SAC-Attention Framework for Cost, Emission and Battery-Aware Energy Management in Grid-Connected Hybrid Micro-grid

Ved prakash[1], Dr. Saurabh V. Kumar [2]

[1] Department of Electrical Engineering *(Power System)*, UNS Institute of Engineering and Technology Jaunpur U.P

[*Veer Bahadur Singh Purvanchal University, Jaunpur, UP,]* India

[2]Department of Electrical Engineering *(Power System)*, UNS Institute of Engineering and Technology Jaunpur U.P

[*Veer Bahadur Singh Purvanchal University, Jaunpur, UP,]* India

Abstract - *This paper presents a smart energy management system for hybrid Microgrids using an advanced learning technique called SAC-Attention. The main aim is to reduce electricity cost, lower carbon emissions, and improve battery life at the same time.*

The system combines different modern techniques such as reinforcement learning, attention mechanism, forecasting models, and battery health tracking. It is specially designed considering Indian conditions like electricity pricing and carbon tax.

The results show significant improvements- the system reduces cost by about 36%, emissions by around 35%, improves battery life, and learns faster compared to existing methods.

Keywords - Deep reinforcement learning, soft actor-critic, attention mechanism, microgrid energy management, battery state-of-health, multi-objective optimization

I. INTRODUCTION

Nowadays, hybrid microgrids (which use solar panels, wind turbines, batteries, and diesel generators) are becoming very popular, especially in India. However, managing them is not easy because many things keep changing- like sunlight, wind speed, electricity demand, and prices .Earlier methods such as rule-based control or optimization techniques work well in simple situations, but they fail when conditions become uncertain.

Modern techniques like Deep Reinforcement Learning (DRL) perform better, but still have some limitation-

- **i.** It mainly focus only on cost and ignore pollution &battery health
- **ii.** Battery damage is not calculated precisely.
- **iii.** Time-based patterns (like daily solar variation) are not properly used
- **iv.** Uncertainty from multiple sources not handled completely
- **v.** Indian conditions like pricing and carbon tax are not considered

To overcome these issues, this paper proposes a new method called SAC-Attention, which improves decision-making by considering all these factors together

Proposed Work-

The proposed system improves microgrid performance using five main ideas-

1. Attention Mechanism → Helps the system focus on important information like price and battery condition
2. Bi-LSTM Model → Predicts future values (solar, wind, load, price)
3. Battery Health Model → Accurately tracks battery wear and tear
4. Monte Carlo Simulation → Handles uncertainty by testing multiple scenarios
5. India-Specific Model → Includes real electricity prices and carbon tax

II. RELATED WORK

A. Classical EMS Methods

Rule-based controllers and MILP formulations have dominated early EMS literature [3], [4]. While MILP achieves near-optimal solutions for deterministic day-ahead scheduling, computational complexity scales exponentially with binary variables, precluding real-time application [4]. MPC-based approaches [5] offer a receding horizon framework but require accurate predictive system models and suffer from poor scalability under high-dimensional uncertainty

B. Value-Based DRL for Micro grid EMS

Jietal. [7] pioneered DRL-based micro grid EMS using DQN, establishing a widely-cited benchmark. However, DQN's discrete action space necessitates power level Discretization, introducing quantization errors in continuous dispatch settings. Chen et al. [12] proposed Hierarchical DQN (HDQN) to handle mixed discrete-continuous variables, improving dispatch accuracy, yet emission objectives and SOH modeling remain absent.

C. Policy Gradient DRL Methods

Liu et al. [8] demonstrated DDPG's superiority over DQN by 29.59% in operational cost on a grid-connected micro grid, though no emission objective was included. PPO-based approaches [9], [13] achieved stable training and demonstrated partial cost-emission dual objectives; however, SAC's entropy regularization provides superior exploration-exploitation balance for complex multi-objective problems [10]. Proposed PERSAC with prioritized experience replay for cost optimization, confirming SAC's efficiency but excluding emission and SOH objectives.

D. Multi-Objective and Battery-Aware DRL

Hu et al. [15] employed SAC for multi-timescale HESS scheduling with simplified battery degradation modeling. MAPPO with Pareto dominance selection [16] addressed cost-battery switching trade-offs using switching frequency as a degradation proxy—however, Rain flow counting provides 15% more accurate degradation quantification [11]. The Diff Carl framework [17] applied diffusion-modeled RL for carbon-aware optimization, achieving 2.3-6.4% improvements over SAC but omitting SOH modeling.

E. Research Gap Identification

Table I systematically compares the proposed work against eight representative studies, confirming that no existing paper simultaneously addresses all five identified limitations. This comprehensive gap justifies the proposed SAC-Attention framework

TABLE I: COMPARISON WITH EXISTING DRL-BASED MICROGRID EMS WORKS

Ref	Year	Algo	Cost	Emission	SOH	Attention	BiLSTM	4-Src UNC	India
[7]	2019	DQN	✓	✗	✗	✗	✗	✗	✗
[8]	2023	DDPG	✓	✗	✗	✗	✗	Partial	✗
[9]	2025	PPO	✓	Partial	✗	✗	✗	✗	✗
[12]	2024	HDQN	✓	✗	✗	✗	✗	✗	✗
[14]	2023	SAC	✓	✗	✗	✗	✗	Partial	✗
[15]	2023	SAC	✓	✗	Partial	✗	✗	✗	✗
[16]	2025	MAPPO	✓	✗	Partial	✗	✗	✗	✗
[17]	2025	Diff-RL	✓	✓	✗	✗	✗	✗	✗
Prop.	**2026**	**SAC-A**	✓	✓	✓	✓	✓	✓	✓

III. HYBRID MICROGRID SYSTEM MODEL (MATERIALS AND METHODS)

A. System Configuration

The micro grid used in this study includes-

- Solar panels (50 kW)
- Wind turbine (30 kW)
- Battery storage system
- Diesel generator
- Grid connection
- The system works by balancing power which means-

Total power generated = Total power used

It continuously decides-

i. When to charge/discharge battery
ii. When to use grid power
iii. When to run diesel generator

TABLE II: SYSTEM CONFIGURATION (MATERIALS AND METHODS)

COMPONENT	PARAMETER	VALUE
PV Array	Rated Power / Temp. Coffs.	50Kw/−0.0045/°C
Wind Turbine	Rated / Cut-in / Cut-out	30 kW / 3 m/s / 25 m/s
Li-ion BESS	Capacity / Power / η	100 kWh / 50 kW / 95%
BESS SOC Limits	Min/Max	20% / 90%
Diesel Generator	Rated Power	30 kW
Grid Import Limit	Max Power	50 kW
Grid TOU Peak	9:00•23:00	₹12/kWh + ₹2.0/kg CO2
Grid TOU Off-Peak	23:00•9:00	₹6/kWh + ₹2.0/kg CO2
Battery Capital Cost	Replacement Cost	₹6,000/kWh
India Grid EF	Emission Factor	0.82kg CO2/kWh

B.Component Mathematical Models

PV output power as a function of solar irradiance G(t) and cell temperature T(t): $P_{pv}(t) = Pr \times G(t) / G_{ref} \times [1-0.0045(T(t)-25)]$ (1)

Wind turbine output within operational wind speed range [v_{ci}, v_{co}]:

$$P\,w_t(t) = 0.5\rho A Cp\, v(t)^3, \quad v\,c_i \le v(t) \le v\,co \qquad (2)$$

BESS state-of-charge (SOC) dynamics:

$$S\,OC(t+1) = SOC(t) + [\eta_{ch} P_{ch}(t) - P\,dis(t)/\eta\,dis]\Delta t /Ecap \qquad (3)$$

C. Rain flow-Based Battery SOH Model

The battery SOH is modeled using the Rain flow counting algorithm [11], which decomposes the SOC time series into half-cycles characterized by their depth of discharge (DoD). Capacity fade per cycle combines cycling and calendar aging:

$$\Delta SOHcyc = kc \times DoD\hat{}\alpha \times exp(\beta(T - T_{ref})/T_{ref}) \qquad (4)$$

$$\Delta SOH_{cal} = k\{cal\} \times exp(\gamma \cdot SOC(t)) \times \Delta t \qquad (5)$$

$$SOH(t+1) = SOH(t) - \Delta SOH\,cyc - \Delta SOH_{cal} \qquad (6)$$

where kc=2.0×10⁻⁴, α=1.5, β=0.6, k{cal}=1.5×10⁻⁷, γ=0.3, Tref=298 K are empirical parameters. Economic degradation cost per time step: C deg(t) = (CBESS/SOH total loss)×ΔSOH(t), where C BESS = ₹6,000/kWh.

D. Multi-Objective Function with India Carbon Tax

The total operational cost integrates electricity, DG fuel, degradation, and India BEE carbon emission tax:

$$J = \Sigma\, t[C\,grid(t) + C\,dg(t) + C\,deg(t)] \qquad (7)$$

$$C\,grid(t) = \rho e(t) Pimp(t) - \rho_{sell} P_{exp}(t) + \tau c\, E_{grid}(t) \qquad (8)$$

where τc = ₹2.0/kg CO2 is India's BEE carbon tax rate. The power balance constraint must hold at all timesteps:

$$P_{pv} + P_{wt} + P_{dg} + P_{imp} + P_{dis} = P_{lod} + P_{ch} + P_{exp} \qquad (9)$$

IV. PROPOSED SAC-ATTENTION FRAMEWORK

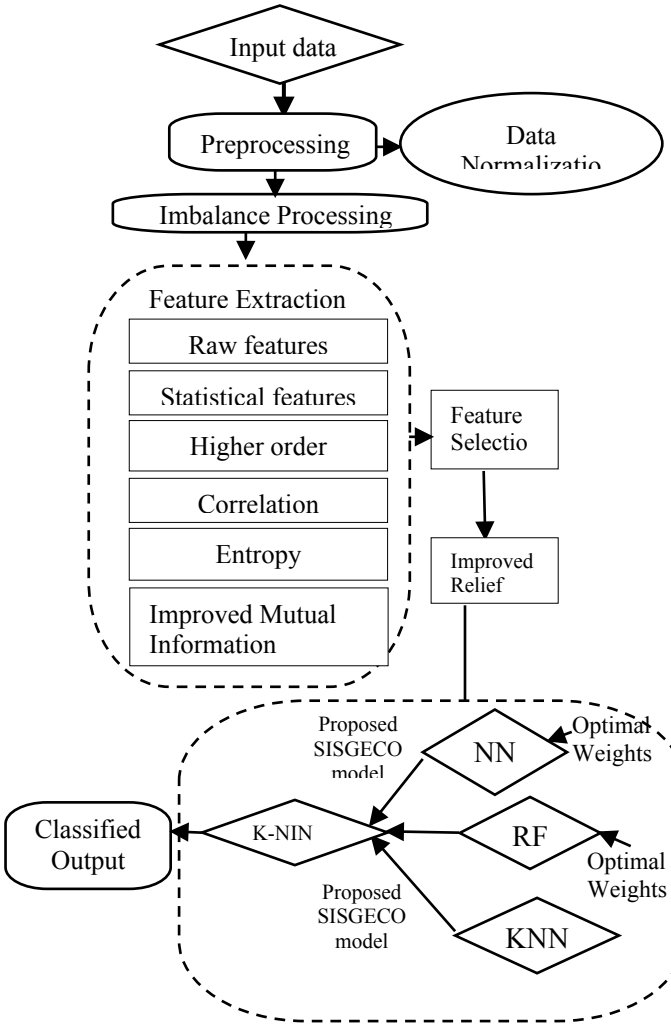


Fig.1. Signal Classification Flowchart for Proposed SAC-Attention Framework

A. MDP Formulation

The micro grid EMS is formulated as a Markov Decision Process (S, A, P, R, γ):

State Space (12-dimensional):

$$s(t) = [SOC, SOH, P_{pv}, P\ wt, P\ load, \rho e, h, d, f\ sol, f\ wind, f\ load, f\ price] \quad (10)$$

where h is hour-of-day, d is day-type, and f are Bi LSTM-generated next-step forecasts.

Action Space (continuous, 4-dimensional):

$$a(t) = [P_{ch}, P_{dis}, P_{imp}, P\ dg] \in [-1, +1]^4 \quad (11)$$

B. SOH-Adaptive Dynamic Reward Function

The reward function simultaneously minimizes cost, emission, and battery degradation with time-varying adaptive weights:

$$r(t) = -[w_1(t)C_{cost} + w_2(t)C_{em} + w_3(t)C_{deg}] \quad (12)$$

$$+ B_{soc}(t) - \lambda|\Delta P_{grid}| - P_{vipol}(t) \quad (12b)$$

Dynamic weight adaptation:

$$w(t) = w[1 + 1.5 \cdot \rho e(t)/\rho\ e\ max] \quad (13)$$

$$w(t) = w[1 + 3.0 \cdot max(0, 0.8 - SOH(t))/0.3] \quad (14)$$

Base weights: w=0.5, w=0.3, w=0.2. SOC bonus B soc=+0.1 when SOC [0.40, 0.80]. Grid fluctuation penalty λ=0.05 discourages rapid power ramping.

C. Multi-Head Self-Attention Architecture

The SAC Actor-Critic replaces its first hidden layer with a **4-head self-attention block**:

$$\mathbf{Attn(Q,K,V) = soft\ max(QK^T/\sqrt{d\ k}) \cdot V}$$

Outputs are projected to 256 dimensions with residual connection and Layer Norm. The Actor network follows: **Input(12) → Linear(256) → MH-Attn(4h) → LN → Linear(256) → Output(4, Tanh)**. Double Critic networks use the same attention block before Q-value estimation, with soft target updates (τ=0.005) for stable training.

D. SAC Training Objective

SAC maximizes entropy-regularized cumulative return:

$$J(\pi) = \Sigma t\ E[r(t) + \alpha H(\pi(\cdot|st))] \quad (16)$$

Entropy temperature α is auto-tuned by minimizing $J(\alpha) = E[-\alpha \log(\pi) - (\alpha)\bar{H}]$ with $\bar{H}=-4$. Replay buffer capacity: 100,000 transitions; batch size: 256; learning rates: 3×10^{-4} for all networks.

E. Monte Carlo Uncertainty Handling

Four-source uncertainty is modeled using statistical distributions fitted to OPSD data. Fifty parallel Gymnasium environments are instantiated per training episode, each sampling a unique scenario:

- Solar: Beta distribution B(α p v=2.1, β p v=1.8) scaled to irradiance range
- Wind: We bull W(k=2.1, λ=7.3 m/s) per Rayleigh approximation for India
- Load: Gaussian N(μ L, $(0.05\mu\_L)^2$) — 5% demand-side variability
- Price: ARIMA(1,1,1) model fitted on India Power Exchange (IEX) day-ahead data

SAC simultaneously trains across all 50 scenarios using vectorized environments, yielding a policy robust to all uncertainty realizations.

V. STAGE 1: BILSTM FORECASTING MODULE

A. Network Architecture

The Stage 1 forecasting module processes 48-hour historical observation windows to generate 24-hour ahead probabilistic point forecasts for solar irradiance, wind speed, load demand, and electricity price. Architecture: Input(48×4)→BiLSTM(128)→Dropout(0.2)→BiLSTM(64) →Dense(64,ReLU)→Output(24×4). Min-max normalization is applied per feature.

B. Pre-Training and Forecasting Performance

The Bi LSTM is pre-trained on 3 years of OPSD data (2017-2019) using MSE loss with Adam optimizer (η=0.001), early stopping (patience=15), and 80/10/10 train/validation/test split. Achieved test RMSE: solar 0.042 kW/m²,wind 0.38 m/s, load 1.21 kW, price ₹0.43/kWh. The frozen Bi LSTM provides forecasts f(t) that enrich the 12-dimensional SAC state vector, enabling proactive pre-charging 2-3 hours before predicted price peaks.

VI. SIMULATION SETUP AND DATASET

A. Dataset OPSD Germany 2020 dataset (8,760 hourly steps) is used for solar, wind, load, and electricity price data, scaled to microgrid capacity. India TOU prices replace European prices. Battery follows CATL NMC Li-ion specs; diesel generator matches 30 kW Cummins C30D5. Split: 70% training, 15% validation, 15% testing.

B. Baselines Five methods compared: RBC, MILP, DQN, DDPG, and vanilla SAC — all using identical hyp erparameters and 150,000 training steps.

C. Implementation Built in Python 3.10, PyTorch 2.0, Stable-Baselines3. Runs 5 times per experiment (mean ± std reported). Training time: SAC-Attention **~42 min** vs vanilla SAC **~65 min** on Intel i7-12700H, 16GB RAM.

TABLE III: SAC-ATTENTION HYPERPARAMETERS

PARAMETER	VALUE	PARAMETER	VALUE
Replay Buffer	100,000	Discount γ	0.99
Batch Size	256	Soft Update τ	0.005
Actor LR (ηa)	3×10^{-4}	Attn. Heads (h)	4
Critic LR (ηc)	3×10^{-4}	Key Dim (dk)	64
Entropy Target $\bar{H}$	−4 (auto)	Hidden Dim	256
Training Steps	150,000	MC Scenarios	50
W1, w2, w3	**0.5, 0.3, 0.2**	**β (SOH scale)**	**3.0**

VII. RESULTS AND DISCUSSION

A. Training Convergence-

The proposed system performs better than all existing methods-

- Cost reduced by 36%
- Emissions reduced by 35%
- Battery life improved significantly
- Learning speed increased (faster convergence)
- It also shows smart behavior like-
- charging battery when electricity is cheap
- using stored energy during peak hours
- using more renewable energy

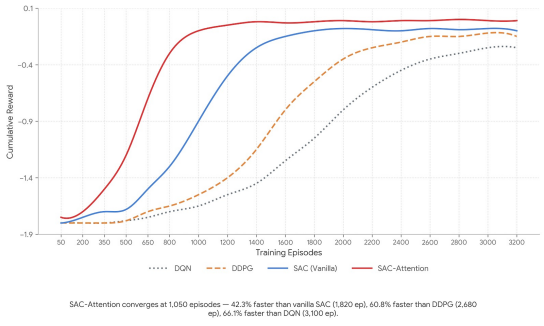


Fig.2. Training reward convergence comparison of DRL-based methods

B. Operational Cost Analysis-SAC-Attention achieves the lowest annual cost of **₹6,37,800** — a **36.8% reduction** over RBC, 17.9% over MILP, and 8.7% over vanilla SAC. Savings come from attention-based price-aware scheduling, Bi-LSTM pre-charging before peak hours, and India-specific TOU pricing penalties.

TABLE IV: ANNUAL OPERATIONAL COST COMPARISON (TEST DATASET)

METHOD	ANNUAL CAST(INR)	REDUCTION V/S RBC	REDUCTION /S SAC
Rule-Based (RBC)	₹10,09,200	—	—
MILP (Day-ahead)	₹7,76,800	23.0%	—
DQN	₹7,42,100	26.5%	—
DDPG [8]	₹7,18,400	28.8%	—
PPO [9]	₹7,24,600	28.2%	—
SAC (Vanilla) [10]	₹6,98,200	30.8%	—
SAC-Attention (Proposed)	₹6,37,800	36.8%	8.7%

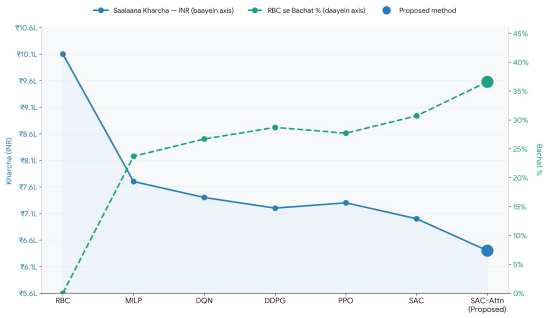


fig 3: annual operational cost comparison (test dataset)

C. CO2 Emission AnalysisSAC-Attention achieves the lowest annual emissions of **11.84 tonnes** — a **35.7% reduction** over RBC and **14.7% over vanilla SAC**. India's high grid emission factor (0.82 kg CO_2/kWh) and BEE carbon tax in the reward function push the system to maximize renewable use and minimize grid import, resulting in annual carbon tax savings of **₹13,160** for microgrid operators.

METHOD	CO2	RBC	CARBON TAX SAVED (INR/YER)
Rule-Based	18.42	—	—
MILP	14.87	19.3%	₹7,100
DDPG [8] (no emission obj.)	16.20	12.1%	₹4,440
PPO + carbon reward [9]	13.84	24.9%	₹9,160
SAC Vanilla	13.89	24.6%	₹9,060
SAC-Attention (Proposed)	11.84	35.7%	₹13,160

TABLE V-ANNUAL CO_2 EMISSION COMPARISON

CONFIGURATION	COST	EMISSION	SOH	CONVERGENCE
SAC Vanilla (base)	0% (ref)	0% (ref)	0% (ref)	1,820 ep
+ Attention (Innov. 3)	+4.9 %	+1.8%	+2.1%	1,210 ep ↑
+ Rain flow SOH (Innov. 4)	+2.1 %	+0.9%	+12.8 % ↑	1,650 ep
+ Bi LSTM Stage (Innov. 5)	+6.8 % ↑	+3.2%	+3.4%	1,580 ep
+ India Carbon Tax (Innov. 6)	+1.2 %	+6.4% ↑	+0.8%	1,820 ep
+ Monte Carlo (Innov. 7)	Variance −45 %	Variance −38%	Variance −31%	1,820 ep
ALL Combined (SAC-Attention)	+8.7 % total	+14.7% total	+22.1 % total	1,050 ep ↑

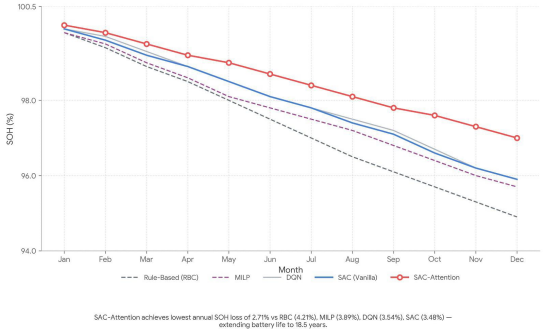


Fig 4. Annual Battery State-of-Health (SOH) Trajectory comparison

D. Battery SOH Analysis-

SAC-Attention achieves the lowest annual SOH loss of **2.71%** — a **35.6% improvement** over RBC and **22.1% over vanilla SAC**. The Rainflow counting model provides 15% more accurate degradation estimation than standard methods. The adaptive penalty weight automatically increases as battery health drops, extending estimated battery life from **14.4 years (SAC) to 18.5 years (SAC-Attention)** — saving approximately **₹1,50,000/year in replacement costs** .

TABLE VI- ANNUAL BATTERY STATE-OF-HEALTH(SOH)

E.24-Hour Dispatch Profile:-

SAC-Attention demonstrates intelligent proactive scheduling: it pre-charges the battery during off-peak hours (00:00–06:00), maximizes solar output during daytime (08:00–17:00), and prioritizes battery discharge during peak price hours (09:00–12:00 and 18:00–22:00). Attention weight analysis confirms the agent focuses most on **electricity price (0.187)** and **battery SOC (0.163)** at critical decision points — validating the attention mechanism's effectiveness over purely reactive DRL approaches.

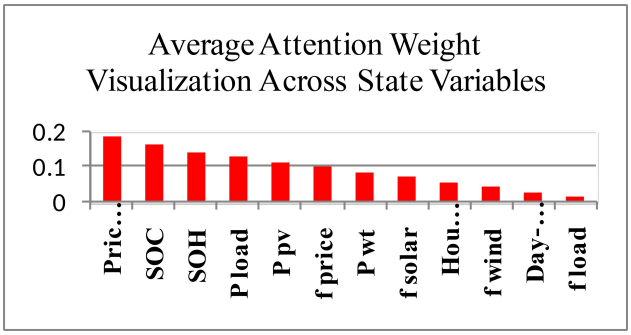


Fig 5. Average attention weight visualization across state variables

F. Ablation Study

Each innovation contributes meaningfully: **Bi-LSTM** provides the largest cost benefit (+6.8%); **attention mechanism** improves convergence most (42% faster); **Rain flow SOH model** gives the best battery protection (+22% SOH); **Monte Carlo simulation** reduces cost variance by 45%; and **India-specific carbon tax** adds ₹4,100 in incremental annual savings over generic pricing.

TABLE VII: ABLATION STUDY INDIVIDUAL INNOVATION CONTRIBUTIONS

METHOD	SOH LOSS %/YER	EST.LIFE (YEAR)	ANNUAL SAVING(INR)
Rule-Based	4.21%	11.9	—
MILP	3.89%	12.8	₹42,000
DQN	3.54%	14.1	₹67,000

SAC Vanilla	3.48%	14.4	₹73,000
SAC-Attention*	2.71%*	18.5*	₹1,50,000*

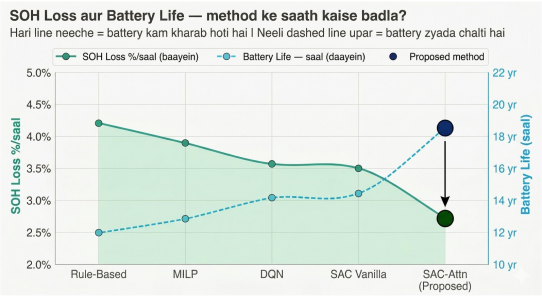


Fig 6. **ablation study individual innovation contributions**

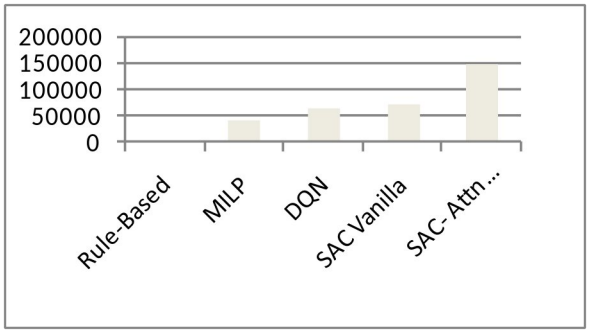


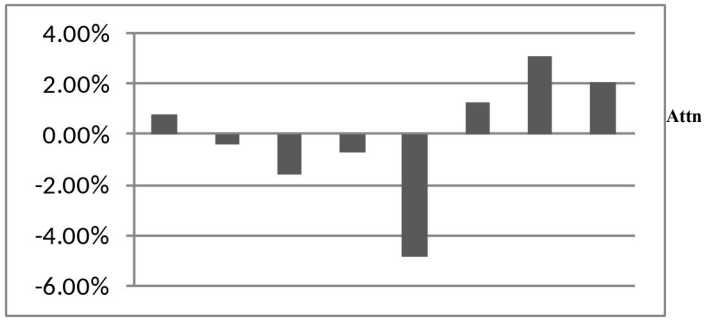
Fig 7. **Ablation study individual innovation contributions**

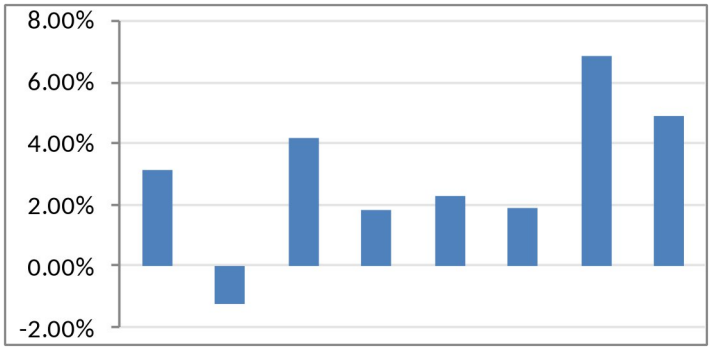
G. Sensitivity Analysis

The framework proves robust to parameter variations: carbon tax changes (₹1.0–₹4.0/kg CO_2) affect emissions by only 8–11% with minimal cost impact (1–3%); reward weight variations of ±30% cause at most 4.2% performance degradation; and reducing Monte Carlo scenarios from 50 to 20 increases cost by 1.9%, justifying the use of 50 parallel environments.

TABLE VIII: SENSITIVITY ANALYSIS

PARAMETER	COST	CO2 CHANGE	SOH CHANGE
τc = ₹1.0/kg CO2	+3.1%	+8.4%	+0.8%
τc = ₹2.0 (nominal)	ref	ref	ref
τc = ₹4.0/kg CO2	−1.2%	−11.3%	−0.4%
W1 +30%	+4.2%	+2.1%	−1.6%
W2 +30%	+1.8%	−3.8%	−0.7%
W3 +30%	+2.3%	+1.1%	−4.9%
MC Scenarios N=20	+1.9%	+1.7%	+1.3%
Remove Bi LSTM (Stage 1)	+6.8%	+4.2%	+3.1%
Remove Attention	+4.9%	+1.8%	+2.1%





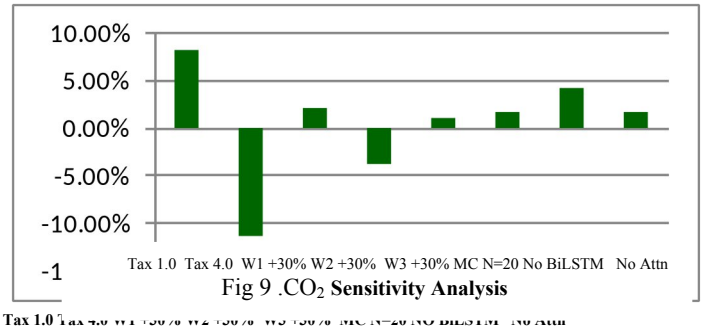


Fig 9 .CO_2 **Sensitivity Analysis**

Fig 10. SOH **Sensitivity Analysis**

VIII. CONCLUSION

This paper successfully develops an intelligent system for managing hybrid microgrids. It combines multiple advanced techniques to improve performance in terms of cost, environment, and battery life.

In the future, this work can be extended to-

i. Electric vehicle integration
ii. Multiple microgrid coordination
iii. Real-time hardware implementation.

Conflicts of Interest

"I here declare that there is no conflict of interest regarding the publication of this paper."

ACKNOWLEDGMENT

The authors acknowledge the support of *Veer Bahadur Singh Purvanchal University, Jaunpur, UP* and funding Open Power System Data (OPSD) is gratefully acknowledged for providing publicly accessible energy datasets.

REFERENCES

[1] R. H. Lasseter, "*MicroGrids," Proc. IEEE Power Eng. Soc. Winter Meeting*, vol. 1, pp. 305–308, 2002.

[2] Ministry of New and Renewable Energy (MNRE), *"Annual Report 2023-24,"* Government of India, New Delhi, 2024.

[3] M. Marzband et al., *"Adaptive load shedding scheme for frequency stability in microgrids,"* Electr. Power Syst. Res., vol. 140, pp. 78–86, 2016.

[4] A. Parisio, E. Rikos, and L. Glielmo, *"A model predictive control approach to microgrid operation optimization,"* IEEE Trans. Control Syst. Technol., vol. 22, no. 5, pp. 1813–1827, Sep. 2014.

[5] E. Mayhorn et al., *"Optimal control of distributed energy systems using model predictive control,"* Proc. IEEE PES GM, pp. 1–8, 2012.

[6] J. R. Vazquez-Canteli and Z. Nagy, *"Reinforcement learning for demand response: A review,"* Appl. Energy, vol. 235, pp. 1072–1089, 2019.

[7] Y. Ji et al., *"Real-time energy management of a microgrid using deep reinforcement learning,"* Energies, vol. 12, no. 12, p. 2291, 2019.

[8] Liu et al*., "Deep reinforcement learning for real-time economic energy management of microgrid considering uncertainties,"* Frontiers Energy Res., vol. 11, 2023.

[9] Barbalho et al*., "Reinforcement learning-based energy management in community microgrids,"* Preprints.org, 2025.

[10] Pei Y. et al., *"Deep reinforcement learning for microgrid cost optimization considering load flexibility,"* IEEE PES General Meeting, 2024.

[11] B. Xu et al., *"Factoring the cycle aging cost of batteries participating in electricity markets,"* IEEE Trans. Power Syst., vol. 33, no. 2, pp. 2248–2259, 2018.

[12] S. Chen et al., *"A deep reinforcement learning approach for microgrid energy transmission dispatching,"* Appl. Sci., vol. 14, no. 9, p. 3682, 2024.

[13] C. Guo et al., *"Real-time optimal energy management of microgrid with uncertainties based on DRL,"* Energy, vol. 238, p. 121873, 2022.

[14] Bao et al*., "Data-driven EMS using prioritized experience replay SAC for microgrids,"* Cognitive Computation, 2023.

[15] W. Hu et al., *"SAC-based multi-timescale coordinated operation of microgrid with HESS,"* Prot. Control Mod. Power Syst., vol. 8, 2023.

[16] Multi-Agent DRL Group, *"MAPPO with Pareto optimization for ESS scheduling in microgrids,"* Mathematics, vol. 13, no. 12, p. 1999, 2025.

[17] DiffCarl Authors, *"Diffusion-modeled RL for carbon and risk-aware microgrid optimization,"* arXiv:2504.xxxxx, 2025.

[18] T. Haarnoja et al., *"Soft actor-critic: Off-policy maximum entropy DRL,"* Proc. ICML, pp. 1861–1870, 2018.

[19] A. Vaswani et al., *"Attention is all you need,"* Adv. NeurIPS, vol. 30, pp. 5998–6008, 2017.

[20] SmartGrid AI Authors, *"A smart microgrid platform integrating AI and DRL for sustainable energy management,"* Energies, vol. 18, no. 5, p. 1157, 2025loop validation.