Artificial Intelligence-Optimized Hybrid Hydrogen–Battery Energy Storage for Renewable Microgrids

UDC:546.11:621.355
Original scientific paper https://doi.org/10.46793/adeletters.2026.5.2.1

Johnson Abiola[1*], Humbulani Simon Phuluwa[2], David Aborisade[3], Muniru Okelola[3], Godwin Igbinigie[4], Ilesanmi Daniyan[1,2,5]

[1]Department of Mechatronics Engineering, College of Engineering, Bells University of Technology, P.M.B. 1015, Ota, Nigeria
[2]Department of Industrial Engineering & Engineering Management, University of South Africa, Florida, 1709, South Africa
[3]Department of Electronic and Electrical Engineering, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Nigeria
[4]Department of Biomedical Engineering, College of Engineering, Bells University of Technology, P.M.B. 1015, Ota, Nigeria
[5]Centre for Artificial Intelligence, Bells University of Technology, P.M.B. 1015, Ota, Nigeria

Abstract:
The transition toward sustainable energy on a global scale is hindered by the intermittent nature of renewables and the high costs of achieving energy independence for off-grid commercial systems. This study develops a deep reinforcement learning (DRL) solution for the optimal management of a hybrid hydrogen-battery energy storage system in a microgrid based on renewables. Taking a study-based approach, drawing on a case study from a commercial hospitality center, the control challenge was modelled as a Markov Decision Process. For the implementation, the Soft Actor-Critic (SAC) learning algorithm was used to regulate power transitions among photovoltaic generation units, batteries, and a hydrogen chain comprising electrolyzers, storage, and fuel cells. The SAC learning-based controller was constructed for minimizing the cost of operations through a reward function that takes into account the degradation costs of batteries as well as the transition costs for seamless operational changes. The SAC-learning-based controller was validated on a high-resolution dataset for a year and was compared with a rule-based controller. It was found that the SAC controller is capable of reducing costs by 2.0%, producing a much smoother transition in power, as well as more intelligent usage of the two storage systems. These findings pave the way for exploration of the DRL for a reliable and affordable transition toward commercial microgrids.

ARTICLE HISTORY
Received: 22 December 2025
Revised: 25 February 2026
Accepted: 9 March 2026
Published: 22 June 2026

KEYWORDS
Energy management, Deep Reinforcement learning, Hybrid energy storage, Hydrogen, Battery, Microgrid, Soft actor-critic algorithm

1. INTRODUCTION

The transition to sustainable energy systems worldwide depends on the widespread adoption of intermittent sources, such as solar photovoltaic (PV) systems [1]. This is most important for the commercial and industrial sectors across the globe, in which the energy cost makes a significant contribution and the issue of power failure is a primary concern. A microgrid system has been identified as a crucial enabler for the integration process, as it manages local energy production, storage, and consumption to improve power reliability and reduce power failures [2]. However, the intermittent nature of the renewables makes

*CONTACT: Johnson Abiola, e-mail: joabiola@bellsuniversity.edu.ng

the task very challenging for the management of the balance between demand and supply [3].

Such a challenge is even more specific to business organizations found in regions with high electricity tariffs and unreliable supply. In Nigeria, for instance, businesses such as hotels and event organizers face operating costs, with the main challenge being high spending on both diesel and electricity. Thus, energy costs are among the most substantial and uncertain operating costs for most organizations. Although a solar PV system is helpful, it is not sufficient to meet the increasing energy needs. Many organizations in emerging economies are still significantly reliant on electricity supply, especially at night, during major events, and when there is little sunlight.

This particular instance highlights the absence of a key requirement: having solar power alone is insufficient. There is a strong requirement to shift the current approach from complementing the energy grid to proactively replacing it with high-capacity energy storage systems. Lithium-ion batteries and the concept of hydrogen storage have mutually complementary properties, which are very appropriately combined for the mentioned aim [4]. Lithium-ion batteries have high power density and the capability for quick reaction, which are suitably applicable for short-term power oscillations [5]. On the other hand, the energy storage system based on hydrogen can include an electrolyzer, a storage tank, as well as a fuel cell, because it has high energy density properties along with the capability for long-term energy storage, which makes it ideally suitable for storing excess energy throughout the night or within the seasons [6]. A H-B ESS, as mentioned, utilizes the collective benefits of the two resources. It also adds another layer of increased complexity—the proper coordination of energy handling across different resources by taking into account the stochastic behaviour of renewable energy sources, as well as the energy demand together [7].

Conventional energy management systems (EMS) struggle to handle such complexities. Rule-based techniques, being simple and less computationally expensive, have flaws in their non-adaptivity and can result in suboptimal performance in novel situations [8]. Optimization techniques, including Model Predictive Control (MPC), have the potential of attaining superior results but involve complex forecasting tasks, which make real-time processing rather difficult [9].

Recently, DRL has been cast as an efficient, model-free approach for tackling challenging control problems [10]. By effectively learning an optimal policy through interaction with the environment, DRL can address uncertainties without requiring explicit system models. The SAC algorithm was recently introduced for continuous control problems due to its improved stability and sample efficiency, which have been credited to an entropy regularization term encouraging exploration [11-13]. Although SAC has demonstrated effectiveness for vehicular energy management applications [14], its use regarding the design of full dynamic models with a focus on control theory for microgrids consisting of grid-scale hybrid hydrogen battery energy storage systems, however, is an area that awaits investigation. This is because existing works either neglect complexities associated with premise dynamics [15] or primarily focus on economic aspects without integrating stability/smoothness directly during the learning phase [16]. This offers an interesting development opportunity regarding the integration of an SAC-based EMS that considers the dynamic constraints, degradation cost, as well as operational requirements introduced by a full-scale H-B ESS at the commercial center.

The aim of this study is to develop a DRL solution for the optimal management of a hybrid hydrogen-battery energy storage system in a microgrid based on renewables. This study is significant in that it contributes to the development of intelligent and sustainable energy systems. The system is adaptive, with real-time decision-making geared towards balancing the volatility of renewable energy from sources such as solar and wind with effective storage capacity. This study is in line with Sustainable Development Goal (SDG) 7, which emphasizes access to modern, cost-effective, and sustainable energy for all by 2030 by increasing renewables. The DRL can adaptively learn and interact with the environment to understand and balance energy needs, demand, and supply against storage capacity. This will enhance cost optimization, grid stability, and efficiency by preventing power fluctuation from the renewables. Furthermore, this study supports the quest for net-zero carbon by maximizing the generation of renewable energy use and minimizing fossil fuel reliance to promote sustainability in microgrids.

Hybrid Energy Storage System (HESS) incorporation into a renewable energy microgrid has gained significant interest in the literature as a reliable means of mitigating solar and wind intermittency. For a more reliable commercial setup, choosing an appropriate Energy Management

Strategy (EMS) is crucial for economic feasibility. An assessment of the development of Energy Management Strategies for H-B ESS aims to evaluate past approaches, including current AI technologies, and identify a research gap this study aims to resolve.

The Traditional Energy Management (TEM) strategies for H-B ESS can be classified into rule-based systems and optimization-based systems.

The Rule-Based Strategy uses fixed thresholds and logic rules. In one form, this scheme gives high priority to the battery for solving power imbalances over a short time period, considering its fast response time and high efficiency, activating the hydrogen chain (electrolyzer or fuel cell) only when the State of Charge (SOC) falls in predetermined high or low limits [8, 17]. Schemes like this have advantages: they are simple, deterministic, and require less computation. However, these methods are not adaptive, are highly dependent on parameter tuning, and generally do not produce global optimal solutions in stochastic or time-varying environments [18]. This will lead to high operational and degradation costs because this method will not adapt to patterns over time.

The Optimization-Based Methods, such as MPC, are a major step forward in this respect. In the MPC approach, energy management is framed as a finite-horizon optimization problem, in which renewable power and load forecasts, along with setpoints, are used to anticipate and optimize, subject to system constraints [9, 19]. The economic advantage provided by the MPC approach over rule-based approaches has been highlighted in research such as Xu [20]. The drawbacks to the MPC approach are its reliance on forecasts, where inaccuracy can result in suboptimal or unfeasible operations for micro-grids subjected to significant volatilities, and the computational challenges to optimize the problem in real time, especially using nonlinear system models [21].

DRL approach (DRL) has also proven to be a powerful, model-free approach to complex sequential decision-making problems that provide an attractive substitute for classical optimization. In contrast with MPC, which demands explicit system knowledge or forecasting, an agent in RL learns an optimal policy by interacting with the environment, thus making it inherently robust against uncertainties and noise [22]. In the initial applications of RL technology in energy control, there were several demonstrations involving value functions, also known as Deep Q-Networks (DQN) [11]. The discretization of the actions, however, led to a loss of the high level of resolution needed in continuous control problems, such as the specification of precise power setpoints for batteries, electrolyzers, or fuel cells.

The SAC algorithm [13] incorporates an entropy term into its reward function, thus improving exploration during learning and obtaining more robust results. Its effectiveness in many fields, including energy management in vehicles, was proven. For example, SAC was employed in a fuel cell hybrid vehicle in [12] and led to a significant reduction in hydrogen consumption using a newly designed reward function. This shows the high potential of SAC in managing multi-goal energy systems that often involve conflicting tasks.

The existing literature review, while still recent and growing in its applications related to stationary microgrids, has some patterns that keep re-emerging:

- It has been noticed that several researchers use reduced models, with constant electrolyzer and fuel cell efficiencies, disregarding the nonlinear behavior of these components. This makes it difficult to assess the realism of the energy management system (EMS) being analyzed.
- One common trend is the use of hierarchical schemes that decouple day-ahead scheduling from real-time control. Even though it is computationally expedient, the performance gap between optimal operation, which may diverge due to unpredictable real-time variations, may be introduced.
- While many DRL-based methods for EMS consider, among other things, minimizing economic costs, very few of them formally integrate control theory notions of stability, smoothness, and hard constraint satisfaction into their reward functions.
- As a consequence, a DRL-agent could potentially act in an economically optimal but highly unstable manner, with many switches and mechanical stresses.

As evidenced, the current literature lacks a unified control-theoretic DRL framework that simultaneously addresses:

i. Realistic system nonlinearities without oversimplification.
ii. Balancing cost, degradation, and efficiency
iii. Operational stability and smoothness directly within the learning objective.

This paper, therefore, bridges the gaps mentioned by proposing a novel SAC-based EMS designed from a control-theoretic perspective. This work moves beyond mere economic optimization to incorporate stability and smoothness directly into the reward function, utilizing a full dynamic model of the H-B ESS that captures the essential nonlinearities of all components. The proposed method is rigorously validated against both rule-based and MPC benchmarks under a real-world operational scenario.

This paper proposed a DRL-based control framework for a hybrid hydrogen–battery microgrid, validated using data from a historical dataset. This study contributes a full dynamic system model of a commercial microgrid, parameterized and validated with one year of high-resolution operational data from an active hospitality center, capturing the nonlinearities and constraints of all components.

It developed an EMS using the Soft Actor-Critic algorithm. The reward function is meticulously designed not only to minimize operational cost but also to enforce battery SOC and hydrogen tank level constraints, penalize component degradation, and encourage smooth control actions for enhanced system lifetime.

It also demonstrates that the proposed SAC controller outperforms conventional rule-based strategies across key metrics, including total operational cost, renewable utilization, and battery stress reduction, providing a clear quantitative assessment of its benefits for the business case.

2. MATERIALS AND METHODS

2.1 System Modeling

The microgrid under study is an islanded system comprising the following:

i. A PV array.
ii. A lithium-ion battery energy storage system (BESS).
iii. A hydrogen chain (electrolyzer, storage tank, and fuel cell).
iv. An electrical load.

The system architecture is shown in Fig. 1. The energy management objective is to control the power flows between these components to meet the load demand at minimum cost, subject to operational constraints. A discrete-time model with a time step Δt is used, where k denotes the index.

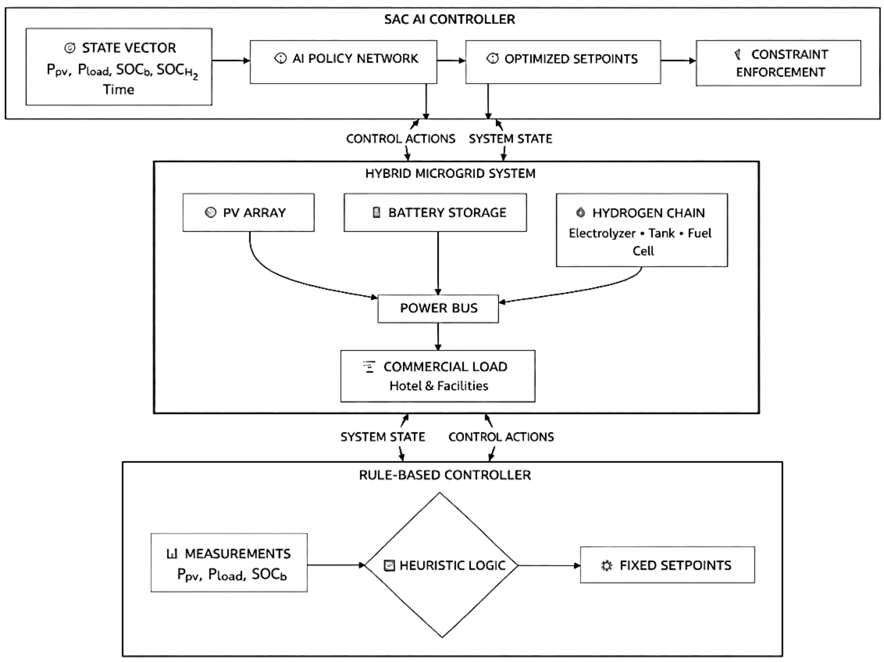

Fig. 1. Architecture of the hybrid hydrogen-battery microgrid

2.1.1 Power Balance

The fundamental constraint for the islanded microgrid is the instantaneous power balance at each time step k [23] is given by Eq. (1).

$$P_{\mathrm{pv},k} + P_{f,k} + P_{b,k}^{\mathrm{dis}} = P_{L,k} + P_{b,k}^{\mathrm{ch}} + P_{e,k} \quad (1)$$

where $P_{\mathrm{pv},k}$ is the PV power, $P_{b,k}^{\mathrm{dis}}$ and $P_{b,k}^{\mathrm{ch}}$ are the battery discharge and charge power, $P_{f,k}$ is the fuel cell output power, $P_{e,k}$ is the electrolyzer power and $P_{L,k}$ is the load demand. Note that all power values are measured in kilowatts (kW).

2.1.2 Battery Storage Model

The battery is modelled as an energy reservoir. Its SOC is denoted as $E_{b,k}$ according to Eq. (2).

$$E_{b,k+1} = E_{b,k} + \left(\eta_b^{\text{ch}} \times P_{b,k}^{\text{ch}} - \frac{1}{\eta_b^{\text{dis}}} \times P_{b,k}^{\text{dis}}\right)\Delta t \tag{2}$$

where η_b^{ch} and η_b^{dis} are the charge and discharge efficiencies, respectively. The battery operation is subjected to the following constraints:

$$0 \le P_{b,k}^{\text{ch}} \le P_b^{\max}, \quad 0 \le P_{b,k}^{\text{dis}} \le P_b^{\max}$$

$$E_b^{\min} \le E_{b,k} \le E_b^{\max}$$

$$P_{b,k}^{\text{ch}} \cdot P_{b,k}^{\text{dis}} = 0 \quad \text{(no simultaneous charge/discharge)}$$

To account for lifetime degradation, a convex cost proportional to the energy throughput is applied [23] as Eq. (3).

$$C_{b,k}^{\text{deg}} = \lambda_b (P_{b,k}^{\text{ch}} + P_{b,k}^{\text{dis}})\Delta t \tag{3}$$

where λ_b is a degradation constant.

2.1.3 Hydrogen Chain (Electrolyzer, Tank, Fuel Cell) Model

The hydrogen chain converts electrical energy to chemical energy and vice versa. The hydrogen tank level (H_k, in kWh using the lower heating value $LHV \approx 33.33\ kWh/kg$) [24] is governed by Eq. (4).

$$H_{k+1} = H_k + \left(\eta_e\, P_{e,k} - \frac{P_{f,k}}{\eta_f}\right)\Delta t \tag{4}$$

where η_e is the electrolyzer efficiency and η_f is the fuel cell efficiency.

The hydrogen storage level is constrained by its physical limits:

$$0 \le P_{e,k} \le P_e^{\max}, \quad 0 \le P_{f,k} \le P_f^{\max}$$

$$H^{\min} \le H_k^{(\text{kWh})} \le H^{\max}$$

2.1.4 Cost and Objective Function

The objectives of the energy management system are to minimize the total operational cost. The stage cost c_k At each time step, it is defined as a weighted sum of several factors given as Eq. (5). These factors are (i) conversion losses, (ii) battery degradation cost, (iii) hydrogen operation cost, and (iv) smoothing penalties to discourage aggressive switching.

$$c_k = C_{\text{loss},k} + C_{b,k}^{\text{deg}} + C_{H,k} + C_{\Delta,k} \tag{5}$$

where $C_{\text{loss},k}$ is the approximate immediate conversion loss energy in the battery and the hydrogen chain, which is given as Eq. (6), $C_{b,k}^{\text{deg}}$ is the battery degradation given as Eq. (7), $C_{H,k}$ is the hydrogen operation cost, representing the opportunity of using electricity for hydrogen production and the operating cost of the fuel cell, this is given as Eq. (8), and $C_{\Delta,k}$ is the smoothing penalty, which discourages excessive power fluctuations in the control signals; this is given in Eq. (9).

$$C_{\text{loss},k} = \lambda_{\text{loss}} \Big[P_{e,k}(1-\eta_e) + P_{f,k}(1-\eta_f) + P_{b,k}^{\text{ch}}(1-\eta_b^{\text{ch}}) + P_{b,k}^{\text{dis}}\left(1-\frac{1}{\eta_b^{\text{dis}}}\right)\Big]\Delta t \tag{6}$$

$$C_{b,k}^{\text{deg}} = \lambda_b\, (P_{b,k}^{\text{ch}} + P_{b,k}^{\text{dis}})^{\alpha}\Delta t \tag{7}$$

$$C_{H,k} = \lambda_e P_{e,k}\Delta t + \lambda_f P_{f,k}\Delta t \tag{8}$$

$$C_{\Delta,k} = \lambda_\Delta \sum_{u\in\mathcal{U}} (u_k - u_{k-1})^2 \tag{9}$$

where α is the temperature parameter controlling the tradeoff between reward and exploration, u is the control action vector, λ_{loss} is the weight for energy conversion losses, λ_e is the associated cost coefficient, λ_f is the specific cost weight, and λ_Δ is the smoothing weight.

The goal is to minimize the total cost $J = \sum_{k=0}^{T-1} c_k$ over the operational horizon.

Fig. 2 shows the solar panels for the hybrid hydrogen-battery microgrid installed in a hotel located at Oke-Anu, Ogbomoso, Ogbomoso North Local Government Area, Oyo State, Nigeria. The building is a hospitality business centre with accommodation, restaurant, event halls, and recreational park facilities.

Fig. 2. The solar panels for the hybrid hydrogen-battery microgrid

2.2 Formulating Deep Reinforcement Learning

The energy management problem is formulated as a Markov Decision Process (MDP), the state s_k and action a_k [25] is defined as follows:

(i) State $s_k = \left[E_{b,k},\ H_k,\ P_{\mathrm{pv},k},\ P_{L,k},\right]$

(ii) Action $a_k = \left[P_{b,k}^{\mathrm{ch}},\ P_{b,k}^{\mathrm{dis}},\ P_{e,k},\ P_{f,k}\right]$

(iii) Reward $r_k = -c_k$, the negative of the stage cost is to align the reward maximization with cost minimization.

2.2.1 Soft Actor-Critic Algorithm

The implemented control strategy is the SAC algorithm, a model-free, off-policy, entropy-regularized reinforcement learning approach optimized for continuous control tasks. SAC was selected for its state-of-the-art performance on continuous control tasks, characterized by stability, sample efficiency, and inherent exploration capabilities enabled by entropy regularization [13].

The SAC algorithm augments the standard reinforcement learning objective with an entropy regularization term to encourage exploration. The maximum entropy objective is given as Eq. (10):

$$J(\pi_\phi) = \sum_t \mathbb{E}_{(s_t,a_t)\sim\rho_{\pi_\phi}}\left[r_t + \alpha\mathcal{H}\left(\pi_\phi(\cdot \mid s_t)\right)\right] \quad (10)$$

where $\mathcal{H}(\pi_\phi(\cdot| s_t)) = -\log \pi_\phi(a_t \mid s_t)$ is the policy entropy, and π_ϕ depicts policy network.

Two independent critic networks $Q_{\theta_1}(s_t, a_t)$ and $Q_{\theta_2}(s_t, a_t)$ are employed to mitigate positive bias during value estimation. The critic loss function for each network i is defined as Eq. (11):

$$J_Q(\theta_i) = \mathbb{E}_{(s_t,a_t,r_t,s_{t+1})\sim\mathcal{D}}\left[\frac{1}{2}\left(Q_{\theta_i}(s_t, a_t) - y_t\right)^2\right] \quad (11)$$

where the target value y_t is computed as Eq. (12).

$$y_t = r_t + \gamma\mathbb{E}_{a_{t+1}\sim\pi_\phi}\left[\min_{i=1,2} Q_{\bar{\theta}_i}(s_{t+1}, a_{t+1}) - \alpha \log \pi_\phi(a_{t+1} \mid s_{t+1})\right] \quad (12)$$

with $\bar{\theta}_i$ denoting the parameters of the target Q-networks which are updated using exponential moving average and γ is the discount factor.

The policy network $\pi_\phi(a_t \mid s_t)$ is optimized to maximize both the expected Q-value and the policy entropy given as Eq. (13).

$$J_\pi(\phi) = \mathbb{E}_{s_t\sim\mathcal{D},a_t\sim\pi_\phi}\left[\alpha \log \pi_\phi(a_t \mid s_t) - \min_{i=1,2} Q_{\theta_i}(s_t, a_t)\right] \quad (13)$$

The gradient update for the policy parameters ϕ is performed using the reparameterization trick, where the stochastic action a_t is represented as Eq. (14):

$$a_t = f_\phi(\epsilon_t; s_t) = \mu_\phi(s_t) + \sigma_\phi(s_t) \odot \epsilon_t, \epsilon_t \sim \mathcal{N}(0, I) \quad (14)$$

where f_ϕ is the reparameterization function, ϵ_t is the noise vector, $\mathcal{N}$ standard normal distribution and I is the identity matrix.

The temperature parameter α is automatically tuned by minimizing Eq. (15).

$$J(\alpha) = \mathbb{E}_{a_t\sim\pi_\phi}\left[-\alpha\left(\log \pi_\phi(a_t \mid s_t) + \mathcal{H}_{\mathrm{target}}\right)\right] \quad (15)$$

where $\mathcal{H}_{\mathrm{target}}$ is a desired target entropy value, typically set proportional to the negative of the action dimension $-\dim(\mathcal{A})$, μ_ϕ is the mean output, σ_ϕ is the standard deviation.

2.2.2 Problem Formulation for DRL Agent

At each discrete time step t, the microgrid environment is represented by a state vector $s_t \in \mathbb{R}^n$, which captures the system's dynamic information. The agent is designed to make optimal decisions based on a complete representation of the microgrid's operational state. The state vector s_t at time t is defined as Eq. 16.

$$s_t = \left[P_{\mathrm{pv}}(t), P_{\mathrm{load}}(t), SOC_b(t), SOC_h(t), t_{hour}(t)\right]^T \quad (16)$$

where $P_{pv}(t)$ is the current PV generation power, $P_{\mathrm{load}}(t)$ is the total load demand (kW), $SOC_b(t)$ is the state of charge of the battery given as Eq. (17), $SOC_h(t)$ is the normalized hydrogen tank level, which is given as Eq. (18), and $t_{hour}(t)$ is the hour-of-the-day encoded as cyclic feature to capture diurnal patterns. All power is measured in kilowatt (kW).

$$SOC_b(t) = \frac{E_b(t)}{E_b^{max}} \quad (17)$$

$$SOC_h(t) = \frac{H}{H^{max}}(t) \quad (18)$$

The agent outputs a continuous, multi-dimensional action vector a_t that directly specifies the power setpoints for the controllable assets. The action vector a_t is given as Eq. (19):

$$a_t = [P_b^{\mathrm{ch/dis}}(t), P_e(t), P_{fc}(t)]^T \quad (19)$$

where $P_b^{\mathrm{ch/dis}}(t)$ is the net battery charge/discharge power (positive = discharge, negative = charge), $P_e(t)$ is the electrolyzer input power for hydrogen production, $P_{fc}(t)$ is the fuel cell output power command during hydrogen conversion. To ensure the feasibility of the generated actions, a custom action projection layer is implemented after the network output. This layer:

i. Scales the normalized network outputs to the physical minimum/maximum limits of each component.
ii. Enforces the no simultaneous charge/discharge constraint for the battery.
iii. Clips the final values to ensure they remain within safe operating boundaries.

2.2.3 Reward Function Design

The reward function is the backbone of the control strategy, meticulously designed to capture the problem's multi-objective nature. It directly mirrors the stage cost c_k defined in Section 2.3, but is formulated as a reward to be maximized. At each time step, the reward function encodes both economic and operational objectives, as given in Eq. (20).

$$r_t = -\left(\lambda_{loss} C_{loss,t} + \lambda_b C_{b,t}^{deg} + \lambda_H C_{H,t} + \lambda_\Delta C_{\Delta,t} + \lambda_{pen} C_{pen,t}\right) \tag{20}$$

where $C_{loss,t}$ is the cost of energy conversion losses in the battery and hydrogen chain, $C_{b,t}^{deg}$ is the battery degradation cost proportional to depth of discharge (DoD), $C_{H,t}$ is the hydrogen operation cost, $C_{\Delta,t}$ is the penalty on the rate-of-discharge of control actions to promote smooth operation, $C_{pen,t}$ is the significant penalty term applied only when state constraints are violated, for example is when $H(t)$ or SOC_b exceed their limits, this ensures safe operation.

The weights $\lambda_.$ are tuned to balance these competing objectives, prioritizing system safety and longevity alongside economic efficiency.

Therefore, the SAC algorithm, as shown in Algorithm 1, is an off-policy actor-critic method that maximizes the expected return while also maximizing the policy's entropy. The controller integrates the learned policy with the microgrid's physical constraints, ensuring that the power balance stated in Eq. (1) is maintained while dynamically optimizing the cost-performance trade-off in real-time. The architecture is shown in Fig. 3.

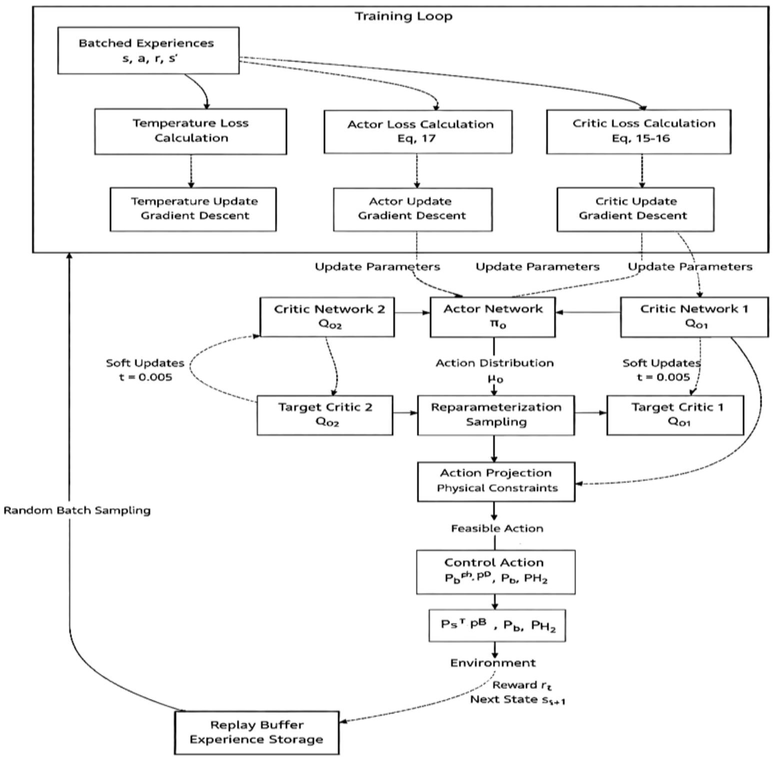

Fig. 3. Soft Actor-Critic AI controller architecture

Algorithm 1: Soft Actor-Critic for Hybrid Hydrogen-Battery Energy Management:

Initialize actor network π_ϕ, critic networks Q_{θ_1}, Q_{θ_2}, and their target networks.

Initialize replay buffer D, temperature parameter α.

for each episode do

Observe initial state s_0.

for each time step t do

Sample action $a_t \sim \pi_\phi(\cdot | s_t)$ using reparameterization (Eq. 14).

Apply action projection to enforce constraints.

Execute a_t, observe r_t, s_{t+1}.
Store (s_t, a_t, r_t, s_{t+1}) in D.
if |D| > $batch_{size}$ then
Sample a random minibatch B from D.
Update critics Q_{θ_1}, Q_{θ_2} by minimizing $J_Q(\theta)$ (Eq. 11).
Update actor ϕ by minimizing $J_\pi(\phi)$ (Eq. 13).
Update temperature α by minimizing J(α) (Eq. 15).
Soft-update target networks: $\bar{\theta} \leftarrow \tau\theta + (1-\tau)\bar{\theta}$.
end if
end for
end for

2.3 Simulation Setup

To evaluate the performance of the proposed SAC-based energy management strategy (EMS), a Python 3.9 simulation environment using the PyTorch library for neural networks was developed. The ordinary differential equations governing component dynamics were discretized and integrated using a forward Euler method with a 15-minute time step.

2.3.1 Dataset Description and Scenario

The microgrid operation was simulated using one year (2023) of real-world 34945 sample historical data with a time resolution of $\Delta t = 0.25$h (15 minutes). The dataset includes the Solar Generation ($P_{\text{pv}}(t)$) , which is the measured solar irradiance data, and Load Demand (cap P sub load , open paren t close paren, which is the commercial building load data consisting of both diurnal and seasonal variations.

The dataset was divided into an 8-month training set, a 2-month validation set for hyperparameter tuning and tracking learning progress, and a 2-month test set for the final performance evaluation reported in this paper. To assess robustness, the test set includes periods of high variability and low renewable generation.

2.3.2 System Parameters

The component sizing for the hybrid microgrid represents a typical commercial-scale installation. The key parameters are summarized in Table 1. The cost function weights (λ) were calibrated to reflect current technology costs and were tuned empirically during preliminary experiments to achieve a balanced performance across all objectives.

Table 1. System simulation parameters

Parameter	Value
Time Step (Δt)	0.25 h
PV Rated Power (P_{pv}^{rated})	150 kW
Battery Max Power ($P_b^{\max}$)	100 kW
Battery Capacity ($E_b^{\max}$)	400 kWh
Battery Min SOC ($SOC_b^{\min}$)	0.2
Battery Efficiency ($\eta_b^{ch}, \eta_b^{dis}$)	0.95
Electrolyzer Max Power ($P_e^{\max}$)	50 kW
Electrolyzer Efficiency (η_e)	0.70
Fuel Cell Max Power ($P_f^{\max}$)	50 kW
Fuel Cell Efficiency (η_f)	0.60
Hydrogen Tank Capacity ($H^{\max}$)	2000 kWh-eq
Degradation Cost Coefficient (λ_b)	0.02 $/kWh
Smoothing Weight (λ_Δ)	0.001

2.3.3 Baseline Controllers

The SAC-based EMS was compared against a rule-based heuristic, which is a priority-based method. Using the instances: (i) If $P_{\text{pv}} > P_{\text{load}}$, there is surplus power, charge battery until full ($SOC_b^{\max}$ is reached), then power the electrolyzer to produce hydrogen, (ii) If $P_{\text{load}} > P_{\text{pv}}$, there is deficit of power, discharge battery first to reach $SOC_b^{\min}$, then use fuel cell, and (iii) Enforce limits and simple thresholds.

2.3.4 Training and Evaluation

The SAC agent was implemented in Python using the PyTorch library. The actor and critic networks are fully connected neural networks with two hidden layers of 256 units each, using ReLU activations. Training was conducted with episodes terminated at 100, each episode comprising 24 hours of operation (i.e., 96 time steps). The key hyperparameters provided in Table 2 were selected through a grid search on the validation set while Table 3 shows how the dataset was splited for model development.

Table 2. DRL Hyperparameters

Hyperparameter	Value
Learning Rate (Actor & Critic)	3×10^{-4}
Discount Factor (γ)	0.99
Replay Buffer Size	1×10^{6}
Batch Size	256
Target Smoothing Coefficient (τ)	0.005
Initial Temperature (α)	0.2
Target Entropy ($\mathcal{H}_{\text{target}}$)	$-\dim(\text{A})$

Table 3. Dataset splits for model development

Dataset	Percentage	Description
Training	67%	Model learning and
Validation	16.5%	Hyperparameter tuning
Test	16.5%	Final performance evaluation

The training was conducted on a workstation with an NVIDIA RTX 3080 GPU. The learned policy from the final training epoch was used for all subsequent testing and comparisons to ensure a fair evaluation.

3. RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the proposed SAC-based energy management system (EMS) for the hybrid hydrogen-battery microgrid. The performance is analyzed using annual simulation results and compared with a conventional rule-based controller (RBC). The next issue discussed involves contextualizing the microgrid environment, followed by an assessment of training performance, and culminating in a comparative analysis of various methodologies for microgrid control strategies.

3.1 Microgrid Data Analysis and Preprocessing

A full year of 2023 data at a high-resolution (15-minute-interval) operational level has been used to conduct a realistic assessment. The dataset summary (Table 4) contains 34,945 samples, providing a strong foundation for both training and testing. Results from the descriptive statistical analysis show that the mean PV power is 32.95 kW and the mean load demand is 93.27 kW (Table 4), indicating a steady energy gap that requires storage solutions. This has been confirmed by the monthly energy balance (Table 5), which shows a net negative energy balance each month. This rate peaks during the period from September to November. The relationship between PV power and load demand (Table 6) is positive and very weak (0.458), indicating that matching solar availability to hotel energy demand is difficult to manage, as the hotel has higher loads on weekdays (Table 7). Fig. 4 shows the microgrid data analysis, while Fig. 5 shows the correlation between PV generation and load demand. Fig. 6 shows the monthly energy balance, while Fig. 7 shows the seasonal variation in PV generation and load demand.

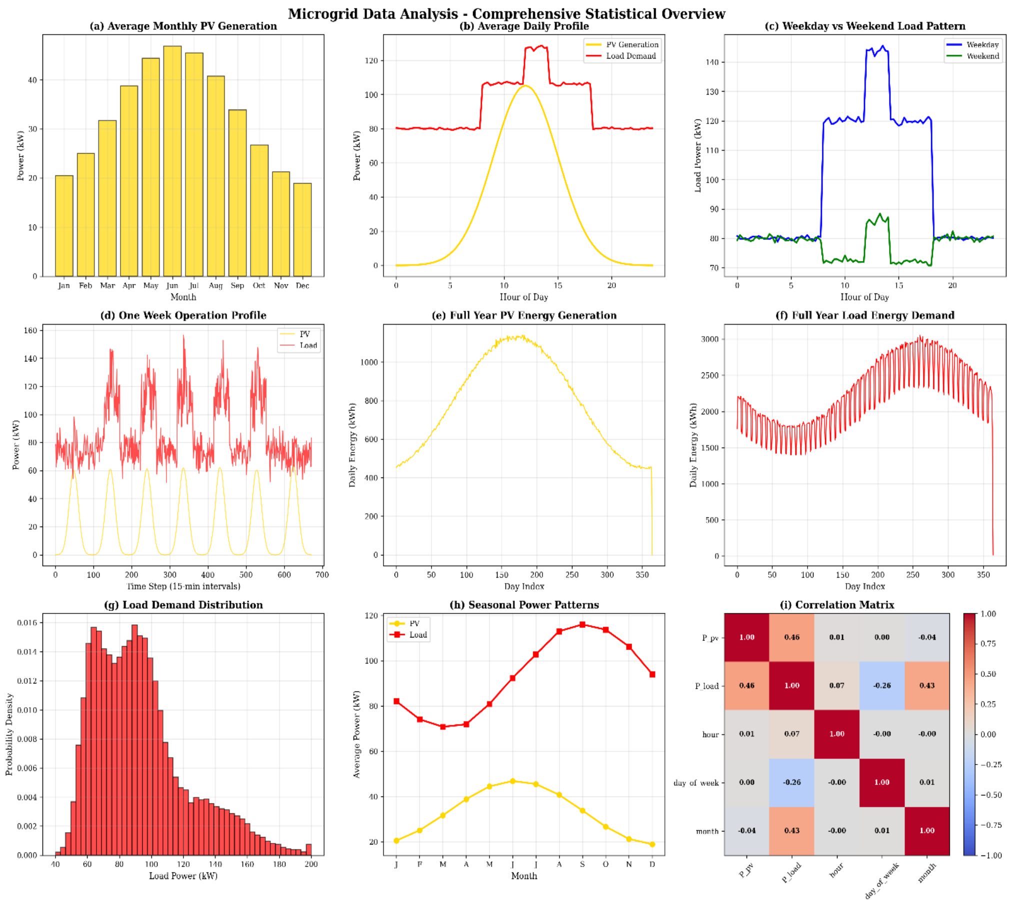

Fig. 4. Microgrid data analysis

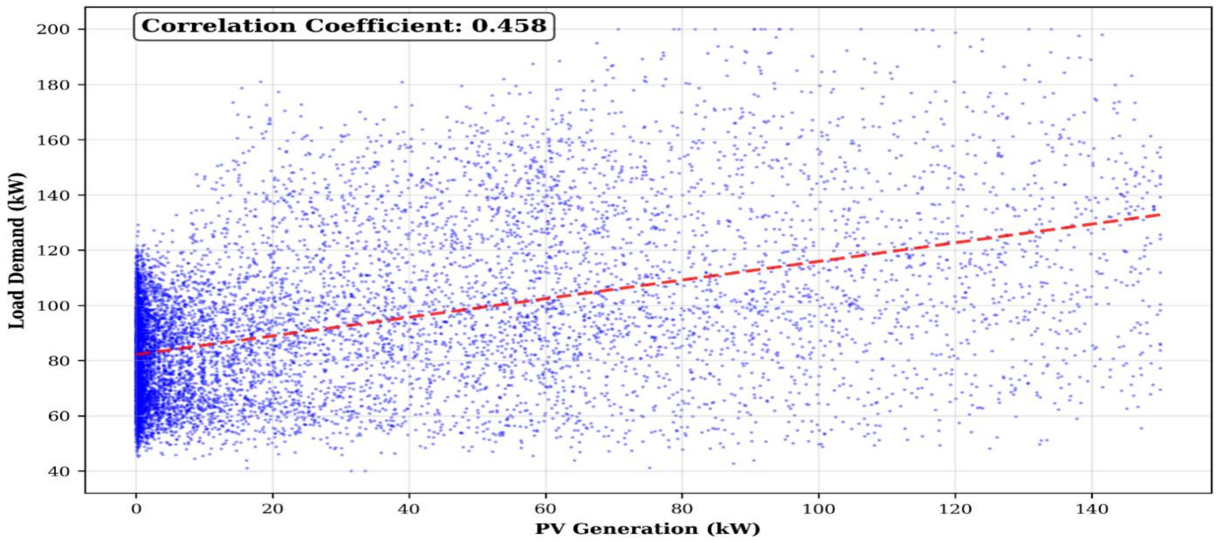

Fig. 5. PV generation vs load demand correlation

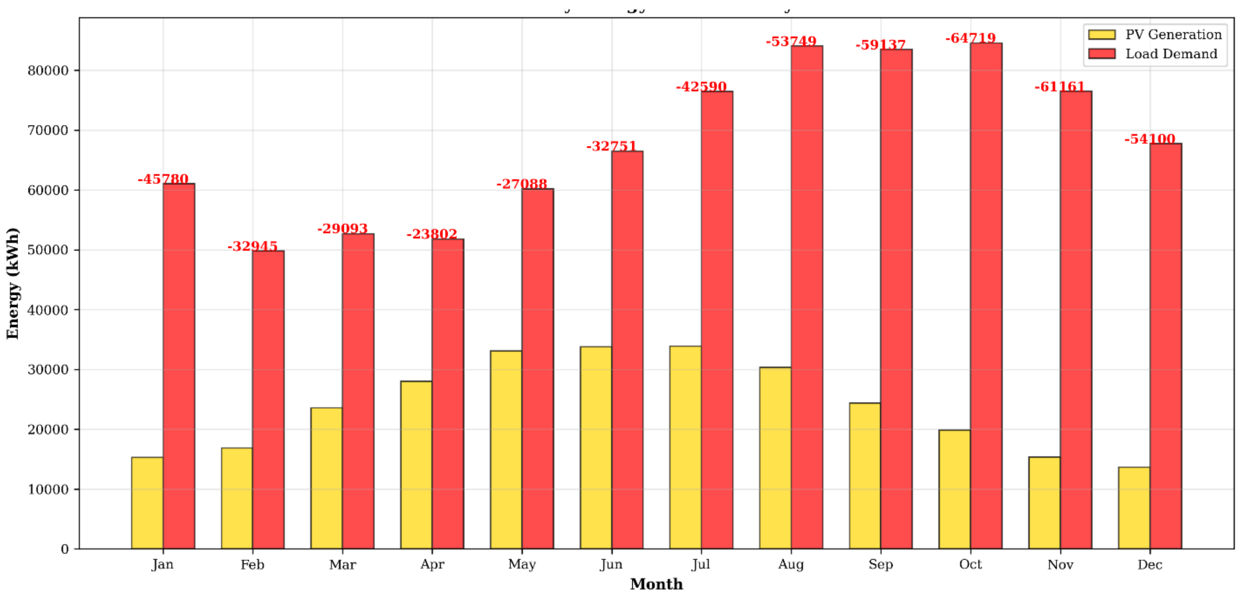

Fig. 6. Monthly energy balance

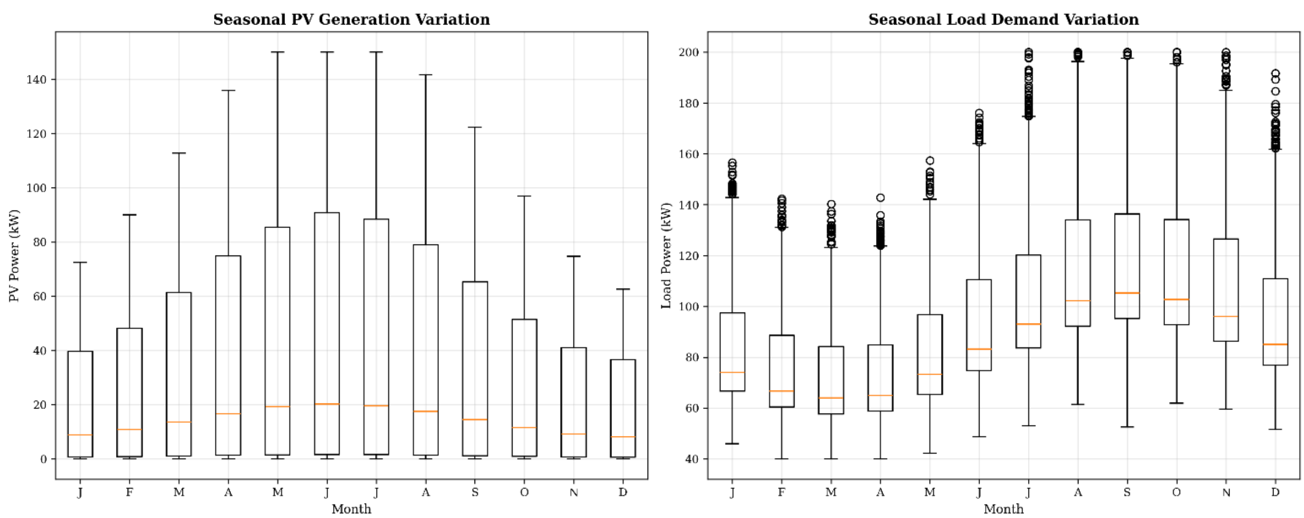

Fig. 7. Seasonal PV generation and load demand variation

Table 4. Power statistics for PV generation and load demand

Statistic	PV Generation (kW)	Load Demand (kW)
Count	34,945	34,945
Mean	32.95	93.27
Std	39.90	29.41
Min	0.02	40.00
25%	1.11	70.63
50%	13.41	88.48
75%	56.95	106.82
Max	150.00	200.00

Table 5. Monthly energy analysis

Month	PV Energy (kWh)	Load Energy (kWh)	Energy Balance (kWh)
1	15,281	61,061	-45,780
2	16,851	49,795	-32,944
3	23,575	52,668	-29,093
4	27,968	51,769	-23,801
5	33,076	60,164	-27,088
6	33,753	66,504	-32,751
7	33,869	76,458	-42,589
8	30,316	84,065	-53,749
9	24,372	83,509	-59,137
10	19,861	84,581	-64,720
11	15,322	76,483	-61,161
12	13,640	67,740	-54,100

Table 6. Correlation analysis

Metric	Value
PV-Load Correlation	0.458

Table 7. Average power by day of week

Day	PV Power (kW)	Load Power (kW)
Monday	32.93	99.34
Tuesday	32.99	99.03
Wednesday	32.94	99.51
Thursday	32.94	99.48
Friday	32.95	99.61
Saturday	32.97	77.94
Sunday	32.94	77.96

3.2 SAC Training Performance and System Operation

The learning process of the SAC agent, shown in Fig. 8, indicates effective learning. The rewards escalate promptly from 3,358.72 at the start to 374.36 in episode 100, signifying efficient convergence toward lower operational costs. The convergence of both the training and validation rewards indicates that the learned policy generalizes effectively and that there are no overfitting issues. The efficient early stop at episode 100 signifies that there is efficient usage of samples.

This is evident in Fig. 8, in which a typical operational cycle of the SAC agent is shown. There is indicative intelligence evident in the SAC controller, wherein the battery is given the utmost importance to counter short-term variations due to its high efficiency, and then careful use of the electrolyzer to exploit surplus solar generation. Then, the fuel cell is activated to provide power during the long PV deficit.

3.3 Comparative Performance Analysis

The comparison between the SAC controller and the Rule-Based Controller is shown in Table 8. The total cost incurred by the SAC controller is $20,655.37, which is 2.0% lower compared to the Rule-Based Controller's cost of $21,084.66. This will result in significant annual savings for the commercial center and address the business problem caused by increased energy use.

A breakdown of the cost components (Table 9 and Fig. 9) shows that there are strategic variations between the two controllers. The RBC incurs low costs related to conversion loss, degradation, and hydrogen costs, which are incurred due to the conservative operation of the hydrogen chain. However, this is accompanied by a high penalty related to power imbalance ($21,051.68) that significantly affects total costs, signifying that there is constant violation of load demand considering the PV and storage limits.

On the other hand, the SAC controller uses all available resource capabilities actively. It has a greater cost for conversion losses ($437.40), battery degradation ($265.88), and costs for hydrogen usage ($502.22) as it acts on available capabilities like the electrolyzer and fuel cell proactively in order to avoid any disparity in power. By this action, it minimizes the power imbalance penalty by $19,423.78. It is clear that the SAC learning agent has properly extracted the trade-off between costs and benefits for the usage of the complete system. Additionally, the SAC controller is able to provide a smoother power outcome with a low value for the smoothing penalty. Such a result decreases mechanical loading on equipment like electrolyzers or fuel cell units and may extend their lifetime.

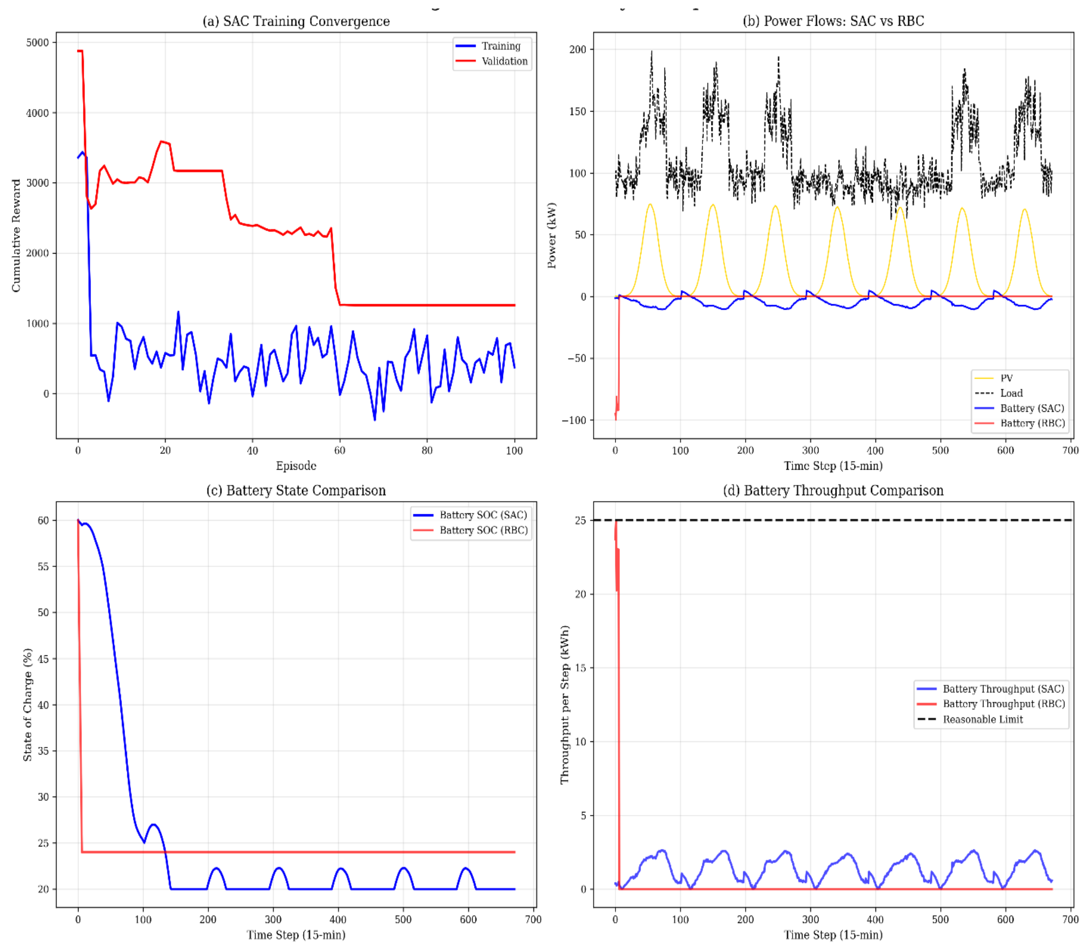

Fig. 8. SAC training performance and system operation

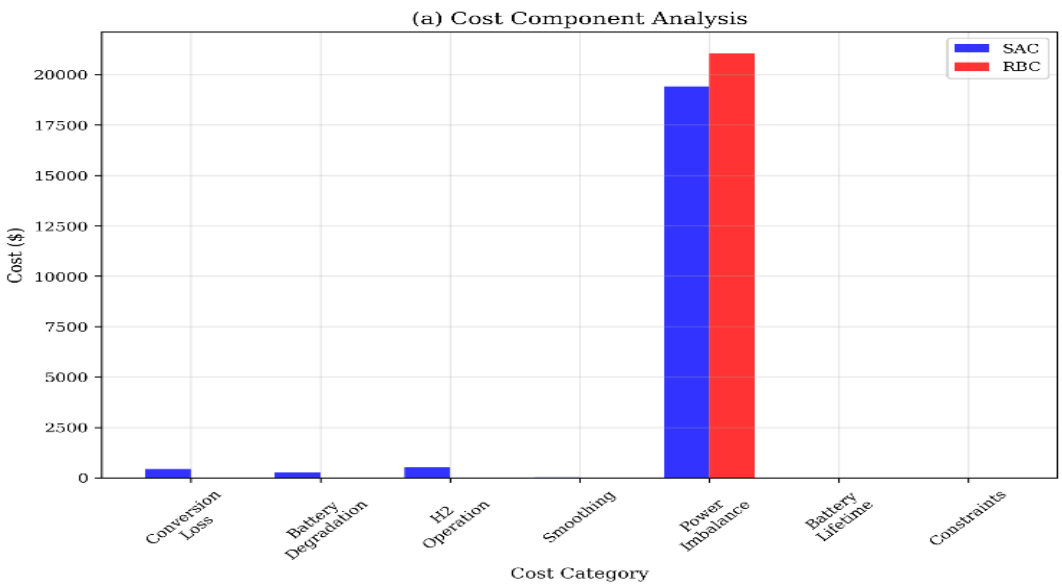

Fig. 9. Cost analysis

Table 8. Overall performance comparison of control strategies

Metric	SAC	RBC	Improvement
Total Cost ($)	20,655.37	21,084.66	2.0%
Battery Throughput (kWh)	2,658.78	136.80	-1,843.6%
Avg Battery SOC (%)	21.86	24.06	-
Avg H_2 Level (%)	11.18	12.44	-
Power Imbalance (kW)	77.695	84.207	-

Table 9. Cost component analysis ($)

Cost Component	SAC	RBC
Conversion Loss	437.40	8.76
Battery Degradation	265.88	13.68
Hydrogen Operation	502.22	9.12
Smoothing Penalty	26.08	1.41
Power Imbalance Penalty	19,423.78	21,051.68
Battery Lifetime Penalty	0.00	0.00
Constraint Penalty	0.00	0.00

3.4 Analysis of Storage System Behaviour

Analysis of storage system dynamics provides further information on the strategies of the respective controllers. The SAC controller shows better average power balance (77.70 kW) than the RBC (84.21 kW) (Table 8), thus proving its superiority in regulating supply-demand balance. The battery throughput for SAC (2,658.78 kWh) is much higher than that of the RBC (136.80 kWh). Even though this increases the cost associated with SAC, it can be termed a trade-off between the benefits. It can be inferred that the SAC controller utilizes the battery with high efficiency for short-term balancing and therefore keeps hydrogen for storage. On the other hand, the RBC controller excessively protects the battery. Hence, its actual potential utilization remains low; therefore, imbalance issues are higher in the case of the RBC controller.

Average SOC for both controllers is approximately at its lowest level (21.86% for SAC and 24.06% for RBC), indicating that both controllers keep trying to save storage capacity in case of potential emergencies, with SAC performing it much better through storage cycling. Hydrogen tank contents are relatively low (11.18% for SAC and 12.44% for RBC), indicating that the control strategy and associated storage capacity do not enable long-term hydrogen storage. Hence, small hydrogen contents in SAC are an indication of its efficient utilization of its fuel cell in order to ensure that there are no power deficits.

3.5 Discussion

The findings prove that DRL-based control has the potential to be more effective in terms of improving energy costs, component degradation, and system smoothness than a standard rule-based approach in determining complex, non-intuitive policies for the control of hybrid storage. The flexibility of the SAC agent in handling several competing objectives—simultaneously trying to minimize costs, component degradation, and maximize system smoothness in a single control solution—is a key strength.

The effectiveness achieved in terms of real-world data analysis based on the challenging commercial load pattern provides evidence of the viability of the approach. The intelligent approach demonstrated in this study provides a generic platform for commercial microgrid operation on weak grids with hydrogen-battery backup. Similarly, Waghmare et al. [26] stressed the need for intelligent control of microgrids to optimize energy use, achieve cost savings, and improve the overall reliability of the microgrids.

It was found that the SAC controller is capable of reducing costs by 2.0%, producing a much smoother transition in power, as well as more intelligent usage of the two storage systems. The findings from this study agree significantly with existing studies. For instance, Ye et al. [27] achieved a 16.02% reduction in scheduling-related costs with the use of the deep deterministic policy gradient (DDPG) reinforcement learning model for solving the scheduling optimization problem, focusing on cost and demand response. Alabdullah and Abido [28] utilized the DQN model to model and simulate the energy management problem. The model demonstrated the capacity to keep the operation cost within 2% of the theoretical optimum. Liang et al. [29] also reported that the use of the DDPG algorithm for intelligent energy scheduling offers significant benefits, including improved renewable energy utilization and cost-effectiveness, compared with conventional control techniques. Zheng et al. [30] employed reinforcement learning to model and simulate the performance of an integrated hydrogen-battery energy storage system. The outcome of the study indicated improved performance in cost efficiency, enhanced power stability, and extended battery life. Furthermore, Zhu et al. [31] achieved cost reduction with 8.43% increase in intraday hydrogen revenue using the DRL model for the simulation of the HESS compared to the use of the baseline model. Kumar et al. [32] achieved optimal and balanced power distribution among the microgrid components using the DRL.

4. CONCLUSION

This work describes the development of a reinforcement learning system geared toward renewable energy management. Specifically, the study considered a hybrid hydrogen battery microgrid, using the Deep Reinforcement Learning paradigm. In response to the pressing concerns of energy costs and reliability for commercial hubs, the research hereby proposes a paradigm shift, not at the supplementation level, but at the extent to which total energy autonomy can be engineered with the AI-optimized storage solution.

The study models the problem using the Soft Actor-Critic algorithm, in a control-theoretic framework, to regulate the intricate energy interactions between the solar energy source, lithium-ion batteries, and the hydrogen energy chain consisting of the electrolyzer, hydrogen tank, and fuel cell. The SAC agent is trained to optimize the agency to learn a continuous-action policy, trading off between cost minimization, degradation, and operational smoothness. Simulations performed on a real-world dataset for a year at a high resolution establish the effectiveness of the EMS based on the SAC algorithm over the traditional rule-based controller. Compared to the traditional rule-based control system, the proposed SAC-based system offers a 2.0% annual cost savings, which is significant for the company.

This saving is achieved mainly by the agent's clever use of the entire hybrid storage system, significantly reducing power imbalances. Analysis shows that the proposed system, based on the SAC algorithm, makes a smart use of the battery for buffering and the hydrogen system for energy shifting. Based on the above discussion, the research work makes a significant contribution toward filling a critical research gap in the existing body of knowledge by offering a holistic control-theoretical framework for DRL, that considers system dynamics and multi-objective optimization without the need for oversimplification.

The findings confirm the significant potential of DRL technology, and, more specifically, the SAC model, to support the economically feasible and operationally viable design and implementation of hybrid hydrogen-battery microgrids. This research has particular significance for for-profit companies operating in areas where the power grid may be costly or unreliable. It is, however, limited to the use of the Soft Actor-Critic algorithm for the optimal management of a hybrid hydrogen-battery energy storage system in a microgrid based on renewables. Future research may involve:

i. A comparative analysis with other reinforcement learning model such as the DDPG. The use of a multi-agent system can also be considered due to the complexity and non-linear problem of microgrids.
ii. Extending the system model to include load-dependent efficiency curves to enhance generalization,
iii. Testing the reuse potential of the policy in different commercial facilities that operate under different loads and conditions to enable scalability; and
iv. Applying the policy to a real-time testbed to verify its performance.

ACKNOWLEDGMENT

The authors acknowledge Bells University of Technology, Ota, Nigeria, where this work was carried out, for creating an enabling environment for this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] M.S. Su'ait, N.A. Ludin, K. Sopian, Introduction to Renewable Energy Technologies and Energy Transition Strategies. *ACS symposium Series*, 1449, 2025: 1–14. https://doi.org/10.1021/bk-2025-1499.ch001

[2] S. Bielecki, T. Skoczkowski, M. Wołowicz, Microgrids as a Tool for Energy Self-Sufficiency. *Sensors,* 25(21), 2025: 6707. https://doi.org/10.3390/s25216707

[3] P. Cosgrove, T. Roulstone, S. Zachary, Intermittency and Periodicity in Net-Zero Renewable Energy Systems with Storage. *Renewable Energy,* 212, 2023: 299–307. https://doi.org/10.1016/j.renene.2023.04.135

[4] P. Parvizi, M. Jalilian, A.M. Amidi, M.R. Zangeneh, J.R. Riba, From Present Innovations to Future Potential: The Promising Journey of Lithium-Ion Batteries. *Micromachines*, 16(2), 2025: 194. https://doi.org/10.3390/mi16020194

[5] R. Areola, A.A. Adebiyi, K. Moloi, Integrated Energy Storage Systems for Enhanced Grid Efficiency: A Comprehensive Review of Technologies and Applications. *Energies,* 18(7), 2025: 1848. https://doi.org/10.3390/en18071848

[6] R.Y. Abdallah, M.F. Shaaban, A.H. Osman, A. Ali, K. Obaideen, L. Albasha, Synergizing Gas and Electric Systems Using Power-To-Hydrogen: Integrated Solutions for Clean and Sustainable Energy Networks. *Smart Cities,* 8(3), 2025: 81. https://doi.org/10.3390/smartcities8030081

[7] Y. Zhao, H. Li, C. Wan, D. Du, B. Chen, Microgrid Energy Management Considering Energy Storage Degradation Cost. *Batteries*, 11(5), 2025: 169. https://doi.org/10.3390/batteries11050169

[8] M.S. Endiz, G. Gökkuş, A.E. Coşgun, H. Demir, A Review of Traditional and Advanced MPPT Approaches for PV Systems Under Uniformly Insolation and Partially Shaded Conditions. *Applied Sciences*, 15(3), 2025: 1031. https://doi.org/10.3390/app15031031

[9] H. Liu, H. Wang, M. Yu, Y. Wang, Y. Luo, Long Short-Term Memory–Model Predictive Control Speed Prediction-Based Double Deep Q-Network Energy Management for Hybrid Electric Vehicle to Enhanced Fuel Economy. *Sensors,* 25(9), 2025: 2784. https://doi.org/10.3390/s25092784

[10] L. Jia, Y. Pei, Recent advances in Multi-Agent Reinforcement Learning for Intelligent Automation and Control of Water Environment Systems. *Machines,* 13(6), 2025: 503. https://doi.org/10.3390/machines13060503

[11] G. Muriithi, S. Chowdhury, Deep Q-network Application for Optimal Energy Management in a Grid-Tied Solar PV-Battery Microgrid. *The Journal of Engineering,* 2022(4), 2022: 422–441. https://doi.org/10.1049/tje2.12128

[12] A. Bäumler, A. Benterki, J. Meng, T. Azib, M. Boukhnifer, Energy Management Strategy based on Reinforcement Learning for Fuel Cell Hybrid Vehicle with a New Reward Function Approach. *2023 IEEE Vehicle Power and Propulsion Conference*, 24-27 October 2023, Milan, Italy, pp.1–6.

[13] E. Mohammadi, D. Ortiz-Arroyo, A.A. Hansen, M. Stokholm-Bjerregaard, S. Gros, A.S. Anand, P. Durdevic. Application of Soft Actor-Critic Algorithms in Optimizing Wastewater Treatment with Time Delays Integration. *Expert Systems with Applications*, 277, 2025: 127180. https://doi.org/10.1016/j.eswa.2025.127180

[14] X. Lü, S. Li, X. He, C. Xie, S. He, Y. Xu, J. Fang, M. Zhang, X. Yang, Hybrid Electric Vehicles: A Review of Energy Management Strategies Based on Model Predictive Control. *Journal of Energy Storage,* 56, 2022: 106112. https://doi.org/10.1016/j.est.2022.106112

[15] A. Dosa, O.A. Olanrewaju, F. Mora-Camino, A Comprehensive Review of Hybrid Renewable Microgrids: Key Design Parameters, Optimization Techniques, and the Role of Demand Response in Enhancing System Flexibility. *Energies,* 18(19), 2025: 5154. https://doi.org/10.3390/en18195154

[16] M.A.P. Ramos, A. Busboom, Systematic Review of Reinforcement Learning in Process Industries: A Contextual and Taxonomic Approach. *Applied Sciences,* 15(24), 2025: 12904. https://doi.org/10.3390/app152412904

[17] M. Aatabe, W. Jenkal, M.I. Mosaad, S.A. Hussien, Stochastic Control for Sustainable Hydrogen Generation in Standalone PV–Battery–PEM Electrolyzer Systems. *Energies,* 18(15), 2025: 3899. https://doi.org/10.3390/en18153899

[18] X. Liu, H. Qi, S. Jia, Y. Guo, Y. Liu, Recent Advances in Optimization Methods for Machine Learning: A Systematic Review. *Mathematics,* 13(13), 2025: 2210. https://doi.org/10.3390/math13132210

[19] P. Michailidis, I. Michailidis, F. Minelli, H.H. Coban, E. Kosmatopoulos, Model Predictive Control for Smart Buildings: Applications and Innovations in Energy Management. *Buildings*, 15(18), 2025: 3298. https://doi.org/10.3390/buildings15183298

[20] J. Xu, Efficiency-Oriented Model Predictive Control: A Novel MPC Strategy to Optimize the Global Process Performance. *Sensors*, 24(17), 2024: 5732. https://doi.org/10.3390/s24175732

[21] Z. Zhang, Z. Wu, D. Rincon, P. Christofides, Real-Time Optimization and Control of Nonlinear Processes Using Machine Learning. *Mathematics,* 7(10), 2019: 890. https://doi.org/10.3390/math7100890

[22] M. Gautam, Deep Reinforcement Learning for Resilient Power and Energy Systems: Progress, Prospects, and Future Avenues. *Electricity,* 4(4), 2023: 336–380. https://doi.org/10.3390/electricity4040020

[23] S. Meziane, T. Ryad, Y.O. Assolami, T.M. Aljohani, A Hierarchical Two-Layer MPC-Supervised strategy for Efficient Inverter-Based Small Microgrid Operation. *Sustainability,* 17(19), 2025: 8729. https://doi.org/10.3390/su17198729

[24] S. Dai, P. Shen, W. Deng, Q. Yu, Hydrogen Energy in Electrical Power Systems: A Review and Future Outlook. *Electronics,* 13(17), 2024: 3370. https://doi.org/10.3390/electronics13173370

[25] D. Deltetto, D. Coraci, G. Pinto, M. S. Piscitelli, A. Capozzoli, Exploring the Potentialities of

Deep Reinforcement Learning for Incentive-Based Demand Response in a Cluster of Small Commercial Buildings. *Energies,* 14(10), 2021: 2933. https://doi.org/10.3390/en14102933

[26] A.V. Waghmare, V.P. Singh, T. Varshney, P. Saanjevikumar, A Systematic Review of Reinforcement Learning-Based Control for Microgrids: Trends, Challenges, and Emerging Algorithms. *Discover Applied Sciences*, 7, 2025: 939. https://doi.org/10.1007/s42452-025-07529-6

[27] J. Ye, X. Wang, Q. Hua, L. Sun, Deep Reinforcement Learning Based Energy Management of a Hybrid Electricity-Heat-Hydrogen Energy System with Demand Response. *Energy,* 305, 2024: 131874. https://doi.org/10.1016/j.energy.2024.131874

[28] M.H. Alabdullah, M.A. Abido, Microgrid Energy Management using Deep Q-Network Reinforcement Learning. *Alexandria Engineering* Journal, 61(11), 2022: 9069–9078. https://doi.org/10.1016/j.aej.2022.02.042

[29] T. Liang, L. Chai, X. Cao, J. Tan, Y. Jing, L. Lv, Real-time Optimization of Large-Scale Hydrogen Production Systems Using Off-Grid Renewable Energy: Scheduling Strategy Based on Deep Reinforcement Learning. *Renewable Energy,* 224, 2024: 120177. https://doi.org/10.1016/j.renene.2024.120177

[30] Y. Zheng, J. Jia, D. An, Energy management for Microgrids with Hybrid Hydrogen-Battery Storage: A Reinforcement Learning Framework Integrated Multi-Objective Dynamic Regulation. *Processes*, 13(8), 2025: 2558. https://doi.org/10.3390/pr13082558

[31] W. Zhu, S. Wen, Q. Zhao, B. Zhang, Y. Huang, M. Zhu, Deep Reinforcement Learning Based Optimal Operation of Low-Carbon Island Microgrid with High Renewables and Hybrid Hydrogen–Energy Storage System. *Journal of Marine Science and Engineering,* 13(2), 2025: 225. https://doi.org/10.3390/jmse13020225

[32] K. Kumar, S. Kwon, S. Bae, Deep Reinforcement Learning-Based Control Strategy for Integration of a Hybrid Energy Storage System in Microgrid. *Journal of Energy Storage,* 108(13), 2025: 114936. https://doi.org/10.1016/j.est.2024.114936