MULTI-OBJECTIVE SCALE OPTIMIZATION OF POWER SYSTEM ENERGY STORAGE BASED ON EFCPPO ALGORITHM

Xiu Zheng, Jiyang Zhang, Jie Yang, Qinghua Liao*
School of Electrical Engineering & Automation, Henan Institute of Technology, Xinxiang, 453003, China

Abstract: With the increasing penetration rate of renewable energy and the growing complexity of multi-energy coupled systems, power system energy storage dispatch faces challenges in terms of economy, environmental protection, and multi-timescale collaborative optimization. To achieve multi-objective and efficient dispatch of energy storage devices in integrated energy systems, this paper proposes an intelligent decision-making framework based on an exploration fallback clip proximal policy optimization algorithm. The new method introduces an exploration fallback mechanism and a hierarchical reinforcement learning structure, and achieves collaborative updating of policy and value through a dual network, constructing a collaborative dispatch mechanism of upper-level macro-planning and lower-level execution. The research results show that the new algorithm exhibits superior performance in both weekday and holiday scenarios. Its peak daily load dispatch is reduced by approximately 200 kW/h-210 kW/h compared to traditional multi-objective optimization methods, with an average decrease in operating costs of 3.1%-5.6% and pollution costs of 3.8%-8.3%. Simultaneously, by coordinating load sharing among electric boilers, gas-fired boilers, and combined heat and power units during off-peak hours, the algorithm effectively reduces the system's total energy consumption and carbon emission intensity. Therefore, the new algorithm demonstrates significant advantages in improving the economic efficiency and environmental friendliness of power system energy storage dispatch. This provides a feasible solution for multi-energy coordinated dispatch under high-proportion renewable energy integration.

Keywords: EFCPPO; Power system; Energy storage; Multi-objective scale; Optimization.

1. Overview

With the transformation of the global energy structure and the large-scale access of renewable energy, the operation mode of the power system is undergoing profound changes. As a key technology to improve the flexibility of the power system and promote the consumption of new energy, the role of energy storage system in peak shaving, frequency regulation and reserve capacity is becoming increasingly prominent [1]. However, traditional power system energy storage dispatch models often fail to ensure the economic operation of the system while taking into account environmental protection and multi-energy synergistic optimization, especially in complex scenarios with multiple time scales and multiple energy forms coupled, where the adaptability and optimization effect of its dispatch strategy are significantly limited. In recent years, reinforcement learning technology, especially Proximal Policy Optimization (PPO) algorithm, has been widely used in power system optimization dispatch due to its excellent performance in continuous control and high-dimensional state space [2-3].

However, traditional PPO algorithms have problems such as strong randomness, slow convergence speed and unstable policy updates in the action selection process, which limits their further application in complex multi-objective energy storage scheduling. Dhiman G et al. proposed a collaborative decision-making framework that integrates artificial intelligence, Multi-objective Optimization (MOO) and Internet of Things to address the problems of low scheduling efficiency and system response lag faced by consumer electronics products in complex multi-objective scenarios in smart cities. Experimental results showed that the proposed method had better comprehensive scheduling performance than existing technologies in typical smart city scenarios, providing new algorithm support and practical path for intelligent management of equipment in highly dynamic urban environments [4]. Moghaddasi K et al. proposed a MOO framework based on dual deep Q network to address the problems of low task offloading efficiency, difficult energy management and insufficient data security in

dx.doi.org/10.17683/ijomam/issue24.36

the vehicle network environment. This model integrated deep neural network and deep learning technology to achieve adaptive decision-making in dynamic network environments. Experiments showed that compared with traditional methods such as deep Q network and deep deterministic policy gradient, the proposed model had significant performance in terms of energy consumption reduction of 26.4%, latency reduction of 6.87% and system cost reduction of 7.41% [5]. Baimbetov D et al. proposed a new architecture for a combined cooling, heating and power system integrating hydrogen fuel cell vehicles and hydrogen storage devices to address the problems of insufficient renewable energy penetration, high carbon emissions, and low multi-energy synergy efficiency in multi-hub energy systems. This study used stochastic programming to handle the uncertainty of multi-energy markets and introduced electricity-to-gas technology and demand response planning to construct a MOO operation framework. The results showed that the proposed model significantly improved the economic and environmental benefits of the system, with operating costs reduced by 32.17% and carbon dioxide emissions reduced by 79.21% [6]. Rajagopalan A et al. proposed an energy management strategy based on an improved iterative mapping adaptive crystal structure algorithm to address the MOO problem of economic and environmental benefits faced by microgrid systems containing renewable energy and Plug-In Hybrid Electric Vehicles (PHEVs) in coordinated scheduling. Simulation results showed that compared with traditional crystal structure algorithms and other evolutionary optimization methods, SaCryStAl showed better comprehensive performance in terms of operating costs, carbon emissions, and computation time, providing an effective optimization path for the operation of microgrids with high penetration PHEV access [7].

Based on this, this study proposes an exploration fallback clip proximal policy optimization (EFCPPO) algorithm to improve its performance in multi-objective scale optimization of power system energy storage. The study innovatively constructs a basic framework for an improved PPO algorithm (EFCPPO) that combines an exploratory mechanism, policy fallback, and gradient clip techniques. By constructing the EFCPPO, the new algorithm optimizes the policy update mechanism and introduces a fallback clip strategy, thereby improving the convergence efficiency and training stability in complex scheduling environments. Simultaneously, the study combines a Hierarchical Reinforcement Learning (HRL) architecture to design a multi-timescale optimization model that coordinates upper-level macro-planning with lower-level real-time execution, achieving joint scheduling and dynamic coordination of multiple energy systems such as electricity, heat, and gas. Through multi-scenario simulation experiments and comparative analysis with various mainstream optimization algorithms, the study verifies the algorithm's comprehensive advantages in reducing system operating costs, improving environmental benefits, and enhancing multi-timescale scheduling flexibility. This research provides new theoretical support and methodological pathways for multi-objective collaborative optimization of power system energy storage, and has significant theoretical and practical value for promoting the intelligent and low-carbon operation of new power systems.

2. Methods

2.1 Construction of EFCPPO Algorithm

Currently, traditional power system energy storage dispatch models cannot simultaneously achieve energy conservation and environmental protection goals while ensuring system operation. Therefore, to realize the economic dispatch of new energy sources in the power system, a new method based on improved PPO is proposed. Since the traditional PPO algorithm often randomly selects action parameters during the power system energy storage optimization process, the entire process becomes random. Therefore, this study proposes EFCPPO based on the traditional PPO algorithm. This method can select the distribution probability as shown in equation (1) [8-9].

$$\kappa_{new} = \arg\max(\kappa(\square S_t)) \tag{1}$$

In equation (1), κ_{new} represents the improved proximal degradation strategy, $\kappa(\square S_t)$ represents the state in the strategy, and $\arg\max$ represents the parameter corresponding to the maximum value.

The basic process of selecting the probability is shown in Figure 1.

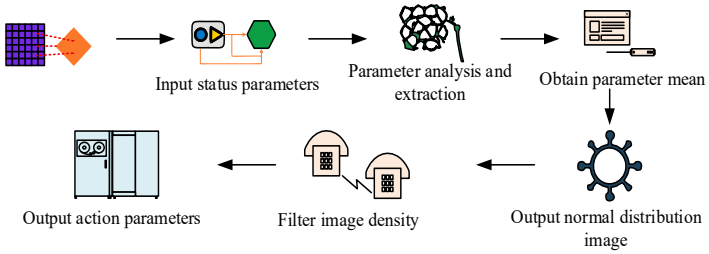

Figure 1: Basic process of selecting probability

As shown in Figure 1, during the action selection process, the algorithm first obtains power system data information through the model, and then inputs the current system state parameters. Then, the network model analyzes and extracts the parameters to obtain the mean and variance of the current parameters. Then, the normal distribution image of the parameters is output based on the overall distribution of the parameters. The highest density value of the normal distribution image is selected. Finally, the action parameters of the current parameters are output. The PPO algorithm uses a pruning strategy to control the magnitude of the policy update, thereby maintaining the stability of the training process. This makes the gradient of the objective function approach zero when the policy difference is too large, effectively avoiding the drastic fluctuation of the policy caused by the parameter update, and ensuring that the difference between the old and new policies is always within a controllable range. Equation (2) is the objective function formula for the new strategy [10-11].

$$H^{p}(\phi)=\frac{1}{D}\sum_{t=1}^{D}H_{t}^{p}(\phi) \tag{2}$$

In equation (2), $H^{p}(\phi)$ represents the overall objective function, ϕ represents the trainable parameters of the policy network, D represents the batch size, and $H_{t}^{p}(\phi)$ represents the individual objective function of the t-th sample. The update of the entire decision-making process is achieved through the gradient change of the objective function. Therefore, the probability of the entire decision-making process has certain limitations compared with the strict pruning range. The study introduces a fallback clip mechanism to bring the probability ratio of out-of-bounds errors back to the preset range. The formula function of the pruning mechanism is shown in equation (3) [12-13].

$$c^{new}(\varphi_i(\phi),1-\iota,1+\iota)=\begin{cases}1-\iota+\alpha A_t,\varphi_i(\phi)\le 1-\iota\\1+\iota-\alpha A_t,\varphi_i(\phi)\ge 1-\iota\\\varphi_i(\phi),\text{other}\end{cases} \tag{3}$$

In equation (3), c^{new} represents the pruning operation, $\varphi_i(\phi)$ represents the probability ratio of the new and old strategies, $1-\iota$ and $1+\iota$ represent the lower and upper bounds of the pruning interval, respectively, α represents the backoff strength coefficient, A_t represents the dominance function, and ι represents the pre-set hyperparameters.

The final network flow incorporating the clip mechanism is shown in Figure 2.

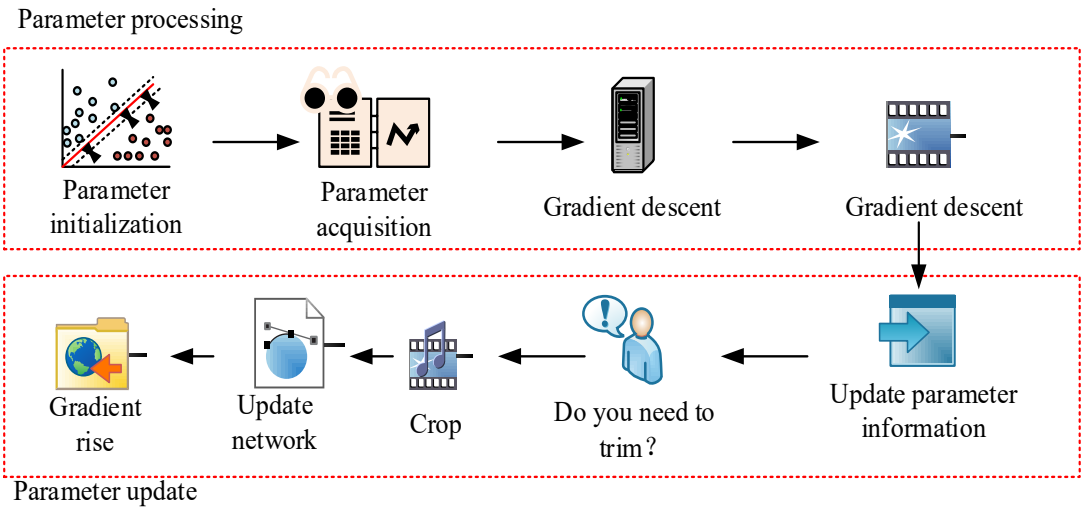

Figure 2: Network process introducing clip mechanism

As shown in Figure 2, during the clip and updating process, the network model first initializes its parameters and then collects data on these parameters through the network structure. The collected parameter data is then updated using gradient descent to update the parameter network information. Next, the network function determines whether the current parameter data needs pruning. If clip is needed, it is performed, and the pruned parameter data is then updated using gradient ascent to update the network. If clip is not needed, the network is directly updated using gradient ascent. Finally, the updated parameter network is used as the output function to terminate the process. Therefore, the entire EFCPPO algorithm's execution flow is shown in Figure 3.

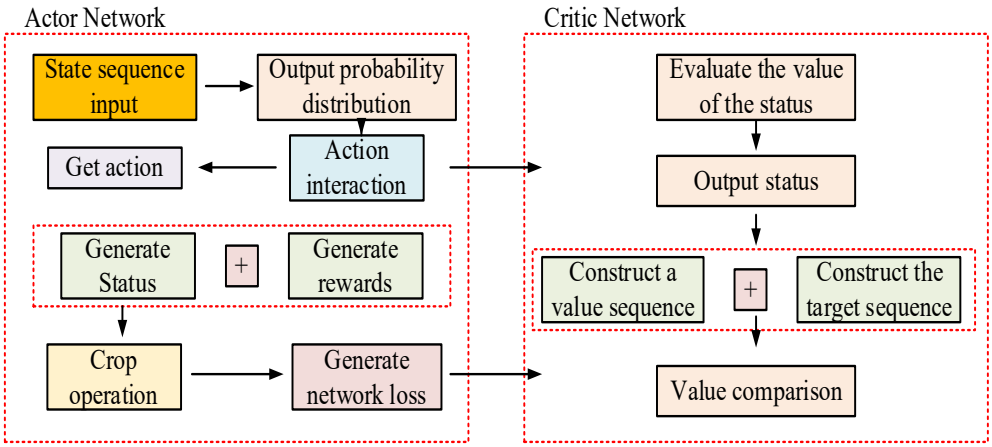

Figure 3: EFCPPO algorithm running process

As shown in Figure 3, the network operation consists of two parts: the Actor network and the Critic network. The Actor network first inputs the state sequence, outputs an action probability distribution, and obtains the specific action through a maximum-value mechanism.

After the action interacts with the environment, a reward and the next state are generated, resulting in a new action sequence and reward sequence. To prevent excessive policy updates, the ratio of new to old values in the network is pruned and, together with the undrained ratio, passed through a minimum-value function to form a conservative policy optimization objective.

This objective is then used to generate the policy network's loss through backpropagation. The Critic network is responsible for evaluating the value of the states. It outputs a value estimate for the input states, forming a value sequence.

A target value sequence is constructed using the new target value. The network's predicted value is compared with the target value. The difference is used to form the Critic network's loss through backpropagation, which is used to improve the accuracy of the value estimate.

2.2 Multi-objective Scale Optimization of System Energy Storage based on EFCPPO Algorithm

The EFCPPO algorithm enables multi-dimensional parameter analysis and updating of power storage systems. However, due to the coupling and efficiency conversion between different energy forms in traditional power storage systems, a single time scale cannot fully analyze the characteristic differences of the storage system. Therefore, this study introduces HRL based on the EFCPPO algorithm. HRL decomposes complex tasks into multiple sub-tasks by constructing a hierarchical architecture of high-level decision-making and low-level execution. The top-level policy is responsible for macro-planning and generating abstract instructions, while the low-level policy completes the action sequence in the specific state space based on the instructions [14-15]. This structure significantly reduces the complexity of the problem through task decomposition, which not only improves the training efficiency and scalability of the agent, but also enhances its decision-making performance in complex environments. This study introduces HRL to perform hierarchical analysis of the regulation of energy storage systems at different time scales. Figure 4 shows the hierarchical energy storage multi-objective scale optimization model.

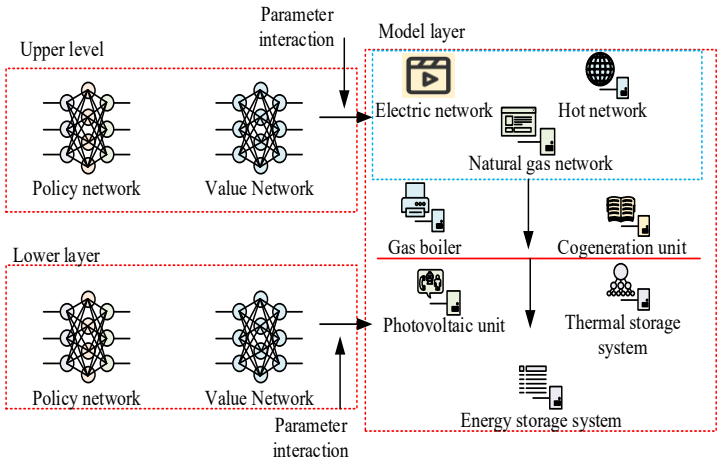

Figure 4: Multi-objective scale optimization model for layered energy storage

As shown in Figure 4, the model consists of three parts: a hierarchical upper and lower layer and a model layer. The model layer is composed of a physical equipment network, including three types of energy subsystems: an electric network, a thermal network, and a natural gas network. The main power grid and the natural gas network serve as energy inputs, connecting energy supply equipment such as electric boilers, gas boilers, combined heat and power units, and photovoltaic units, as well as energy storage units such as thermal storage systems and electric storage systems, ultimately meeting user load demands. The upper-layer network outputs upper-layer actions through its policy network, and these decisions affect the lower-layer state space.

At the same time, the lower-layer network generates lower-layer actions through its own policy network, and these execution results affect the upper-layer state space, forming feedback. The upper and lower layers are equipped with value networks for value function evaluation. Through the interaction of the two-layer policy-value networks, collaborative optimization scheduling at different time scales is achieved. The calculation of the multi-objective scale parameters of the upper-layer model is shown in equation (4).

$$S_t^1 = \left\{ J_l(t), t, c_g(t), P_e(t) \right\} \tag{4}$$

In equation (4), S_t^1 represents the agent state space at time t, $J_l(t)$ represents the heat load power at time t, t represents time, $c_g(t)$ represents the electricity price at time t, and $P_e(t)$ represents the energy storage system power at time t. The agent space action function is shown in equation (5) [16-17].

$$a_t^1 = \left\{ J_C(t), J_E(t), P_T(t) \right\} \tag{5}$$

In equation (5), a_t^1 represents the action space function of the agent at time t, $J_C(t)$ represents the heat power generated by the cogeneration unit at time t, $J_E(t)$ represents the heat power generated by the electric boiler at time t, and $P_T(t)$ represents the heat power generated by the energy storage system at time t. The reward function formula of the upper-level model is shown in equation (6).

$$w_t^1 = \mu_1 w_{t,1} + \mu_2 w_{t,2} \tag{6}$$

In equation (6), w_t^1 represents the reward function at time t, μ_1 and μ_2 both represent weighting factors, and $w_{t,1}$ and $w_{t,2}$ represent the operating cost and low-carbon cost of the upper layer, respectively S_t^2. The state space formula of the lower layer is shown in equation (7).

$$S_t^2 = \left\{ P_l(t), P_p(t), c_g(t), t, J_C(t), J_E(t), J_G(t), P_T(t) \right\} \tag{7}$$

In equation (7), S_t^2 represents the state-space function of the lower layer at time t, $P_l(t)$ represents the electrical load at time t, $P_p(t)$ represents the photovoltaic power generation at time t, and $J_G(t)$ represents the heating power of the gas boiler at time t. The action-space function of the lower layer is shown in equation (8) [18-19].

$$a_t^2 = \left\{ P_e(t) \right\} \tag{8}$$

In equation (8), a_t^2 represents the action space function at time t. The formula for the lower-level establishment function is shown in equation (9).

$$w_t^2 = -(\varphi_1 C_E(t) + \varphi_2 \lambda_c (A_1 P_2(t))) \tag{9}$$

In equation (9), w_t^2 represents the reward function of the lower layer, φ_1 and φ_2 represent the weighting factors, $C_E(t)$ represents the battery operating cost at time t, λ_c represents the carbon emission system, A_1 represents the energy price system, and $P_2(t)$ represents the gas boiler power consumption at time t. At the same time, in order to prevent policy bias, the study also analyzed it through low-level policy networks. The formula for the network reward function of the low-level strategy is shown in equation (10).

$$q_{low}(m,n,i) = -\left\| m - i \right\|^2 \tag{10}$$

In equation (10), $q_{low}(m,n,i)$ represents the reward function value of the low-level policy network. Among them, m represents the given current state, n represents the execution action, and i represents the goal of high-level strategy formulation.

Simultaneously addressing the stationarity of goals by introducing behavior transfer and goal transfer.

The objective reward function formula is shown in equation (11).

$$q_{high}(m,i)=\begin{cases}1, if \left\|m-i\right\| \le \varphi \\ -1, otherwise\end{cases} \tag{11}$$

In equation (11), $q_{high}(m,i)$ represents the reward function of the high-level, φ represents the threshold, and it is determined whether the current high-level reward function has reached the target value. The state space expressions for different parameters are shown in equation (12).

$$m=\begin{cases}p_a(1), p_d(1) \\ p_a(t), p_d(t)\end{cases} \tag{13}$$

In equation (13), $p_a(1)$ represents the total power generation, $p_d(1)$ represents the total charge demand, $p_a(t)$ represents the total power consumption at time t, and $p_d(t)$ represents the total power load demand at time t. The expression of action space parameters is shown in equation (14).

$$n=U(t) \tag{14}$$

In equation (14), A represents the total power output of the action space. Because the EFCPPO model exhibits slow global convergence during MOO of the energy storage system, this study also introduces a group relative measurement optimization method.

Figure 5 shows the operation flow of the strategy optimization method.

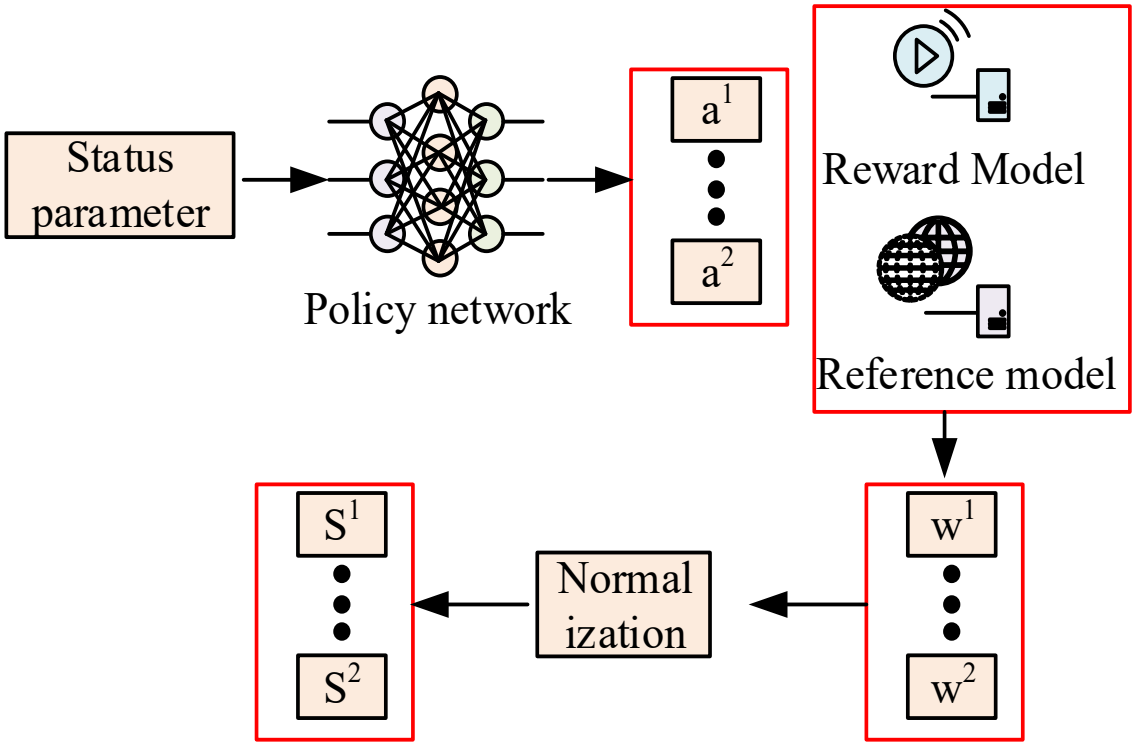

Figure 5: Operation process of strategy optimization method

As shown in Figure 5, the strategy optimization method first inputs multiple cue samples, then inputs the sample parameters into the strategy network and reference model. The strategy network is responsible for generating the corresponding action responses, while the reference model provides the distribution constraints for the strategy output. Subsequently, the reward model calculates the original reward value for each group of action responses.

The reward value is normalized by the grouped relative reward module, transforming it into a relative reward with comparative significance. Finally, the final reward parameters are output by optimizing the strategy network parameters.

The operation flow of the time-scale optimization method for the energy storage system after introducing the two-layer network update is shown in Figure 6.

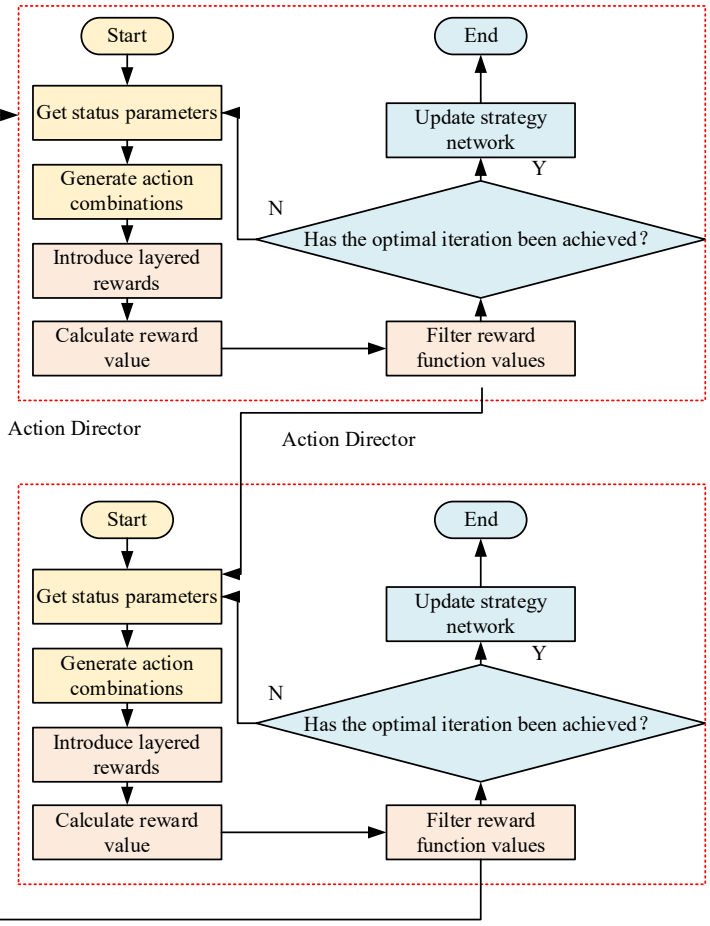

Figure 6: Operation process of time scale optimization method for energy storage system

As shown in Figure 6, the network update process first acquires the system's state parameters, then generates state-action combinations based on these parameters. Next, a tiered reward system is introduced, and the reward value is calculated. Then, the action selection process filters for the most advantageous group reward function values. Finally, it determines whether the optimal iteration has been reached; if so, the policy network is updated; otherwise, state values are reacquired. The action selection in the upper and lower layers directly involves action callbacks to obtain action state values. The operation flow of the upper and lower layers is identical. Through mutual updates between the upper and lower layers, more action states can be obtained, enabling the entire network parameter acquisition process to capture a wider range of state and action states, thus improving the optimization effect.

Develop deep reinforcement learning agents using TensorFlow/PyTorch framework in Python environment, and achieve collaborative optimization of adaptive energy management and intelligent maintenance decision-making through deep Q-network algorithm. The two perform joint simulation through an interface: MATLAB/Simulink outputs real-time equipment status to the Python agent, which selects the optimal action based on the current state and returns to the simulation model, forming a closed-loop iteration of "perception decision optimization". By repeatedly learning from a large number of training scenarios, the comprehensive optimal path for energy efficiency, cost, and reliability throughout the entire lifecycle is ultimately obtained.

3. Results and Analysis

3.1 Comprehensive Dispatch and Optimization Analysis of Power Storage System

To verify the optimization performance of the EFCPPO algorithm in integrated energy systems, this study constructed a test system model of a standard 33-node distribution network and a 16-node thermal network. The photovoltaic power output data and building thermal load curves required for the system were generated using typical meteorological data from Beijing, China, through EnergyPlus building energy consumption simulation software. The grid purchase price and natural gas price parameters referenced energy price data for North China published by the China Electricity Council. The carbon trading price used was the historical settlement price of the Chinese carbon emission trading market. The unit parameters used in this study were as follows: for cogeneration units, fuel cost coefficients were set to -0.1778 and -0.1698, carbon emission intensity was 1.7819, electrical power output range was

100-600kW, thermal power output upper limit was 500kW, and ramp rate was 200kW; for the energy storage system, the rated power was set to 200kW, capacity range was 0-500kWh, charge/discharge efficiency was 99%, and
self-discharge rate was 10%.
The thermal storage system has a capacity range of 0-500 kWh, a thermal storage/release power limit of 100 kW, and a heat loss coefficient of 10%. The gas-fired boiler had a rated thermal power of 200 kW and a thermal efficiency of 90%. The electric boiler had a rated electric power of 150 kW and an electrothermal conversion efficiency of 97%. The grid interaction power was limited to 300 kW. The analysis model's operation under load time is shown in Figure 7, where the grid load during weekdays and holidays is illustrated.

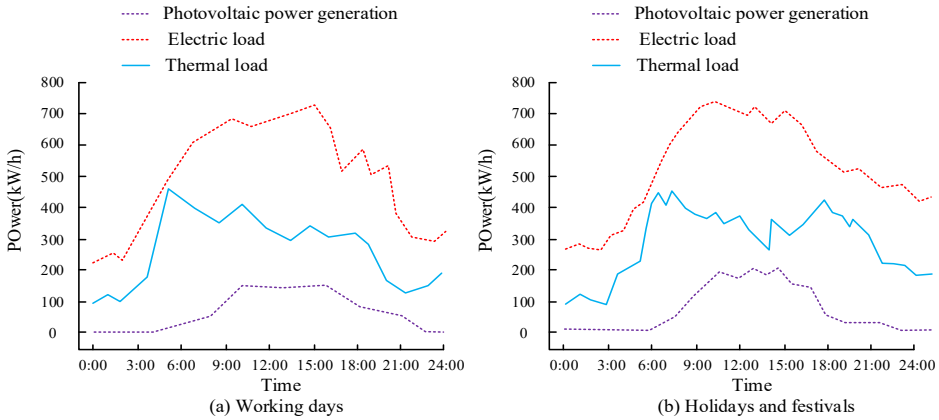

Figure 7: Operating load of power grid energy storage system

As shown in Figure 7(a), the electrical load of the power storage system was at its highest during operation, reaching 710 kW/h, while the highest energy storage load of photovoltaic power generation was only 150 kW/h, a decrease of 560 kW/h compared to the electrical load. This indicates that photovoltaic power generation cannot fully guarantee the operation of the power system throughout the weekday. As shown in Figure 7(b), all three types of electrical load values increased significantly during holidays, indicating a surge in electricity consumption during holidays, which poses a more severe challenge to the optimization of the entire system's energy storage. This study compared and analyzed the multi-objective scheduling results for different dates, comparing the Improved Whale Optimization Algorithm (IWOA), the Multi-objective Jellyfish Search Algorithm (MOISA), and MOO. The results of the electricity load optimization after different algorithms were analyzed and compared, as shown in Figure 8.

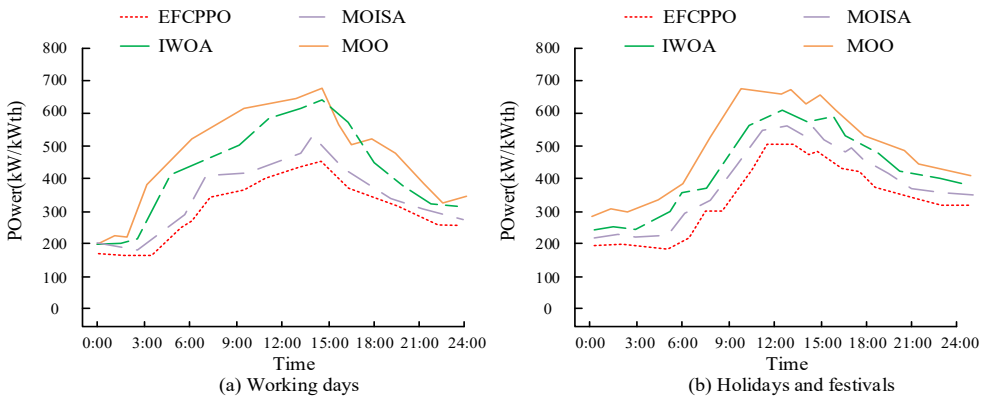

Figure 8: Comparison results of different optimization algorithms

As shown in Figure 8(a), during weekday optimized scheduling, the EFCPPO algorithm used in this study had a lower operating load value, with a maximum load value of only 420 kW/h. In contrast, the MOO algorithm had a relatively higher electrical load value, reaching a maximum of 630 kW/h, an improvement of 210 kW/h compared to the EFCPPO algorithm. This demonstrates that when using different algorithms for scheduling, the EFCPPO algorithm has a lower optimized operating load value, requires less energy, and saves on operating costs. This is because the EFCPPO algorithm incorporates a multi-layer network, which improves the selection of state parameters. As shown in Figure 8(b), during holiday scheduling optimization, the EFCPPO algorithm reduced the operating load value by approximately 200 kW/h compared to the MOO algorithm. This indicates that the EFCPPO algorithm also performs relatively well in holiday scheduling.

The comparative analysis of the operating costs of different algorithms is shown in Table 1.

Table 1. Comparison of operating costs of different algorithms

Day	Operating Cost (yuan)				Pollution Cost (yuan)			
Algorithm Name	IWOA	MOJSA	MOO	EFCPPO	IWOA	MOJSA	MOO	EFCPPO
Day 1	2140	2090	2040	2020	21.7	21.2	20.7	20.1
Day 2	2140	2090	2045	2030	21.8	21.3	20.8	20.3
Day 3	2140	2090	2045	2015	21.9	21.4	20.9	20.2
Day 4	2145	2095	2050	2040	22.1	21.6	21.1	20.5
Day 5	2135	2085	2030	2005	21.8	21.3	20.8	20.0
Day 6	2140	2090	2045	2025	21.9	21.4	20.9	20.4
Day 7	2155	2105	2050	2035	22.3	21.8	21.3	20.6
Day 8	2140	2090	2040	2018	21.7	21.2	20.7	20.1
Day 9	2140	2090	2040	2022	21.8	21.3	20.8	20.3
Day 10	2138	2088	2038	2010	21.9	21.4	20.9	20.2

As shown in Table 1, in the Day 1 test, the operating cost of EFCPPO was 2020 yuan, a decrease of approximately 5.6% compared to IWOA's 2140 yuan. The pollution cost was 20.1 yuan, a decrease of approximately 7.4% compared to IWOA's 21.7 yuan. In the Day 2 scenario, the operating cost of EFCPPO was 2030 yuan, a decrease of approximately 2.9% compared to MOJSA's 2090 yuan. The pollution cost was 20.3 yuan, a decrease of approximately 2.4% compared to MOO's 20.8 yuan. This is because the EFCPPO algorithm, by integrating constraint processing mechanisms and strategy optimization frameworks, can dynamically adapt to the impact of weather changes on system scheduling, thereby achieving more refined multi-objective collaborative optimization. In subsequent continuous tests, EFCPPO maintained the lowest values in terms of both operating cost and pollution cost, especially in the Day 5 scenario, where its operating cost was reduced by approximately 1.2% compared to MOO, and its pollution cost was reduced by approximately 8.3% compared to IWOA. This indicates that in scheduling optimization under various weather conditions, EFCPPO has more stable cost control capabilities and environmental benefits.

3.2 EFCPPO Algorithm Execution Scheduling Optimization Test

To analyze the performance of the algorithm's scheduling optimization after introducing different modules, this study compared and tested the performance of the algorithm with different modules. The same power storage system as described above was used for the test system performance. The results of testing the scheduling effects of different modules are shown in Figure 9.

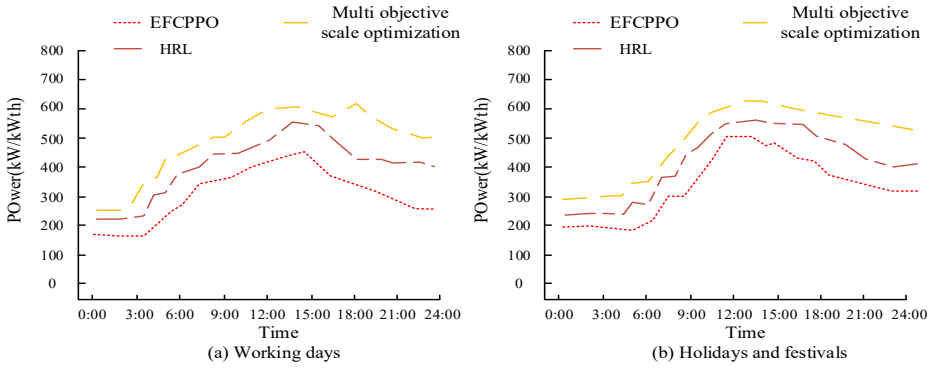

Figure 9: Comparison results of optimization of different models

As shown in Figure 9(a), the EFCPPO algorithm exhibited the best optimization performance among different models, while the multi-objective scale optimization performed the worst. Its highest scheduled electrical load value reached 580 kW/h, a reduction of approximately 160 kW/h compared to the EFCPPO algorithm. This demonstrates a significant improvement in the overall MOO performance after incorporating different modules. Figure 9(b) shows that the EFCPPO algorithm

achieves a lower optimized load value during holiday operations, reducing it by 180 kW/h compared to the multi-objective scale optimization. This indicates that the EFCPPO algorithm demonstrates excellent optimization performance during both holiday and weekday operations. Figure 10 compares the scheduling optimization system results before introducing different modules.

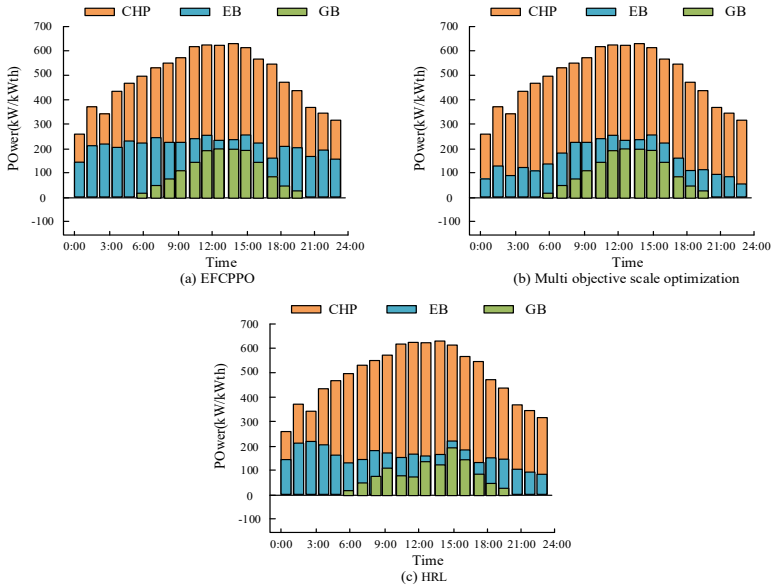

Figure 10: Optimization scheduling results of different modules

Since the operating load and operating cost of CHP accounted for the largest proportion in the power storage system, optimizing the scheduling of this system could reduce the power load and operating cost. As shown in Figures 10(a), 10(b), and 10(c), the EFCPPO algorithm performed relatively well in scheduling among the three models. CHP had a relatively low operating load. Compared to multi-objective scale optimization methods, the EFCPPO algorithm reduced CHP's operating energy consumption by scheduling more EB system energy consumption between 0:00 and 9:00. Furthermore, compared with HRL, the EFCPPO algorithm reduced the operating energy consumption of the CHP system by scheduling more EB and GB system energy consumption, thus reducing the overall system's operating energy consumption and cost. Therefore, the EFCPPO algorithm used in this study has better system optimization and scheduling performance. The comparative analysis of the operating costs of different modules is shown in Table 2.

Table 2. Comparison of operating costs of different modules

Day	Operating Cost (yuan)			Pollution Cost (yuan)		
	Multi-objective Scale Opt.	HRL	EFCPPO	Multi-objective Scale Opt.	HRL	EFCPPO
Day 1	285	2110	2020	20.9	21.3	20.1
Day 2	2095	2120	2030	21.0	21.5	20.3
Day 3	2070	2100	2015	21.1	21.6	20.2
Day 4	2100	2140	2040	21.3	21.8	20.5
Day 5	2060	2085	2005	20.8	21.2	20.0
Day 6	2080	2115	2025	21.0	21.4	20.4
Day 7	2095	2130	2035	21.4	21.9	20.6
Day 8	2075	2105	2018	20.9	21.3	20.1
Day 9	2082	2112	2022	21.0	21.4	20.3
Day 10	2065	2090	2010	21.1	21.5	20.2

As shown in Table 2, in the Day 1 test, the operating cost of EFCPPO was 2020 yuan, which is about 3.1% lower than the 2085 yuan of the multi-objective scale optimization method and about 4.3% lower than the 2110 yuan of the HRL algorithm. In terms of pollution cost control, EFCPPO achieved 20.1 yuan, which is about 3.8% lower than the 20.9 yuan of the multi-objective scale optimization method and about 5.6% lower than the 21.3 yuan of the HRL algorithm. In the extreme test scenario of Day 5, the advantages of EFCPPO were even more obvious. Its operating cost was 2005 yuan, which is about 3.8% lower than the 2085 yuan of the HRL algorithm, and its pollution cost was 20.0 yuan, which is about 3.8% lower than the 20.8 yuan of the multi-objective scale optimization method. This may be because the EFCPPO algorithm, by constructing a multi-objective constraint processing framework, can effectively

balance economic operation and environmental protection requirements, while utilizing the exploratory capabilities of deep reinforcement learning to achieve a dynamic optimization strategy. During a 10-working-day testing period, EFCPPO maintained the best performance across all 20 cost metrics, with its operating costs averaging 2.8%-4.1% lower than the other two methods. A comprehensive comparison shows that EFCPPO significantly outperforms traditional MOO methods and reinforcement learning algorithms in both reducing total system cost and improving environmental benefits.

Conclusions

To achieve multi-objective collaborative optimization of energy storage in power systems, this study proposes a multi-timescale optimization framework based on the EFCPPO algorithm. The new method constructs an Actor-Critic dual-network structure to achieve collaborative updates of policies and values, introduces an exploration backoff pruning mechanism to enhance training stability, and combines HRL to decompose the scheduling task into upper and lower layer decisions, thereby achieving a balanced optimization of operational economy and low-carbon objectives in complex multi-energy systems. The results show that, under typical weekday scenarios, the proposed EFCPPO algorithm achieves a maximum load scheduling value of only 420 kW/h, a reduction of approximately 33.3% compared to the 630 kW/h of traditional MOO methods. Regarding operating costs, the EFCPPO algorithm achieved an average cost of 2022 yuan in a 10-day continuous test, a reduction of 5.5%, 3.2%, and 1.2% compared to the IWOA, MOJSA, and MOO algorithms, respectively, with pollution costs also decreasing by 7.4%, 5.8%, and 3.6%. In module comparisons, EFCPPO further reduces operating costs by 3.1%–4.3% and pollution costs by 3.8%–5.6% compared to multi-objective scale optimization and the HRL method. This demonstrates that the proposed EFCPPO algorithm exhibits better economic efficiency, environmental friendliness, and dispatch adaptability in MOO of energy storage in power systems. Although the study validated the effectiveness of the EFCPPO algorithm in a typical integrated electricity-heat-gas energy system, certain limitations remain. For example, the test scenario is primarily based on meteorological and price data from North China, and the model's generalization ability across different regions and energy structures needs further verification. Future research will expand to multi-regional and multi-energy market environments and explore its integration with other advanced reinforcement learning algorithms.

Funding

The research is supported by: 2025 Henan Provincial Science and Technology Research Project: Research on Key Technologies of UAV Remote Sensing Intelligent Monitoring System for Smart Agriculture, Project No.: 252102210011; Henan Provincial Science and Technology Research Program Project ---- Research on Key Technologies of Multi-Time Scale Cooperative Optimization Scheduling for Source-Grid-Load-Storage Based on DTW Algorithm, No.: 252102241020; 2026 Henan Provincial Universities Key Scientific Research Project: Research on the Application of Load Forecasting Integrating Machine Learning Methods and HGS Algorithm in Combined Cycle Power Generation, No.: 26B470001.

References

[1] Ankar S. J., & Pinkymol K. P. Optimal sizing and energy management of electric vehicle hybrid energy storage systems with multi-objective optimization criterion. IEEE Transactions on Vehicular Technology, 2024, 73(8), 11082-11096.DOI: 10.1109/TVT.2024.3372137.

[2] Wakgra F. G., Kar B., Tadele S. B., Shen S. H., & Khan A. U. Multi-objective offloading optimization in mec and vehicular-fog systems: A distributed-td3 approach. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(11), 16897-16909. https://doi.org/10.1109/TITS.2024.3409367.

[3] Güven A. F., Yörükeren N., & Mengi O. Ö. Multi-objective optimization and sustainable design: a performance comparison of metaheuristic algorithms used for on-grid and off-grid hybrid energy systems. Neural Computing and Applications, 2024, 36(13), 7559-7594. https://doi.org/10.1007/s00521-024-09585-2.

[4] Dhiman G., & Alghamdi N. S. Smose: Artificial intelligence-based smart city framework using multi-objective and iot approach for consumer electronics application. IEEE Transactions on Consumer Electronics, 2024, 70(1), 3848-3855. DOI: 10.1109/TCE.2024.3363720

[5] Moghaddasi K., Rajabi S., & Gharehchopogh F. S. Multi-objective secure task offloading strategy for blockchain-enabled IoV-MEC systems: a double deep Q-network approach. IEEe Access, 2024, 12(1), 3437-3463. DOI: 10.1109/ACCESS.2023.3348513.

[6] Baimbetov D., Salem M., Syrlybekkyzy S., Talantovich N. E., Muralev Y., & Nazari M. A. Optimal Scheduling of a Multi-Energy Hub With Renewables, Hydrogen Vehicles, and Storage Systems. IEEE Access. 2025, 12(10):161709-161728 DOI: 10.1109/ACCESS.2025.3609439.

[7] Rajagopalan A., Nagarajan K., Bajaj M., Uthayakumar S., Prokop L., & Blazek V. Multi-objective energy management in a renewable and EV-integrated microgrid using

an iterative map-based self-adaptive crystal structure algorithm. Scientific Reports, 2024, 14(1), 15652. https://doi.org/10.1038/s41598-024-66644-3.

[8] Zhou Y., Lei L., Zhao X., You L., Sun, Y., & Chatzinotas S. Decomposition and meta-DRL based multi-objective optimization for asynchronous federated learning in 6G-satellite systems. IEEE Journal on Selected Areas in Communications, 2024, 42(5), 1115-1129. https://doi.org/10.1109/JSAC.2024.3365902

[9] Akbari E., Faraji Naghibi A., Veisi M., Shahparnia A., & Pirouzi S. Multi-objective economic operation of smart distribution network with renewable-flexible virtual power plants considering voltage security index. Scientific reports, 2024, 14(1), 19136-19138. https://doi.org/10.1038/s41598-024-70095-1.

[10] Pandya S. B., Kalita K., Čep R., Jangir P., Chohan J. S., & Abualigah L. Multi-objective snow ablation optimization algorithm: An elementary vision for security-constrained optimal power flow problem incorporating wind energy source with FACTS devices. International Journal of Computational Intelligence Systems, 2024, 17(1), 33-34. https://doi.org/10.1007/s44196-024-00415-w.

[11] Islam M. M., Yu T., Giannoccaro G., Mi Y., La Scala M., Nasab M. R., & Wang J. Improving reliability and stability of the power systems: A comprehensive review on the role of energy storage systems to enhance flexibility. IEEE Access, 2024, 12(9), 152738-152765.DOI: 10.1109/ACCESS.2024.3476959.

[12] Younesi A., Oustad E., Abolnejadian M., Ansari M., & Ejlali A. Moticps: Energy optimization on multi-objective task scheduling in IoT-integrated cyber-physical systems. IEEE Transactions on Sustainable Computing.2025,2(1): 744-755DOI: 10.1109/TSUSC.2024.3525090.

[13] Li L., Sun Y., Han Y., & Chen W. Seasonal hydrogen energy storage sizing: Two-stage economic-safety optimization for integrated energy systems in northwest China. IScience,2024, 27(9):5-7. https://doi.org/10.1016/j.isci.2024.110691

[14] Wen X., Shen Q., Zheng W., & Zhang H. AI-driven solar energy generation and smart grid integration: A holistic approach to enhancing renewable energy efficiency. Academia Nexus Journal, 2024, 3(2):14-15. https://academianexusjournal.com/index.php/anj/article/view/9.

[15] Cheng, Y., Cao, Z., Zhang, X., Cao, Q., & Zhang, D. (2024). Multi objective dynamic task scheduling optimization algorithm based on deep reinforcement learning. The Journal of Supercomputing, 80(5), 6917-6945. https://doi.org/10.1007/s11227-023-05794-z.

[16] Agajie E. F., Agajie T. F., Amoussou I., Fopah-Lele A., Nsanyuy W. B., Khan B.,... & Tanyi E. Optimization of off-grid hybrid renewable energy systems for cost-effective and reliable power supply in Gaita Selassie Ethiopia. Scientific Reports, 2024, 14(1), 10929-10930. https://doi.org/10.1038/s41598-024-61783-z.

[17] Song F., Deng M., Liu Y., Ye, F. Energy-efficient trajectory optimization with wireless charging in UAV-assisted MEC based on multi-objective reinforcement learning. IEEE Transactions on Mobile Computing, 2024, 23(12), 10867-10884. DOI: 10.1109/TMC.2024.3384405.

[18] Li, J., Fang, Z., Wang, Q., Zhang, M., Li, Y., & Zhang, W. Optimal operation with dynamic partitioning strategy for centralized shared energy storage station with integration of large-scale renewable energy. Journal of Modern Power Systems and Clean Energy, 2024, 12(2), 359-370. DOI: 10.35833/MPCE.2023.000345.

[19] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari. Machine Learning Methodology for Identifying Vehicles Using Image Processing. AIA, 2023,1(3):170-178, https://doi.org/10.47852/bonviewAIA3202833.