Furthermore, the system successfully prevented three major unplanned outages, reducing direct economic losses by approximately RMB 4.5 million and avoiding renewable energy curtailment losses estimated at RMB 12 million. Automated analysis also reduced manual inspection workloads by approximately 600 labor-hours annually, resulting in labor cost savings of RMB 0.6 million. Overall, the annual economic benefit of the system reached approximately RMB 23.9 million, while the total investment cost was only RMB 4.2 million. The static payback period was approximately 2.1 months, demonstrating excellent economic viability.

In addition, the system significantly improved operational safety and equipment reliability, reducing the risk of catastrophic failures such as stator insulation breakdown and rotor damage. Enhanced reliability further strengthened the grid's peak-shaving and frequency-regulation capability, thereby supporting large-scale renewable energy integration. Owing to its scalability and transferability, the proposed technical framework also provides valuable references for intelligent operation and maintenance of other large rotating machinery systems.

Calculation Methodology for Economic Benefits

All economic benefits reported in this study were calculated based on the LCC framework, with input data derived from the power station's actual financial and operational records. The detailed calculation methods are as follows.

1. Annual Maintenance Cost Savings

The maintenance cost per unit before deployment (C_{base}) and after deployment (C_{proposed}) are calculated as

$$C_{base} = N_{scheduled} \times C_{overhaul} + N_{unplanned} \times C_{emergency} \tag{37}$$

where $N_{\text{scheduled}}$ and N_{planned} are the annual number of scheduled/unplanned maintenance events; C_{overhaul} and $C_{\text{emergency}}$ are the average costs per scheduled overhaul and emergency repair, respectively; and C_{system} is the annualized cost of the proposed system (including hardware depreciation and software maintenance) (see Table 16).

The annual savings are

$$\Delta C_{maintenance} = C_{base} - C_{proposed} \tag{38}$$

Table 16. Input parameters for maintenance cost calculation.

Parameter	Value	Data Source
Scheduled overhauls per unit per year (before)	2.0	Power station maintenance logs (2022–2023)
Scheduled overhauls per unit per year (after)	1.4	Projected based on condition-based maintenance
Unplanned outages per unit per year (before)	1.2	Power station fault records (2022–2023)
Unplanned outages per unit per year (after)	0.12	Achieved during the 12-month deployment
Average cost per scheduled overhaul	RMB 3.2 M	Power station financial reports
Average cost per emergency repair	RMB 4.5 M	Power station financial reports
Annualized system cost (4 units)	RMB 0.6 M	Project investment (RMB 4.2 M/7-year lifecycle)

2. Losses Avoided from Unplanned Outages (ΔC_{outage})

Each unplanned outage incurs direct losses, including repair labor, replacement parts, and replacement power procurement:

$$\Delta C_{outage} = (N^{before}_{unplanned} - N^{after}_{unplanned}) \times C^{total}_{peroutage} \tag{39}$$

where $C^{total}_{peroutage}$ includes both direct repair costs (C_{repair}) and indirect losses due to power replacement and grid imbalance penalties (see Table 17).

Table 17. Input parameters for outage loss calculation.

Parameter	Value	Data Source
Direct repair cost per outage	RMB 1.5 M	Power station financial reports
Indirect loss per outage (power replacement)	RMB 2.8 M	Grid settlement data
Total loss per unplanned outage	RMB 4.3 M	Sum of above

3. Renewable Curtailment Avoidance ($\Delta C_{curtailment}$)

By improving the reliability of PSH units, the system avoids forced curtailment of wind and solar power that would otherwise occur due to insufficient grid regulation capacity:

$$\Delta C_{curtailment} = E_{avoided} \times P_{curtailment\_price} \tag{40}$$

where E_{avoided} is the amount of renewable energy (in MWh) that was successfully integrated due to improved PSH availability and $P_{\text{curtailment\_price}}$ is the average tariff of curtailed renewable energy (obtained from the regional grid operator) (see Table 18).

Table 18. Input parameters for curtailment loss calculation.

Parameter	Value	Data Source
Avoided renewable curtailment	85,000 MWh/year	Grid dispatch data
Average curtailment tariff	RMB 141/MWh	Regional grid policy
Total curtailment loss avoided	RMB 12.0 M/year	Product of above

4. Total Annual Economic Benefit

$$\Delta C_{curtailment} = E_{avoided} \times P_{curtailment\_price} \tag{41}$$

where ΔC_{labor} represents labor cost savings due to reduced manual inspection workload (calculated as 600 labor-hours × RMB 1000/hour = RMB 0.6 M/year). Based on the above methodology, the total annual economic benefit is RMB 23.9 M and the static payback period is RMB 4.2 M/RMB 23.9 M ≈ 0.18 years (approximately 2.1 months).

4.5. Discussion on Limitations and Practical Applicability

While the proposed intelligent diagnosis and decision-support system has demonstrated promising performance through both experimental validation and 12-month field deployment, several limitations should be acknowledged to provide a balanced assessment of its practical applicability and guide future improvements.

4.5.1. Scalability and Computational Requirements

The current system was validated on four 300 MW pumped storage units at a single power station. Although the total model size is relatively compact (approximately 0.19 M parameters), the training phase requires a GPU-accelerated computing environment (NVIDIA RTX A6600) and approximately 10 h of training time. For large-scale deployment across multiple stations with dozens of units, the centralized training architecture may face scalability challenges in terms of data storage, network bandwidth, and computational resource allocation.

Mitigation and future direction: We envision adopting a federated learning or edge-cloud collaborative architecture, where edge devices perform local inference and only send degraded or anomalous data segments to the cloud for model retraining. This would significantly reduce data transmission and centralized computational burdens while preserving diagnostic accuracy.

4.5.2. Robustness Against Sensor Failures and Missing Data

The proposed framework currently assumes that all monitoring sensors (PD couplers, full-flux probes, FBG accelerometers, and electrical transducers) are fully operational at all times. However, in actual power station environments, sensor drift, intermittent disconnection, or complete failure may occur due to strong electromagnetic interference, mechanical vibration, or aging of sensing components. The current model does not incorporate explicit mechanisms for handling missing or corrupted data streams.

Mitigation and future direction: Future work will investigate data imputation strategies (e.g., generative adversarial networks or matrix completion) to reconstruct missing sensor readings. Additionally, a multi-sensor redundancy design and sensor-fault detection module could be incorporated to alert operators when monitoring data quality deteriorates beyond acceptable levels.

4.5.3. Transferability to Other Pumped Storage Stations

The model was trained and validated using data from a specific 300 MW unit with particular design characteristics (e.g., rated speed, pole count, and cooling configuration).

Pumped storage stations with different unit capacities (e.g., 100 MW or 500 MW), manufacturers, or hydraulic conditions may exhibit distinct fault signatures and degradation patterns. Consequently, direct application of the pre-trained model to other stations without adaptation may result in degraded performance.

Mitigation and future direction: Transfer learning and domain adaptation techniques will be explored to reduce the performance gap when deploying the model to new stations. Specifically, we plan to investigate few-shot learning approaches that require only a small amount of local data to fine-tune the pre-trained model for site-specific conditions. Moreover, the proposed framework is designed with modular components (preprocessing, feature extraction, diagnosis, and RUL prediction), each of which can be independently re-calibrated using local data, facilitating adaptation without complete retraining.

4.5.4. Generalization to Unseen Operating Conditions

Although the system has been validated across four operating modes (pumping, generating, phase modulation, and mode transitions), it was evaluated under relatively steady operational schedules. Extreme operating conditions—such as emergency start-ups, grid black-start scenarios, or operation under severely imbalanced grid voltages—were not explicitly included in the training or validation datasets. The model's performance under such conditions remains unknown.

Mitigation and future direction: We will continue to collect data under more diverse and extreme operating scenarios to progressively expand the model's operational envelope. Online learning strategies, where the model updates itself incrementally as new data become available, are also being considered to adapt to evolving operating patterns without periodic offline retraining.

4.5.5. Limitation of Current Fault Coverage

The current framework focuses on three dominant fault types (stator insulation degradation, rotor inter-turn short circuits, and stator end-winding vibration), which together account for 82.6% of failures in PSH units. However, other failure modes—such as bearing wear, governor system faults, and cooling system failures—are not addressed. Furthermore, compound faults (where two or more fault types co-exist and interact) have not been specifically studied, and the model's performance under such coupled conditions remains to be verified.

Mitigation and future direction: Future research will incorporate additional sensor modalities and fault signatures to expand the coverage of fault types. For compound faults, we plan to investigate multi-label classification approaches and causal inference methods to disentangle the interactions between concurrent fault mechanisms.

4.5.6. Summary of Practical Applicability

Despite these limitations, we emphasize that the proposed system has demonstrated reliable operation over 12 continuous months in an active power station, achieving fault warnings 5–10 days in advance and reducing unplanned outages by 90%. These results confirm that within its current scope (300 MW units, three dominant fault types, and standard operating conditions), the framework provides a practically viable solution for intelligent condition-based maintenance. The limitations identified above represent opportunities for future research rather than fatal flaws, and we have outlined concrete pathways to address each in ongoing and planned work.

5. Conclusions

With the increasing penetration of renewable energy in modern power systems, pumped storage units have become indispensable regulating assets for maintaining grid stability and enhancing renewable energy integration. However, their unique operating characteristics, including bidirectional operation, frequent start-stop cycles, and deep load regulation, expose the equipment to significantly higher fault risks than those encountered in conventional generator units. Consequently, traditional time-based maintenance strategies are no longer adequate for ensuring reliable and economical operation.

To address the challenges associated with multi-source data fusion, early fault detection, and maintenance decision-making, this study proposed an integrated intelligent operation and maintenance framework incorporating condition monitoring, intelligent diagnosis, and decision support. The framework was validated through both simulation studies and practical engineering applications in a pumped storage power station.

The major contributions of this work can be summarized as follows: (1) an integrated condition monitoring framework with event-triggered adaptive time alignment and hierarchical joint denoising for heterogeneous multi-source data; (2) a three-stage hybrid intelligent fault diagnosis framework combining CVMD-based signal decomposition, KPCA-based nonlinear feature fusion, and attention-enhanced LSTM for temporal learning; and (3) a predictive maintenance decision-support framework based on quantitative health assessment and remaining useful life prediction.

Experimental and engineering validation results demonstrate that the proposed system achieved an overall fault diagnosis accuracy of $94.2\% \pm 0.7\%$, provided fault warnings 5 to 10 days in advance, and reduced annual maintenance costs by approximately 19.8%.

Beyond the specific application to PSH units, the proposed framework offers generalizable value for other large rotating machinery systems, such as wind turbine generators, gas turbines, and centrifugal compressors, which share similar characteristics of multi-sensor monitoring, non-stationary operation, and degradation-driven failures. The modular design of the framework (preprocessing, feature extraction, diagnosis, and prognostics) allows independent re-calibration or replacement of each module using site-specific data. The core diagnostic architecture, particularly the attention-enhanced temporal learning mechanism, is directly transferable to any machinery with available multi-sensor time-series data. For practical deployment, the primary adaptation effort would focus on re-parameterizing the CVMD and KPCA modules according to the dominant frequency bands and fault characteristics of the target equipment.

Future work will focus on three specific directions to further enhance the framework‘s applicability and robustness. First, transfer learning and few-shot adaptation will be investigated to enable rapid cross-station deployment with limited local data, reducing the reliance on large-scale labeled datasets from each new site. Second, lightweight model deployment on edge computing terminals will be pursued through knowledge distillation and model pruning techniques, allowing real-time inference directly on IEDs without dependence on centralized GPU servers. Third, compound fault diagnosis will be addressed by extending the current single-label classification framework to multi-label and causal inference formulations, enabling the disentanglement of concurrent and interacting fault mechanisms.

Author Contributions: Conceptualization, X.L. and J.Z.; methodology, X.L.; software, X.L.; validation, X.L., J.B. and B.D.; formal analysis, T.L.; investigation, X.Y.; resources, J.Z.; data curation, B.D.; writing—original draft preparation, X.L.; writing—review and editing, J.Z., J.B. and X.Y.; visualization, T.L.; supervision, J.Z.; project administration, J.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 2026 Science and Technology Project of Inner Mongolia Power (Group) Co., Ltd., on 26 March 2026.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Xuan Liu, Jie Bai, Bingjie Dou, Tianyu Liu, Xiaohui Yang were employed by the Inner Mongolia Power (Group) Co., Ltd. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Feng, X.; Huang, R.; Xu, L.; Li, Y.; Liu, X.; Zeng, F.; Wang, Z. Dam-Axis Siting with Improved Adaptive Variable Neighborhood Search Algorithm. *Infrastructures* **2026**, *11*, 182. [CrossRef]
2. Chen, X.; Yu, S.; Hong, Q.; Wang, Y.-J.; Wang, Y.; Li, P. Assessment of Excavation-Induced Soil and Ecological Degradation in Pumped-Storage Hydropower Construction Areas Using Field Measurements and Time-Series Remote Sensing. *Appl. Sci.* **2026**, *16*, 5173. [CrossRef]
3. Hou, W.; Wei, H. Data-driven robust day-ahead unit commitment model for hydro/thermal/wind/photovoltaic/nuclear power systems. *Int. J. Electr. Power Energy Syst.* **2021**, *125*, 106427. [CrossRef]
4. Han, X.; Zhang, J.; Qin, X.; Gao, J.; Zhao, Y.; Zhao, Z.; Cheng, C. A Method for Formulating Delivery Curves of Clean Energy Bases Considering Load Demand of Receiving Provinces. *Energies* **2026**, *19*, 2445. [CrossRef]
5. El Mansour, J.; Hendrick, P.; Hajjaji, A.; Belhora, F. A Study of Erosion–Cavitation Inception Synergy in Seawater Centrifugal Pumps. *Processes* **2026**, *14*, 1438. [CrossRef]
6. Kozakowski, T.; Kozioł, M.; Koniuszy, A.; Tkaczyk, K. Modern Energy Storage Methods and Technologies: Comparison, Case Study and Analysis of the Impact on Power Grid Stabilization. *Sustainability* **2026**, *18*, 2659. [CrossRef]
7. Magableh, S.K.; Magableh, M.A.; Dawaghreh, O.M.; Wang, C. Advanced Energy Management and Dynamic Stability Assessment of a Utility-Scale Grid-Connected Hybrid PV–PSH–BES System. *Electronics* **2026**, *15*, 384. [CrossRef]
8. Chen, L.; Li, J.; Deng, L.; Xie, E.; Yan, X.; Hao, G.; Chen, H.; Xue, H.; Zhong, Z.; Kan, K. Hydraulic Instability Characteristics of Pumped-Storage Units During the Transition from Hot Standby to Power Generation. *Water* **2026**, *18*, 61. [CrossRef]
9. Lee, S.-H. Electrical Conductivity as an Inline Monitor for Aqueous Precipitation and Crystallization: Mechanistic Interpretability and a Model-Implementation Blueprint. *Minerals* **2026**, *16*, 658. [CrossRef]
10. Thobejane, L.; Thango, B. On the Role of Feature Extraction in Transformer PD Severity Classification: A Controlled Comparison of PCA and Autoencoder Models. *Machines* **2026**, *14*, 708. [CrossRef]
11. *IEC 60270:2015*; High-Voltage Test Techniques–Partial Discharge Measurements. International Electrotechnical Commission: Geneva, Switzerland, 2015.
12. *GB/T 7354-2018*; High-Voltage Test Techniques–Partial Discharge Measurements. Standards Press of China: Beijing, China, 2018.
13. Sivasubramanian, K.; Bharath, R.R.; Vajravelu, L.K.; Kumar D, M.; Pamarthi, J. Assessment of IL-6 and IL-8 Levels and Other Bio Markers in Predicting Dengue Severity Across Serotypes. *Pathogens* **2026**, *15*, 434. [CrossRef] [PubMed]
14. Aguas, R.; Dorigatti, I.; Coudeville, L.; Luxemburger, C.; Ferguson, N.M. Cross-serotype interactions and disease outcome prediction of dengue infections in Vietnam. *Sci. Rep.* **2019**, *9*, 9395. [CrossRef] [PubMed]
15. Choe, S.; Faizan Shaikh, M.; Byun, N.; Kim, K.; Kim, K.; Chen, A.; Nysveen, A.; Ehya, H.; Bin Lee, S. Silicon Controlled Rectifier (SCR)-Embedded Identification of Field Winding, Damper Bar, and Eccentricity Faults in Wound Field Synchronous Machines. *IEEE Trans. Ind. Appl.* **2026**. [CrossRef]
16. Winardi, A.; Sudiarto, B.; Jaya, I.; Baskoro, Y. A Practical Approach for Estimating Short Interturn Fault Severity in Turbogenerator Rotor Using RSO Measurement. *IEEE Trans. Dielectr. Electr. Insul.* **2026**. [CrossRef]
17. Shaikh, M.F.; Battulga, B.; Chen, A.; Nysveen, A.; Ehya, H.; Kim, K. Sensitive detection of shorted field winding turns in salient pole synchronous machines with high frequency pole drop test. *IEEE Trans. Energy Convers.* **2025**, *40*, 1686–1689. [CrossRef]
18. Shaikh, M.F.; Kim, H.J.; Byun, N.; Lee, S.; Kim, K.; Kong, T.; Ko, B.; Kim, K.; Park, H. A simple offline test for detecting rotor faults in wound field synchronous machines. *IEEE Trans. Energy Convers.* **2023**, *38*, 2980–2983. [CrossRef]
19. Xu, S.; Li, X.; Wang, T.; Wang, T.; Liu, H. Fiber Bragg grating pressure sensors: A review. *Opt. Eng.* **2023**, *62*, 010902. [CrossRef]
20. Shadab, A.; Raghuwanshi, S.K.; Kumar, S. Advances in micro-fabricated fiber Bragg grating for detection of physical, chemical, and biological parameters—A review. *IEEE Sens. J.* **2022**, *22*, 15650–15660. [CrossRef]
21. Bonopera, M. Fiber-Bragg-Grating-Based Displacement Sensors: Review of Recent Advances. *Materials* **2022**, *15*, 5561. [CrossRef] [PubMed]
22. Dong, S.; Huang, T.; Chen, L.; Yin, S.; Chang, Q.; Han, D.; Li, Y.; Zhang, B.; Liao, H.; Chen, F. Fiber Bragg Grating Force Sensors for Minimally Invasive Surgery: State of the Art, Challenges, and Opportunities. *SmartBot* **2026**, e70040. [CrossRef]
23. Zhao, H.; Gong, Z.; Gan, K.; Gan, Y.; Xing, H.; Wang, S. Supervised kernel principal component analysis-polynomial chaos-Kriging for high-dimensional surrogate modelling and optimization. *Knowl.-Based Syst.* **2024**, *305*, 112617. [CrossRef]

24. Yu, Y.; Zhou, C.; Wang, Y.; Lu, T.; Fan, Z. Power System state prediction method based on improved long short-term memory considering renewable energy uncertainty. *Eng. Appl. Artif. Intell.* **2026**, *172*, 114313. [CrossRef]
25. Pervez, A.; Hasan, S.U.; Norrish, A.R. Convolutional neural networks in paediatric fracture detection: Pooled evidence from a systematic review and meta-analysis. *Eur. Radiol.* **2026**, *36*, 6312–6324. [CrossRef] [PubMed]
26. Özarslan, H.; Uluocak, I. CNN-LSTM and CNN-GRU based estimation of NOx conversion efficiency in diesel engine exhaust system. *Fuel* **2026**, *405*, 136807. [CrossRef]
27. Zhang, W.; Suo, W.; Song, J.; Cao, W. Physics-informed neural networks (PINNs) as intelligent computing technique for solving partial differential equations: Limitation and future prospects. *Sci. China Phys. Mech. Astron.* **2026**, *69*, 214602. [CrossRef]
28. Wang, Z.; Yu, X.; Guo, Y.; Yang, J.; Qiao, Z. Research on a novel weak fault detection method based on vibrational resonance in high-dimensional chaotic system, and variational mode decomposition. *Chaos Solitons Fractals* **2026**, *208*, 118084. [CrossRef]
29. Deng, Z.; Miao, B.; Chen, Q.; Chen, J.; Tong, C.; Liu, H.; Deendarlianto; Suwarno; Wang, H.; Chan, S. Degradation prediction and remaining useful life estimation of PEMFCs: Mechanisms, methods, datasets, and challenges. *Renew. Sustain. Energy Rev.* **2026**, *229*, 116598. [CrossRef]
30. Kaib, M.T.H.; Kouadri, A.; Harkat, M.F.; Bensmail, A.; Mansouri, M. Data size reduction approach for nonlinear process monitoring refinement using Kernel PCA technique. *Expert Syst. Appl.* **2025**, *274*, 126975. [CrossRef]
31. Shahgholi, M.; Awrejcewicz, J. Vibration control of a parametrically excited asymmetrical rotor-active magnetic bearings system with oil-film effect. *Nonlinear Dyn.* **2025**, *113*, 17855–17881. [CrossRef]
32. Wang, P.; Liu, Y.; Liu, Z. A fault diagnosis method for rotating machinery in nuclear power plants based on long short-term memory and temporal convolutional networks. *Ann. Nucl. Energy* **2025**, *213*, 111092. [CrossRef]
33. Wang, H.; Wang, H.; Tang, X. A Review of Deep Learning in Rotating Machinery Fault Diagnosis and Its Prospects for Port Applications. *Appl. Sci.* **2025**, *15*, 11303. [CrossRef]
34. Wang, H.; Wu, S.; Yu, F.; Bi, Y.; Xu, Z. Study on remaining useful life prediction of sliding bearings in nuclear power plant shielded pumps based on nearest similar distance particle filtering. *Ann. Nucl. Energy* **2025**, *223*, 111625. [CrossRef]
35. Irfan, M.; Tahir, T.; Deilami, S.; Huang, S.; Veettil, B.P. Novel artificial intelligence driven model-based control framework for solar–electric vehicle home energy optimization. *Eng. Appl. Artif. Intell.* **2026**, *177*, 114868. [CrossRef]
36. *ISO 20816-5:2018*; Mechanical Vibration—Measurement and Evaluation of Machine Vibration—Part 5: Machine Sets in Hydraulic Power Generating and Pump-Storage Plants. International Organization for Standardization: Geneva, Switzerland, 2018.
37. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
38. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.
39. Vaswani, A.; Shazeer, N.; Parmar, N.; Kaiser, L.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30*; Curran Associates, Inc.: Red Hook, NJ, USA, 2017.