Modeling and Optimization of Power Grid Intelligent Restoration Strategy Combining Diffusion Models and Reinforcement Learning

Jun Li
School of Software Engineering, Jiangxi University of Software Professional Technology, Nanchang 330041, Jiangxi, China
Corresponding author: Jun Li (e-mail: 13330077138@163.com)

ABSTRACT With the increasing proportion of renewable energy integration in modern power systems, grid fault recovery faces greater complexity and uncertainty, posing significant challenges for intelligent energy transmission infrastructures and reliable operation in electromagnetic power networks. To address these issues, this paper proposes a collaborative decision-making model (CDM-RL) that combines diffusion models (DM) and reinforcement learning (RL) for intelligent power grid restoration. The proposed framework generates physically feasible fault scenarios through a denoising diffusion probabilistic model (DDPM) and integrates graph neural networks (GNN) with proximal policy optimization (PPO) to achieve efficient restoration control under complex operating conditions. An alternating training mechanism is further introduced to enhance the generalization capability of the model across diverse fault scenarios. Experimental evaluation on the IEEE 33-bus system demonstrates that CDM-RL significantly outperforms conventional approaches, achieving an initial power restoration time of 8.9 s, a load recovery rate of 93.1%, and an average of only 3.9 switching operations while maintaining superior cross-scenario stability of 85.3%. The proposed AI-driven framework improves grid self-healing capability and provides an effective technical pathway for enhancing the resilience and operational reliability of intelligent electromagnetic energy transmission and distribution systems.

INDEX TERMS power grid fault recovery, diffusion model, reinforcement learning, intelligent energy systems, electromagnetic power networks

I. INTRODUCTION

As the proportion of renewable energy access in new power systems continues to increase [1-2], the grid structure is becoming increasingly complex[3-4], and the uncertainty of operating status has increased significantly[5-6], which creates particularly acute challenges for sectors like the industry that rely on an extremely stable and reliable power supply for continuous, precision manufacturing operations. Traditional power grid fault recovery strategies mostly rely on fixed rules or single optimization methods [7-8]. When faced with diverse and dynamically changing fault scenarios, it is often difficult to achieve efficient and stable recovery control. Especially in the context of a high proportion of renewable energy access, the volatility of grid operation status increases and the fault propagation path becomes more complex, which poses severe challenges to the real-time and adaptability of intelligent recovery strategies [9-10]. Therefore, it is urgent to explore new technical means to enhance the self-healing capability of the power grid and ensure the safe and stable operation of the power system.

This paper focuses on the key issues in the grid fault restoration process, including grid restoration efficiency[11], grid load restoration quality[12-13] and restoration cost[14]. Specifically, traditional power grid optimization strategies have significant limitations when dealing with complex topologies and dynamic fault evolution, and it is difficult to balance rapid response with physical feasibility constraints [15-17]. Therefore, how to construct an intelligent recovery strategy that is both physically feasible and highly adaptable has become a core issue in current research, particularly when considering the need for quick, precise, and reliable power restoration to high-priority industrial loads.

In recent years, Reinforcement Learning (RL) models have been widely used in power grid fault restoration tasks [18-19], but they still have certain defects. For example, Liao et al. [20] comprehensively analyzed the key applications of Graph Neural Networks (GNNs) in power systems, such as fault scenario applications, time series prediction, power flow calculation, and data generation. At the same time, the defects of GNN were also pointed out, such as the inability to effectively process graph structured data in dynamic structures; Pinciroli et al. [21] proposed an RL model based on Proximal Policy Optimization (PPO), which effectively solved the problems of operation and maintenance optimization in power systems and had good policy stability and adaptability. However, the study did not verify the generalization ability of the model. In addition, some studies use traditional optimization models such as Mixed Integer Linear Programming (MILP) to model recovery strategies. Although these models can guarantee local optimal solutions, they have poor adaptability in uncertain environments and are difficult to meet real-time requirements. For example, Er-Rahmadi et al. [22] proposed a data-driven MILP model that effectively improved the accuracy and computational efficiency of power grid fault diagnosis. However, the operating conditions simulated by the model were relatively ideal and did not fully consider the common nonlinear physical constraints and dynamic uncertainties in the power grid, which limited its direct applicability in complex power systems.The strategy proposed in this paper focuses on network reconfiguration, that is, by optimizing the combination of normally closed/normally open switches to form a new radial power supply topology, so as to isolate faults and restore power supply to non-faulty areas; it does not involve the operation of voltage regulation devices or the dispatching of distributed generation output.

II. ALGORITHM DESIGN

In response to the challenges of scenario complexity and strategy adaptability faced by power grid fault recovery, this study constructs a collaborative optimization model CDM-RL that combines diffusion models and RL models, and designs it from aspects such as fault modeling, physical constraint embedding, decision architecture, training mechanism, environment simulation, and collaborative strategy. Firstly, this paper adopts DDPM with U-Net (U-Net: Convolutional Networks for Biomedical Image Segmentation) architecture, extracts topological features through graph convolutional network (GCN), combines the fast decoupled load flow (FDLF) algorithm to quantify voltage over-limit and branch overload constraints, and incorporates diffusion loss to form an enhanced objective function. Secondly, this paper constructs an Actor-Critic decision-making architecture that combines GAT and PPO, and uses a multi-head attention mechanism to aggregate topological information. Subsequently, this paper designs a multi-objective reward function with dynamic weight adjustment to collaboratively optimize the recovery time, power supply coverage, and switching cost. Furthermore, this paper establishes an alternating training framework. The Diffusion Models priority stage strengthens the generation ability through online difficult example mining. The RL model priority stage introduces adversarial perturbation samples and gradient coordination strategies, combined with dual time scale update rules to ensure training stability. Next, the simulation environment builds a three-level architecture based on OpenDSS (Open Distribution System Simulator) and Gym (Gym Interface for Reinforcement Learning), integrating sliding window z-score normalization and binary action space to form a dynamic interactive closed loop under the constraints of physical laws. Finally, each part achieves deep collaboration through residual constraints of power flow equations, minimization of gradient inner products, and cross-stage parameter freezing strategies to achieve efficient and intelligent decision-making. The overall algorithm design framework of this study is shown in Figure 1.

DESIGN OF FAULT SCENARIO MODELING MODULE

This module uses DDPM with U-Net structure to model and generate power grid fault scenarios. This module aims to construct diverse fault samples with physical feasibility to support subsequent recovery strategy training and evaluation.

The input data includes historical grid operation data, topology information and typical fault records. The original data is standardized and used as the input features of the diffusion model. In the diffusion process, the forward diffusion process is defined as gradually adding Gaussian noise to the original grid state x0, generating an intermediate state sequence x1, x2, $\cdots$, xT, satisfying:

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \tag{1}$$

In formula (1), β_t is the preset noise scheduling parameter, which controls the noise intensity at each time step.

The back diffusion process is parameterized by a neural network, and the goal is to approximate the real data distribution by learning the denoising function $\epsilon_\theta(x_t, t)$:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\right) \tag{2}$$

The training goal is to minimize the mean square error between the predicted noise ϵ_θ and the actual added noise ϵ:

$$\mathcal{L} = E_{x_0,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right] \tag{3}$$

In order to ensure the physical feasibility of the generated fault scenarios, the power flow equation constraint is introduced as an auxiliary loss term. Specifically, after each generation, the FDLF algorithm is called to verify the electrical consistency of the generated samples, and the degree of violation of physical constraints such as voltage limit and branch overload is quantified as a penalty term $\mathcal{L}_{phys}$, which is weighted and fused with the original diffusion loss:

$$\mathcal{L}_{total} = \mathcal{L} + \lambda \cdot \mathcal{L}_{phys} \tag{4}$$

In formula (4), λ is the weight coefficient.

In the generation phase, the initial state $x_T \sim \mathcal{N}(0, I)$ is sampled from the standard normal distribution, and the

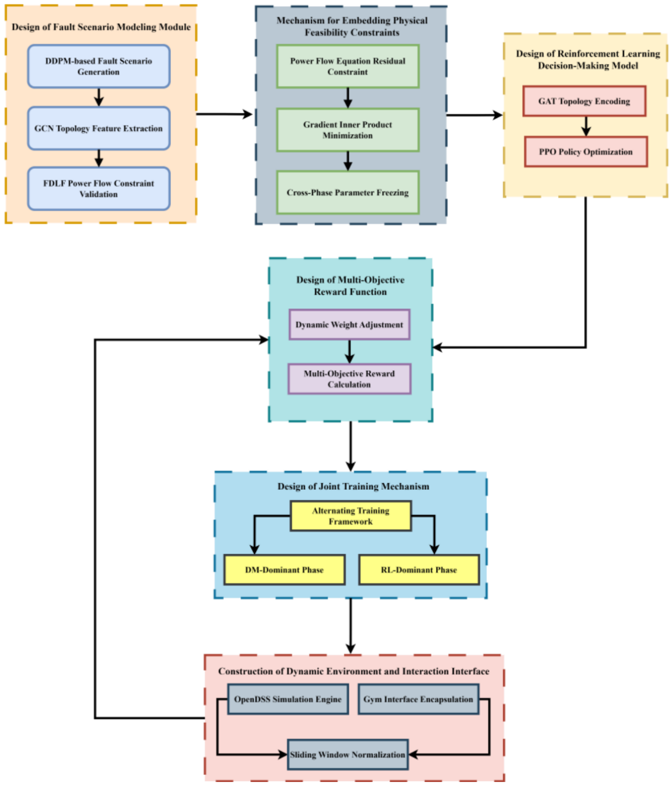

FIGURE 1. Algorithm design framework

denoising steps are performed sequentially to output the complete grid state x_0'. Then, combined with the preset fault mode, local disturbances are injected into the generated state to simulate the real fault evolution process. The final output fault scenario includes key variables such as node voltage, branch flow, and switch status for use by subsequent intelligent decision-making modules.

PHYSICAL FEASIBILITY CONSTRAINT EMBEDDING MECHANISM

To ensure that the fault scenarios generated by the diffusion model meet the physical laws of power grid operation, this paper introduces a physical guided diffusion mechanism in the denoising process and embeds the power flow equation into the loss function as a regularization term. This mechanism achieves boundary control of key variables such as voltage and power by constraining the noise prediction direction, thereby improving the engineering usability of the generated samples.

To ensure that the generated samples strictly conform to the physical laws of the power grid, this paper explicitly embeds the power flow equation residuals and the degree of violation of the operating boundary into the loss function of the diffusion model. Specifically, for the power grid state output in each denoising step, the Fast Decoupled Power Flow (FDLF) algorithm is first called to calculate the power balance residuals of all nodes and the branch power flow values; if the node voltage exceeds [0.95, 1.05] pu or the line power flow exceeds its thermal stability limit, the excess is added to the total loss in the form of a secondary penalty. This mechanism enables the diffusion model not only to learn the data distribution, but also to actively correct the generation direction that violates physical constraints through gradient backpropagation, thereby ensuring the consistency of voltage, power flow and topological logic at the sample level. At each step $t \to t-1$ of the back-diffusion process, the noise $\epsilon_\theta(x_t, t)$ predicted by the neural network is used to reconstruct the state x_{t-1}:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z \tag{5}$$

In formula (5), $z \sim \mathcal{N}(0, I)$, $\alpha_t = 1-\beta_t$, $\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s$, and σ_t are sampling noise coefficients. The residual function $r(x; A)$ of the power flow equation is constructed in the form of the standard AC power flow equation group:

$$P_i = \sum_{j=1}^{N} V_i V_j \left(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}\right) \tag{6}$$

$$Q_i = \sum_{j=1}^{N} V_i V_j \left(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}\right) \tag{7}$$

The power flow error term is defined as:

$$\mathcal{L}_{pf}(x) = \|r(x; A)\|_2^2 \tag{8}$$

Considering the upper and lower limits of voltage amplitude $V_{\min} \le V_i \le V_{\max}$ and the line transmission capacity limit $|S_{ij}| \leq S_{ij}^{\max}$, the boundary penalty term is constructed:

$$\begin{aligned}\mathcal{L}_{bound}(x) = \sum_{i=1}^{N} \big[& \max(0, V_{\min} - V_i)^2 \\ & + \max(0, V_i - V_{\max})^2 \big] \\ & + \sum_{(i,j)\in E} \max\left(0, |S_{ij}| - S_{ij}^{\max}\right)^2\end{aligned} \tag{9}$$

The above two items are added to the diffusion model loss function to form an enhanced objective function:

$$\begin{aligned}\mathcal{L}_{total} =& E_{x_0,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right] \\ & + \lambda_1 \cdot \mathcal{L}_{pf}(x_t) + \lambda_2 \cdot \mathcal{L}_{bound}(x_t)\end{aligned} \tag{10}$$

In formula (10), λ_1 and λ_2 are weight hyperparameters.

RL DECISION MODEL DESIGN

The RL decision model is an intelligent decision model for power grid fault recovery based on GNN and PPO. The model aims to extract structured features from complex, high-dimensional power grid topologies and make efficient and feasible switching operation decisions in a dynamic environment.

This model models the power grid as a heterogeneous graph, where node feature vectors include voltage amplitude (per unit), active/reactive power injection, and load priority identifiers, while edge features include line impedance, current power flow value, and switch state (0/1). After aggregating neighbor information through a multi-layer graph attention network (GAT), the resulting node embedding not only encodes local electrical states but also incorporates global topological connectivity. This graph embedding, after global pooling and MLP mapping, forms the state input of the PPO agent, directly determining the switch action probability distribution output by its Actor network and the state value estimated by the Critic network. This ensures that the decision-making respects both the power grid topology constraints and prioritizes the restoration of high-priority industrial loads.

The input grid state is represented as a graph structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V}$ represents the bus and the edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents the line connection relationship. Each node $v_i \in \mathcal{V}$ corresponds to a feature vector $h_i^{(0)} \in R^{d_h}$, which contains electrical measurement information such as line voltage amplitude V_i, injected power P_i, Q_i, etc.; each edge $e_{ij} \in E$ features $m_{ij}^{(0)} \in R^{d_m}$ describing line parameters and current power flow status. The adjacency matrix $A \in \{0,1\}^{N\times N}$ represents the topological connection.

A multi-layer GAT is used as an encoder to extract the topological features of the power grid. The l-th layer GAT is defined as follows:

$$e_{ij}^{(l)} = \text{LeakyReLU}\left(a^T\left[W^{(l)}h_i^{(l)} \| W^{(l)}h_j^{(l)}\right]\right) \tag{11}$$

$$\alpha_{ij}^{(l)} = \frac{\exp\left(e_{ij}^{(l)}\right)}{\sum_{k\in N(i)} \exp\left(e_{ik}^{(l)}\right)} \tag{12}$$

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in N(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}\right) \tag{13}$$

In formula (11), $a \in R^{2d}$ is the attention coefficient vector, $W^{(l)} \in R^{d\times d}$ is the trainable weight matrix. In formula (12), $N(i)$ represents the neighbor set of node i. In formula (13), $\sigma(\cdot)$ is the activation function. By stacking multiple layers of GAT, the final output node embedding $H^{(L)} \in R^{N\times d}$ integrates local topology and global operation status information.

The output graph embedding is used as the state input of the PPO Agent. The state space is defined as:

$$s_t = \text{MLP}_{state}\left(\text{GlobalPool}\left(H^{(L)}\right)\right) \in R^{d_s} \tag{14}$$

In formula (14), GlobalPool represents the graph-level pooling operation (such as mean or maximum pooling), and MLP (Multi-Layer Perceptron) represents the mapping to the policy network input dimension. Actor-Critic architecture is used to achieve policy optimization. The probability distribution of Actor output action is:

$$\pi_\theta(a|s) = \text{Softmax}\left(\text{MLP}_\theta(s)\right) \tag{15}$$

The critic estimates the state value function as:

$$V_\phi(s) = \text{MLP}_\phi(s) \tag{16}$$

The PPO objective function is defined as:

$$\mathcal{L}_{PPO}(\theta) = E_t\left[\min\left(\rho_t(\theta)A_t, \text{clip}(\rho_t(\theta), 1-\epsilon, 1+\epsilon)A_t\right)\right] \tag{17}$$

In formula (17), $\rho_t(\theta)$ is the probability ratio, A_t is the generalized advantage estimate, and ϵ is the clipping range hyperparameter.

DESIGN OF MULTI-OBJECTIVE REWARD FUNCTION

In order to guide the RL decision model to take into account multiple optimization objectives during the power grid fault recovery process, this paper designs a weighted multi-objective reward function, which is as follows:

$$R_t = w_1 \cdot R_{time,t} + w_2 \cdot R_{coverage,t} - w_3 \cdot R_{switch,t} \tag{18}$$

In formula (18), R_t is the comprehensive reward value of the tth step, $R_{time,t}$, $R_{coverage,t}$ and $R_{switch,t}$ correspond to the sub-items of minimizing the recovery time, maximizing the power supply coverage and minimizing the number of switch switching, respectively. w_1, w_2, $w_3 \in [0,1]$ are normalized weight coefficients, satisfying $w_1+w_2+w_3 = 1$.

The dynamic adjustment weight changes during the model training phase are shown in Table 1.

TABLE 1. Dynamically adjusted weight changes during the training phase

Training Episode	w_1	w_2	w_3
Episode 1	0.35	0.35	0.30
Episode 2	0.36	0.36	0.28
Episode 3	0.38	0.37	0.25
Episode 4	0.40	0.38	0.22
Episode 5	0.43	0.39	0.18
Episode 6	0.46	0.40	0.14
Episode 7	0.49	0.40	0.11
Episode 8	0.52	0.39	0.09
Episode 9	0.55	0.37	0.08
Episode 10	0.58	0.36	0.06

From the results in Table 1, as the training rounds advance from the 1st round to the 10th round, the weight w1 used to measure the recovery time shows a clear upward trend, and the weight value increases from 0.35 to 0.58; the weight w2 used to measure the power supply coverage rate shows a fluctuating upward trend, and the final weight value (in the 10th round) is 0.36; while the weight w3 used to control the switch switching cost shows a continuous downward trend, and the weight value in the 10th round is only 0.06. The above results show that with the deepening of training rounds, the model has been able to achieve fast recovery and high power supply coverage more stably, so the penalty weight for switching actions is further compressed to avoid additional losses caused by frequent switching. This dynamic weight adjustment mechanism helps to achieve the optimal restoration strategy that takes into account efficiency, stability and economy in complex power grid environments.

JOINT TRAINING MECHANISM DESIGN

The alternating training framework is used to establish a closed-loop interaction mechanism between the diffusion model and the RL model to achieve mutual feedback and coordinated optimization between the two during the training process. The training process is divided into two alternating phases: first, the RL model is fixed and the diffusion models are trained to generate more challenging failure scenarios; then the diffusion models are fixed and the RL model is trained to adapt to these newly generated samples, gradually improving the overall performance through multiple rounds of iterations.

The output uncertainty of renewable energy is explicitly modeled through diverse scenarios generated by the diffusion model. During the training phase, each failure scenario generated by the DM model includes randomly fluctuating DG output and load demand (consisting of historical statistical distributions), and the RL agent learns its policy in these environments with different uncertainties. This process essentially constitutes a domain randomization training: the policy is optimized in a large number of physically feasible but randomly perturbed scenarios, thereby gaining robustness to unseen uncertainty patterns, rather than relying on a single deterministic state assumption.

In the parameter updating process, a dual time scale update rule is adopted, combined with momentum term to improve the model's memory and adjustment ability for difficult examples. The dynamic resource allocation problem is modeled as the following optimization form:

$$\rho = \rho_0 + \lambda \cdot |\mathcal{L}_{DDPM} - \mathcal{L}_{RL}| \tag{19}$$

In formula (19), ρ is the ratio of diffusion models to RL training rounds.

After entering the RL model-first training phase, the Diffusion Models generator parameters and physical constraint embedding modules are fixed to focus on optimizing the policy network. To improve the robustness of the strategy, the top 10% high-confidence scenarios are selected from the samples generated by the diffusion model, and the residual perturbations of the power flow equation are superimposed to construct a set of adversarial scenarios with feasible boundaries but challenging. The perturbation generation method is as follows:

$$\delta = \Delta_{pp} \cdot \mathrm{sign}\left(\nabla_x \mathcal{L}_{phys}\right) \tag{20}$$

In formula (20), Δ_{pp} is the preset physical error threshold. On this basis, the gradient coordination strategy is implemented. When a conflict between the policy gradient and the DM generated gradient is detected, a simple weighted correction is used:

$$\nabla J'_{RL} = (1-\alpha)\nabla J_{RL} + \alpha \nabla L_{phys} \tag{21}$$

In formula (21), $\alpha \in [0, 1]$ is the conflict weight coefficient.

DYNAMIC ENVIRONMENT CONSTRUCTION AND INTERACTIVE INTERFACE

The control action space in this simulation environment is limited to binary operations (opening/closing) of tie switches and sectionalizing switches, excluding capacitor switching, OLTC tuning, or DG power adjustment, in order to focus on topology reconfiguration as the core recovery mechanism.

This paper constructs a dynamic simulation environment for intelligent power grid restoration based on OpenDSS, Python API (Application Programming Interface) and Gym framework to achieve efficient collaboration between high-fidelity physical simulation and reinforcement learning training. The system adopts a three-level architecture consisting of simulation engine layer, data interaction layer and interface encapsulation layer. Each layer communicates through standardized protocols to ensure the stability and scalability of the system.

This simulation assumes that after fault isolation, the system state (including all node voltages, branch power flows, and real-time DG/load values) can be fully obtained through advanced measurement systems (such as AMI and PMU), i.e., it is fully observable. Although real-world systems may have communication delays or missing measurements, this paper focuses on the upper limit performance of the strategy under ideal observability; robust extensions to partial observability will be a future research direction. At the simulation engine level, the IEEE 33 (Institute of Electrical and Electronics Engineers) node distribution system is converted into an OpenDSS model, the reference voltage is set to 12.66 kV, the total load is set to 3.715 + j2.3 MVA, and the parameters of components such as lines and transformers are configured. Using Fault objects to simulate typical faults and set the operation mode and frequency to match the actual operation of the power grid. The power flow calculation is driven by the built-in solver to generate time-series state data that conforms to physical laws and supports strategy evaluation and optimization. The data interaction layer implements millisecond-level bidirectional communication between Python and OpenDSS through the COM (Component Object Model) interface. In terms of state observation, key electrical quantities such as node voltage and branch power flow are collected and organized into structured vectors. The sliding window z-score normalization method is introduced to eliminate dimensional differences and improve the stability of the algorithm. The expression is:

$$x_{norm} = \frac{x - \mu_{window}}{\sigma_{window} + \epsilon} \tag{22}$$

In formula (22), μ_{window} and σ_{window} are the mean and standard deviation of the data in the most recent steps, and ϵ is a small constant to prevent division by zero.

The interface encapsulation layer strictly follows the Gym v26 specification to achieve standardized access to the reinforcement learning framework. When the environment is initialized, the fault state is cleared, the load curve is loaded, and N-1 safety topologies are randomly generated, and the initial observation vector containing the grid operation status is returned. The step function parses the action command and triggers the power flow update, and calculates the reward based on the multi-objective reward function. When the power supply coverage exceeds the preset value or the switch operation exceeds the limit, the round is terminated immediately.

III. EXPERIMENT AND VERIFICATION

EXPERIMENTAL DESIGN

This experiment is based on the IEEE 33-node distribution network standard system to verify the intelligent recovery performance of the proposed CDM-RL model in complex fault scenarios. The dataset used in the experiment is a self-built power grid fault recovery dataset based on the IEEE 33-node system architecture. The experimental platform is built on the OpenDSS simulation environment, and Python is used to implement model training and strategy deployment. The topology and parameter configuration of the IEEE 33-node system, including line impedance, load distribution and substation access method, are constructed in OpenDSS to ensure that the model conforms to the operating characteristics of the actual distribution network. On this basis, by constructing a diverse set of faults, the performance of traditional optimization methods, a single RL model and the CDM-RL model proposed in this paper in intelligent recovery tasks are compared to comprehensively evaluate the effectiveness and practicality of the proposed collaborative optimization framework. Among them, MILP is selected as the representative of the traditional optimization model, and GNN-PPO (Graph Neural Network with Proximal Policy Optimization) and GAT-DQN (Graph Attention Network with Deep Q-Network) are selected as the representatives of the single RL model.

GRID RESTORATION EFFICIENCY EVALUATION

Figure 2 shows the performance of each model in the grid restoration efficiency evaluation. The evaluation indicators include: Time to First Service Restoration (TFSR), Average Restoration Time (ART), Average Decision Delay (ADD), and Action Execution Consistency (AEC). Among them, the horizontal axis represents different models, and the vertical axis corresponds to the value of each indicator.

Judging from the data in Figure 2, the CDM-RL model has the best overall performance. In terms of the TFSR indicator, the CDM-RL model has the lowest value of 8.9 seconds, followed by the GAT-DQN model (11.2 seconds) and the GNN-PPO model (13.5 seconds), while the MILP model has the highest value of 18.6 seconds; in the ART indicator, the CDM-RL model is also at the lowest level, at only 29.5 seconds. In comparison, the GAT-DQN model is 37.8 seconds, the GNN-PPO model is 42.1 seconds, and the MILP model is as high as 54.3 seconds; in the ADD indicator that evaluates the efficiency of strategy execution, the value of the CDM-RL model is 220.7 milliseconds, which is significantly lower than the 315.6 milliseconds of the GAT-DQN model and the 405.2 milliseconds of the GNN-PPO model. The MILP model has the highest ADD value, reaching 920.3 milliseconds. In terms of the AEC indicator that measures the stability of the control strategy, the CDM-RL model has the highest value, at 91.7%, much higher than the 84.6% of the GAT-DQN model and the 80.3% of the GNN-PPO model. The MILP model has the lowest AEC, at only 72.4%.

GRID LOAD RESTORATION QUALITY ASSESSMENT

Figure 3 shows the comprehensive performance of each model in grid load restoration quality assessment.

The evaluation indicators include: Load Restoration Rate (LRR); Critical Load Restoration Rate (CLRR); and Low Voltage Area Recovery Rate (LVARR). Among them, the horizontal axis represents different evaluation indicators, and the vertical axis represents the value of the correspond-

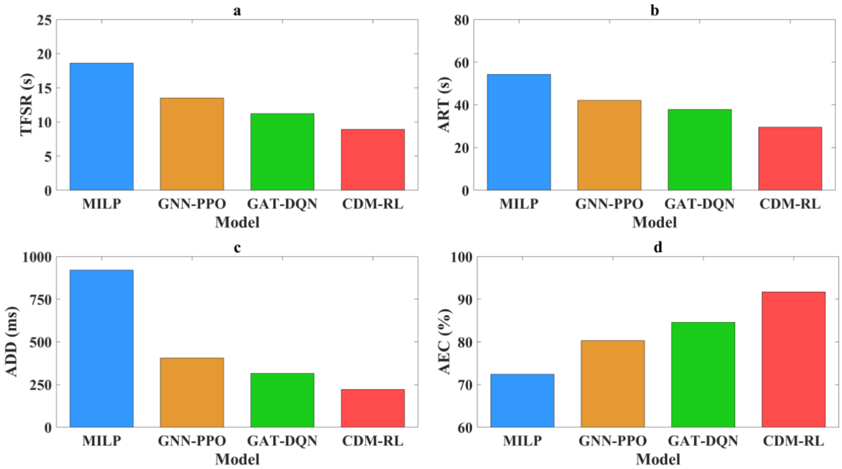

FIGURE 2. Performance of each model in grid restoration efficiency evaluation. Figure 2a TFSR; Figure 2b ART; Figure 2c ADD; Figure 2d AEC

ing indicator. The higher the value, the better the recovery effect.

As can be seen from Figure 3, the CDM-RL model achieves the highest values in all three indicators, with LRR, CLRR and LVARR values of 93.1%, 90.5% and 91.7% respectively, which are significantly higher than other models, showing its stronger coverage and priority response capabilities for various load areas during fault recovery; the three indicators of the GAT-DQN model are 87.6%, 81.3% and 84.5% respectively, which is the second best overall performance, better than the 83.5%, 76.8% and 79.2% of the GNN-PPO model, reflecting that the GAT-DQN model has certain advantages in modeling complex power grid topology and optimizing restoration paths; in contrast, the MILP model has the lowest values, which are 78.2%, 65.4%, and 72.9%, respectively, indicating that the traditional optimization method has limited recovery capabilities when facing variable failure scenarios. Overall, the CDM-RL model performs better than other models in all indicators, reflecting its comprehensive advantages in recovery efficiency, control stability and load balancing.

SWITCHING OPERATION COST AND STRATEGIC EFFICIENCY EVALUATION

In order to comprehensively evaluate the economic efficiency and strategic efficiency of each model in the switching operation during the power grid fault recovery process, this section starts from the economic dimension and strategic efficiency dimension. Among them, the following indicators are selected for analysis in the economic dimension: Average Switching Actions, Total Switching Cost and Tie-switch Proportion; the following indicators are selected for quantitative analysis in the strategic efficiency dimension: Ineffective Action Rate and Ratio of Optimal Path.

Table 2 shows the performance of three key indicators of each model in terms of switch operation economy, namely Average Switching Actions, Total Switching Cost and Tie-switch Proportion.

TABLE 2. Comparison results of various models on switch operation economic indicators

Model	Average Switching Actions	Total Switching Cost	Tie-switch Proportion
MILP	5.8	23.6	0.45
GNN-PPO	5.1	20.4	0.41
GAT-DQN	4.5	18.2	0.38
CDM-RL	3.9	15.7	0.35

From the data in Table 2, the CDM-RL model outperforms other models in all three indicators and shows better economic performance. In terms of the average number of switch switching, the CDM-RL model is 3.9 times, which is significantly lower than the 5.8 times of the MILP model, the 5.1 times of the GNN-PPO model, and the 4.5 times of the GAT-DQN model, indicating that the CDM-RL model requires fewer operations during the recovery process; in terms of the total cost of switching operations, the total cost of switching operations of the CDM-RL model is 15.7, which is much lower than 23.6 of the MILP model, 20.4 of the GNN-PPO model, and 18.2 of the GAT-DQN model, indicating that the CDM-RL model has more advantages in terms of overall operation costs; in terms of the proportion of tie switches, the proportion of tie switches in the MILP model is 0.45, the proportion of tie switches in the GNN-PPO model is 0.41, and the proportion of tie switches in the GAT-DQN model drops to 0.38. In contrast, the proportion of tie switches in the CDM-RL model is the lowest, at only 0.35.

Figure 4 shows the performance of each model on two operational efficiency indicators, namely, Ineffective Action Rate and Ratio of Optimal Path. Among them, the left Y-axis represents the invalid action rate. The lower the value, the less trial and error behavior the model has during execution and the more stable the decision. The right Y-axis represents the optimal ratio of the recovery path. The higher the value, the closer the selected recovery path is to the theoretical optimal solution.

From the data in Figure 4, the CDM-RL model reaches the optimal value in both indicators. Its invalid action rate is

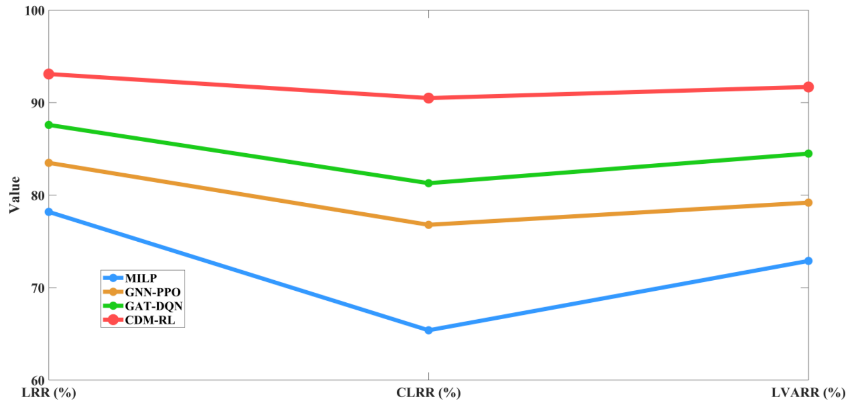

FIGURE 3. Performance of grid load restoration quality evaluation of each model

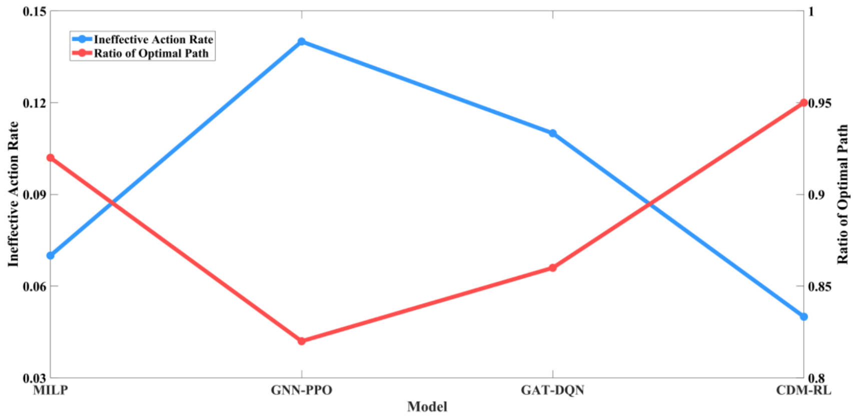

FIGURE 4. Test results of each model on indicators related to the operational efficiency dimension

only 0.05, the lowest among all models. At the same time, the optimal recovery path ratio reaches 0.95, the highest among all models. This shows that the CDM-RL model not only has high policy stability, but also is closest to the ideal state in path planning. In contrast, the invalid action rate of the MILP model is 0.07, slightly higher than the CDM-RL model, but lower than the GNN-PPO model (0.14) and the GAT-DQN model (0.11), and its optimal recovery path ratio is 0.92, second only to the CDM-RL model, but higher than the GNN-PPO model (0.82) and the GAT-DQN model (0.86), showing that traditional optimization methods still have strong path planning capabilities in deterministic scenarios.

STRATEGY SUCCESS RATE EVALUATION

Figure 5 shows the Strategy Success Rate (SSR) and Average Strategy Success Rate (Average SSR) of each model under three typical failure scenarios (single point failure, multiple point failure, cascading failure).

As can be seen from Figure 5, the CDM-RL model reaches the highest values in all four indicators. Its strategy success rate under single-point failure is 94.6%, the strategy success rate under multiple-point failure is 87.3%, and the strategy success rate under cascading failure is 81.2%, the average strategy success rate reaches 87.7%, showing its stronger adaptability and strategy stability under various failure scenarios. In contrast, the strategy success rate of the MILP model under single-point failure, strategy success rate under multiple-point failure, strategy success rate under cascading failure, and average strategy success rate are 83.1%, 72.5%, 66.8%, and 74.1%, respectively, which are the lowest among all models, reflecting the limitations of traditional optimization methods in complex fault recovery tasks. The strategy success rates of the GNN-PPO model under single-point, multi-point, and cascading failures are 85.4%, 76.2%, and 69.5%, respectively, and the average strategy success rate is 77.0%. The overall performance is better than the MILP model, but significantly weaker than the CDM-RL model. The strategy success rate of the GAT-DQN model under single-point failure, multiple-point failure, cascading failure, and average strategy success rate are 89.4%, 80.2%, 74.5%, and 81.4%, respectively. Its overall performance is slightly better than the GNN-PPO model, but there is still a certain gap with the CDM-RL model. Overall, the CDM-RL model shows better overall performance in improving the success rate of recovery strategies by virtue of the synergistic mechanism of diffusion models and RL models.

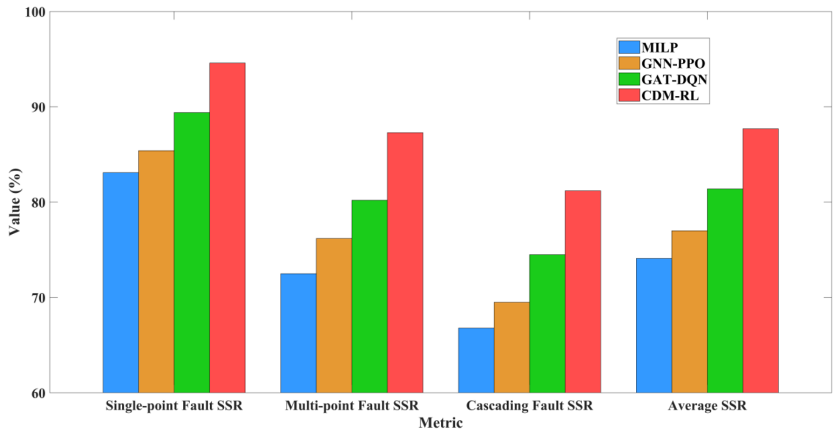

FIGURE 5. Strategy success rate evaluation results of each model

VOLTAGE STABILITY EVALUATION

In order to comprehensively evaluate the control ability of each model on voltage stability in the grid intelligent restoration task, this paper selects the following four key indicators for analysis: Minimum Bus Voltage (MBV), Voltage Recovery Time (VRT), Integral of Voltage Deviation (IVD), and Percentage of Low Voltage Nodes (LVN). The voltage stability evaluation results of each model are shown in Figure 6.

From the data in Figure 6, the MBV of the MILP model is 0.83 p.u., which is the lowest among the four models, indicating that its ability to maintain the minimum voltage level of the system during fault recovery is relatively weak; while the MBV of the CDM-RL model is the highest, reaching 0.95 p.u., which means that its recovery strategy can more effectively ensure the voltage stability of key nodes in the power grid. In terms of VRT, the VRT of CDM-RL is 6.7 s, which is significantly lower than 14.6s of the GNN-PPO model, 10.4s of the GAT-DQN model, and 18.2s of the MILP model, reflecting its faster voltage recovery response speed. The IVD indicator shows that the IVD value of CDM-RL is the smallest, which is 0.17 p.u.·s. In comparison, the MILP model reaches 0.48 p.u.·s, which means that during the entire recovery process, the CDM-RL model has more precise control over the deviation of the system voltage from the ideal state. In terms of LVN indicators, the LVN of the CDM-RL model is only 7.3%, which is much lower than 32.5% of MILP and the other two RL models, showing that it has more advantages in reducing the overall low voltage risk. Overall, the CDM-RL model outperforms the other three models in all four indicators, demonstrating stronger voltage stability control capabilities.

NETWORK LOSS CHANGE ASSESSMENT

In order to comprehensively evaluate the ability of each model to regulate network loss changes in the power grid fault recovery task, this paper selects the following four key indicators for analysis: Active Loss Reduction Rate (ALRR), Maximum Branch Loss Proportion (MBLP), Loss Fluctuation Magnitude (LFM) and Post-recovery Loss Deviation (PLD). The network loss change evaluation results of each model are shown in Figure 7.

As can be seen from Figure 7, in terms of the ALRR index, the performance of the MILP model is 62.4%, the GNN-PPO model is 73.6%, the GAT-DQN model is improved to 78.1%, and the CDM-RL model reaches the highest 84.9%; in terms of MBLP index, the MILP model is 28.7%, the GNN-PPO model drops to 22.5%, the GAT-DQN model further drops to 19.3%, and the CDM-RL model is the lowest at 15.2%; in terms of LFM index, the LFM value of the MILP model is 1.85kW, GNN-PPO drops to 1.34kW, GAT-DQN is 1.12kW, and CDM-RL is the lowest, which is 0.83kW; in terms of PLD indicators, the values of MILP, GNN-PPO, GAT-DQN and CDM-RL models are 0.93kW, 0.61kW, 0.47kW and 0.28kW, respectively, showing a trend of gradual optimization. Overall, as the model evolves from traditional optimization methods to RL and then to the CDM-RL model proposed in this paper, all indicators show an improving trend, reflecting the effect of more advanced modeling methods on improving the loss control capability during power grid fault recovery. It also further verifies the effectiveness and advancement of the CDM-RL model in improving the self-healing capability of the power grid and optimizing the recovery strategy.

EVALUATION OF SCENARIO DIVERSITY ADAPTABILITY

In order to fully verify the generalization ability of each model in unknown or untrained fault scenarios, this paper selects the following three key indicators for analysis: Cross-scenario Policy Stability (CPS), Unseen Scenario Restoration Delay (USRD), and Policy Failure Boundary (PFB). The performance of each model in the evaluation of scene diversity adaptability is shown in Figure 8.

From the data in Figure 8, the CPS of the MILP model is 61.2%, the USRD is 22.4s, and the PFB is 0.48p.u., which is relatively low overall. This shows that when facing unknown or untrained fault scenarios, its strategy stability is weak, its response speed is slow, and it is more likely to fail. The GNN-PPO model has improved in all three indicators, with CPS increased to 69.5%, USRD reduced

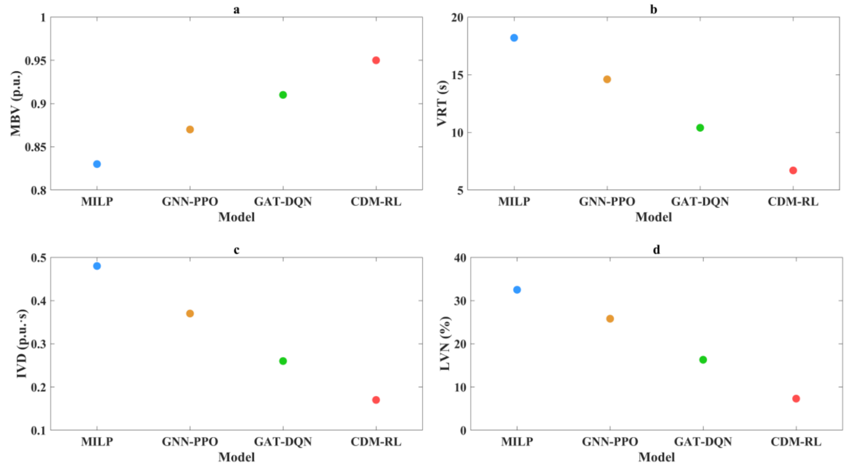

FIGURE 6. Voltage stability evaluation results of each model. Figure 6a MVB; Figure 6b VRT; Figure 6c IVD; Figure 6d LNN

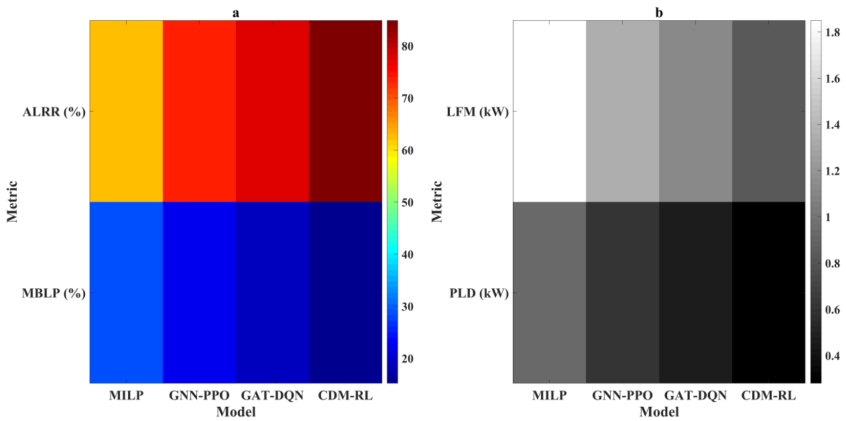

FIGURE 7. Network loss change evaluation results of each model. Figure 7a ALRR and MBLP; Figure 7b LFM and PLD

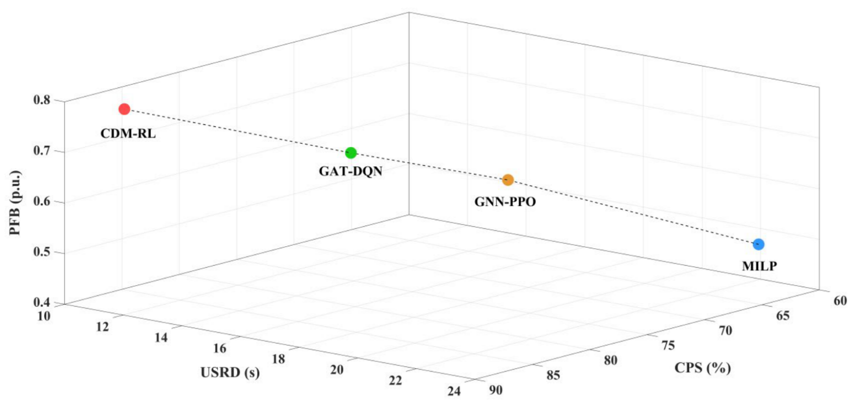

FIGURE 8. Evaluation results of the scene diversity adaptability of each model

to 17.1s, and PFB increased to 0.60 p.u., indicating that the design based on the GNN and PPO algorithm has more advantages in generalization ability and recovery efficiency than the traditional optimization method. The GAT-DQN model further improves the performance, with a CPS of 76.8%, a USRD of 14.6s, and a PFB of 0.67 p.u., showing better adaptability to complex topological structures and variable failure modes after the introduction of the attention mechanism. In contrast, the CDM-RL model achieves the highest level in all indicators, with a CPS of 85.3%, the lowest USRD (10.2s), and the highest PFB (0.76 p.u.), reflecting the leading advantage of the CDM-RL model in comprehensive performance.

To verify the synergistic gains of the alternating training mechanism, this paper conducted additional ablation experiments: a pure GNN-PPO agent (denoted as "PPO-Hist") was constructed, trained only using historical fault records and traditional Monte Carlo generated scenarios. The results show that on the same test set, PPO-Hist's cross-scenario policy stability (CPS) is 71.6%, and its first power recovery time is 12.3 seconds, significantly inferior to CDM-RL's 85.3% and 8.9 seconds, respectively. This indicates that the diverse and physically constrained challenging scenarios generated by the diffusion model effectively improve the generalization ability and robustness of the RL policy, confirming the necessity of DM-RL co-training.

Although this paper validates the effectiveness of CDM-RL on an IEEE 33-node system, scaling it to larger-scale networks (such as those with hundreds of nodes) still faces computational challenges. The main bottlenecks lie in the overhead of generating high-dimensional grid states using the diffusion model (O(N2)) and solving OpenDSS power flow problems. These can be mitigated in the future through graph coarsening, partitioned parallel generation, and decoupling DM pre-training from RL fine-tuning. Preliminary experiments show that on a 118-node system, a partitioned training strategy can keep the alternating training time of a single round of DM-RL within an acceptable range (<2 hours), providing a possible path for industrial-scale deployment.

IV. CONCLUSION

In order to solve the complexity and uncertainty of grid fault recovery caused by the increase of the renewable energy access ratio in new power systems, this paper proposes a collaborative decision-making model (CDM-RL) that combines diffusion models and RL models, which is vital for ensuring the rapid, stable restoration of power critical to continuous and sensitive industrial operations. Experimental results show that the performance of the CDM-RL model on the IEEE 33-node system is significantly better than that of the traditional optimization model and the single RL model. It excels in key indicators such as restoration efficiency, power supply coverage, and switch operation cost, and comprehensively improves the efficiency, stability, and adaptability of power grid fault recovery. The CDM-RL model proposed in this paper provides a theoretical basis and technical path for the intelligent restoration of new power systems in uncertain environments, thereby ensuring the crucial stability and reliability of the electricity supply for critical industrial consumers.

AUTHOR CONTRIBUTIONS

Jun Li designed, collected and analyzed the data, and drafted the manuscript. Jun Li conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Jun Li participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] T. Egeland-Eriksen, A. Hajizadeh, and S. Sartori, "Hydrogen-based systems for integration of renewable energy in power systems: Achievements and perspectives," International journal of hydrogen energy, vol. 46, no. 63, pp. 31963-31983, 2021, doi: 10.1016/j.ijhydene.2021.06.218.

[2] Y. Shu, G. Chen, J. He, and F. Zhang, "Building a new electric power system based on new energy sources," Strategic Study of Chinese Academy of Engineering, vol. 23, no. 6, pp. 61-69, 2021, doi: 10.15302/J-SSCAE-2021.06.003.

[3] H. Zhang, W. Xiang, W. Lin, and J. Wen, "Grid forming converters in renewable energy sources dominated power grid: Control strategy, stability, application, and challenges," Journal of modern power systems and clean energy, vol. 9, no. 6, pp. 1239-1256, 2021, doi: 10.35833/MPCE.2021.000257.

[4] Q. Li, B. Ren, W. Tang, D. Wang, C. Wang, and Z. Lv, "Analyzing the inertia of power grid systems comprising diverse conventional and renewable energy sources," Energy Reports, vol. 8, pp. 15095-15105, 2022, doi: 10.1016/j.egyr.2022.11.022.

[5] C. D. Iweh, S. Gyamfi, E. Tanyi, and E. Effah-Donyina, "Distributed generation and renewable energy integration into the grid: Prerequisites, push factors, practical options, issues and merits," Energies, vol. 14, no. 17, pp. 5375-5408, 2021, doi: 10.3390/en14175375.

[6] A. A. Kebede, T. Kalogiannis, J. Van Mierlo, and M. Berecibar, "A comprehensive review of stationary energy storage devices for large scale renewable energy sources grid integration," Renewable and sustainable energy re-views, vol. 159, pp. 112213-112231, 2022, doi: 10.1016/j.rser.2022.112213.

[7] Q. Liu, V. Hagenmeyer, and H. B. Keller, "A review of rule learning-based intrusion detection systems and their pro-spects in smart grids," Ieee Access, vol. 9, pp. 57542-57564, 2021, doi: 10.1109/ACCESS.2021.3071263.

[8] J. Yan, B. Hu, C. Shao, W. Huang, Y. Sun, W. Zhang, et al., "Scheduling post-disaster power system repair with in-complete failure information: A learning-to-rank approach," IEEE Transactions on Power Systems, vol. 37, no. 6, pp. 4630-4641, 2022, doi: 10.1109/TPWRS.2022.3149983.

[9] D. Liu, X. Zhang, and K. T. Chi, "Effects of high level of penetration of renewable energy sources on cascading failure of modern power systems," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 12, no. 1, pp. 98-106, 2022, doi: 10.1109/JETCAS.2022.3147487.

[10] W. Strielkowski, L. Civín, E. Tarkhanova, M. Tvaronavičienė, and Y. Petrenko, "Renewable energy in the sustainable development of electrical power sector: A review," Energies, vol. 14, no. 24, pp. 8240-8263, 2021, doi: 10.3390/en14248240.

[11] J. Hu, X. Liu, M. Shahidehpour, and S. Xia, "Optimal operation of energy hubs with large-scale distributed energy resources for distribution network congestion management," IEEE Transactions on Sustainable Energy, vol. 12, no. 3, pp. 1755-1765, 2021, doi: 10.1109/TSTE.2021.3064375.

[12] L. Lv, Z. Wu, J. Zhang, L. Zhang, Z. Tan, and Z. Tian, "A VMD and LSTM based hybrid model of load forecasting for power grid security," IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 6474-6482, 2021, doi: 10.1109/TII.2021.3130237.

[13] C. Wang, P. Ju, F. Wu, X. Pan, and Z. Wang, "A systematic review on power system resilience from the perspective of generation, network, and load," Renewable and Sustainable Energy Reviews, vol. 167, pp. 112567-112587, 2022, doi: 10.1016/j.rser.2022.112567.

[14] P. Denholm, D. J. Arent, S. F. Baldwin, D. E. Bilello, G. L. Brinkman, J. M. Cochran, et al., "The challenges of achieving a 100% renewable electricity system in the United States," Joule, vol. 5, no. 6, pp. 1331-1352, 2021, doi: 10.1016/j.joule.2021.03.028.

[15] A. M. Nassef, M. A. Abdelkareem, H. M. Maghrabie, and A. Baroutaji, "Review of metaheuristic optimization algo-rithms for power systems problems," Sustainability, vol. 15, no. 12, pp. 9434-9460, 2023, doi: 10.3390/su15129434.

[16] M. Mahdavi, H. H. Alhelou, A. Bagheri, S. Z. Djokic, and R. A. V. Ramos, "A comprehensive review of metaheuristic methods for the reconfiguration of electric power distribution systems and comparison with a novel approach based on efficient genetic algorithm," IEEE Access, vol. 9, pp. 122872-122906, 2021, doi: 10.1109/ACCESS.2021.3109247.

[17] S. A. Shezan, M. F. Ishraque, G. M. Shafiullah, I. Kamwa, L. C. Paul, S. M. Muyeen, et al., "Optimization and control of solar-wind islanded hybrid microgrid by using heuristic and deterministic optimization algorithms and fuzzy logic controller," Energy reports, vol. 10, pp. 3272-3288, 2023, doi: 10.1016/j.egyr.2023.10.016.

[18] X. Ding, X. Liao, W. Cui, X. Meng, R. Liu, Q. Ye, et al., "A Deep Reinforcement Learning Optimization Method Con-sidering Network Node Failures," Energies, vol. 17, no. 17, pp. -4483, 2024, doi: 10.3390/en17174471.

[19] M. M. Hosseini and M. Parvania, "Resilient operation of distribution grids using deep reinforcement learning," IEEE Transactions on Industrial Informatics, vol. 18, no. 3, pp. 2100-2109, 2021, doi: 10.1109/TII.2021.3086080.

[20] W. Liao, B. Bak-Jensen, J. R. Pillai, Y. Wang, and Y. Wang, "A review of graph neural networks and their applications in power systems," Journal of Modern Power Systems and Clean Energy, vol. 10, no. 2, pp. 345-360, 2021, doi: 10.35833/MPCE.2021.000058.

[21] L. Pinciroli, P. Baraldi, G. Ballabio, M. Compare, and E. Zio, "Deep reinforcement learning based on proximal policy optimization for the maintenance of a wind farm with multiple crews," Energies, vol. 14, no. 20, pp. 6743-6759, 2021, doi: 10.3390/en14206743.

[22] B. Er-Rahmadi and T. Ma, "Data-driven mixed-Integer linear programming-based optimisation for efficient failure detection in large-scale distributed systems," European Journal of Operational Research, vol. 303, no. 1, pp. 337-353, 2022, doi: 10.1016/j.ejor.2022.02.006.