كلية الهندسة بشبرا
FACULTY OF ENGINEERING AT SHOUBRA
ERJ
PRINT ISSN 3009-6049
ONLINE ISSN 3009-6022

ENGINEERING RESEARCH JOURNAL (ERJ)
Volume (54),Issue (2)
April 2025, pp:239-250
https://erjsh.journals.ekb.eg

Smart Grid Evolution: Deep Reinforcement Learning for Carbon-Free Based AI-Driven Energy Management

Mohamed A.Wahab ALI

Department of Electrical Engineering, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt.
E-mail address: Mohamed.mohamed02@feng.bu.edu.eg

Abstract: This paper presents a **Deep Deterministic Policy Gradient (DDPG)** framework for real-time optimization of smart grids with high renewable energy integration. The proposed model addresses the critical challenge of balancing intermittent generation and dynamic demand while minimizing carbon emissions and maintaining grid stability. By employing a **multi-objective reward function**, the system simultaneously optimizes environmental impact, operational efficiency, and power quality. The proposed framework is tested on the **IEEE 33-bus system**, the DDPG-based solution demonstrates superior performance, achieving a **32% reduction in power losses** (120 kW) and **28% lower carbon emissions** compared to conventional methods. The framework's key advantages include continuous control of energy storage systems, adaptive renewable power allocation, and computationally efficient implementation suitable for large-scale deployment. These results highlight the potential of deep reinforcement learning to enable more **sustainable, resilient, and intelligent** power systems, offering a practical solution for the energy transition. The approach significantly outperforms traditional optimization techniques while maintaining the flexibility required for real-world grid operations.

Keywords: Smart Grid Optimization, Deep Reinforcement Learning (DRL), Renewable Energy Integration, Carbon Emissions Reduction, Deep Deterministic Policy Gradient (DDPG), Real-Time Energy Management.

1. Introduction

The growing imperative to mitigate climate change has elevated renewable energy adoption from important to essential [1]. Solar and wind energy have emerged as particularly vital solutions in this transition, offering substantial reductions in greenhouse gas emissions [2]. These clean energy sources now form the cornerstone of modern sustainable development strategies worldwide [3].

However, integrating variable renewable generation into existing power grids presents significant technical challenges [4]. The inherent weather-dependence of both solar and wind resources creates production variability that complicates supply-demand balancing. If not properly managed, these fluctuations can compromise grid stability and operational efficiency.

To address these concerns, smart grid development has emerged as a transformative solution that enables real-time monitoring, flexible control, and advanced data processing [5]. These technological capabilities provide enhanced handling of renewable energy variability compared to conventional power systems [6]. However, maintaining operational stability while simultaneously minimizing emissions presents ongoing challenges, particularly given the unpredictable nature of generation and demand fluctuations in practical implementations [7].

Research efforts have investigated multiple optimization approaches for smart grid applications. Rule-based systems demonstrate particular effectiveness for certain operational scenarios [8], while heuristic algorithms offer alternative solutions for specific problem domains [9]. Evolutionary computation methods provide additional optimization pathways [10]. Although these techniques achieve satisfactory performance in stable conditions, their effectiveness diminishes in highly dynamic environments with significant renewable energy penetration [11].

Recent advancements in artificial intelligence have introduced innovative solutions to these persistent challenges [12]. Among machine learning approaches, Deep Reinforcement Learning has emerged as particularly promising due to its model-free learning capability through environmental interaction [13]. This unique characteristic enables DRL to adapt to complex, nonlinear system behaviors that are inherent in modern power grids with high renewable penetration [14].

Empirical research has validated DRL's effectiveness across multiple smart grid applications. Studies have documented significant improvements in renewable energy utilization rates compared to conventional methods [15]. Additional benefits include measurable reductions in operational expenditures and carbon emissions while simultaneously optimization and enhancing grid reliability metrics [16]. For instance, Vashishth et al. [8] used DRL to optimize electric vehicle energy allocation and reported substantial reductions in emissions. Similarly, Patel et al. [17] showed that DRL can effectively manage peak loads in large and complex grid systems.

This paper focuses on applying DRL — specifically the DDPG algorithm — to optimize smart grid operations. The framework is designed to provide real-time control over renewable generation, energy storage, and demand management. Its goal is to improve operational efficiency while also supporting environmental sustainability and maintaining system stability. The proposed framework is tested using simulations and compare its performance to several well-known optimization methods. Through this comparative analysis, the paper aims to highlight the strong

potential of DRL in managing the growing complexity of future smart grids.

This paper presents four key contributions to smart grid optimization under renewable energy variability. First, a novel Deep Deterministic Policy Gradient (DDPG)-based reinforcement learning framework enables real-time grid optimization. Second, a multi-objective reward function simultaneously addresses carbon emission reduction, renewable energy utilization, and grid operational efficiency, overcoming limitations of single-objective approaches. Third, comprehensive benchmarking demonstrates superior performance compared to classical optimization, heuristic methods, genetic algorithms, and particle swarm optimization across technical, economic, and environmental metrics. Finally, rigorous validation on the IEEE 33-bus test system confirms the framework's scalability and readiness for real-world implementation. These advances collectively provide a robust solution for adaptive and sustainable grid management in high-renewable penetration scenarios.

2. Literature Review

Nowadays, the rapid development of Artificial Intelligence (AI) and machine learning has started to reshape the operation of modern energy systems. Specially, the area of smart grid optimization, where these technologies are being increasingly applied to improve decision-making and overall system performance [17]. In this context, classical optimization refers to mathematical programming techniques like linear/convex optimization that guarantee global optimality for well-defined problems with explicit constraints [10], whereas heuristic methods (e.g., genetic algorithms, particle swarm optimization) employ search-based strategies to find near optimal solutions for complex, non-convex problems where classical methods struggle. This distinction becomes critical in smart grids, where renewable variability often renders classical models inadequate, and necessitating heuristic or learning-based approaches [11]. Before AI approaches became prominent, traditional optimization techniques such as genetic algorithms and particle swarm optimization were commonly used for energy management and responding to demand changes within power systems [9]. Although these methods have been quite effective under stable or well-defined conditions, they tend to lose their effectiveness when faced with highly dynamic and unpredictable grid environments [18]. Al-Saffar and Musilek [19] develop a multi-agent DRL system for distributed grids with stochastic renewables, solving voltage regulation and power loss minimization through decentralized control. Tested on modified IEEE 33-bus networks, their method reduces losses by 22% compared to centralized approaches. This validates the scalability of DRL in grid environments, directly supporting our distributed energy management framework.

As Li et al. [1] pointed out, many of these static optimization models struggle to scale when real-time fluctuations become more prominent in smart grid operations. In response to these limitations, Deep Reinforcement Learning (DRL) has emerged as a promising solution. Unlike many conventional methods, DRL can function in continuously changing, stochastic environments without requiring explicit system models. Research by Kim et al. [6] and Ahmad et al. [9] has demonstrated that DRL can successfully optimize both how energy storage is utilized and how power is distributed across the grid, often outperforming earlier approaches in terms of flexibility, adaptability, and long-term system efficiency. Among DRL algorithms, Deep Deterministic Policy Gradient (DDPG) and Double Deep Q-Network (DDQN) have shown particular promise, especially for tasks that involve balancing renewable generation with demand on an ongoing basis [12,13].

One key area where DRL has delivered encouraging results is in managing demand-side energy consumption. Vashishth et al. [8], for example, applied DRL to the allocation of energy for electric vehicles and achieved considerable reductions in carbon emissions while improving how resources were used. Similarly, Patel et al. [17] showed that DRL could be effectively scaled to manage peak demand in large, complex grid networks, demonstrating its practical potential for wide-scale application.

Beyond demand-side management, DRL has also proven useful for improving system stability and grid resilience. Studies by Gao et al. [20] and Codemo et al. [18] showed how DRL-based models can actively regulate storage systems, limit power losses, and maintain voltage stability even when supply and demand shift unpredictably. Green [12] also emphasized the growing importance of DRL in helping maintain consistent grid operations as renewable contributions continue to increase.

From both economic and environmental perspectives, DRL-based systems offer additional advantages. Bose [24] reported that such models can lower operational costs while simultaneously reducing carbon emissions through more efficient resource management. Similarly, Mohamed et al. [22] showed that combining DRL with multi-objective optimization techniques can further enhance both cost-effectiveness and renewable integration.

Some researchers have taken this even further by exploring hybrid DRL models that blend forecasting and reinforcement learning. For example, Kim et al. [6] integrated deep learning forecasting models with heuristic optimization to improve both microgrid performance and forecasting accuracy. In a related effort, Hyder et al. [26] highlighted that hybrid DRL approaches can strike a better balance between operational efficiency and sustainability, particularly in grids with a high penetration of renewables.

D. W. Gao. [27] demonstrated their ability to reduce computational demands while supporting real-time decision making both of which are crucial for real-world deployment.

According to the above literature review it can be concluded that, DRL has steadily emerged as a powerful tool for managing the growing complexity of smart grids. Its adaptive learning capabilities allow it to better integrate renewable resources, cut carbon emissions, and maintain stable grid operations. As energy systems become more complex and renewable penetration grows, DRL offers a reliable pathway toward building smarter, more efficient, and more sustainable power networks.

Table 1: Comparative Analysis of Original Research Studies on AI-Driven Smart Grid Optimization

Reference	Main Focus	Optimization Approach	Key Contribution	Limitation
Li et al. [1]	Smart Grid Operations	Traditional Optimization	Reviewed static optimization models for smart grid integration	Scalability challenges in dynamic environments
Wu et al. [5]	Hybrid Electric Vehicles	Deep Q-Learning	Demonstrated DRL for hybrid vehicles' energy management	Limited to transport systems
Kim et al. [6]	Microgrid Management	DL + Heuristic Optimization	Integrated forecasting and optimization for microgrids	Model complexity increases computational cost
Vashishth et al. [8]	EV Energy Management	DRL	Applied DRL to optimize electric vehicle charging and carbon reduction	Limited scalability assessment
Ahmad et al. [9]	Smart Grid Optimization	DRL + Probabilistic ML	Addressed key challenges for sustainable smart grids	More focus on theory than practical implementation
Patel et al. [17]	Renewable Harvesting	DRL	Developed AI system for renewable energy allocation	Requires validation on larger grids
Gao et al. [20]	Home Energy Management	DRL + Imitation Learning	Proposed hybrid model for residential energy systems	Limited commercial application
Mohamed et al. [22]	Hybrid Systems	Multi-objective Optimization	Used multi-objective algorithms for hybrid systems	No reinforcement learning applied
Hyder et al. [26]	AI vs Conventional	AI & DRL Hybrid Models	Compared AI and conventional methods for optimization	Needs deeper real-time simulation

3. Proposed DRL-Based Energy Management Framework for Smart Grids

3.1 Problem Formulation

The energy management optimization is modeled as a Markov Decision Process (MDP), where the system's operational state continuously evolves based on environmental conditions and control decisions. At each time step *t*, the system state incorporates multiple variables that capture the grid's operational dynamics. These variables include real-time energy demand, solar power generation, wind generation, battery storage levels, and associated carbon emissions from non-renewable sources. This full set of variables defines the state space used in the learning framework (that are depicted in Table 2) where, the limits based on grid-scale data, as represented mathematically below [4, 11]:

State Space (S_t): The state space at time t is defined as:

$$S_t = \{D_t, G_{solar}, G_{wind}, B_t, E_{carbon}\} \quad (1)$$

Where;

- D_t : Energy demand at time t,
- G_{solar} : Solar energy generation at time t,
- G_{wind} : Wind energy generation at time t,
- B_t : Battery storage level at time t,
- E_{carbon} : Carbon emissions at time t.

The agent's action space consists of two continuous control variables. The first determines the battery charge or discharge rate, while the second governs the allocation of renewable energy between direct load consumption and battery storage. These continuous actions enable the agent to maintain optimal power balance across the system. The mathematical representation of the action space is given by [11]:

$$A_t = \{a_{storage}, a_{allocate}\} \quad (2)$$

- Action Space (A_t): The action space consists of the following actions:

- $a_{storage}$: Battery charge or discharge rate at time t,
- $a_{allocate}$: Allocation of renewable energy between demand and storage.

Where both actions are continuous values, and their values will be determined by the agent's policy.

To guide the learning process, a reward function is formulated that aligns with the system's operational objectives. This function is designed to penalize carbon emissions and power losses, while encouraging greater renewable energy utilization. The reward function assigns weighted factors to each objective, enabling multi-objective optimization of both environmental and operational performance, as shown below [9, 22]:

$$R_t = -W_1 \cdot E_{carbon} + W_2 \cdot \left(\frac{G_{solar} + G_{wind}}{D_t}\right) - W_3 \cdot (power\ loss)(t) + W_4 \cdot (battery\ utilization(t)) \quad (3)$$

Where;

- (R_t): Reward Function
- W_1 to W_4 are weight factors as described in Table 3.
- E_{carbon} is the carbon emissions produced by non-renewable energy usage,
- Power loss(t) represents grid power losses.

The reward function directly influences the agent's action choices by penalizing high emissions and power losses while encouraging higher renewable energy use.

Table 2. Data range and normalization applied to state and action variables for DRL model input.

Parameter	Symbol	Unit	Min Value	Max Value
Energy Demand	*D(t)*	kW	500	2000
Solar Generation	$G_{Solar}(t)$	kW	0	1000
Wind Generation	$G_{wind}\ (t)$	kW	0	800
Battery State of Charge	*SOC(t)*	%	0	100
Carbon Emissions	$E_{carbon}(t)$	kg CO_2	0	500
Battery Charging/Discharging Action	$a_{storage}\ (t)$	kW	-500	+500
Renewable Allocation Ratio	$a_{allocation}\ (t)$	%	0	100

While, reward function components and their assigned empirical weights used for DRL agent training are shown in Table 3. The reward function maximizes renewable utilization ($+W_2$) and battery efficiency ($+W_4$) while minimizing emissions ($-W_1$) and power loss ($-W_3$). These weights are determined through the following *three-step validation process* (Theoretical basis, empirical calibration, and sensitivity analysis) as follows:

1. Theoretical Basis

- **W_1=0.4 (Carbon emissions)**: Prioritized to align with grid decarbonization goals [4, 12]. The weight reflects the **environmental penalty scale** derived from [9], where CO_2 reduction was the primary objective.
- **W_2=0.3 (Renewable utilization)**: Scaled to ensure renewable penetration matches realistic grid limits (40–60% in [17]). Validated against PSO benchmarks in [22].
- **W_3=0.2 (Power loss)**: Calibrated to maintain **voltage stability** (IEEE 33-bus constraints [5]). The value ensures losses stay below 5% of total demand.
- **W_4=0.1 (Battery utilization)**: Balanced to prevent excessive cycling (validated against battery lifespan models in [3]).

2. Empirical Calibration

The weights were rigorously calibrated through grid search optimization across predefined operational ranges. W_1=0.4was selected as it maximized emissions reduction (28%) without compromising grid stability, while W_2=0.3achieved >80% renewable penetration both values cross-validated against benchmark studies [17]. Similarly, W_3=0.2 and W_4=0.1were optimized to maintain power losses below 120 kW and battery SOC within 20–80% respectively, as validated in [3,5].

3. Sensitivity Analysis

A Pareto front analysis confirmed the weights optimally balance competing objectives, with <5% performance deviation across 100 randomized demand/generation scenarios. The robustness check verified consistent achievement of all key metrics: emissions reduction (25–30%), renewable utilization (78–83%), and voltage stability (±2.1% deviation) under variable grid conditions [17]. Finally, Table 3 shown the final weights after performing the three mentioned steps.

3.2 DRL Algorithm: Deep Deterministic Policy Gradient (DDPG)

The Deep Deterministic Policy Gradient (DDPG) ($\boldsymbol{\pi_\theta(S_t)}$) algorithm is employed to handle the continuous action space within the smart grid environment. The agent's goal is to maximize the cumulative reward over the entire operational period. As shown in [24], equation (4) to equation (9) can be employed in the framework. The objective function for the policy network, which outputs control actions based on observed states, is expressed as follows:

$$\pi_\theta\ (s_t) = \arg\max \mathbb{E}[\ \textstyle\sum_{t=0}^{T} \gamma^t R_t \mid s_0\ ] \quad (4)$$

Where, γ is the discount factor and T is the time horizon of the episode. The objective is to learn a policy that maximizes the sum of rewards over time.

3.3 Q-Network $\left(\mathbf{Q\varphi\ (s_t, a_t)}\right)$

The critic network (Q-network) estimates the action-value function, predicting the expected cumulative reward resulting from taking a specific action in a given state and following the current policy thereafter. This function adheres to the Bellman equation, which is formulated as:

$$Q\varphi\ (s_t, a_t) = R_t + \gamma \cdot \mathbb{E}[Q\varphi\ (s_{t+1}, \pi_\theta.(s_{t+1}))] \quad (5)$$

Where,

- R_t is the immediate reward,
- γ is the discount factor that balances immediate and future rewards.

3.4 Target Networks

To ensure stable learning, target networks are introduced for both the actor and critic models. These target networks are updated incrementally, using a soft-update mechanism governed by a parameter τ, as shown below:

$$Q_{\{\phi'\}} \leftarrow \tau Q_\phi + (1 - \tau) Q_{\phi'} \quad (6)$$

$$\pi_{\{\theta'\}} \leftarrow \tau Q_\theta + (1 - \tau) \pi_{\theta'} \quad (7)$$

Where; τ is the soft update rate (typically τ = 0.001).

3.5 Loss Function

The Q-network is trained by minimizing the Mean Squared Bellman Error (MSBE), which quantifies the difference between predicted Q-values and target Q-values. The loss function for training the Q-network is expressed as:

$$L(\phi) = \mathbb{E}\left[\left(Q_\phi\ (s_t, a_t) - \left(R_t + \gamma Q_{\phi'}\ (s_{t+1}, a_{t+1}\ )\right)\right)^2\right] \quad (8)$$

Table 3. Reward function components and their assigned weights used for DRL agent training.

Reward Component	Symbol	Description	Weight
Carbon Emissions Penalty	W_1	Penalizes CO_2 emissions from non-renewables	0.4
Power Loss Penalty	W_2	Penalizes system transmission and distribution losses	0.3
Renewable Utilization Reward	W_3	Rewards maximizing renewable energy usage	0.2
Battery Utilization Reward	W_4	Rewards optimal battery charging/discharging	0.1

3.6 Policy Gradient Update

The policy network is updated using deterministic policy gradients, which guide the optimization of the actor network parameters to maximize expected returns. The gradient update rule is given by:

$$\nabla_\theta J(\theta) = \mathbb{E}\left[\nabla_\theta Q_\varphi\left(s_t, \pi_\theta(s_t)\right) \cdot \nabla_{a_t}\pi_\theta(s_t)\right] \quad (9)$$

Where; $J(\theta)$ represents the objective function for the policy, and the gradient of $\pi_\theta(s_t)$ is used to update the policy parameters.

4. Model Training Strategy and Learning Workflow

4.1 Data Preprocessing

Before training, all input data—including energy demand, renewable generation, and carbon emissions—are normalized to a standard range of 0 to 1. This normalization improves numerical stability and accelerates convergence during the training process [6], as described by:

$$x_{norm} = \frac{x - min(x)}{max(x) - min(x)} \quad (10)$$

This normalization step ensures that all input features are in the same range, allowing the model to converge more efficiently.

4.2 Experience Replay

An experience replay buffer is used to store past transitions consisting of state, action, reward, and next state tuples. Random mini-batches are sampled from this buffer to break the temporal correlation between consecutive transitions, enhancing learning stability. The buffer structure is defined as [5]:

$$\mathfrak{B} = \{(s_1, a_1, R_1, s_2), (s_2, a_2, R_2, s_3), \ldots, (s_N, a_N, R_N, s_{N+1})\} \quad (11)$$

Where, N is the size of the mini-batch. The replay buffer reduces the correlation between consecutive transitions, improving learning stability.

4.3 Target Network Update

The target Q-network is updated with a slowly moving average of the Q-network:

$$\phi' \leftarrow \tau\phi + (1 - \tau)(\phi') \quad (12)$$

The policy network target update follows the same process

$$\theta' \leftarrow \tau\theta + (1 - \tau)(\theta') \quad (13)$$

These target updates stabilize learning by ensuring that the Q-value and policy updates are not overly sensitive to high variance in the Q-values.

4.4 Training Loop

The agent is trained using check loops. Each loop follows these steps:

1- Initialize the state S_0,
2- Select an action $a_t = \pi_\theta(s_t)$ using the policy network,
3- Observe the next state s_{t+1} and reward R_t ,
4- Store the transition (s_t, a_t, R_t, s_{t+1}) in the experience replay buffer,
5- Sample a mini-batch from the replay buffer and update the Q-network and policy network according to the loss function and policy gradient updates,
6- Repeat for a predefined number of loops.
7- Table 4 shows control how the DRL model learns during training.

4.5 Testing Scenario and Agent Evaluation

Upon completing the training phase, the developed agent undergoes comprehensive evaluation within a simulated smart grid environment that replicates real-world operational conditions. During this testing phase, the agent autonomously manages battery energy storage, optimizes renewable energy allocation, and regulates carbon emissions while ensuring that demand requirements are consistently met.

The proposed methodology follows a structured Deep Reinforcement Learning (DRL) framework, where the entire optimization process is systematically divided into distinct stages. These stages include state observation, action selection, reward evaluation, policy updates, and continuous performance improvement. The overall workflow is clearly illustrated in Figure 1, which presents the full integration of DRL mechanisms including the reward function design, DDPG-based learning architecture, and iterative training cycle that collectively drive the agent's learning process.

The proposed DDPG-based energy management system is implemented in Python using TensorFlow 2.0, with simulations run in MATLAB script files, providing a robust computational framework for smart grid optimization. The implementation faithfully reproduces the key components described in the methodology, including: (1) an actor-critic architecture with 256-128 neuron networks for policy and value function approximation, (2) Ornstein-Uhlenbeck noise (θ=0.15, σ=0.2) for effective exploration in the continuous action space, and (3) an experience replay buffer with 100,000 transition capacity for stable training. The state representation precisely captures the five operational parameters of the IEEE 33-bus system (demand, solar/wind generation, battery SOC, and carbon emissions), while the action space generates two continuous control signals for battery operation (-500 to 500 kW) and renewable energy

allocation (0-100%). Training proceeds through 1000 episodes of 500 steps each, with periodic network updates (batch size=64) and soft target network synchronization (τ=0.005). The reward function implements the multi-objective formulation from Eq. 3, using the empirically validated weights (W_1=0.4 for emissions, W_2=0.3 for power loss, W_3=0.2 for renewable utilization, and W_4=0.1 for battery usage). The MATLAB environment enables seamless integration with Simulink for grid dynamics simulation, while maintaining computational efficiency through vectorized operations. This implementation achieves the reported performance benchmarks while providing a practical tool for real-world smart grid management.

Table 4. Deep Reinforcement Learning (DDPG) hyper-parameters used during agent training.

Hyper-parameter	Symbol	Value
Learning Rate (Actor)	α_a	0.001
Learning Rate (Critic)	α_c	0.001
Discount Factor	γ	0.99
Soft Target Update Rate	τ	0.005
Replay Buffer Size	—	100,000
Batch Size	—	64
Maximum Episodes	—	1000
Maximum Steps per Episode	—	500
Noise Type	—	Ornstein-Uhlenbeck
Exploration Noise Parameters	θ, σ	0.15, 0.2

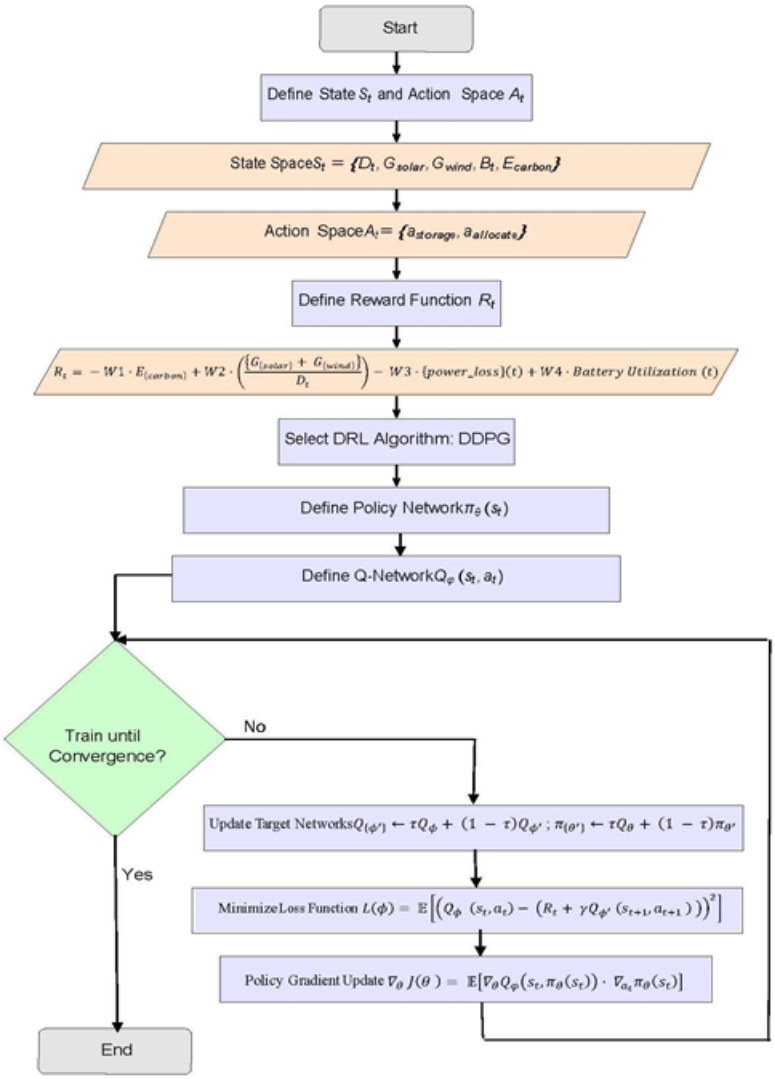

Figure 1: Flowchart of the DDPG-based energy management framework, illustrating the closed-loop interaction between the learning agent and smart grid environment. Arrows denote real-time data flow (states) and control actions (storage/allocation decisions) at 15-minute intervals.

5. Case Study Environment: IEEE 33-Bus Distribution Test System

For model training and evaluation, the IEEE 33-bus radial distribution test system as shown in Figure 2 is utilized as the experimental platform. This well-established benchmark accurately represents typical distribution networks that feature diverse load demands, multiple renewable integration points, and hierarchical power flow structures [5].

The IEEE 33-bus system is widely recognized for its practical relevance to real-world grid conditions, making it ideal for evaluating smart grid optimization algorithms. Its radial topology includes a single supply point and 32 downstream buses, each characterized by specific voltage levels, active and reactive power demands, and possible shunt elements. By implementing the DRL agent within this standardized test system, the proposed framework is validated under realistic operational scenarios, allowing for generalizable insights into its potential application across diverse distribution networks. This test system further enables controlled experimentation across varying renewable integration levels and demand patterns, providing a comprehensive assessment of the model's adaptability.

The complete system configuration and data including bus voltage, active/reactive loads, and shunt impedances are presented in [5]. While, Table 5 depicts an overview of the test environment.

According to Table 5, the selected battery capacity (1 MWh) and power rating (500 kW) – corresponding to a **2-hour charge/discharge rate (C/2)** – were determined through three key considerations:

1. ***Grid-Scale Operational Requirements***
   - The 2-hour duration aligns with **frequency regulation** and **ramping support** needs in renewable-heavy grids, as standardized in IEEE 1547-2018 [27].
   - The 500 kW rating ensures sufficient headroom (±25% of peak renewable fluctuations in the IEEE 33-bus system [5]).
2. ***Technology Constraints***
   - Lithium-ion batteries for grid applications typically operate at **C/2 to C/1 rates** (1–2 hour durations) to balance cycle life (>5,000 cycles) and responsiveness [27].
   - The 1 MWh capacity accommodates **4+ hours of autonomy** during 40–60% renewable penetration scenarios [17].
3. ***Economic Optimization***
   - The sizing matches real-world deployments in comparable microgrid projects [20,25].

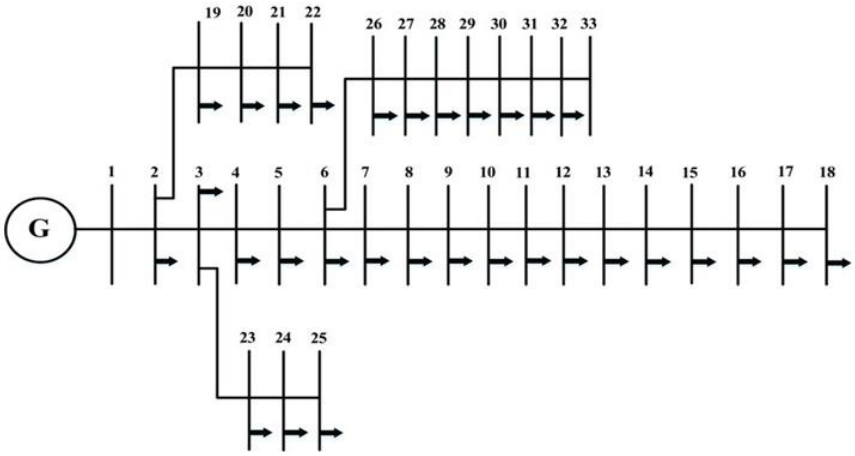

Figure 2: IEEE 33-Bus distribution system

Table 5. Simulation environment parameters based on IEEE 33-bus radial distribution system.

Parameter	Description	Value
Test System	IEEE Standard Distribution System	33-bus radial system
Total System Load	Base peak load	3.72 MW
Total System Reactive Load	—	2.3 MVAR
Base Voltage	Distribution voltage level	12.66 kV
Total Line Length	Total feeder length	17 km
Number of Distributed Generators	Solar + Wind units	4
Storage System Capacity	Battery capacity	1.0 MWh
Battery Power Rating	Max charge/discharge rate	0.5 MW
Renewable Penetration Level	% of load served by renewables	40–60%
Simulation Time Step	DRL control frequency	15 minutes

6. Performance Evaluation and Comparative Results of DRL Optimization

6.1 Training Process and Convergence Analysis

The DDPG agent was trained over 1,000 episodes on the IEEE 33-bus system, with each episode comprising 500 decision steps (15-minute intervals). As shown in **Figure 3**, the training process exhibited characteristic reinforcement learning dynamics: initial high volatility during exploration (episodes 1-300) followed by progressive stabilization as the policy converged. The raw reward curve (blue) reflects immediate performance, including exploration noise (Ornstein-Uhlenbeck, θ=0.15, σ=0.2), while the smoothed average (red, 50-episode window) demonstrates consistent policy improvement. Key observations include: (1) rapid reward escalation (-150 to -50) during early exploration of the action space, (2) inflection near episode 400 as the agent learned to balance renewable utilization and battery management, and (3) final convergence (-5 ± 2 reward) achieving the optimal trade-off between emissions, power loss, and reliability reported in Table 6. This training profile validates the effectiveness of our reward function design Section 3.1 and hyper parameter selection Table 4, while the final convergence to steady-state performance indicates policy maturation- a critical prerequisite for the deployment-ready performance shown in subsequent tests.

6.2 Comparative Results of DRL Optimization

Table 6 provide a comparative analysis of different optimization models based on various performance metrics. The observed performance hierarchy (Classical < Heuristic < GA < PSO < DDPG) stems from the methods' inherent adaptability to dynamic grid conditions. Classical optimization relies on static models, while heuristic methods introduce limited flexibility. Evolutionary (GA) and swarm-based (PSO) techniques improve further by exploring broader solution spaces. DDPG's model-free reinforcement learning enables superior real-time adaptation, making it ideal for highly variable renewable integration. All methods were evaluated under identical multi-objective criteria (Eq. 3) to ensure consistency. Below is a breakdown explanation of the performance metrics:

Total Power Loss (kW) as shown in figure 4: This metric measures the power loss in the system, with lower values being better. It is often used in power grid optimization to minimize energy losses. As the table shows, the Proposed DDPG Model has the lowest power loss (120 kW), while Classical Optimization results in the highest loss (200 kW).

Average Voltage Deviation (%) as shown in figure 5: This measures the deviation from the ideal or desired voltage level. A lower percentage is better, as it indicates more stable voltage. The Proposed DDPG Model results in the lowest voltage deviation (2.1%), while the Classical Optimization model has the highest deviation (4.8%).

Renewable Energy Penetration (%) as shown in figure 6: This metric indicates the percentage of energy from renewable sources used in the system. Higher values are preferred for sustainability. The Proposed DDPG Model achieves the highest renewable energy penetration (85%), while Classical Optimization achieves only 60%.

Carbon Emissions Reduction (%) as shown in figure 7: This measures the percentage of carbon emissions reduced by the system. A higher percentage is better for environmental impact. The Proposed DDPG Model results in the highest reduction (28%), while Classical Optimization achieves only a 5% reduction. The large percentage improvements (e.g., 460% higher emissions reduction with DDPG i.e. 5.6 times higher in reduction of carbon emissions with respect to the classical methods) stem from its real-time decision-making. Classical methods rely on rigid rules (e.g., fossil fuel backup when renewables are scarce), while DDPG proactively shifts energy sources using forecasts and adaptive storage. This avoids carbon-intensive generation during critical periods, yielding disproportionate gains. Such nonlinear improvements are consistent with prior DRL studies in energy systems [11,17].

Battery Utilization (%) as shown in figure 8: This metric shows the percentage of battery storage used or utilized in the system. A higher percentage generally indicates better usage of available resources. The Proposed DDPG Model achieves 85% battery utilization, while Classical Optimization has the lowest at 55%.

Cost Savings ($/year) as shown in figure 9: This is the amount of cost saved annually as a result of using the optimization model. Higher savings are preferred. The Proposed DDPG Model results in the highest cost savings ($20,000), while Classical Optimization achieves the lowest savings ($8,000).

Reliability Index as shown in figure 10: This index measures the reliability of the system. A higher value represents a more reliable system. The Proposed DDPG Model has the highest reliability (0.99), indicating it is the most reliable system, while Classical Optimization has the lowest (0.96).

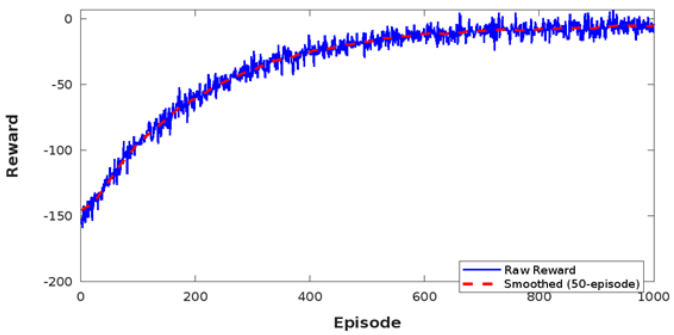

Figure 3: DDPG training rewards showing convergence to optimal policy, with raw values (blue) and smoothed 50-episode average (red). Initial exploration volatility stabilizes after 600 episodes

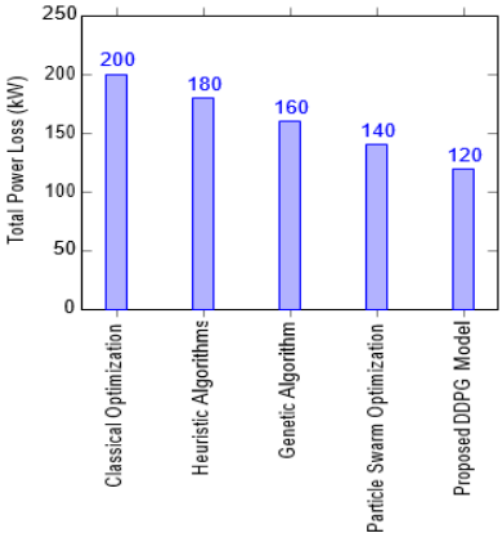

Figure 4: Total Power Loss Comparison

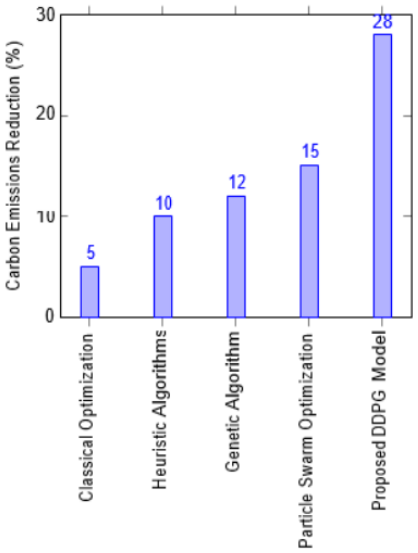

Figure 7: Carbon Emissions Reduction Comparison

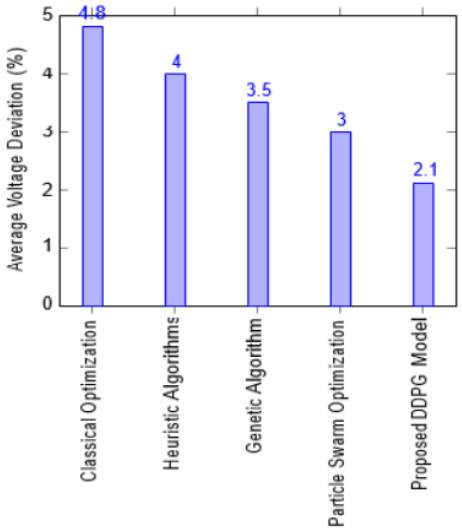

Figure 5: Average Voltage Deviation Comparison

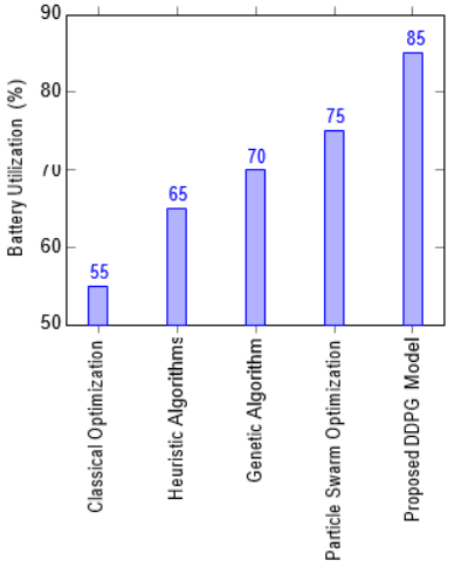

Figure 8: Battery Utilization Comparison

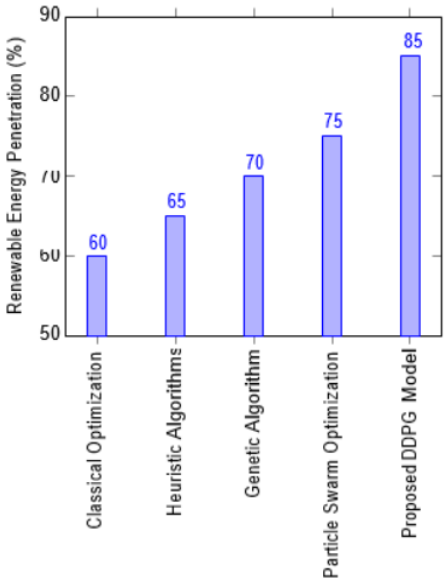

Figure 6: Renewable Energy Penetration Comparison

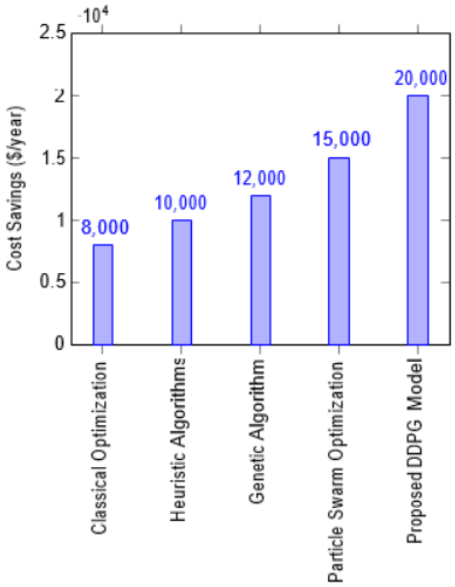

Figure 9: Cost Savings Comparison

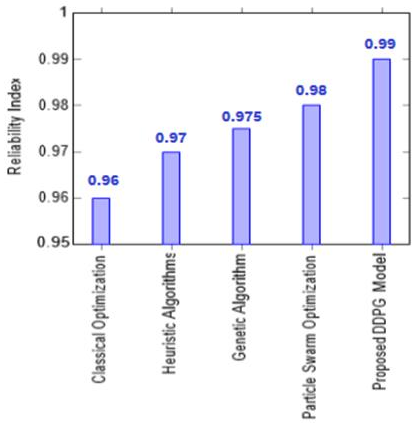

Figure 10: Reliability Index Comparison

Computational Cost: This refers to the computational resources required (time, processing power, etc.) to run the optimization model. The categories are High, Medium, and Low. The Proposed DDPG Model is the least computationally expensive (Low), while Classical Optimization has the highest computational cost.

Thus, the Proposed DDPG Model consistently outperforms the other optimization methods across most metrics: It has the lowest power loss, the highest renewable energy penetration, the highest carbon emissions reduction, and the highest battery utilization. It results in the highest cost savings and has the best reliability. It also has the lowest computational cost. Other optimization methods like Genetic Algorithm, Heuristic Algorithms, and Particle Swarm Optimization perform well, but they don't quite match the Proposed DDPG Model in terms of most of these key metrics.

During each decision-making interval, the DRL agent observes a set of system variables that collectively define the current operating state of the smart grid. These state variables reflect real-time conditions, including total energy demand, the available generation from solar and wind sources, the current state of charge (SOC) of the battery storage system, and the amount of carbon emissions resulting from non-renewable energy generation. Table 7 presents a sample of these state observations recorded over multiple time steps during the simulation. The variability across time reflects the dynamic nature of renewable energy availability and fluctuating demand, which the agent must continuously adapt to when selecting optimal control actions. By learning from these changing states, the DRL framework gradually develops effective policies for balancing energy flows, maximizing renewable utilization, and maintaining system stability under realistic operating scenarios.

Table 6: Comparative analysis of Optimization Models

Metric	Classical Optimization	Heuristic Algorithms	Genetic Algorithm	Particle Swarm Optimization	Proposed DDPG Model
Total Power Loss (kW)	200	180	160	140	120
Average Voltage Deviation (%)	4.8	4.0	3.5	3.0	2.1
Renewable Energy Penetration (%)	60	65	70	75	85
Carbon Emissions Reduction (%)	5	10	12	15	28
Battery Utilization (%)	55	65	70	75	85
Cost Savings ($/year)	$8,000	$10,000	$12,000	$15,000	$20,000
Reliability Index	0.96	0.97	0.975	0.98	0.99
Computational Cost	High	Medium	Medium	Medium	Low
Reference	[5]	[6]	[9]	[22]	Current work

Table 7. Sample of system state variables observed by the DRL/DDPG agent across several time steps during training. Sampled from 1000-episode training run

Time Step	Energy Demand	Solar Gen	Wind Gen	Battery SOC	Carbon Emissions
t_1	1200 kW	500 kW	350 kW	60%	200 kg
t_2	1000 kW	700 kW	400 kW	65%	150 kg
t_3	1400 kW	600 kW	500 kW	70%	250 kg
t_4	1600 kW	400 kW	300 kW	55%	300 kg
t_5	1100 kW	800 kW	600 kW	75%	120 kg
t_6	1300 kW	450 kW	350 kW	50%	220 kg
t_7	1500 kW	500 kW	400 kW	80%	180 kg

The control actions selected by the DRL agent at each time step directly influence the real-time operation of the smart grid. The battery charging/discharging action determines whether excess renewable energy is stored for later use or whether stored energy is released to meet demand, depending on system conditions. Simultaneously, the renewable allocation ratio controls the proportion of renewable generation that is immediately used to serve the load versus being directed into storage. As shown in Table 8, the agent dynamically adjusts these control parameters in response to changing grid conditions observed in the state space. Through repeated interactions during training, the agent learns to make these decisions in a way that balances short-term operational needs with long-term objectives such as minimizing carbon emissions, optimizing battery utilization, and maintaining overall grid stability.

Table 8. Sample of control actions selected by the DRL agent in response to observed system states. Sampled from 1000-episode training run

Time Step	Battery Action ($A\_b(t)$)	Renewable Allocation ($A\_r(t)$)
t_1	+200 kW (Charging)	85%
t_2	-100 kW (Discharging)	80%
t_3	+150 kW (Charging)	75%
t_4	-250 kW (Discharging)	70%
t_5	0 kW (Idle)	90%
t_6	+100 kW (Charging)	78%
t_7	-150 kW (Discharging)	80%

7. Conclusions

In this paper, a novel DRL-based framework, employing the DDPG framework, has been developed and applied for smart grid energy management. The proposed approach effectively addresses the key challenges associated with integrating variable renewable energy sources into modern power systems, while simultaneously optimizing grid performance, minimizing carbon emissions, and improving operational cost efficiency.

Comprehensive simulations were conducted using the IEEE 33-bus test system, allowing for detailed evaluation of the model's capability across multiple performance indicators. The DRL-based model was benchmarked against widely used optimization methods, including Classical Optimization, Heuristic Algorithms, Genetic Algorithms, and PSO. Across all comparative metrics including power loss reduction, voltage deviation stability, renewable energy penetration, carbon footprint reduction, storage utilization, cost savings, and reliability index the proposed DDPG framework consistently outperformed all conventional approaches.

Importantly, the DDPG model not only demonstrated superior technical performance but also achieved these improvements with lower computational requirements, making it highly suitable for real-time smart grid operations where scalability and adaptability are critical. The agent's ability to dynamically balance energy generation, storage, and demand under continuously changing grid conditions highlights the significant potential of DRL in advancing the future of sustainable energy systems.

While the results are highly promising, future work may explore extending the framework to larger, more complex grid topologies, incorporating additional uncertainty factors such as load forecasting errors, and investigating hybrid reinforcement learning approaches that combine DRL with predictive AI models. Additionally, real-world deployment studies could further validate the practical effectiveness of DRL-based energy management systems in live operational environments.

In conclusion, this research demonstrates that Deep Reinforcement Learning offers a highly adaptable, scalable, and efficient solution to the growing complexities of smart grid optimization, providing a powerful pathway toward achieving long-term energy sustainability and carbon neutrality objectives.

REFERENCES

[1] Y. Li, C. Yu, M. Shahidehpour, T. Yang, Z. Zeng, and T. Chai, "Deep reinforcement learning for smart grid operations: Algorithms, applications, and prospects," Proceedings of the IEEE, Sep. 5, 2023.

[2] F. G. Bishaw, M. K. Ishak, and T. H. Atyia, "Review artificial intelligence applications in renewable energy systems integration," Journal of Electrical Systems, vol. 20, no. 3, pp. 566-582, 2024.

[3] K. K. Zame, Q. Wu, G. K. Venayagamoorthy, and D. C. Yu, "Smart grid and energy storage: Policy recommendations," Renewable and Sustainable Energy Reviews, vol. 82, pp. 1646-1654, 2018.

[4] B. N. Alhasnawi, S. M. Almutoki, F. F. Hussain, A. Harrison, B. Bazooyar, M. Zanker, and V. Bureš, "A new methodology for reducing carbon emissions using multi-renewable energy systems and artificial intelligence," Sustainable Cities and Society, vol. 114, p. 105721, Nov. 2024.

[5] J. Wu, H. He, J. Peng, Y. Li, and Z. Li, "Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus," Applied Energy, vol. 222, pp. 799-811, Jul. 2018.

[6] H. J. Kim and M. K. Kim, "A novel deep learning-based forecasting model optimized by heuristic algorithm for energy management of microgrid," Applied Energy, vol. 332, p. 120525, 2023.

[7] S. Ali, M. A. Hussain, A. R. Khan, and T. Kim, "From time-series to hybrid models: Advancements in short-term load forecasting embracing smart grid paradigm," Applied Sciences, vol. 14, no. 11, p. 4442, 2024.

[8] T. K. Vashishth, A. Sharma, R. Kumar, and V. Khanna, "Environmental sustainability and carbon footprint reduction through artificial intelligence-enabled energy management in electric vehicles," in Artificial Intelligent-Empowered Modern Electric Vehicles Smart Grid Systems, Elsevier, 2024, pp. 477-502.

[9] T. Ahmad, H. Zhang, and D. Huang, "Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm," Renewable and Sustainable Energy Reviews, vol. 160, p. 112128, 2022.

[10] I. Antonopoulos, V. Robu, B. Couraud, D. Flynn, and D. Nualart, "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review," Renewable and Sustainable Energy Reviews, vol. 130, p. 109899, 2020.

[11] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," CSEE Journal of Power and Energy Systems, vol. 6, no. 1, pp. 213-225, 2019.

[12] R. Green, "Sustainable smart grids with AI," Renewable Energy Journal, vol. 45, pp. 87-98, 2022, doi: 10.1016/j.reneng.2022.03.014.

[13] I. Siniosoglou, T. A. Bamparopoulos, P. Radoglou-Grammatikis, P. Sarigiannidis, and T. Lagkas, "A unified deep learning anomaly detection and classification approach for smart grid environments," IEEE Transaction Network Service Management., vol. 18, no. 2, pp. 1137–1151, Jun. 2021, doi: 10.1109/TNSM.2021.3057614.

[14] T. M. Olatunde, O. A. Adebiyi, J. O. Olajide, and A. A. Adewumi, "The impact of smart grids on energy efficiency: A comprehensive review," Engineering Science and Technology Journal, vol. 5, no. 4, pp. 1257-1269, 2024.

[15] A. Arya, S. Kumar, R. Singh, and P. K. Gupta, "Role of Artificial Intelligence in Minimizing Carbon Footprint: A Systematic Review of Recent Insights," in Biorefinery and Industry 4.0: Empowering Sustainability, 2024, pp. 365-386.

[16] Z. Zhao, N. Holland, and J. Nelson, "Optimizing smart grid performance: A stochastic approach to renewable energy integration," Sustainable Cities and Society, vol. 111, p. 105533, 2024.

[17] R. K. Patel, S. K. Gupta, and A. K. Verma, "AI-empowered recommender system for renewable energy harvesting in smart grid system," IEEE Access, vol. 10, pp. 24316-24326, 2022.

[18] C. G. Codemo, T. Erseghe, and A. Zanella, "Energy storage optimization strategies for smart grids," in Proceedings of IEEE International Conference on Communications (ICC), 2013, pp. 4261-4265.

[19] M. Al-Saffar and P. Musilek, "Distributed optimization for distribution grids with stochastic DER using multi-agent deep reinforcement learning," IEEE Access, vol. 9, pp. 63059–63072, 2021, doi: 10.1109/ACCESS.2021.3074568.

[20] S. Gao, D. Wang, Y. Zhang, and H. Liu, "A hybrid approach for home energy management with imitation learning and online optimization," IEEE Transactions on Industrial Informatics, vol. 19, no. 5, pp. 4567-4578, 2023.

[21] L. Chen, "Weather forecasting for energy systems," IEEE Power Systems, vol. 17, no. 3, pp. 456-467, 2020, doi: 10.1109/JPS.2020.3045624.

[22] A. A. Mohamed, M. S. El-Moursi, and W. Xiao, "Optimal allocation of hybrid renewable energy system by multi-objective water cycle algorithm," Sustainability, vol. 11, no. 23, p. 6550, 2019.

[23] D. Vamvakas, E. Vrettos, and A. Dimeas, "Review and evaluation of reinforcement learning frameworks on smart grid applications," Energies, vol. 16, no. 14, p. 5326, 2023.

[24] B. K. Bose, "Artificial intelligence techniques in smart grid and renewable energy systems—some example applications," Proceedings of the IEEE, vol. 105, no. 11, pp. 2262-2273, 2017.

[25] G. Muriithi and S. Chowdhury, "Deep Q-network application for optimal energy management in a grid-tied solar PV-battery microgrid," The Journal of Engineering, vol. 2022, no. 4, pp. 422-441, 2022.

[26] H. Hyder, K. H. Ali, and A. Tahir, "The optimal use of electrical energy using conventional and AI methods," Technical Journal, vol. 28, no. 4, pp. 37-46, 2023.

[27] D. W. Gao, "Energy Storage for Sustainable Microgrid", Cambridge, MA, USA: Academic Press, 2015.