EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

A deep reinforcement learning-based adaptive multivariate state estimation fault diagnosis method for a photovoltaic storage grid-connected system under a high proportion of new energy integration

Yiguo Guo[1,2], Zhiwei Wang[1,2], Fumu Lu[1,2], Yimu Fu[1,2], Zhiqing Wang[3], Jinghui Meng[3,*]

[1]Shandong Zhiyuan Electric Power Design and Consulting Co., Ltd., Jinan China
[2]Economic & Technology Research Institute, State Grid Shandong Electric Power Company, Jinan China
[3]North China Electric Power University, Beijing, China

Abstract

With the increasing penetration of photovoltaic and other renewable energy sources in power systems, the operational characteristics of photovoltaic storage grid-connected systems have become increasingly complex. The stochastic nature of renewable energy output and the alternating charging/discharging states of energy storage systems due to diurnal cycles lead to significant time-varying operational conditions, under which the performance of conventional multivariate state estimation methods for fault diagnosis markedly deteriorates. To address this, targeting high-penetration renewable energy integration scenarios, this paper proposes an adaptive state estimation-based fault diagnosis method for photovoltaic-storage grid-connected systems using deep reinforcement learning. Within the deep reinforcement learning framework, an agent is deployed to adaptively match and update modeling samples according to the real-time system state, thereby enhancing the adaptability of state estimation and the fault diagnosis accuracy. Test results under multivariate typical operational scenarios demonstrate that, compared to existing advanced fault diagnosis methods, the proposed approach achieves an average improvement of 10% in fault detection rate and reduces the missed detection rate by 3%, validating its superior adaptability and diagnostic performance in systems with complex time-varying characteristics. This research provides a new technical pathway for ensuring the safe and stable operation of photovoltaic-storage grid-connected systems under high-penetration renewable energy integration, contributing to enhanced intelligent operation and maintenance capabilities and improved fault response.

Keywords: deep agent reinforcement learning; high proportion of new energy integration; photovoltaic storage grid connected system; fault diagnosis; adaptive multivariate state estimation.

Received on 17 September 2026, accepted on 19 December 2025, published on 15 April 2026

doi: 10.4108/ew.12145

1. Introduction

As the penetration rate of renewable energy sources such as photovoltaics continues to rise [1-2], photovoltaic-storage grid-connected systems have become a vital

*Corresponding author. Email: mengjh@ncepu.edu.cn

component of modern power systems. By integrating flexible intermittent photovoltaic generation, controllable energy reservoir units, these systems effectively enhance the grid's ability to accommodate renewable energy and improve overall stability [3-4]. Bai et al [5] investigated control strategies for battery storage in VSG-based wind-storage grid-connected systems, analyzing the characteristics of such systems under renewable energy integration. Jiang et al [6] proposed a multiparameter frequency stability management scheme for photovoltaic-storage systems that accounts for frequency regulation dead zones, and conducted in-depth research on factors affecting system frequency stability. In fact, as the proportion of renewable energy integration continues to increase, the operation characteristics of the system undergo profound changes: the random fluctuations in photovoltaic output intertwine with the day-night cycle of charge-discharge state switching on the storage side, resulting in a highly time-varying operation state for the system. This presents new challenges for the stable and safe operation.

As a key technology for guaranteeing the security, stability, and efficient operation of systems, fault diagnosis has become a current research hotspot [7-8]. In the field of fault detection, numerous algorithms have been developed in recent years. These can be broadly classified into two categories: statistical analyses [9] and machine learning algorithms [10]. Among these, multivariable statistical inference-based methods detect changes in system status by establishing relationships between operation data and system health conditions. Such methods do not require fault samples, making them more suitable for practical applications. Furthermore, multivariate statistical-based fault detection methods can be subdivided into parametric [11] and non-parametric approaches [12]. Parametric methods require constructing precise parametric diagnostic models for the target system, whereas non-parametric methods rely solely on extensive data without model structures. Common parametric multivariate statistical fault detection techniques include Principal Component Analysis (PCA) [13], Independent Component Analysis (ICA) [14], and Partial Least Squares (PLS) [15]. Liu et al [16] suggested a Dynamic Concurrent Kernel Partial Least Squares approach that handles nonlinear and dynamic industrial processes with kernel techniques. Compared to parametric multivariate methods, nonparametric multivariate fault detection avoids modeling each operating mode, better meeting the demands of complex system fault detection—such as photovoltaic-storage grid-connected systems with high renewable energy penetration. Widely applied nonparametric multivariate failure diagnosis methods include: Multivariate State Estimation (MSET) [17], Gaussian Mixture Models (GMM) [18], and Sequential Probability Ratio Test (SPRT) [19]. Among them, MSET is more practical and feasible in real-world applications due to its lack of complex parameter searches and model structure assumptions. MSET primarily utilizes historical operation data (the history matrix) to represent the estimated health information of the system. The estimates are derived from computing a linearly scaled sum over the historical operation records and the present measurement results using the least squares approach. Lv et al. [20] transformed the problem of updating the history matrix into an optimization problem.

In fact, the operation of high-proportion renewable energy photovoltaic-storage grid-connected systems is complex and multi-step. Consequently, the entire system comprises multiple devices, each responsible for distinct operation subtasks. Simultaneously, they coordinate to complete the entire power generation and grid-connection process. Therefore, the system may experience various failures, including photovoltaic panel failures, energy storage side failures, grid-connection failures, and others. Moreover, the operation characteristics of high-penetration renewable energy photovoltaic-storage grid-connected systems are time-varying rather than static. This arises from the interplay between the random fluctuations in photovoltaic output and the daily cycle of charge-discharge state transitions on the storage side. However, the historical matrix update methods in existing MSET approaches are poorly suited for this context. Their update objective functions and criteria require expert knowledge, posing significant challenges for accurate fault diagnosis in high-penetration renewable energy photovoltaic-storage grid-connected systems.

In summary, this paper proposes an adaptive state estimation and fault diagnosis method based on deep agent reinforcement learning for high-penetration renewable energy photovoltaic-storage grid-connected systems. Considering the time-varying characteristics of such systems, performing state estimation based on all samples in the historical matrix is inappropriate. When the photovoltaic-storage grid-connected system operates during daylight hours, utilizing nighttime operation samples would lead to inaccurate state estimation and erroneous fault diagnosis results. Furthermore, employing the entire historical matrix increases calculation time. Therefore, this research suggests an effective update strategy based on deep agent reinforcement learning. This strategy transforms the historical matrix update problem into an agent's sample selection decision. In particular, the historical memory issue can be modelled as a Markov decision process. The state consists of data samples containing attribute and target variables collected for high-penetration renewable energy photovoltaic storage grid-connected systems, while the reward reflects the accuracy of classification algorithms. In essence, the proposed approach simulates the operation of high-penetration renewable energy photovoltaic storage grid-connected systems within a deep reinforcement learning framework, aiming to maximize reward. Through expert-free agent learning in a dynamic environment, representative model samples most matching the agent decision can be chosen from the historical matrix, enabling adaptable historical matrix updates. Finally, based on the updated historical matrix, a more effective adaptive state estimation fault diagnosis model for high-penetration renewable energy

photovoltaic storage grid-connected systems can be established.

2. Construction of Grid-Connected Systems with High Proportion of New Energy Photovoltaic Storage

2.1 System Architecture Design

The photovoltaic-storage microgrid system architecture constructed is shown in Figure 1. This system is composed of multiple functional subsystems working in concert: the photovoltaic power generation subsystem handles energy harvesting and preliminary conversion, with its core comprising photovoltaic modules and supporting DC-AC conversion equipment; the energy storage subsystem is responsible for energy regulation and storage management, featuring energy storage components and reversible voltage regulation modules; the grid-connection interface unit relies on power electronic conversion equipment to enable energy exchange between the system and the external grid. For operation control, the PV generation section employs Maximum Power Point Tracking (MPPT) technology to maximise energy harvesting and incorporates power limiting capabilities. The energy storage segment utilizes a dual-closed-loop control strategy to ensure safe and stable charging/discharging processes. The grid-connection component applies Virtual Synchronous Generator (VSG) control technology to align its output characteristics with grid integration requirements. Considering the combined effects of environmental factors on PV output, constraints on energy storage capacity and operation status, and the fluctuating characteristics of load demand, this system requires a multidimensional coordinated control strategy to maximize overall operation efficiency and maintain system stability.

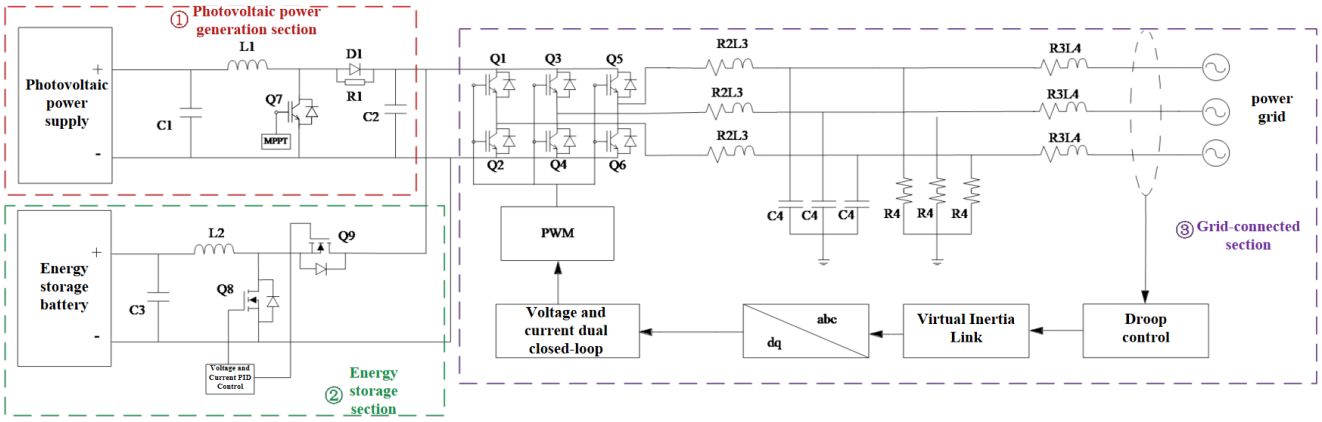


Figure 1. Structure of photovoltaic storage grid-connected system under a high proportion of new energy integration

2.2 Principles of Photovoltaic MPPT Control

MPPT control serves as a crucial technology for optimizing energy output in photovoltaic systems. Its core principle lies in dynamically adjusting circuit impedance to ensure photovoltaic cells consistently operate at their maximum power point, thereby enhancing the overall energy conversion efficiency. The output characteristics of photovoltaic cells exhibit significant nonlinearity. Their current–voltage (I–V) response function and power–voltage (P–V) response function shift with changes to environmental parameters. Under Standard Test Conditions, The P-V graph of a solar panel exhibits a max power point, which represents the max power point the system should be tracked. Typical algorithms currently employed for MPPT control include constant voltage tracking, conductance increment method, and disturbance observation method. In this study, the voltage ratio approach is selected as the MPPT implementation strategy, and the corresponding circuit topology is shown in Figure 2.

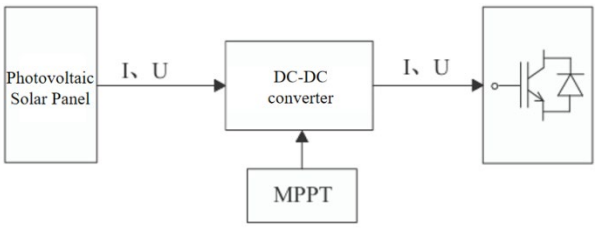


Figure 2. Photovoltaic diagram

2.3 Overall Model of Energy Storage Circuit Based on VSG Control

The energy storage circuit employs a buck-boost converter as the core bidirectional DC/DC energy conversion component. This converter features a single-switch non-isolated structure, enabling flexible adjustment of the output voltage above or below the input voltage level. For the control strategy, this paper adopts a voltage-current dual closed-loop control architecture for the buck-boost converter, as illustrated in Figure 3. Within this structure, the outer voltage loop precisely regulates the output voltage, ensuring voltage stability during the energy storage system charging and discharging processes. This effectively prevents overcharging and over-discharging while mitigating interference from grid-side voltage fluctuations. The inner current loop rapidly tracks dynamic load changes, enabling precise control of charging and discharging currents. This approach optimizes the system's energy conversion efficiency while preventing overcurrent damage to the energy storage components. This dual-closed-loop control system significantly enhances overall dynamic response speed, stability, and reliability. It improves the system flexibility to different load profiles and energy storage components, thereby meeting the functional requirements of bidirectional charging and discharging.

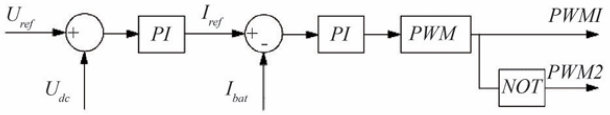


Figure 3. Energy storage circuit control strategy

The key droop control, based on the set active and reactive power reference values, employs the mathematical expression for power control in Equation (1). This equation can be used to calculate the frequency and amplitude of the output voltage.

$$\begin{cases} U = U_0 - nQ \\ f = f_n + m\left(P_n - P\right) \end{cases} \tag{1}$$

In this equation: U_0 is the voltage amplitude when reactive power is zero; n and m are the droop coefficients for reactive and active power, respectively; f_n is the rated frequency; P_n is the micro power source active power output; Q and P are the reactive and active power outputs of the inverter power supply, respectively.

Grid frequency fluctuations primarily stem from transient imbalances between power generation and consumption. Traditional synchronous generators store mechanical energy through the rotational mass of their rotors, suppressing speed variations by adjusting mechanical power input. However, grid-connected inverters, as static power electronic devices lacking mechanical rotating components, cannot adaptively balance power fluctuations via mechanical torque and thus struggle to directly support grid frequency stability. Consequently, Virtual Synchronous Generator (VSG) control technology must be introduced to simulate the inertia characteristics of synchronous motors. In conventional units, inertia response is provided by mechanical rotors, exhibiting relatively slow dynamic processes. In contrast, the main circuit of a VSG consists of power electronic devices, with its equivalent inertia derived from distributed power sources or energy storage units on the DC side, enabling faster response capabilities. By establishing a mathematical model simulating the rotor motion of a synchronous generator, the virtual rotor dynamics equation for the VSG can be derived, as specifically expressed in Equation (2).

$$\begin{cases} \dfrac{d\theta}{dt} = w_o \\ J\dfrac{dw_o}{dt} = \dfrac{P_m}{w_s} - \dfrac{P_o}{w_s} - D\left(w_o - w_s\right) \end{cases} \tag{2}$$

In the equation: θ is the work angle; w_o is the output angular velocity; J is the virtual inertia; w_s is the rated angular velocity; D is the damping; P_m and P_o are the mechanical power and electromagnetic power, respectively.

Figure 4 illustrates the system control process: Grid voltage and current signals are first fed into the power calculation unit for processing, yielding instantaneous active and reactive power values. Active power is input to the virtual rotor model and virtual governor module, which compute the system power angle and angular frequency. This process eliminates the need for a phase-locked loop, significantly enhancing the system's dynamic response performance. Concurrently, reactive power is processed by the virtual exciter to generate corresponding potential signals. Subsequently, the obtained power angle and potential signals undergo Park transformation to decouple voltage and current components. Utilizing dual-closed-loop control of current and voltage, PWM drive signals are generated to ultimately control the inverter output. This control scheme integrates droop control with a virtual inertia element. Its structural design closely aligns with the operation mechanism of synchronous generators, thereby fully reproducing the dynamic characteristics of traditional synchronous motors.

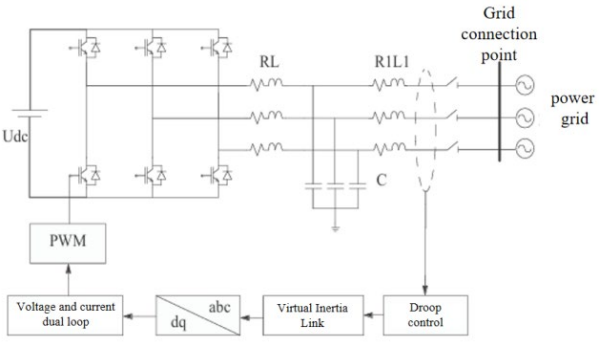


Figure 4. Energy storage circuit model

3. Fault Diagnosis Methods

3.1 Multi-State Estimation Fault Diagnosis Model

Multivariate State Estimation (MSET), as a typical nonparametric mathematical fault diagnosis approach, considers that the system has M variables. It represents the system state at time t_i as:

$$z(t_i) = z_i = \left[z_{1,i}, z_{2,i}, \ldots, z_{M,i}\right]^{\mathrm{T}} \tag{3}$$

Collect n normal operation samples to construct the historical matrix:

$$\boldsymbol{D} = \left[z_1, z_2, \ldots, z_n\right] = \begin{bmatrix} z_{1,1}, z_{1,2}, \cdots, z_{1,n} \\ z_{2,1}, z_{2,2}, \cdots, z_{2,n} \\ \vdots \\ z_{M,1}, z_{M,2}, \cdots, z_{M,n} \end{bmatrix} \tag{4}$$

Every column in the historical matrix $\boldsymbol{D}$ means the vector of M variables at different time points, while each column denotes a normal historical sample at a distinct time point. In the MSET model, for the estimate vector z_{obs} with estimates, its estimated vector z_{est} is computed through linear weighting by the historical matrix $\boldsymbol{D}$:

$$z_{\mathrm{est}} = \boldsymbol{D} \cdot \boldsymbol{w} = w_1 z_1 + w_2 z_2 + \ldots + w_n z_n \tag{5}$$

The weight coefficient $\boldsymbol{w}$ can be obtained by minimising the quadratic residual error $\varepsilon_{\mathrm{est}}$.

$$\varepsilon_{\mathrm{est}} = \left(z_{\mathrm{est}} - z_{\mathrm{obs}}\right)^{\mathrm{T}} \left(z_{\mathrm{est}} - z_{\mathrm{obs}}\right) \tag{6}$$

To minimize the squared estimation error, calculate the gradient of $\boldsymbol{w}$ and set it to zero:

$$\frac{d\varepsilon_{\mathrm{est}}}{d\boldsymbol{w}} = \frac{d\left(z_{\mathrm{est}} - z_{\mathrm{obs}}\right)^{\mathrm{T}} \left(z_{\mathrm{est}} - z_{\mathrm{obs}}\right)}{d\boldsymbol{w}} = 2\boldsymbol{D}^{\mathrm{T}}\left(\boldsymbol{D}\boldsymbol{w} - z_{\mathrm{obs}}\right) = 0 \tag{7}$$

The weight coefficient $\boldsymbol{w}$ can be determined:

$$\boldsymbol{w} = \left(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\right)^{-1} \cdot \left(\boldsymbol{D}^{\mathrm{T}} z_{\mathrm{obs}}\right) \tag{8}$$

The final result is the estimated value z_{est} of the observation vector z_{obs}.

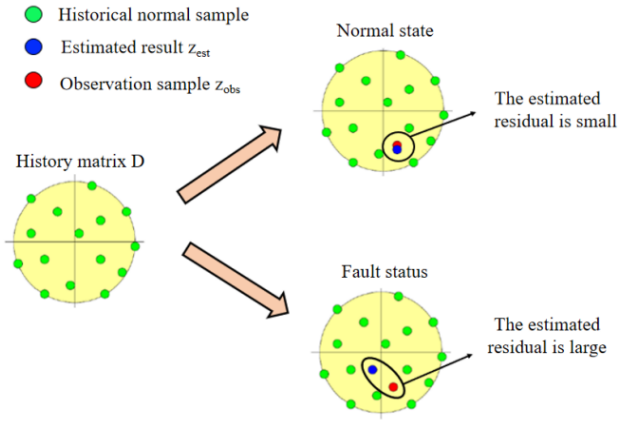


Figure 5. MSET fault diagnosis

As shown in Figure 5, if the system performs correctly, the observed sample z_{obs} resembles the correct samples in the historical matrix $\boldsymbol{D}$. Consequently, the observed sample can be accurately recovered using the historical matrix $\boldsymbol{D}$, enabling precise state estimation through the MSET model. Conversely, if a fault occurs, the MSET model will fail to accurately predict the observed sample, leading to significant prediction errors. Therefore, MSET achieves fault detection.

3.2 Adaptive State Estimation Fault Diagnosis Method based on Deep Multi-Agent Reinforcement Learning

As mentioned above, MSET-based algorithms primarily utilise historical matrices for parameter calculation, making the fault diagnosis accuracy influenced in this matrix. In fact, high-penetration renewable energy photovoltaic-storage grid-connected systems exhibit time-varying characteristics. Therefore, to enhance fault diagnosis performance, adaptive updating of the historical matrix is necessary. Updating historical matrices can be viewed as a sampling design issue, that is, choosing samples of historical matrices that most closely match the system's actual operating conditions, while dropping redundant samples. Nevertheless, establishing sample collection methods based on expert knowledge limits the practical application of MSET-based methods.

To overcome this aforementioned challenge, we developed an adaptive state estimation fault detection approach using deep multi-agent reinforcement training, as illustrated in Figure 6. This method primarily comprises two components: 1) offline training, and 2) online estimation. The initial section details the learning approach, consisting of the general learning framework, learning steps, specific formulas, and descriptions of states, actions, and rewards. The second section presents the online calculation for fault detection.

3.2.1 Offline training

Given a training dataset $\boldsymbol{D}_{tra} = \{\boldsymbol{x}_t, y_t\}_{t=1}^{T}$, where T denotes the sample size, y_t=0 indicates the normal sample, and y_t=1 indicates the fault sample. The training process for this adaptive state identification fault diagnosis approach using deep reinforcement learning is illustrated by Figure 7.

To achieve historical adaptive updates of the historical matrix, the agent selects pattern samples from the historical matrix to construct estimation models with superior failure detection capability. In this case, active pattern sample generation becomes an agent planning issue and can be described as a Markov Decision Process (MDP). To resolve it, the Soft Actor-Critic (SAC) algorithm is used. In this algorithm, the agent consists of a value function and an actor-critic function. This actor uses a deterministic action decision process to select training samples, while the critic network evaluates the quality of this process, providing guidance to the actor network during training. Through this approach, the SAC algorithm fully leverages the interaction information between the agent and the high-

penetration renewable energy photovoltaic storage grid-connected system, achieving better training results.

We define the state s, the reward r, the agent policy π given parameter θ, the value function Q given parameter ω, and the action a. The value loss function is utilized to compute the gradients of the parameters of the agent's actor network:

$$L_\pi(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{s}_t \sim R, \boldsymbol{a}_t \sim \mu_\theta}\left[\alpha log\left(\pi_\theta\left(\boldsymbol{a}_t \middle| \boldsymbol{s}_t\right)\right) - Q_\omega\left(\boldsymbol{s}_t, \boldsymbol{a}_t\right)\right] \quad (9)$$

where, $\boldsymbol{a}_t$ and $\boldsymbol{s}_t$ denote the action and state of the actor at step t, respectively. R is the memory buffer, where every element in R represents $(\boldsymbol{s}, \boldsymbol{s}', \boldsymbol{a}, r)$ and $\boldsymbol{s}'$ is the agent's subsequent action.

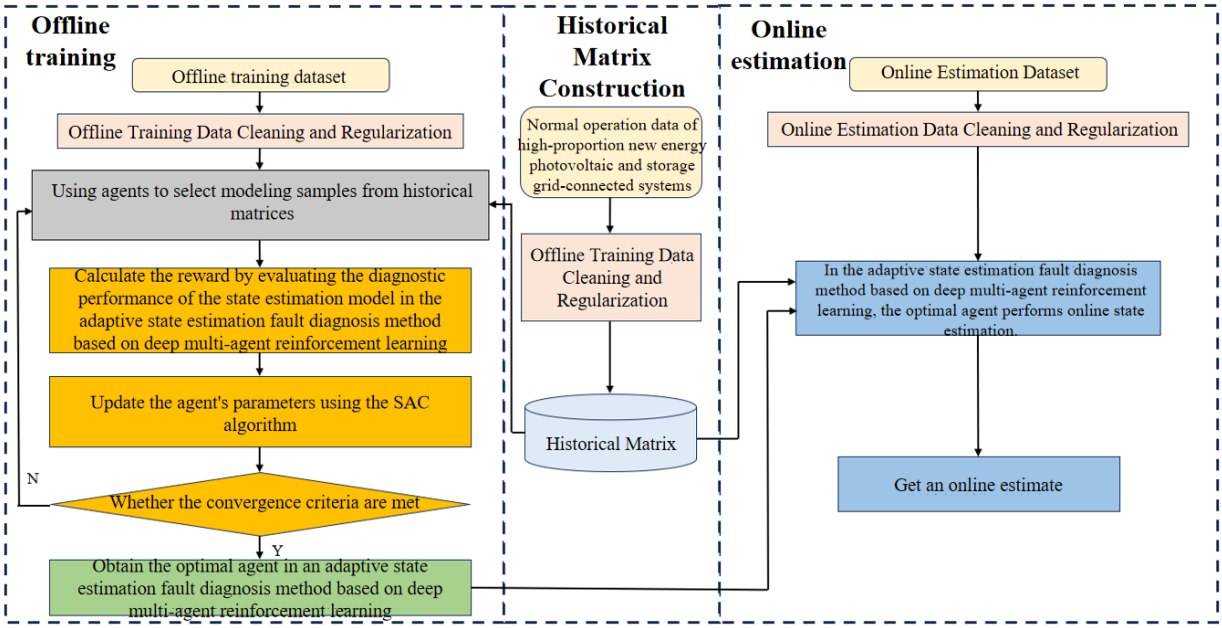


Figure 6. Overall framework of the proposed method

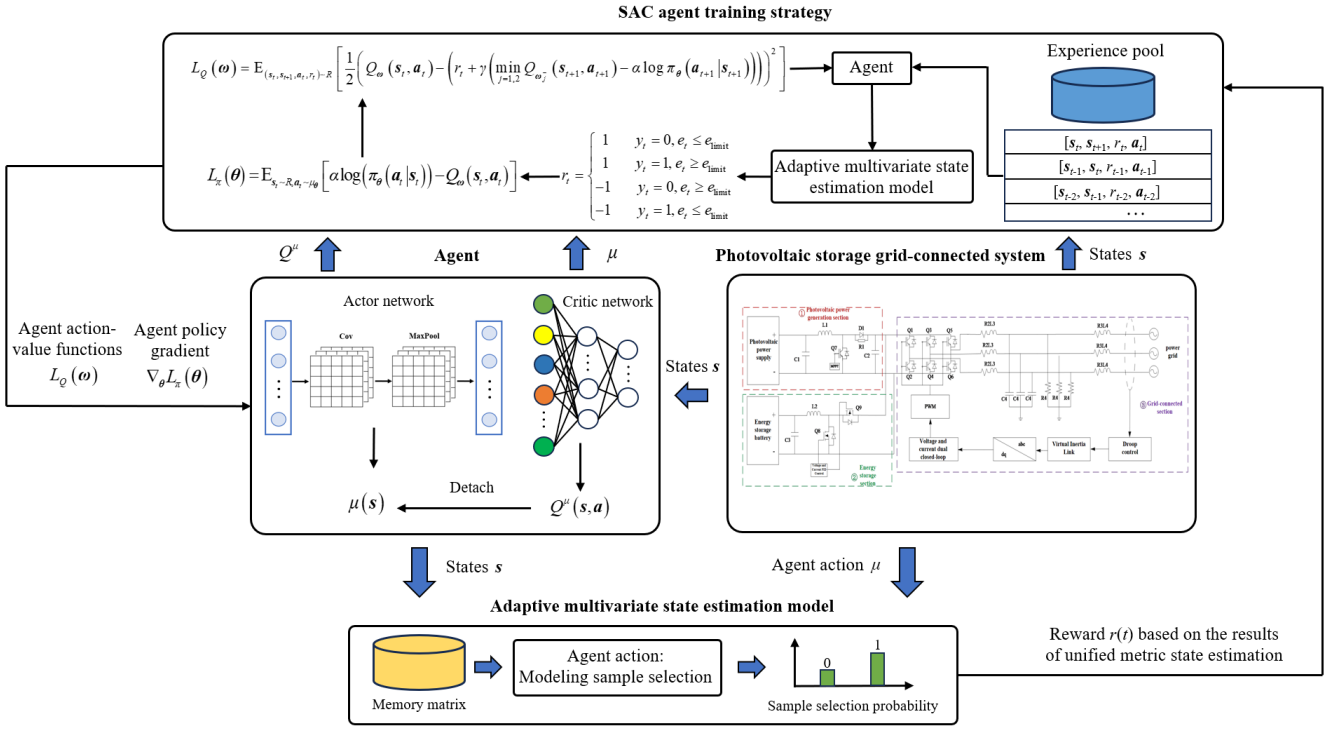


Figure 7. Training process of the proposed method

The loss function for the critic network is:

$$L_Q(\omega)=\mathrm{E}_{(s_t,s_{t+1},a_t,r_t)\square R}\left[\frac{1}{2}\left(Q_\omega(s_t,a_t)-\left(r_t+\gamma\left(\min_{j=1,2}Q_{\omega_j^-}(s_{t+1},a_{t+1})-\alpha\log\pi_\theta(a_{t+1}|s_{t+1})\right)\right)^2\right] \quad (10)$$

Among these, Q_{ω^-} is the goal Q-network with parameter ω^-, which enhances the stability of the learning progress. γ is the decay parameter to balance the ratio between long-term and short-term reward returns.

Given the previous analysis, active model selection can be characterised as an MDP. Specifically, the key components of the MDP are described next:

1) State: The agent's state is denoted as s. In the adaptive state estimation fault detection algorithm using deep multi-agent reinforcement training, it is assumed that the agent can monitor every states $\boldsymbol{s}_t$ (t=1, 2,…, T) of the high-proportion renewable energy photovoltaic-storage grid-connected system. The combination of process variables within the system and the corresponding labels at time t is how each state is described, expressed as follows:

$$\boldsymbol{s}_t=\boldsymbol{X}\oplus Y \quad (11)$$

where $\boldsymbol{X}$ and Y represent the inputs and labels of the testing data.

To embed $\boldsymbol{s}_t$ within the reinforcement learning framework, a convolutional neuronet is employed to produce deep representations $\boldsymbol{h}_t$.

$$\boldsymbol{h}_t=Relu\left(W*Relu\left(\boldsymbol{s}_t\right)\right) \quad (12)$$

Here, W denotes the parameters of the CNN, * represents the conjugate layer, and Relu(·) denotes the activation function.

2) Action: Action $\boldsymbol{a}$ is derived from the agent's strategy π. Within the context of adaptive state estimation, fault detection using deep reinforcement learning involves the agent adaptively selecting representative samples from the history matrix for state detection. The model is the actor network output, which produces an n-D spatial output (where n is the size of the history matrix). Each dimension's value ranges between 0-1, representing the probability of sample selection. The actor network comprises three completely connected layers. The action at time step t can be expressed as:

$$\boldsymbol{a}_t=\mathrm{sigmoid}\left(\pi\left(\boldsymbol{h}_t\right)\right) \quad (13)$$

The sigmoid function is a normalization function that maps outputs to the range [0, 1]. $\boldsymbol{a}_t$ is a vector with n elements, where $a_{i,t}$ represents the i-th element of $\boldsymbol{a}_t$, denoting the probability that the i-th sample in the history matrix was selected. Therefore, prior to performing state estimation on grid-connected systems with high proportions of new energy photovoltaic storage, the intelligent agent will adaptively select a model template. Specifically, it will choose the template from the historical matrix that closest matches the system operating state to create the state prediction template.

3) Reward: For adaptive state estimation fault diagnosis methods using deep multi-agent reinforcement learning, the objective is to maximize the fault diagnosis accuracy of the state estimation algorithm by adaptively selecting suitable training templates from historical matrices. Based on this, we design a fault diagnosis performance-based unified metric to calculate reward returns. The reward at time step T is computed in the way:

$$r_t=\begin{cases}1 & y_t=0, e_t\le e_{\text{limit}}\\ 1 & y_t=1, e_t\ge e_{\text{limit}}\\ -1 & y_t=0, e_t\ge e_{\text{limit}}\\ -1 & y_t=1, e_t\le e_{\text{limit}}\end{cases} \quad (14)$$

$$e_t=f_{\text{est}}\left(\boldsymbol{D}_t\right) \quad (15)$$

Here, e_t denotes the output failure of the state estimation model $f_{\text{est}}(\cdot)$ at time t, $\boldsymbol{D}_t$ represents the updated history matrix at time t, e_{limit} is the fault diagnosis control limit, and y_t indicates the system state of the high-proportion renewable energy photovoltaic storage grid-connected system at time t. y_t=0 indicates normal system operation, while y_t=1 indicates a system fault. Rewards are given during training for correctly detecting normal and fault samples. Conversely, penalties are imposed if normal samples are misclassified as faults or if fault samples are missed.

After defining all components, the SAC algorithm is employed for recurrent training.

3.2.2 Online Estimation

After offline learning, stochastic action π^* can be observed from the deterministic agent. Given an observation sample $\boldsymbol{x}_{\text{obs}}$ from the system under test, the optimal agent will dynamically select an appropriate model sample dataset from the history matrix via Equation (16).

$$\boldsymbol{a}^*=\mathrm{sigmoid}\left(\pi^*\left(\boldsymbol{x}_{\text{obs}}\right)\right) \quad (15)$$

For the best action $\boldsymbol{a}^*$, the error evaluation is computed by:

$$e_{\text{obs}}=f_{\text{est}}\left(\boldsymbol{D}^*\right) \quad (16)$$

Among these, $\boldsymbol{D}^*$ selects the optimal template dataset from the historical matrix $\boldsymbol{D}$.

Finally, the estimated error e_{obs} is obtained and compared against the fault diagnosis control limit e_{limit}. If e_{obs} exceeds e_{limit}, the method rises an alarm signal indicating a system fault.

4. Results and Discussion

4.1 Comparative Methods and Diagnostic Criteria

The aim is to show the advantage of the suggested approach in fault diagnosis; two advanced diagnostic approaches, LDKPCA and MMO-MSE, are employed as comparison methods. Table 1 lists the hyperparameters for these methods.

LDKPCA: This is a nonlinear static fault detection approach devised by Gao et al. [21]. It integrates the strengths of nonlinear dynamic principal component analysis and local keep projection.

MMO-MSET: This is an advanced nonparametric fault diagnosis approach proposed by Lv et al. [20] employing a historical matrix optimization strategy for computational

time reduction and fault diagnosis performance enhancement.

Table 1. Hyperparameters of different methods.

Method	hyperparameters	References
LDKPCA	Weighting factor: γ_1=0.5; γ_2=0.5; γ_3=0.5	[21]
MMO-MSET	Nuclear function: RBF	[20]
This method	Networks in Deep Reinforcement Learning Frameworks: Convolutional Neural Network Layers, Fully Connected Layers, Number of Selected Localization Models: K=15	-

Fault alarm rate (FAR) and fault detection rate (FDR) are adopted as evaluation metrics:

$$\mathrm{FDR}=\frac{N\left(e>e_{\mathrm{limit}}\,|\,z=1\right)}{N\left(z=1\right)}\times 100\% \tag{17}$$

$$\mathrm{FAR}=\frac{N\left(e>e_{\mathrm{limit}}\,|\,z=0\right)}{N\left(z=0\right)}\times 100\% \tag{18}$$

Here, $N()$ denotes the set of samples fulfilling the conditions within the parentheses. $N(z=0)$ represents the number of normal samples, while $N(z=1)$ indicates the quantity of faulty samples. Additionally, $N\left(e>e_{\mathrm{limit}}\,|\,z=1\right)$ signifies the quantity of faulty samples correctly diagnosed, and $N\left(e>e_{\mathrm{limit}}\,|\,z=0\right)$ represents the quantity of normal samples falsely identified as faulty.

4.2 System Fault Description

To assess the fault diagnosis accuracy of various approaches, three types of typical faults were designed on the generation side, storage side, and grid-connection side respectively. These faults simulate issues arising from performance degradation due to long-term operation, leading to failures in various system components during operation. The faults are described as follows:

Source-side fault: Hot-spot effect in photovoltaic modules, where photovoltaic cells become energy-consuming loads during power generation, leading to localized overheating.

Storage-side faults: Capacity degradation. After prolonged operation, the capacity and internal resistance variations among individual cells within the battery pack increase, adversely affecting overall performance and usable capacity.

Grid-connected side fault: Single-phase open circuit, causing the system to operate in an unbalanced state.

4.3 Model Training and Hyperparameter Selection

In fact, the performance of the agent is closely linked to the actor and critic networks within the deep reinforcement learning framework. To identify optimal network hyperparameters, the neurons and convolutional layers number in the actor network were varied, as shown in Figure 8(a). As the agent network was trained with 3 convolutional filters and 32 features, fairly good rewards were achieved. Therefore, fixing this optimum agent network, we adjusted the size of completely linked levels and the quantity of neurons in the critic network, as shown in Figure 8(b). The critic network performed best when the number of layers and neurons were set to 2 and 64, respectively. To illustrate the reinforcement learning process, Figure 9 displays the reward returns during training, using a training dataset of 2000 normal samples and 2000 fault samples. According to the definition of reward function (14), the maximum reward achievable when all samples are successfully diagnosed is 4000. Figure 9 shows that the reward increases steadily throughout training, rising more rapidly during the first training phase. As the model converges, the reward gradually stabilizes near 4000.

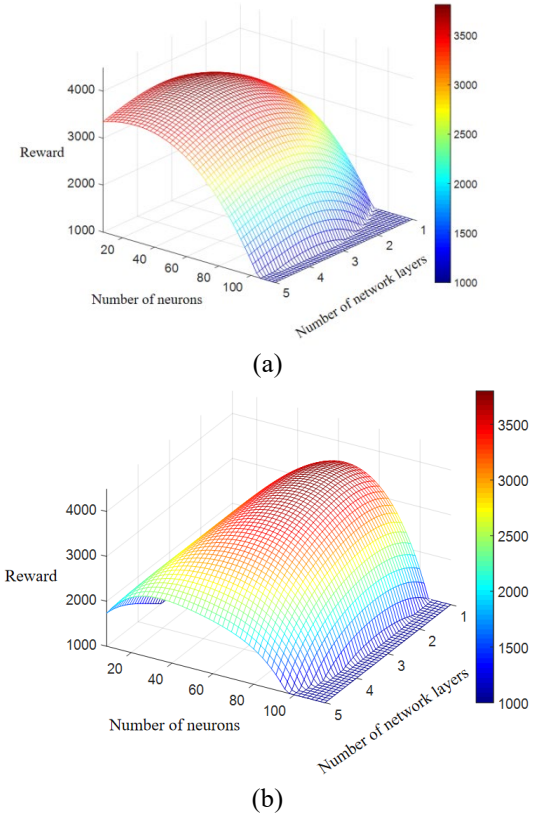


Figure 8. Key parameters selection (a) Actor network; (b) Critic network

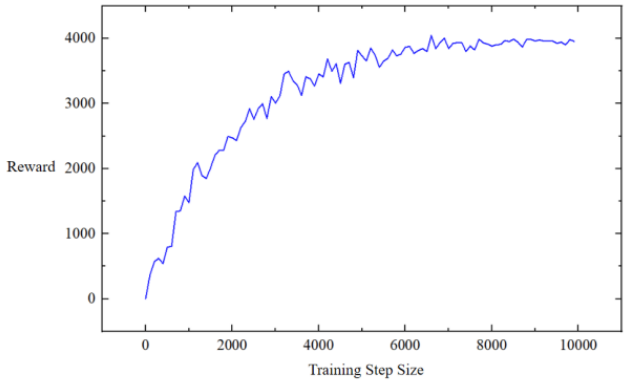


Figure 9. The reward during the training process

4.4 Fault Diagnosis Result Verification

Tables 2 and 3 present the fault diagnosis results for the two methods. The suggested approach outperforms the other two enhanced fault detection approaches in both FDR and FAR. Compared to LDKPCA and MMO-MSET, the average FDR of the proposed method improved by 2.16% and 0.76%, while the overall FAR declined by 2.88% and 1.13%. Furthermore, the introduction of agents in deep reinforcement learning enables the proposed method to adaptively select optimal model samples for state estimation, thereby enhancing fault diagnosis performance. Under three typical fault scenarios: source-side, storage-side, and grid-connected measurement, the suggested approach attains the most effective FAR and FDR. These fault diagnosis results in high-penetration renewable energy photovoltaic-storage grid-connected systems demonstrate the superiority of the proposed approach. Furthermore, to better illustrate the results, Figures 10-12 further present the diagnostic outcomes of different methods under various faults.

Table 2. FDR of different methods on three typical system faults.

	LDKPCA	MMO-MSET	The proposed method
Source-side fault	95.67%	98.33%	99.33%
Storage-side fault	97.20%	98.40%	99.07%
Grid-connected side fault	96.20%	98.20%	98.80%
Average	96.91%	98.31%	99.07%

Table 3. FAR of different methods on three typical system faults.

	LDKPCA	MMO-MSET	The proposed method
Source-side fault	7.83%	1.83%	1.17%
Storage-side fault	4.93%	2.53%	1.20%
Grid-connected side fault	3.40%	2.20%	0.80%
Average	3.94%	2.19%	1.06%

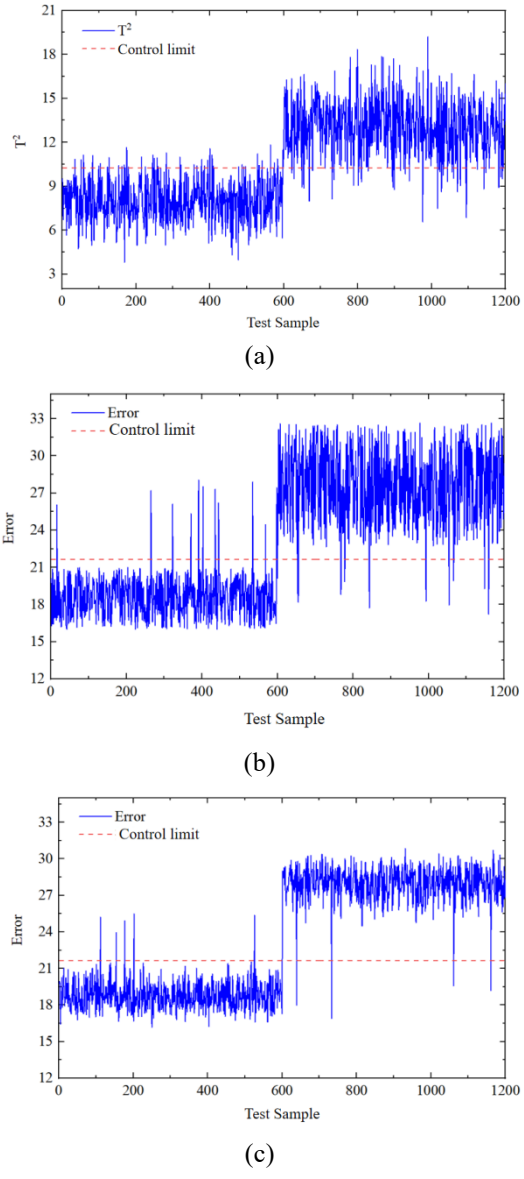


Figure 10. The source side fault diagnosis results of different methods (a)LDKPCA; (b)MMO-MSET; (c)proposed method

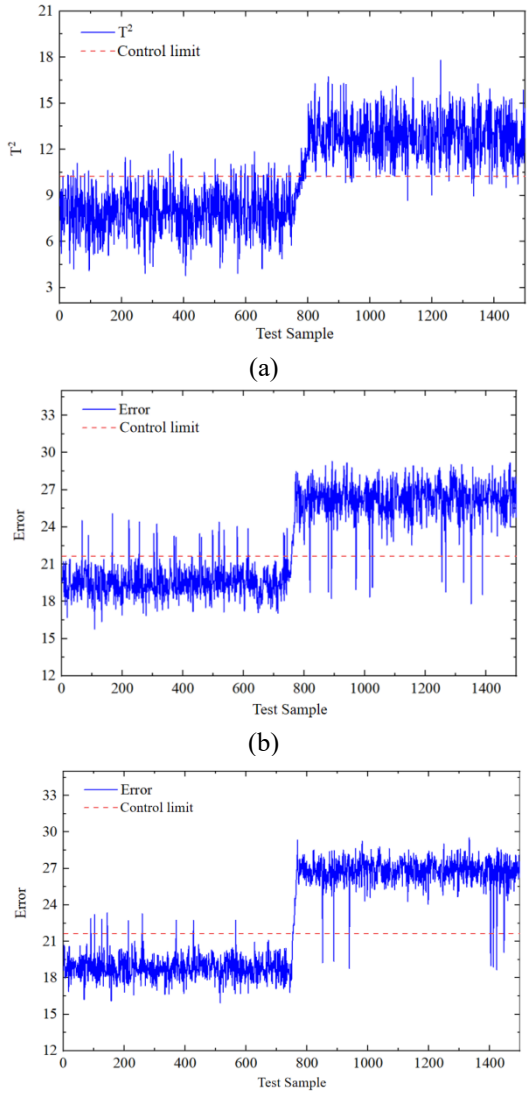
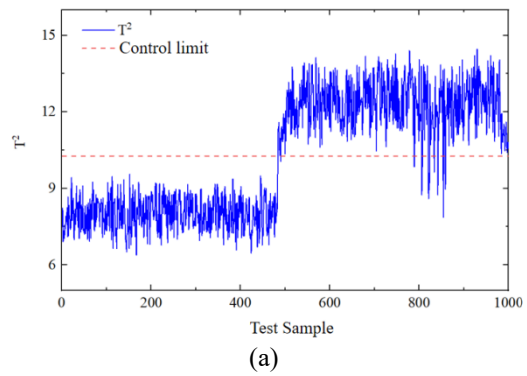


Figure 11. The energy storage side fault diagnosis results of different methods (a)LDKPCA; (b)MMO-MSET; (c)proposed method



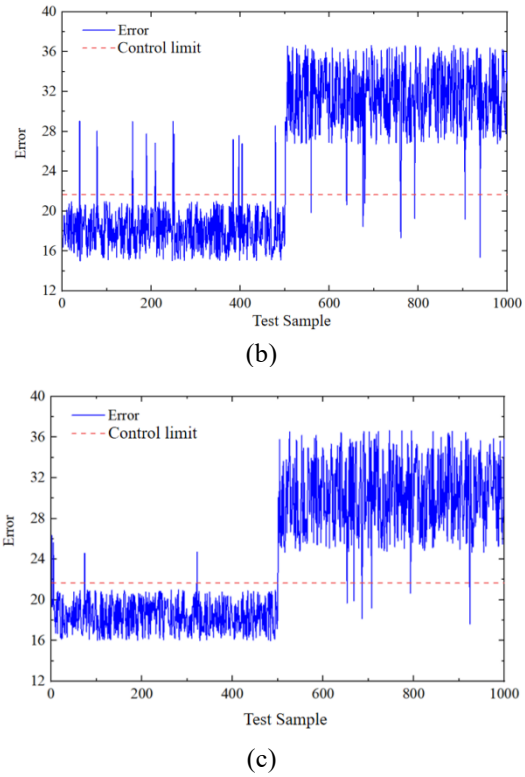


Figure 12. The grid side fault diagnosis results of different methods (a)LDKPCA; (b)MMO-MSET; (c)proposed method

To further validate the effectiveness of adaptive modeling sample selection, multiple samples were collected for state estimation across three distinct operation scenarios of high-penetration renewable energy photovoltaic-storage grid-connected systems: low renewable energy penetration, medium renewable energy penetration, and high renewable energy penetration. Figure 13 illustrates the results of adaptive model selection. The algorithm in the proposed approach can dynamically adapt to the actual operation state of the system to select the appropriate model for updating the history matrix. Consequently, leveraging the agent's adaptive update strategy, the presented approach demonstrates excellent fault detection performance. Specifically, it selects samples most relevant to the present state of the high-penetration renewable energy photovoltaic-storage grid-connected system for estimation while filtering out redundant and irrelevant samples.

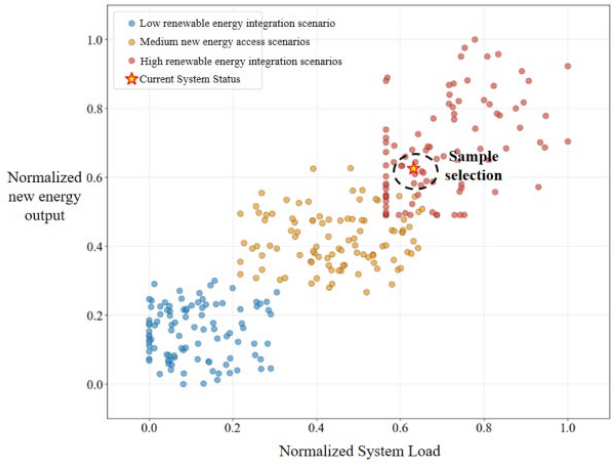


Figure 13. Modeling sample adaption selection results

5. Conclusion

1) Compared with traditional fault diagnosis methods, this proposed fault detection approach using deep reinforcement training achieves adaptive historical matrix update through intelligent agent decision-making mechanism, and can dynamically track the complex state changes of the photovoltaic-storage grid-connected system under high proportion of new energy access. The random fluctuations in photovoltaic output and the frequent charging and discharging switching of energy storage systems during day night cycles result in significant time-varying operating conditions, making it difficult for traditional methods to effectively track such dynamic characteristics. This method uses intelligent agents to perceive the real-time operating status of the system and autonomously adjust the modeling sample set accordingly, enabling the state estimation model to continuously adapt to the rapid state transition of the system under the dominance of new energy, thereby significantly improving the reliability and efficiency of fault detection.

2) In comparative experiments under multiple typical operating scenarios, the performance was compared with two advanced fault diagnosis methods, LDKPCA and MMO-MSET. The findings indicated that the suggested approach enhanced the fault detection rate (FDR) by an average of 2.16% and 0.76%, respectively, while reducing the failure alarm rate (FAR) by an average of 2.88% and 1.13%, respectively. Such performance improvement fully demonstrates the proposed method's superiority in dealing with complex and time-varying operating conditions in high proportion new energy solar storage systems. The strong volatility and multi-mode operation characteristics of photovoltaics and energy storage pose higher requirements for the reliability and robustness of fault diagnosis. The approach presented in this study achieves precise tracking and rapid response of system status through reinforcement learning mechanism, effectively reducing misjudgements and omissions caused by uncertainty in new energy output and energy storage state switching, and providing more reliable technical support for the stable operation of high proportion new energy grid connected systems.

Acknowledgements.

This work was supported by the Research on Safety Enhancement and Control Strategies for Smart Distribution Networks with High Proportion of Distributed Renewable Energy Sources (SGSDJY00ZYJS2500232).

References

[1] Hsu HW, Fan ZW. Multi-faceted procurement with mixed integer linear programming for corporate 100 % renewable energy goal. Energy, 2025, 320.
[2] Efthymiou C, Khan A, Assimakopoulos MN, Santamouris M. Urban pavement-mounted photovoltaics as renewable energy systems for energy generation and microclimate control. Solar energy, 2025, 299.
[3] Sanda MG, Emam M, Ookawara S, Hassan H. Techno-enviro-economic evaluation of on-grid and off-grid hybrid photovoltaics and vertical axis wind turbines system with battery storage for street lighting application. Journal of cleaner production, 2025, 491.
[4] Wesonga R, Tutesigensi A, Moodley K. Advancing net zero carbon construction: A techno-economic and environmental analysis of onsite microgrids and prosumer energy adoption. Applied energy, 2025, 398.
[5] Bai WW, Fang YY, Li XY, Li WW. Additional self-disturbance rejection control strategy for battery storage in VSG-based wind power grid-connected system.New energy power control technology, 2024, 46(6).
[6] Jiang SQ, Zhang HF, Fu G, Xin YC, Wang LX. Coordinated Control Strategies for Enhancing Frequency Stability of Photovoltaic and Storage Networking Systems. Electric Power Construction, 2025, 46(8).
[7] Zhang ZW, Jiao ZH, Li YJ, Shao MY, Dai XJ. Intelligent fault diagnosis of bearings driven by double-level data fusion based on multichannel sample fusion and feature fusion under time-varying speed conditions. Reliab Eng Syst Saf 2024, 251.
[8] Xue JY, Zhang TS, Ye H. KPI-oriented process monitoring based on causal-weighted partial least squares. Inform Sciences 2025, 689.
[9] Obanya PO, Coetzer RLJ, Olivier CP, Verster T. Variable contribution analysis in multivariate process monitoring using permutation entropy. Comput Ind Eng 2024, 190.
[10] Chen C, Wang T, Lu KJ, Liu Y, Cheng LL. Compact convolutional transformers- generative adversarial network for compound fault diagnosis of industrial robot. Eng Appl Artif Intell 2024, 138.
[11] Yang C, Li Y, Chen QJ. A novel two-Stage fault-detection method based on constrained RVM and integrating LDA with minimax probability machine. IEEE Trans Ind Informat 2023, 19(3).
[12] Jin CC, Chen X. An end-to-end framework combining time-frequency expert knowledge and modified transformer networks for vibration signal classification. Expert Syst Appl 2021, 171.

[13] Dong YN, Qin SJ. A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. J Process Control 2018, 67.
[14] Harrou F, Kini KR, Madakyaru M, Sun Y. Sensor fault detection and diagnosis in photovoltaic systems using Hellinger Distance and Individual Conditional Expectation analysis. Solar energy 2025, 298.
[15] Xue JY, Zhang TS, Ye H. KPI-oriented process monitoring based on causal-weighted partial least squares. Inform Sciences 2025, 689: 121470.
[16] Liu HB, Yang J, Zhang YC, Yang C. Monitoring of wastewater treatment processes using dynamic concurrent kernel partial least squares. Process Safety and Environmental Protection 2021, 147: 274-282.
[17] Su SY, Sun YC, Li LB, Peng C, Zhang H, Zhang TT. Risk warning for aircraft bleed air system with multivariate state estimation technique. J Aerosp Inform Syst 2022.
[18] Zhang S, Wang F, Tan S, Wang S, Chang Y. Novel Monitoring Strategy Combining the Advantages of the Multiple Modeling Strategy and Gaussian Mixture Model for Multimode Processes. Ind Eng Chem Res 2015, 54(47).
[19] Gao GL, Zhong YM, Gao SS, Gao BB. Double-Channel Sequential Probability Ratio Test for Failure Detection in Multisensor Integrated Systems. IEEE Trans Instrum Meas 2021, 70.
[20] Lv Y, Fang F, Yang TT, Romero CE. An early fault detection method for induced draft fans based on MSET with informative memory matrix selection. ISA Trans 2020, 102.
[21] Gao XY, Zhang Y, Zhou JF. Improved dynamic kernel PCA based on local preserving projections and its application for electric submersible pump fault diagnosis. Canadian Journal of Chemica Engineering. 2023, 101: 8.