



Contents list available at CBIORE journal website

**International Journal of Renewable Energy Development**

Journal homepage: <https://ijred.cbiorc.id>



Research Article

# Energy optimization management of microgrid using improved soft actor-critic algorithm

Zhiwen Yu<sup>1\*</sup>, Wenjie Zheng<sup>1</sup>, Kaiwen Zeng<sup>1</sup>, Ruifeng Zhao<sup>1</sup>, Yanxu Zhang<sup>2</sup>, Mengdi Zeng<sup>2</sup>

<sup>1</sup>Electric Power Dispatching & Control Center of Guangdong Power Grid, Guangzhou 510600, China

<sup>2</sup>Guangdong Provincial Key Laboratory of Smart Grid New Technology Enterprises, China Southern Power Grid Technology Co., Ltd., Guangzhou 510180, China.

**Abstract.** To tackle the challenges associated with variability and uncertainty in distributed power generation, as well as the complexities of solving high-dimensional energy management mathematical models in microgrid energy optimization, a microgrid energy optimization management method is proposed based on an improved soft actor-critic algorithm. In the proposed method, the improved soft actor-critic algorithm employs an entropy-based objective function to encourage target exploration without assigning significantly higher probabilities to any part of the action space, which can simplify the analysis process of distributed power generation variability and uncertainty while effectively mitigating the convergence fragility issues in solving the high-dimensional mathematical model of microgrid energy management. The effectiveness of the proposed method is validated through a case study analysis of microgrid energy optimization management. The results revealed an increase of 51.20%, 52.38%, 13.43%, 16.50%, 58.26%, and 36.33% in the total profits of a microgrid compared with the Deep Q-network algorithm, the state-action-reward-state-action algorithm, the proximal policy optimization algorithm, the ant-colony based algorithm, a microgrid energy optimization management strategy based on the genetic algorithm and the fuzzy inference system, and the theoretical retailer strategy, respectively. Additionally, compared with other methods and strategies, the proposed method can learn more optimal microgrid energy management behaviors and anticipate fluctuations in electricity prices and demand.

**Keywords:** Energy optimization management; Electricity rate; Microgrid; Reinforcement learning; Soft actor-critic algorithm



@ The author(s). Published by CBIORE. This is an open access article under the CC BY-SA license (<http://creativecommons.org/licenses/by-sa/4.0/>).

Received: 1<sup>st</sup> Dec 2023; Revised: 26<sup>th</sup> January 2024; Accepted: 20<sup>th</sup> Feb 2024; Available online: 28<sup>th</sup> February 2024

## 1. Introduction

Energy supply and management have always been one of the significant challenges faced by modern society. With the rapid development of renewable energy and the widespread adoption of distributed energy sources, microgrids (MGs) has garnered extensive attention as an emerging paradigm in energy systems (Saeed *et al.*, 2023; Kheiter *et al.*, 2022). MGs, often incorporating renewable energy sources and energy storage, have emerged as viable solutions to address the challenges of conventional centralized power systems (Ahmad *et al.*, 2023; Mostefa *et al.*, 2023). MGs, decentralized energy systems that can operate independently or in conjunction with the main grid, also offer a promising avenue for enhancing energy efficiency, resilience, and sustainability, which makes them well-suited for supporting peak load management on the main power grid (Mostefa *et al.*, 2023; Lagouir *et al.*, 2021).

Efficient energy management is critical for the seamless functioning of MGs, especially in the context of increasing renewable energy penetration (Alzahrani *et al.*, 2023; Shezan *et al.*, 2023). Energy optimization not only ensures the reliable and stable operation of MGs but also contributes to cost reduction, environmental sustainability, and grid resilience. In traditional power systems, centralized energy management and control strategies are typically sufficient to meet the requirements for

their efficient operation (Tajjour *et al.*, 2023; Dey *et al.*, 2023). However, in MGs, due to the constraints imposed by diverse energy resources and constantly changing power demands, traditional energy optimization management methods are no longer applicable (Alamir *et al.*, 2023; Zhang *et al.*, 2023). Therefore, as the complexity of MG systems grows, the need for intelligent and adaptive algorithms becomes paramount to address the intricacies associated with energy generation, storage, and consumption (Jahani *et al.*, 2023; El Bourakadi *et al.*, 2022).

Hou *et al.* (2023) focused on a proactive energy management and collaborative optimization method using multi-stakeholder game used for a community MGs with multiple renewable energy sources. Simulation results demonstrated that this approach can effectively enhance the economic viability of electricity consumption for users and the on-site integration capacity of renewable energy. Gao *et al.* (2023) developed a remote island MG energy management strategy based on master-slave games. The economic and practical efficiency of this strategy was validated through simulation examples. Leonori *et al.* (2020) introduced a MG energy optimization management strategy based on genetic algorithms. The effectiveness of the model and algorithm was verified through a small-scale MG case study. The case study results indicated that the model achieved optimal energy management for the MG.

\* Corresponding author  
Email: [Zhiwen\\_yu2023@126.com](mailto:Zhiwen_yu2023@126.com) (Z. Yu)

Datta *et al.* (2023) presented an energy management scheme for multi-MG systems, encompassing distributed resource scheduling, renewable energy integration, and plug-in electric vehicle penetration. The presented method was validated on a test system comprising residential, commercial, and industrial MGs, with simulation results confirming its effectiveness. Alamir *et al.* (2023b) developed an improved version of the artificial rabbit optimization algorithm, enhanced with quantum mechanics and the Monte Carlo method for optimal MG energy scheduling. Simulation results affirmed the effectiveness of the developed approach in reducing operation costs and maximizing energy usage efficiency. Rodriguez *et al.* (2023) outlined a MG energy management system utilizing fuzzy logic control. Simulation results demonstrated that the developed method efficiently harnessed available solar energy resources while extending batteries' life through the efficient control of their over-charging and deep discharging.

While the aforementioned studies have achieved commendable results in MG energy management, they necessitate a comprehensive understanding of the dynamic current and subsequent states of the MG. In recent years, reinforcement learning (RL) has garnered extensive research interest as a machine learning method and has shown remarkable success in addressing complex control issues (Ibrahim *et al.*, 2023; Cavus *et al.*, 2023). The soft actor-critic (SAC) algorithm, rooted in RL, has demonstrated remarkable success in optimizing complex systems (Kim *et al.*, 2023; Dong *et al.*, 2024). Originally designed for robotic control tasks, SAC has shown versatility and adaptability, making it an attractive candidate for application in MG energy management (Bao *et al.*, 2023; Topa *et al.*, 2021).

To address the challenges present in the aforementioned researches, this paper introduces an energy optimization management approach based on an improved SAC algorithm. The main contributions of this research is summarized as follows. By incorporating entropy into the objective function, the algorithm strikes a delicate balance between exploration and exploitation. Unlike traditional methods that assign high probabilities to certain actions, the entropy-based approach maintains a level of uncertainty, allowing the algorithm to explore a broader action range. This innovative strategy enhances the adaptability of the improved SAC algorithm, making it particularly well-suited for dynamic environments where the optimal action may change over time. The incorporation of entropy not only fosters exploration but also contributes to the robustness and versatility of the algorithm in real-world applications. Furthermore, the significant breakthrough of the proposed method also lies in its ability to

effectively address the challenges associated with MG energy management. The inherent variability and uncertainty of DGs pose formidable obstacles to achieving reliable and efficient energy management. The conventional mathematical models used in MG energy management often suffer from convergence brittleness, making them less adaptive to dynamic operating conditions. Through the application of the proposed method, these challenges are mitigated. The algorithm demonstrates a remarkable capacity to handle variability and uncertainty, providing a more resilient and robust solution to the intricacies of MG energy management. This breakthrough not only improves the stability of MG operations but also opens up new possibilities for the integration of renewable energy sources. In addition, in the realm of electricity price prediction for MGs, the proposed method presents a noteworthy departure from conventional approaches. Unlike methods that require an exhaustive understanding of MG dynamics for accurate predictions, the proposed approach takes a more streamlined and efficient route. The necessity for a comprehensive solution to MG dynamics is alleviated, streamlining the prediction process. This not only reduces the computational complexity associated with forecasting purchase and sale prices but also enhances the applicability of the method to a wider range of scenarios. By focusing on key features and leveraging the inherent adaptability of the SAC algorithm, the proposed approach achieves accurate and efficient electricity price predictions without the need for an overly intricate understanding of MG dynamics. This innovation marks a significant advancement in the field of MG energy economics, making it more accessible and practical for real-world implementation.

The aim of this research is to effectively optimize energy management within MGs and enhance the reliability and economic efficiency of MG energy management, during tackling the challenges associated with variability and uncertainty in distributed power generation, as well as the complexities of solving high-dimensional energy management mathematical models in microgrid energy optimization.

## 2. Energy management method of microgrid based on an improved soft actor-critic algorithm

### 2.1 Microgrid architecture and configuration

The schematic diagram of a MG under investigation, situated in the local area, is depicted in Fig. 1. This MG configuration encompasses several integral components:

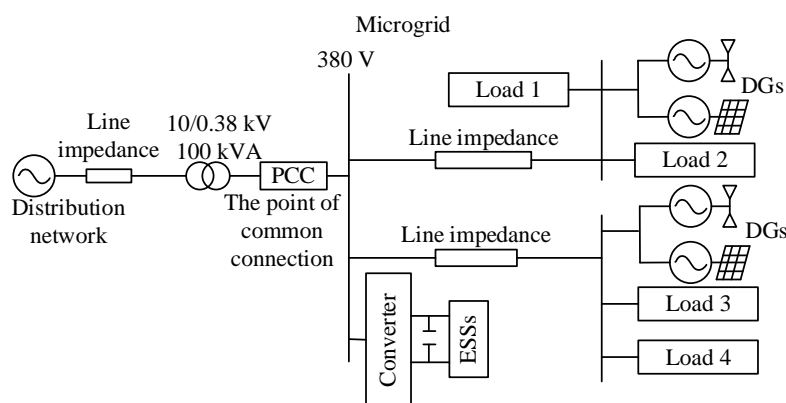


Fig. 1 Microgrid schematic diagram.

**Table 1**  
Parameters of DGs.

DGs	Diesel generator	PV	WT	BESS
Parameters	Maximum generation power=20 kW Minimum generation power=4 kW	Maximum generation power=10kW	Maximum generation power=10kW	Capacity=10kWh Maximum discharging power=10kW Maximum charging power=10kW

distributed generators (DGs), energy storage systems (ESSs), temperature control loads (TCLs), and residential loads.

The analyzed MG in Figure 1 is interconnected with various DGs, including a diesel generator, four photovoltaic generator systems (PVs), four wind turbines (WTs), and three battery energy storage systems (BESSs). In Fig. 1, a single diesel generator with a maximum capacity of 20 kW is considered in the energy optimization model of MG. The parameter values associated with DGs are presented in Table 1.

## 2.2 Energy network components

The MG is interconnected with the distribution network (DN), facilitating the continuous exchange of electrical energy within the electricity market through the distribution network. Effective bidirectional communication, established between each component within the MG and the energy management system (EMS) is utilized to enable the exchange of vital information, including electricity pricing, battery charging/discharging status, and power generation statistic data (Zhang *et al.*, 2021). The EMS efficiently utilizes smart agents to relay control signals to the various components, manage on/off operations for TCLs, charge/discharge procedures for ESSs, and send or receive control signals of purchasing/selling energy from/for the DN.

### 2.2.1. Energy storage systems

ESSs play a pivotal role in modern energy management, providing a mean to store and release energy when needed. These systems contribute to grid stability, enable the integration of renewable energy sources, and enhance the overall reliability of energy supply. The ESSs studied in the MG under investigation refer to the community-level ESSs rather than individual residential battery ESSs. The capacity of an ESS utilized is designed to meet a minimum of 2 hours' worth of energy demands for all users within the MG. At each time step  $t$ , the dynamic variation in energy within an ESS is modeled as follows

$$B_t = B_{t-1} + \eta_c C_t - \frac{D_t}{\eta_d} \quad (1)$$

where  $B_t \in [0, B_{max}]$  represents the energy stored by the analyzed ESS at a given time step  $t$ .  $B_{max}$  stands for the maximum capacity of the analyzed ESS.  $\eta_c \in [0, 1]$  and  $C_t \in [0, C_{max}]$  represent the charging efficiency and power of the analyzed ESS, respectively.  $\eta_d \in [0, 1]$  and  $D_t \in [0, D_{max}]$  denote the discharging efficiency and power of the analyzed ESS respectively.  $C_{max}$  and  $D_{max}$  represent the upper limits of  $C_t$  and  $D_t$ , respectively. The state of charge (SOC) of the analyzed ESS is defined as follows:

$$B_{BSC,t} = \frac{B_t}{B_{max}} \quad (2)$$

ESSs examine their discharging conditions and release available electrical energy. If the ESSs cannot fully supply the requested electrical energy as per EMS's request, the remaining amount is automatically provided by the DN.

### 2.2.2. Distributed generators

DGs are small-scale power generation sources that are decentralized and located close to the end-users or within the DN. Unlike traditional centralized power plants, DGs contribute to local generation and can operate independently or in conjunction with the main grid. These systems offer various benefits, including improved grid resilience, increased energy efficiency, and support for renewable energy integration. The MG is equipped with DGs capable of producing varying amounts of electrical energy based on weather conditions. The utilized DGs does not rely on a generative model analysis but rather employs actual distributed generation data, such as wind power data obtained from NREL. DGs share real-time generation data with EMS ( $G_t$ ) and can directly supply power to the local grid.

### 2.2.3. Distribution network

When DGs' supply falls shortly, the DN can promptly provide electrical energy to the MG. When there is an excess of electrical energy, the DN can also accept surplus electrical energy from the MG. The energy transactions between the DN and DGs are carried out in real-time by adjusting the fluctuations in electrical energy pricing in the electricity market, where the upward and downward price adjustments are denoted as  $K_t^u$  and  $K_t^d$ , respectively.

To establish the priority power source during electricity supply shortages and the priority discharge source during electricity surpluses, EMS exclusively manages the electrical switches of the DN. Following  $t$ , the EMS of a DG receives information regarding the amount of electricity ( $E_t$ ) to be purchased or sold to the DN. A positive value of  $E_t$  indicates purchasing electrical energy from the market, while a negative value of  $E_t$  signifies selling electrical energy to the market.

### 2.2.4. Temperature control loads

TCLs refer to systems or devices that are designed to regulate and maintain a specific temperature within a given space. These loads are prevalent in various sectors, including residential, commercial, and industrial settings, and they play a crucial role in ensuring comfort, preserving products, and supporting industrial processes. TCLs adhere to the principle of heat conservation. TCLs can serve as a significant source of flexibility in energy provision. TCLs can be directly controlled at  $t$  by the agents, and the energy cost incurred by users to maintain indoor comfort is accounted for  $C_{gen}$ . The controller for TCLs receives on/off operation signals from the agents. The TCL controllers perform on/off operations by examining temperature constraints and based on the following conditions.

$$a_{b,t}^i = \begin{cases} 0 & \text{if } T_t^i > T_{max}^i \\ a_t^i & \text{if } T_{min}^i < T_t^i < T_{max}^i \\ 1 & \text{if } T_t^i < T_{min}^i \end{cases} \quad (3)$$

where  $a_{b,t}^i$  represents the on/off operation command given by a TCL controller.  $a_t^i$  denotes on/off operation command when  $T_{min}^i < T_t^i < T_{max}^i$ .  $T_t^i$  is the operating temperature of the analyzed

TCL (heat pump, water heater, refrigerator, etc.)  $i$  at the time step  $t$ , or the indoor temperature in the room where the TCL (air conditioner)  $i$  is located at the time step  $t$ .  $T_{max}^i$  and  $T_{min}^i$  respectively stand for the maximum and minimum temperatures determined by users. The research conducted primarily focuses on air conditioner-type TCLs, and the dynamic temperature variation process in the room where the analyzed TCL is located is modeled as follows:

$$\dot{T}_t^i = \frac{1}{C_a^i}(T_t^0 - T_t^i) + \frac{1}{C_m^i}(T_{m,t}^i - T_t^i) + L_{acl}^i U_{b,t}^i + q^i \quad (4)$$

$$\dot{T}_{m,t}^i = \frac{1}{C_m^i}(T_t^i - T_{m,t}^i) \quad (5)$$

where  $T_t^i$  represents the indoor temperature.  $T_{m,t}^i$  is the temperature of the TCL's output air,  $T_t^0$  is the outdoor temperature,  $C_a^i$  and  $C_m^i$  represent the thermal mass of the air and building materials, respectively.  $q^i$  denotes an uncontrollable heat load within the analyzed building.  $L_{acl}^i$  represents the TCL's rated output power.

### 2.2.5. Residential loads

The analyzed residential loads represent household electricity demands that cannot be directly controlled within the MG. The analyzed residential loads exhibit daily fluctuations, with their variable energy consumption influenced by electricity pricing. The modeling of electricity load  $L_t^j$  in household  $j$  over time step  $t$  is as follows:

$$L_t^j = L_{b,t} - S_{SL,t}^j + S_{PB,t}^j \quad (6)$$

$$S_{SL,t}^j = L_{b,t} \times \beta_j \times \delta_t \quad (7)$$

where  $L_{b,t} > 0$  represents the daily baseline load power (Nakabi et al, 2019). The sensitivity parameter  $\beta_i$  (where  $\beta_i \in [0,1]$ ) indicates the proportion of load power variation relative to the price fluctuation magnitude.  $S_{SL,t}^j$  is the transferred loads, and  $\delta_t$  is the change in electricity pricing at the time step  $t$  compared to the previous moment.  $S_{SL,t}^j$  is a positive value for high electricity prices ( $\delta_t > 0$ ) and a negative value for low electricity prices ( $\delta_t < 0$ ).  $S_{PB,t}^j$  corresponds to the load amount transferred from the previous time period to the current one. For a given time, positively transferred load charges must operate after a certain time delay, while negatively transferred load charges will be retained in the subsequent time steps.

### 2.3. Energy management objective

The goal of energy optimization management in MGs is to identify the optimal strategy  $\pi^*$  that satisfies the following conditions.

$$J^* = \max_{\pi} J_{\pi} = \max_{\pi} E\{r_t | s_t\} \quad (8)$$

where  $J_{\pi}$  represents the expected total profit for MG under a given control strategy  $\pi$ .  $E\{r_t | s_t\}$  denotes the average value of  $r_t$  under a certain state  $s_t$ .

In the training process of a RL algorithm, four elements are primarily employed: the state space  $S$ , the action space  $A$ , the state transition probabilities  $p$ , and the value function  $r$  (Homod et al, 2022; Moos et al, 2022). During the learning process, the agent interacts with the system and an operation is carried out at  $a_t \in A \subseteq \mathbb{R}^{n_A}$ , where  $a_t$  represents the action executed at time step  $t$ . Subsequently, the system is transitioned from  $s_t \in S \subseteq \mathbb{R}^{n_S}$  to its subsequent state  $s_{t+1}$  (Du et al, 2022; Shavandi et al, 2022).

Assuming for a given strategy  $\pi$ , the value function over a certain time interval ( $V^{\pi}(s_t)$ ) depends on the sequence of state transitions ( $T(s, a, s') = p(s_{t+1} = s' | s_t = s, a_t = a)$ ). The discounted sum of future profits generated is represented as (Singh et al, 2023; Tian et al, 2023)

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots$$

The value function of the expected state discounted under the strategy  $\pi$  in the given state  $s_t$  is described as (Park et al, 2022; Zhu et al, 2022)

$$V^{\pi}(s_t) = E_{\pi}\{R_t | s_t\} = E_{\pi}\{r_t + \gamma V^{\pi}(s_{t+1})\} \quad (9)$$

where  $\gamma \in (0,1]$  is the discount factor that weights future profits.  $V^*(s_t)$  is used to represent the maximum discounted profit that the agent obtains when execute the optimal strategy starting from a state  $s_t$ . From this, it can be seen that (Alavizadeh et al, 2022; Kosuru et al, 2022)

$$\pi^* = \operatorname{argmax}_{\pi} V^{\pi}(s_t) \quad (10)$$

Meanwhile,  $Q$  expresses the expected returns of choosing an action to be operated in  $s_t$  and following the subsequent decision  $\pi$  thereafter.

$$Q^{\pi}(s_t, a_t) = r_t + \gamma V^{\pi}(s_{t+1}) \quad (11)$$

### 2.4. The optimization technique

The soft actor-critic (SAC) algorithm, based on the maximum entropy RL framework, is considered as a non-policy algorithm (Sun et al, 2022). It is capable of handling continuous action spaces, thus enhancing its applicability to various control problems. The action-critic architecture is employed by SAC (Zheng et al, 2023).

The action-critic architecture employs two distinct deep neural networks to approximate the function  $Q$  and the state value function  $V$  (Zheng et al, 2023). The actor maps the current state to what it deems the optimal action, while the critic evaluates actions by computing the value function. The SAC algorithm operates within the maximum entropy RL framework, aiming to maximize the expected reward and entropy, which is to

$$J_{\pi} = \operatorname{argmax}_{\pi} E_{\pi} \left\{ \sum_{t=0}^T \gamma^t (r_t + \alpha H_t^{\pi}) \right\} \quad (12)$$

where  $H^{\pi}$  represents the Shannon entropy term, indicating the agent's uncertainty when taking random actions.  $\alpha$  is the regularization coefficient, signifying the importance of the entropy term on the rewards. Generally, when considering traditional reinforcement learning algorithms,  $\alpha$  is set to zero. Maximizing this objective function ensures that the agent is explicitly encouraged to explore new strategies while also preventing it from making suboptimal actions (Xu et al, 2022).

In the context of the studied method for performing energy optimization in MGs, SAC algorithm primarily utilizes three network functions. These are the state value function  $V$  parameterized by  $\psi$ , the smoothing function  $Q$  parameterized by  $\theta$ , and the policy function  $\pi$  parameterized by  $\psi$  (Huang et al, 2021).

Firstly, the state value function is trained by minimizing the following error (Tightiz et al, 2023).

$$J_V(\psi) = E_{s_t} \left[ \frac{1}{2} \left( V_{\psi}(s_t) - E_{a_t \sim \pi_{\theta}} (Q_{\theta}(s_t, a_t) - \log \pi_{\psi}(a_t | s_t)) \right)^2 \right] \quad (13)$$

That is, the training of the state value function  $V$  involves minimizing the squared difference between the predicted value of the function  $Q$  and its expected predicted value, along with maximizing the entropy of the policy function  $\pi$  (measured by the negative logarithm of the policy function) to the greatest extent possible. To train the approximate function for the policy  $\pi$ , the following error is minimized (Hu *et al*, 2022).

$$J_{\pi}(\phi) = E_{s_t} \left[ D_{KL} \left( \pi_{\phi}(s_t) \parallel \frac{\exp(Q_{\theta}(s_t))}{Z_{\theta}(s_t)} \right) \right] \quad (14)$$

where the term  $D_{KL}$  is the Kullback-Leibler divergence. By seeking the minimum value of the above objective function, it ensures that the policy function's distribution becomes more similar to the exponential distribution form of normalized function  $Q$  by normalization function  $Z$ .

To improve the performance of the SAC algorithm and minimize this objective, reparameterization is employed using  $a_t = f_{\phi}(\mathbb{Q}_t; s_t)$ . This technique is used to ensure that the policy sampling process is a differentiable one. The parameterized policy can be represented as follows (Xiong *et al*, 2023):

$$J_{\pi}(\phi) = E_{s_t, \mathbb{Q}_t} [lg \pi_{\phi}(f_{\phi}(\mathbb{Q}_t, s_t) | s_t) - Q_{\theta}(s_t, f_{\phi}(\mathbb{Q}_t, s_t) | s_t)] \quad (15)$$

Since the normalization function  $Z$  is independent of the parameter  $\phi$ , it is discarded. The unbiased estimate of the gradient for the aforementioned objective is as follows:

$$\hat{\nabla}_{\phi} J_{\pi}(\phi) = \nabla_{\phi} lg \pi_{\phi}(a_t | s_t) + (\nabla_{a_t} lg \pi_{\phi}(a_t | s_t) - \nabla_{a_t} Q(s_t, a_t)) \nabla_{\phi} f_{\phi}(\mathbb{Q}_t, s_t) \quad (16)$$

## 2.5. Simulation test configuration

The proposed energy management method was implemented using the OpenAI Gym tool in a MG operational environment described in the second section (Huang *et al*, 2022). The MG energy management process was run for a total of 10 days within the constructed MG operational environment. This process was divided into multiple stages, with each stage lasting for one day. During each stage, one day was randomly selected from the dataset of 10 days. Each time step  $t$  represents an hour, resulting in 24 time steps per day.

Initially, the current state of the MG is determined based on the average power consumption state of the TCLs ( $S_{SoC,t}$ ), the charging/discharging status ( $B_{BSC,t}$ ) of the ESSs, the utility meter state ( $C_t^b$ ), temperature ( $T_t^i$ ), generated power ( $G_t$ ), electricity price ( $K_t$ ) in the power market, time step  $t$ , and the current load value ( $L_{b,t}$ ). These variables are organized using a vector representation.

$$s_t = [S_{SoC,t}, B_{BSC,t}, C_t^b, T_t^i, G_t, K_t, L_{b,t}, t] \quad (17)$$

Next, a reward function  $r_t$  is designed to maximize the economic profit of the MG. The reward function  $r_t$  represents the operating profit margin, which is the income obtained from selling electrical energy to users and the DN minus the costs associated with DN generated, purchased, and transmitted power. The reward function  $r_t$  is defined as follows:

$$r_t = C_{L,t} - C_{cost,t} \quad (18)$$

$$C_{L,t} = P_t \sum_{Loads} L_t^i - C_{gen} \sum_{TCLs} L_{TCL}^i u_{b,t}^i + K_t^b E_t^V \quad (19)$$

$$C_{cost,t} = (K_t^a + C_{trimp}) E_t^C + C_{trexp} E_t^V \quad (20)$$

where  $K_t^a$  and  $K_t^b$  are the up- and down-regulation rates respectively, which are the rates for selling electrical energy to and buying electrical energy from the DN, respectively.  $G_t$ ,  $E_t^V$ , and  $E_t^C$  represent the amount of electrical energy generated, sold to and purchased from the DN, respectively.  $C_{gen}$  is the energy generation cost charged to customers with controlled loads.  $u_{b,t}^i$  is the binary variable corresponding to the on or off actions for the TCLs.  $C_{trimp}$  and  $C_{trexp}$  are the costs associated with importing and exporting electrical energy to the distribution network, respectively.

To evaluate the effectiveness of the proposed method, it was compared with two energy management methods:

i) Strategy 1: A MG energy optimization management strategy based on the genetic algorithm and the fuzzy inference system (Leonori *et al*, 2020). Leonori *et al*. (2020) proved that this strategy is capable of obtaining optimal information regarding electricity generation, electricity consumption, electricity prices, and temperature. The model of Strategy 1 involves the integration of a Microgrid (MG) energy optimization management strategy utilizing both genetic algorithms and fuzzy inference systems (FIS). Here's a breakdown of the model: 1) Hierarchical Genetic Algorithm (GA) and Fuzzy Inference System (FIS): The Energy Management System (EMS) of Strategy 1 is synthesized through a hierarchical genetic algorithm and fuzzy inference system (FIS). This hybrid approach combines the optimization capabilities of genetic algorithms with the decision-making flexibility of fuzzy logic, 2) Fuzzy Inference System (FIS) Design: The FIS is responsible for defining the consequent part of each rule, tuning the membership functions (MFs) position and shape, setting the rule weights, and eliminating input MFs deemed ineffective. This ensures that the FIS accurately captures the system's behavior and effectively guides decision-making, 3) Optimization Parameters Setting: Once the EMS of Strategy 1 is designed and its optimization parameters are set, the model proceeds to implement various strategies for tuning the FIS parameters. This iterative process allows for refinement and optimization of the FIS to enhance its performance in managing MG energy, 4) Tuning FIS Parameters: Different strategies are employed to tune the FIS parameters, which may include adjusting the shape and position of membership functions, optimizing rule weights, and removing redundant input membership functions. These strategies aim to improve the accuracy and efficiency of the FIS in making energy management decisions. In summary, the model utilizes a hybrid approach combining genetic algorithms and fuzzy inference systems to develop an Energy Management System for a Microgrid. The FIS plays a central role in decision-making, with its parameters optimized through iterative tuning processes to enhance the overall performance of the energy optimization strategy.

ii) Strategy 2: Theoretical retailers. These retailers purchase an exact amount of electrical energy in a day-ahead electricity market and sell it to a simulated user group at market prices. The model of strategy 2 involves several key components: 1) Theoretical Retailers: The model features theoretical retailers as virtual entities within the simulation framework. These retailers do not physically exist but are instead conceptual representations used to simulate the behavior of actual market participants, 2) Procurement Process: Within the model, theoretical retailers engage in the process of purchasing a predetermined quantity of electrical energy. This procurement occurs in the day-

ahead electricity market, where transactions for future delivery are conducted based on forecasted demand and supply conditions, 3) Exact Quantity Procurement: The model specifies that theoretical retailers procure an exact amount of electrical energy. This suggests a precise allocation of resources, indicating a predetermined quantity agreed upon by the retailers based on their anticipated demand or contractual obligations, 4) Day-Ahead Electricity Market: The procurement process takes place within the day-ahead electricity market, a segment of the wholesale electricity market where buyers and sellers trade electricity contracts for delivery on the following day. This market allows participants to plan and secure their energy needs in advance, 5) Energy Distribution to Simulated User Group: After procuring electrical energy, theoretical retailers distribute it to a simulated user group. This user group represents consumers or end-users within the simulation environment and serves as the recipient of the electricity supplied by the retailers, 6) Pricing Mechanism: The electrical energy is sold to the simulated user group at market prices. This implies that the pricing mechanism within the simulation model is based on prevailing market conditions, reflecting the supply-demand dynamics and pricing signals observed in real-world electricity markets, 7) Overall, the model simulates the behavior of theoretical retailers in procuring electrical energy from the day-ahead electricity market and subsequently distributing it to a simulated user group, all while adhering to market prices.

These comparisons were made to assess the performance of the proposed method relative to these alternative energy management strategies. To compare the performance of the proposed method with other DL approaches, the energy management results of the proposed method were also compared with two other DL algorithms utilized for energy optimization management of MG: the Deep Q-network (DQN) algorithm (Alabdullah *et al*, 2022) and the state-action-reward-state-action (SARSA) algorithm (Nakabi *et al*, 2021). This comparison was conducted to evaluate how the proposed method fares relative to these other DL algorithms in MG energy management. To further verify the effectiveness of the proposed method, the energy management results of the proposed method were compared with two latest energy optimization management methods for MGs: the proximal policy optimization algorithm (PPO) (Guo *et al*, 2022) and the ant-colony based algorithm (ACO) (Suresh *et al*, 2023).

The proposed energy optimization management method for MGs is simulated, tested, and analyzed according to the defined energy management objective. The data used to verify

the effectiveness of the proposed method is collecting from a MG power by a subsidiary of the Southern Power Grid Company. The data collected from June 1 2023 to July 30 2023. The simulation test platform used is a computer equipped with an Intel Core i5@2.30GHz processor, 8GB RAM, and 1TB hard disk space. The analysis of the designed energy optimization management model is performed using MATLAB software, with Gurobi selected as the solver. To obtain average values for the test results, each simulation test is repeated 30 times and each result represents the average of these repeated tests.

### 3. Simulation results and analysis

#### 3.1 Energy optimization effect comparison with two strategies and three DL algorithms

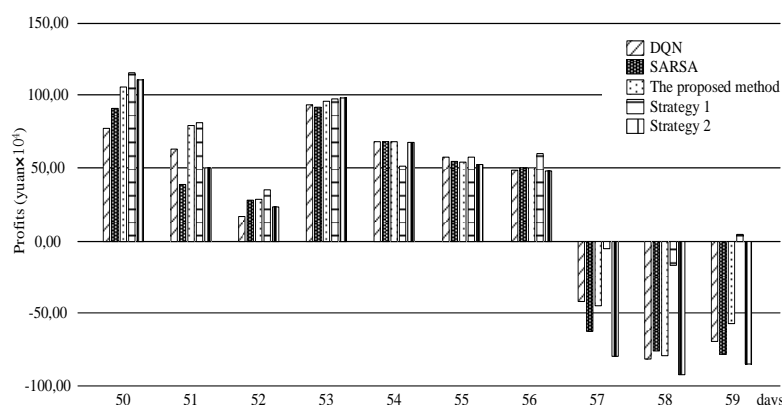
##### 3.1.1 Total profit analysis

The proposed method, DQN, SARSA, Strategy 1, and Strategy 2 were tested on the MG depicted in Fig. 1, and the daily average profit data for 10 days, from day 50 to day 59, were recorded. The comparative analysis of estimated daily earnings from employing various energy management methods over a span of ten consecutive days is depicted in Fig. 2. The total profit histograms and statistical values for 10 days for the proposed method, DQN, SARSA, Strategy 1, and Strategy 2 are shown in Fig. 3, respectively.

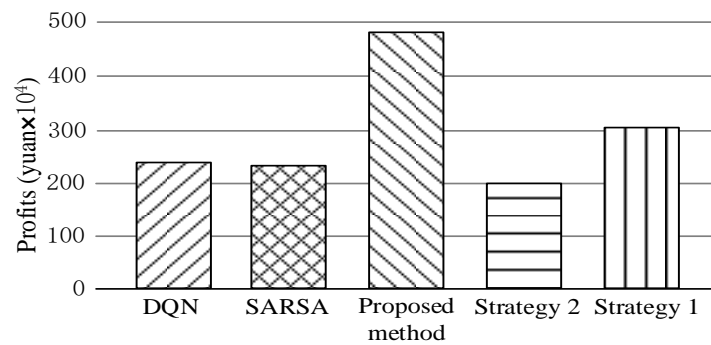
As presented in Fig. 3, a visual representation highlights the noteworthy success of the proposed method in terms of average profitability when compared to prominent alternatives such as DQN, SARSA, Strategy 1, and even the retailer strategy (Strategy 2). The compelling evidence from this figure underscores the superior financial outcomes achieved through the application of the proposed method.

Notably, the proposed method consistently outshines its counterparts, namely DQN, SARSA, Strategy 1, and Strategy 2, showcasing its sustained advantage in profitability. This temporal perspective provides a robust understanding of the method's performance over an extended period. While Strategy 1 exhibits commendable results in the domain of MG energy management, its efficacy is contingent upon a comprehensive comprehension of both current and subsequent states within the network dynamics. This nuanced requirement sets it apart as a strategy demanding a deeper understanding of the intricacies associated with the MG, underlining the unique strengths of the proposed method in offering superior profitability without the need for an exhaustive understanding of the network dynamics.

As shown in Fig. 3, only the total profit and the daily benefits of strategy 1 is close to the proposed method. In order to further analyze the advantages of the proposed method, in the following analysis, the following MG energy optimization



**Fig. 2** Daily benefits from running different energy management methods or strategies.



**Fig. 3** Total profit histograms for 10 days after running the proposed method, DQN, SARSA, Strategy 1, and Strategy 2.

management effects of the proposed method is only compared with that of Strategy 1.

### 3.1.2 TCL energy distribution and ESS charging status

In Fig. 4(a) and Fig. 4(b), the dynamic interplay between TCL energy distribution and the status of ESSs during charging and discharging cycles is meticulously illustrated for both the proposed method and Strategy 1. These visual representations provide valuable insights into the operational intricacies of the MG under the influence of these two distinct energy management approaches. Examining Fig. 4, a striking resemblance in behaviors becomes apparent on the 50th day of the MG energy management simulation. This noteworthy observation underscores the similarity in outcomes between the proposed method and Strategy 1. The synchronized patterns in TCL energy distribution and ESS charging/discharging status hint at the comparable effectiveness of both strategies in optimizing MG operations on this specific day. This convergence of results highlights a key aspect of the proposed method – its

capacity to achieve performance akin to the more complex Strategy 1. This similarity in behavior signifies the robustness of the proposed method in delivering outcomes on par with a strategy that, as mentioned earlier, necessitates a comprehensive understanding of the network dynamics. The visual evidence in Fig. 4 serves as a compelling testament to the efficacy and reliability of the proposed method in MG energy management.

The nuanced disparity between the utilization of the proposed method and Strategy 1 becomes apparent in their respective energy allocation strategies. Notably, the proposed method strategically directs a greater accumulation of energy within the TCLs, emphasizing its commitment to leveraging this resource for optimized MG performance. In contrast, Strategy 1 adopts a divergent approach, allocating a larger share of the available energy to the ESSs, indicating a preference for storing energy within these systems.

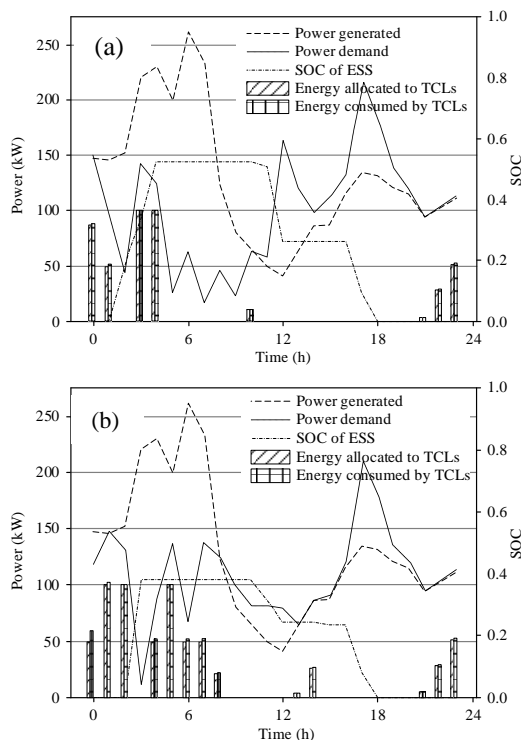
Furthermore, a shared characteristic emerges between the proposed method and Strategy 1. Both strategies exhibit a proactive stance in deciding to store a substantial amount of electrical energy during periods characterized by abundant renewable energy generation and a lack of peak demand. This commonality underscores the importance of aligning energy storage decisions with favorable conditions, showcasing a strategic awareness of optimizing resource utilization.

In essence, the proposed method's emphasis on TCL energy accumulation and Strategy 1's inclination towards ESS energy allocation highlight the distinct yet equally effective paths these strategies take in responding to the dynamic energy landscape. The shared commitment to capitalize on renewable energy surpluses accentuates the sophistication embedded in their decision-making processes, contributing to the overall efficacy of MG energy management.

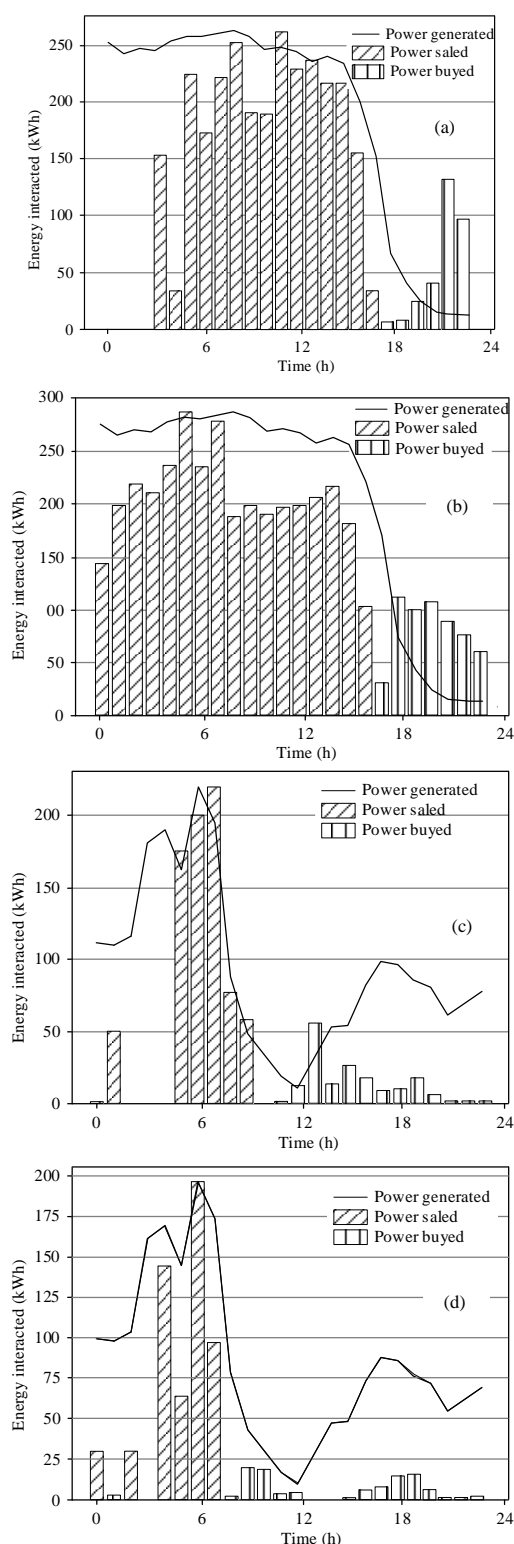
### 3.1.3 The complexities of electricity prices and demand dynamics

Fig. 5 provides a visual narrative of the electricity transactions with the DN under both Strategy 1 and the proposed method. The power curve of renewable energy generation, showcasing the potential for both storage and sale to the DN, is also depicted in this comprehensive illustration.

A discernible trend emerges from the analysis of Fig. 5: the proposed method exhibits superior performance in energy scheduling compared to Strategy 1. This is evident even on days marked by substantial fluctuations in electricity generation and consumption. The visual representation underscores the capability of the proposed method to adapt and optimize MG energy management behaviors, outshining Strategy 1 in its capacity to navigate the complexities of electricity prices and demand dynamics.



**Fig. 4** TCL energy distribution and ESS charging status. (a) The proposed method; (b) Strategy 1.



**Fig. 5** Electrical energy interacted with the main network. (a) Electrical energy interacting with the distribution network on day 50 when the proposed method is adopted; (b) Electrical energy interacting with the distribution network on day 50 when strategy 1 is adopted; (c) Electrical energy interacting with the distribution network on day 56 when the proposed method is adopted; (d) Electrical energy interacting with the distribution network on day 56 when strategy 1 is adopted.

The significance of this superiority lies in the proposed method's ability to learn and anticipate fluctuations in electricity

prices and demand. By doing so, it positions itself as a more adaptive and forward-thinking solution for MG energy optimization. This adaptability is crucial in real-world scenarios where the energy landscape is characterized by variability, allowing the proposed method to make more informed and strategic decisions in response to changing conditions. Fig. 5 serves as compelling visual evidence of the proposed method's prowess in achieving optimal energy scheduling outcomes, even in the face of challenging fluctuations in the MG's energy dynamics.

Indeed, In the intricate landscape of MG energy optimization management, Strategy 1 distinguishes itself by relying on a comprehensive understanding of the MG dynamics. This thorough comprehension empowers Strategy 1 to delve into the intricate details, predicting not only the purchase and sale prices of electricity but also the availability of these critical resources. This requirement for an extensive solution underscores the complexity and depth of Strategy 1's decision-making process, as it navigates the nuances of market dynamics and resource availability.

In stark contrast, the proposed method takes a more streamlined approach by operating without these constraints. Unlike Strategy 1, the proposed method bypasses the need for an exhaustive solution to MG dynamics. This divergence in approach positions the proposed method as a more agile and adaptable solution, capable of optimizing energy management without the burden of predicting market prices or resource availability with the same level of granularity.

This distinction not only highlights the operational variance between the two strategies but also underscores a key strength of the proposed method. Its ability to operate efficiently without the need for detailed predictions of market dynamics allows for a more flexible and practical application in real-world scenarios, where uncertainties and fluctuations are inherent. This characteristic sets the proposed method apart as a pragmatic and effective tool for MG energy optimization.

### 3.2 Energy optimization effect comparison with two latest research works

The proposed method, PPO, and ACO were tested on the MG depicted in Fig. 1, and the daily average profit data for 10 days, from day 50 to day 59, were recorded. The total profit histograms and statistical values for 10 days for the proposed method, PPO, and ACO are presented in Tab. 2. The PPO utilizes historical data on energy consumption and renewable energy generation to continuously update the network and learn an optimal policy. To ensure efficient and stable action selection, a clipped surrogate loss function is employed in PPO. The ACO draws inspiration from ant foraging behavior to effectively explore the solution space and identify optimal configurations for energy distribution within a microgrid system. Through iterative exploration, solution construction, and pheromone-based communication, ACO facilitates the discovery of high-quality solutions that maximize energy efficiency while meeting predefined optimization objectives.

The analysis presented in Table 2 unveils a significant breakthrough achieved by the proposed method, particularly evident in its substantial improvement in average profitability when contrasted with both PPO and ACO methodologies. This observation not only highlights the effectiveness of the proposed approach but also accentuates its capacity to deliver superior financial results in comparison to established techniques. By quantifying and juxtaposing the profitability metrics in a structured format, Table 2 offers a clear and objective assessment of the proposed method's performance, serving as a cornerstone for deeper analysis and interpretation.



**Table 2**

Total profit statistical values for 10 days after running the proposed method, PPO, and ACO.

Methods or strategies	Total profit
PPO	415.6
ACO	400.9
Proposed method	480.1

Moreover, the insights gleaned from Table 2 provide compelling evidence that further reinforces the superior financial outcomes achieved through the application of the proposed method. The meticulous examination of the data presented in the table sheds light on the magnitude of the improvement in profitability, underscoring the tangible benefits brought about by adopting the proposed approach. This evidence not only bolsters the credibility of the findings but also serves to deepen our understanding of the underlying mechanisms driving the observed results.

The combination of the findings from Table 2 and the subsequent analysis offers a comprehensive perspective on the remarkable success of the proposed method in enhancing profitability. By providing empirical evidence of its superiority over established methodologies such as PPO and ACO, the analysis reaffirms the efficacy of the proposed approach in achieving superior financial outcomes. This in-depth exploration not only enriches our understanding of the observed trends but also underscores the transformative potential of the proposed method in optimizing financial performance within the specified context. In summary, the proposed method consistently outshines its counterparts, namely DQN, SARSA, Strategy 1, Strategy 2, PPO, and ACO, as shown in Fig. 3 and Table 2.

## 5. Conclusions

A MG energy optimization management method is proposed based on an improved soft actor-critic (SAC) algorithm. To validate the effectiveness of the proposed approach, a comparison was conducted between the MG energy optimization management results achieved by the proposed method and those obtained through other energy optimization management methods or strategies. The comparative results reveal that the average profitability of the MG when implementing the proposed method surpasses those of the Deep Q-network DQN algorithm, the State-action-reward-state-action algorithm, the retailer strategy, the genetic algorithm, the proximal policy optimization algorithm, and the ant-colony based algorithm. Additionally, when employing the proposed method, the MG exhibits superior average profitability over a span of 10 days in comparison to other methods strategies. Although MG energy management performs well when Strategy 1 is employed, it necessitates a thorough understanding of the current and subsequent network dynamics. Furthermore, in the MG energy management simulation on the 50th day, similar behaviors are observed between the results of MG energy optimization management when the proposed method and Strategy 1 are employed. However, Strategy 1 necessitates a comprehensive solution of the MG's dynamics, enabling it to predict the purchase and sale prices of electricity, as well as the availability of electrical energy, while the proposed method operates without such constraints.

**Author Contributions:** Conceptualization, Zhiwen Yu, Wenjie Zheng and Mengdi Zeng; Data curation, Zhiwen Yu and Kaiwen Zeng; Formal analysis, Zhiwen Yu, Kaiwen Zeng, Ruifeng Zhao, Yanxu Zhang and Mengdi Zeng; Funding acquisition, Zhiwen Yu; Investigation, Zhiwen

Yu, Wenjie Zheng, Kaiwen Zeng and Ruifeng Zhao; Methodology, Zhiwen Yu, Kaiwen Zeng and Yanxu Zhang; Resources, Zhiwen Yu, Wenjie Zheng and Ruifeng Zhao; Software, Zhiwen Yu, Wenjie Zheng and Yanxu Zhang; Validation, Zhiwen Yu, Ruifeng Zhao, Yanxu Zhang and Mengdi Zeng; Visualization, Zhiwen Yu, Wenjie Zheng, Kaiwen Zeng and Yanxu Zhang; Writing – original draft, Zhiwen Yu, Wenjie Zheng and Mengdi Zeng; Writing – review & editing, Zhiwen Yu.

**Funding:** This research was funded by the China Southern Power Grid technology project (036000KK52210054(GDKJXM20210063)).

**Conflicts of Interest:** The authors declare no conflict of interest

## References

- Ahmad, S., Shafiullah, M., Ahmed, C. B., & Alowafeer, M. (2023). A Review of Microgrid Energy Management and Control Strategies. *IEEE Access*, 11, 21729-21757. <https://doi.org/10.1109/ACCESS.2023.3248511>
- Alabdullah, M. H., & Abido, M. A. (2022). Microgrid energy management using deep Q-network reinforcement learning. *Alexandria Engineering Journal*, 61(11), 9069-9078. <https://doi.org/10.1016/j.aej.2022.02.042>
- Alamir, N., Kamel, S., Hassan, M. H., & Abdelkader, S. M. (2023). An improved weighted mean of vectors algorithm for microgrid energy management considering demand response. *Neural Computing and Applications*, 35(28), 20749-20770. <https://link.springer.com/article/10.1007/s00521-023-08813-5>
- Alavizadeh, H., Alavizadeh, H., & Jang-Jaccard, J. (2022). Deep Q-learning based reinforcement learning approach for network intrusion detection. *Computers*, 11(3), 41. <https://doi.org/10.3390/computers11030041>
- Alzahrani, A., Sajjad, K., Hafeez, G., Murawwat, S., Khan, S., & Khan, F. A. (2023). Real-time energy optimization and scheduling of buildings integrated with renewable microgrid. *Applied Energy*, 335, 120640. <https://doi.org/10.1016/j.apenergy.2023.120640>
- Bao, G., & Xu, R. (2023). A Data-Driven Energy Management Strategy Based on Deep Reinforcement Learning for Microgrid Systems. *Cognitive Computation*, 15(2), 739-750. <https://doi.org/10.1007/s12559-022-10106-3>
- Barnawi, A. B. (2023). Development and Analysis of Optimization Algorithm for Demand-Side Management Considering Optimal Generation Scheduling and Power Flow in Grid-Connected AC/DC Microgrid. *Sustainability*, 15(21), 15671. <https://doi.org/10.3390/su152115671>
- Cavus, M., Allahham, A., Adhikari, K., Zangiabadi, M., & Giaouris, D. (2023). Energy Management of Grid-Connected Microgrids Using an Optimal Systems Approach. *IEEE Access*, 11, 9907-9919. <https://doi.org/10.1109/ACCESS.2023.3239135>
- Datta, J., & Das, D. (2023). Energy management of multi-microgrids with renewables and electric vehicles considering price-elasticity based demand response: A bi-level hybrid optimization approach. *Sustainable Cities and Society*, 99, 104908. <https://doi.org/10.1016/j.scs.2023.104908>
- Dey, B., Misra, S., & Marquez, F. P. G. (2023). Microgrid system energy management with demand response program for clean and economical operation. *Applied Energy*, 334, 120717. <https://doi.org/10.1016/j.seta.2023.103377>
- Dong, Y., Zhang, H., Wang, C., & Zhou, X. (2024). Soft actor-critic DRL algorithm for interval optimal dispatch of integrated energy systems with uncertainty in demand response and renewable energy. *Engineering Applications of Artificial Intelligence*, 127, 107230. <https://doi.org/10.1016/j.engappai.2023.107230>
- Du, A., & Ghavidel, A. (2022). Parameterized deep reinforcement learning-enabled maintenance decision-support and life-cycle risk assessment for highway bridge portfolios. *Structural Safety*, 97, 102221. <https://doi.org/10.1016/j.strusafe.2022.102221>
- El Bourakadi, D., Yahyaoui, A., & Boumhidi, J. (2022). Intelligent energy management for micro-grid based on deep learning LSTM

- prediction model and fuzzy decision-making. *Sustainable Computing: Informatics and Systems*, 35, 100709. <https://doi.org/10.1016/j.suscom.2022.100709>
- Gao, X. Y., Chen, B., & Huang, S. B. (2023). Energy management strategy for pelagic island microgrid considering leader-follower game. *Proceedings of the CSU-EPSA*, 35(02): 1-9. <https://doi.org/10.19635/j.cnki.csu-epsa.001036>
- Guo, C., Wang, X., Zheng, Y., & Zhang, F. (2022). Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning. *Energy*, 238, 121873. <https://doi.org/10.1016/j.energy.2021.121873>
- Han, X., Mu, C., Yan, J., & Niu, Z. (2023). An autonomous control technology based on deep reinforcement learning for optimal active power dispatch. *International Journal of Electrical Power & Energy Systems*, 145, 108686. <https://doi.org/10.1016/j.ijepes.2022.108686>
- Hou, Y. Y., Zeng, J., Luo, Y., & Liu, J. F. (2023). Research on collaborative and optimization methods of active energy management in community microgrid. *Power System Technology*, 47(04), 1548-1557. <https://doi.org/10.13335/j.1000-3673.pst.2022.1361>
- Huang, Y., Li, G., Chen, C., Bian, Y., Qian, T., & Bie, Z. (2022). Resilient distribution networks by microgrid formation using deep reinforcement learning. *IEEE Transactions on Smart Grid*, 13(6), 4918-4930. <https://doi.org/10.1109/TSG.2022.3179593>
- Hu, C., Cai, Z., Zhang, Y., Yan, R., Cai, Y., & Cen, B. (2022). A soft actor-critic deep reinforcement learning method for multi-timescale coordinated operation of microgrids. *Protection and Control of Modern Power Systems*, 7(1), 1-10. <https://doi.org/10.1186/s41601-022-00252-z>
- Huang, J., Bi, Y., Qin, L., Zang, G., & Li, H. (2021). Design parameters and dynamic performance analysis of a high efficient solar-ground source cooling system using parabolic trough collector. *International Journal of Sustainable Energy*, 40(3), 253-282. <https://doi.org/10.1080/14786451.2020.1806841>
- Ibrahim, M. M., Hasanien, H. M., Farag, H. E., & Orman, W. A. (2023). Energy management of multi-area islanded hybrid microgrids: a stochastic approach. *IEEE Access*, 11, 101409 - 101424. <https://doi.org/10.1109/ACCESS.2023.3313259>
- Jahani, A., Zare, K., & Khanli, L. M. (2023). Short-term load forecasting for microgrid energy management system using hybrid SPM-LSTM. *Sustainable Cities and Society*, 98, 104775. <https://doi.org/10.1016/j.scs.2023.104775>
- Kheiter, A., Souag, S., Chaouch, A., Boukourt, A., Bekkouche, B., & Guezgouz, M. (2022). Energy Management Strategy Based on Marine Predators Algorithm for Grid-Connected Microgrid. *International Journal of Renewable Energy Development*, 11(3), 751-765. <https://doi.org/10.14710/ijred.2022.42797>
- Kim, B., & Kim, J. (2023). Efficient Difficulty Level Balancing in Match-3 Puzzle Games: A Comparative Study of Proximal Policy Optimization and Soft Actor-Critic Algorithms. *Electronics*, 12(21), 4456. <https://doi.org/10.3390/electronics12214456>
- Kosuru, V. S. R., & Venkitaraman, A. K. (2022). Developing a deep Q-learning and neural network framework for trajectory planning. *European Journal of Engineering and Technology Research*, 7(6), 148-157. <https://doi.org/10.24018/ejeng.2022.7.6.2944>
- Lagouir, M., Badri, A., & Sayouti, Y. (2021). Solving Multi-Objective Energy Management of a DC Microgrid using Multi-Objective Multiverse Optimization. *International Journal of Renewable Energy Development*, 10(4), 911-922. <https://doi.org/10.14710/ijred.2021.38909>
- Leonori, S., Paschero, M., Mascioli, F. M. F., & Rizzi, A. (2020). Optimization strategies for Microgrid energy management systems by Genetic Algorithms. *Applied Soft Computing*, 86, 105903. <https://doi.org/10.1016/j.asoc.2019.105903>
- Moos, J., Hansel, K., Abdulsamad, H., Stark, S., Clever, D., & Peters, J. (2022). Robust reinforcement learning: A review of foundations and recent advances. *Machine Learning and Knowledge Extraction*, 4(1), 276-315. <https://doi.org/10.3390/make4010013>
- Mostefa, A., Belalia, K., Lantri, T., Boulouiha, H. M., & Allali, A. (2023). A four-line active shunt filter to enhance the power quality in a microgrid. *International Journal of Renewable Energy Development*, 12(3), 488-498. <https://doi.org/10.14710/ijred.2023.50270>
- Park, S., Pozzi, A., Whitmeyer, M., Perez, H., Kandel, A., Kim, G., & Moura, S. (2022). A deep reinforcement learning framework for fast charging of li-ion batteries. *IEEE Transactions on Transportation Electrification*, 8(2), 2770-2784. <https://doi.org/10.1109/TTE.2022.3140316>
- Nakabi, T. A., & Toivanen, P. (2019). An ANN-based model for learning individual customer behavior in response to electricity prices. *Sustainable Energy, Grids and Networks*, 18, 100212. <https://doi.org/10.1016/j.segan.2019.100212>
- Nakabi, T. A., & Toivanen, P. (2021). Deep reinforcement learning for energy management in a microgrid with flexible demand. *Sustainable Energy, Grids and Networks*, 25, 100413. <https://doi.org/10.1016/j.segan.2020.100413>
- NREL, Wind data and tools, <https://www.nrel.gov/wind/data-tools.html>
- Rodriguez, M., Arcos-Aviles, D., & Martinez, W. (2023). Fuzzy logic-based energy management for isolated microgrid using meta-heuristic optimization algorithms. *Applied Energy*, 335, 120771. <https://doi.org/10.1016/j.apenergy.2023.120771>
- Saeed, M. H., Rana, S., El-Bayeh, C. Z., & Fangzong, W. (2023). Demand response based microgrid's economic dispatch. *International Journal of Renewable Energy Development*, 12(4): 749-759. <https://doi.org/10.14710/ijred.2023.49165>
- Shavandi, A., & Khedmati, M. (2022). A multi-agent deep reinforcement learning framework for algorithmic trading in financial markets. *Expert Systems with Applications*, 208, 118124. <https://doi.org/10.1016/j.eswa.2022.118124>
- Shezan, S. A., Kamwa, I., Ishraque, M. F., Mueen, S. M., Hasan, K. N., Saidur, R., ... & Al-Sulaiman, F. A. (2023). Evaluation of different optimization techniques and control strategies of hybrid microgrid: a review. *Energies*, 16(4), 1792. <https://doi.org/10.3390/en16041792>
- Singh, V., Nanavati, B., Kar, A. K., & Gupta, A. (2023). How to maximize clicks for display advertisement in digital marketing? A reinforcement learning approach. *Information Systems Frontiers*, 25(4), 1621-1638. <https://doi.org/10.1007/s10796-022-10314-0>
- Suresh, V., Janik, P., Jasinski, M., Guerrero, J. M., & Leonowicz, Z. (2023). Microgrid energy management using metaheuristic optimization algorithms. *Applied Soft Computing*, 134, 109981. <https://doi.org/10.1016/j.asoc.2022.109981>
- Sun, W., Zou, Y., Zhang, X., Guo, N., Zhang, B., & Du, G. (2022). High robustness energy management strategy of hybrid electric vehicle based on improved soft actor-critic deep reinforcement learning. *Energy*, 281, 124806. <https://doi.org/10.1016/j.energy.2022.124806>
- Tajjour, S., & Chandel, S. S. (2023). A comprehensive review on sustainable energy management systems for optimal operation of future-generation of solar microgrids. *Sustainable Energy Technologies and Assessments*, 58, 103377. <https://doi.org/10.1016/j.seta.2023.103377>
- Topa Gavilema, Á. O., Álvarez, J. D., Torres Moreno, J. L., & García, M. P. (2021). Towards optimal management in microgrids: An overview. *Energies*, 14(16), 5202. <https://doi.org/10.3390/en14165202>
- Tian, Y., Li, X., Ma, H., Zhang, X., Tan, K. C., & Jin, Y. (2022). Deep reinforcement learning based adaptive operator selection for evolutionary multi-objective optimization. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(4): 1051 - 1064. <https://doi.org/10.1109/TETCI.2022.3146882>
- Tightiz, L., & Yang, H. (2021). Resilience microgrid as power system integrity protection scheme element with reinforcement learning based management. *IEEE Access*, 9, 83963-83975. <https://doi.org/10.1109/ACCESS.2021.3087491>
- Topa Gavilema, Á. O., Álvarez, J. D., Torres Moreno, J. L., & García, M. P. (2021). Towards optimal management in microgrids: An overview. *Energies*, 14(16), 5202. <https://doi.org/10.3390/en14165202>
- Zhang, B., Hu, W., Ghias, A. M., Xu, X., & Chen, Z. (2023). Multi-agent deep reinforcement learning based distributed control architecture for interconnected multi-energy microgrid energy management and optimization. *Energy Conversion and*

- Management, 277, 116647. <https://doi.org/10.1016/j.enconman.2022.116647>
- Xiong, K., Cao, D., Zhang, G., Chen, Z., & Hu, W. (2023). Coordinated volt/VAR control for photovoltaic inverters: A soft actor-critic enhanced droop control approach. *International Journal of Electrical Power & Energy Systems*, 149, 109019. <https://doi.org/10.1016/j.ijepes.2023.109019>
- Xu, H., Wu, Q., Wen, J., & Yang, Z. (2022). Joint bidding and pricing for electricity retailers based on multi-task deep reinforcement learning. *International journal of electrical power & energy systems*, 138, 107897. <https://doi.org/10.1016/j.ijepes.2021.107897>
- Zhang, Z., Dou, C., Yue, D., & Zhang, B. (2021). Predictive voltage hierarchical controller design for islanded microgrids under limited communication. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69(2), 933-945. <https://doi.org/10.1109/TCSI.2021.3117048>
- Zheng, Y., Tao, J., Sun, Q., Sun, H., Chen, Z., & Sun, M. (2023). Deep reinforcement learning based active disturbance rejection load frequency control of multi-area interconnected power systems with renewable energy. *Journal of the Franklin Institute*, 360(17), 13908-13931. <https://doi.org/10.1360/N092017-00182>
- Zhu, Q., Su, S., Tang, T., Liu, W., Zhang, Z., & Tian, Q. (2022). An eco-driving algorithm for trains through distributing energy: A Q-Learning approach. *ISA transactions*, 122, 24-37. <https://doi.org/10.1016/j.isatra.2021.04.036>

1  
2  
3  
4  
5



© 2024. The Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA) International License (<http://creativecommons.org/licenses/by-sa/4.0/>)