East Asian Journal of Multidisciplinary Research (EAJMR)
Vol. 5, No. 10, 2026: 4091 - 4110

Digital Twin-Assisted PPO-Based Energy Management for PV–BESS Operation in Active Distribution Networks Under Uncertainty

MinhPhong Le
Faculty of Electrical and Electronic, Thu Duc College of Technology, Viet Nam
Corresponding Author: Le MinhPhong minhphongle232@gmail.com

A R T I C L E I N F O

Keywords: Digital Twin, Deep Reinforcement Learning, Proximal Policy Optimization, Battery Energy Storage System, Active Distribution Network

Received : 12, August
Revised : 13, September
Accepted: 17, October

©2026 Le : This is an open-access article distributed under the terms of the Creative Commons Atribusi 4.0 Internasional.

A B S T R A C T

This study proposes a Digital Twin-assisted Proximal Policy Optimization (PPO) framework for PV–BESS energy management in active distribution networks under uncertainty. The Digital Twin continuously synchronizes operating information and validates the network state through AC power-flow calculations, while the PPO agent determines adaptive BESS charging and discharging actions. Monte Carlo scenarios are used to represent variations in PV generation and load demand, while operating cost, voltage deviation, and battery degradation are incorporated into the control objective. The framework is evaluated on the IEEE 33-bus distribution system with 60% PV penetration and a 1.5 MWh BESS. Results show a 22.2% reduction in operating cost compared with rule-based control, together with improved voltage regulation, reduced ESS cycling, and lower operating-cost variability under stochastic operating conditions.

DOI prefix: https://doi.org/10.55927/eajmr.v5i10.17
ISSN-E: 2828 – 1519 (Online)
https://eajmrformosa.my.id/index.php/eajmr

INTRODUCTION

The increasing penetration of photovoltaic (PV) generation is changing the operating characteristics of distribution networks. Unlike conventional systems designed mainly for one-way power delivery, active distribution networks (ADNs) with distributed renewable resources experience bidirectional power flows, variable generation, and stronger interactions between local generation and demand. Variations in solar irradiation and load demand can consequently affect feeder power flows and bus-voltage profiles, creating challenges for reliable network operation (Ardebili et al. 2021), (Liu, Zhang, and Li 2021), (Zhou, Yang, and Shao 2021), (Musa and Ghosh 2023). These conditions require energy-management strategies that can adapt to continuously changing operating states rather than relying only on predetermined schedules.

Battery energy storage systems (BESSs) provide an effective means of mitigating renewable variability. By shifting energy between operating periods, BESSs can absorb surplus PV generation and supply stored energy when renewable output is insufficient, thereby improving renewable-energy utilization and supporting network operation (Mohamed, Kaddoum, and Hassan 2026), (Kabir, Halder, and Ray 2024), (Nneoma Ugwu et al. 2026). However, coordinated PV–BESS operation involves more than determining charging and discharging actions. The control strategy must consider network conditions, grid power exchange, battery state-of-charge (SOC), operating limits, and battery utilization simultaneously (Kumari et al. 2023), (Guerrero et al. 2011).

Model-based approaches such as optimal power flow (OPF) and model predictive control (MPC) can explicitly incorporate electrical constraints and operational objectives (Musa and Ghosh 2023), (Parisio, Rikos, and Glielmo 2014). Their effectiveness, however, depends on the accuracy of system models and operating forecasts. Significant deviations in PV generation or load demand can reduce the consistency of the resulting decisions, while repeatedly solving optimization problems under rapidly changing conditions may increase the computational burden. These limitations have encouraged the application of deep reinforcement learning (DRL), which treats energy management as a sequential decision-making problem (Glavic et al. 2021), (Zhang, Wang, and Chen 2021). A trained DRL agent learns a policy that maps system states to control actions, making it suitable for continuously adjusting BESS operation according to renewable generation, demand, SOC, and network conditions.

Nevertheless, the effectiveness of DRL depends strongly on the quality and physical consistency of the information provided to the agent. Digital Twin (DT) technology offers a promising solution by maintaining an updated virtual representation of the physical energy system. A DT can integrate system measurements with a virtual model to support monitoring, simulation, state synchronization, and operational decision-making (Liu, Zhang, and Li 2021), (Guerrero et al. 2011), (Wang et al. 2021). Its application to renewable-energy and grid-connected systems has also demonstrated potential for improving system representation and operational management (Liu, Zhang, and Li 2021), (Nneoma Ugwu et al. 2026), (Wang et al. 2024). Recent investigations have further demonstrated the potential of DT technology in renewable-energy systems and

grid-connected microgrids (Musa and Ghosh 2023), (Kabir, Halder, and Ray 2024), (Parisio, Rikos, and Glielmo 2014), (Glavic et al. 2021). When incorporated into an intelligent energy management framework, the virtual model can provide an intermediate environment in which the current operating condition is reproduced and checked using the underlying electrical model. This capability is particularly relevant to DRL-based control, because an inaccurate system state may lead the learning agent to generate inappropriate control actions.

Despite these developments, the coordinated use of DT, DRL, and degradation-aware BESS management for PV-rich ADNs remains insufficiently explored. This study addresses this gap by integrating a MATLAB/MATPOWER-based DT with a PPO controller for PV–BESS energy management under stochastic PV and load conditions. The framework considers operating cost, voltage deviation, and battery degradation within a closed-loop control process.

The practical significance of this research is associated with the simultaneous treatment of three factors that strongly affect renewable-rich distribution-network operation: operating-cost variability, voltage regulation, and battery utilization. In particular, increasing PV penetration can reduce grid energy purchases but may also increase voltage deviations and operating uncertainty when renewable output changes rapidly. A BESS can mitigate these effects, but excessive charging and discharging may increase energy throughput and degradation-related operating costs. Therefore, an effective energy-management strategy should not optimize economic performance independently from network voltage and battery operating conditions. The proposed DT-assisted PPO framework addresses this combined requirement by using AC power-flow-validated network states as inputs to the controller and by incorporating operating cost, voltage deviation, and degradation-related penalties into the control objective. The resulting evaluation provides quantitative evidence of the practical value of this integrated approach, including a 22.2% reduction in operating cost, 51.9% lower maximum voltage deviation, and 55.9% lower average voltage deviation compared with rule-based control.

The main contributions are as follows:

i) A Digital Twin-assisted PPO framework is developed for coordinated PV–BESS energy management using AC power-flow-validated network states.
ii) Monte Carlo scenarios are incorporated to represent PV and load uncertainty, while operating cost, voltage deviation, and battery degradation are considered in the PPO reward.
iii) A systematic evaluation is performed using economic, voltage, BESS-operation, statistical, and uncertainty-sensitivity indicators.
iv) The proposed PPO-DRL+DT strategy is systematically compared with RBC, OPF, and standalone PPO-DRL to assess its effectiveness under uncertain operating conditions.

THEORETICAL REVIEW

The operation of PV-integrated active distribution networks involves the interaction of renewable generation, electrical demand, and energy storage. Variations in PV output and load demand directly affect feeder power flow and bus-voltage conditions. BESS provides a controllable resource for balancing these variations through coordinated charging and discharging, while its SOC and operating limits determine the feasible control range. Battery utilization should also be considered because frequent cycling and unfavorable SOC conditions can increase degradation-related operating costs.

For sequential energy-management problems, DRL represents the network operating condition as a state and determines an appropriate control action through a learned policy. In this framework, PPO is used to obtain continuous BESS control actions. Its actor–critic structure separates action generation from value estimation, while the clipped policy update limits excessive changes during training and supports stable policy optimization.

Digital Twin technology complements the learning controller by maintaining a virtual representation of the distribution network. The virtual model receives operating information and reproduces the corresponding electrical state through AC power-flow calculations. The resulting synchronized state provides the DRL controller with network-level information, including PV generation, load demand, BESS SOC, and bus voltages, thereby linking the learned control policy with the electrical behavior of the distribution system.

In this study, these concepts are integrated into a closed-loop energy-management framework in which the Digital Twin represents the network environment and PPO determines the BESS control action. This combination provides the theoretical basis for coordinated PV–BESS operation under stochastic operating conditions while accounting for network constraints and battery-related effects.

METHODOLOGY

PV–ESS Distribution Network Configuration

The proposed energy management framework is developed using the IEEE 33-bus active distribution network integrated with photovoltaic (PV) generation and a battery energy storage system (BESS). The network operates at 12.66 kV with a peak demand of 3.72 MW and 2.30 MVAr. PV penetration is set at 60% of the peak load, while the BESS has an energy capacity of 1.5 MWh and a maximum charging/discharging power of ±0.5 MW. The scheduling horizon is 24 h with a 15-min control interval, corresponding to 96 operating steps. Bus voltages are constrained within 0.95–1.05 p.u.

As shown in Figure 1, the control architecture consists of three interconnected components: the IEEE 33-bus distribution network, the MATLAB/MATPOWER-based Digital Twin, and the PPO-based DRL controller. Operating measurements, including load demand, PV generation, BESS SOC, and bus voltages, are transferred to the Digital Twin for state reconstruction and AC power-flow validation. The resulting synchronized state is then provided to the PPO controller to determine the BESS charging/discharging action.

The resulting closed-loop architecture enables the BESS control policy to respond to the current network condition through continuous state synchronization and feedback, rather than relying solely on predetermined ESS schedules.

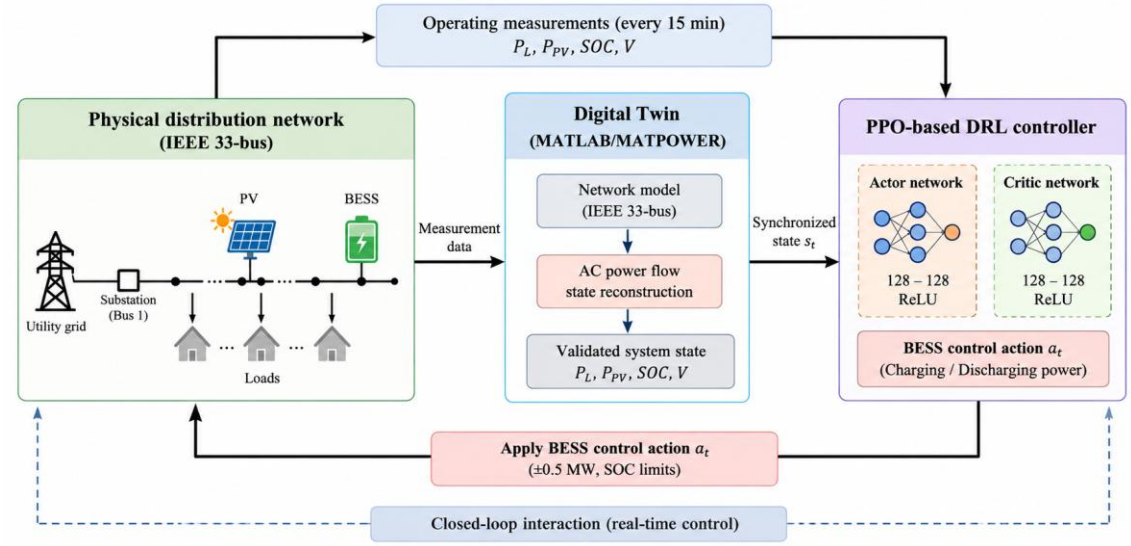


Figure 1. Closed-loop Digital Twin-assisted PPO architecture for PV–BESS energy management.

Digital Twin-Based State Synchronization

The Digital Twin (DT) serves as the interface between the distribution-network model and the reinforcement-learning controller. Rather than allowing the DRL agent to interact directly with an unconstrained data-driven representation, the DT reconstructs the network operating state and verifies it using the electrical model. In this study, the synchronized information is defined using load demand, PV output, BESS SOC, and network voltage information.

At each control interval tt, the measurement vector is defined as

$$x_t = [P_t^{Load}, P_t^{PV}, SOC_t, V_t] \quad ((1)$$

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 ttt.

The DT maps the measured operating information into a synchronized virtual state:

$$x_t^{DT} = f_{DT}(x_t; \theta_{DT}) \quad ((2)$$

Where $f_{DT}(\cdot)$ represents the DT state-reconstruction process, x_t represents the measurement vector, x_t^{DT} $\hat{s_t}$ denotes the estimated virtual state and θ_{DT} represents the Digital Twin model parameters.

The virtual distribution system is implemented in MATLAB/MATPOWER using AC power-flow calculations. Network topology, load demand, PV generation, and BESS operating conditions are updated according to the current operating data. The calculated state is then checked against the electrical constraints before being supplied to the PPO agent.

The quality of synchronization is evaluated through the discrepancy between the measured physical state and its DT representation:

$$e_t = \| x_t - \tilde{x}_t \| \quad ((3)$$

where $\tilde{x}_t$ denotes the corresponding state reconstructed by the Digital Twin.

After the PPO agent generates an ESS control action a_t, the action is applied to the virtual network and the resulting electrical state is recalculated. The updated state is then returned to the DT, forming a closed-loop interaction:

$$x_t \rightarrow s_t^{DT} \rightarrow a_t \rightarrow x_{t+1} \quad (4)$$

In this study, the Digital Twin is implemented as a simulation-based cyber–physical environment rather than a hardware testbed. The virtual network is continuously updated using operating measurements and validated through AC power-flow calculations at each 15-min control interval. Therefore, the Digital Twin is used to emulate state synchronization and closed-loop interaction between the distribution network and the DRL controller, while physical hardware implementation is beyond the scope of this study.

Stochastic Operating Data Generation

Because PV production and load demand vary with operating conditions, Deterministic input data alone cannot adequately characterize the controller response to variable PV generation and load demand. A stochastic dataset is therefore generated from the IEEE 33-bus operating profiles.

The basic dataset is constructed with a 15-min resolution over a 24-h period, giving 96 time steps per daily operating cycle. Each sample contains load demand, PV output, bus voltage, and BESS SOC information.

PV uncertainty is introduced by perturbing the forecast PV output according to

$$P_{PV,t}^{(k)} = \hat{P}_{PV,t}\left(1 + \delta_{PV,t}^{(k)}\right) \quad (5)$$

where $\hat{P}_{PV,t}$ is the forecast value and $\delta_{PV,t}^{(k)}$ represents the uncertainty associated with scenario k.

Similarly, the load profile is modified as

$$P_{L,t}^{(k)} = \hat{P}_{L,t}\left(1 + \delta_{L,t}^{(k)}\right) \quad (6)$$

The uncertainty ranges used in the original study are ±20% for PV generation and ±15% for load demand. A total of 100 Monte Carlo scenarios are generated for stochastic training and subsequent policy evaluation.

For each scenario ω_k, the resulting PV and load trajectories are passed through the Digital Twin. The corresponding AC power-flow solution provides the voltage profile and other network states required by the DRL environment.

The scenario-based performance is subsequently characterized using the expected value

$$E[J] = \frac{1}{N_s}\sum_{k=1}^{N_s} J(\omega_k) \quad (7)$$

and the corresponding variance

$$Var[J] = \frac{1}{N_s}\sum_{k=1}^{N_s}[J(\omega_k) - E[J]]^2 \quad (8)$$

where $N_s = 100$ is the number of Monte Carlo scenarios and $J(\omega_k)$ denotes the operating performance under scenario k. These two indicators are used to assess both average performance and sensitivity to stochastic operating conditions.

Electrical and Battery Operating Models

AC Distribution-Network Model

The electrical behavior of the distribution network is represented using nonlinear AC power-flow equations. For each bus i, the active and reactive power balances are determined from the nodal voltage magnitudes, voltage angles, network admittance, local demand, PV injection, and ESS injection.

The active and reactive power balances can generally be expressed as

$$P_i = V_i \sum_j V_j \left[G_{ij}\cos(\theta_i - \theta_j) + B_{ij}\sin(\theta_i - \theta_j)\right] \quad (9)$$

$$Q_i = V_i \sum_j V_j \left[G_{ij}\sin(\theta_i - \theta_j) - B_{ij}\cos(\theta_i - \theta_j)\right] \quad (10)$$

Here, V_i and θ_i are the voltage magnitude and angle at bus i, while G_{ij} and B_{ij} are the conductance and susceptance elements of the network admittance matrix.

The voltage-security constraint is imposed as

$$V_i^{\min} \le V_i \le V_i^{\max} \quad (11)$$

with

$$V_i^{\min} = 0.95\ \text{p.u.}, \qquad V_i^{\max} = 1.05\ \text{p.u.} \quad (12)$$

The AC power-flow solution is recalculated after every BESS control action, thereby ensuring that the control policy is evaluated against the electrical response of the distribution network rather than an independent battery model.

BESS State and Operating Constraints

The BESS energy state is represented through its SOC. The SOC transition is determined by the charging and discharging power and the corresponding efficiencies:

$$SOC_{t+1} = SOC_t + \frac{\eta_c P_t^{ch} \Delta t}{E_B} - \frac{P_t^{dis} \Delta t}{\eta_d E_B} \quad (13)$$

where E_B is the nominal battery energy capacity, η_c and η_d are the charging and discharging efficiencies, and Δt is the control interval.

The BESS must satisfy the operating constraints

$$SOC^{\min} \le SOC_t \le SOC^{\max} \quad (14)$$

and

$$-P_B^{\max} \le P_{B,t} \le P_B^{\max} \quad (15)$$

The sign convention adopted in the DRL formulation defines positive $P_{B,t}$ as charging and negative $P_{B,t}$ as discharging.

Battery Degradation Representation

Battery aging is explicitly incorporated into the energy-management formulation. Instead of using a computationally intensive electrochemical aging

model, the study employs a surrogate degradation representation that accounts for energy throughput, deviation from the preferred SOC region, and SOC fluctuations.

The degradation penalty can be represented as

$$D_{ESS,t} = \alpha|P_{B,t}|\Delta t + \beta|SOC_t - SOC_{ref}| + \kappa|\Delta SOC_t| \quad (16)$$

where α represents the cycling-related degradation component, β penalizes operation away from the preferred SOC region, SOC_{ref} is the reference SOC, and κ accounts for frequent SOC fluctuations.

The adopted coefficients are

$$\alpha = 1.0, \qquad \beta = 0.20, \qquad \kappa = 0.05.$$

This formulation provides a computationally efficient means of introducing battery-aging effects into the DRL control environment with low computational overhead. In this study, the surrogate degradation model is adopted because of its lower computational burden compared with more detailed electrochemical or rainflow-based approaches.

Reinforcement Learning Formulation

The BESS energy-management problem is formulated as a Markov Decision Process (MDP). The Digital Twin constitutes the environment, while the PPO agent determines the BESS operating command at each time step. The formulation allows the controller to learn a control policy from repeated interactions with the stochastic network environment.

The MDP is defined by the tuple

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, R, \gamma) \quad (17)$$

where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ represents the transition dynamics, R is the reward function, and γ is the discount factor.

The state vector is defined as

$$s_t = [P_{L,t}, P_{PV,t}, SOC_t, \boldsymbol{V}_t] \quad (18)$$

Thus, the agent receives information on the current load, PV generation, BESS energy state, and distribution-network voltage condition.

The control action corresponds to the BESS charging/discharging command:

$$a_t = P_{B,t} \quad (19)$$

Positive values indicate charging, whereas negative values represent discharging. The action is subsequently passed to the Digital Twin, where the AC network state and battery state are updated.

The reward function is constructed to penalize operating cost, voltage deviation, and battery degradation:

$$r_t = -\left[C_{op,t} + w_1 V_{dev,t} + w_2 D_{ESS,t}\right] \quad (20)$$

where $C_{op,t}$ is the operating-cost component, $V_{dev,t}$ represents voltage deviation, $D_{ESS,t}$ is the degradation penalty, and w_1 and w_2 are weighting coefficients.

The operational-cost component considers electricity purchased from the utility grid together with the BESS operating cost. The resulting formulation establishes a direct trade-off between short-term economic performance, voltage regulation, and battery lifetime preservation.

The cumulative discounted objective is therefore expressed as

$$(\pi) = E_{\pi}\left[\sum_{t=0}^{T-1} \gamma^{t} r_{t}\right] \quad (21)$$

where π denotes the control policy and T is the number of control intervals.

This formulation allows the agent to account for the future consequences of its present charging or discharging decisions, particularly the influence of intensive ESS operation on battery degradation.

PPO Controller Design

Proximal Policy Optimization is selected as the DRL algorithm for solving the continuous ESS control problem. The PPO implementation uses separate actor and critic networks. The actor generates the BESS control command, whereas the critic estimates the state-value function used for policy improvement.

The input layer contains four principal information groups:

1. load demand;
2. PV generation;
3. BESS SOC; and
4. network voltage information.

Both actor and critic networks contain two fully connected hidden layers with 128 neurons per layer. ReLU activation functions are used in the hidden layers. The actor produces the ESS charging/discharging command, while the critic estimates the value of the current state.

The PPO policy update is based on the clipped surrogate objective:

$L^{CLIP}(\theta) = E_t[\min(r_t(\theta)A_t, \text{clip}(r_t(\theta),1-\epsilon, 1+\epsilon)A_t)]$ (22)

where $r_t(\theta)$ is the probability ratio between the updated and previous policies, A_t is the estimated advantage, and ϵ is the clipping parameter.

The clipping mechanism limits excessively large policy changes during each update and thereby supports stable policy learning.

The principal PPO settings retained from the original study are summarized Table 1 below.

Table 1. PPO Hyperparameter Settings

Parameter	Value
Learning rate	1×10^{-4}
Discount factor, γ	0.99
Clipping factor, ϵ	0.20
Batch size	64
Training epochs/update	10
Hidden neurons/layer	128
Number of training episodes	2000
Optimizer	Adam

These parameters are used consistently throughout the training procedure.

Training and Validation Procedure

The training and validation process is designed to enable the PPO agent to learn an adaptive BESS control policy under stochastic operating conditions. As summarized in Figure 2, the procedure consists of two main stages: PPO training and policy validation.

During the training stage, the IEEE 33-bus network, PV system, BESS, and model parameters are first initialized. A set of 100 stochastic operating scenarios is then generated using Monte Carlo sampling. For each scenario, the initial state is defined by the load demand, PV generation, BESS SOC, and network voltage profile. At each 15-min control interval, the Digital Twin updates the operating conditions and performs an AC power-flow calculation to obtain the validated system state.

The synchronized state is provided to the PPO controller, which determines the BESS charging or discharging action. The selected action is applied to the BESS and distribution-network model, after which the battery SOC and electrical state are updated. The corresponding reward is calculated from the operating cost, voltage deviation, and battery degradation:

$$s_t \rightarrow \pi_\theta(a_t|s_t) \rightarrow a_t \rightarrow DT \rightarrow s_{t+1} \qquad (23)$$

The transition (s_t, a_t, r_t, s_{t+1}) is then used for PPO policy optimization. The actor and critic networks are updated using the clipped PPO objective and the Adam optimizer. This process is repeated over the 96 control intervals of each operating episode and continued for 2000 training episodes.

After training, the learned PPO policy is fixed and evaluated using the stochastic operating scenarios. The validation procedure follows the same Digital Twin-based simulation process but without further policy updating. The performance is assessed using economic, electrical, battery-operation, and statistical indicators. In addition, the controller is tested under different PV uncertainty levels to examine its robustness under increasingly variable renewable generation.

The complete training and validation procedure is illustrated in Figure 2.

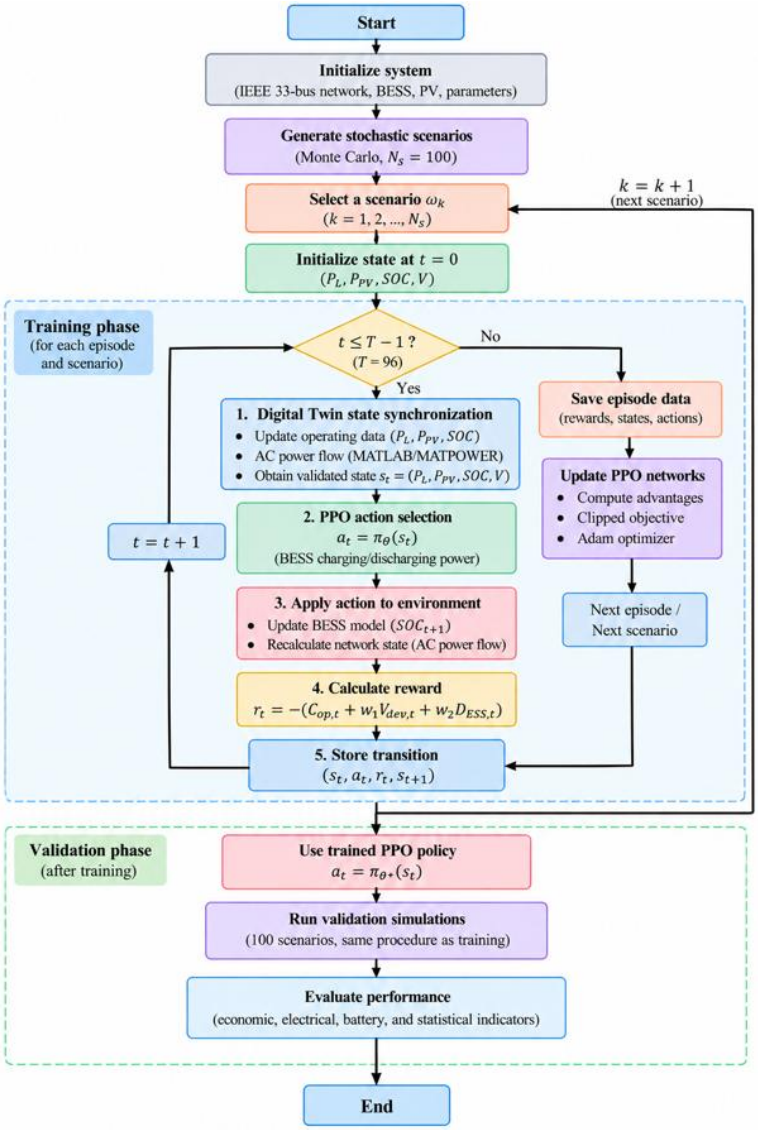


Figure 2. Flowchart of the Digital Twin-Assisted PPO Training and Validation Procedure

Data Analysis and Performance Evaluation

The generated data are analyzed at both time-step and scenario levels. Each 24-h scenario contains 96 intervals of load, PV generation, BESS SOC, BESS power, and bus-voltage data. After each PPO action, the Digital Twin performs AC power-flow calculations and records the resulting system states.

For each of the 100 Monte Carlo scenarios, operating cost, voltage deviation, ESS cycling, SOC variance, and voltage-violation rate are calculated. Mean cost, standard deviation, coefficient of variation (CV), and the best–worst cost range are then used to assess performance consistency:

$$\mathrm{CV} = \frac{\sigma_C}{\mu_C} 100\% \qquad (24)$$

where μ_C and σ_C are the mean and standard deviation of daily operating cost, respectively. Comparative improvements are calculated relative to RBC.

Finally, PV uncertainty is increased from 10% to 40% to examine its effect on operating cost and controller performance.

RESULTS

Simulation Setup

The proposed Digital Twin-assisted PPO framework was evaluated on the IEEE 33-bus active distribution network using MATLAB/MATPOWER. The test system operates at 12.66 kV with a peak demand of 3.72 MW and 2.30 MVAr. PV penetration was set to 60% of the peak load, while the BESS capacity and power rating were 1.5 MWh and ±0.5 MW, respectively. The control interval was 15 min over a 24-h scheduling horizon. A total of 100 Monte Carlo scenarios were considered to represent stochastic variations in PV generation and load demand.

The proposed DRL+DT strategy was compared with three reference approaches: Rule-Based Control (RBC), deterministic Optimal Power Flow (OPF), and standalone PPO-based DRL without Digital Twin integration. The evaluation considered economic performance, voltage regulation, ESS operating behavior, and robustness under stochastic conditions.

PPO Training Convergence and Economic–Voltage Performance

The learning behavior of the PPO controller was first examined over 2000 training episodes. The average episode reward improved from approximately −1250 at the beginning of training to around −295 after 2000 episodes. The rapid improvement during the initial training stage indicates that the agent progressively learned more effective BESS control actions. After approximately 1800 episodes, the reward became relatively stable, indicating convergence of the learned policy.

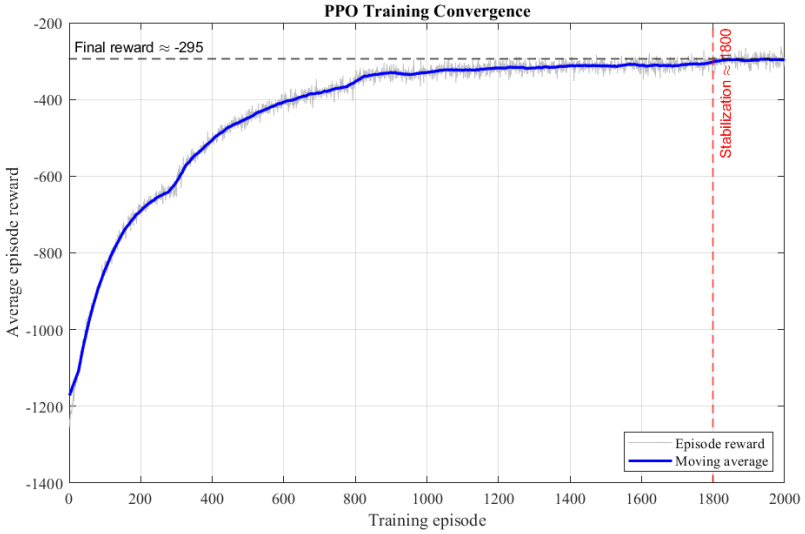


Figure 3. PPO training convergence curve.

The convergence behavior demonstrates that the PPO agent can learn a stable control policy through repeated interaction with the Digital Twin environment. The use of synchronized network states enables the agent to evaluate its actions using the corresponding electrical operating condition rather than relying only on the battery state.

The economic and voltage-regulation performance of the four control strategies is then summarized using normalized improvement indicators. The analysis was conducted by first obtaining the absolute performance of each

control strategy over the evaluated operating horizon and then normalizing the results with respect to the RBC benchmark. This procedure allows the relative benefits of OPF, PPO-DRL, and PPO-DRL+DT to be compared using the same reference condition. Lower operating cost and lower voltage deviation are treated as improvements, and the corresponding percentage reductions are calculated from the absolute simulation results. Table 2 presents the improvement relative to RBC.

Table 2. Relative performance improvement compared with RBC.

Method	Cost improvement (%)	Maximum voltage-deviation improvement (%)	Average voltage-deviation improvement (%)
OPF	13.1	30.9	38.2
PPO-DRL	16.9	39.5	47.1
PPO-DRL + DT	22.2	51.9	55.9

The proposed DRL+DT configuration provides the largest improvement for all three indicators. In particular, the operating cost is reduced by 22.2% relative to RBC, while the maximum and average voltage deviations are reduced by approximately 51.9% and 55.9%, respectively. These normalized indicators are calculated from the corresponding absolute simulation results.

The corresponding voltage profiles and operating-cost comparison are illustrated in Figure 4.

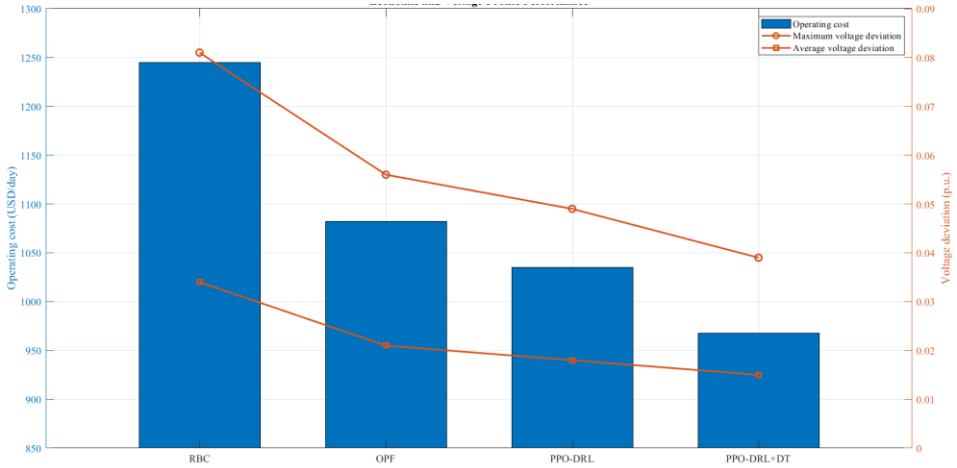


Figure 4. Voltage-profile and operating-cost comparison of the evaluated energy-management strategies.

Figure 4 illustrates the corresponding differences in operating cost and voltage deviation among the four control strategies. The PPO-DRL+DT configuration exhibits the lowest values for both indicators.

ESS Operating Characteristics and Voltage Reliability

The effect of the proposed controller on ESS utilization and network reliability was further examined under the 100 stochastic operating scenarios. To provide a different representation from the original results, Table 3 reports the relative reduction of the main indicators with respect to RBC.

Table 3. Relative Reduction of ESS and Robustness Indicators

Method	ESS-cycle reduction (%)	SOC-variance reduction (%)	Cost-variability reduction (%)	Voltage-violation reduction (%)
OPF	25.0	26.3	25.8	46.0
PPO-DRL	32.1	42.1	35.2	66.9
PPO-DRL + DT	**42.9**	**52.6**	**51.6**	**81.5**

The reduced cycling and lower SOC variation obtained with PPO-DRL+DT indicate a more controlled pattern of BESS utilization, which is consistent with the degradation-aware formulation adopted in the controller. The average number of ESS cycles decreases from 2.8 cycles/day for RBC to 1.6 cycles/day for PPO-DRL+DT. The lower SOC variance also indicates more controlled BESS operation under stochastic conditions.

This behavior is important from a battery-management perspective because unnecessary charge/discharge activity increases energy throughput and may accelerate degradation. The lower SOC variance also indicates that the controller maintains a more controlled battery operating state under stochastic conditions.

The corresponding ESS trajectories, Monte Carlo cost distribution, and voltage-deviation behavior are presented in Figure 5.

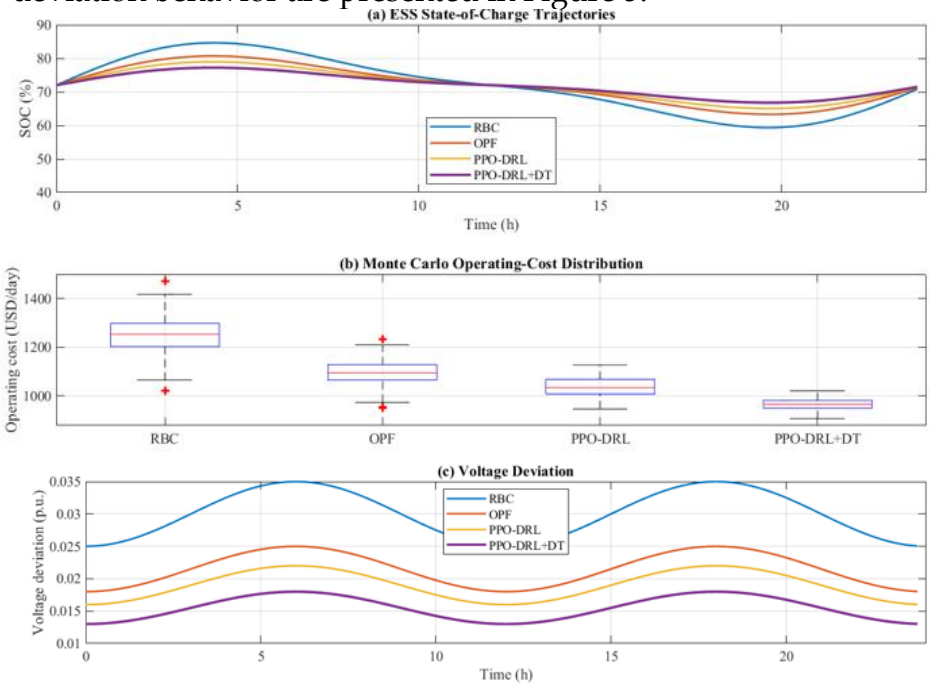


Figure 5. ESS operation and system robustness under stochastic operating conditions.

The graphical results complement Table 3 by showing the interaction between ESS operation and network performance. The reduced cycling behavior of the PPO-DRL+DT controller is accompanied by lower cost dispersion and improved voltage behavior, demonstrating that the controller does not obtain economic improvement simply by excessively exploiting the BESS.

Statistical Validation of Operating Cost

The consistency of the operating results was further evaluated using the 100 stochastic scenarios. Instead of repeating the original absolute-value comparison, the coefficient of variation (CV) was calculated as equation (24)

Table 4. Statistical Characteristics of Operating Cost

Method	Mean cost (USD/day)	Standard deviation (USD/day)	Coefficient of variation (%)	Best–worst range (USD/day)
RBC	1245	76	6.10	206
OPF	1082	52	4.81	145
PPO-DRL	1035	38	3.67	123
PPO-DRL + DT	**968**	**21**	**2.17**	**67**

The PPO-DRL+DT approach has the smallest coefficient of variation, approximately 2.17%, compared with 6.10% for RBC. It also produces the narrowest best–worst cost range, indicating more consistent operating-cost performance under stochastic conditions. The Monte Carlo cost distribution already included in Figure 5 can also be used to visually support the statistical characteristics reported in Table 4. Therefore, a separate figure for Table 4 is not necessary and helps avoid unnecessary duplication.

Sensitivity to PV Generation Uncertainty

The robustness of the learned control policy was further assessed by gradually increasing the PV uncertainty level from 10% to 40%. The corresponding operating-cost behavior is summarized in Table 5.

Table 5. Operating-Cost Variation under Different PV Uncertainty Levels

PV uncertainty (%)	Operating cost (USD/day)	Cost increase relative to 10% (%)
10	942	0.0
20	968	2.8
30	1014	7.6
40	1087	15.4

The operating cost increases as PV uncertainty becomes more severe. This trend is expected because greater uncertainty in renewable generation increases the difficulty of maintaining power balance and coordinating ESS operation. The

results indicate a gradual increase in operating cost as PV uncertainty becomes more severe, without an abrupt change over the tested range.

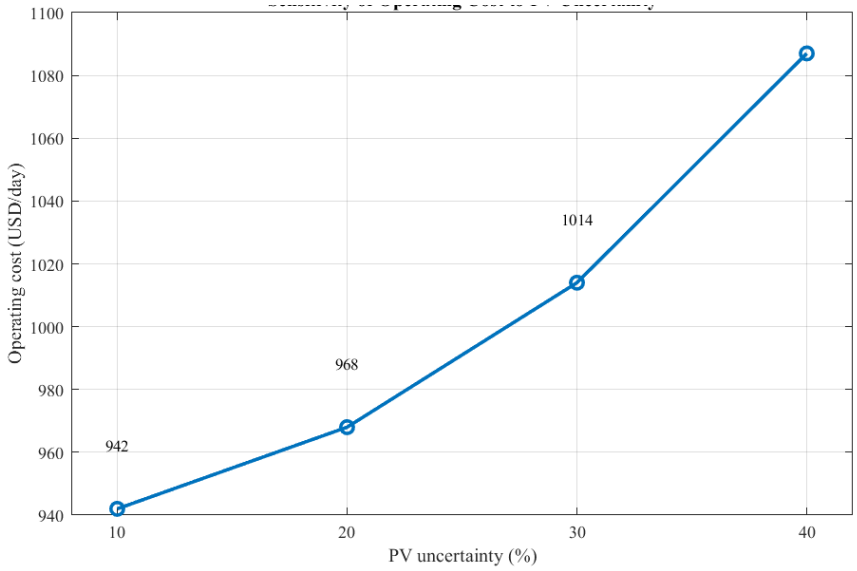


Figure 6. Sensitivity of operating cost to PV generation uncertainty.

Figure 6 shows a gradual increase in operating cost as PV uncertainty increases from 10% to 40%. The results indicate that increasing renewable uncertainty has a measurable but gradual effect on the economic performance of the learned controller.

Comparison with Existing Energy-Management Strategies

Table 6 compares RBC, OPF, standalone DRL, and the proposed PPO-DRL+DT framework using economic, voltage, ESS, and robustness indicators. The proposed DRL+DT achieves the lowest operational cost of 968 USD/day, corresponding to a 22.2% reduction relative to RBC. It also obtains the lowest maximum and average voltage deviations of 0.039 and 0.015 p.u., respectively.

Table 6. Comprehensive Comparison of Energy-Management Strategies

Method	Cost Reduction (%)	Max. Voltage Deviation (p.u.)	Avg. Voltage Deviation (p.u.)	Avg. ESS Cycles/Day	SOC Variance	Mean Cost under Uncertainty (USD/day)	Voltage Violations (%)
RBC	0.0	0.081	0.034	2.8	0.19	1310	12.4
OPF	13.1	0.056	0.021	2.1	0.14	1150	6.7
DRL	16.9	0.049	0.018	1.9	0.11	1085	4.1
PPO-DRL+DT	**+22.2**	**0.039**	**0.015**	**1.6**	**0.09**	**1020**	**2.3**

Table 6 provides a consolidated comparison of the evaluated energy-management strategies in terms of economic performance, voltage regulation, and BESS operation. The proposed PPO-DRL+DT framework achieves the lowest maximum and average voltage deviations, at 0.039 and 0.015 p.u., respectively.

It also reduces the average BESS cycling frequency to 1.6 cycles/day and the SOC variance to 0.09, indicating more controlled battery operation. Under stochastic operating conditions, the framework achieves a mean operating cost of 1020 USD/day and a voltage violation rate of only 2.3%. These results demonstrate improved economic performance, voltage regulation, BESS utilization, and operational robustness compared with RBC, OPF, and standalone PPO-DRL+DT

DISCUSSION

The results reveal a clear advantage of combining Digital Twin-based state synchronization with PPO rather than relying on PPO alone. While standalone PPO-DRL reduces operating cost by 16.9% relative to RBC, the proposed PPO-DRL+DT achieves a 22.2% reduction. More importantly, maximum and average voltage deviations decrease by 51.9% and 55.9%, respectively. This improvement is consistent with the reported effectiveness of DRL for sequential energy-management problems (Glavic et al. 2021; Zhang, Wang, and Chen 2021), while indicating that AC power-flow-validated network states can further improve control decisions.

The Digital Twin provides an important link between the learned policy and the actual electrical behavior represented by the network model. Previous studies have emphasized DT-based monitoring, state synchronization, and energy-system optimization (Liu, Zhang, and Li 2021; Zhou, Yang, and Shao 2021; Kabir, Halder, and Ray 2024; Wang et al. 2024). In the present framework, this capability is directly embedded in the control loop. The PPO agent therefore responds not only to PV output and BESS SOC, but also to the resulting network voltage conditions. This provides a plausible explanation for the lower voltage deviations obtained with PPO-DRL+DT.

The BESS results further strengthen this finding. Average ESS cycling decreases from 2.8 to 1.6 cycles/day, while SOC variance decreases from 0.19 to 0.09. These results are consistent with degradation-aware battery-management studies that emphasize limiting unnecessary battery utilization (Xu et al. 2018; Chen et al. 2025). The proposed formulation therefore achieves better economic and voltage performance without increasing battery cycling.

Under stochastic conditions, the proposed method also shows greater consistency. Its mean operating cost is 968 USD/day with a CV of 2.17%, compared with 1245 USD/day and 6.10% for RBC. Moreover, increasing PV uncertainty from 10% to 40% raises the operating cost gradually from 942 to 1087 USD/day. These results suggest that the proposed controller can maintain relatively stable performance as renewable uncertainty increases, although broader uncertainty ranges and larger networks require further investigation.

Compared with conventional OPF and MPC approaches, which repeatedly solve optimization problems for operating decisions (Parisio, Rikos, and Glielmo 2014), the proposed framework learns a reusable control policy while retaining network-state validation through the Digital Twin. The main value of the proposed approach therefore lies in the integration of DRL adaptability, Digital Twin-based electrical validation, and degradation-aware BESS control within one closed-loop framework. Nevertheless, the present Digital Twin is simulation-

based rather than a physical testbed; practical deployment will require further validation of communication latency, model mismatch, hardware implementation, and scalability.

CONCLUSIONS AND RECOMMENDATIONS

This study developed a Digital Twin-assisted PPO framework for PV–BESS energy management in an active distribution network under stochastic operating conditions. The framework combines synchronized network-state reconstruction, AC power-flow validation, Monte Carlo uncertainty representation, and degradation-aware BESS control. Evaluation on the IEEE 33-bus system demonstrated reduced operating cost, improved voltage regulation, lower ESS cycling, and more consistent performance compared with RBC, OPF, and standalone PPO-based DRL. The results indicate that Digital Twin-assisted state synchronization can improve the effectiveness of PPO-based BESS control when network operating conditions vary over time. For practical implementation, reliable measurement synchronization, accurate network modeling, communication performance, and appropriate BESS operating constraints should be considered to maintain physically feasible control actions.

FURTHER STUDY

Future work should extend the proposed framework to larger active distribution networks and more diverse operating conditions. The integration of electric vehicles, demand response, and multiple distributed energy resources can also be investigated. Further studies should examine communication latency, model mismatch, and practical cyber–physical implementation to assess the scalability and real-time applicability of the Digital Twin–DRL framework.

ACKNOWLEDGMENT

The author gratefully acknowledges the Faculty of Electrical and Electronic Engineering, Thu Duc College of Technology, Vietnam, for providing the academic environment and institutional support for this study.

REFERENCES

Ardebili, A A, M Zappatore, A.I.H.A. Ramadan, A Longo, and A Ficarella. 2024. "Digital Twins of Smart Energy Systems: A Systematic Literature Review on Enablers, Design, Management and Computational Challenges." *Energy Informatics* 7(1, Art. no. 94). doi:10.1186/s42162-024-00385-5.

B. Xu, A. Oudalov, A. Ulbig, G. Andersson, and D. S. Kirschen. 2018. "Modeling of Lithium-Ion Battery Degradation for Cell Life Assessment." *IEEE Transactions on Smart Grid* 9(2): 1131–1140. doi:10.1109/TSG.2016.2578950.

B. Xu, Y. Shi, D. S. Kirschen, and B. Zhang. 2018. "Optimal Battery Participation in Frequency Regulation Markets Considering Degradation Cost." *IEEE Transactions on Power Systems* 33(5): 4908–4918. doi:10.1109/TPWRS.2018.2799437.

Chen, J, X Wang, M Zhou, and Y Fu. 2025. "Battery Degradation-Aware Deep Reinforcement Learning for Energy Storage Scheduling in Distribution Networks." *IEEE Transactions on Sustainable Energy* 16(1): 215–227.

doi:10.1109/TSTE.2024.3478912.

Glavic, M, R Moreno, T Bi, and M Shahidehpour. 2021. "Reinforcement Learning for Electric Power System Decision and Control: Past Considerations and Perspectives." *IEEE Transactions on Smart Grid* 12(6): 5128–5142. doi:10.1109/TSG.2021.3070934.

Guerrero, J M, J C Vasquez, J Matas, L G Vicuña, and M Castilla. 2011. "Hierarchical Control of Droop-Controlled AC and DC Microgrids—A General Approach toward Standardization." *IEEE Transactions on Industrial Electronics* 58(1): 158–172. doi:10.1109/TIE.2010.2066534.

Kabir, Md Rafiul, Dipal Halder, and Sandip Ray. 2024. "Digital Twins for IoT-Driven Energy Systems: A Survey." *IEEE Access* 12: 177123–43. doi:10.1109/ACCESS.2024.3506660.

Kumari, N, A Sharma, B Tran, N Chilamkurti, and D Alahakoon. 2023. "A Comprehensive Review of Digital Twin Technology for Grid-Connected Microgrid Systems." *Energies* 16(14): 5525. doi:10.3390/en16145525.

Liu, Z, Y Zhang, and H Li. 2021. "Digital Twin for Power Systems: State-of-the-Art and Future Trends." *Renewable and Sustainable Energy Reviews* 152: 111675. doi:10.1016/j.rser.2021.111675.

Mohamed, Nahed Belhadj, Georges Kaddoum, and Md Zoheb Hassan. 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(February): 591–611. doi:10.1109/TMLCN.2026.3669837.

Musa, Usman Ibrahim, and Sukanta Ghosh. 2023. "Advancing Digital Twin through the Integration of New AI Algorithms." *International Research Journal of Engineering and Technology*: 922–32. https://www.researchgate.net/publication/369693333.

Nneoma Ugwu, Chinyere, Fabian Chukwudi Ogenyi, Jovita Nnenna Ugwu, Paul Chima Okechukwu Ugwu, and Micheal Ben Okon. 2026. "Integration of AI-Driven Digital Twins for Real-Time Optimization of Renewable Energy Grids." *Frontiers in Energy Research* 14(February): 1–13. doi:10.3389/fenrg.2026.1748233.

Parisio, A, E Rikos, and L Glielmo. 2014. "A Model Predictive Control Approach to Microgrid Operation Optimization." *IEEE Transactions on Control Systems Technology* 22(5): 1813–1827,. doi:10.1109/TCST.2013.2295737.

Ton, T N, P M Le, and T M Le. 2026. "Multi-Objective Optimization of Distributed Generation Placement and Sizing in Active Distribution Networks Considering Harmonic Distortion." *International Journal of Electrical and Computer Engineering* 16(2): 598–607,. doi:10.11591/ijece.v16i2.pp598-607.

W. F. Mbasso, A. Harrison, I. Dagal, P. Jangir, M. Khishe, H. Kotb, M. S. Shaikh, A. Smerat, E. F. Donfack, and R. Kumar. 2025. "Digital Twins in Renewable Energy Systems: A Comprehensive Review of Concepts, Applications, and Future Directions." *International Journal of Electrical Power & Energy Systems* 61(101814). doi:10.1016/j.esr.2025.101814.

Wang, Y, X Liu, J Zhao, and M Shahidehpour. 2024. "Digital Twin-Driven Resilient Operation of Active Distribution Networks With High Renewable

Penetration." *IEEE Transactions on Smart Grid* 15(2): 1648–1661,. doi:10.1109/TSG.2023.3334869.

Zhang, Z, J Wang, and B Chen. 2021. "Deep Reinforcement Learning for Energy Management in Microgrids." *Applied Energy* 288: 116618. doi:10.1016/j.apenergy.2021.116618.

Zhou, K, S Yang, and Z Shao. 2021. "Digital Twin-Based Optimization for Energy Systems." *Applied Energy* 285: 116419. doi:10.1016/j.apenergy.2020.116419.