UJEAS

UNIZIK JOURNAL OF ENGINEERING AND APPLIED SCIENCES

UNIZIK Journal of Engineering and Applied Sciences 6(5), December (2026), 4122-4139
Journal homepage: https://journals.unizik.edu.ng/index.php/ujeas
PRINT ISSN: 2992-4383 || ONLINE ISSN: 2992-4391

DEEP REINFORCEMENT LEARNING-BASED ADAPTIVE ENERGY MANAGEMENT FOR LITHIUM-ION BATTERY ENERGY STORAGE SYSTEMS

Prince Nonso Nwankwo*

Department of Computer Engineering, Federal Polytechnic, Oko

Christiana Chikodi Okezie

Department of Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka

Nosike Anthony Isizoh

Department of Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka

**Corresponding Author's E-mail: nonso.nwankwo@federalpolyoko.edu.ng*

Abstract

Effective energy management is important for the reliable operation of lithium-ion battery energy storage systems, especially when solar generation and electricity demand vary. This study develops a Deep Reinforcement Learning-based energy management approach using a Deep Deterministic Policy Gradient (DDPG) controller for a photovoltaic–battery energy storage system. The controller uses battery State of Charge (SOC), battery temperature, photovoltaic power, and load demand to determine appropriate charging and discharging actions. To create a realistic operating environment, solar resource data for Ekwulobia, Anambra State, Nigeria, obtained from NASA POWER, were combined with the Ausgrid 2012–2013 residential electricity demand dataset. The combined dataset contained 52,560 observations at 10-minute intervals and was divided chronologically into training, validation, and unseen testing periods. The DDPG controller was tested using the unseen May–June 2013 data and compared with a conventional photovoltaic-priority rule-based energy management system (RB-EMS). During testing, the DDPG controller achieved 575.34 kWh of battery charging, 521.35 kWh of discharging, and 1,096.69 kWh of total battery energy throughput. Grid import was 1,096.86 kWh, representing a 3.39% reduction compared with the RB-EMS, while battery energy throughput was 6.88% lower. The controller maintained SOC, temperature, and voltage within the defined limits without constraint violations. Overall, the results show that DDPG can provide adaptive battery control while reducing grid dependence and battery energy throughput under changing solar and load conditions.

Keywords: Deep Reinforcement Learning, Deep Deterministic Policy Gradient (DDPG), Rule- Based Energy Management System (RB- EMS), Adaptive Energy Management (AEM), Lithium-Ion Battery.

1. Introduction

1.1 Background

Lithium-ion batteries are employed extensively for energy storage, being adaptable to flexible and efficient energy storage applications. The operation of battery systems in combination with photovoltaic (PV) systems is, however, affected by changes in solar generation and power demand. Thus, efficient energy management is crucial to determine when and how much energy to store, discharge, or supply. Deep Reinforcement Learning (DRL) has been widely applied to battery and photovoltaic energy management, since it can learn control actions from the changing operating conditions. Li et al. (2024), for example, developed a DDPG-based online scheduling approach for photovoltaic systems with battery energy storage using continuous control actions for battery charging and discharging. Real et al. (2024) applied deep reinforcement learning to photovoltaic–battery system optimisation together with load forecasting, while Chen et al. (2024) evaluated different DRL approaches for grid-connected photovoltaic–battery systems under changing electricity-price and photovoltaic conditions.

Other studies have also investigated DDPG for energy-storage control. Shyni and Kowsalya (2024) developed an LSTM–DDPG power-management strategy for a heterogeneous energy-storage system and considered SOC limits under different renewable-generation and load conditions. Cardo-Miota et al. (2025) investigated continuous-action reinforcement learning for residential PV–battery energy management by comparing TD3 and DDPG. More recently, Gao et al. (2026) examined building-integrated PV–battery control using a Decision Transformer and compared its performance with a rule-based controller and several deep reinforcement learning approaches, including SAC, PPO, and TD3. These studies show the continued development of learning-based approaches for managing battery storage under changing renewable-generation and electricity-demand conditions.

However, differences remain in the state variables, datasets, battery constraints, control objectives, and evaluation procedures used in existing studies. In particular, there is a need to examine how a DDPG controller performs when battery SOC, temperature, photovoltaic generation, and load demand are considered together and the learned policy is tested under an operating period that was not used during training. In this study, the controller was evaluated using a chronological data arrangement, with July 2012 to March 2013 used for training, April 2013 used for validation, and May–June 2013 reserved as the unseen test period. This arrangement ensured that the controller was tested on a period that was not used during training or validation. This provides a separate period for examining the controller's response to changes in solar generation and residential electricity demand.

The research gap addressed in this study is the need for a DDPG-based adaptive energy-management approach that combines battery SOC, battery temperature, photovoltaic power, and load demand within a continuous-control framework and evaluates the learned policy against a conventional energy-management strategy under the same unseen solar–load conditions. To address this gap, this study develops and evaluates a Deep Reinforcement Learning-Based Adaptive Energy Management (AEM) approach using a Deep Deterministic Policy Gradient (DDPG) controller. The operating environment combines NASA POWER solar-resource data for Ekwulobia, Anambra State, Nigeria, with the Ausgrid residential electricity-demand dataset. The DDPG controller is evaluated on the unseen test period and compared with a conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS) under the same operating conditions.

The innovative aspect of the study is the combination of local solar-resource data, residential load demand, four operating-state variables, continuous DDPG control, chronological unseen-period testing, and direct comparison with a conventional rule-based strategy. This approach provides a practical way of assessing adaptive battery and grid energy management under changing photovoltaic and residential-load conditions.

1.2 Problem Statement

The combined use of photovoltaic generation and battery energy storage leads to continuous fluctuations in power generation and load demand, making battery energy management a challenging control problem. The battery shall respond to these changes while maintaining proper operating conditions and avoiding overcharge or overdischarge. Deep reinforcement learning has been applied for the energy management of PV-battery systems; however, most studies are limited to specific operating conditions or predefined control strategies (Li et al., 2024; Chen et al., 2024). One important area of further evaluation is the ability of a controller to adjust its charging and discharging decisions to changes in battery SOC, temperature, photovoltaic generation, and load demand while maintaining the battery within defined operating limits. Therefore, it is relevant to investigate the possibility of a continuous-action reinforcement learning controller to learn proper battery-management decisions as the solar-generation and residential-load conditions change and offer measurable improvements compared to a traditional energy-management strategy.

The problem is addressed in this study by proposing a Deep Reinforcement Learning-Based Adaptive Energy Management for the lithium-ion battery energy storage system using a Deep Deterministic Policy Gradient (DDPG) controller. The controller generates continuous charging and discharging actions of the battery; and is evaluated using an extended solar–load operating dataset, and its performance is compared against a conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS).

1.3 Aim of the Study

The study aims to develop and evaluate a Deep Reinforcement Learning-Based Adaptive Energy Management approach for lithium-ion battery energy storage systems.

1.4 Contribution of the Study

The main contribution of this study is the development and evaluation of a DDPG-based adaptive energy-management framework for continuous battery control. The framework uses battery SOC, temperature, photovoltaic power, and load demand as operating-state variables and generates continuous charging and discharging actions. The study further evaluates the learned controller under an extended solar–load operating environment developed from NASA POWER solar-resource data and Ausgrid residential electricity-demand data. The DDPG controller is evaluated using battery operating behaviour, energy utilisation, grid interaction, and operating constraints, and its performance is compared with a conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS).

The study therefore provides an experimental basis for assessing the use of deep reinforcement learning for adaptive lithium-ion battery energy management under changing solar-generation and residential-load conditions.

2. RELATED WORK

2.1 Deep Reinforcement Learning for Battery Energy Management

Deep Reinforcement Learning (DRL) has been applied more and more in the energy management domain since it allows a controller to learn control decisions from the operating conditions of an energy system. This is especially useful in photovoltaic-battery systems where solar production and electricity demand change throughout the day.

Fan and Wang (2023) investigated DRL for community battery scheduling under uncertainties in photovoltaic generation, load demand, and electricity prices. Their work showed how reinforcement learning can be used to adapt battery scheduling decisions to changing energy conditions.

Li et al. (2024) developed a DDPG-based online scheduling approach for photovoltaic systems with battery energy storage. Their method used continuous state and action spaces to determine battery charging and discharging decisions. The study is particularly relevant to the present work because DDPG allows the controller to produce continuous battery-control actions rather than selecting from a small number of predefined actions.

Xu et al. (2023) investigated reinforcement-learning-based optimisation of grid-connected residential PV–battery systems and examined the use of DRL algorithms, including DDPG, for energy-management applications. Their work demonstrated the application of learning-based control to PV–battery operation under changing generation and demand conditions.

Chen et al. (2024) evaluated Deep Q-Network (DQN), Deep Deterministic Policy Gradient (DDPG), and Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithms for grid-connected photovoltaic–battery energy management under dynamic operating conditions. Their results showed that DRL can adapt battery operation to variations in the operating environment, while also highlighting differences between the algorithms in terms of their control performance.

Real et al. (2024) investigated the optimisation of a photovoltaic–battery system using deep reinforcement learning together with load forecasting. The study shows the growing interest in using information about future operating conditions to improve battery energy-management decisions.

Hussain and Musilek (2024) also examined DRL-based energy management for buildings equipped with photovoltaic generation and energy storage. Their work further demonstrates the use of learning-based approaches for energy-management problems involving variable renewable generation and demand. More recent studies have continued to extend DRL-based energy management towards practical and safety-related issues.

Yang et al. (2025), for example, proposed a physics-shielded DRL approach that incorporated battery operating limits and battery-health considerations into microgrid energy management. Their work highlights the importance of maintaining battery operation within safe limits when applying reinforcement learning to energy-storage systems.

2.2 DDPG-Based Energy-Storage Control

DDPG is particularly suitable for energy-storage control because it supports continuous action spaces. This allows a controller to determine the magnitude and direction of battery power rather than choosing only between discrete actions such as charging, discharging, or idle.

Li et al. (2024) demonstrated this application through continuous PV–battery scheduling and charging/discharging control. Their work provides a direct example of using DDPG for photovoltaic energy-storage management.

Shyni and Kowsalya (2024) developed an LSTM–DDPG power-management strategy for a heterogeneous energy-storage system consisting of a battery and supercapacitor in a standalone DC microgrid. DDPG was used for power-allocation decisions, while LSTM was used to extract information from observed system states. The study also considered SOC limits to reduce excessive charging and discharging and evaluated the controller under different solar, wind, and load conditions.

Rezaeimozafar et al. (2024) investigated a hybrid heuristic–reinforcement-learning approach for real-time control of residential behind-the-meter PV–battery systems. This work is relevant to the present study because it considers battery operation in a residential energy-management setting rather than treating battery control only as an isolated optimisation problem.

Other studies have compared DDPG with alternative reinforcement-learning algorithms for PV–battery energy management. Chen et al. (2024) compared Deep Q-Network (DQN), Deep Deterministic Policy Gradient (DDPG), and Twin Delayed Deep Deterministic Policy Gradient (TD3) for grid-connected PV–battery energy management, while Cardo-Miota et al. (2025) compared TD3 and DDPG for residential PV–battery energy management. These studies show that different reinforcement-learning algorithms can be applied to continuous battery-control problems, with their performance depending on the operating conditions and control objectives considered.

More recently, Gao et al. (2026) investigated the control of building-integrated photovoltaic and battery systems using a Decision Transformer and compared its performance with a rule-based controller and several deep reinforcement learning methods, including Soft Actor-Critic (SAC), Proximal Policy Optimisation (PPO), and Twin Delayed Deep Deterministic Policy Gradient (TD3). Their study considered changing photovoltaic generation and building electricity demand and demonstrated the application of advanced learning-based controllers to PV battery energy management. The authors reported an 18.5% reduction in operating cost compared with the rule-based controller and consistent performance under different initial battery conditions. The study also discussed the challenges associated with stability and long-horizon optimisation in conventional DRL approaches.

Overall, previous studies demonstrate the potential of DRL, including continuous-action methods such as DDPG, for PV–battery energy management under changing generation and demand. However, existing studies differ in their choice of operating-state variables, data sources, control objectives, battery constraints, and evaluation methods. Building on this body of work, the present study develops a DDPG-based adaptive energy-management approach that uses battery SOC, temperature, photovoltaic power, and load demand as the controller state variables. The performance of the controller is benchmarked against a conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS) using an extended solar–load operating environment developed from NASA POWER solar-resource data for Ekwulobia, Aguata Local Government Area, Anambra State, Nigeria, and the Ausgrid residential electricity-demand dataset.

2.3 Research Gap

The literature reviewed shows that the application of deep reinforcement learning to photovoltaic-battery energy management is increasing, including DDPG-based continuous control, scheduling, and energy-storage optimisation. Previous studies have addressed various objectives and operating conditions, including electricity price optimisation, load forecasting, grid interaction, heterogeneous energy storage systems, and integrated energy systems (Li et al., 2024; Chen et al., 2024; Real et al., 2024; Shyni & Kowsalya, 2024; Gao et al., 2026). However, some differences still exist in the operating-state variables, data sources, control objectives, battery constraints, and evaluation methods adopted in these studies. Specifically, an adaptive energy-management scheme that jointly considers battery SOC, temperature, photovoltaic power, and load demand in determining continuous battery-control actions under varying solar generation and residential load scenarios is needed.

This paper addresses this gap by proposing and testing a DDPG-based adaptive energy-management strategy for a lithium-ion battery energy storage system. The controller is evaluated in an extended solar-load operating environment generated from NASA POWER solar-resource data for Ekwulobia, Anambra State, Nigeria, and the Ausgrid residential electricity-demand dataset. Moreover, its performance is compared with a conventional PV-priority Rule-Based Energy Management System (RB-EMS) under identical operating conditions. The comparison is in terms of battery energy throughput, grid interaction, SOC behaviour, and compliance with defined battery operating constraints.

3. METHODOLOGY

3.1 Research Framework

The proposed energy-management framework consists of a photovoltaic (PV) generation system, lithium-ion battery energy storage system (BESS), electrical load, and a Deep Deterministic Policy Gradient (DDPG) controller. The PV–battery system provides the operating environment in which the DDPG agent learns battery charging and discharging decisions under changing solar generation and electricity demand. The operating environment was developed using NASA POWER solar-resource data for Ekwulobia, Anambra State, Nigeria**,** combined with the Ausgrid 2012–2013 residential electricity-demand dataset. The two datasets were processed, temporally aligned, and resampled to 10-minute intervals, producing a final solar–load operating dataset containing 52,560 observations. The dataset provides the changing photovoltaic generation and residential load conditions required for evaluating the adaptive controller.

The DDPG controller uses Battery State of Charge (SOC), battery temperature, photovoltaic power, and load demand as the state variables. Based on the current state of the system, the controller generates a continuous battery-control action that determines the required charging or discharging response. The controller operates through a closed-loop interaction with the simulated environment, where each action produces a new battery state and a corresponding reward. The trained DDPG controller is evaluated on the unseen testing period and compared with a conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS) under the same operating conditions. The comparison is used to assess battery energy throughput, grid interaction, SOC behaviour, and compliance with the defined battery operating constraints.

A one-day preliminary solar–battery dataset obtained from an on-site installation in Ekwulobia was used during the initial development and debugging of the battery-control environment. These preliminary observations were not included in the final DDPG training, validation, or unseen-test evaluation.

3.2 Photovoltaic–Battery Operating Model

The interaction between photovoltaic generation, load demand, and battery operation was represented using physically constrained mathematical relationships to maintain realistic battery behaviour throughout the DDPG simulation. These relationships were used to determine battery power, battery current, and the evolution of Battery State of Charge (SOC) at each time step.

3.2.1 PV–Load Power Balance

The battery power at each time step was determined from the difference between photovoltaic generation and load demand:

$$P_{bat}(t) = P_{PV}(t) - P_{load}(t) \quad \ldots\ldots\ldots\ldots (1)$$

Where P_{PV} is the photovoltaic power, P_{load} is the load demand, and P_{bat} is the resulting battery power. When $P_{PV} > P_{load}$, the available surplus is used for battery charging. When $P_{load} > P_{PV}$, the battery supplies part of the load demand and operates in discharge mode. The remaining demand, where applicable, is supplied by the grid.

3.2.2 Battery Power–Current Relationship

Battery current was derived from the electrical relationship between power and battery voltage:

$$I_{bat(t)} = \frac{P_{bat(t)}}{V_{bat(t)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where I_{bat} is the battery current, and V_{bat} is the battery voltage. This relationship preserves consistency between power, voltage, and current throughout the simulation.

3.2.3 State of Charge (SOC) Update

The Battery State of Charge (SOC) was updated using a Coulomb-counting approximation:

$$SOC(t+1) = SOC(t) + \frac{\eta_c I_{bat(t)\Delta t}}{C_{Ah}} \ldots\ldots (3)$$

Where η_c is the charging efficiency, Δt is the sampling interval, and C_{Ah} is the nominal battery capacity in ampere-hours. The SOC was constrained within the defined operating limits of the battery. This formulation provides a consistent representation of SOC changes during battery charging and discharging and supports the evaluation of the DDPG controller under changing photovoltaic generation and load demand.

3.3 Reinforcement Learning Problem Formulation

3.3.1 Markov Decision Process (MDP)

The battery energy-management problem was formulated as a Markov Decision Process (MDP), represented by the tuple:

$$M = (S, A, R, \gamma)$$

Where S represents the state space, A represents the action space, R denotes the reward function, and γ is the discount factor used to determine the importance of future rewards. At each time step, the DDPG agent receives information about the current battery and energy-system conditions and uses this information to select a continuous control action. The environment then responds to the selected action by updating the battery and energy-system state and providing a reward. Through this repeated interaction, the agent learns a control policy that seeks to maximise the cumulative reward while maintaining the battery within the defined operating conditions.
The main components of the MDP are defined as follows:

- **State:** The state consists of battery SOC, battery temperature, photovoltaic power generation, and load demand.
- **Action:** The action is a continuous control signal that determines the battery charging or discharging response.
- **Reward:** The reward provides feedback to the agent based on the resulting battery and energy-system operation and guides the agent towards desirable operating behaviour.
- **State transition:** The state transition represents the change in battery and energy-system conditions resulting from the selected control action.

The MDP formulation provides the mathematical basis for the DDPG controller and enables the battery energy-management problem to be treated as a sequential decision-making task.

3.3.2 State Space

The state vector provided to the DDPG controller was defined as:

$$s_t = [SOC(t), T(t), P_{PV}(t), P_{load}(t)] \ldots\ldots (4)$$

Where $SOC(t)$ represents the Battery State of Charge, $T(t)$ is the battery temperature, $P_{pv}(t)$ is the photovoltaic power generation, and $P_{load}d(t)$ is the electrical load demand. Together, these variables describe the battery condition and the prevailing generation and load conditions. The state is updated at each time step and provided to the DDPG agent, allowing the controller to adjust its action as SOC, temperature, solar generation, and load demand change during operation.

3.3.3 Action Space

The DDPG controller uses a continuous action space because battery charging and discharging require different levels of control rather than simply selecting between fixed operating states. The control action at time t is expressed as:

$$a_t = \mu\,(s_t \mid \theta^{\mu}) \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Where a_t represents the continuous battery control action, s_t is the current system state, and $\mu(\cdot/\theta^{\mu})$ represents the actor policy parameterised by θ^{μ}. The action is constrained within the defined battery operating limits to prevent excessive charging or discharging.
This continuous-action formulation allows the DDPG agent to adjust the battery control level according to the current SOC, battery temperature, photovoltaic generation, and load demand.

3.3.4 Bellman Optimality and Reward Function

The DDPG control problem is based on the Bellman principle, which relates the reward received from the current action to the expected return from future actions. The optimal action-value function can be expressed as:

$$Q^{*}(s, a) = R(s, a) + \gamma Q^{*}(s', \mu^{*}(s')) \ldots\ldots (6)$$

Where $Q*(s, a)$ represents the optimal action-value function, $R(s, a)$ is the immediate reward obtained by taking action a in state s, γ is the discount factor ($0 < \gamma < 1$), s' represents the next system state, and $\mu^{*}(s')$ represents the action generated by the optimal actor policy for the next state. The equation shows that the value of the current action depends on both the immediate reward and the estimated future return obtained by following the learned policy.

For the battery energy-management problem, the reward function was designed to guide the DDPG agent towards efficient and safe battery operation. The formulation used in the study is:

$$R_t = \alpha\eta_t - \beta D_t - \gamma T_t \ldots\ldots\ldots\ldots\ldots (7)$$

Where η_t represents the charging or discharging efficiency, D_t represents the degradation or capacity-fade measure, and T_t represents the thermal condition at time t. The weighting factors α, β, and γ determine the relative contribution of efficiency, degradation, and thermal conditions to the overall reward.

The reward is provided to the DDPG agent after each control action. Through repeated interaction with the operating environment, the agent learns to favour actions that improve battery operation while reducing undesirable conditions such as excessive degradation and thermal stress. This reward formulation therefore links the learning process to the main operating requirements of the battery energy-management system.

3.3.5 State Transition

The state transition describes how the battery and energy-system conditions change after the DDPG controller applies a control action. At each time step, the agent observes the current state s_t, selects an action a_t, receives the corresponding reward r_t, and the environment evolves to the next state s_{t+1}.
This interaction can be represented as:

$$S_{t+1} = f(s_t, a_t, w_t) \ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

Where $f(\cdot)$ represents the system dynamics and w_t represents the prevailing operating conditions, including photovoltaic generation and load demand. The resulting state is returned to the agent, allowing it to continuously adapt its control policy during the learning process.

3.4 Continuous-Action DDPG Framework for Battery Control

The Deep Deterministic Policy Gradient (DDPG) algorithm was employed to provide continuous battery-control actions. DDPG is an actor–critic, model-free reinforcement-learning algorithm designed for continuous action spaces (Lillicrap et al., 2015). The DDPG framework employs target actor and critic networks to improve learning stability. The target value used for critic learning is:

$$y_t = R_t + \gamma Q_{target}(s_{t+1}, \mu_{target}(s_{t+1})) \ldots\ldots \quad (9)$$

Where R_t is the immediate reward and γ is the discount factor. This formulation follows the DDPG target-value structure described in the original algorithm and implemented in the proposed framework. The use of DDPG enables the controller to learn continuous battery-control policies that respond to changing PV generation, load demand, SOC, and temperature conditions.

3.5 DDPG Training and Evaluation Procedure

The DDPG controller was trained by allowing the agent to interact repeatedly with the photovoltaic–battery operating environment. The final solar–load dataset was divided into training, validation, and unseen testing periods so that the time sequence of the data was maintained and future observations were not used during training. Data from July 2012 to March 2013 were used for training, April 2013 was used for validation, while May to June 2013 were reserved for the unseen test period.

We implemented the DDPG controller with the Stable-Baselines3 framework, and it was trained for 250000 timesteps. Main training settings were: learning rate 0.0001, replay-buffer size 100,000, batch size 256, learning-start threshold 1,000 timesteps, discount factor $\gamma = 0.99$, and target-network update coefficient $\tau = 0.005$. To encourage exploration of different continuous control actions, Gaussian action noise with a standard deviation of 0.10 was added during training. Experience replay and target networks have also been used to increase the stability of the learning process.

After training, the learned DDPG policy was evaluated using the unseen May–June 2013 test period. The evaluation examined how the controller responded to changes in photovoltaic generation and residential load demand, with particular attention to battery control behaviour, energy utilisation, grid interaction, and compliance with the defined battery operating constraints. For comparison, the trained DDPG controller was evaluated against a conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS) using the same unseen test conditions. The comparison considered battery energy throughput, grid interaction, SOC behaviour, and compliance with the defined battery operating limits. The detailed results of the DDPG evaluation and its comparison with the RB-EMS are presented in Section 4**.**

3.6 Performance Evaluation Metrics

The performance of the DDPG controller was evaluated using measures that reflect its battery-control behaviour, energy utilisation, grid interaction, and compliance with the defined operating constraints. The evaluation was carried out on the unseen May–June 2013 test period to assess the performance of the trained controller under operating conditions that were not used during training. The main performance measures included battery charging energy, battery discharging energy, total battery energy throughput, grid import, and grid export. These measures were used to examine how the controller managed energy between photovoltaic generation, the battery, the electrical load, and the grid.

Battery operating behaviour was further assessed using the SOC profile, battery temperature, and battery voltage. The observed ranges of these variables were compared with the defined operating limits to determine whether the controller maintained the battery within acceptable conditions during the test period. Constraint violations, where applicable, were also recorded.

The DDPG controller was additionally compared with the conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS) under the same unseen test conditions. The comparison focused on battery energy throughput and grid interaction, while SOC, temperature, voltage, and constraint compliance were used to assess battery operating behaviour and safety. These measures provide a basis for evaluating the effectiveness of the proposed DDPG-based energy-management strategy and for determining its performance relative to the conventional RB-EMS. The detailed results are presented and discussed in Section 4.

4. RESULTS AND DISCUSSION

4.1 Overview of Results and Discussion

This section presents and discusses the results obtained from the implementation and evaluation of the proposed Deep Reinforcement Learning-Based Adaptive Energy Management approach using the Deep Deterministic Policy Gradient (DDPG) controller. The evaluation was carried out using the final solar–load operating dataset developed from NASA POWER solar-resource data for Ekwulobia, Anambra State, Nigeria, and the Ausgrid residential electricity-demand dataset.

The results are presented in relation to the objectives of the study and cover the DDPG training performance, performance on the unseen test period, battery energy utilisation, grid interaction, battery operating behaviour, constraint compliance, and comparison with a conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS). The DDPG controller and RB-EMS were evaluated under the same PV generation, residential load, battery configuration, and operating constraints, allowing their performance to be compared under identical operating conditions.

4.2 DDPG-Based Adaptive Energy Management Results

During the unseen May–June 2013 test period, the trained DDPG controller was evaluated under changing photovoltaic generation and residential electricity demand. The test period contained 8,784 observations at 10-minute intervals, corresponding to 8,783 sequential state transitions used for DDPG evaluation.

The analysis focused on the controller's battery charging and discharging behaviour, energy utilisation, grid interaction, SOC behaviour, and compliance with the defined battery operating constraints. The DDPG results were subsequently compared with those of the conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS) under the same operating conditions.

4.2.1 Solar and Load Conditions During the Unseen Test Period

Figure 1 shows the PV generation and residential load demand during the unseen May–June 2013 test period. PV generation varied from about 0 kW during non-solar periods to approximately 7.4 kW at periods of high solar availability, while load demand generally remained below 2 kW.
The variation between PV generation and load demand created changing operating conditions for the battery. When PV generation was higher than the load, excess energy was available for battery charging, whereas periods of lower PV generation relative to the load increased the need for battery discharge and grid supply. The PV profile was derived from NASA POWER data for Ekwulobia, Anambra State, while the load profile was obtained from the Ausgrid residential General Consumption dataset. These changing solar-generation and load conditions formed the basis for evaluating the adaptive response of the DDPG controller during the unseen test period.

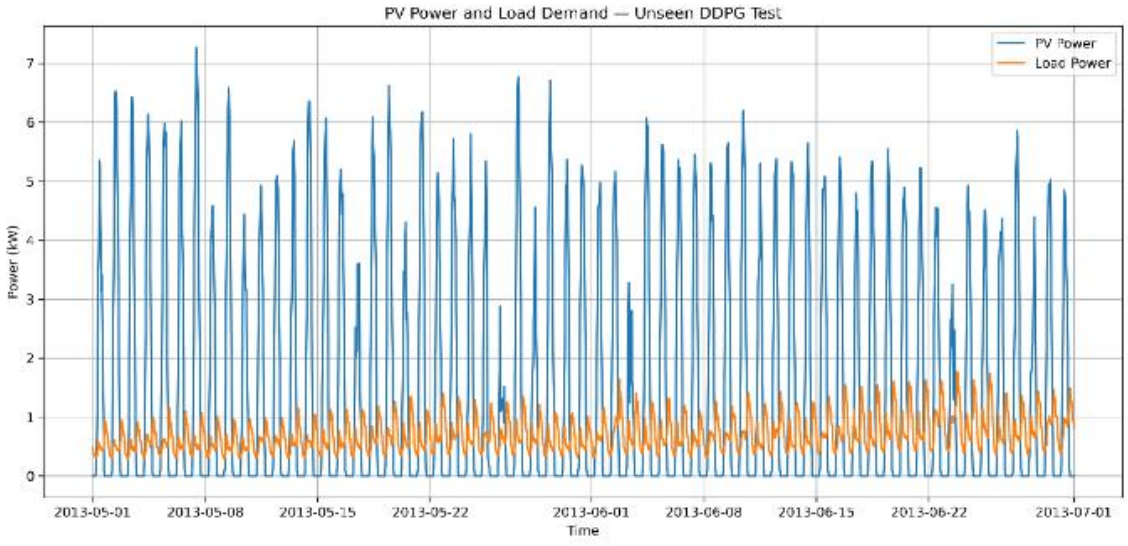

Figure 1: Photovoltaic Generation and Residential Load Demand during the Unseen May–June 2013 DDPG Test Period

4.2.2 DDPG Battery Operating Performance

We examined the battery response during the unseen May–June 2013 test period using the SOC, temperature, and charging/discharging profiles shown in Figures 2(a)–2(c). Table 1 summarises the main battery operating conditions.

Figure 2(a**)** shows the battery SOC throughout the test period. The SOC repeatedly increased and decreased in response to changes in PV generation and electricity demand. It reached a maximum of 95.00% and a minimum of 21.30%, remaining within the defined 20–95% operating range throughout the evaluation. The repeated charging and discharging cycles reflect the changing balance between PV generation and residential load demand.

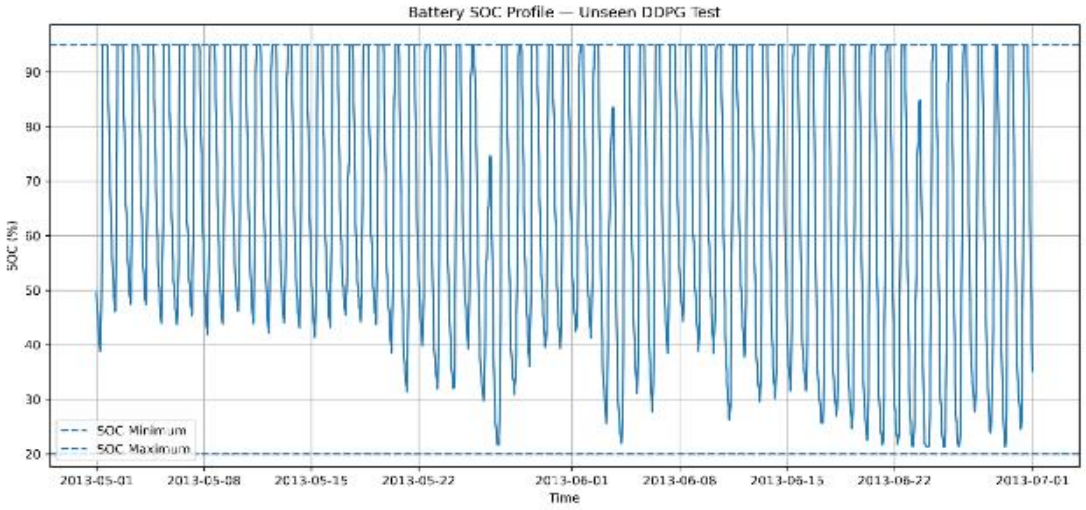

Figure 2(a): **Battery State of Charge during the Unseen May–June 2013 DDPG Test Period**

Figure 2(b) shows the battery temperature during the same evaluation period. The temperature varies continuously, with values ranging from 21.97°C to 31.51°C. The repeated temperature fluctuations correspond to changes in battery operating conditions during charging and discharging. Higher temperature values occur at different points throughout the test period, while the temperature subsequently decreases as the operating conditions change.

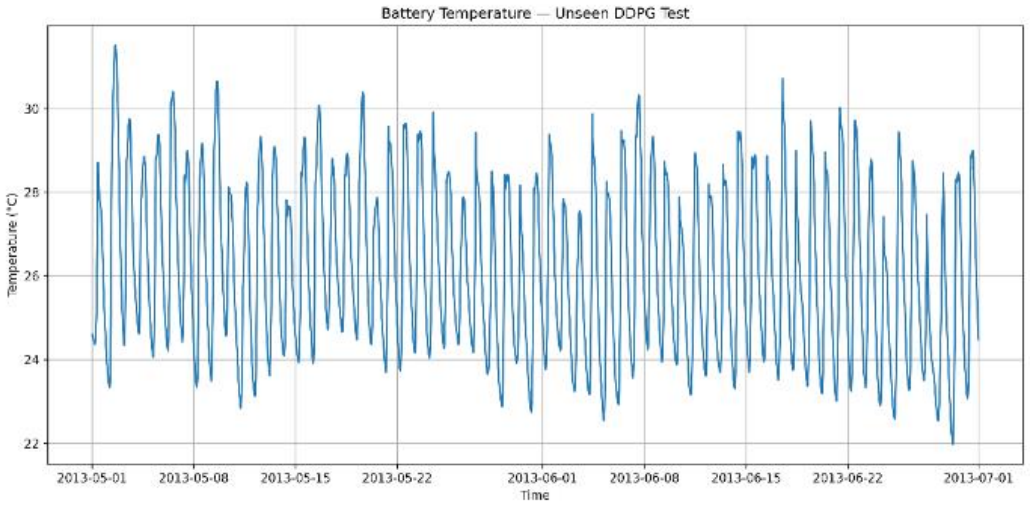

Figure 2(b): Battery Temperature during the Unseen May–June 2013 DDPG Test Period

Importantly, the observed temperature remained within the defined operating limits, and no temperature constraint violation was recorded during the unseen test. The temperature profile therefore indicates that the learned control policy operated without producing excessive thermal conditions under the evaluated solar–battery operating environment.

Figure 2(c) shows the battery charging and discharging power under DDPG control during the unseen test period. With the control convention used in this study, positive power represents battery charging, while negative power represents discharging. The figure shows repeated transitions between charging and discharging as PV generation and electricity demand change. Charging power reached the imposed maximum of 5.00 kW at several points, while discharging power reached approximately −1.77 kW. The alternating positive and negative power patterns reflect periods when available PV energy was sufficient to charge the battery and periods when battery energy was required to support the load. Overall, the results show that the DDPG controller adjusted both the direction and magnitude of battery power in response to the changing operating conditions.

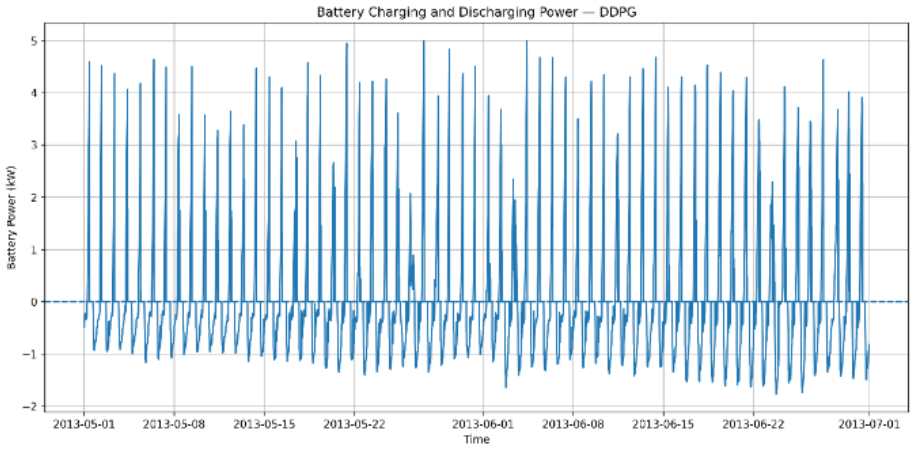

Figure 2(c): **Battery Charging and Discharging Power under DDPG Control during the Unseen Test Period**

Table 1 summarises the principal battery operating conditions recorded during the unseen DDPG test. The results show that the battery SOC remained between 21.30% and 95.00%, thereby remaining within the defined 20–95% operating range. The maximum observed SOC of 95% reflects the controller's operating constraint during the test and does not imply that the physical $LiFePO_4$ battery cannot be charged to 100%. The battery temperature varied from 21.97°C to 31.51°C, indicating moderate thermal variation throughout the evaluation. The battery voltage remained between 49.99 V and 53.83 V, while the battery power ranged from −1.77 kW to 5.00 kW, consistent with the defined maximum charging power.

Table 1 also shows that the DDPG control action ranged from −0.997 to 1.000, with a mean action of approximately 0.096. The wide action range indicates that the controller made use of the continuous action space, generating both charging and discharging commands of different magnitudes according to the prevailing operating conditions. The positive mean action should not, by itself, be interpreted as an overall predominance of charging because the action distribution and the resulting battery power also depend on the operating conditions and the controller's response throughout the test period.

Table 1: Battery Operating Conditions under DDPG Control

Parameter	Value	Unit
Initial SOC	49.43	%
Final SOC	35.26	%
Minimum SOC	21.30	%
Maximum SOC	95.00	%
Minimum Temperature	21.97	°C
Maximum Temperature	31.51	°C
Minimum Voltage	49.99	V
Maximum Voltage	53.83	V
Minimum Battery Power	−1.77	kW
Maximum Battery Power	5.00	kW
Minimum DDPG Action	−0.997	
Maximum DDPG Action	1.000	
Mean DDPG Action	0.096	

Overall, the values presented in Table 1 indicate that the DDPG controller maintained the battery within the defined operating limits while continuously adjusting battery power in response to changing solar generation and load demand. In particular, the absence of recorded SOC, temperature, voltage, and battery-power constraint violations during the unseen evaluation provides evidence that the learned control policy operated within the specified constraints under the conditions tested.

4.2.3 DDPG Control Action Behaviour

Figure 3 shows the DDPG control actions during the unseen May–June 2013 test period. The actions vary continuously between approximately −1 and +1, with positive values representing charging and negative values representing discharging. The figure shows frequent changes in both the direction and magnitude of the control action as the PV generation, load demand, and battery conditions change. Several actions reach close to +1.0 and −1.0, while many others fall between these limits. Values close to zero indicate periods when only a small battery-control action was required.

Overall, the results demonstrate the continuous nature of the DDPG control strategy, allowing the controller to vary the battery-control level rather than relying only on fixed switching decisions.

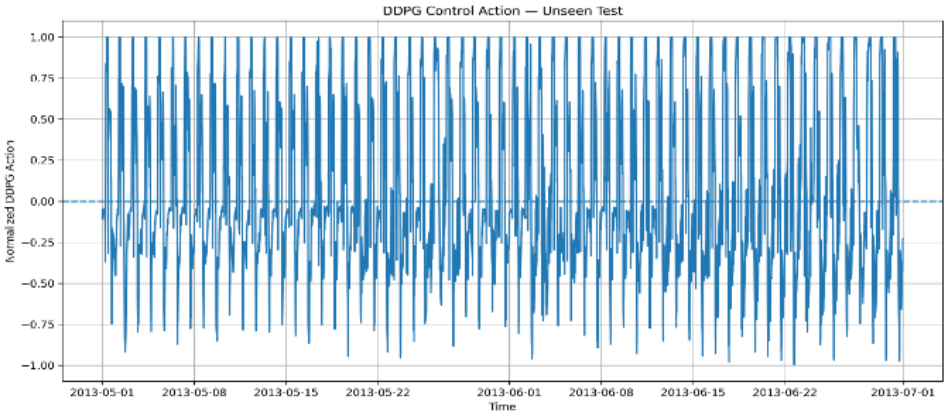

Figure 3: DDPG Control Actions during the Unseen May–June 2013 Test Period

The quantitative action distribution is presented in Table 2. Out of the recorded test intervals, 3,848 contained charging actions, 4,928 contained discharging actions, and 7 were classified as idle actions. However, the actual battery response occurred during 1,830 charging intervals and 4,488 discharging intervals. This difference shows that a commanded action does not necessarily result in battery charging or discharging after the operating conditions and battery constraints are applied. Table 2 also shows 4,816 load-deficit intervals and 3,967 PV-surplus intervals. Actual discharging occurred during 4,488 of the load-deficit intervals, giving a discharge utilisation of 93.19%, while actual charging occurred during 1,830 PV-surplus intervals, giving a charging utilisation of 46.13%.

Overall, Figure 3 and Table 2 show that the DDPG controller continuously adjusted its control action according to the changing solar–load conditions, producing both charging and discharging responses rather than relying on a fixed switching rule.

Table 2: DDPG Control Action Utilisation during the Unseen May–June 2013 Test Period

Metric	Value	Unit
Charging DDPG actions	3848.0	intervals
Discharging DDPG actions	4928.0	intervals
Idle DDPG actions	7.0	intervals
Actual charging intervals	1830.0	intervals
Actual discharging intervals	4488.0	intervals
Load-deficit intervals	4816.0	intervals
PV-surplus intervals	3967.0	intervals
Discharge utilisation	93.18936877076412	%
Charging utilisation	46.13057726241493	%

4.2.4 Grid Interaction and Energy Management Performance

The grid interaction and energy management behaviour of the DDPG controller is presented in Figure 4(a) and Figure 4(b), while the main quantitative results are summarised in Table 3. These results show how the controller managed the available PV energy, battery storage, and interaction with the electricity grid during the unseen May–June 2013 test period.

Figure 4(a) shows the variation in grid power exchange throughout the test period. Positive values represent periods of grid export, while negative values represent grid import. The figure shows a repeated pattern of grid export during periods of higher PV availability and grid import when the available PV generation and battery support were insufficient to meet the load. The repeated transitions between import and export demonstrate that grid interaction changed with the prevailing solar–load conditions rather than remaining constant throughout the test period.

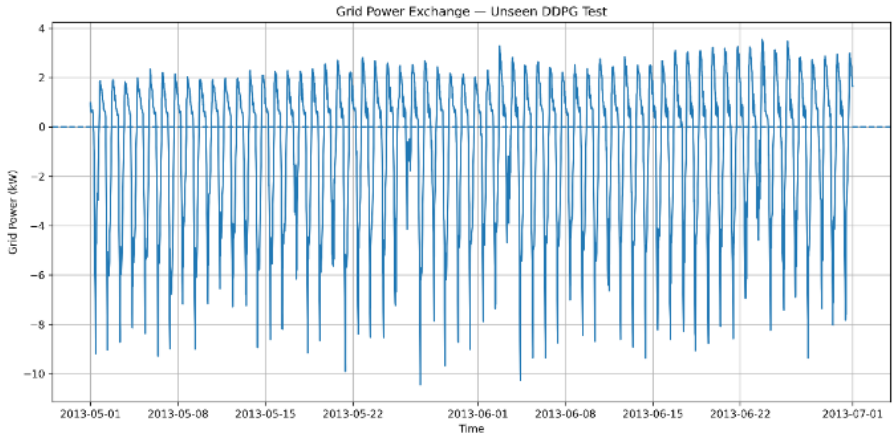

Figure 4(a): Grid Power Exchange during the Unseen May–June 2013 DDPG Test Period

Figure 4(b) provides a broader view of the energy flows during the same test period. The figure shows that the system generated approximately 2,192.48 kWh of PV energy, while the total load demand was approximately 1,020.20 kWh. The battery absorbed 575.34 kWh of energy through charging and supplied approximately 521.35 kWh through discharging. The remaining energy was exchanged with the grid, resulting in approximately 1,096.86 kWh of grid import and 2,323.13 kWh of grid export.

The relatively large grid-export value shown in Figure 4(b) is consistent with the PV generation profile, where substantial periods of PV generation exceeded the immediate load and the battery's charging requirements. During periods of insufficient PV generation, the battery provided energy support, and the grid supplied the remaining demand.

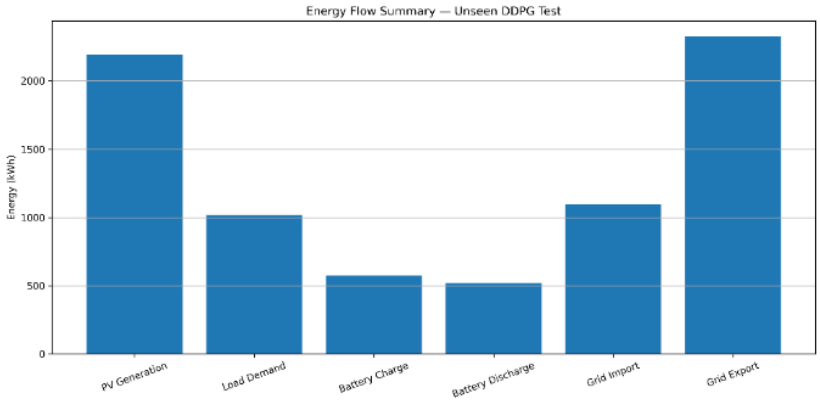

Figure 4(b): Energy Flow Summary during the Unseen May–June 2013 DDPG Test Period

Table 3 summarises the overall quantitative performance of the DDPG controller. The results are based on 8,783 recorded test transitions. The battery charging and discharging energies give a total battery energy throughput of approximately 1,096.69 kWh during the evaluation period. The controller therefore used the battery repeatedly for energy shifting while maintaining the defined operating conditions.

The battery SOC remained between 21.30% and 95.00%, while the battery temperature varied between 21.97°C and 31.51°C. The battery voltage remained between 49.99 V and 53.83 V, and no SOC, temperature, voltage, or battery-power constraint violations were recorded during the test. These results indicate that the energy-management actions shown in Figure 4(a) and Figure 4(b) were performed within the operating limits defined for the DDPG environment.

Overall, Figure 4(a), Figure 4(b), and Table 3 show that the DDPG controller managed energy exchange among the PV system, battery, load, and grid throughout the unseen test period. The controller responded to periods of PV surplus by allowing energy to be directed towards battery charging and grid export, while during periods of PV deficit, battery discharge and grid import supported the load.

Table 3: Overall DDPG Performance during the Unseen May–June 2013 Test Period

Metric	Value	Unit
Evaluation transitions	8783.0	transitions
Total reward	-11065.462777379144	reward units
PV energy	2192.479758	kWh
Load energy	1020.2045677777776	kWh
Battery charging energy	575.3438713567974	kWh
Battery discharging energy	521.3482941550651	kWh
Grid import energy	1096.8622436883984	kWh
Grid export energy	2323.1330111123525	kWh

4.2.5 DDPG Constraint Compliance

The constraint compliance of the DDPG controller during the unseen May–June 2013 test period is presented in Table 4. The table compares the defined operating limits with the minimum and maximum values observed during the evaluation. For the SOC, the defined operating range was 20–95%. The observed SOC varied from 21.30% to 95.00%, with zero recorded violations. This shows that the controller maintained the battery within the specified SOC operating range during the test. The battery temperature remained between 21.97°C and 31.51°C, compared with the defined range of 15–40°C. Similarly, the battery voltage varied from 49.99 V to 53.83 V, remaining within the specified 48–58.4 V range.

For battery power, the permitted range was −5.00 to +5.00 kW. The observed battery power ranged from approximately −1.77 kW to +5.00 kW, with no recorded violations. The maximum observed value reached the defined charging-power limit of 5.00 kW, but did not exceed it.

Table 4: Constraint Validation of the DDPG Controller during the Unseen May–June 2013 Test Period

Constraint	Allowed range	Minimum observed	Maximum observed	Violations	Status
SOC	20–95%	21.299350565199777	95.0	0	PASS
Temperature	15–40 °C	21.97178426615108	31.510498186 78419	0	PASS
Voltage	48–58.4 V	49.98605084480705	53.828935546 875	0	PASS
Battery power	±5.00 kW	-1.7725666666666666	5.0	0	PASS

Overall, Table 4 shows zero recorded violations for SOC, temperature, voltage, and battery-power constraints during the unseen test period. Therefore, under the evaluated solar–load conditions, the DDPG controller maintained the battery within the predefined operating limits.

4.2.6 Monthly DDPG Performance

Figure 5 presents the monthly energy performance of the DDPG controller during May and June 2013, identified as the unseen test period. The results indicate that PV generation decreased from May to June, while load demand increased.

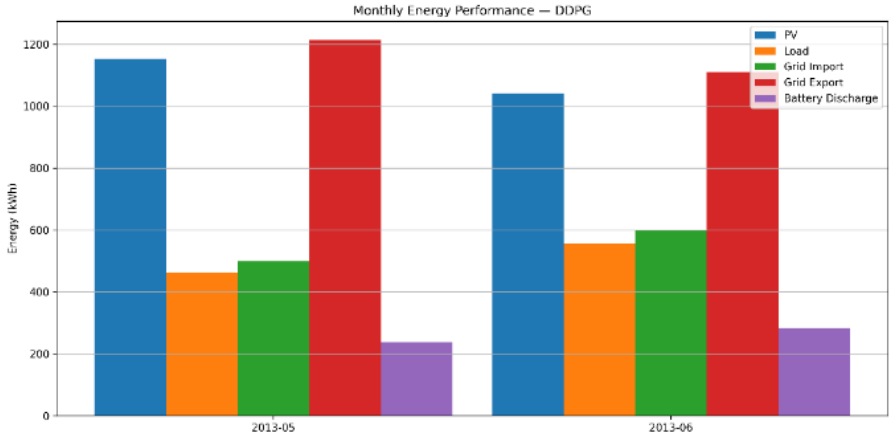

Figure 5: Monthly Energy Performance of the DDPG Controller during the Unseen Test Period

Over the same period, grid import increased and grid export decreased. Battery discharge also increased, as indicated by the corresponding bars in the figure. These trends suggest that the reduction in PV generation, combined with the increase in load demand, contributed to a greater reliance on grid electricity during June. Overall, the results illustrate the variation in energy flows under changing generation and demand conditions during the test period.

4.3 Comparison of Deep Deterministic Policy Gradient (DDPG)-Based Adaptive Control and Conventional Photovoltaic (PV)-Priority Rule-Based Energy Management System (RB-EMS)

To further evaluate the performance of the proposed DDPG-based adaptive controller, its results were compared with those of a conventional PV-priority Rule-Based Energy Management System (RB-EMS). Both controllers were evaluated using the same unseen May–June 2013 test period, with the same 7.56 kW PV system, 15 kWh $LiFePO_4$ battery, 51.2 V nominal battery voltage, 10-minute control interval, PV generation profile, and residential load profile. The comparison considers the main aspects of battery energy management, including grid import, grid export, battery charging, battery discharging, battery energy throughput, SOC behaviour, battery temperature, battery voltage, and compliance with the defined battery operating constraints. This allows the behaviour of the learned DDPG controller to be examined alongside a conventional control strategy under the same operating conditions.

The RB-EMS uses predefined PV-priority rules for charging and discharging, whereas the DDPG controller learns its control actions from the defined reward function and operating state. The comparison therefore provides a basis for examining the differences in energy utilisation, grid interaction, battery usage, and constraint compliance between the two control approaches.

4.3.1 Overall DDPG and RB-EMS Performance

Table 5 presents the overall performance of the proposed Deep Deterministic Policy Gradient (DDPG) controller and the conventional Rule-Based Energy Management System (RB-EMS) during the unseen May–June 2013 test period.

Table 5: Overall Performance Comparison of DDPG and Conventional RB-EMS during the Unseen May–June 2013 Test Period

Parameter	DDPG	Conventional RB-EMS
Test transitions	8783	8783
PV energy (kWh)	2192.48	2192.48
Load energy (kWh)	1020.20	1020.20
Battery charging (kWh)	575.34	617.96
Battery discharging (kWh)	521.35	559.81
Battery throughput (kWh)	1096.69	1177.77
Grid import (kWh)	1096.86	1135.32
Grid export (kWh)	2323.13	2365.75
Initial SOC (%)	49.43	49.43
Final SOC (%)	35.26	35.26
Minimum SOC (%)	21.30	20.00
Maximum SOC (%)	95.00	95.00
Minimum temperature (°C)	21.97	21.97
Maximum temperature (°C)	31.51	31.51
Minimum voltage (V)	49.99	49.93
Maximum voltage (V)	53.83	53.83
Minimum battery power (kW)	-1.77	-1.77
Maximum battery power (kW)	5.00	5.00
Discharge utilisation (%)	93.23	94.91
Charging utilisation (%)	46.16	47.54

Both controllers were evaluated using the same 8,783 test transitions, PV generation profile, load profile, battery configuration, and operating constraints. As shown in Table 5, both controllers operated with the same 2,192.48 kWh of PV energy and 1,020.20 kWh of load demand, since the input solar and load conditions were identical. The difference therefore comes from how each controller managed the available energy. The DDPG controller charged the battery with 575.34 kWh and discharged 521.35 kWh, giving a total battery throughput of 1,096.69 kWh. In comparison, the RB-EMS charged the battery with 617.96 kWh and discharged 559.81 kWh, resulting in a higher battery throughput of 1,177.77 kWh. Thus, the DDPG controller used less battery energy over the test

period. The difference is also reflected in grid interaction. The DDPG controller recorded 1,096.86 kWh of grid import, compared with 1,135.32 kWh for the RB-EMS. Grid export was 2,323.13 kWh for DDPG and 2,365.75 kWh for the RB-EMS. The results therefore show that the DDPG controller achieved lower grid import while also using less total battery throughput during the evaluated period.

The battery operating conditions were broadly similar for both controllers. The DDPG SOC ranged from 21.30% to 95.00%, while the RB-EMS ranged from 20.00% to 95.00%. Maximum battery temperature was 31.51°C for DDPG and 31.51°C for RB-EMS, while the maximum battery voltage was 53.83 V for both controllers. The discharge and charging utilisation values were slightly higher for the RB-EMS, at 94.91% and 47.54%, respectively, compared with 93.23% and 46.16% for DDPG.

Overall, Table 5 shows that the two controllers operated under the same solar and load conditions but produced different battery and grid-management patterns. The DDPG controller recorded lower grid import and lower battery energy throughput, while the conventional RB-EMS showed slightly higher battery charging and discharging utilisation.

4.3.2 Comparison of Battery SOC Behaviour

Figure 6 compares the battery SOC profiles produced by the DDPG controller and the conventional RB-EMS throughout the unseen May–June 2013 test period. Both controllers show repeated charging and discharging cycles, with the SOC frequently rising to the 95% upper operating limit and then falling as stored energy is used. The DDPG profile generally follows a similar pattern to the RB-EMS, although differences can be observed at several points in the test period. The DDPG controller recorded a minimum SOC of 21.30%, while the RB-EMS reached the lower limit of 20.00%. Both controllers reached a maximum SOC of 95.00%. Thus, both operated within the defined 20–95% SOC range, with DDPG maintaining a slightly higher minimum SOC.

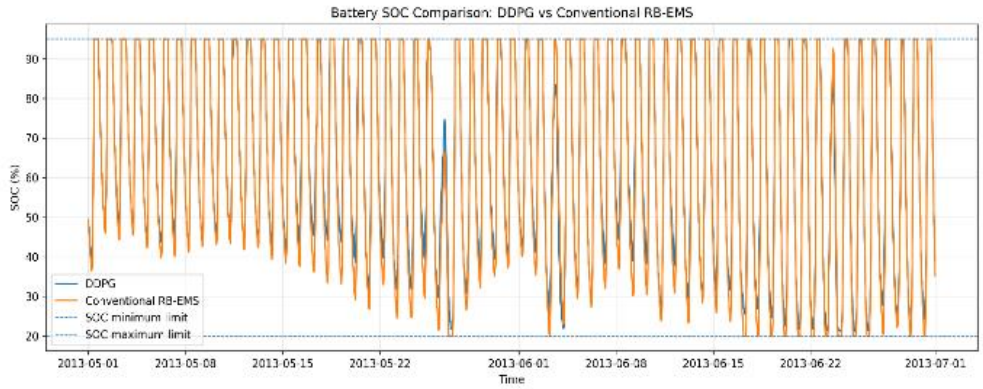

Figure 6: Comparison of Battery SOC under DDPG and Conventional RB-EMS Control

Figure 7 shows the cumulative grid import over the test period. The two curves remain close during much of the evaluation, but the RB-EMS curve finishes above the DDPG curve. The final values are 1,135.32 kWh for RB-EMS and 1,096.86 kWh for DDPG, giving a difference of 38.46 kWh, or 3.39% lower grid import for DDPG. The results indicate that the DDPG controller achieved slightly lower grid dependence over the evaluated test period while responding to the same PV and load conditions as the RB-EMS.

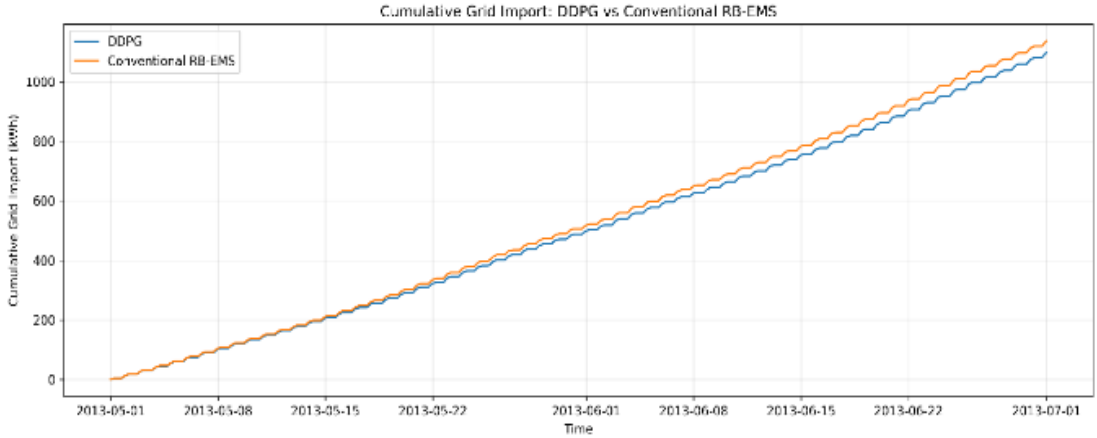

Figure 7: Cumulative Grid Import under DDPG and Conventional RB-EMS Control

4.3.3 Comparison of Battery Energy Throughput

Figure 8 shows the cumulative battery throughput of the DDPG and conventional RB-EMS controllers throughout the unseen May–June 2013 test period. The two curves follow a similar increasing pattern, but the RB-EMS curve gradually moves above the DDPG curve, indicating greater cumulative battery use. By the end of the test period, the DDPG controller recorded 1,096.69 kWh of battery throughput, compared with 1,177.77 kWh for the RB-EMS. This represents a difference of 81.08 kWh, corresponding to 6.88% lower battery throughput under DDPG.

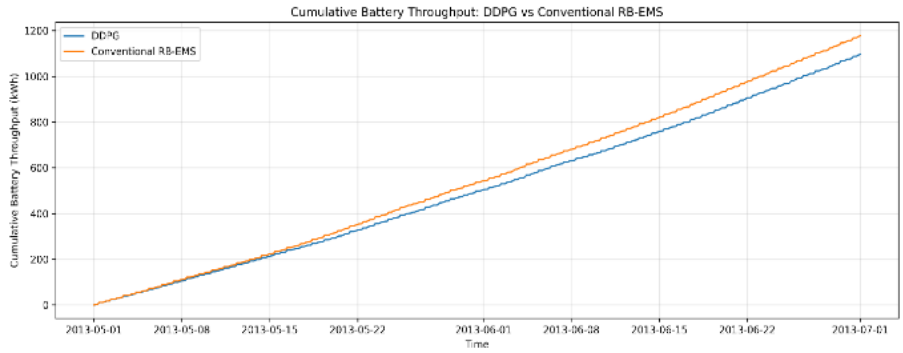

Figure 8: Cumulative Battery Energy Throughput under DDPG and Conventional RB-EMS Control

The figure therefore shows that the DDPG controller did not achieve lower grid import by increasing battery usage. Instead, it recorded lower grid import together with lower cumulative battery throughput under the evaluated test conditions.

4.3.4 Percentage Difference between DDPG and RB-EMS

Table 6 summarises the percentage differences between the DDPG controller and the conventional RB-EMS for the main performance indicators during the unseen May–June 2013 test period. The DDPG controller recorded 6.88% lower battery throughput, with battery charging and discharging also lower by 6.90% and 6.87%, respectively. Grid import was 3.39% lower under DDPG, while grid export was 1.80% lower. The charging and discharging utilisation values were also slightly lower under DDPG, by 2.92% and 1.77%, respectively.

Table 6: Percentage Differences between DDPG and Conventional RB-EMS

Metric	RB-EMS	DDPG	DDPG_vs_RB_EMS_difference_percent
Battery charging (kWh)	617.96	575.34	-6.90
Battery discharging (kWh)	559.81	521.35	-6.87
Battery throughput (kWh)	1177.77	1096.69	-6.88
Grid import (kWh)	1135.32	1096.86	-3.39
Grid export (kWh)	2365.75	2323.13	-1.80
Discharge utilisation (%)	94.91	93.23	-1.77
Charging utilisation (%)	47.54	46.16	-2.92

Overall, the table shows that the DDPG controller achieved lower grid import while using less battery energy during the evaluated test period.

4.4 Discussion of DDPG and RB-EMS Results

The comparison results demonstrate the practical impact of the proposed DDPG-based adaptive control strategy under the unseen May–June 2013 test conditions. Although both controllers operated with the same PV generation, load demand, battery configuration, and operating constraints, their energy-management behaviour was different. The most important result is that the DDPG controller reduced grid import by 3.39%, from 1,135.32 kWh under the conventional RB-EMS to 1,096.86 kWh. At the same time, battery throughput decreased by 6.88%, from 1,177.77 kWh to 1,096.69 kWh. This indicates that the lower grid dependence was achieved without increasing overall battery energy throughput. The battery results further show that DDPG used 6.90% less charging energy and 6.87% less discharging energy than the RB-EMS. Although the RB-EMS recorded slightly higher charging and discharging utilisation, the DDPG controller operated with lower overall battery throughput. This difference reflects the distinct operating behaviour produced by the learned control policy compared with the predefined rule-based strategy.

The grid results provide further evidence of the difference between the two approaches. While both controllers responded to periods of grid import and export, the cumulative grid-import profile showed that DDPG consistently remained below the RB-EMS toward the end of the test period. Grid export was also slightly lower under DDPG, by 1.80%, indicating that the controller's behaviour involved a different balance between battery utilisation and grid energy exchange. Importantly, these differences were achieved while maintaining the evaluated battery operating constraints. Both controllers recorded zero SOC, temperature, voltage, and battery-power constraint violations during the unseen test period. The DDPG controller therefore demonstrated adaptive battery and grid management under the tested conditions without recorded constraint violations.

Overall, the results show the impact of the proposed DDPG control strategy: compared with the conventional RB-EMS, it achieved lower grid import and lower battery energy throughput while maintaining the defined operating constraints. These findings provide quantitative evidence that the learned control approach can produce a different and more controlled pattern of battery and grid energy management under the evaluated solar–load conditions.

5. CONCLUSION

5.1 Conclusion

This study developed and evaluated a Deep Reinforcement Learning-Based Adaptive Energy Management approach for a lithium-ion battery energy storage system using a Deep Deterministic Policy Gradient (DDPG) controller. The controller was designed to adjust battery operation as photovoltaic generation, residential load demand, and battery conditions changed. The DDPG controller used battery SOC, battery temperature, photovoltaic power, and load demand as the main state variables for making its control decisions. A chronological solar–load dataset was prepared using NASA POWER solar-resource data for Ekwulobia, Anambra State, Nigeria, and the Ausgrid residential electricity-demand dataset. The controller was trained on the earlier period and tested on the unseen May–June 2013 period.

During the test period, the controller changed the battery charging and discharging level according to the available PV power and load demand. The battery SOC ranged from 21.30% to 95.00%, while the recorded battery temperature, voltage, and power remained within the specified operating limits. No SOC, temperature, voltage, or battery-power constraint violations were recorded during the test. The DDPG results were also compared with those of the conventional photovoltaic-priority Rule-Based Energy Management System (RB-EMS) under the same test conditions. DDPG recorded 1,096.86 kWh of grid import, compared with 1,135.32 kWh for RB-EMS, representing a 3.39% reduction. Battery energy throughput was also reduced from 1,177.77 kWh with RB-EMS to 1,096.69 kWh with DDPG, a reduction of 6.88%. These results show that the DDPG controller achieved lower grid import while using less total battery energy during the evaluated period.

The study therefore shows that DDPG can be used to provide continuous and adaptive control of a lithium-ion battery in a PV–battery energy-storage system. The results also show that the controller can be evaluated using local solar-resource conditions together with residential load data, while maintaining the specified battery operating constraints.

6.0 Recommendation

Based on the findings of this study, further work can extend the study to longer periods and different seasonal conditions, additional load profiles, other battery operating conditions, and real-time implementation on battery-management hardware. The performance of DDPG can also be compared with other continuous-action reinforcement-learning methods to examine their behaviour under a wider range of operating conditions.

Acknowledgements

The authors acknowledge the assistance and support of the Chief Executive Officer (CEO), KENSTAN SOLAR LTD, Ekwulobia, Aguata Local Government Area, Anambra State, Nigeria, for providing access to the on-site solar–battery installation data used during the initial development and debugging of the battery-control environment. This support contributed to the preliminary development and testing of the proposed battery management system.

Nomenclature

V_b = Battery voltage (V)
I_b = Battery current (A)
T_b = Battery temperature (°C)
P_{PV} = Photovoltaic power (kW)
P_L) = Load power (kW)

E_{PV} = Photovoltaic energy generated (kWh)
E_L = Load energy consumption (kWh)
E_{imp} = Grid import energy (kWh)
E_{exp} = Grid export energy (kWh)
C = Battery capacity (Ah)
t = Time (s or h)
s_t = State at time (t)
a_t = Action at time (t)
r_t = Reward at time (t)
γ = Discount factor (dimensionless);
MAE = Mean Absolute Error;
RMSE = Root Mean Square Error;
R^2 = Coefficient of determination.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors confirm that no generative AI or AI-assisted technologies were employed during the writing of this manuscript. All content was produced solely by the authors.

References

Cardo-Miota, J., Khadem, S., & Bahloul, M., 2025. Deep reinforcement learning based electricity bill minimization strategy for residential prosumer. Mathematics and Computers in Simulation, 238, 296–305. https://doi.org/10.1016/j.matcom.2025.06.011.

Chen, Q., Kuang, Z., Liu, X., & Zhang, T., 2024. Application-oriented assessment of grid-connected PV-battery system with deep reinforcement learning in buildings considering electricity price dynamics. Applied Energy, 364, 123163. https://doi.org/10.1016/j.apenergy.2024.123163.

Fan, J., & Wang, H., 2023. Deep reinforcement learning for community battery scheduling under uncertainties of load, PV generation, and energy prices. In J. Zhao & X. Shen (Eds.), 2023 IEEE the 7th Conference on Energy Internet and Energy System Integration (EI2 2023) (pp. 4871–4876). IEEE. https://doi.org/10.1109/EI259745.2023.10513300.

Gao, Y., Hu, Z., Otomo, J., & Yan, K., 2026. Net-zero ready building control: Benchmarking Decision Transformers against deep reinforcement learning for PV–battery buildings. Energy, 360, 141616. https://doi.org/10.1016/j.energy.2026.141616.

Hussain, A., & Musilek, P., 2024. Energy management of buildings with energy storage and solar photovoltaic: A diversity in experience approach for deep reinforcement learning agents. Energy and AI, 15, 100313. https://doi.org/10.1016/j.egyai.2023.100313.

Li, Y., Wu, J., & Pan, Y., 2024. Deep reinforcement learning for online scheduling of photovoltaic systems with battery energy storage systems. Intelligent and Converged Networks, 5(1), 28–41.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D., 2015. Continuous control with deep reinforcement learning. arXiv. https://doi.org/10.48550/arXiv.1509.02971

Real, A. C., Luz, G. P., Sousa, J. M. C., Brito, M. C., & Vieira, S. M., 2024. Optimization of a photovoltaic-battery system using deep reinforcement learning and load forecasting. Energy and AI, 16, 100347. https://doi.org/10.1016/j.egyai.2024.100347.

Rezaeimozafar, M., Duffy, M., Monaghan, R. F. D., & Barrett, E., 2024. A hybrid heuristic-reinforcement learning-based real-time control model for residential behind-the-meter PV-battery systems. Applied Energy, 355, 122244. https://doi.org/10.1016/j.apenergy.2023.122244.

Shyni, R., & Kowsalya, M. (2024). An LSTM-DDPG framework power management strategy for a heterogeneous energy storage system in a standalone DC microgrid. Journal of Energy Storage, 104, 114437.

Xu, Y., Gao, W., Li, Y., & Xiao, F., 2023. Operational optimization for the grid-connected residential photovoltaic-battery system using model-based reinforcement learning. Journal of Building Engineering, 73, 106774. https://doi.org/10.1016/j.jobe.2023.106774.

Yang, J., Du, S., Chen, P., Liu, S., & Chen, B., 2025. Physics-shielded deep reinforcement learning for safe energy management of microgrids with battery health consideration. Journal of the Franklin Institute, 362(16), 108112. https://doi.org/10.1016/j.jfranklin.2025.108112