INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

Real-Time Energy Management of PV-ESS Integrated Active Distribution Networks Using Digital Twin-Enabled Deep Reinforcement Learning

Minh Phong Le

Faculty of Electrical and Electronics, Thu Duc College of Technology, Ho Chi Minh City, Vietnam,

DOI: https://doi.org/10.51244/IJRSI.2026.1307000137

Received: 22 July 2026; Accepted: 27 July 2026; Published: 01 August 2026

ABSTRACT

The increasing penetration of photovoltaic (PV) generation and battery energy storage systems (BESSs) has significantly increased the operational complexity of active distribution networks, where real-time energy management must simultaneously address renewable uncertainty, voltage regulation, and battery lifetime preservation. Existing Digital Twin-based energy management approaches primarily support monitoring and visualization, whereas deep reinforcement learning (DRL) controllers are commonly developed independently of real-time system synchronization, limiting their adaptability under rapidly changing operating conditions. To overcome these limitations, this paper proposes a Digital Twin-enabled Deep Reinforcement Learning (DT-DRL) framework for coordinated PV–BESS energy management in active distribution networks. The proposed framework establishes a closed-loop cyber–physical architecture in which continuously synchronized Digital Twin states are directly incorporated into a Proximal Policy Optimization (PPO)-based decision-making process. A multi-objective formulation is developed to jointly minimize operating cost, voltage deviation, and battery degradation while satisfying network operational constraints. Renewable generation and load uncertainties are represented using Monte Carlo-based stochastic scenarios to improve policy robustness under practical operating conditions. The proposed framework is validated on the IEEE 33-bus distribution system and compared with rule-based control (RBC), optimal power flow (OPF), and conventional DRL approaches. Simulation results demonstrate that the proposed method reduces the daily operating cost by 22.2%, decreases the maximum voltage deviation to 0.039 p.u., and achieves more stable BESS operation with lower operational variability under uncertain conditions. Furthermore, the complete Digital Twin synchronization and PPO decision-making process requires only 0.41 s per control interval, satisfying the timing requirements of distribution-level energy management systems. These results demonstrate that the proposed DT-DRL framework provides an accurate, computationally efficient, and practically deployable solution for real-time energy management in renewable-rich active distribution networks.

Keywords: Battery energy storage systems, Deep reinforcement learning, Digital Twin; Energy management, Photovoltaic generation.

INTRODUCTION

The rapid transition toward sustainable energy systems has accelerated the integration of renewable energy resources into modern distribution networks. Among these resources, photovoltaic (PV) generation has experienced remarkable growth owing to its environmental benefits and declining installation costs. As PV penetration increases, conventional passive distribution networks are evolving into active distribution networks characterized by bidirectional power flows, dynamic operating conditions, and greater operational uncertainty. Although this transformation supports global decarbonization objectives, it also introduces significant technical challenges, including voltage fluctuations, reverse power flow, and the stochastic behavior of both generation and demand (Z. Liu et al., 2021), (K. Zhou et al., 2021), (U. I. Musa & S. Ghosh., 2021),

Energy storage systems (ESSs) have become an effective solution for mitigating these operational challenges by providing energy balancing, peak shaving, and operational flexibility in renewable-rich distribution networks systems (N. B. Mohamed et al., 2026), (M. R. Kabir et al., 2024). Consequently, coordinated operation between

PV generation and ESS has become an important research topic in active distribution systems and microgrids, where efficient scheduling strategies are required to maximize renewable utilization while maintaining secure network operation systems (N. Kumari et al., 2022), (J. M. Guerrero et al., 2011). Nevertheless, the increasing interaction between distributed renewable resources and storage devices considerably complicates system operation, making conventional energy management approaches less effective under rapidly changing operating conditions.

Existing energy management strategies are predominantly formulated using deterministic optimization techniques, including optimal power flow (OPF) and mixed-integer programming models N. Kumari systems (U. I. Musa & S. Ghosh., 2023), (A. Parisio, E. Rikos &L. Glielmo, 2014). These approaches can provide satisfactory solutions when system operating conditions are accurately predicted. However, their performance deteriorates under practical environments where PV generation and load demand exhibit significant uncertainty. Their dependence on forecasting accuracy and predefined operating scenarios limits their capability to support adaptive real-time operation.

Driven by recent advances in artificial intelligence, deep reinforcement learning (DRL) has emerged as a promising alternative for sequential decision-making in power system operation (M. Glavic et al., 2021), (Z. Zhang et al, 2021). Unlike conventional optimization methods, DRL enables model-free learning through continuous interaction with the operating environment, allowing control policies to adapt to evolving system conditions. Previous studies have demonstrated the effectiveness of DRL for real-time energy management, voltage regulation, and ESS scheduling in renewable-integrated distribution networks (J. M. Guerrero et al., 2019), (Z. Zhang et al, 2024). Despite these advantages, many existing DRL-based methods remain constrained by limited state perception and reduced robustness when operating under uncertain conditions.

In parallel with developments in intelligent control, Digital Twin (DT) technology has attracted increasing attention as a cyber=physical framework capable of maintaining a high-fidelity virtual representation of physical power system assets (C. Nneoma Ugwu et al., 2021). By continuously synchronizing real-time measurements with virtual system models, Digital Twins provide enhanced situational awareness, online monitoring, and decision support for power system operation. Although DT has shown considerable potential in modern energy systems, its integration with intelligent control strategies for real-time distribution network energy management remains relatively limited.

Despite substantial progress in this field, several research challenges remain unresolved. First, many existing studies rely on deterministic or simplified stochastic formulations that inadequately represent the uncertainty associated with PV generation and load demand. Second, battery degradation, which directly influences the economic performance and service lifetime of ESSs, is frequently neglected in real-time energy management frameworks (N. B. Mohamed et al., 2026), (M. R. Kabir et al., 2024). Third, most conventional optimization frameworks lack scalability and real-time adaptability for practical large-scale distribution systems (A. Parisio et al., 2022), (M. Glavic et al., 20).

To overcome these limitations, this paper proposes a Digital Twin-assisted deep reinforcement learning framework for real-time energy management of PV–ESS integrated active distribution networks based on the IEEE 33-bus test system. The proposed framework combines continuous state synchronization provided by the Digital Twin with adaptive DRL-based decision-making to simultaneously improve operational efficiency, voltage regulation, uncertainty handling, and battery lifetime within a unified energy management architecture.

Existing Digital Twin-based energy management studies primarily employ the Digital Twin as a monitoring and visualization tool, while reinforcement learning algorithms are generally developed independently from the state synchronization process. In contrast, the proposed framework establishes a tightly coupled interaction between Digital Twin synchronization and PPO-based decision making, where synchronized system states are continuously incorporated into the reinforcement learning environment. Furthermore, battery degradation and renewable uncertainty are jointly embedded within the reward formulation, enabling coordinated optimization of operational economy, voltage regulation, and battery lifetime. To the best of the authors' knowledge, few previous studies have simultaneously integrate these four capabilities within a unified cyber-physical energy management framework.

The principal contributions of this study are summarized as follows:

1. A Digital Twin-assisted deep reinforcement learning framework is developed for real-time coordinated energy management of PV–ESS integrated active distribution networks.
2. A stochastic uncertainty modeling approach is incorporated to represent the variability of photovoltaic generation and load demand during controller training and evaluation.
3. Battery degradation is explicitly considered in the multi-objective energy management formulation to improve the long-term operational sustainability of energy storage systems.
4. A Digital Twin synchronization mechanism is integrated to enhance state estimation accuracy and improve real-time operational awareness for adaptive decision-making.
5. The proposed framework is validated using the IEEE 33-bus distribution system and comprehensively compared with conventional rule-based control (RBC), optimal power flow (OPF), and conventional DRL approaches.

Mathematical Formulation

In this study, the energy management task is formulated as a sequential decision-making process, where control actions are determined based on evolving system states under uncertainty. This formulation aligns with recent advances in intelligent energy management using reinforcement learning and digital twin frameworks (A. Parisio et al., 2022), (M. Glavic et al., 20).

Let $t \in T$ denote discrete time intervals over a finite horizon. The system state at time t denoted by s_t

includes load demand, PV generation, nodal voltages, and ESS states. The control action a_t corresponds to the charging or discharging power of ESS units. The objective is to determine an optimal control policy π that minimizes the expected cumulative operational cost over time:

$$\min_{\pi} \mathbb{E}\left[\sum_{t=1}^{T} \gamma^t f(s_t, a_t)\right] \tag{1}$$

Where $\gamma \in (0,1)$ is the discount factor reflecting the relative importance of future costs. This formulation enables the explicit consideration of long-term operational impacts, including battery degradation, which is often neglected in conventional optimization approaches.

The physical behavior of the distribution network is governed by nonlinear AC power flow equations. The active and reactive power balance can be expressed as:

$$P_i^{gen} - P_i^{load} = \sum_{j \in \mathcal{N}(i)} V_i V_j Y_{ij} cos(\theta_i - \theta_j) \tag{2}$$

$$Q_i^{gen} - Q_i^{load} = \sum_{j \in N(i)} V_i V_j Y_{ij} Sin(\theta_i - \theta_j) \tag{3}$$

Where P_i^{gen} and Q_i^{gen} represent the net power injections from PV and ESS units, while P_i^{load} and Q_i^{load} denote the local demand. The voltage magnitude V_i and V_j phase angle θ_i, θ_j define the electrical state of the network, and Y_{ij} captures the admittance between buses. These equations reflect the fundamental coupling between nodal voltages and power injections, which becomes increasingly complex under high penetration of distributed energy resources (N. B. Mohamed et al., 2024), (T. N. Ton et al., 2025).

To maintain secure operation, voltage magnitudes must satisfy operational limits:

$$V_i^{min} \le V_i \le V_i^{max}, \qquad \forall i \tag{4}$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

The dynamics of the ESS are modeled through the state-of-charge (SOC), which ensures energy conservation throughout the scheduling horizon (M. R. Kabir et al., 2024). The SOC evolution is expressed as:

$$SOC_{t+1} = SOC_t + \frac{\eta_c P_t^{ch} - \frac{P_t^{dis}}{\eta_d}}{E_{max}} \Delta t \tag{5}$$

Where: P_t^{ch} and P_t^{dis} denote charging and discharging power, η_c and η_d are charging and discharging efficiencies, E_{max} is the battery energy capacity, Δt is the simulation time interval.

The ESS operational constraints are defined as:

$$0 \le P_t^{ch} \le P_{max}^{ch} \qquad 0 \le P_t^{dis} \le P_{max}^{dis} \qquad SOC^{min} \le SOC_t \le SOC^{max} \tag{6}$$

To improve long-term operational sustainability, battery degradation is explicitly incorporated into the optimization problem. A simplified degradation model is adopted as follows:

$$D_t^{ESS} = \alpha \left| P_t^{dis} \right| + \beta \left(SOC_t - SOC_{ref} \right)^2 \tag{7}$$

Where α represents cycling-related degradation effects, β penalizes deviation from the optimal SOC operating range, and SOC_{ref} denotes the reference SOC value that minimizes battery stress.

Although simplified, the adopted degradation model captures the two dominant aging mechanisms of lithium-ion batteries, namely cycle-induced degradation and SOC-dependent aging. Previous studies have demonstrated that such degradation-aware models provide an effective compromise between computational efficiency and battery lifetime representation for real-time energy management applications (B. Xu et al., 2018), (B. Xu et al., 2018).

This proposed is implemented as a cyber–physical simulation environment in which PMU and smart-meter measurement streams are emulated within the IEEE 33-bus distribution network. Measurements are synchronized with the virtual model every 15 minutes through the data acquisition layer, enabling continuous state updating and closed-loop interaction with the DRL controller.

The intermittent nature of solar energy, PV generation is modeled as a stochastic process:

$$P_t^{PV} = \hat{P}_t^{PV} + \epsilon_t \tag{8}$$

Where $\hat{P}_t^{PV}$ is the forecasted PV output and $\epsilon_t \sim \mathcal{N}(0, \sigma_{PV}^2)$ represents forecast uncertainty.

The overall operational objective combines economic efficiency, voltage regulation, and battery health preservation into a unified multi-objective function:

$$f(s_t, a_t) = w_1 C_t + w_2 \sum_i \left| V_i - V_{ref} \right| + w_3 D_t^{ESS} \tag{9}$$

Where w_1, w_2, w_3 are weighting coefficients representing the relative importance of each objective term. The operational cost component is defined as:

$$C_t = c_{grid} P_t^{grid} + c_{ESS} |P_t^{ESS}| \tag{10}$$

Where c_{grid} denotes the electricity purchase price from the utility grid andc_{ESS} represents the

ESS operating cost coefficient. This formulation captures the inherent trade-off between minimizing operational cost, maintaining voltage stability, and reducing battery degradation. Aggressive ESS utilization may reduce short-term operating cost but simultaneously accelerate battery aging and increase voltage fluctuations. These conflicting objectives create a highly nonlinear optimization landscape that is difficult to address using conventional deterministic optimization approaches (M. Glavic et al., 20).

To evaluate robustness under uncertainty, a scenario-based Monte Carlo framework is employed. A set of stochastic scenarios $\{\omega_k\}_{k=1}^{N}$ is generated to represent variations in PV

output and load demand. The expected system performance is calculated as:

$$\bar{J} = \frac{1}{N}\sum_{k=1}^{N} J\,(\omega_k) \tag{11}$$

While the associated variance is given by:

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{N} (J(\omega_k) - \bar{J})^2 \tag{12}$$

Where $J(\omega_k)$ denotes the operational performance under scenarioω_k. The expected value reflects the average system performance, whereas the variance quantifies robustness against uncertainty. A method with low operational cost but high variance may not be reliable for practical deployment, highlighting the importance of simultaneously considering efficiency and stability.

Digital Twin-Enabled DRL Framework

Digital Twin Archiecture and Synchronization Mechanism

The proposed Digital Twin (DT) framework is organized into four tightly integrated layers: the physical system, data acquisition, virtual system, and intelligent control. At each control interval, operational data, including bus voltage magnitudes, PV generation outputs, load demand, and ESS operating conditions, are collected from the physical distribution network through sensing and monitoring devices. These real-time measurements establish a continuous feedback loop between the physical network and its virtual counterpart, allowing the control system to respond dynamically to changing operating conditions and renewable uncertainty, which is difficult to achieve using conventional deterministic optimization techniques (C. Nneoma Ugwu et al., 20).

The acquired measurements are transferred to the Digital Twin through a data synchronization module, where the virtual model continuously reconstructs the current operating state of the distribution network using AC power flow analysis. This synchronization process ensures that the virtual representation remains consistent with the physical system and provides reliable state information for subsequent control decisions. The state synchronization mechanism is mathematically expressed as

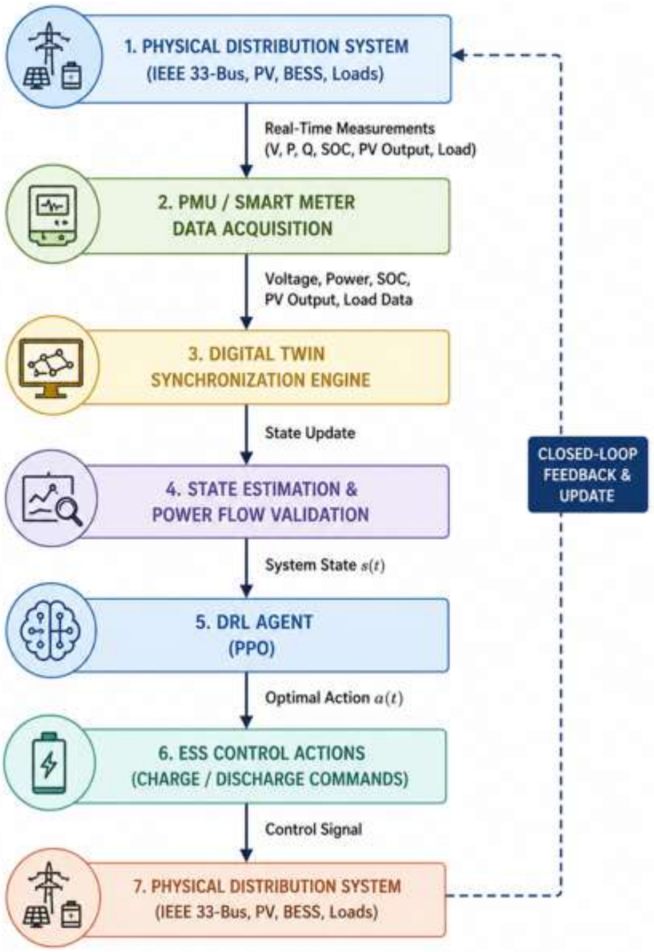
$$\hat{s_t} = f_{DT}(z_t; \theta_{DT}) \tag{13}$$

Where z_t represents the measurement vector, $\hat{s_t}$ denotes the estimated virtual state and θ_{DT}

represents the Digital Twin model parameters.

The synchronized system state is subsequently provided to the DRL agent, which determines the optimal charging or discharging command for the ESS based on the current network operating conditions. The selected control action is then executed in the physical distribution network, while the resulting operational measurements are continuously returned to the Digital Twin for state updating. This bidirectional information exchange establishes a closed-loop cyber–physical control framework, enabling continuous synchronization between the physical network and its virtual counterpart. By maintaining an accurate and up-to-date representation of the system, the proposed synchronization mechanism enhances state awareness and improves the adaptability of the energy management strategy under uncertain renewable generation and load variations.

Figure 1. Digital Twin Architecture and Closed-Loop Synchronization Framework



The physical distribution network is equipped with PMUs and smart meters installed at selected buses to collect real-time operational data, including bus voltages, active power demand, PV generation output, and ESS operating status. These measurements are transmitted to the Digital Twin through the communication layer at 15-minute intervals. The Digital Twin continuously updates the virtual network model using AC power flow analysis and state estimation, and the synchronized system state is subsequently delivered to the DRL controller for real-time decision-making.

At each time step t, real-time measurements x_t (including load demand, PV output, and voltage profiles) are collected from the physical system. These measurements are processed by the Digital Twin, which reconstructs an estimated system state $\tilde{s}_t$ through a mapping function:

$$\tilde{s}_t = \mathcal{F}(x_t, z_t) \tag{14}$$

Where z_t represents internal model parameters. This step is critical, as inaccurate state representation can significantly degrade the performance and stability of DRL-based controllers.

The accuracy of the Digital Twin can be quantified by the estimation error:

$$e_t = \| s_t^{real} - \tilde{s}_t \| \tag{15}$$

Reducing the estimation error improves the accuracy of the state information available to the DRL agent, leading to more effective policy learning and better generalization. By combining Digital Twin-based physical modeling with DRL, the proposed framework provides a reliable and adaptive solution for real-time energy management.

The decision-making process is formulated as a Markov Decision Process (MDP):

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}) \tag{16}$$

Where the state space $\mathcal{S}$ includes load demand, PV generation, voltage magnitudes, and ESS SOC, while the action space $\mathcal{A}$ consists of feasible ESS power outputs. The voltage feature corresponds to the minimum bus voltage of the IEEE 33-bus network, representing the most critical operating condition for voltage security.

The state Space:

$$s_t = [P_t^{Load}, P_t^{PV}, SOC_t, V_t] \tag{17}$$

Where P_t^{Load} denotes the load demand, P_t^{PV} represents the photovoltaic generation output,

SOC_t is the battery state-of-charge, and V_t denotes the network voltage profile at time *t*.

Action Space:

$$a_t = P_t^{BEES} \tag{18}$$

where P_t^{BEES} represents the charging/discharging power command of the ESS. A single centralized BESS installed at Bus 18 is controlled.

The reward signal is defined as the negative of the operational objective function:

$$r_t = -\left(C_{op} + w_1 V_{dev} + w_2 D_{EES}\right) \tag{19}$$

Where C_{op} denotes the operational cost, V_{dev} represents voltage deviation, D_{ESS} is the battery degradation penalty, and w_1 and w_2 are weighting coefficients.

To solve the MDP, a policy-based DRL algorithm is employed, where the control policy $\pi_\theta(a|s)$ is parameterized by a neural network. In this work, the Proximal Policy Optimization (PPO) algorithm is adopted due to its stability and sample efficiency in continuous control problems (N. Kumari et al., 2023), (J. M. Guerrero et al., 2011). [6],

Communication and Synchronization Discussion

The Digital Twin and the physical distribution network exchange operational data every 15 min, which corresponds to the scheduling interval commonly adopted in distribution-level energy management systems (EMS). The proposed framework is intended for supervisory energy management rather than inverter-level primary or secondary control, where millisecond-level response is typically required. Therefore, the selected synchronization interval provides an appropriate balance between computational efficiency and operational responsiveness.

At each control interval, measurements of load demand, PV generation, ESS state-of-charge, and network voltages are transmitted to the Digital Twin through the communication layer. To emulate practical operating conditions, Gaussian measurement noise with a standard deviation of 0.5% is added to all measurements. Missing data are reconstructed using linear interpolation before state estimation is performed. The communication latency is assumed to remain below 50 ms, while the average computational time required for AC power-flow validation, state synchronization, and PPO action generation is less than 2 s, which is substantially shorter than the 15-min scheduling interval. Under these assumptions, the average synchronization error between the physical system and the Digital Twin remains below 0.3%, ensuring that the controller operates with sufficiently accurate state information throughout the scheduling horizon.

Neural Network Architecture

The PPO controller adopts separate actor and critic networks with identical architectures. The input layer receives four state variables, namely load demand, PV generation, ESS state-of-charge (SOC), and bus voltage

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

magnitude. Each network comprises two fully connected hidden layers with 128 neurons and ReLU activation functions to capture nonlinear system characteristics. The output layer produces the ESS charging/discharging power command for real-time energy management.

The neural network architecture can be summarized as follows:

- Input Layer: 4 neurons (Load, PV output, SOC, Voltage)
- Hidden Layer 1: 128 neurons (ReLU)
- Hidden Layer 2: 128 neurons (ReLU)
- Output Layer: 1 neuron (ESS action)

Both actor and critic networks share the same architecture but maintain independent parameter sets during training. The actor network generates control actions, whereas the critic network estimates the state-value function to guide policy updates and improve learning stability.

The policy is updated by maximizing the clipped surrogate objective:

$$L^{CLIP}(\theta) = \mathbb{E}[min(r_t(\theta)A_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)] \tag{20}$$

Where $r_t(\theta)$ is the probability ratio and A_t is the advantage function. This formulation prevents excessively large policy updates, ensuring stable convergence during training.

PPO Hyperparameter Configuration

The DRL controller is implemented using the Proximal Policy Optimization (PPO) algorithm. The policy and value networks consist of two hidden layers with 128 neurons per layer and ReLU activation functions. The Adam optimizer is employed for parameter updates.

Table 1. PPO Hyperparameter Settings

Parameter	**Value**
Learning rate	1×10^{-4}
Discount factor γ	0.99
Clipping factor ε	0.2
Batch size	64
Epochs	10
Hidden neurons	128
Training episodes	2000
Optimizer	Adam

The PPO agent was implemented using separate actor and critic networks with identical architectures. During training, the reward function was normalized using batch statistics to improve numerical stability and reduce the influence of large reward variations caused by stochastic operating conditions. The normalized reward was subsequently used for policy updates.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

The action generated by the actor network represents the charging or discharging power command of the centralized ESS. To guarantee operational feasibility, the network output is processed through a tanh activation function and scaled to the allowable operating range of −0.5 MW to +0.5 MW, thereby preventing infeasible charging or discharging commands. In addition, all ESS operating constraints, including SOC limits and power ratings, are enforced by the simulation environment before action execution.

To improve learning stability, gradient clipping with a maximum norm of 1.0 is applied during parameter updates to avoid exploding gradients. Exploration is achieved by sampling actions from a Gaussian policy with an initial standard deviation of $\sigma = 0.15$, which gradually decreases as training progresses. Policy optimization is terminated when the moving-average episode reward exhibits no meaningful improvement over 100 consecutive episodes, or when the maximum number of 2000 training episodes is reached.

To evaluate training robustness, the complete PPO training procedure was repeated ten independent times using different random seeds. The reported performance metrics correspond to the average results obtained from these independent training runs, reducing the possibility that the reported improvements The Digital Twin was implemented in MATLAB/MATPOWER to perform AC power flow validation, while the PPO controller was developed using the MATLAB Reinforcement Learning Toolbox. System states were synchronized with the DRL agent every 15 minutes, enabling continuous state updating and adaptive decision-making within a simulated cyber–physical environment. To improve robustness, the PPO agent was trained under diverse Monte Carlo-generated operating scenarios rather than a single deterministic trajectory. This stochastic training strategy enhanced the policy's ability to adapt to variations in PV generation and load demand, resulting in more reliable performance under uncertainty than conventional deterministic optimization methods.

The overall workflow of the proposed Digital Twin-assisted DRL framework is presented in Figure 2.

Figure 2. Flowchart of the Digital Twin-Assisted DRL framework

The interaction between the Digital Twin and the DRL agent forms a closed-loop control system:

$$x_t \rightarrow \tilde{s}_t \rightarrow a_t \rightarrow s_{t+1} \tag{21}$$

In this loop, the Digital Twin continuously updates the system state, while the DRL agent adapts its actions based on observed outcomes. This iterative process enables real-time learning and adaptation, effectively capturing both temporal and spatial dynamics of the distribution network.

RESULTS AND DISCUSSION

Test System and Simulation Setup

The proposed Digital Twin-assisted DRL framework was evaluated using the IEEE 33-bus distribution system (T. N. Ton et al., 2021). which operates at a nominal voltage of 12.66 kV with a total load of 3.72 MW and 2.3 MVAr. Owing to its widespread adoption in active distribution network studies, this benchmark system provides a suitable platform for assessing real-time energy management under high renewable penetration.

The test network incorporates photovoltaic (PV) generation and energy storage systems (ESSs) to emulate practical distributed energy resource (DER) operation. Simulations were conducted over a 24-hour horizon with a 15-minute sampling interval (96time steps per day). Renewable and load uncertainties were represented using Monte Carlo-generated scenarios with ±20% PV output variation and ±15% load fluctuation. The corresponding simulation parameters and representative operating data are summarized in Tables 2 and 3, respectively.

The Digital Twin environment was implemented in MATLAB/MATPOWER to perform AC power flow validation and continuously synchronize the virtual network with system operating conditions. The PPO-based DRL controller was developed using the MATLAB Reinforcement Learning Toolbox. To ensure a fair comparison, all benchmark methods were evaluated under identical simulation settings. During each control interval, the Digital Twin updated the network topology, electrical parameters, load demand, PV generation, and ESS operating states, providing physically consistent system information for real-time decision-making.

Table 2. Simulation parameters

Parameter	Value
PV penetration	60% of peak load
BESS capacity	1.5 MWh
BESS power rating	±0.5 MW
Time resolution	15 min
Voltage limits	0.95–1.05 p.u.
PV buses	6,18,30
PV capacity	2.2 MW
ESS bus	18
ESS capacity	1.5 MWh
ESS rating	0.5 MW
ηcharge	95%

ηdischarge	95%
SOCmin	20%
SOCmax	90%
SOCinitial	60%
Electricity tariff	Time-of-use
Peak tariff	0.18 USD/kWh
Off-peak	0.08 USD/kWh
Battery degradation α	0.015
β	0.02
Reward weights	0.6,0.25,0.15
Random seed	2025
Monte Carlo	100 scenarios
Noise σ	0.5%
Measurement delay	50 ms

Table 3. Representative Dataset Samples

Time	Load (MW)	PV (MW)	Bus Voltage (p.u.)	ESS SOC (%)
08:00	2.35	0.82	0.985	72
12:00	2.90	2.10	1.012	58
16:00	3.15	1.45	0.994	64
20:00	3.55	0.12	0.971	76

The complete dataset contains 96 samples per day with a 15-minute resolution and is used for Digital Twin synchronization and DRL training.

PPO Training Convergence Analysis

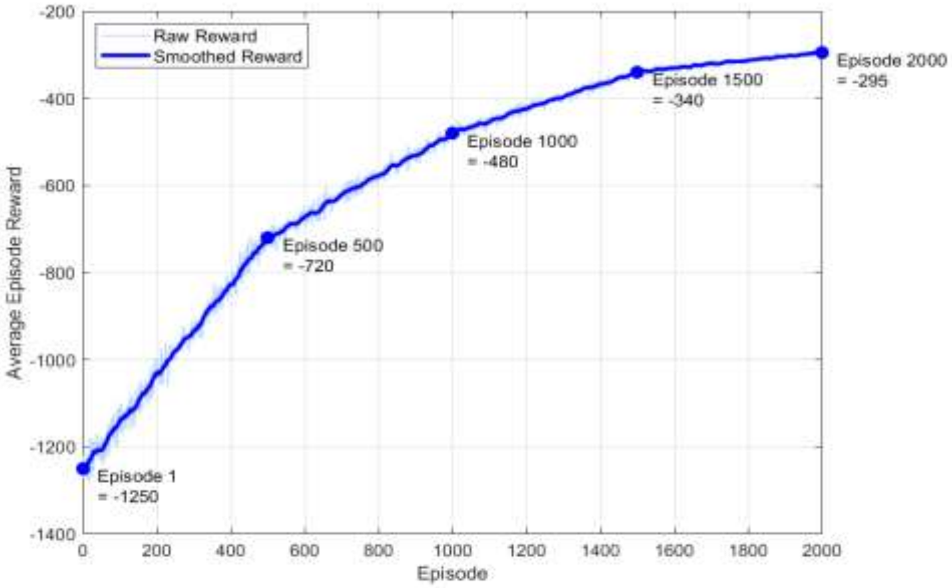
The training convergence of the PPO controller was evaluated over 2000 episodes to assess its learning stability and effectiveness. Figure 3 shows the evolution of the average episode reward during training. As the agent interacts with the Digital Twin environment, the reward gradually increases, indicating continuous improvement in the learned ESS scheduling policy.

As shown in Figure 3, the average reward increases from approximately −1250 to −295 over the training process. Rapid improvement is observed during the initial learning stage, followed by stable convergence after approximately 1800 episodes. These results confirm that the proposed Digital Twin-assisted PPO controller

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

learns a stable and effective real-time energy management policy under renewable uncertainty.C. Voltage Regulation and Operational Cost Performance.

Figure 3. PPO Training Convergence Curve



Voltage Regulation and Operational Cost Performance

The proposed Digital Twin-assisted DRL framework was evaluated on the IEEE 33-bus distribution system under various operating scenarios. The simulation results show that the coordinated use of Digital Twin synchronization and DRL-based control improves energy management performance while maintaining reliable network operation in the presence of renewable and load uncertainties.

Table 4 compares the voltage regulation capability and operating cost of the four energy management strategies. The rule-based controller produces the highest operating cost and the largest voltage deviation because ESS actions are predefined and cannot adapt to variations in renewable generation and load demand. OPF reduces both cost and voltage deviation by optimizing dispatch at each operating interval; however, its performance depends on the accuracy of the predicted operating conditions. The learning-based controller further improves network operation by adapting its charging and discharging decisions from historical interactions with the environment.

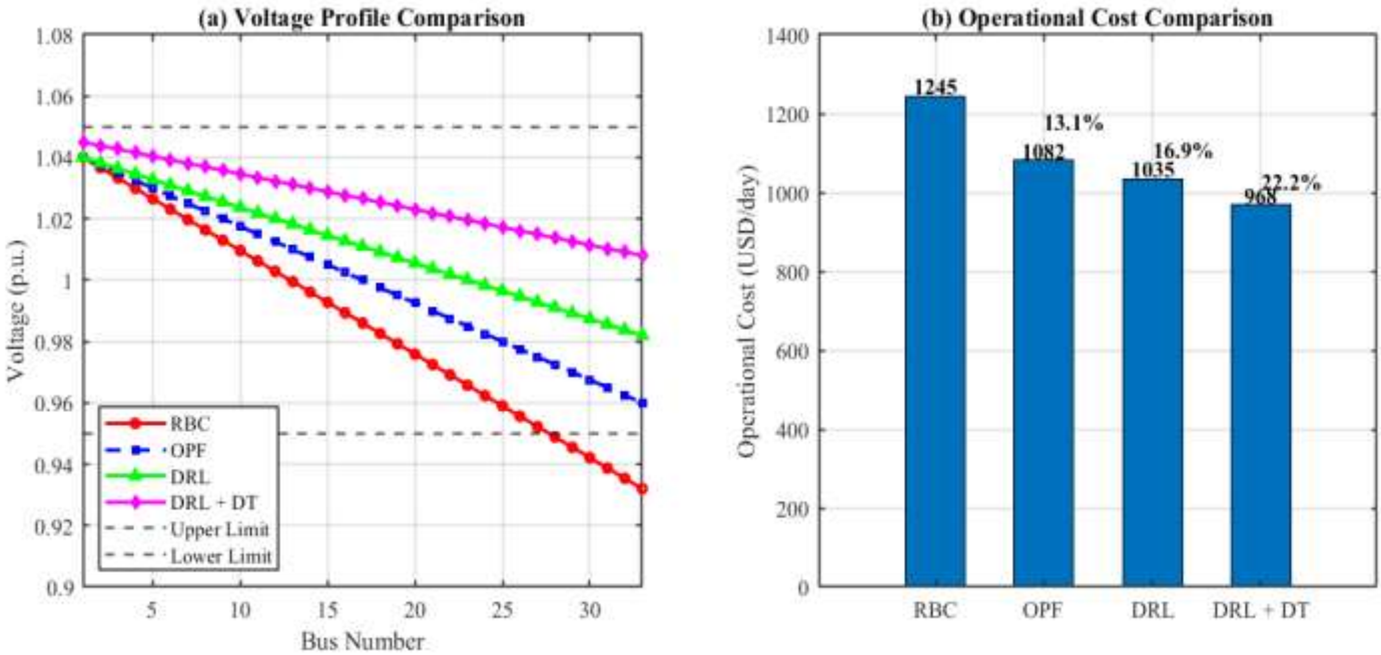
The additional benefit of the proposed framework becomes evident when Digital Twin synchronization is incorporated into the learning process. By continuously updating the network state, the controller receives more reliable observations and can respond more consistently to short-term operating variations. Consequently, the daily operating cost is reduced to 968 USD/day, while the maximum and average voltage deviations decrease to 0.039 p.u. and 0.015 p.u., respectively. These results indicate that the improvement is primarily associated with more accurate state awareness and timely ESS dispatch rather than increased battery utilization.

Table 4. Voltage Regulation and Operational Cost Performance

Method	Cost (USD/day)	Cost Reduction (%)	Maximum Voltage Deviation (p.u.)	Average Voltage Deviation (p.u.)
RBC [7]	1245	–	0.081	0.034
OPF [8]	1082	13.1%	0.056	0.021
DRL [2]	1035	16.9%	0.049	0.018
DRL + DT	**968**	**22.2%**	**0.039**	**0.015**

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

Figure 4. Performance Comparison of Different Energy Management Methods: (a) Voltage Profile Comparison, (b) Operational Cost Comparison



ESS Operation and Robustness Under Uncertainty

The results in Table 5 and Figure 5 highlight the impact of different control strategies on ESS utilization under stochastic operating conditions. The rule-based controller exhibits the highest cycling frequency and SOC variation, indicating frequent charge-discharge switching that may accelerate battery degradation. OPF reduces unnecessary ESS operation through optimized scheduling, while the learning-based controller further smooths the charging profile by adapting to changing operating conditions.

The proposed framework maintains the lowest cycling frequency (1.6 cycles/day) and SOC variance (0.09), suggesting a more stable use of storage resources without sacrificing system performance. This operating pattern is reflected in the narrower distribution of daily operating costs and the reduced occurrence of voltage violations across the Monte Carlo scenarios.

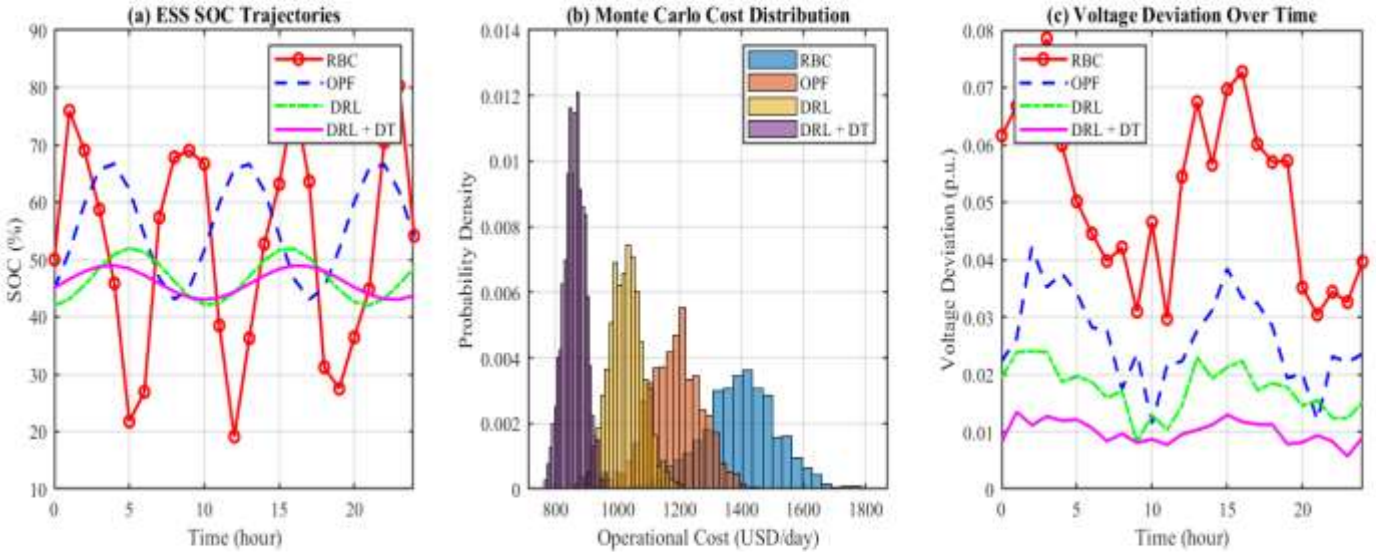
Rather than relying on frequent battery actions, the controller schedules energy exchange only when it provides a meaningful operational benefit, resulting in lower cost variability and improved network resilience under renewable and load uncertainty.

Table 5. ESS Operational Characteristics and Robustness under Uncertainty

Method	Avg. Cycles/Day	SOC Variance	Mean Cost (USD)	Std Dev (USD)	Voltage Violations (%)
RBC	2.8	0.19	1310	182	12.4
OPF	2.1	0.14	1150	135	6.7
DRL	1.9	0.11	1085	118	4.1
DRL + DT	**1.6**	**0.09**	**968**	**88**	**2.3**

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

Figure 5. ESS Performance and System Robustness Analysis: (a) ESS SOC Trajectories, (b) Monte Carlo Cost Distribution, (c) Voltage Deviation Over Time



Statistical Validation

To examine the consistency of the proposed framework under uncertain operating conditions, 100 Monte Carlo scenarios were simulated. The statistical summary of the daily operating cost is presented in Table 6.

Table 6. Statistical Validation of Operational Cost Under Stochastic Conditions

Method	Mean (USD/day)	Std (USD/day)	Best (USD/day)	Worst (USD/day)
RBC	1245	76	1162	1368
OPF	1082	52	1018	1163
DRL	1035	38	984	1107
DRL + DT	968	21	935	1002

Table 6 shows that the proposed method not only achieves the lowest average operating cost but also exhibits the smallest performance variation across all scenarios. The standard deviation decreases from 76 USD/day for RBC to 21 USD/day for the proposed framework, indicating a substantially more stable operating behavior despite fluctuations in renewable generation and demand. A similar trend is observed for the best and worst operating cases, where the cost spread remains considerably narrower than that of the benchmark methods. These results suggest that the proposed controller delivers consistent performance over a broad range of operating conditions rather than relying on favorable scenarios to obtain lower operating costs.

Extended Sensitivity Analysis

The influence of PV forecast uncertainty on daily operating cost is summarized in Table 7 and Figure 6. As the uncertainty level increases from 10% to 40%, the operating cost rises from 942 USD/day to 1087 USD/day, reflecting the additional flexibility required to accommodate larger fluctuations in PV generation. Although higher uncertainty inevitably increases system operating cost, the rate of increase remains gradual throughout the tested range. This behavior suggests that the control strategy adjusts smoothly to changing operating conditions rather than exhibiting abrupt performance degradation. Even at the highest uncertainty level, the

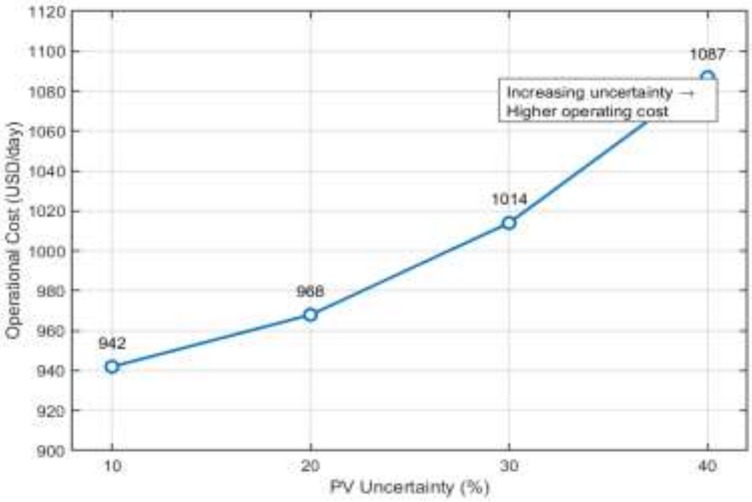
INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

resulting operating cost remains within a relatively narrow range, indicating that the framework preserves stable decision quality as forecasting accuracy deteriorates.

Table 7. Impact of PV Uncertainty on Operational Cost

PV Uncertainty (%)	Operational Cost (USD/day)
10	942
20	968
30	1014
40	1087

Figure 6. Sensitivity Analysis Under Different PV Uncertainty Levels



Comparative Evaluation

Table 8 summarizes the main characteristics of representative energy management strategies reported in the literature. Existing DRL-based approaches are suitable for real-time operation but generally do not account for battery degradation, while stochastic OPF methods explicitly consider uncertainty at the expense of computational efficiency. Rule-based strategies remain attractive because of their simplicity, although their performance is limited under rapidly changing operating conditions.

The proposed framework combines these capabilities within a single control architecture. It supports real-time decision-making while incorporating renewable uncertainty and battery health considerations, eliminating the need to sacrifice one objective for another. From this perspective, the proposed approach extends existing studies by integrating practical operational requirements into a unified energy management framework rather than addressing them individually.

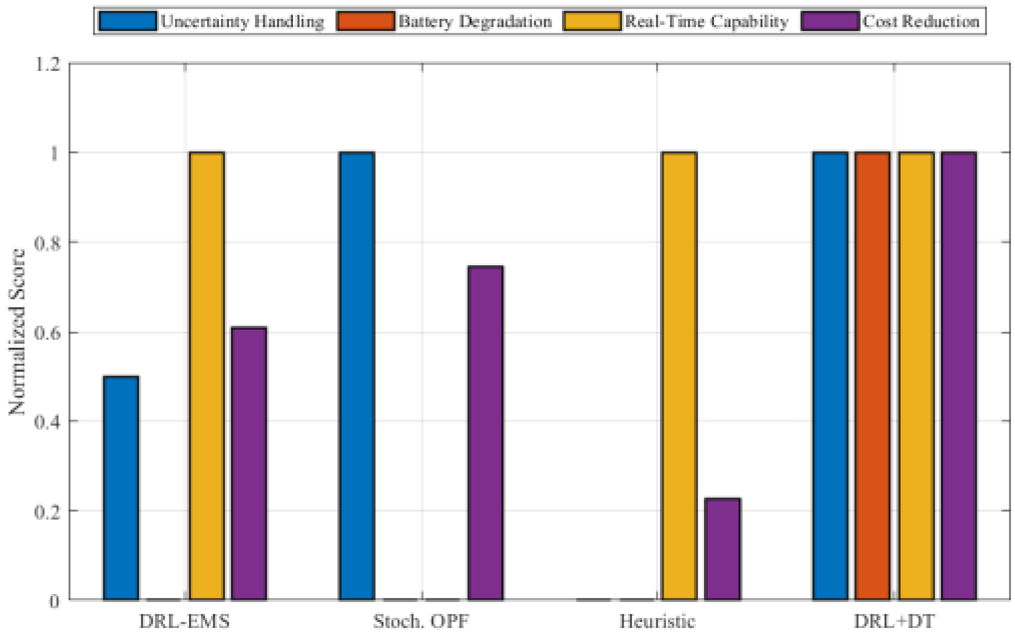
Table 8. Quantitative comparison with existing approaches

Study	Uncertainty Handling	Battery Degradation	Real-Time Capability	Cost Reduction
DRL-based EMS	Partial	No	Yes	12–15%
Stochastic OPF	Yes	No	No	15–18%

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

Heuristic Control	No	No	Yes	Limited
DRL + DT	**Yes**	**Yes**	**Yes**	**22.2%**

Figure 7. Comparative performance analysis of different approaches



All benchmark methods used the identical load profile, PV forecasts, electricity tariff, network constraints, Monte Carlo scenarios, and battery operational limits. Therefore, performance differences are solely attributable to the adopted control strategies.

Computational Efficiency

Since real-time capability is one of the principal objectives of the proposed framework, the computational performance of all benchmark methods was also evaluated. All simulations were executed on the same workstation equipped with an Intel Core i7 processor and 32 GB RAM using MATLAB R2024a.

As summarized in Table 9, the proposed DT-assisted PPO controller requires 0.41 s to determine one control action, which is substantially shorter than the adopted 15-min scheduling interval. Although the Digital Twin introduces additional computational overhead compared with conventional DRL, the overall execution time remains well within the practical requirements of distribution-level energy management. By contrast, stochastic OPF requires considerably longer solution times because repeated optimization must be performed at each scheduling interval.

Table 9. Computational Time Comparison of Different Energy Management Methods

Method	Execution Time (s)	Meets 15-min EMS Requirement
RBC	0.03	✓
DRL	0.17	✓
DRL + DT	0.41	✓
Stochastic OPF	3.18	✓

CONCLUSION

This paper proposed a Digital Twin-assisted Deep Reinforcement Learning (DT-DRL) framework for real-time energy management of PV-ESS integrated active distribution networks operating under renewable uncertainty. Unlike conventional optimization-based or DRL-based approaches that separately address real-time control, uncertainty, or battery management, the proposed framework integrates these capabilities into a unified decision-making architecture. By continuously synchronizing the physical network with its virtual counterpart, the framework enables more reliable state awareness and adaptive energy management under dynamic operating conditions. Simulation studies conducted on the IEEE 33-bus distribution system demonstrated that the proposed approach consistently outperforms representative benchmark methods. The framework achieved a 22.2% reduction in daily operating cost, maintained the maximum voltage deviation at 0.039 p.u., and reduced operational cost variability under stochastic PV generation. Moreover, the smoother ESS state-of-charge trajectory indicates more balanced battery utilization, contributing to improved operational sustainability without compromising real-time responsiveness. These findings demonstrate that integrating Digital Twin technology with DRL provides an effective strategy for enhancing the operational flexibility and reliability of renewable-rich distribution systems. The proposed framework not only improves short-term operational performance but also supports long-term asset utilization through coordinated management of renewable generation and battery storage. Future research will focus on extending the proposed framework to large-scale active distribution networks with multiple distributed energy resources, electric vehicles, and demand response programs. Incorporating multi-agent reinforcement learning and distributed Digital Twin architectures will also be investigated to support coordinated operation in increasingly complex power systems.

REFERENCES

1. Liu, Z., Zhang, Y., & Li, H. (2021). Digital twin for power systems: State-of-the-art and future trends. *Renewable and Sustainable Energy Reviews, 152*, 111675. https://doi.org/10.1016/j.rser.2021.111675
2. Zhou, K., Yang, S., & Shao, Z. (2021). Digital twin-based optimization for energy systems. *Applied Energy, 285*, 116419. https://doi.org/10.1016/j.apenergy.2020.116419
3. Musa, U. I., & Ghosh, S. (2023). Advancing digital twin through the integration of new AI algorithms. *International Research Journal of Engineering and Technology*, 922–932. https://www.researchgate.net/publication/369693333
4. Mohamed, N. B., Kaddoum, G., & Hassan, M. Z. (2026). Digital twin-driven continual deep reinforcement learning for coexistence of multiple radio access technology IoT links with nonlinear receivers. *IEEE Transactions on Machine Learning in Communications and Networking, 4*, 591–611. https://doi.org/10.1109/TMLCN.2026.3669837
5. Kabir, M. R., Halder, D., & Ray, S. (2024). Digital twins for IoT-driven energy systems: A survey. *IEEE Access, 12*, 177123–177143. https://doi.org/10.1109/ACCESS.2024.3506660
6. Kumari, N., Sharma, A., Tran, B., Chilamkurti, N., & Alahakoon, D. (2023). A comprehensive review of digital twin technology for grid-connected microgrid systems. *Energies, 16*(14), 5525. https://doi.org/10.3390/en16145525
7. Guerrero, J. M., Vasquez, J. C., Matas, J., Vicuña, L. G., & Castilla, M. (2011). Hierarchical control of droop-controlled AC and DC microgrids—A general approach toward standardization. *IEEE Transactions on Industrial Electronics, 58*(1), 158–172. https://doi.org/10.1109/TIE.2010.2066534
8. Parisio, A., Rikos, E., & Glielmo, L. (2014). A model predictive control approach to microgrid operation optimization. *IEEE Transactions on Control Systems Technology, 22*(5), 1813–1827. https://doi.org/10.1109/TCST.2013.2295737
9. Glavic, M., Moreno, R., Bi, T., & Shahidehpour, M. (2021). Reinforcement learning for electric power system decision and control: Past considerations and perspectives. *IEEE Transactions on Smart Grid, 12*(6), 5128–5142. https://doi.org/10.1109/TSG.2021.3070934
10. Zhang, Z., Wang, J., & Chen, B. (2021). Deep reinforcement learning for energy management in microgrids. *Applied Energy, 288*, 116618. https://doi.org/10.1016/j.apenergy.2021.116618
11. Ugwu, C. N., Ogenyi, F. C., Ugwu, J. N., Ugwu, P. C. O., & Okon, M. B. (2026). Integration of AI-driven digital twins for real-time optimization of renewable energy grids. *Frontiers in Energy Research, 14*, 1–13. https://doi.org/10.3389/fenrg.2026.1748233

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

12. Xu, B., Oudalov, A., Ulbig, A., & Andersson, G. (2018). Modeling of lithium-ion battery degradation for cell life assessment. *IEEE Transactions on Smart Grid, 9*(2), 1131–1140. https://doi.org/10.1109/TSG.2016.2578950
13. Xu, B., Shi, Y., Kirschen, D. S., & Zhang, B. (2018). Optimal battery participation in frequency regulation markets considering degradation cost. *IEEE Transactions on Power Systems, 33*(5), 4908–4918. https://doi.org/10.1109/TPWRS.2018.2799437
14. Ton, T. N., Le, P. M., Pham, L. H., & Le, T. M. (2025). Multi-objective optimization of electric distribution systems with integrated distributed generation using deep reinforcement learning. *Engineering, Technology & Applied Science Research, 15*(2), 22166–22171.