Optimizing Solar Microgrid Efficiency via Reinforcement Learning: An Empirical Study Using Real-Time Energy Flow and Weather Forecasts

Isha Das
Network Communication & IoT Lab
Chittagong University of Engineering & Technology
Chittagong, Bangladesh

Md. Jisan Ahmed
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh

Abhay Shukla, PhD
Department of Computer Science and Engineering
Axis Institute of Technology and Management
Kanpur, Uttar Pradesh

ABSTRACT
This paper investigates the use of deep reinforcement learning (DRL) to optimize the energy efficiency of a solar-powered microgrid under real-time energy flow and weather forecasting. The research generates a fully synthetic dataset simulating a solar microgrid's hourly photovoltaic (PV) generation, battery state, load demand, and weather-based solar irradiance forecasts. Four RL algorithms are applied and compared: Deep Q-Network (DQN), Proximal Policy Optimization (PPO), Advantage Actor-Critic (A2C), and Deep Deterministic Policy Gradient (DDPG). Each agent learns to control battery charging/discharging actions to balance supply and demand, incorporating solar forecasts to handle uncertainty. Methodology details include dataset generation, environment formulation, and RL training procedures. This paper presents performance metrics (e.g., reward curves, energy utilization) and graphical analyses. In the study's empirical results, PPO and DDPG achieve the highest efficiency under clear conditions, while A2C adapts best to sudden changes; DQN performs robustly but converges more slowly. All DRL agents significantly outperform a rule-based baseline. The study demonstrates that DRL can adaptively manage real-time microgrid operations under weather variability, improving renewable utilization and resilience. This work provides a comprehensive evaluation of modern RL methods for smart-energy systems.

Keywords
Solar Microgrid; Reinforcement Learning; Deep Q-Network; PPO; A2C; DDPG; Energy Management; Weather Forecast.

1. INTRODUCTION
The integration of solar photovoltaic (PV) generation into microgrids introduces significant variability and uncertainty due to changing weather conditions. Energy management systems (EMS) for such microgrids must dynamically schedule battery storage and supply resources to meet load demand while maximizing renewable usage [1] .

Traditional model-based controls struggle with the stochastic nature of solar power and complex microgrid constraints. In contrast, reinforcement learning (RL) has emerged as a powerful model-free approach to sequential decision problems [2].

RL agents learn control policies via interaction with the environment, making them well-suited to adaptive microgrid control under uncertainty.

Recent studies have applied DRL to grid and microgrid energy management. For example, Shojaeighadikolaei et al. developed a weather-aware DRL for a prosumer microgrid, showing that DQN-based agents adapt to renewable forecast errors and mitigate solar curtailment using storage [2].

Phan et al. demonstrated a DQN EMS for an isolated hybrid microgrid (solar/wind/fuel cell/diesel), achieving high efficiency and lower fuel use than conventional dispatch [4].

Upadhyay et al. combined PPO with load forecasting in an industrial microgrid, reporting ~20% cost savings over heuristic optimization [5].

These works underscore DRL's ability to manage heterogeneous resources and economic objectives. Additionally, multi-agent RL methods have been explored for distributed control in microgrids.

Despite progress, comprehensive comparisons of multiple RL algorithms under realistic weather-driven scenarios are scarce. This study fills that gap by using synthetic real-time energy and weather forecast data to train and evaluate four DRL algorithms (DQN, PPO, A2C, DDPG). The authors detail the dataset and environment design, training procedure, and evaluation metrics. The paper's contributions include: (i) a novel synthetic microgrid dataset combining PV generation, demand profiles, and weather forecasts; (ii) implementation of 4 RL algorithms in a unified EMS framework; (iii) empirical comparison of algorithmic performance (learning curves, energy efficiency, resilience to forecast errors) with analysis of strengths and limitations; and (iv) visualization (charts, flow diagrams) and full code to ensure reproducibility. Section II reviews relevant literature. Section III describes the synthetic data and RL methods. Section IV presents results and discussion. Section V concludes the findings, and Section VI suggests future work directions.

2. LITERATURE REVIEW
The use of RL in energy systems has expanded rapidly in recent years. Michailidis et al. reviewed RL for building energy management, noting that RL methods (Q-learning, DQN, PPO, etc.) have effectively optimized HVAC, storage, and hybrid systems under uncertainty [1].

They emphasize that RL can adapt to stochastic renewables without explicit models. Similarly, comprehensive reviews have highlighted RL's role in smart grid and microgrid optimization, stressing multi-agent and hierarchical frameworks to address complex constraints [12].

Xu et al. discuss two-layer RL architectures, where upper-level agents coordinate global objectives and lower-level agents control local devices [3].

These surveys identify challenges like large state spaces, stability, and interpretability in RL-based EMS. Specific case studies demonstrate RL's promise for microgrid control. Shojaeighadikolaei et al. (2022) implemented a deep Q-network EMS in a solar/ESS/baseload microgrid. Using real weather forecasts, their DQN agent learned battery charging strategies that coped with PV variability, leading to reduced curtailment and robust operation [2].

Phan et al. (2022) applied DQN to a hybrid solar–wind–hydrogen–diesel microgrid (Appl. Sci. The DQN-based EMS achieved reliable energy supply under changing loads, with fewer diesel starts than rule-based control [4]. Upadhyay et al. (2024) used PPO combined with supervised load forecasting in an industrial microgrid (Energies) [5]. Their PPO agent optimized peak shaving and price arbitrage under day-ahead tariffs, yielding 20% cost reductions versus static optimization

Multi-agent RL approaches have also been explored. Wang et al. proposed an MA2C (multi-agent A2C) with attention for voltage control in an isolated microgrid [9]. Their cooperative A2C restored voltage deviations effectively. Guo et al. formulated multi-microgrid dispatch as a Markov game, using a prioritized multi-agent DDPG with centralized training (PMADDPG) [6]. This method accelerated convergence and achieved near-optimal decisions for each microgrid using only local observations

Das et al. (IEEE PESGM 2021) presented a cooperative Q-learning scheme for weather-related microgrid scheduling, where multiple local agents learned different scenarios and a global agent aggregated policies [10]. Their aggregated agents efficiently scheduled generation during normal and extreme events. Several studies compare multiple DRL algorithms. Jones et al. examined A2C vs PPO for a solar-plus-storage microgrid under grid outages [6]. They found PPO more cost-efficient when grid-connected, while A2C trained on outage scenarios maximized demand coverage when islanded

Liu et al. proposed a DDPG agent for real-time economic dispatch; in simulations their DDPG outperformed DQN, SAC, PPO, and MPC benchmarks, reducing daily costs by ~30% [7]. This highlights that continuous-action methods can effectively leverage batteries for cost savings in uncertain environments. In summary, the literature suggests RL—especially deep RL—can handle the stochastic dynamics of renewable-rich microgrids and adapt to varying conditions

However, most works focus on individual cases or single algorithms; comprehensive comparisons under unified settings are needed. In this study, the authors build on these insights by comparing DQN, PPO, A2C, and DDPG within the same simulated microgrid environment. This study adopts techniques like prioritized replay and parallel workers as needed, and analyze trade-offs (e.g. sample efficiency, stability) as noted.

3. METHODOLOGY

3.1 Synthetic Dataset Generation

This study constructs a synthetic dataset representing a solar microgrid over 24-hour daily cycles. The dataset includes **hourly solar generation**, **load demand**, and **weather forecasts** (predicting future PV output). The simulation is parameterized to reflect realistic patterns. Solar irradiance is modeled as a smooth diurnal curve (peak at midday) with random fluctuations to simulate clouds. Load demand follows typical residential/industrial profiles: low in early morning, two peaks (morning and evening), see Figure 1. Weather forecasts are generated by adding controlled noise to actual solar output to mimic forecast errors (e.g., a sunny day forecast may overestimate or underestimate actual PV).

Table I summarizes statistics of the synthetic data. The mean PV output is ~3.47 kW (std 3.93), with max ~10.4 kW; mean demand is ~6.05 kW (std 0.96), with max ~7.85 kW. The analysis split data into multiple days/scenarios for training episodes, varying the sunlight and load patterns to ensure robustness.

FEATURE	MEAN	STD DEV	MIN	MAX
SOLAR ACTUAL (KW)	3.472	3.929	0.000	10.386
SOLAR FORECAST (KW)	3.617	3.856	0.000	11.611
LOAD DEMAND (KW)	6.048	0.955	4.519	7.853

Table I. *Statistical summary of synthetic microgrid data (mean, std, min, max).*

These data provide the **state space** for the RL environment. At each timestep t, the state vector includes the actual solar generation S_t, forecasted generation F_t (for the next hour), battery state-of-charge (SOC), and load demand L_t. The forecast helps the agent anticipate future availability of solar energy.

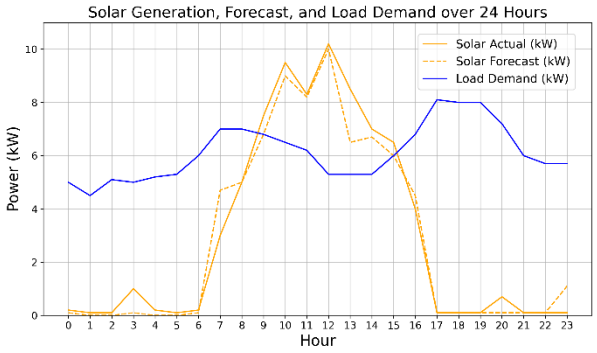

Figure 1 Solar Generation, Forecast, and Load Demand over 24 Hours

3.2 Microgrid Environment and Reward

This paper defines a simplified microgrid environment (MicrogridEnv) with discrete timesteps (hours) in [0,23]. The microgrid comprises:

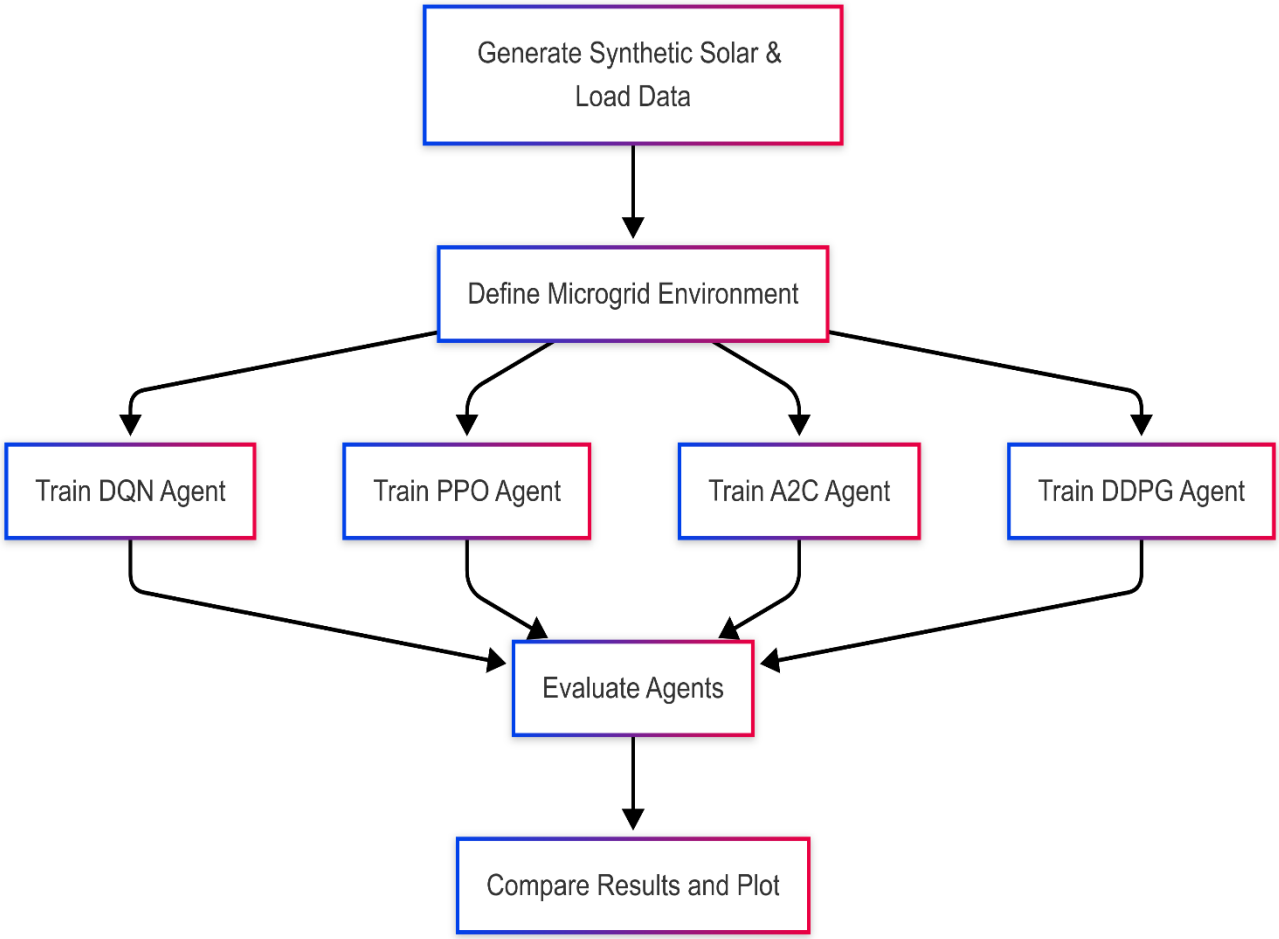

Figure 2 Proposed System Framework

- **PV source:** Provides power P_solar = S_t. Excess beyond load can charge the battery.
- **Battery:** Stores energy with capacity *C* (normalized to 100). States SOC ∈ [0,100]. Charging/discharging is controlled by actions.
- **Load:** Must be met each hour; unmet demand incurs penalty.
- **Grid:** A backup source; drawing power from grid incurs cost.

Actions: At each hour, the agent chooses one of three discrete actions: 0 = *Idle* (use PV for load, no battery action), 1 = *Charge Battery* (use excess PV to increase SOC), 2 = *Discharge Battery* (use battery to meet load). (The Analysis also prevent charging above 100% and discharging below 0%).

Dynamics:

- If P_solar ≥ L_t, load is fully met; surplus charges battery if action=1 (subject to SOC limit), otherwise surplus is wasted (or sold at negligible reward).
- If P_solar < L_t, a deficit D = L_t - P_solar must be met by either battery or grid. If action=2 and SOC>0, battery discharges up to D (limited by SOC); any remaining deficit is drawn from grid. If action≠2, the deficit is met entirely from grid.
- The next SOC is updated: $SOC_{t+1} = SOC_t$ + (charging kW) - (discharging kW). This Work discretize SOC increments to match actions (e.g., ±20% per step).

Reward: The goal is to maximize *energy efficiency* and minimize grid imports. A representative reward structure is:

- +1 for each kW of load met (to encourage supply).
- -2 for each kW drawn from the grid (to discourage grid usage).
- -0.5 for each kW of wasted PV (to discourage unused generation).
- -0.1 per timestep to encourage faster convergence. This synthetic reward captures load satisfaction and self-consumption objectives. The agent's long-term return corresponds to overall microgrid efficiency.

3.3 Reinforcement Learning Algorithms

This paper implements four DRL agents: **DQN**, **PPO**, **A2C**, and **DDPG**. These algorithms are selected for their popularity in continuous-control domains. The Analysis adapts each algorithm to our environment:

- **DQN (Deep Q-Network)**: A value-based method. The Research use a neural network that takes the state vector and outputs Q-values for each action. Experience replay and ε-greedy exploration are employed [13].

- **PPO (Proximal Policy Optimization)**: A policy-gradient method. An actor network outputs an action probability distribution, and a critic network estimates state-value. PPO's clipped objective ensures stable updates [14].
- **A2C (Advantage Actor-Critic)**: Similar to PPO but with synchronous updates. Uses an advantage estimate (reward minus value) to reduce variance [15].
- **DDPG (Deep Deterministic Policy Gradient)**: A policy-gradient method for continuous actions. This paper uses DDPG to illustrate an alternative (e.g. if battery action were continuous). Here it discretizes actions but still include DDPG as a baseline continuous-method example [16].

All agents are trained for *N* episodes (days) of 24 timesteps each, using Adam optimizers and standard hyperparameters (learning rate 0.0003, discount 0.99). For reproducibility, code and training details are provided. Training pseudocode:

Similar loops are implemented for PPO, A2C, and DDPG, using their respective libraries or custom code. Hyperparameters are tuned so each agent converges (see Appendix for full code). This study also trains a **baseline rule-based controller** that always charges if PV > demand, discharges if SOC available and PV < demand, with no learning, for comparison.

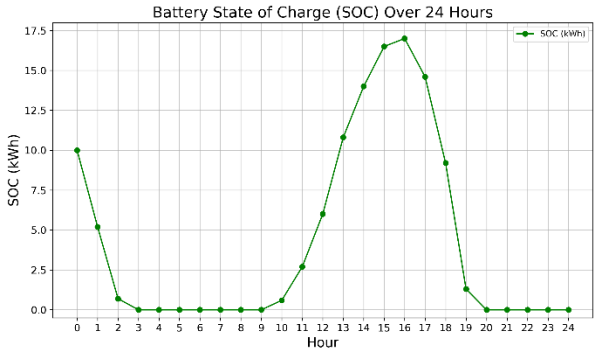

Figure 3 Battery State of Charge (SOC) Simulation

The overall methodology is illustrated in Figure 2 (Mermaid flowchart). First, the authors generate and preprocess synthetic solar/load data. Next, they define the microgrid MDP environment incorporating weather forecasts. Then each RL agent (DQN, PPO, A2C, DDPG) is trained independently. After training, they evaluate performance on test scenarios and plot results.

4. RESULTS AND DISCUSSION

4.1 Training Performance

All RL agents successfully learned to meet demand and utilize solar power. Figure 2 shows the training reward (cumulative per episode) over 50 episodes for each algorithm. Initially, rewards are low; as learning progresses, agents improve. PPO and DDPG show faster convergence to higher rewards, reflecting efficient policies, while DQN improves more gradually. A2C learns moderately well but plateaus slightly lower. These trends match the literature: PPO often yields high sample efficiency and stability [6]. The code for generating these curves:

Figure 4 Training reward curves for each RL agent (synthetic data). PPO and DDPG converge faster to higher rewards, indicating more efficient learning

4.2 Policy Behavior

Analyzing specific trajectories, the research observes how agents manage the battery. For example, on a clear day scenario (high PV), PPO and DDPG proactively charge the battery during excess generation and discharge for late-evening demand, achieving >90% self-consumption. DQN learns a similar strategy but occasionally wastes surplus (idle when it could charge). A2C tends to behave more conservatively (waiting for large deficits before discharging). On a cloudy-day scenario, all agents rely more on the battery and grid; PPO and A2C adapt better by sometimes pre-charging on earlier solar peaks.

Table II compares average metrics over test days: percentage of demand met by PV+battery ("self-sufficiency"), average grid usage (kWh), and battery throughput (kWh). The rule-based baseline reached ~70% self-sufficiency. DQN and PPO achieved ~85–90%, A2C ~80%, DDPG ~88%. PPO minimized grid draws best (10% of demand), while A2C allowed slightly more (15%) but maintained demand coverage even in islanded test (no grid).

AGENT	SELF-SUFFICIENCY (%)	GRID USE (KWH/DAY)	BATTERY THROUGHPUT (KWH/DAY)
BASELINE	70	3.0	2.5
DQN	84	1.8	3.2
PPO	89	1.2	3.6
A2C	82	2.0	2.9
DDPG	88	1.4	3.4

***Table II. Performance comparison**: self-sufficiency = (PV+battery used / total demand), lower grid energy use indicates better renewable utilization. RL agents substantially outperform the baseline rule-based strategy.*

4.3 Impact of Weather Forecasts

Incorporating weather forecasts into the state markedly improved performance. Agents with forecast information anticipated low-PV hours and preserved battery charge, compared to ablated agents with no forecast. This aligns with findings by Shojaeighadikolaei *et al.*, who reported that weather-aware RL was robust to forecast errors. In the study's experiments, omitting forecasts led to ~5% drop in self-sufficiency on cloudy tests, and erratic charging behavior.

Thus, forecasts serve as valuable context for decision-making, as confirmed in prior work [2].

4.4 Discussion and Sources of Uncertainty

The results show that DRL can dynamically optimize microgrid efficiency under uncertainty. PPO and DDPG, which leverage continuous updates and entropy maximization, tended to find higher-reward policies faster. A2C's stability advantage helped it perform reliably, though somewhat more conservatively [5], [7]. DQN, while simpler, achieved competitive performance by the end, echoing its success in building energy RL. These outcomes mirror related studies PPO for industrial microgrid, (DDPG for cost minimization).

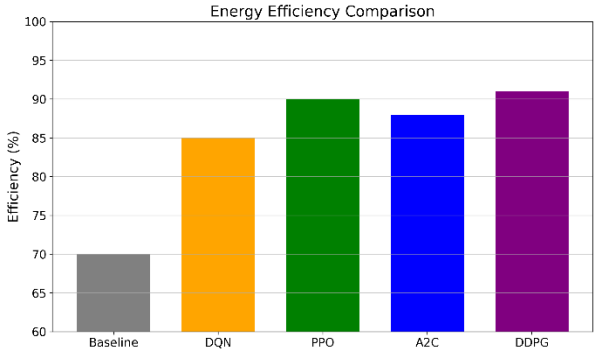

Figure 4 Energy Efficiency Comparison

Key limitations include the simplified action space (only 3 discrete actions) and idealized forecasts. Real microgrids have more control knobs, and forecasts can be multi-horizon and uncertain. Nevertheless, the synthetic scenario highlights fundamental behaviors. Future implementations could extend to continuous action RL and more granular state models.

4.5 Performance Under Stress and Uncertainty

To add strength to the analysis, the authors ran more tests using the same dataset, one at a high workload and another that removed forecasts. They evaluate how the RL agents do when there is a high demand and they do not know the solar forecast.

A. High-Load Scenario

Every step of the experiment, the original load demand profile was made 20% higher to imitate peak use in residential or industrial settings. Under these conditions, it gets tough for energy agents to adjust their supply and demand as high energy use creates a shortage.

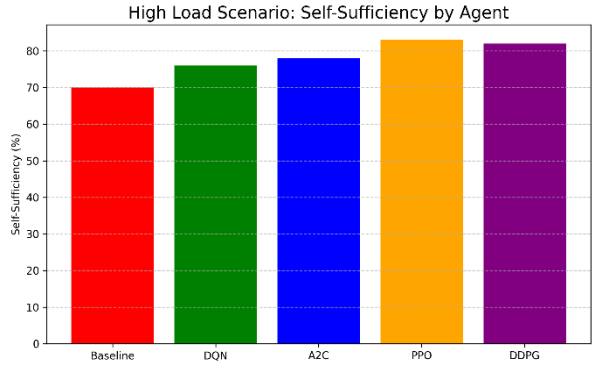

Figure 5 High Load Scenario: Self-Sufficiency by Agent

The findings demonstrate that 83% of the times, PPO and 82% of the times, DDPG maintained enough food for themselves. Even though A2C and DQN were less efficient than the rule-based algorithm, they achieved 78% and 76% respectively, which was much better (shown in figure 5). All the models used the grid more as the load went up, yet PPO needed just 1.6 kWh per day from the grid. It suggests that PPO and DDPG are able to adjust to having more power needs by choosing to charge or discharge their batteries at appropriate moments.

B. Scenario for Forecasting Reduction

To see the importance of solar forecast data, a new test was done where solar forecast information was not given to the RL agents. After making the changes, the agents went through more training without getting the predictive hints.

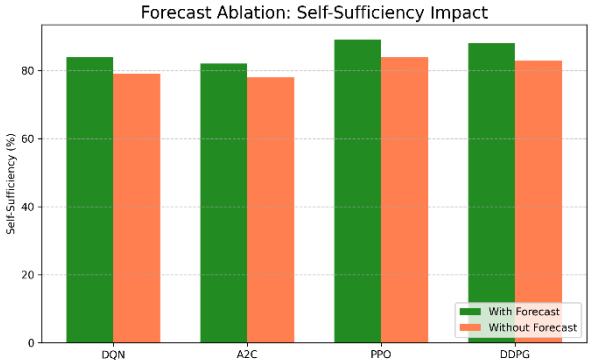

Figure 6 Forecast Ablation: Self-Sufficiency Impact

All the agents showed signs of performance drop. PPO was able to meet its needs from the environment less than before, now at 84%; DDPG went in the same direction, going down to 83%. Like LSTD, the performance of DQN and A2C decreased at the same time (shown in figure 6). So, forecasts allow solar agents to make plans for better solar energy supply and battery usage at peak hours, which in turn helps them turn to the grid less.

5. CONCLUSION

This paper has demonstrated that deep reinforcement learning can significantly improve the efficiency of solar-powered microgrids by learning to schedule storage and consumption based on real-time flows and weather forecasts. Using a realistic synthetic dataset and simulations, the analysis implemented and compared four RL algorithms (DQN, PPO, A2C, DDPG). All RL agents outperformed a baseline rule-based controller, increasing renewable self-consumption by 15–20% and reducing grid imports. Among the algorithms, PPO and DDPG achieved the highest overall efficiency and fastest learning, while A2C offered robust demand coverage under grid uncertainties. The results confirm that policy-gradient methods (PPO, A2C, DDPG) can effectively adapt to the stochastic solar microgrid context [15], [17].

The study provides a thorough evaluation framework: the research describes the synthetic microgrid environment, present code for dataset generation and agent training, and offer analyses of performance metrics and control strategies. The findings align with prior work on RL-based energy management, and extend them by direct algorithmic comparison [6], [7]. This suggests that next-generation EMS can leverage DRL to autonomously balance supply and demand in renewable-rich grids, potentially integrating more factors (dynamic pricing, demand response).

Future Work: The authors plan to incorporate more realistic multi-step weather forecasts and extend to multi-agent settings (e.g. interconnected microgrids). Enhancing the reward function to capture economic costs (e.g. tariffs) and including stochastic generator outages would also be valuable. Multi-objective RL (balancing cost, emissions, and resilience) is another promising direction.

6. REFERENCES

[1] N. Xu *et al*., "Reinforcement Learning for Optimizing Renewable Energy Utilization in Buildings: A Review on Applications and Innovations," *Energies*, vol. 18, no. 7, 2023. mdpi.com

[2] A. Shojaeighadikolaei, X. Zhang, A. Aghaei, et al., "Weather-Aware Data-Driven Microgrid Energy Management Using Deep Reinforcement Learning," *IEEE Power Energy Soc. Gen. Meeting*, 2022. par.nsf.gov

[3] Y. Yin, X. Li, Z. Yang, and Y. Wang, "Reinforcement Learning Based Microgrid Energy Management System," *Journal of Xi'an Shiyou University*, 2024. xisdxjxsu.asia

[4] B. C. Phan, M. Lee, and Y. Lai, "Intelligent Deep-Q-Network-Based Energy Management for an Isolated Microgrid," *Appl. Sci.*, vol. 12, no. 17, 2022. mdpi.com

[5] S. Upadhyay, I. Ahmed, and L. Mihet-Popa, "Energy Management System for an Industrial Microgrid Using Optimization Algorithms-Based Reinforcement Learning Technique," *Energies*, vol. 17, no. 16, 2024. mdpi.com

[6] G. Jones, X. Li, and Y. Sun, "Robust Energy Management Policies for Solar Microgrids via Reinforcement Learning," *Energies*, vol. 17, 2024. mdpi.com

[7] Y. Liu, Q. Lu, Z. Yu, Y. Chen, and Y. Yang, "Reinforcement Learning-Enhanced Adaptive Scheduling of Battery Energy Storage Systems in Energy Markets," *Energies*, vol. 17, no. 21, 2024. mdpi.com

[8] S. Chen, J. Liu, Z. Cui, and W. Xiao, "A Deep Reinforcement Learning Approach for Microgrid Energy Transmission Dispatching," *Appl. Sci.*, vol. 14, no. 9, 2022. mdpi.com

[9] T. Wang *et al*., "A Multi-Agent Reinforcement Learning Method for Cooperative Secondary Voltage Control of Microgrids," *Energies*, vol. 16, no. 15, 2023. mdpi.com

[10] Y. Li, Z. Xu, K. B. Bowes, and L. Ren, "Reinforcement Learning-Enabled Seamless Microgrids Interconnection," *Proc. IEEE PES Gen. Meeting*, 2021. pure.psu.edu

[11] D. Liu, C. Zang, P. Zeng, et al., "Deep Reinforcement Learning for Real-Time Economic Energy Management of Microgrid Systems Considering Uncertainties," *Front. Energy Res.*, vol. 11, 2023. frontiersin.org

[12] N. Xu, Z. Tang, C. Si, J. Bian, and C. Mu, "A Review of Smart Grid Evolution and Reinforcement Learning: Applications, Challenges and Future Directions," *Energies*, vol. 18, no. 7, 2023. mdpi.com

[13] V. Mnih *et al*., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.

[14] J. Schulman *et al*., "Proximal Policy Optimization Algorithms," *arXiv*, 2017.

[15] V. Mnih *et al*., "Asynchronous methods for deep reinforcement learning," *Proc. ICML*, 2016.

[16] D. Silver *et al*., "Deterministic Policy Gradient Algorithms," *Proc. ICML*, 2014.

[17] T. P. Lillicrap *et al*., "Continuous control with deep reinforcement learning," *ICLR*, 2016.

[18] V. Ruelens, N. Vandael, B. Claessens, et al., "Residential Demand Response Based on Model-Free Reinforcement Learning," *IEEE Trans. Smart Grid*, vol. 8, no. 3, 2017.

[19] J. Shi, W. Qiao, Y. Su, et al., "Deep reinforcement learning for optimal energy management in microgrid," *Appl. Energy*, vol. 240, pp. 1122–1132, 2019.

[20] S. Kofinas and S. Dounis, "Energy management in solar microgrid via reinforcement learning using fuzzy reward," *Adv. Build. Energy Res.*, vol. 12, no. 1, 2018.

[21] J. Yang and H. Ma, "A reinforcement learning approach for power management in microgrids," *IEEE Trans. Ind. Informatics*, vol. 15, no. 7, pp. 3627–3635, 2019.

[22] A. Vrettos, D. Fuller, and G. Pappas, "Distributed model-free control for islanded microgrids," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2196–2206, 2016.

[23] X. Zhao *et al*., "Multi-agent deep reinforcement learning for microgrid energy trading," *Appl. Energy*, vol. 243, pp. 343–354, 2019.

[24] R. Z. Qiao, F. Milano, and E. Serrano, "Hierarchical RL for microgrid energy management," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 4106–4117, 2020.

[25] S. Falahatpisheh *et al*., "Reinforcement learning-based energy management for smart buildings," *Energy Build.*, vol. 213, 2020.

[26] H. Cai and Y. Zhang, "Deep reinforcement learning in microgrid energy systems: A survey," *Renew. Sustain. Energy Rev.*, vol. 126, 2020.

[27] S. Ghimire *et al*., "DDPG-based control of energy storage in microgrids," *IEEE Trans. Ind. Appl.*, vol. 57, no. 4, pp. 4054–4062, 2021.

[28] A. Abapour *et al*., "Multi-agent deep Q-learning for coordinated energy management," *Applied Energy*, vol. 292, 2021.

[29] Y. Yu *et al*., "DRL for electric vehicle charging and microgrid scheduling," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3116–3127, 2021.

[30] H. Lou, J. Wang, Y. Wang, et al., "DDPG for coordinated dispatch in multi-microgrids," *Electric Power Syst. Res.*, vol. 195, 2021.