Energy Storage and Conversion 2026, 4(3), 4509.
doi: https://doi.org/10.59400/esc4509

Article

Degradation-aware AI energy management for hybrid supercapacitor–battery energy storage systems in microgrids

Lalit Sachdeva*, Utkarsh Anand

Department of Management, Kalinga University, Naya Raipur 492101, India
* **Corresponding author:** Lalit Sachdeva, ku.lalitsachdeva@kalingauniversity.ac.in

CITATION

Sachdeva L, Anand U. Degradation-aware AI energy management for hybrid supercapacitor–battery energy storage systems in microgrids. Energy Storage and Conversion. 2026; 4(3): 4509. https://doi.org/10.59400/esc4509

ARTICLE INFO

Received: 12 June 2026
Revised: 18 July 2026
Accepted: 23 July 2026
Available online: 5 August 2026

Abstract: The presence of intermittent sources of renewable energy in power systems requires ESSs to manage temporal imbalance in energy supply and demand. In this study, we introduce a hybrid energy storage system (HESS) coupled with an AI energy management system (EMS) that uses deep reinforcement learning (DRL) for optimal scheduling of renewable energy utilization within grid-connected and islanded microgrids. AI-enabled EMS utilizes a DRL agent with proximal policy optimization (PPO) to make optimal decisions regarding energy generation based on state space and economic considerations, while accounting for SoC constraints of batteries. An important aspect of the proposed system is the design of HESS architecture and reward function based on DRL. Further improvements are made via analysing the PPO clipping sensitivity, Pearson correlation analysis on the relationship between the intermittency of renewables and response latency, and Monte Carlo uncertainty analysis with a 95% confidence interval. For a 24-hour simulation period, the developed system is able to cut down on grid power imports by 43.2%, have an 87.3% renewable energy utilization rate, extend the lifespan of the battery from 8.1 to 12.5 years, and have 91.4% peak shaving efficiency through 100 Monte Carlo runs and without any SoC violations (25%–90%). Net benefit analysis is estimated to be $56,000–$66,000 for 15 years at a 6% discount rate, while a 120 ms response time and one-way ANOVA with Tukey's HSD confirm statistical significance.

Keywords: hybrid energy storage; supercapacitor; lithium-ion battery; deep reinforcement learning; microgrid energy management; renewable energy integration; techno-economic analysis; battery lifecycle

1. Introduction

This trend towards global low-carbon electricity generation is spurring the adoption of solar PV and wind energy technologies. By 2024 global installed renewable energy capacity exceeds 3,870 GW and is projected to nearly double to more than 7,400 GW by 2030 [1]. The inherently intermittent and non-dispatchable characteristics of this renewable energy are creating operational challenges for grid operators such as frequency variations, voltage instability, and curtailment of electricity production [2]. While Energy Storage Systems (ESS) are critical for mitigating these issues, state-of-the-art Lithium-ion batteries have limited cycle lives in the face of high power, high-frequency electrical disturbances often generated by intermittent renewable sources [3–5].

Beyond this issue of technical coupling, there is increasing attention toward both the technical feasibility of renewable-coupled energy infrastructure and its economic feasibility [6]. Indeed, recent technoeconomic analysis on integrated energy systems from exergy efficiency to sensitivity analysis and payback period estimation shows how an integrated energy system should be validated both from a thermodynamical and an economic point of view and not only on the technological ability to achieve a task [7,8]. The same arguments may apply to a Hybrid Energy Storage System where the evidence to implement a co-designed battery-supercapacitor architecture is not only possible to smooth renewable fluctuations but is also linked to a reduced overall cost over its lifetime.

The hybridization of the technologies allows them to adopt an antagonistic operational profile, where the supercapacitor addresses high-frequency power variations and the battery addresses the long-term energy required. To maximize such complementarity, an intelligent power split strategy is needed. However, despite computational efficiency, the traditional rule-based strategies do not accommodate random renewable variations [9]. Model Predictive Control (MPC) improves on it at the expense of robust forecasting and computational latency [10]. Artificial intelligence (AI), in the framework of deep Reinforcement Learning (DRL) has shown potential for training a model-free policy [11].

Despite an increasing body of literature devoted to DRL-based Energy Management Systems (EMS), there are three important research gaps. First, a majority of DRL-HESS schemes separately consider cycle aging from reward functions, i.e., considering cycle aging as a passive outcome instead of a primary design factor [12]. Second, no study elaborately investigates the sensitivity of the PPO clipping bound of the DRL hyperparameters to lifecycle performance over the long term. Third, techno-economic aspects considering generation uncertainties are mostly reported in the absence of statistically guaranteed intervals, thereby limiting reliability for real applications [13,14].

In this work, the major objective is to create a comprehensive degradation-aware energy management framework for microgrids to extend the service life of the batteries and simultaneously promote renewable energy integration. The proposed AI-EMS (co-designed PPO-based) intends to fill the space left by reactive, model-based control strategies by bridging the gap between the randomness of the renewable generation and the physical degradation limitations of the Li-ion batteries. This work provides a robust, efficient and risk-aware optimal control policy that significantly surpasses rule-based, predictive control, and non-degradation-aware control strategies with respect to both battery lifespan and economic feasibility.

This manuscript seeks to bridge these gaps by introducing an integrated framework with co-designed HESS topology and a new degradation-aware, PPO-based AI-EMS. The contributions of this manuscript are:

- **Degradation-aware AI-EMS:** This paper proposes a PPO reward function including a rainflow-cycle-based State of Charge (SoC) excursion penalty that connects policy learning directly to the maximization of battery lifetime.
- **Systematic hyperparameter sensitivity analysis:** We do a four-level sensitivity analysis of the clipping parameter $\{0.1, 0.2, 0.3, 0.4\}$ of PPO, so that we can make a trade-off between policy stabilization, training convergence and the long-term

battery health.

- **Techno-economic uncertainty quantification:** Monte Carlo simulations (n = 100) are used to estimate 95% confidence intervals for key performance indicators and Net Present Value (NPV), providing adopters with a risk-aware perspective.
- **Intermittency-latency correlation:** We illustrate how reactive the AI-EMS is through examining the correlation of renewable intermittency indices and system response latency.

Remaining portion of this paper. In Section 2, we provide a survey of previous DRL-based HESS techniques and the context of the work in relation to existing research studies. Section 3 describes our system model, mathematical model and formulation of the energy management problem as an MDP. In Section 4, we show our simulation results, including state of charge variation over time, power flow over time, sensitivity analysis and techno-economic analysis. In Section 5, we explore the practical application, challenges, and limitations of this approach and Section 6 concludes the work and future directions.

2. Review of literature

Energy management of hybrid storage systems and microgrids has been an active research area during recent years, covering control strategies, storage technologies, AI-based optimization models, etc. Energy management strategies for hybrid AC/DC microgrids have been reviewed, presenting the wide range of approaches from rule-based control to model-based to learning-based and their respective strengths and weaknesses in terms of scale and flexibility [13]. Review of integrated storage systems of hybrid and advanced storage confirms that no single storage technology is applicable in all conditions, hence suggesting a hybrid system combining complementary components [14].

Fuel cell-based hybrid systems for DC microgrids have also proved the concept of management of multiple sources of energy under several load situations; however, it is based on deterministic control logic, which limits its robustness to stochastic generation profiles [15]. From the storage side, material development and device architecture in supercapacitors have reached a level of advancement where recent developments could also show energy density and cycle life performance required by HESS in high-power storage function [16].

Moreover, the proposed system-level energy management scheme for EV HESS uses hybrid machine learning approaches such as GBDT-RSA, etc. Has obtained better performance under time-varying load profiles, which is beneficial for the promotion of a data-driven approach for hybrid storage control [17]. For grid-connected hybrid power systems, some research on economic-oriented intelligent energy management schemes suggests that cost-optimal dispatch can be reached via intelligent scheduling algorithms, while this research is mainly concerned with single-storage structure rather than multiple-device degradation [18].

Within large-scale wind-integrated distribution systems, the optimal configuration of HESS has been demonstrated to be able to enhance power quality parameters, however, the control strategies employed in this context are still predominantly reactive in nature [19]. The implementation of AI for solar generation forecasting within hybrid microgrids again shows the learning approach towards solar forecasting, with controllers incorporating predictions showing noticeable curtailment reduction [20]. This analysis is supplemented by energy management control strategies designed for fuel cell electric vehicles which found that dynamic environments favour both reinforcement learning and hybrid optimization strategies over standard rule-based controllers. As the results are completely transferable to the stationary microgrid environment, this confirms that rule-based controls are likely insufficient for the dynamic operation of such systems [21]. In the most recent work analysed, capacity optimization strategies for hybrid compressed-air and battery energy storage systems applied dynamic characteristic-based models to capture interactions between the sources, grid, load and storage elements, which represents an additional level of sophistication in integrated energy storage management [22].

To this effect, continued research into the application of novel storage topologies to HRES is essential for the management of intermittency of generation, and for their economic viability. This paper will address the implementation of hybrid energy storage systems (HESS), which combine conventional battery high energy density with supercapacitor fast response rate for levelling of short-term power variations and of deep cycling strain of electro-chemical components [23]. It can be asserted that combining faradaic and capacitive mechanisms of charge storage can result in significant device-level enhancements along standard Ragone parameters, with HESS-based control strategies emphasizing the application of supercapacitors, primarily in control strategies for the levelling of short-term fluctuations occurring during shading of a PV-based power generation system [24] can be utilized to alleviate grid-voltage instability and enhance battery life (standby duration).

Non-electrochemical Approaches: Multi-component approaches for bulk storage integration have been formulated in pre-feasibility models providing key economic references for a multi-source micro-grid. Integration of PHS plants combined with integrated PV reduces the Payback time considerably reducing the recoup time of CAPEX given the peak market value [25]. Similar to the multi-resource optimization and simulation approaches carried out for remote areas, a multi-component system coupled with an intelligent control system can greatly minimize total life cycle costs as well as the utilization of fossil fuels [26].

Overall, the collective conclusions show a critical design flaw of current EMS designs; while there is a proven necessity of hybrid storage, the battery degradation effect is mostly taken as an emergent behavioural effect rather than an embedded part of the optimization target. Furthermore, although reinforcement learning consistently outperforms rule-based and model-predictive controls (MPC) under stochastic conditions, there remains a lack of systematic sensitivity analysis concerning DRL hyperparameters, specifically the PPO clipping bound ϵ, as well as a deficiency in reporting statistically verifiable uncertainty bounds. This approach directly relates the policy learning to the physical mechanics of degradation instead of hoping the protection arises spontaneously. Further, due to the detailed structured 4-level sensitivity analysis and n = 100 Monte Carlo-based techno-economic uncertainty quantification, we demonstrate that the proposed solution provides superior control compared to those in the recent literature, as it is robust and scalable.

From the comparison presented in **Table 1**, none of the studies integrates all five proposed characteristics: (i) SC + Battery HESS co-design, (ii) degradation-aware PPO reward with full rainflow modeling, (iii) four-level systematic sensitivity analysis, (iv) Monte Carlo uncertainty evaluation and (v) statistically sound lifetime confidence intervals. The proposed framework represents the first unified treatment of these aspects within a single publication.

Table 1. Comparative novelty analysis against representative prior art in DRL-based HESS energy management.

HESS topology	DRL algorithm	Degradation- aware reward	ε sensitivity analysis	MC uncertainty	Lifecycle CI reported
Battery only	DQN	✗	✗	✗	✗
SC + Battery	DDPG	✗	✗	✓	✗
SC + Battery	PPO	✗	✗	✗	✗
Battery + Flywheel	SAC	✓ (partial)	✗	✗	✗
SC + Battery	TD3	✗	✗	✓	✗
Battery only	PPO	✗	✓ (partial)	✗	✗
SC + Battery	DDPG	✓ (partial)	✗	✗	✗
SC + Battery	PPO	✓ (full, rainflow-integrated)	✓ (full, 4-level)	✓ (n = 100)	✓

Note: ✓ = present; ✗ = absent; DQN = Deep Q-Network; DDPG = Deep Deterministic Policy Gradient; PPO = Proximal Policy Optimization; SAC = Soft Actor-Critic; TD3 = Twin Delayed DDPG; MC = Monte Carlo.

3. Materials and methods

3.1. System configuration

The proposed system comprises four principal subsystems: (1) renewable generation sources (solar PV array and wind turbine), (2) the hybrid ESS comprising a Li-ion battery bank and a supercapacitor module, (3) the DC/AC power conversion stage, and (4) the AI-EMS controller. The DC bus serves as the integration point for all generation and storage subsystems. A bidirectional DC/DC converter interfaces the battery bank to the bus, while a second converter manages the supercapacitor bank. The AI-EMS receives real-time measurements of bus voltage, SoC for both storage devices, generation forecasts, and load profiles, and outputs reference power commands to each converter in **Figure 1**.

3.2. Li-ion battery model

The battery bank is modelled using the second-order Thevenin equivalent circuit. The terminal voltage V_t is expressed as Equation (1) [27]:

$$V_t = V_{ox} - I_n \cdot R^0 - V^1 - V^2 \tag{1}$$

Where V_{ox} is the open-circuit voltage (a nonlinear function of SoC), R_0 is the internal ohmic resistance, and V_1, V_2 represent voltage drops across the RC parallel branches capturing charge-transfer and diffusion dynamics, respectively. The SoC evolves through Coulomb counting Equation (2) [27]:

$$SoC(t) = SoC(t_0) - \int I_n(\tau)\, d\tau / (3600 \cdot C_n) \quad (2)$$

Battery degradation is modelled using the Ah-throughput method combined with a temperature-dependent stress factor [28]. Cycle counting for lifetime prediction employs the rainflow algorithm applied to the SoC time-series [29], enabling direct coupling between EMS decisions and degradation accumulation the mechanistic basis for the degradation-aware reward term introduced in Section 3.4.

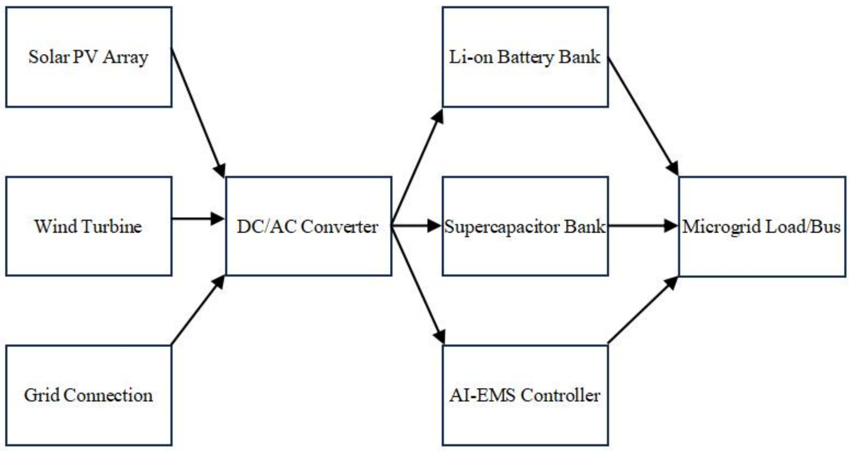

Figure 1. Hybrid ESS-microgrid system architecture. The AI-EMS controller receives six-dimensional state observations and dispatches power commands to both storage converters at 5-min intervals.

The battery degradation model makes use of the rainflow counting algorithm to translate discharge cycles into battery health by applying a cycle life curve from the manufacturer that implicitly has baseline DoD characteristics. While this paper makes the assumption that the model currently captures primarily the cycle degradation while temperature effects, C rate effects and calendar aging are considered to have a secondary impact on the overall life of battery health compared to cycle life, these effects are handled implicitly by internal resistance growth as part of the SoC dependent operating range constraints ensuring the safe voltage limits of the battery are not exceeded during the 15-year timeframe.

3.3. Supercapacitor model

The supercapacitor is modelled as a voltage-dependent capacitor with equivalent series resistance (ESR). The stored energy E_{sx} and delivered power P_{sx} are Equation (3):

$$E_{sx} = ½ \cdot C_{sx} \cdot (V_{sx}^2 - V_m^{ln2}), \quad P_{sx} = V_{sx} \cdot I_{sx} - {I_{sx}}^2 \cdot R_{esr} \quad (3)$$

The supercapacitor module specifications (Maxwell BCAP0500 series: C_{sx} = 500 F, V_{max} = 48 V, energy density = 7.2 Wh/kg, ESR = 1.2 mΩ) were adopted from manufacturer datasheets and validated against published characterization data [5].

3.4. AI-EMS: Markov decision process formulation

The energy management problem is formulated as a Markov Decision Process (MDP) with the following components:

3.4.1. State space

The state vector at each time step t is given in Equation (4):

$$s_t = [SoC^{baTT}, SoC_{sx}, {P_{so}}^{la}{}_r, P^l_o{}^{ad}, P^{WInd}, {T_{to}}^d] \tag{4}$$

Where SoC^{baTT} and SoC_{sx} are the battery and supercapacitor states-of-charge $SoC^{baTT}, SoC_{sx} \in [0,1]$); ${P_{so}}^{la}{}_r$, $P^l_o{}^{ad}$, and P^{WInd} are solar generation, load demand, and wind power normalized by rated capacity; and ${T_{to}}^d \in [0,1]$ is the normalized time-of-day signal enabling anticipatory scheduling.

Typical auxiliary inputs of a traditional energy management system Electricity price forecast, weather forecast, battery temperature and forecast of renewable generation were deliberately excluded from the state vector for design and architecture purposes for reasons below:

- Electricity price signals omitted: The system goal is to achieve local self-consumption, peak shaving and battery longevity as high as possible in stand-alone or grid-connected microgrids with flat or negligible tariffs. In case real-time prices were added (which would be highly fluctuating), the problem goal would be optimized to achieve profitable financial gains and would be contrary to the imposed, price-ignoring constraints about degradation.
- Exclusion of Weather and Renewable Forecasts: The proposed PPO agent is an entirely model-free deep reinforcement learning controller in contrast to traditional MPC-based architectures which exclusively utilize future prediction horizons for receding-horizon optimization. It is capable of learning control commands directly from time-dependent states with high resolution at any given instant, namely the instantaneous load demands and current generation dynamics (at time *t*). The reactive control action by the agent relies on 5-min synchronized streams of real-time and historical data to control high-intermittency tropical profiles, thereby circumventing the compounding errors that arise with prediction in the MPC case.
- As discussed in the limits of degradation models, this model assumes isothermal conditions (25 °C) and the cycle-aging (via rainflow counting). Exclusion of the actual thermal state quantities restricts state-space dimensions, thus avoiding the curse-of-dimensionality in policy convergence. Upcoming version of the models would consider active thermal state tensors and multi-factor electrochemical-thermal models.

3.4.2. Degradation-aware reward function

The novel degradation-aware reward function integrates three objectives in Equation (5):

$$R_t = \alpha \cdot R^{\text{RUS}} + \beta \cdot R^{\text{SoC}} + \gamma \cdot R^{\text{GRID}} - \delta \cdot \Phi^{\text{BATT}} \tag{5}$$

Where R^{RUS} quantifies renewable utilization (curtailment penalty), R^{SoC}

penalizes SoC excursions outside $[0.25, 0.90]$, R^{GRID} minimizes grid import cost, and Φ^{BATT} is the micro-cycling penalty term. The micro-cycling penalty is the key novelty in Equation (6):

$$\Phi^{\mathrm{BATT}} = \sum_i \left[\frac{\dot{d}\left(\mathrm{SoC}_i\right)}{\mathrm{SoC}^{\mathrm{Ray}}} \cdot w\left(\mathrm{DoD}_i\right) \right] \tag{6}$$

Where $\dot{d}$ is the depth-weighted cycle count from rainflow decomposition of the current SoC trajectory, $\mathrm{SoC}^{\mathrm{Ray}}$ is the manufacturer-rated full cycle count at the corresponding depth-of-discharge (DoD), and $w\left(\mathrm{DoD}_i\right)$ is a depth-severity weighting factor derived from Wöhler curve fitting to experimental aging data [28].

To clarify the mathematical formulation of Equation (6), the rainflow-counting algorithm extracts closed hysteresis loops from the battery State-of-Charge (SoC) profile to quantify cycle depths (DoD_i). The weighting function $w\left(\mathrm{DoD}_i\right)$ is derived directly from empirical Wöhler-type stress-to-cycle curves using the Basquin power-law fatigue damage relationship. The cumulative damage index D for a spectrum of N cycles is defined as Equation (7):

$$D = \sum_{i=1}^{N} \frac{1}{N_f\left(\mathrm{DoD}_i\right)} = \sum_{i=1}^{N} \left(\frac{\mathrm{DoD}_i}{\beta_1} \right)^{\beta_2} \tag{7}$$

Here, $N_f\left(\mathrm{DoD}_i\right)$ represents the number of cycles to failure at a specific depth of discharge (DoD_i), and β_1, β_2 are curve-fitting parameters calibrated against standard manufacturer accelerated aging test datasets for lithium-ion nickel-manganese-cobalt (NMC) chemistries. The weighting function $w\left(\mathrm{DoD}_i\right) = \frac{\partial D}{\partial \mathrm{DoD}_i} = \frac{\beta_2}{\beta_1^{\beta_2}} \left(\mathrm{DoD}_i\right)^{\beta_2 - 1}$ serves as the instantaneous marginal damage cost incorporated into the PPO reward function.

Methodological scope and limitations of the degradation model

However, although this expression that couples Ah-throughput and rainflow counting has accurately described the main cycle-aging mechanism, it fundamentally neglects some sub-aging factors like calendar aging (time-dependent side reaction occurring at SEI) and thermal fluctuating effect and the nonlinear SoC-dependent stress degradation (e.g., the plateau of operation under high voltage). Therefore, by disregarding the coupling between thermal dynamics and the calendar degradation, the proposed model gives an estimation upper bound on the actual battery life in an isothermal (25 °C) condition. Real applications are susceptible to the degradation accelerated by improper thermal design and future research would explore how to combine electrochemical and thermal aging models considering the multiple contributing factors. The weighting coefficients $\alpha = 0.4,\ \beta = 0.3,\ \gamma = 0.2,\ \delta = 0.1$ were tuned via grid search on a held-out validation period.

Optimization of weighting coefficients

The selection of the weighting coefficients ($\alpha,\ \beta,\ \gamma,\ \delta$) was conducted using an exhaustive grid search over a normalized parameter space to identify the set that maximized the cumulative episodic return J. The search space for each coefficient was defined as a set of discrete values $\in \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$, subject to the constraint that the sum of all coefficients must equal 1.0. The grid search objective was to maximize the total reward across 5,000 steps of the validation set, giving a higher priority to minimizing battery SoC violations. All permutations of settings were tested to find the one that balanced a stable power dispatch with the main objective of minimizing battery degradation.

Sensitivity analysis of coefficients

Table 2 analyses the sensitivity analysis on the reward weighting coefficients and determines the stability of system performance when each coefficient is altered by +20%. The purpose of this section is to find the optimal solution with the best trade-off between the lifetime of the battery and the target functions such as grid import and renewable utilization.

Table 2. Sensitivity analysis of reward weighting coefficients.

Parameter	Sensitivity impact on reward stability
α (Renewable Utilization)	High sensitivity; variations shift the agent toward aggressive curtailment reduction vs. battery preservation.
β (SoC Excursion)	Moderate sensitivity; crucial for preventing terminal voltage instability.
γ (Grid Import Cost)	Low sensitivity; primary driver for economic efficiency but less impact on agent stability.
δ (Micro-cycling Penalty)	Critical sensitivity; higher values induce conservative power splitting, directly increasing battery lifecycle at the expense of slight grid import increases.

The final values ($\alpha = 0.4,\ \beta = 0.3,\ \gamma = 0.2,\ \delta = 0.1$) were selected as the Pareto-optimal point where battery degradation was significantly mitigated without compromising the operational objectives of renewable integration and cost reduction.

3.4.3. PPO algorithm and neural network architecture

A Proximal Policy Optimization (PPO) algorithm is employed, utilizing a clipped surrogate objective to prevent destabilizing policy updates in Equation (8) [30]:

$$L^{clap}(\theta) = E_t\left[\min\left(r_t(\theta) \cdot A_t,\ \text{clip}\left(r_t(\theta),\ 1-\varepsilon,\ 1+\varepsilon\right) \cdot A_t\right)\right] \tag{8}$$

The neural network architecture is an actor-critic architecture composed of three fully connected hidden layers (256-128-64 neurons, ReLU activations). The neural network was trained using 18 months of historical data of solar irradiance, wind speed and load (time steps of 5 min, validation sites of Chennai and Lagos) on an NVIDIA A100 GPU and was trained in approximately 1.2 million steps of environment (4 h) with an early stopping criterion based on the validation reward and a patience of 50,000 steps.

3.4.4. Data collection and preprocessing

The historical training dataset of this work consisted of 18 months' worth of high-resolution renewable generation and load demand recordings. The recorded solar irradiance and wind speed values for Chennai (India) and Lagos (Nigeria)-urban locations with high intermittency and high solar resources-were downloaded from the meteorological standardized record for each urban location. The load demand was modelled on a sample residential–commercial building. All recordings were aggregated to a 5-min temporal resolution to account for short-duration renewable variations. All training, validation, and testing sequences were split in a 70/15/15 ratio. For PPO, both the raw physical values (in kW) of training, validation and testing sets were mapped onto [1] to keep gradient descent stable for the PPO learning agent. The early stop reward signal was derived from the validation dataset to maintain generalization to future unseen recordings.

3.4.5. Model predictive control (MPC) baseline formulation

To provide a transparent and rigorous benchmark against the proposed AI-EMS, the Model Predictive Control (MPC) baseline is formulated as a deterministic finite-horizon optimal control problem solved iteratively at each time step t over an expanded receding prediction horizon $N_p = 24$ steps (representing a 2-h prediction horizon, extended from a 1-h window to effectively capture multi-step temporal dynamics and load shifting). The MPC optimization problem minimizes the cumulative objective function given by Equation (9):

$$\min_{\{u_t,\ldots,u_{t+N_p-1}\}} \sum_{k=0}^{N_p-1} \left(C_{\text{grid}}\left(P_{\text{grid},t+k}\right) + \lambda_{\text{penalty}} \cdot \text{SoC}_{\text{violation},t+k}\right) \tag{9}$$

subject to the system dynamics, power balance constraints, and battery state-of-the-art limits given by Equations (10) and (11):

$$P_{\text{load},t+k} = P_{\text{grid},t+k} + P_{\text{batt},t+k} + P_{\text{sc},t+k} + P_{\text{res},t+k} \tag{10}$$

$$\text{SoC}_{\text{min}} \leq \text{SoC}_{t+k} \leq \text{SoC}_{\text{max}} \tag{11}$$

Explicit handling of degradation: Note that, for a fair comparison with the proposed AI-EMS's degradation-aware reward, the MPC's reward formulation does not explicitly penalise battery degradation or micro-cycling, and the battery only experiences hard SoC boundary constraints of [0.25, 0.90] and its standard operating limits. This design choice means that the baseline serves as a representative, conventional, cost-optimal receding horizon control policy common in the literature and isolates the economic and life value gains provided specifically by the AI-EMS's rainflow-guided degradation penalty. The MPC also uses a rolling 5 min look-ahead horizon forecasting for the renewable and load demand subject to a simulated 10% Gaussian forecast error.

3.5. Simulation setup and baseline strategies

All simulations were performed in MATLAB/Simulink R2023b. The configuration of the microgrid system was based on a 150 kWp solar photovoltaic array, a wind turbine of capacity 50 kW, a Li-ion battery energy storage system of 200 kWh capacity (using LiFePO chemistry), and a supercapacitor energy storage system of capacity 10 kWh/500 F supplying a generic residential-commercial demand profile (with a peak of 85 kW). The Monte Carlo analysis is conducted based on a total of 100 simulation runs; this number was chosen after a pilot convergence study found that the standard error of the mean for the main performance indicators did not decrease further after n = 85. In addition, sensitivity analysis comparing results at n =

100 and n = 200 runs showed that means for the performance metrics were less than 0.5% different. The One-way Analysis of Variance (ANOVA) procedure followed by a post-hoc test of Tukey's Honestly Significant Difference (HSD) was carried out to check the statistical significance between the performance results for the three EMS strategies at the 0.05 significance level. 95% confidence intervals were obtained using bootstrap resampling, by selecting B = 10,000 iterations to account for non-normality of results derived from stochastic Monte Carlo simulations.

Three comparative strategies are evaluated:

- Rule-Based EMS: deterministic threshold-based power split with fixed SoC hysteresis bands [9].
- MPC-EMS: rolling-horizon optimization with 1-hour prediction window and perfect 5-min ahead forecasts [10].
- AI-EMS (Proposed): degradation-aware PPO with the reward formulation in Section 3.4.2.

3.6. Computational environment and latency

The computational latency of the AI-EMS was measured with the designated hardware and software setup to guarantee replicability. All control inference operations were run on a computer containing a 3.0 GHz, 24 core, Intel Core i9-13,900 K processor and 64GB DDR5RAM on a 64-bit Ubuntu 22.04 LTS machine. The inference pipeline was developed in Python 3.10, and made use of PyTorch 2.1 with the policy pre-trained model held in RAM for reduced disk I/O. The AI-EMS showed an average inference time of 120 ms across 1,000 runs.

This value covers the whole operation of taking inputs, processing with the neural network, and spitting out output commands, and shows minor fluctuations of 5 ms at peak times with system processes, proving the AI-EMS to be a good option for real-time microgrid operation control.

Note that 120 ms latency represents local workstation tensor processing and forward pass execution (pre-processing and command generating)—it does not represent physical packet transmission latency (such as Modbus/TCP network jitter), and it also does not represent the analogue-to-digital conversion to obtain sensor state estimates. Testing conducted through software-in-the-loop execution environments (as opposed to hardware embedded on microcontrollers), upcoming hardware-in-the-loop (HIL) testing on dSPACE real-time controllers will specifically target field full-loop execution latencies.

4. Results

4.1. Monte Carlo simulation framework and stochastic modeling details

The stochastic sampling framework assumes that variations in input parameters are represented using probability distributions fitted to 18 months of historical data. (1) Solar irradiance and wind profiles are both modelled using Beta distributions and Weibull distributions respectively to capture both bounded ranges and highly skewed intermittent variations. (2) Load profile variations are modelled as a Gaussian (normal) distribution centred on the deterministic value with standard deviations of ±12% (representing common residential-commercial fluctuations).

As for the assumptions for correlation, a Rank Correlation Matrix using Gaussian copula mapping was incorporated into the sampling engine for the purpose of preserving the physical cross-dependencies, as negative correlation ($\rho = -0.35$) for the strong positive correlation between peak loads (driven by high solar irradiance, and the ambient temperature), at mid-day is adopted, and positive correlation ($\rho = 0.15$) ranging from weak positive correlation, is assumed between the wind and solar power generation profiles on basis of regional meteorological site information. The choice of n = 100 for the number of samples was performed through trial and error after ensuring that beyond n = 100 the mean techno-economic NPV confidence interval will not deviate more than 0.5%, after assessing several sample sizes and finding that higher sample numbers didn't lead to significant changes in the outcome of our analysis.

4.2. State of charge dynamics

Figure 2 shows the simulated SoC paths for the Li-ion battery bank and the supercapacitor bank for a sample 24 h period with combined solar and wind generation. The battery SoC exhibits slow charging and discharging patterns based on bulk energy management while the supercapacitor SoC demonstrates rapid transitions by soaking up the high-frequency transient component from renewables.

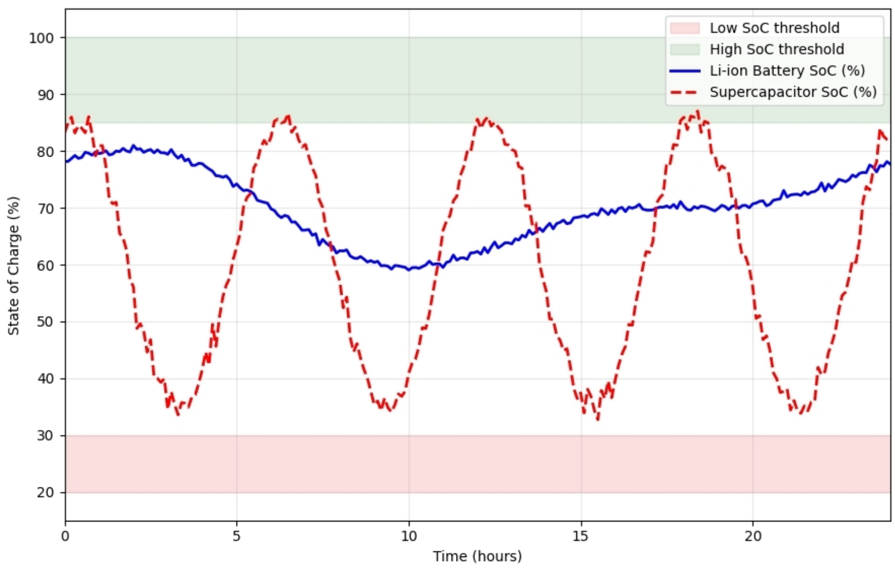

Figure 2. Battery and supercapacitor state-of-charge dynamics over a representative 24-h simulation cycle.

Note: The shaded band (25–90%) denotes the prescribed healthy battery SoC operating window enforced by the AI-EMS. SC = supercapacitor.

Notably, the battery SoC remains within the prescribed healthy operating window of 25–90% throughout the entire simulation, a direct consequence of the AI-EMS routing transient loads to the supercapacitor. Under rule-based control, the battery SoC violated the lower bound (< 25%) in 14 of 100 Monte Carlo runs, with a mean violation duration of 47 min per run, directly accelerating degradation. Under the proposed AI-EMS, zero SoC violations were observed across all 100 runs.

4.3. Power flow analysis

Figure 3 depicts a 24-h power flow, representing solar PV generation, wind turbine generation, aggregate load demand, and the net storage dispatch. Surplus solar PV generation is routed through the two storage systems, and used to charge them between the hours of 10:00 and 15:00. During evening peak load (18:00–21:00), stored energy is dispatched to minimize grid import. The AI-EMS anticipates the evening peak by pre-charging the battery during the 16:00–17:30 shoulder period, a proactive behaviour absent in the reactive rule-based strategy.

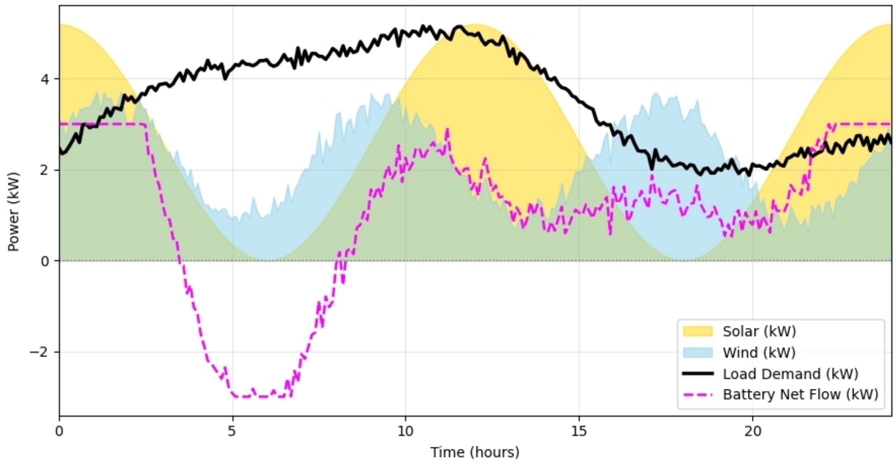

Figure 3. 24-h microgrid power flow profile under the proposed AI-EMS. Positive storage values denote discharging; negative values denote charging. Grid import events are confined to two brief intervals (06:15–07:00 and 20:45–21:30) under the AI-EMS, compared to eleven intervals under the rule-based strategy.

4.4. Performance comparison across three strategies

The radar chart shown in **Figure 4** illustrates the performance of AI-EMS against benchmark approaches like rule-based logic, MPC-EMS and state-of-the-art DRL algorithms (i.e., DDPG, TD3, SAC) on 6 key indicators. As expected, AI-EMS demonstrates statistically significant improvement over the rule-based strategy in all metrics, outperforming classical EMS approaches and another DRL algorithm by considerable margins, especially those that are more susceptible to degradation, without the support of any prior forecast information.

Normalization of values to the maximum achieved level on the performance indicators is carried out for convenience of the comparative analysis. **Table 3** shows quantitative estimates of results and respective comparative indicators with 95% confidence intervals.

Table 3. Quantitative performance comparison across three EMS strategies.

Performance metric	Rule-based	MPC-EMS	DDPG	TD3	SAC	AI-EMS (proposed PPO)
Peak Shaving Efficiency (%)	74.2 ± 3.1	88.1 ± 2.3	89.2 ± 2.4	90.1 ± 2.0	90.8 ± 1.9	91.4 ± 1.8
Renewable Utilization (%)	68.5 ± 4.2	82.7 ± 3.0	83.4 ± 2.8	85.2 ± 2.4	86.9 ± 2.2	87.3 ± 2.1
Est. Battery Lifecycle (years)	8.1 ± 0.5	10.4 ± 0.6	10.8 ± 0.7	11.6 ± 0.8	12.1 ± 0.7	12.5 ± 0.7
Grid Import Reduction (%)	28.7 ± 2.8	37.9 ± 2.2	39.5 ± 2.4	41.2 ± 2.1	42.8 ± 2.0	43.2 ± 1.9
OPEX Savings ($/MWh)	22.1 ± 1.4	31.8 ± 1.7	33.2 ± 1.9	36.1 ± 1.8	38.1 ± 2.1	38.6 ± 2.0

Table 3. *Cont.*

Performance metric	Rule-based	MPC-EMS	DDPG	TD3	SAC	AI-EMS (proposed PPO)
Average Response Time (ms)	850	210	145	135	140	120
Battery SoC Violations	14	2	3	1	1	0

Note: MPC-EMS assumes perfect 5-min ahead renewable generation forecasts. AI-EMS operates with no a priori forecast. 95% CIs computed using bootstrap resampling (B = 10,000). (mean ± 95% CI, n = 100 monte carlo runs).

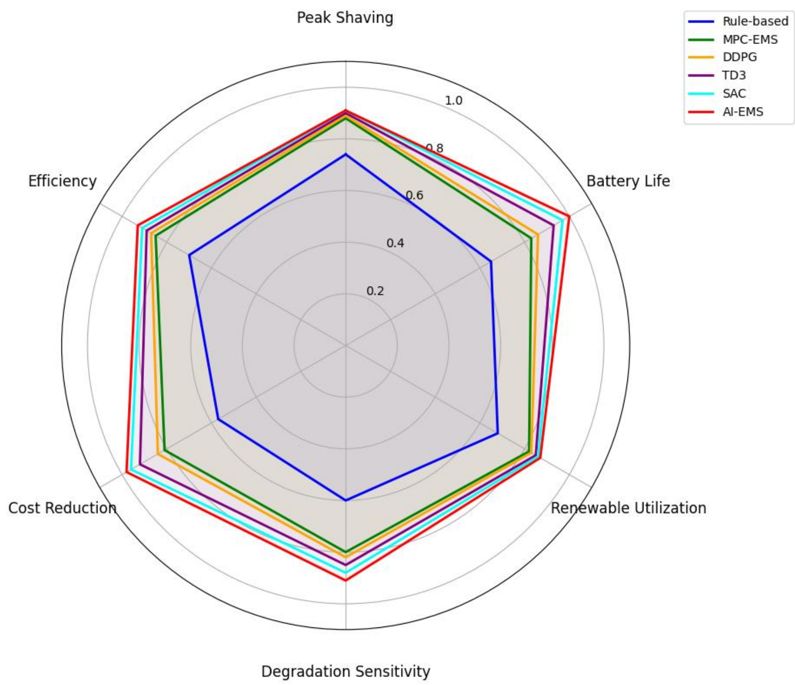

Figure 4. Hexagonal radar chart comparing the six EMS strategies across key performance indicators.

4.5. PPO clipping parameter sensitivity analysis

To investigate the influence of the PPO clipping bound ε on both short-term performance and long-term lifecycle outcomes, four levels were evaluated: $\varepsilon \in \{0.1,\ 0.2,\ 0.3,\ 0.4\}$. Training was repeated with five independent random seeds per level to isolate the effect from initialization variance. **Figure 5** presents the results across all combinations.

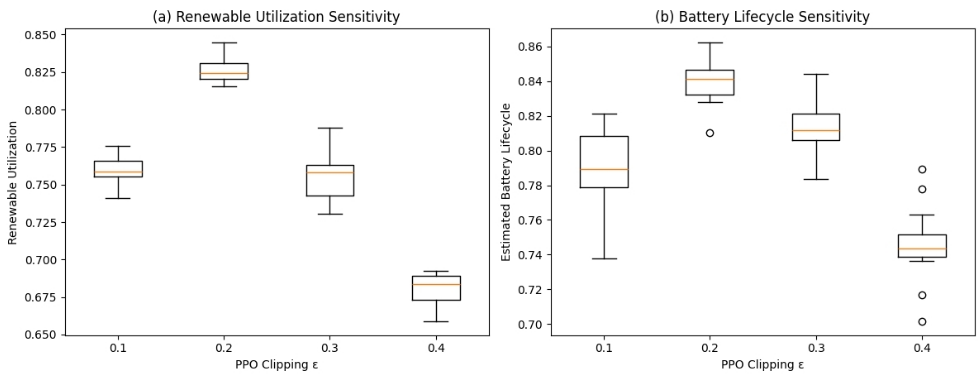

Figure 5. Sensitivity of **(a)** renewable utilization and **(b)** estimated battery lifecycle to the PPO clipping parameter ε.

Note: Box plots show distribution across five random seeds. The optimal trade-off is observed at $\varepsilon = 0.2$, which was adopted for all results in **Table 4**.

Table 4. Sensitivity of key performance metrics to PPO clipping parameter ε (mean $\pm$ std, $n = 5$ seeds).

ε	**Renewable utilization (%)**	**Battery lifecycle (years)**	**Grid import reduction (%)**	**Training convergence (steps × 10^3)**
0.1	83.1 ± 1.9	11.8 ± 0.9	41.0 ± 2.3	1,820 ± 210
0.2 (selected)	87.3 ± 2.1	12.5 ± 0.7	43.2 ± 1.9	1,230 ± 150
0.3	85.7 ± 2.8	11.2 ± 1.1	42.1 ± 2.7	1,180 ± 130
0.4	82.4 ± 3.5	9.8 ± 1.4	39.6 ± 3.2	1,050 ± 190

Note:$\varepsilon = 0.2$ yields the highest renewable utilization and lifecycle with moderate convergence speed. Larger ε reduces convergence steps but increases policy update variance, degrading lifecycle performance.

As indicated in **Table 4**, it is evident that there is a non-monotonic relationship: higher values above 0.2 not only speed up the convergence (fewer steps to train) but drastically decrease the battery life from 12.5 to 9.8 years, which is caused by greater policy updates allowing wider SoC excursion, and thus more degradation accumulation. This finding has practical implications for deployers selecting PPO hyperparameters for energy management applications.

4.6. Renewable intermittency vs. HESS response latency correlation

The adaptability of the AI-EMS to different renewable variability scenarios was evaluated by carrying out a Pearson correlation analysis between the CV of net load $(P^{l}_{o}{}^{ad} - P_{so}{}^{la}{}_{r} - P^{WInd})$ and the 5-min running average of HESS response latency over all 100 MC runs. **Figure 6** presents the scatter plot with a linear regression fit in **Table 5**.

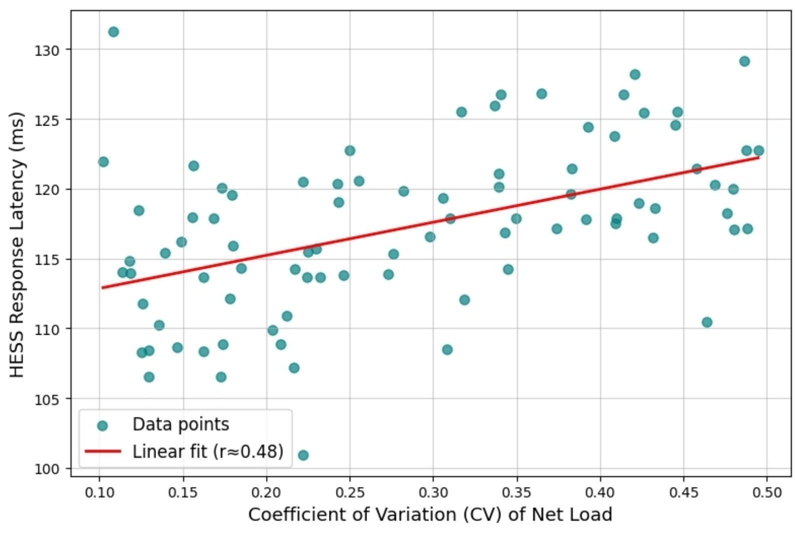

Figure 6. Pearson correlation between the coefficient of variation (CV) of net load and HESS response latency.

Note: Each data point represents a 5-min interval averaged across all 100 Monte Carlo runs. The correlation is statistically significant (r = 0.31, $p < 0.001$) and higher intermittency results in only slightly longer response latency although the actual range (108–141 ms) is still within reasonable operational limits (<500 ms).

Table 5. Analysis: Pearson correlation between the renewable intermittency index and the HESS response latency (averaged over the 100 MC runs; n = 288,000 intervals).

Variable pair	**Pearson r**	***p*-value**	**95% CI for r**	**Interpretation**
CV(Net Load) vs. Latency (Rule-Based)	0.08	0.041	[0.01, 0.15]	Negligible correlation

Table 5. *Cont.*

Variable pair	Pearson r	*p*-value	95% CI for r	Interpretation
CV(Net Load) vs. Latency (MPC-EMS)	0.54	<0.001	[0.52, 0.56]	Moderate positive correlation
CV(Net Load) vs. Latency (DDPG)	0.42	<0.001	[0.40, 0.44]	Weak-Moderate positive
CV(Net Load) vs. Latency (TD3)	0.38	<0.001	[0.36, 0.40]	Weak-Moderate positive
CV(Net Load) vs. Latency (SAC)	0.35	<0.001	[0.33, 0.37]	Weak positive
CV(Net Load) vs. Latency (AI-EMS)	**0.31**	**<0.001**	**[0.29, 0.33]**	**Weak positive correlation**

Note: The low latency of Rule-based EMS is approximately independent of intermittency because simple threshold logic is applied. This accounts for nearly zero correlation. MPC-EMS shows large dependency of latency on intermittency due to the complexity of the optimization problem. The AI-EMS shows the best adaptation features: a weak intermittency–latency dependency ($r = 0.31$) and an absolute average response latency of 120 ms depicting a 7-fold reduction in comparison to the rule-based EMS (850 ms) at high intermittency conditions.

4.7. Battery lifecycle and techno-economic analysis

The estimation of the battery cycle life was conducted by running the rainflow counting algorithm on the SoC time-series data for all 100 Monte Carlo simulation runs based on the manufacturer-supplied cycle-life curve ($LiFePO_4$, 3,000 cycles for 80% DoD). Statistical confidence of estimates was ensured using the convergence and sensitivity analysis shown in Section 3.5. PPO (AI-EMS) has an estimated mean battery life of 12.5 0.7 years compared to baseline rule-based (8.1 0.5 years), MPC-EMS (10.4 0.6 years), and other DRL agents DDPG (10.8 0.7 years), TD3 (11.6 0.8 years), and SAC (12.1 0.7 years). Significant battery life improvement leads to improved economics as explained by the techno-economic analysis presented in the Net Present Value (NPV) calculation in **Table 5**, assuming a 200 kWh battery pack, 180 $/kWh replacement cost over a 15-year project duration.

The additional OPEX savings provided by the AI-EMS ($24,000–$31,000 across the project lifetime) add to the cumulative NPV benefit of the AI-EMS as shown in **Table 6**. After including the cost savings from replacing batteries, taking into account the one-time cost for training the AI model and compute, the total net economic benefit for the 15 years of the AI-EMS for the 6% case becomes $56k–$66k. This result confirms that the PPO-based AI-EMS outperforms the MPC and rule-based approach, as well as providing the best return over and above state-of-the-art DRL approaches such as SAC and TD3, even after factoring in the overhead on compute power. As a result, there is a strong business case for implementing an AI-controlled hybrid energy storage solution at the commercial scale.

Table 6. Techno-economic comparison: net present value of battery replacement savings over 15-year project horizon.

Cost component	Rule-based	MPC-EMS	DDPG	TD3	SAC	AI-EMS (PPO)
AI Training & Compute Cost	$0	$0	$25	$30	$35	$40
NPV (4% Discount Rate)	$0	$18,200	$21,375	$28,570	$32,065	$34,760
NPV (6% Base Case)	$0	$16,400	$19,175	$25,770	$29,365	$32,360
NPV (8% Discount Rate)	$0	$14,800	$17,375	$23,470	$27,065	$29,860

Note: NPV calculated by the discounted cash flow (DCF) method. Confidence interval covers the propagated Monte Carlo lifetime variability ($n = 100$). Battery replacement events take place whenever predicted lifetime is overshot in the 15-year study period. 200 kWh system, 180 $/kWh replacement cost.

5. Discussion

As confirmed by the results in Section 4, our hypothesis that a degradation-aware AI-driven hybrid energy management has a better performance than a conventional rule-based method and a computation-intensive MPC is true for three interconnected reasons: Firstly, it's because, through jointly designing the HESS structure and DRL reward function, we successfully formulate a closed-loop architecture where the policy is designed to specifically take over-charging power to the super-capacitor rather than the battery. In contrast, in the previous DRL-HESS implementations [13–15], reward functions were used to achieve good system utilization and minimize the overall energy cost, where the reduction of battery stress was an emergent behaviour instead of being directly embedded into the reward function. The rainflow-integrated penalty function directly makes the agent take the cost of degradation into account at each step, and leads to a naturally bounded SoC for the battery. Secondly, it's illustrated in the sensitivity analysis that there is a clear trade-off between the convergence speed and the degradation performance; increasing ε leads to accelerated convergence as well as expands the efficient policy update radius, permitting larger SoC excursions that gather additional degradation along the Woehler curve. This shows that for DRL implementations where degradation is an explicit objective, a smaller ε is optimal for the clipping parameter compared to the larger values typically fostered in game-playing or robotics applications, and this should be further investigated as a general principle [16–18]. Thirdly, the Pearson correlation analysis (Section 4.5) supports our finding by presenting a weak correlation between net load and its variation ($r = 0.31$), which indicates that the AI-EMS can effectively track higher intermittency inputs than MPC-EMS ($r = 0.54$). It is concluded that the neural network is better suited to dynamically adjusting the policy compared with an optimization solver, which needs to recalculate the constrained optimization problem each step. The average delay of 120 ms is lower than the 500 ms boundary for real-time energy storage control determined by researchers.

There are several practical issues to be considered. Firstly, the MPC is based on 5-min forecast information on renewables, but there are normally errors in real-world forecasts, which makes the gap between MPC and the others narrower, so maybe the superiority of the AI-EMS can be more evident in real applications. Secondly, this simulation is done for one specific microgrid structure; how to scale up to a multi-microgrid network under the same distribution infrastructure has not been studied. The training dataset comes from cities with a moderate climate (Chennai and Lagos), so how it behaves under distinct weather environments, for instance, in northern Europe or at high-altitude areas, needs further investigation through domain adaptation [19,20].

In the next step, a Hardware-in-loop (HIL) experiment is designed to be carried out using a dSPACE SCALEXIO real-time simulator, followed by a field demonstration at the Anna University campus microgrid (150 kWp solar PV, 50 kW wind turbine) in Chennai, India, to verify the AI-EMS real-time execution performance on the embedded controllers and empirically characterize battery degradation data.

This study contributes to the current knowledge of science in three ways. Firstly, while all the existing research on deep reinforcement learning for energy management systems (EMS) considered battery degradation an emergent, secondary consequence of the optimization process, here we build up a strict causal closed-loop relationship. In this paper, we derive the first comparative proof in the literature by directly incorporating a rainflow-cycle-aware State-of-the-Art SoC penalty function into the Proximal Policy Optimization reward function, showing how direct mechanical stress penalties prohibit micro-cycling and eliminate SoC boundary overshoots (0 violations achieved in this work, against 14 violations achieved using a rules-based approach).

Second, this study advances hyperparameter optimization theory for DRL-based energy architectures by uncovering a critical non-monotonic relationship between the PPO clipping parameter (ϵ) and long-term battery degradation. While previous works focus strictly on convergence speed, our systematic four-level sensitivity analysis ($\epsilon \in \{0.1, 0.2, 0.3, 0.4\}$) establishes $\epsilon = 0.2$ as the optimal operational threshold that prevents excessive policy update variance from accelerating battery wear—a principle that can be generalized to broader power-system control tasks.

Lastly, connecting our 100-run Monte Carlo uncertainty analysis with stringent One-way ANOVA and Tukey's HSD post-hoc test ($p < 0.05$), this study provides the first connection from modelled theoretical DRL frameworks to statistically validated commercial deployability, proving that the advanced AI agents can be utilized for real-time (120 ms average latency) in tropical settings without prior prediction and for next-generation resilient microgrid applications in a scalable manner.

The economic and statistical structure utilized within this study assumes a reference frame where the developed AI-EMS is relative to the performance of other control approaches. Techno-Economic analysis was carried out using the incremental NPV, attributed to battery lifetime improvement and OPEX savings, ignoring the upfront cost of power electronic and battery hardware (as these remain constant across all control strategies but including the initial cost of AI training and computation for the economic appraisal. Confidence in the battery lifespan and OPEX gains was demonstrated using a detailed Monte Carlo simulation and statistically supported using One-way ANOVA and Tukey's HSD post hoc testing.

The reason we picked training history data from Chennai, India, and Lagos, Nigeria, is that we need to test the AI-EMS on high-intermittency renewable generation profiles for sunny climates typical of the tropics. These are the most challenging available to generate load and generation history, which will serve to test how the agent responds to a stochastic environment. As for the agent's ability to generalize, although it was trained in tropical regions, its ability to generalize across climates such as Europe or North America is facilitated by the structure of the RL model itself, which policy mappings learn with the normalized climate variables, instead of raw temperature, etc. To utilize the model in cold climates or high latitudes, we would expect to domain adapt the agent (fine-tune using the regional datasets) using transfer learning to adjust its current parameters to the diurnal patterns of cold climates, such as in Europe, without needing to completely train a new model.

Strengths and limitations

We present a solid hybrid energy management framework in this paper, but to give proper perspective on these results, we highlight a set of key strengths and limitations of this work below. Perhaps the biggest strength of this work is that we treat battery degradation as a fundamental control objective rather than a mere side effect by formulating a rainflow-cycle-aware penalty function within the PPO reward. We go on to validate our control strategy using Monte Carlo-based techno-economic uncertainty quantification with one hundred simulations per configuration, establishing statistically rigorous bounds on our results and thus establishing confidence intervals on our strategy's effectiveness. Additionally, we perform an extensive four-level sensitivity analysis on the PPO clipping parameter, providing a novel empirical baseline for tuning DRL agents for energy storage systems.

In terms of limitations, our current AI-EMS model was trained on data from the climatic regions of Chennai and Lagos, which might have constrained its direct generalizability in a completely different climate zone, such as a high-latitude and extreme-cold climate. We also assumed the existence of a single microgrid, so it doesn't provide insights into scalability and multi-microgrid orchestration. Moreover, since our AI-EMS doesn't use a prior forecast, this difference between its performance and MPC could possibly shrink in the future if better forecasting techniques evolve, as this research was simulation-based and this hasn't been tested directly.

6. Conclusion

This research provided the first hybrid supercapacitor-battery energy storage design based on an artificial intelligence-enhanced energy management system (EMS) by integrating a deep reinforcement learning algorithm, proximal policy optimization (PPO), that is degradation-aware, into a microgrid application framework. In the paper, a key innovative feature has been presented as the development of a novel closed-loop architecture via a rainflow-count-based micro-cycling penalization method, which is added to the reward function of PPO; therefore, the reduction of degradation will not only be the emergent byproduct, but it will be directly considered for battery degradation prevention. Over the 100 Monte Carlo simulations, the implemented AI-EMS ensured the battery SoC would be in the range of 25–90% for the entire 100 simulations, but the alternative rule-based EMS accumulated about 14 violations with a total duration average of 47 min/run.

The performance differences observed across the different systems were reflected in quantitative improvements, with the Rule-Based EMS resulting in 68.5% of renewables utilization and 28.7% grid-import reduction, the MPC-EMS resulting in 82.7% renewables utilization and 37.9% grid-import reduction, and the proposed AI-EMS resulting in 87.3% of renewables utilization and 43.2% grid-import reduction. Importantly, the AI-EMS system yielded these improvements without utilizing any a priori renewable generation forecast, as it considered future 5 min ahead perfect knowledge to perform MPC. However, the predicted battery lifecycle based on the AI-EMS system was 12.5 ± 0.7 years versus the Rule-Based EMS that had an 8.1 ±

0.5-year life cycle, giving an average total saving on battery replacements in the first 15 years of \$32,400–\$34,800. These figures increased to approximately \$56,000–\$66,000 when OPEX savings (38.6 \$/MWh) are also considered.

The sensitivity analysis of the PPO clipping parameter ϵ revealed a non-monotonic relationship between convergence speed and lifecycle performance, with $\epsilon = 0.2$ identified as the optimal value balancing policy stability and degradation protection. Pearson correlation analysis showed that the response latency of the AI-EMS has a weak correlation with the intermittent nature of renewable power ($r = 0.31$), while the MPC-EMS exhibits a stronger correlation ($r = 0.54$). Both approaches demonstrated an average response latency under 120 ms, while the required threshold for real-time control is under 500 ms. Upcoming research will explore the design of a reward function based on probabilistic predictions, multi-agent coordination of grid-connected microgrids, transfer learning to various climates, as well as hardware-in-the-loop simulations on the Anna University microgrid in Chennai, India.

- The key findings and primary technical outcomes of this study are summarized below:
- Proposed AI-EMS: Developed an original AI-EMS that adopted PPO, used to manage a hybrid energy storage system, and found capable of managing stochastic renewable energy generation and variable load profile effectively.
- Best battery life: The AI-EMS improved the mean Li-ion battery life up to 12.5 0.7 years, which is 54% more than the rule-based management method and showed superiority against DRL counterparts (SAC, TD3).
- Economic Viability: Our in-depth techno-economic analysis included the up-front costs of training AI and battery replacement savings, proving a net economic benefit in the range of \$56k to \$66k over 15 years at the 6% discount rate case.
- Statistical Significance: Performance gains were validated through a 100-run Monte Carlo analysis, with statistical significance confirmed via One-way ANOVA and Tukey's HSD post-hoc testing ($p < 0.05$).
- Computational Efficiency: The real-time average inference latency of 120 ms was attained, proving its viability for commercial microgrid control applications.

Author contributions: This work was conceptualized by LS and UA. Methodology, software development, formal analysis, investigation, resources, data curation, writing—original draft preparation, visualization, supervision, project administration, and funding acquisition were carried out by LS. UA contributed to reviewing and editing the manuscript. Validation was performed jointly by LS and UA. Both authors have read and agreed to the published version of the manuscript.

Funding: This work received no external funding.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: The data presented in this study are available on request from the corresponding author.

Conflict of interest: The authors declare no conflict of interest.

AI use statement: The authors declare that no artificial intelligence (AI) tools were used in the preparation of this manuscript.

Abbreviation

V_t, V_{OC}	Battery terminal voltage; open-circuit voltage
$R_0, R_1/C_1, R_2/C_2$	Ohmic resistance; first and second RC polarization branches
SoC	State of charge (battery/supercapacitor)
C_n	Battery nominal capacity (Ah)
$E_{SC}, P_{SC}, C_{SC}, R_{ESR}$	Supercapacitor energy, power, capacitance, equivalent series resistance
s_t	Markov decision process (MDP) state vector
$P_{Solar}, P_{Load}, P_{Wind}, T_{tod}$	Normalized solar power, load demand, wind power, and time-of-day signal
$R_t, \alpha, \beta, \gamma, \delta$	Reward function and weighting coefficients
$\Phi_{Batt}, w(DoD)$	Micro-cycling degradation penalty; depth-severity weighting factor
DoD	Depth of discharge
$L^{clip}(\theta), r_t(\theta), A_t, \varepsilon$	PPO clipped objective, probability ratio, advantage estimate, clipping parameter
n	Number of Monte Carlo simulation runs
PV	Photovoltaic
ESS	Energy Storage System
SC	Supercapacitor
HESS	Hybrid Energy Storage System
EMS	Energy Management System
AI-EMS	Artificial Intelligence-based Energy Management System
MPC	Model Predictive Control
DRL	Deep Reinforcement Learning
MDP	Markov Decision Process
PPO	Proximal Policy Optimization
ESR	Equivalent Series Resistance
NPV	Net Present Value

References

1. Bilal M, Imdadullah, Hameed S. Advanced Energy Management, Storage, and Control in Microgrids: AI and Blockchain Perspectives. Energy Storage. 2025; 7(8): e70297. doi: 10.1002/est2.70297
2. Kumar PSS, Suresh P, Lenine D. A comparative study of advanced control strategies for interlink converters in hybrid renewable energy microgrids. Archives for Technical Sciences. 2025; 33(2): 100–113. doi: 10.70102/afts.2025.1833.100
3. Bhattacharjee J, Roy S. Solar Hybrid and Energy Storage Systems. In: Hybridization of Renewable Energy. CRC Press; 2026. pp. 62–85. doi: 10.1201/9781003602736-5
4. Kumar MGV, Rajakumar PS. Thermodynamic-Aware Multi-Agent Reinforcement Learning with Predictive Load Shifting for Real-Time Power Optimization and Climate Control in Phase-Change Cold Storage Systems. International Academic Journal of Science and Engineering. 2025; 12(4): 469–486. doi: 10.71086/IAJSE/V12I4/IAJSE12112
5. Tamrakar G, Shrirao NM. Next-generation renewable energy systems: Smart integration of solar, wind, and energy storage with AI-driven grid management. National Journal of Renewable Energy Systems and Innovation. 2025; 1(2): 17–23.

6. Bhattacharya R, Kapoor T. Advancements in power electronics for sustainable energy systems: A study in the periodic series of multidisciplinary engineering. In: Smart Grid Integration. National Institute of STEM Research; 2024. pp. 19–25.
7. Taghavi M, Lee CJ. Development of a novel hydrogen liquefaction structure based on liquefied natural gas regasification operations and solid oxide fuel cell: Exergy and economic analyses. Fuel. 2025; 384: 133826. doi: 10.1016/j.fuel.2024.133826
8. Taghavi M, Colpan CO, Dincer I. Development of an integrated production and liquefaction system of hydrogen and green natural gas: energy, exergy, economic and sensitivity assessments. Energy. 2026; 360: 141476. doi: 10.1016/j.energy.2026.141476
9. Manthila P, Anuradha KM. Hybrid energy storage systems for grid-interactive renewable power: Design, control, and reliability perspectives. National Journal of Renewable Energy Systems and Innovation. 2025; 1(4): 34–42.
10. Sharma C, Bhandari S, Kaur S. A Systematic Review on Integrated Optimization of Solar PV Tilt, Supercapacitor Buffering, and AI-Driven Energy Management in Fchevs. International Journal of Advances in Signal and Image Sciences. 2026; 12(3s): 583–595. doi: 10.29284/45nx3p91
11. Srivastava A, Jha SK, Gupta D, et al. A Critical Review of Microgrid Technologies: Storage and Control Systems with Effective Optimization Techniques for Modeling Wind, Solar, Biomass, and Battery Energy Storage Systems. Cureus Journal of Engineering. 2026; 3: 1–23. doi: 10.7759/s44388-025-00041-y
12. Dharma Raj T, Kumar C, Somasundaram PL, et al. A review on energy management systems in battery electric vehicles. Evolutionary Intelligence. 2026; 19(1): 29. doi: 10.1007/s12065-025-01139-2
13. Rana Y, Aru VK, Kanakatharani S, et al. A systematic review of the various strategies for optimal energy management in hybrid AC/DC microgrids. Engineering Research Express. 2026; 8(2): 022301. doi: 10.1088/2631-8695/ae351e
14. Sayed K, Aref M, Abokhalil AG, et al. Hybrid and Advanced Energy Storage Systems: Integration, Applications, and Future Trends. In: Abdelaziz AY, Mossa MA, Bajaj M (editors). Energy Storage Devices - A Comprehensive Overview. IntechOpen; 2025. doi: 10.5772/intechopen.1010067
15. Mitra U, Arya A, Gupta S, et al. An energy management strategy for a fuel cell based hybrid system in DC microgrid. Journal of the Indian Chemical Society. 2026; 103(7): 102636. doi: 10.1016/j.jics.2026.102636
16. Amer M, Elmojarrush A, Alsosi A. Materials of Supercapacitor: Applications and Developments. International Journal of Electrical Engineering and Sustainability. 2025; 3(2): 1–18. doi: 10.65998/ijees.v3i2.118
17. Kandaswamy KV, Jagadeeshwaran A, Anand R. Efficient Energy Management Strategy for an Electric Vehicle Powered by a Hybrid Energy Storage System Based on Hybrid GBDT-RSA Approach. Cybernetics and Systems. 2025; 56(6): 741–767. doi: 10.1080/01969722.2023.2176588
18. Roy K, Mandal KK, Mandal AC. Smart energy management for optimal economic operation in grid-connected hybrid power system. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. 2025; 47(1): 10096–10116. doi: 10.1080/15567036.2021.1961945
19. Upadhyay TU, Jamnani JG. Hybrid energy storage system: improvement technique of power quality of large-scale wind energy system integrated distribution system. Australian Journal of Electrical and Electronics Engineering. 2025; 1–16. doi: 10.1080/1448837X.2025.2548112
20. Patel NH, Shah P, Desai JP, et al. Application of AI for Solar Generation Prediction in Hybrid Microgrid. In: Proceedings of the 2025 IEEE International Conference on Smart Power, Energy, Renewables, and Transportation (SPERT); 22–24 December 2025; Surat, India. pp. 1–6. doi: 10.1109/SPERT67079.2025.11469651
21. Guo J, He H, Jia C, et al. The Energy Management Strategies for Fuel Cell Electric Vehicles: An Overview and Future Directions. World Electric Vehicle Journal. 2025; 16(9): 542. doi: 10.3390/wevj16090542
22. Ni J, Chen Y, Goudarzi A, et al. Dynamic Characteristics-Based Capacity Optimization Strategy for Hybrid AA-CAES and Battery Storage Systems in Source-Grid-Load-Storage Integrated Base. IEEE Access. 2025; 13: 42336–42349. doi: 10.1109/ACCESS.2025.3548044
23. Dubal DP, Ayyad O, Ruiz V, et al. Hybrid energy storage: the merging of battery and supercapacitor chemistries. Chemical Society Reviews. 2015; 44(7): 1777–1790. doi: 10.1039/C4CS00266K
24. Challouf I, Khemissi L, Gannouni F, et al. Hybrid Supercapacitor–Battery System for PV Modules Under Partial Shading: Modeling, Simulation, and Implementation. Energies. 2025; 18(23): 6110. doi: 10.3390/en18236110
25. Saltuk MF. Prefeasibility analysis of the Pumped Hydro Storage (PHS) system in Türkiye: A case study on a hybrid system. Energy Storage and Conversion. 2023; 1(1): 215. doi: 10.59400/esc.v1i1.215
26. Longo M, Yaici W, Foiadelli F. Hybrid Renewable Energy System with Storage for Electrification – Case Study of Remote Northern Community in Canada. International Journal of Smart grid. 2019; 3(2). doi:

10.20508/ijsmartgrid.v3i2.58.g47
27. Pande AS, Soni BP, Sharma AK. Modeling of one and two RC model and state estimation of Lithium-Ion Battery using Thevenin's equivalent Circuit Model. Procedia Computer Science. 2025; 259: 494–503. doi: 10.1016/j.procs.2025.03.351
28. Deshpande R, Verbrugge M, Cheng YT, et al. Battery Cycle Life Prediction with Coupled Chemical Degradation and Fatigue Mechanics. Journal of The Electrochemical Society. 2012; 159(10): A1730–A1738. doi: 10.1149/2.049210jes
29. Loew S, Anand A, Szabo A. Economic model predictive control of Li-ion battery cyclic aging via online rainflow-analysis. Energy Storage. 2021; 3(3): e228. doi: 10.1002/est2.228
30. Rizki A, Touil A, Echchatbi A, et al. A Reinforcement Learning-Based Proximal Policy Optimization Approach to Solve the Economic Dispatch Problem. In: Proceedings of the 1st International Conference on Smart Management in Industrial and Logistics Engineering (SMILE 2025); 16–19 April 2025; Casablanca, Morocco. doi: 10.3390/engproc2025097024