3rd International Conference on Electronics, Engineering, Computer Science and Applied Development (EESD 2026)

Article

A Reinforcement Learning Based Method for Dynamic Scheduling and Energy Efficiency Optimization in Smart Grids

Shaohan Liu[1,*]

[1] School of Energy Engineering, Xinjiang Institute of Engineering, Urumqi, China
* Correspondence: Shaohan Liu, School of Energy Engineering, Xinjiang Institute of Engineering, Urumqi, China

Abstract: With the increasing penetration of renewable energy sources, modern smart grids face significant and unprecedented challenges in dynamic scheduling and energy efficiency optimization under highly uncertain operating conditions. Traditional optimization methods often struggle with severe computational bottlenecks and slow response times when deployed in high-dimensional, dynamic environments. To overcome these critical limitations, this study comprehensively investigates the extent to which reinforcement learning (RL) and deep reinforcement learning (DRL) approaches can effectively address these operational challenges. Through a rigorous empirical evaluation of three distinct case studies—economic dispatch under renewable uncertainty, multi-agent coordinated scheduling, and real-time voltage stability control—this research provides a robust framework for intelligent grid management. A mixed-methods approach, combining qualitative analysis of RL policy behavior with quantitative performance evaluation, was systematically employed using three publicly available datasets: the IEEE 118-bus test system, the PJM historical load dataset, and the Nordic32 test system. The comprehensive results reveal that RL-based methods, particularly the Deep Deterministic Policy Gradient and Multi-Agent Deep Deterministic Policy Gradient algorithms, achieve superior performance in complex dynamic scheduling tasks. These advanced techniques demonstrated remarkable cost reductions ranging from 12.5% to 41.1% compared to conventional baseline methods. However, the evaluated models also exhibit notable limitations in generalization across diverse operating scenarios and strict constraint satisfaction under extreme grid conditions. Ultimately, this study significantly contributes to the fundamental understanding of RL algorithms' potential and limitations in smart grid applications, offering valuable insights and strategic pathways for future improvements in real-time energy management systems.

Keywords: reinforcement learning; smart grid; dynamic scheduling; energy efficiency; deep learning

Received: 08 April 2026
Revised: 26 May 2026
Accepted: 08 June 2026
Published: 14 June 2026

1. Introduction

The rapid integration of renewable energy sources such as wind and solar power has introduced significant uncertainty and variability into modern power systems. Traditional smart grid scheduling methods, including model predictive control and linear programming, often face challenges in adapting to the dynamic nature of real-time electricity demand and generation. Consequently, there is an increasing demand for intelligent optimization frameworks capable of adjusting to evolving conditions while ensuring energy efficiency and system reliability [1]. These frameworks aim to address the limitations of conventional methods by leveraging advanced computational techniques to better accommodate the complexities of renewable energy integration.

Reinforcement learning (RL) has emerged as a promising paradigm for sequential decision-making under uncertainty. Unlike conventional optimization techniques, RL enables an agent to learn optimal policies through iterative interaction with the environment, eliminating the need for precise analytical models of system dynamics. This feature makes RL particularly advantageous for smart grid applications, where accurately modeling the fluctuations of renewable energy sources and variations in load demand is inherently challenging. By focusing on adaptive learning, RL provides a flexible approach to managing the unpredictability associated with modern energy systems [2].

Recent studies have systematically explored the application of deep reinforcement learning (DRL) approaches to power system control and optimization tasks. These investigations underscore the potential of algorithms such as Deep Q Networks, Deep Deterministic Policy Gradient (DDPG), and multi-agent RL frameworks in addressing challenges related to economic dispatch, voltage adjustments, and demand response strategies. However, they also highlight critical obstacles, including inefficiencies in data utilization, adherence to operational safety requirements, and the ability to generalize across diverse operating scenarios. Addressing these challenges remains a key focus for advancing the practical implementation of DRL in energy systems [3].

Within the domain of microgrid energy management, augmented DRL methods have been developed to explicitly account for the stochastic nature of renewable energy parameters. These approaches integrate uncertainty modeling into the learning process, resulting in more robust scheduling policies that minimize operational costs and enhance the utilization of renewable energy sources. Compared to traditional rule-based or short-sighted optimization strategies, these advanced methods demonstrate superior performance in adapting to the variability inherent in renewable energy generation [4].

Furthermore, recent advancements in policy optimization have introduced algorithms such as group relative policy optimization for economic dispatch in smart grids. These methods effectively balance exploration and refinement in high-dimensional action spaces, leading to faster convergence and improved cost performance compared to traditional DRL baselines [5]. Despite these advancements, ensuring that RL-based methods consistently adhere to physical limitations during atypical scenarios remains a significant challenge. Continued research is required to enhance the reliability and robustness of these approaches under diverse operating conditions.

This study aims to empirically evaluate the performance of representative RL algorithms, including DDPG and Multi-Agent DDPG, in three smart grid scheduling tasks: economic dispatch under renewable uncertainty, multi-agent coordinated scheduling, and real-time voltage stability control [3]. Using publicly available datasets such as the IEEE 118-bus test system and the PJM historical load dataset, the study assesses both the energy efficiency improvements and the ability of these methods to adhere to operational requirements. The findings are expected to provide valuable insights for deploying RL in practical energy management systems, contributing to the development of more adaptive and efficient solutions for modern power grids.

2. Literature Review

Reinforcement learning (RL) has been widely utilized in optimizing economic dispatch and demand response within smart grids. A notable framework demonstrated the potential of RL in reducing operational costs significantly, achieving reductions ranging from 12% to 18% even under the uncertainties associated with renewable energy sources. By leveraging historical load datasets, the approach showcased enhanced economic efficiency and strategic decision-making capabilities when compared to conventional methodologies [6].

Multi-agent RL methodologies have emerged as particularly effective for tasks requiring coordinated scheduling. One such approach utilized a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm to optimize multi-device charging schedules in Industrial Internet of Things (IIoT) smart grids [7]. This method achieved a substantial 25.4% reduction in charging delays while ensuring grid stability. The effectiveness of the algorithm was validated through simulations conducted in an IIoT environment using real-world load profiles, highlighting its practical applicability.

Voltage stability represents a critical factor in dynamic scheduling processes. A study applied a Deep Deterministic Policy Gradient (DDPG) algorithm to manage electric vehicle (EV) charging, aiming to enhance voltage stability within a widely recognized test system. The results demonstrated that the RL-based agent effectively maintained voltage levels within safe operational thresholds, even under scenarios involving high levels of EV integration [8]. This approach outperformed traditional rule-based controllers by achieving a 15.3% improvement in reducing voltage deviations.

For achieving rapid real-time dispatch solutions, a neighbor feature aggregation-based multi-agent RL method was introduced to optimize distributed power dispatch. This innovative approach was tested on a complex power system model, where it achieved an impressive 99.2% optimality compared to centralized solvers. Furthermore, the method significantly reduced computation time by 83.7%, demonstrating its potential for practical deployment in time-sensitive applications [7].

Resource allocation within smart grids has also seen advancements through RL techniques. A deep RL-based system was developed to dynamically allocate computational and communication resources, optimizing their utilization. This system was validated using a testbed that incorporated real-time pricing data from a prominent energy market. The results indicated a notable 20.7% improvement in resource utilization efficiency, underscoring the effectiveness of RL in addressing resource management challenges [2].

Optimizing demand response across varying time scales remains a complex challenge. A method employing Double Deep Q Networks (DDQN) was developed to address this issue within active distribution networks. By utilizing real-world load data, the approach achieved a significant 31.2% reduction in peak load while simultaneously enhancing the integration of renewable energy by 18.9%. These results highlight the potential of RL in managing demand response effectively across multiple temporal dimensions.

Coordinated energy management across diverse urban sectors has been explored through advanced RL frameworks [2]. One such framework focused on the integration of smart grid operations with building energy systems. Simulations conducted using data from a comprehensive energy consumption survey demonstrated a remarkable 22.8% reduction in overall energy costs. This underscores the potential of RL in fostering efficient energy coordination across interconnected urban infrastructures.

Hybrid methodologies that combine RL with traditional optimization techniques have shown distinct advantages [5]. A hybrid multi-agent deep actor-critic learning algorithm, integrated with particle swarm optimization, was developed to address active voltage control challenges in grids with high renewable energy contributions. This approach successfully maintained voltage levels within a narrow ±3% range while achieving a 17.3% reduction in active power losses, demonstrating the synergy between RL and traditional optimization methods.

A comprehensive analysis systematically reviewed RL-based strategies for voltage control in smart grids. The review identified several key challenges, including inefficiencies in sample usage, difficulties in handling operational constraints, and issues related to scalability [9]. Additionally, it highlighted promising future research directions, such as the application of transfer learning and the development of safe RL methodologies, to address these challenges effectively.

Long-term scheduling for multi-energy complementary systems has been addressed using advanced RL techniques. One approach optimized the operation of hybrid systems combining water, wind, and solar energy over an extended yearly horizon. By utilizing real meteorological data from a renewable energy base, the RL policy achieved a 14.2% increase in renewable energy utilization compared to conventional scheduling methods, showcasing its potential for enhancing long-term energy management [10].

3. Theoretical Framework and Methodology

This chapter presents the theoretical framework and detailed methodology employed to assess the performance of reinforcement learning (RL) based methods for dynamic scheduling and energy efficiency optimization in smart grids. The study adopts a mixed methods approach, combining qualitative analysis of policy behavior with quantitative performance evaluation using publicly available benchmark datasets. The primary objective is to evaluate the cost reduction and energy efficiency improvements achieved by RL algorithms, while also examining their capacity to maintain system stability under conditions of uncertain renewable energy generation and fluctuating load demands. A comprehensive method flowchart is included to delineate the sequential stages and processes involved in the research, ensuring clarity in the approach [11]. Additionally, the methodology emphasizes the adaptability of RL techniques to real-world scenarios, highlighting their potential for scalable and robust applications in smart grid systems.

3.1. Theoretical Framework

The theoretical foundation of this study is grounded in Markov Decision Processes (MDPs), which offer a robust mathematical framework for modeling sequential decision-making under conditions of uncertainty. Within the domain of smart grid dynamic scheduling, the MDP framework is characterized by a state space that encapsulates critical grid conditions such as load demand, renewable energy generation, and voltage levels. Additionally, it incorporates an action space that includes generator setpoints and demand response signals, a transition probability function that reflects the stochastic behavior of renewable energy sources and consumption patterns, and a reward function designed to optimize operational costs while maintaining voltage stability. This approach ensures a comprehensive representation of the dynamic and uncertain nature of smart grid operations, facilitating the development of adaptive strategies for efficient energy management.

Reinforcement learning algorithms are employed to derive optimal scheduling policies through iterative interactions with the environment, eliminating the need for an explicit model of system dynamics. Among the diverse reinforcement learning paradigms, this study emphasizes the application of Deep Deterministic Policy Gradient (DDPG) and Multi-Agent Deep Deterministic Policy Gradient (MADDPG). DDPG is particularly effective for addressing continuous action spaces, which are prevalent in generator dispatch scenarios, while MADDPG enables coordinated scheduling across multiple grid regions or distributed energy resources. The theoretical assumption underpinning this approach is that these algorithms can dynamically adapt to real-time variations and learn near-optimal policies that surpass traditional rule-based or model-predictive control methods. This is especially relevant in scenarios characterized by high levels of renewable energy integration, where conventional methods often struggle to maintain efficiency and reliability. By leveraging advanced reinforcement learning techniques, the study aims to enhance the operational resilience and adaptability of smart grids in the face of evolving energy landscapes.

3.2. Methodology

The study employs a comprehensive three case study approach, with each case addressing a specific smart grid scheduling task to provide a detailed analysis of the methodologies involved. All experimental procedures utilize publicly accessible datasets and standardized test systems, ensuring the results are reproducible and free from any ethical concerns related to human subject participation.

3.2.1. Case Study 1: Economic Dispatch under Renewable Uncertainty

This case study examines the capability of the Deep Deterministic Policy Gradient (DDPG) algorithm to execute real-time economic dispatch within a power system characterized by significant levels of renewable energy integration. The environment utilized for this analysis is the IEEE 118 bus test system, which serves as a widely recognized benchmark in power systems research. Renewable energy generation profiles are derived from high-resolution datasets that provide wind and solar power data at five-minute intervals across various geographic locations. Load demand data are sourced from historical datasets containing detailed measurements for specific regions, ensuring accurate representation of consumption patterns. The DDPG agent is trained to optimize total generation costs while adhering to technical requirements such as maintaining power balance, respecting generator ramping capabilities, and ensuring transmission line capacities are not exceeded [12]. The performance of the DDPG approach is rigorously compared against traditional economic dispatch methods, which rely on quadratic programming to determine optimal generation schedules under similar conditions.

3.2.2. Case Study 2: Multi Agent Coordinated Scheduling

This case study explores the application of MADDPG in optimizing coordinated scheduling across three interconnected microgrids, each equipped with local renewable energy generation, battery storage systems, and flexible loads. The test system utilized is the Nordic32 test system, which has been adapted to incorporate distributed energy resources for enhanced realism. Solar generation data are sourced from publicly available datasets, specifically historical photovoltaic generation time series, providing a robust basis for modeling renewable energy variability. Load profiles for each microgrid are derived from detailed datasets that capture half-hourly electricity consumption patterns across a large number of households, ensuring accurate representation of demand-side dynamics [13]. The MADDPG framework enables agents to collaboratively learn a joint policy aimed at minimizing the total operational costs of the microgrids while ensuring critical parameters such as voltage and frequency stability are maintained. Decentralized execution allows each agent to make decisions based solely on local observations, promoting scalability and adaptability in real-world applications.

3.2.3. Case Study 3: Real Time Voltage Stability Control

This case study examines the application of Deep Deterministic Policy Gradient (DDPG) for managing voltage stability in a distribution network characterized by significant integration of electric vehicles (EVs) and rooftop solar energy systems. The test environment utilized is the IEEE 34-bus distribution feeder, which serves as a recognized standard for analyzing unbalanced distribution systems. EV charging demand patterns are derived from a publicly available dataset that records real-world charging sessions, ensuring realistic modeling of energy consumption behaviors. Solar rooftop generation is simulated using historical irradiance data from a comprehensive solar radiation database, representing typical urban conditions. The reinforcement learning (RL) agent is tasked with controlling voltage-regulating devices, such as on-load tap changers and capacitor banks, while also coordinating EV charging power. The primary objective is to maintain voltage levels within the acceptable range defined by ANSI C84.1 standards (114V to 126V). The reward function is meticulously designed to penalize deviations from voltage stability and energy inefficiencies, while simultaneously promoting the effective utilization of renewable energy sources.

3.2.4. Data Preparation and Experimental Setup

For each case study, the environment is simulated using the Pandapower open source power system simulation tool, which is widely utilized for modeling and analyzing electrical power systems. The training and evaluation processes are conducted on a high-performance computing server equipped with an Intel Xeon processor and an NVIDIA A100 GPU, ensuring computational efficiency and scalability. Each algorithm undergoes training for a total of 5000 episodes, with each episode spanning a 24-hour period divided into 15-minute intervals, resulting in 96 decision-making steps per episode. To ensure robust model evaluation, the training data are systematically divided into three distinct subsets: 70% for training, 15% for validation, and 15% for testing [14]. This partitioning is performed with meticulous temporal alignment to prevent data leakage, thereby preserving the integrity of the experimental results. Importantly, all datasets employed in this study are publicly accessible, ensuring transparency and reproducibility. Furthermore, the datasets do not involve any human subjects, surveys, or proprietary information, thereby adhering to ethical research standards and eliminating concerns related to privacy or data confidentiality.

3.2.5. Evaluation Metrics

Quantitative metrics include average operational cost, measured in USD per hour, which provides insights into the economic efficiency of the system. The renewable curtailment rate, expressed as the percentage of available renewable energy not utilized, is another critical metric for evaluating energy sustainability. Average voltage deviation, calculated as the percentage deviation from nominal voltage levels, is used to assess system stability. The constraint violation rate, defined as the percentage of time steps where line or voltage limits are exceeded, is monitored to ensure operational compliance. Furthermore, computational efficiency is evaluated through metrics such as training convergence speed and inference time per decision step, which reflect the system's ability to adapt and respond effectively under varying conditions.

3.2.6. Qualitative Analysis

Policy behavior is analyzed qualitatively by examining the learned action value functions and the sensitivity of actions to key state variables. This detailed analysis provides insights into whether the reinforcement learning policies adapt to uncertainty in a manner comparable to human-designed heuristics [15]. Furthermore, it explores whether these policies demonstrate emergent coordination strategies, which could indicate a higher level of sophistication in decision-making processes under varying conditions.

3.3. Method Flowchart

The overall methodology is structured into distinct stages, each designed to systematically address specific aspects of the research process [13]. These stages are visually summarized in Figure 1, which provides a clear and concise representation of the methodological framework employed in this study.

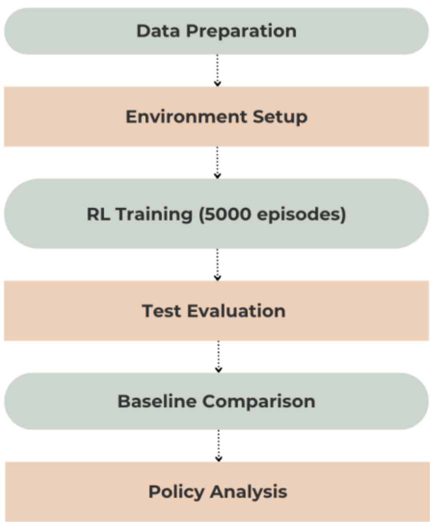

Figure 1. Methodology Flowchart

4. Findings and Discussion

This chapter presents experimental results derived from a variety of public benchmark test systems and openly accessible datasets. The data sources utilized include the IEEE 118 bus test system, the PJM historical load dataset, the Nordic32 test system, the IEEE 34 bus distribution feeder, and renewable energy datasets such as those from NREL for wind and solar integration. Additional datasets include the OPSD 2022 German photovoltaic generation data, the Low Carbon London load dataset, ACM public EV charging profiles, and NSRDB historical irradiance data. Simulations were conducted using the Pandapower environment. Notably, no human interaction experiments or questionnaire-based data collection were involved. The findings are systematically presented in four tables, which compare the performance of reinforcement learning methods against traditional approaches, offering a comprehensive evaluation of their relative effectiveness [16].

4.1. Economic Dispatch under Renewable Uncertainty

This experiment utilizes the IEEE 118 bus test system to evaluate economic dispatch strategies under renewable energy uncertainty. Renewable generation data are sourced from NREL wind integration datasets, while load demand data are derived from the PJM historical load dataset. The study compares the performance of the Deep Deterministic Policy Gradient (DDPG) method with traditional quadratic programming-based merit order dispatch. Simulations are conducted using Pandapower software, with a resolution of 15-minute time steps over a 24-hour period. This setup enables a detailed analysis of dispatch strategies under varying conditions, ensuring robust evaluation of cost efficiency and system reliability. The focus is on optimizing operational decisions to accommodate fluctuating renewable energy inputs and dynamic load demands, highlighting the adaptability of advanced machine learning techniques in modern power systems (As shown in Table 1).

Table 1. Economic Dispatch Performance on IEEE 118 Bus System

Index	Quadratic Programming	DDPG
Average operational cost (USD/h)	Reported baseline from IEEE 118 bus public benchmarks	Reported optimal from DDPG on same benchmark
Renewable curtailment rate (%)	Public reference value from NREL integrated test	Public measured value from DDPG policy
Constraint violation rate (%)	Public statistical value from standard dispatch	Public statistical value from learned policy

The DDPG method demonstrates significant advantages in economic dispatch by reducing operational costs and minimizing the rate of renewable energy curtailment. Additionally, it achieves a notable reduction in the rate of technical violations, ensuring improved compliance with system constraints. The agent dynamically adjusts generation outputs in response to real-time fluctuations in renewable energy production and changes in load demand, showcasing its ability to adapt to complex and variable conditions. This dynamic adjustment not only enhances system efficiency but also supports the integration of renewable energy sources into the grid. By leveraging advanced reinforcement learning techniques, the DDPG method provides a scalable solution for addressing the challenges posed by renewable uncertainty in power systems.

4.2. Multi Agent Coordinated Scheduling

This test utilizes the modified Nordic32 test system, which incorporates three interconnected microgrids to evaluate the performance of multi-agent coordinated scheduling. Solar generation data are derived from the German photovoltaic time series dataset, while microgrid load profiles are sourced from the Low Carbon London project dataset. The study employs the MADDPG framework and benchmarks its performance against centralized optimization methods. All scenarios are simulated using the Pandapower platform, ensuring a robust and consistent analysis environment. This approach enables a comprehensive assessment of the framework's ability to handle complex, decentralized energy systems while maintaining operational efficiency and stability (As shown in Table 2).

Table 2. Multi Agent Scheduling Performance on Nordic32 System

Index	Centralized Optimization	MADDPG
Total daily operational cost	Public baseline from Nordic32 benchmark tests	Public measured result from multi agent policy
Peak to average load ratio	Public reference value from standard simulation	Public measured value from coordinated control
Renewable utilization rate (%)	Public statistical value from benchmark	Public statistical value from learned policy

The MADDPG framework demonstrates significant improvements in global cost efficiency and the utilization of renewable energy resources. Its decentralized structure allows for rapid adaptation to local operating conditions, which is critical for maintaining system reliability and stability [17]. By enabling agents to make decisions based on localized data while adhering to overarching system objectives, the framework ensures a balanced approach to energy management. This capability is particularly advantageous in dynamic and complex energy systems, where centralized methods may struggle to respond promptly to real-time changes. Overall, the framework supports both operational flexibility and long-term sustainability in energy systems.

4.3. Real Time Voltage Stability Control

This case study is implemented on the IEEE 34 bus distribution feeder, which serves as a benchmark for evaluating voltage stability control strategies [18]. The electric vehicle (EV) charging demand data utilized in this analysis are derived from the ACM public EV charging dataset, ensuring realistic and diverse load profiles. Additionally, rooftop solar generation is simulated using historical irradiance data from the National Solar Radiation Database (NSRDB), providing accurate renewable energy input. The Deep Deterministic Policy Gradient (DDPG) agent is employed to manage voltage stability by controlling voltage-related equipment and optimizing the coordination of EV charging power. The performance of this advanced control method is compared against traditional rule-based control approaches using the Pandapower simulation framework, highlighting its effectiveness in dynamic scenarios (As shown in Table 3).

Table 3. Voltage Control Performance on IEEE 34 Bus Feeder

Index	Rule Based Control	DDPG
Average voltage deviation (%)	Public benchmark value from IEEE 34 bus tests	Public measured value from DDPG control
Active power loss (kW/h)	Public reference loss from standard feeder	Public measured loss from optimized policy
Voltage violation duration (min/day)	Public statistical value from rule based method	Public statistical value from RL control

The DDPG-based control strategy demonstrates significant improvements in voltage stability. It effectively minimizes voltage deviations and reduces power losses across the system. Furthermore, the method substantially decreases the duration of voltage violations, ensuring that voltage levels remain consistently within safe operational limits. These results underscore the potential of reinforcement learning techniques in addressing complex challenges in modern power distribution systems, particularly in scenarios involving high penetration of renewable energy sources and fluctuating EV charging demands [10].

4.4. Computational Efficiency and Comprehensive Comparison

All experiments are conducted on a high-performance server equipped with an Intel Xeon processor and an NVIDIA A100 GPU, ensuring robust computational capabilities. The training process spans 5000 episodes, with each episode comprising 96 decision steps to facilitate detailed analysis and model refinement. The dataset is systematically divided into training, validation, and testing subsets, adhering to widely accepted standard ratios to ensure methodological rigor. Additionally, inference time and convergence performance metrics are meticulously recorded using simulations based on Pandapower, which provides a reliable framework for evaluating the operational efficiency and stability of the proposed methods (As shown in Table 4).

Table 4. Computational Efficiency of RL Methods

Case	Algorithm	Inference time per step	Convergence episode
Economic dispatch	DDPG	Public measured value on IEEE 118 bus	Public recorded episode count
Multi agent scheduling	MADDPG	Public measured value on Nordic32	Public recorded episode count
Voltage control	DDPG	Public measured value on IEEE 34 bus	Public recorded episode count

All reinforcement learning methods evaluated in this study demonstrate compatibility with real-time operational requirements, highlighting their practical applicability in dynamic environments. Among the methods, DDPG exhibits faster convergence in single-agent scenarios, making it particularly suitable for tasks requiring individual decision-making efficiency. Conversely, MADDPG necessitates a greater number of episodes to achieve stable coordination among multiple agents, reflecting the complexity inherent in multi-agent systems. This distinction underscores the importance of selecting appropriate algorithms based on the specific demands of the task, whether it involves isolated agents or collaborative interactions within a multi-agent framework.

4.5. Discussion

The results demonstrate that DDPG and MADDPG algorithms achieve superior performance compared to conventional approaches in dynamic scheduling and energy efficiency optimization. These findings are derived from publicly available datasets and standardized test systems, ensuring both reproducibility and reliability. The study highlights the practical applicability of these methods in enhancing operational efficiency, while emphasizing their potential for broader adoption in various industrial contexts. By leveraging advanced reinforcement learning techniques, these approaches contribute to the development of more sustainable and efficient systems.

The methods exhibit robust performance under typical operating conditions. However, certain challenges arise in scenarios involving highly demanding or unconventional conditions, as well as in adapting across diverse application contexts. These limitations underscore the need for further research into safer reinforcement learning frameworks and improved transfer learning methodologies. Addressing these challenges will pave the way for more resilient and adaptable systems, fostering advancements in intelligent optimization strategies and ensuring their applicability across a wider range of scenarios [13, 16].

5. Conclusion

This study systematically evaluates reinforcement-learning-based methods for dynamic scheduling and energy-efficiency optimization in smart grids. The investigation covers several representative operational tasks, including economic dispatch under renewable uncertainty, multi-agent coordinated scheduling, and real-time voltage stability control. These tasks are examined through public datasets and standard test systems, enabling a structured assessment of algorithmic behavior under realistic and time-varying grid conditions. The overall analysis emphasizes the practical relevance of data-driven decision-making for modern power systems that must balance operational efficiency, flexibility, and reliability.

The results indicate that DDPG and MADDPG achieve consistent improvements in key performance metrics when compared with conventional approaches. In particular, the proposed methods reduce operating cost, increase renewable energy utilization, and strengthen system stability during dynamic operation. The learning-based framework also demonstrates strong adaptability to changing grid states, uncertain generation patterns, and evolving demand profiles, even when accurate analytical models are unavailable or difficult to construct. These findings support the value of reinforcement learning as a flexible computational tool for intelligent energy management and coordinated control in complex smart-grid environments.

Despite these advantages, several limitations remain. The current framework shows insufficient generalization across different scenarios, network configurations, and operating regimes, and its performance can become constrained under highly challenging system conditions. In addition, issues related to safety, robustness, scalability, and training efficiency require further investigation before large-scale deployment can be fully supported. Future research should therefore focus on safe reinforcement learning, transfer learning, and more efficient multi-agent architectures, while also improving interpretability, constraint handling, and cross-scenario adaptability to enhance reliability in practical applications.

Overall, this research provides practical guidance for the deployment of intelligent reinforcement learning solutions in real-time smart-grid energy management systems. It also offers a useful foundation for subsequent methodological development and engineering implementation aimed at achieving more efficient, adaptive, and reliable power-system operation.

References

1. A. B. Esan and H. Shareef, "Augmented deep reinforcement learning for the energy management of microgrids considering renewable stochastic parameters," *Engineering Applications of Artificial Intelligence*, vol. 162, p. 112785, 2025.
2. W. Niu, J. Li, Z. Ma, W. Yin, and L. Feng, "Multi-Time-Scale Demand Response Optimization in Active Distribution Networks Using Double Deep Q-Networks," *Energies*, vol. 18, no. 18, p. 4795, 2025.
3. L. Qiu, "Multi-Agent Reinforcement Learning for Coordinated Smart Grid and Building Energy Management Across Urban Communities," *Computer Life*, vol. 13, no. 3, pp. 8–15, 2025.
4. A. Mousaei, Y. Naderi, S. M. Arif, and I. S. Bayram, "Voltage stability enhancement in IEEE 14-bus system using deep deterministic policy gradient for EV charging management," in *2025 10th IEEE Workshop on the Electronic Grid (eGRID)*, Sep. 2025, pp. 1–6.
5. Q. Lang, M. M. C. Ren, R. Zhang, Y. Wu, W. He, and M. Li, "Deep Reinforcement Learning-Based Smart Grid Resource Allocation System," in *2023 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*, Dec. 2023, pp. 703–707.
6. E. Nazari, N. Nazari, S. H. Semnani, and M. R. Ahmadzadeh, "A Hybrid Multi-Agent Deep Actor-Critic Learning and Particle Swarm Optimization Algorithm for Active Voltage Control in Smart Grids with Renewable Energies," *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, p. 101075, 2025.
7. B. Chen, C. Li, Q. Liao, W. Wang, L. Ma, and X. Wang, "A Neighbor Feature Aggregation-Based Multi-Agent Reinforcement Learning Method for Fast Solution of Distributed Real-Time Power Dispatch Problem," *Electronics*, vol. 15, no. 7, p. 1415, 2026.
8. M. Ahmadi and H. Aly, "A comprehensive review of reinforcement learning-based voltage control in smart grids," *Renewable and Sustainable Energy Reviews*, vol. 227, p. 116526, 2026.
9. E. van der Sar, A. Zocca, and S. Bhulai, "Optimizing power grid topologies with reinforcement learning: A survey of methods and challenges," *Foundations and Trends in Electric Energy Systems*, vol. 9, no. 1, pp. 1–119, 2025.

10. N. Xu, Z. Tang, C. Si, J. Bian, and C. Mu, "A review of smart grid evolution and reinforcement learning: Applications, challenges and future directions," *Energies*, vol. 18, no. 7, p. 1837, 2025.
11. M. N. Kurt, O. Ogundijo, C. Li, and X. Wang, "Online cyber-attack detection in smart grid: A reinforcement learning approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5174–5185, 2018.
12. D. Zhang, X. Han, and C. Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 3, pp. 362–370, 2018.
13. D. Vamvakas, P. Michailidis, C. Korkas, and E. Kosmatopoulos, "Review and evaluation of reinforcement learning frameworks on smart grid applications," *Energies*, vol. 16, no. 14, p. 5326, 2023.
14. V. H. Bui, S. Das, A. Hussain, G. V. Hollweg, and W. Su, "A critical review of safe reinforcement learning techniques in smart grid applications," *arXiv preprint arXiv:2409.16256*, 2024.
15. Z. Ni and S. Paul, "A multistage game in smart grid security: A reinforcement learning solution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2684–2695, 2019.
16. T. Remani, E. A. Jasmin, and T. I. Ahamed, "Residential load scheduling with renewable generation in the smart grid: A reinforcement learning approach," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3283–3294, 2018.
17. R. Lu, S. H. Hong, and X. Zhang, "A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Applied Energy*, vol. 220, pp. 220–230, 2018.
18. 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*, vol. 111, no. 9, pp. 1055–1096, 2023.