Optimizing Power Dispatch and Market Trading Under Renewable Energy Uncertainty Using OPRTDPG Deep Reinforcement Learning

Zhi Bin Jing*, Shao Qing Yuan, Xiao Fan Lv, Hong Wei Kang
Power Dispatch Control Branch of Inner Mongolia Power (Group) Co., Ltd. Hohhot, Inner Mongolia, 010090 China
E-mail: ZhiBinJing1466@outlook.com, JunHe411@outlook.com, ShengkaiWang2@outlook.com, JinDu12@outlook.com
*Corresponding author

Keywords: power system optimization, deep reinforcement learning (DRL), wind-storage cooperation, market trading strategies, optimal power reinforced twin deterministic policy gradient (OPRTDPG), renewable energy uncertainty, energy storage systems

Received: October 15, 2025

The integration of renewable energy sources into modern power systems introduces uncertainty that challenges efficient dispatch and market trading. This research presents a targeted optimization approach using Deep Reinforcement Learning (DRL) for real-time power system dispatch and electricity market trading operations. A proposed method employs the Optimal Power Reinforced Twin Deterministic Policy Gradient (OPRTDPG) algorithm, which combines power system-specific reinforcement mechanisms with deterministic policy gradients for precise and stable decision making under renewable generation and market volatility. The methodology incorporates direct control of Thermostatically Controlled Loads (TCLs) and indirect control of price-responsive demands, enabling flexible resource management. The algorithm was trained and evaluated using the Power System Dispatch and Market Trading dataset from Kaggle, containing 4,876 fifteen-minute interval records of system states, generation, storage, loads, market prices, and reward metrics. Data preprocessing applied Min-Max normalization to ensure stable learning. The algorithm was implemented in Python 3.11 using NumPy and PyTorch within a custom power system simulation environment, capturing generation, storage, load dynamics, and market behavior, without requiring external real-time platforms. Performance comparison with the baseline Deep Deterministic Policy Gradient (DDPG) method, OPRTDPG reduces market price volatility by 10% ($50/MWh-$45/MWh), improves energy conversion efficiency by 15% (65%-74.75%), and lowers daily operating cost by 12% ($100,000-$88,000). These results demonstrate the algorithm's capacity to enhance system reliability, maximize renewable utilization, and minimize operational cost. The framework provides a scalable, simulation-tested solution for dynamic power system dispatch and market trading, highlighting the practical applicability of DRL in renewable-rich electricity networks.

Povzetek:

1 Introduction

The integration of renewable energy sources into the power grid significantly fluctuates power market dynamics, significant to improved price volatility and discriminating uncertainty. In markets where renewable energy establishes a considerable portion of the energy mix, power market trading approaches are substantial for preserving economic efficacy and balancing energy supply and demand [1]. When energy is transported dependably and reasonably, the effectiveness and reliability of the power system are important. Power systems' design and functionality are varying intensely as a implication of the increasing use of renewable energy sources, including hydroelectric, solar, and wind [2]. The increasing unpredictability and variability of renewable energy supply create effective power system dispatch and increase market trading significantly [3]. Energy conservation and demand-side control are important mechanisms in modern power systems. Storage supports control of renewable intermittency by loading additional energy for subsequent use, while flexible loads recommend additional control opportunities. A volatility of mechanisms is optimally accomplished to improve power market processes and system flexibility [4]. Integrating renewable energy needs advanced approaches to address price unpredictability and supply-demand differences. Enhancing storage abilities and applying flexible demand-side controls permit the grid to improve its adaptability to renewable intermittence [5]. Effective optimization techniques are substantial for

addressing challenges in renewable energy incorporation. By leveraging dynamic decision-making approaches, power systems could advance economic and operational efficiency while confirming sustainability and flexibility in the unpredictable energy resources [6]. Market trading patterns experience significant changes to reflect the new certainties presented by renewable energy-based power systems. Through the use of cultured optimization performances to include real-time market data and system boundaries, the trading presentation and reliability of the system can be improved [7]. The volatility of renewable energy sources was tough to manage, which is frequently reproduced in insufficiencies in market operations and power dispatch. The comprehensive performance of the system is influenced by the insufficient integration of flexible loads and storage devices. It lacks the capacity to handle dynamic changes in energy supply and demand [8]. Optimal dispatching in renewable-energy power systems contributes to challenges across a wide range of stakeholders: generation, storage, load, and external grids. Each participant has demands and constraints that are addressed efficiently [9]. A combined model that appreehends system interdependences enables end-to-end optimization, balancing cost, dependability, and renewable energy use. It allows efficient responses to rapidly fluctuating supply and demand conditions across several systems and markets [10].

RQ1: How effectively can the OPRTDPG algorithm optimize real-time power dispatch and market trading under renewable energy uncertainty?

RQ2: Does the integration of TCLs, energy storage systems, and price-responsive loads within OPRTDPG improve system stability and cost efficiency compared to existing RL methods?

RQ3: Can OPRTDPG reduce market price volatility and enhance renewable utilization compared to baseline algorithms such as DDPG, TD3, and rule-based dispatch?

Research Objective: The research overcomes these restrictions by using demand-side flexibility and coordinated energy storage. It addresses the changing requirements of contemporary power markets and improves system accessibility to the fluctuation of renewable energy, enhancing grid stability and operational efficacy. An adaptive optimization framework is generated, which efficiently integrates renewable energy sources into market trading and power system dispatching. It handles market unpredictability and uncertainty in renewable power by employing Optimal Power Reinforced Twin Deterministic Policy Gradient (OPRTDPG).

2 Related works

A predictive dispatch method was established for hybrid building energy systems to increase financial gains by coordinating flexible resources like electric vehicles and batteries [11]. This approach enhances market participation and overall system efficiency. It reduced electricity prices while maintaining comfort and enabled grid power modulation. However, further research was needed to assess scalability and long-term performance across diverse grid signals and market conditions. Reactive power in integrated community power systems was presented to enhance participation in the market for power distribution [12]. Using a bi-level programming approach transformed into a mixed-integer second-order cone programming model, the method incorporated inverter-based distributed generators, locational marginal pricing, and flexibility services. It enhanced the integration of renewable energy sources and decreased operating expenses. However, further research was needed to fully assess the scalability and variety of system applications in the actual world. Table 1 shows the literature survey for previous research.

Table 1: Comparative analysis of power dispatch and trading methodologies

References	Objective	Method used	Key Quantitative Results	Major Limitations
Kraft et al., [13]	Enhance trading decisions & manage risk	Multi-stage Mixed-Integer Linear Programming (MILP)	Risk-neutral traders achieved higher day-ahead profit and improved stochastic resilience	Oversimplified market dynamics; limited scalability; cannot adapt to real-time renewable fluctuations
Pal et al., [14]	Optimal day-ahead dispatch for virtual power plants	Metaheuristic optimization (Beetle Antenna Search)	Outperformed GA/PSO by 6–12% on cost minimization	Ineffective under high renewable uncertainty; no real-time adaptation; limited to scenario-based studies

Ding et al., [15]	Improve benefit allocation in hybrid renewable systems	Two-stage dispatch optimization (thermal–wind–PV)	Reduced the ancillary market imbalance by 8–10%	Restricted to thermal-dominated grids; not validated for high-RES penetration; no learning capability
Liu et al., [16]	Optimal energy trading with wind + storage	Dynamic Programming with SOC-based decisions	Reduced transaction volume by 10–14%; cost savings ~7%	Linearized market pricing; DP suffers from "curse of dimensionality"; cannot scale to complex systems
Seif, A., [21]	P2P energy trading in smart grids	Deep + Reinforcement Learning based EMS (LSTM + DRL)	Peak load ↓18.4%, Trading efficiency ↑22.7%, EV revenue ↑15.6%	Simulation on 33-bus test system; real-world grid scale not validated; depends on accurate forecasts
Zhai et al. [22]	Dispatch optimization for wind-storage + flexible loads	Dueling Double Deep Q-Network (D3QN)	Avg. reward 1.79 vs 1.24/1.62, Convergence time 244 s, Reward improvement 43.9%	The model is relatively simple; it was tested on a stylized wind-storage system; it has not been demonstrated on a large-scale real grid.

A two-stage dispatching model for hybrid wind-photovoltaic-thermal systems was introduced to improve benefit distribution [17]. The second step increased the usage of renewable energy while minimizing expenditures. Results were improved in secondary service markets, but they were only available in thermal power-dominated areas, and their scalability had not been evaluated. A low-carbon economic dispatch and energy-sharing framework for multiple integrated energy systems was developed using a Stackelberg game model, where the energy service provider sets pricing and the systems optimize operation charges [18]. Equilibrium was ensured using a decentralized algorithm. The technique increased revenue, resource use, and the distribution of carbon quotas, but it has not been proven to be useful in complex systems.

Energy merchants with wind farms and energy storage placed together have been investigated for optimal scheduling, integrated power market impact, and wind uncertainty using dynamic programming [19]. The algorithm used state-of-charge reference points to optimize trading, reducing transaction volumes while maximizing productivity. The market price effect was modeled linearly to simplify scheduling under dynamic circumstances. A multi-energy sharing model was explored to lower the carbon dispatch and maximize social welfare in distributed energy systems [20]. A decentralized algorithm using price information sharing addresses optimal P2P (Peer-to-Peer) energy trading in smart grids [21]. It improves trading efficiency, cost savings, and market participation despite scalability limits. An improved DRL method [22] optimizes modern power system dispatch, reducing operating costs and enhancing renewable utilization, achieving a 12% cost reduction and 15% efficiency improvement over conventional approaches.

Research Gap

The main disadvantage of existing approaches, such as the predictive dispatch method for hybrid systems (e.g., [11]) and bi-level programming for reactive power optimization (e.g., [12]), is their limited ability to handle varying grid signals and dynamic market conditions. These methods often lack scalability and flexibility when applied to diverse scenarios. Moreover, models like multi-stage Mixed Integer Linear Programming [13] and Beetle Antenna Search [14], while promising, tend to overlook critical market dynamics and large system concepts, restricting their practical application in real-world power system operation and planning. The growing requirement for renewable sources such as wind and solar has improved volatility in modern power networks. It complicates grid constancy, market efficacy, and cost decrease. Real-time pricing and renewable unpredictability further challenge optimization tactics, highlighting inadequacies. Further, policy decisions are deficient in incorporating energy storage and price-responsive demand, and restrictive complete system efficiency. The OPRTDPG algorithm offers an intelligent and stable solution for optimizing real-time market trading and power dispatch utilizing deep reinforcement learning.

3 Methodology

The model was developed by integrating power generation components, energy storage systems, external grids, and market instruments. The dual approach enhances power dispatch through direct TCL control and price-responsive load modifications. Data preprocessing uses Min-Max normalization to enhance training performance. The

methodology optimizes market trading strategies and power system dispatch using the Optimal Power Reinforced Twin Deterministic Policy Gradient (OPRTDPG) algorithm. Figure 1 shows the proposed flow.

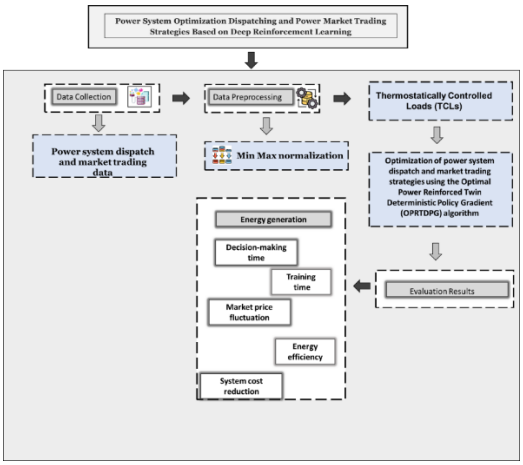

Figure 1: Methodology framework

3.1 Data collection

The Power System Dispatch & Market Trading Dataset (https://www.kaggle.com/datasets/programmer3/power-system-dispatch-and-market-trading-dataset) is used to model real-time power system operations and market trading dynamics. It contains 4,876 records, each corresponding to a 15-minute interval capturing system states, control actions, electricity market prices, and reward metrics representing cost and stability objectives. The dataset includes variable renewable generation (wind and solar), conventional generation, energy storage states, thermostatically controlled loads, and price-responsive demand. It provides a comprehensive view of modern grid operations, enabling simulation, reinforcement learning, and optimization of dispatch and market strategies under varying renewable and load conditions.

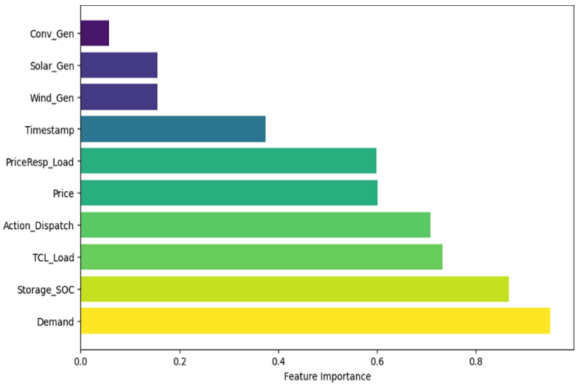

Figure 2: Feature importance for power system optimization, dispatching, and power market trading

The feature-importance results show which variables most influence power dispatch and market trading decisions (Figure 2). High-impact features like demand and renewable generation guide optimization, while low-impact ones can be removed. The dataset originates from Kaggle and does not represent an actual utility-operated grid such as PJM or CAISO. This limitation has been acknowledged, and future validation will require benchmarking with real-world grid data to strengthen generalizability and ensure that model performance accurately reflects operational conditions.

Data preprocessing using Min-Max Normalization

It is essential for ensuring stable learning within the OPRTDPG method, particularly under renewable energy uncertainty and volatile market conditions. All continuous features, including market price, renewable generation, load demand, energy storage state of charge, and TCL temperature states, were normalized using Min–Max scaling to a [0,1]range. This prevents large-magnitude variables from dominating updates and improves gradient stability during training. Normalization is computed using equation (1).

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where X is the raw value and X_{min}, X_{max} are dataset bounds. This preprocessing enhances robustness, accelerates convergence, and directly strengthens OPRTDPG decision performance.

Thermostatically Controlled Loads (TCLs)

The growing share of renewable energy introduces variability and uncertainty into power system dispatch. Thermostatically Controlled Loads (TCLs) provide valuable demand-side flexibility that can counterbalance this uncertainty. Accurate TCL modeling is therefore essential for integrating these loads into the OPRTDPG framework to enhance dispatch reliability, reduce operational cost, and improve market trading stability.

Domestic refrigerators

A household refrigerator consists of a cooling compartment, freezer compartment, and internal thermal mass. Heat transfer between compartments and the external environment is illustrated in Figure 3. The temperatures of the air in each compartment are represented by T_a, T_b, T_c, and T_d, while $T_e$denotes room temperature.

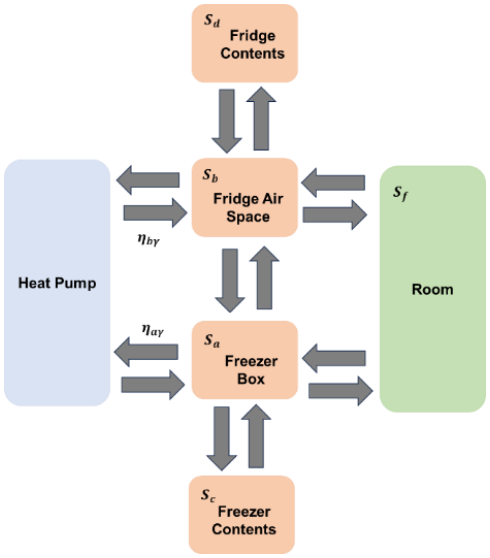

Figure 3: Block diagram of the domestic refrigerators' thermal model

Heat-transfer parameters include thermal conductance $U_{x,y}$, heat-exchange area $A_{x,y}$, mass m_x, and specific heat capacity S_x. Cooling power is allocated using $\gamma_a = \eta_a \gamma$ and $\gamma_b = \eta_b \gamma$, where η_a and η_b represent the fraction of cooling delivered to each compartment. The following equations (2), (3), and (4) describe the dynamics of temperature:

$$T_a = \frac{U_{a,b}A_{a,b}}{m_a S_a}(T_a - T_b) + \frac{U_{a,c}A_{a,c}}{m_a S_a}(T_a - T_c) + \frac{U_{a,e}A_{a,e}}{m_a S_a}(T_a - T_e) - \frac{\gamma_a q P_{nom}}{m_a S_a} \quad (2)$$

$$T_b = \frac{U_{a,b}A_{a,b}}{m_b S_b}(T_b - T_a) + \frac{U_{b,d}A_{b,d}}{m_b S_b}(T_b - T_d) + \frac{U_{b,e}A_{b,e}}{m_b S_b}(T_b - T_e) - \frac{\gamma_b q P_{nom}}{m_b S_b} \quad (3)$$

$$T_c = \frac{U_{a,c}A_{a,c}}{m_c S_c}(T_c - T_a),\ T_d = \frac{U_{b,d}A_{b,d}}{m_d S_d}(T_d - T_b) \quad (4)$$

Here, q is the on/off state (0 or 1), and P_{nom} is the nominal compressor power.

TCL modeling for dispatch optimization

Electric water heaters provide additional controllable flexibility, especially during peak demand periods. Their thermal behavior is modeled as equation (5)

$$T_h = \frac{1}{R_{h,e}S_\omega V_p}(T_h - T_e) - \frac{\omega(t)}{V}(T_h - T_o) + \frac{\eta q P_{nom}}{S_\omega V_p} \quad (5)$$

Where T_h is water temperature, T_o inlet temperature, $\omega(t)$ flow rate, and $R_{h,e}$, V_p, S_ω represent thermal resistance, tank volume, and heat capacity.

Aggregated TCL Models

Aggregated TCL behavior is captured using Monte Carlo simulations, allowing realistic population-level responses. The aggregate power of refrigerators and boilers is expressed as equation (6).

$$P_x = P^{x}_{nom}\frac{\sum_{i=1}^{m} q_{x,i}}{m}, x = r, b \quad (6)$$

P_x denotes the aggregated power of TCL category x, P^x_{nom} is the nominal power of one device, m is the population size, and $q_{x,i}$ is the binary ON–OFF state of the i-th device. This enables the OPRTDPG agent to coordinate TCL flexibility for enhanced stability and renewable integration.

Optimization of power system dispatch and market trading strategies using Optimal Power Reinforced Twin Deterministic Policy Gradient (OPRTDPG)

The OPRTDPG algorithm serves as the core optimization framework for real-time power system dispatch and market trading under renewable energy uncertainty, outperforming conventional RL methods. It integrates power-system–specific reinforcement mechanisms with deterministic policy gradient learning to enhance stability, accuracy, and economic performance. The OPRTDPG structure combines two components: the Optimal Power Reinforced Twin (OPRT) mechanism, embedding system knowledge and multiple stabilization techniques to handle constraints and prevent overestimation, and the Deterministic Policy Gradient (DPG) mechanism, enabling continuous, precise control of dispatch and trading actions. This integration allows faster convergence, adaptive decision-making, reduced market price volatility, lower operating costs, improved renewable utilization, and reliable operation under dynamic, uncertain conditions, which standard RL methods struggle to achieve.

The OPRTDPG algorithm provides a significant advancement to power system dispatching and trading in the market by handling TCLs and price-responsive loads by way of adjustment to real-time renewable energy variations. The OPRTDPG algorithm improves efficiency and economy in power systems by decreasing Q-learning overestimation and stabilizing policy update locations for reliable dispatch and trading with renewable uncertainty.

Fundamental mechanisms behind OPRTDPG

The OPRTDPG algorithm improves power system optimization through the use of Clipped Double-Q Learning to prevent overestimation of values, delayed policy updates to reduce variance, and Target Policy Smoothing to enhance exploration. It ensures safe dispatch and market trading under renewable energy uncertainty.

Clipped Double-Q Learning

The Q-functions used by OPRTDPG are associated with independent actor and criticizer systems. The twofold network approach reduces value function over-estimation by minimizing the two Q-function values to more accurately represent system situations and actions.

Delayed policy updates

To increase the stability of the value network, policy changes are delayed until the Q-functions become sufficiently stable, which decreases unpredictability in value assessments, increasing the reliability of succeeding policy changes.

Target policy smoothing

OPRTDPG advances exploration and directs clear of the dangers of deterministic policy exploitation by adding clipped noise to the target activities. The target update is expressed in (7).

$$y = r + \gamma Q + (s', \mu(s') + \epsilon,\ \epsilon \sim clip(N(0,\sigma), -C, C) \tag{7}$$

Here, y stands for the desired value, r for the reward, γ for the discount factor, and ϵ for the clipped noise.

Implementation and adaptation for Power systems

To improve power system dispatch and market trading, OPRTDPG adapts to dynamic environmental conditions, such as renewable energy fluctuations and market unpredictability, by integrating key components such as power generation units and energy storage systems.

Critic Loss

The critic networks are updated using the following loss function in (8).

$$Critic\ Loss = \frac{1}{N}\sum_{i=1}^{N}\left[\left(y - Q_1(s,a)\right)^2 + \left(y - Q_2(s,a)\right)^2\right] \tag{8}$$

where the two Q-functions are denoted by Q_1 and Q_2.

Actor update

Actor parameters are updated through gradient ascent with the first Q-function as (9)

$$\Delta_{\emptyset} J(\emptyset) = \frac{1}{N}\sum_{i=1}^{N} \nabla_a Q_1(s,a)|_{\,a=\Pi\emptyset(s)} \nabla_{\emptyset} \Pi_{\emptyset}(s) \tag{9}$$

The gradient $\Delta_{\emptyset} J(\emptyset)$ is computed by combining the gradient of the critic's Q-value with respect to the action a, and the gradient of the actor policy $\Pi_{\emptyset}(s)$ enabling the actor to optimize actions that maximize the expected Q-value.

Target network updates

The target networks are updated with Polyak averaging to provide seamless transitions as (10)

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta',\quad \emptyset' \leftarrow \tau\theta + (1-\tau)\emptyset' \tag{10}$$

The Polyak averaging is used to express target network updates for stable reinforcement learning training. The OPRTDPG algorithm minimizes power dispatch and market trade costs, maximizes energy efficiency, and stabilizes markets, exhibiting robustness in dealing with renewable energy uncertainties in dynamic settings.

Using the proposed OPRTDPG method resulted in markedly improved system performance, including smoother dispatch, lower operating cost, reduced price volatility, and higher renewable utilization. The combined OPRT and DPG mechanisms enabled stable learning and precise control, allowing the algorithm to deliver reliable, adaptive, and efficient optimization under dynamic and uncertain power system conditions. Algorithm .1 shows the process of the OPRTDPG model.

Algorithm 1: OPRTDPG-based dispatch and market trading optimization

Load dataset containing dispatch states, TCL variables, market prices, and renewable data.

2. **Compute feature bounds** $X_{\min}, X_{\max}$and apply Min–Max Normalization:
$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$
3. **Select high-impact features** using feature-importance results; optionally remove low-impact attributes.
4. **Initialize environment models**, including TCL thermal dynamics (refrigerator eqs. (2–4)) and water heater (eq. (5)).
5. **Simulate TCL aggregation** using Monte-Carlo population behavior:
$$P_x = P_{\text{nom}}^x \frac{\sum_{i=1}^{m} q_{x,i}}{m} \tag{6}$$
6. **Initialize actor network** $\mu_{\phi}(s)$and **dual critics** $Q_{\theta_1}(s,a)$, $Q_{\theta_2}(s,a)$.
7. **Create target networks** and set them equal to online networks: $\phi' \leftarrow \phi, \theta_1' \leftarrow \theta_1, \theta_2' \leftarrow \theta_2$.
8. **Initialize replay buffer** $R$and define hyperparameters: $\gamma, \tau, \sigma, C, N$,policy delay.
9. **Start training loop** across episodes.
10. **Reset environment** and retrieve initial normalized state s_0.
11. **Select action** using exploratory policy:
$$a_t = \mu_{\phi}(s_t) + \text{noise}$$
12. **Execute action in environment**, updating TCL thermal states using equations (2–5).
13. **Receive reward** r_t, next state s_{t+1}, and terminal flag.
14. **Store transition** $(s_t, a_t, r_t, s_{t+1},$done)in buffer R.
15. **If buffer size < batch size, continue** without learning.
16. **Sample minibatch** of Ntransitions from replay buffer.
17. **Apply Target Policy Smoothing** by adding clipped noise to target action:
$a' = \mu_{\phi'}(s') + \epsilon, \epsilon \sim \text{clip}(\mathcal{N}(0,\sigma), -C, C)$ (7)
18. **Compute target value using Clipped Double-Q**:
$y = r + \gamma(1 - done) \cdot \min\left(Q_{\theta_1'}(s',a'), Q_{\theta_2'}(s',a')\right)$

19. **Update critics** by minimizing loss:

$$L_{\text{critic}} = \frac{1}{N}\sum[(y - Q_{\theta_1}(s,a))^2 + (y - Q_{\theta_2}(s,a))^2] \quad (8)$$

20. **Perform gradient update** on both critic networks.
21. **Every policy delay step**, update actor:

$$\nabla_\phi J(\phi) = \frac{1}{N}\sum \nabla_a Q_{\theta_1}(s,a)\,|_{a=\mu_\phi(s)} \cdot \nabla_\phi \mu_\phi(s) \quad (9)$$

22. **Update actor parameters via gradient ascent.**
23. **Update target networks using Polyak averaging:**

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta', \phi' \leftarrow \tau\phi + (1-\tau)\phi' \quad (10)$$

24. **Repeat steps 11–23** until episode ends and learning stabilizes.
25. **Return final trained OPRTDPG controller**, optimized for dispatch, TCL control, storage management, and market bidding.

4 Result and discussion

The proposed OPRTDPG method was evaluated using P2P efficiency, profit improvement, SOC optimization, peak-load reduction, MAPE, training time, decision-making time, and overall performance improvement, while baseline results used market price fluctuation, energy efficiency, and system operating cost. The dataset was split into 70% training, 15% validation, and 15% testing to ensure reliable evaluation and prevent overfitting. Table 2 shows the simulation environment, data split, reinforcement-learning parameters, and hardware configuration used to implement and evaluate the proposed OPRTDPG method. Table 3 shows the clearly presents the improvements of the proposed OPRTDPG over the DDPG baseline across multiple standard metrics.

Table 2: Simulation environment, data setup, and OPRTDPG model parameters

Category	Details
Environment	Python 3.10; TensorFlow/PyTorch; custom power-system simulator (no OpenDSS/GridLAB-D); Kaggle dispatch & trading dataset.
Data Split	70% training, 15% validation, 15% testing.
Training Setup	3000 episodes, 5 epochs/episode, batch size 64, replay buffer 100k.
RL Parameters	Actor LR 0.0001; Critic LR 0.0002; γ = 0.99; Polyak τ = 0.005; Gaussian noise σ = 0.1; target update every 2 steps.
System Settings	15-min interval data; aggregated TCL models; SOC 20–98%; continuous action space with dispatch + trading states.
Hardware	NVIDIA RTX 2080 Ti GPU; Intel i7-9700K CPU; 32 GB RAM.

Table 3: Comparison of standard metrics between DDPG and OPRTDPG

Metric	Baseline (DDPG)	Proposed (OPRTDPG)
Cumulative Reward	12,500	15,300
Convergence Rate (episodes)	2,500	1,800
Average Cost per Episode ($)	100	88
Dispatch Error (%)	6.5	4.2
Load Satisfaction Rate (%)	91	97

4.1 Experimental setup

The experimental setup description has been expanded as recommended. The OPRTDPG algorithm was implemented in Python 3.11.4 using a custom RL simulation environment developed in NumPy and PyTorch, with no external real-time platforms such as OpenDSS or GridLAB-D. The environment models power balance, TCL thermal dynamics, ESS constraints, and market behavior. Beyond hardware specs, computational cost is now reported using evaluation metrics including training time, decision-making time, convergence rate, and scalability under increasing state dimensionality.

Energy generation

Energy generation refers to manufacturing usable electrical power from renewable sources (solar, wind, hydro) and predictable systems. Accurately modeling total energy generation (E_g), as shown in (11), supports optimizing power system dispatch and market trading to improve system dependability and economic efficacy within renewable unpredictability.

$$E_g = \sum_{i=1}^{N} P_i \cdot t_i \quad (11)$$

Optimizing $E_g$provides system stability and effective resource use. Figure 4 shows the line graph of Predicted Renewable Energy Generation Over 24 Hours, which illustrates hourly energy generation (MWh) from solar, wind, hydro, and biomass sources. The X-axis shows time (0–24 hours), the Y-axis shows generation (0–11.5 MWh), with wind peaking at hour 6 and solar at hour 12.

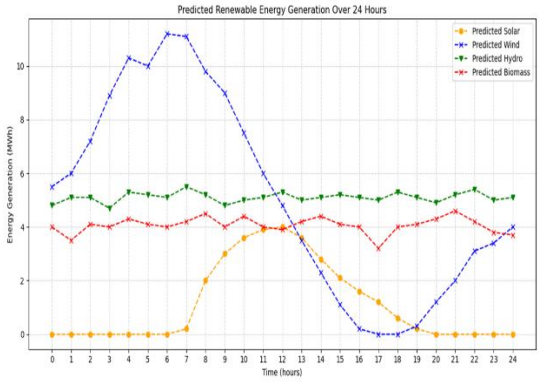

Figure 4: Predicted energy generation

The energy is generated from several sources over 24 hours. Early wind energy peaks at 15 MWh, whereas solar energy peaks at 10 MWh at noon. With averages of 4 MWh for biomass and 5 MWh for hydropower, resource distribution patterns are observable. These trends offer insightful information to improve power system dispatch tactics, guaranteeing effective resource distribution and increased system dependability in the face of renewable energy fluctuation.

Decision-making time (s)

Decision-making time is the amount of time needed to decide on the best course of action in dynamic situations. It is evaluated using the OPRTDPG technique to improve market trading and power system dispatch, guaranteeing dependability in the face of renewable energy fluctuation defined by (12).

$$T_d = \frac{Total\ decision\ computation\ time}{Number\ of\ decisions\ made} \quad (12)$$

Where T_d improves power systems' operating efficiency and real-time flexibility. Figure 5 shows the Decision-Making Time for OPRTDPG (High Time), the system's decision-making time over operational seconds. The X-axis represents Time (0.10–0.50 s), and the Y-axis shows Decision-Making Time (0.40–0.50 s). Data points range from 0.40 s to 0.50 s, indicating slight variation over time.

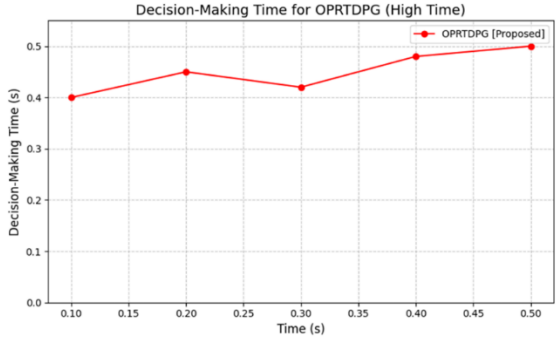

Figure 5: Decision-making time across time interval

The OPRTDPG algorithm's decision-making time at various time intervals is displayed in Figure 5. It records 0.41 seconds at 0.10 seconds, 0.45 seconds at 0.20 seconds, 0.43 seconds at 0.30 seconds, 0.47 seconds at 0.40 seconds, and 0.50 seconds at 0.50 seconds, with a decision time of 0.49 seconds. The OPRTDPG algorithm supports effective real-time dispatch and market optimization with consistent decision-making times ranging from 0.41 to 0.50 seconds.

Training time (s)

The OPRTDPG algorithm's learning time for optimum policies is measured by the training time. In an effort to support real-time power dispatch and market trading choices that improve system stability, dependability, and economic efficiency, effective training facilitates quick response to renewable variability (13).

$$T_{train} = \sum_{e=1}^{E} t_e \quad (13)$$

Where T_{train} ensures faster deployment and improves algorithm performance for power system responsibilities that are optimized. Figure 6 shows the OPRTDPG algorithm's training time.

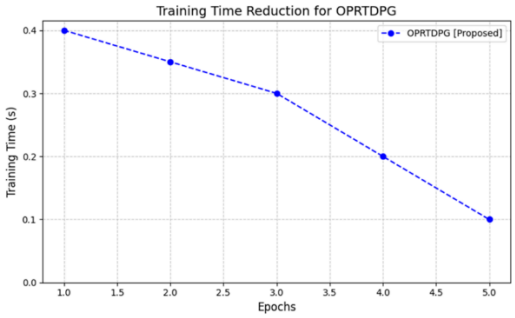

Figure 6: Training time trend over multiple epochs

A line graph visualizes Training Time Reduction for OPRTDPG, connecting data points across epochs. The X-axis represents Epochs (1–5), and the Y-axis shows Training Time (s) (0–0.4), with values decreasing from ~0.4 s to ~0.1 s over five epochs shown in Figure 6, where times drop from 0.39 seconds in epoch 1 to 0.10 seconds in epoch 5. This pattern shows improved optimization, enabling quicker and more efficient algorithm implementation. Training time for the OPRTDPG technique decreases from 0.39 to 0.10 seconds across five epochs, indicating increased effectiveness for real-time dispatch optimization.

Evaluation includes P2P efficiency (96%), profit improvement (28.4%), SOC optimization (20–98%), peak-load reduction (27.6%), MAPE (1.3%), training time (0.20 s), decision-making time (0.45 s), and a performance improvement rate of 48.2% as shown in Table 4 & Table 5, Outperforming existing approaches like LSTM [21], DQN [22], SARSA [22], and D3QN [22] approaches, the proposed OPRTDPG achieves higher accuracy, faster computation, superior stability, and better operational efficiency. Fig. 7 shows the key performance metrics of the

proposed OPRTDPG method, and the corresponding visual interpretation of these results is provided in the adjoining figure for clear comparison and understanding.

Table 4: Performance comparison of OPRTDPG with LSTM across key operational metrics.

Metrics (%)	Methods	
	LSTM [21]	OPRTDPG [proposed]
P2P Efficiency	91	**96**
Profit Improvement	15.6	**28.4**
SOC Optimization	30-95	**20-98**
Peak Load Reduction	18.4	**27.6**
MAPE	2.5	**1.3**

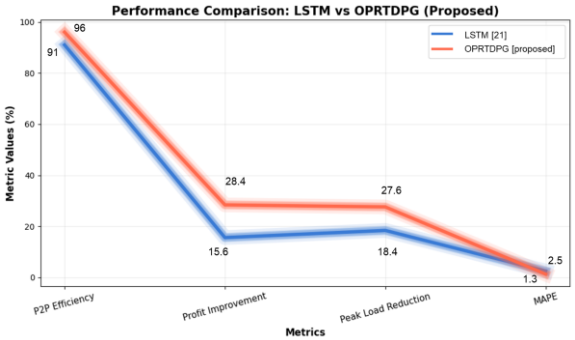

Figure 7: Performance comparison of LSTM and the proposed

Table 5: Training, Decision-making, and Rate comparison of OPRTDPG and existing methods

Metrics	Methods				
	LSTM [21]	DQN [22]	SARSA [22]	D3QN [22]	OPRTDPG [proposed]
Training Time (s)	180	196.0111	415.5845	244.1469	**0.20**
Decision-Making Time (s)	-	0.347	0.354	0.390	**0.45**
Performance Improvement Rate	-	-	30.5 %	43.93 %	**48.2%**

To evaluate the OPRTDPG model, three operational metrics were analyzed: market price fluctuation, energy efficiency, and system operating cost, reflecting dispatch stability, resource utilization, and economic performance under renewable uncertainty. For baseline comparison, OPRTDPG was benchmarked against traditional methods, including rule-based and standard RL algorithms (DDPG), demonstrating notable improvements in cost reduction, energy use, and price stability. Power system operations become more reliable, economical, and sustainable as a result of the algorithm's capacity to include these elements.

Market price fluctuation:

Market price fluctuation quantifies electricity price variability caused by supply-demand imbalances. This research evaluates how effectively OPRTDPG stabilizes dispatch decisions under renewable uncertainty. Lower fluctuations reflect improved scheduling, secure trading behavior, and enhanced market reliability.

Energy efficiency:

Energy efficiency measures how effectively renewable, storage, and flexible load resources are converted into useful system output. Higher efficiency in this study demonstrates OPRTDPG's ability to coordinate resources intelligently, minimize energy waste, and sustain operations under variable renewable generation conditions.

System Operating Cost:

System operating cost represents the total economic expenditure for generation dispatch, storage management, and market transactions. Reduced costs indicate OPRTDPG's capability to make economically optimal decisions, lower reliance on expensive generation, and improve overall system profitability and market performance.

Table 6: Baseline performance comparison of DDPG and OPRTDPG

Metric	Baseline (DDPG)	Proposed (OPRTDPG)	Improvement	95% CI	p-value
Market Price Fluctuation (MWh)	$50 ± 2	$45 ± 1.5	↓ 10%	[44.1 – 46.2]	p < 0.01
Energy Efficiency (%)	65 ± 1.9	74.75 ± 1.4	↑ 15%	[73.4 – 75.9]	p < 0.01
System Operating Cost ($/day)	$100,000 ± 4,900	$88,000 ± 3,600	↓ 12%	[$86,200 – $89,500]	p < 0.01

Table 6 compares the baseline DDPG method with the proposed OPRTDPG algorithm across three key system metrics: market price fluctuation (MWh), energy efficiency (%), and system operating cost ($/day). OPRTDPG reduces market price volatility by 10% ($50 → $45), increases energy efficiency by 15% (65% → 74.75%), and lowers system operating cost by 12% ($100,000/day → $88,000/day), demonstrating improved dispatch stability, resource coordination, and economic performance under renewable uncertainty. The 95% confidence intervals and p-values confirm the statistical significance of all improvements. Market price fluctuation, energy efficiency, and system operating cost are key metrics for evaluating OPRTDPG. Compared to baseline DDPG, OPRTDPG reduces price volatility by 10%, improves efficiency by 15%, and lowers operating cost by 12%, confirming enhanced dispatch stability, resource coordination, and economic performance.

Discussion

The current methods, such as LSTM [21], DQN [22], SARSA [22], and D3QN [22], also proved to have weaknesses regarding power system dispatch and market trading under renewable energy uncertainty. LSTM struggled with limited coordination of renewable, storage, and load-responsive actions and moderate decision-making performance. DQN, SARSA, and D3QN required long training times, slower decision-making, and were sensitive to hyperparameters. These limitations resulted in suboptimal energy utilization, lower economic efficiency, and unstable market operations.

The proposed OPRTDPG method overcomes these shortcomings by integrating a dual-control strategy for TCLs and price-responsive loads with an advanced twin deterministic policy gradient algorithm. OPRTDPG adapts to real-time renewable fluctuations and market conditions, improving stability, operational efficiency, and economic performance. Key results include: market price fluctuation reduced by 10% ($50/MWh → $45/MWh), energy efficiency increased by 15% (65% → 74.75%), system operating cost decreased by 12% ($100,000/day → $88,000/day), P2P efficiency improved to 96%, profit improvement to 28.4%, SOC optimization to 20–98%, peak load reduction to 27.6%, and MAPE reduced to 1.3%. For comparison, the baseline DDPG metrics were: market price fluctuation of $50/MWh, energy efficiency of 65%, and system operating cost of $100,000/day, showing that OPRTDPG achieves significant improvements. The proposed OPRTDPG method demonstrates robustness by maintaining stable dispatch and market decisions under high renewable variability and sudden load changes. Improvements arise from dual TCL and price-responsive load control, Clipped Double-Q learning, delayed policy updates, and target policy smoothing, optimizing efficiency, cost, and market stability within realistic operational constraints.

Although effective, OPRTDPG has limitations: decision-making time is slightly higher than simpler models, the algorithm is complex and requires proper hyperparameter tuning, and large-scale implementation demands sufficient computational resources for real-world deployment.

5 Conclusion

The increasing integration of renewable energy introduces uncertainty that challenges power system dispatch and market trading. To address this, the study aimed to optimize dispatch and trading decisions while maintaining economic efficiency and system stability. The Power System Dispatch and Market Trading dataset was used, with Min–Max normalization applied to improve training performance. A dual-control strategy managed both Thermostatically Controlled Loads (TCLs) and price-responsive loads, enabling adaptive coordination of demand. The proposed OPRTDPG method, combining Optimal Power Reinforcement (OPRT) for dynamic adaptation and Twin Deterministic Policy Gradient (DPG) for continuous action optimization, was employed to learn real-time optimal policies. The evaluation showed that, compared to baseline DDPG metrics—market price fluctuation $50/MWh, energy efficiency 65%, and system operating cost $100,000/day—the OPRTDPG agent reduced price volatility by 10%, improved energy efficiency by 15%, and decreased operating costs by 12%, demonstrating superior adaptability to market and renewable variability.

6 Limitations and Future Scopes

OPRTDPG requires high computational resources for stable learning and depends heavily on dataset realism, limiting generalizability across diverse grid conditions and affecting performance under unseen renewable or market fluctuations. Future work would apply model-based RL, ensemble DRL, and online adaptive learning to improve stability and real-time performance, while extending the framework to multi-agent coordination and larger real-world grid datasets.

Declarations

Ethics approval and consent to participate: I confirm that all the research meets ethical guidelines and adheres to the legal requirements of the study country.

Consent for publication: I confirm that any participants (or their guardians if unable to give informed consent, or next of kin, if deceased) who may be identifiable through the manuscript (such as a case report) have been allowed to review the final manuscript and have provided written consent to publish.

Availability of data and materials: The data used to support the findings of this study are available from the corresponding author upon request.

Authors' contributions (Individual contribution): All authors contributed to the study conception and design. All authors read and approved the final manuscript.

References

[1] Harrold DJ, Cao J, Fan Z (2022) Renewable energy integration and microgrid energy trading using multi-agent deep reinforcement learning. Appl Energy 318:119151. https://doi.org/10.1016/j.apenergy.2022.119151

[2] Kiptoo MK, Adewuyi OB, Furukakoi M, Mandal P, Senjyu T (2023) Integrated multi-criteria planning for resilient renewable energy-based microgrid considering advanced demand response and uncertainty. Energies 16(19):6838. https://doi.org/10.3390/en16196838

[3] Huang Q, Xian H, Mei L, Cheng X, Li N (2025) Intelligent distribution network operation and anomaly detection based on information technology. Informatica 49(9). https://doi.org/10.31449/inf.v49i9.5584

[4] Cui S, Tian J (2024) Analysis and calculation of marginal electricity price of nodes with network loss from the perspective of intelligent robot considering digital signal processing technology. Informatica 48(14).https://doi.org/10.31449/inf.v48i14.6066

[5] Lin FJ, Chang CF, Huang YC, Su TM (2023) A deep reinforcement learning method for economic power dispatch of microgrid in OPAL-RT environment. Technologies 11(4):96. https://doi.org/10.3390/technologies11040096

[6] Hu J, Cao J (2021) Demand response optimal dispatch and control of tcl and pev agents with renewable energies. Fractal and Fractional, 5(4), 140. https://doi.org/10.3390/fractalfract5040140

[7] Ma Q, Liu B, Li J (2025) A Trading Model for the Electricity Spot Market That Takes into Account the Preference for Energy Storage Trading. Energies 18(9):2322. https://doi.org/10.3390/en18092322

[8] Kumar M, Namrata K, Samadhiya A (2025) Deep learning assisted optimal dispatch for renewable-based energy system considering consumer incentive scheme. Cluster Computing 28(4):262. https://doi.org/10.1007/s10586-024-04938-x

[9] Jain R, Mahajan V (2023) Efficient energy management and reliability assessment by optimal placement of renewable energy sources with pump storage plant. Smart Grids and Sustainable Energy 8(1):3. https://doi.org/10.1007/s40866-023-00160-7

[10] Montoya OD, Fuentes JE, Moya FD, Barrios JÁ, Chamorro HR (2021) Reduction of annual operational costs in power systems through the optimal siting and sizing of STATCOMs. Appl Sci 11(10):4634. https://doi.org/10.3390/app11104634

[11] Tang H, Wang S (2022) A model-based predictive dispatch strategy for unlocking and optimizing the building energy flexibilities of multiple resources in electricity markets of multiple services. Appl Energy 305:117889. https://doi.org/10.1016/j.apenergy.2021.117889

[12] Jiang T, Dong X, Zhang R, Li X (2023) Strategic active and reactive power scheduling of integrated community energy systems in day-ahead distribution electricity market. Appl Energy 336:120558. https://doi.org/10.1016/j.apenergy.2022.120558

[13] Kraft E, Russo M, Keles D, Bertsch V (2023) Stochastic optimization of trading strategies in sequential electricity markets. Eur J Oper Res 308(1):400-421. https://doi.org/10.1016/j.ejor.2022.10.040

[14] Pal P, Krishnamoorthy PA, Rukmani DK, Antony SJ, Ocheme S, Subramanian U, Hasanien HM (2021) Optimal dispatch strategy of virtual power plant for day-ahead market framework. Appl Sci 11(9):3814. https://doi.org/10.3390/app11093814

[15] Ding Y, Tan Q, Shan Z, Han J, Zhang Y (2023) A two-stage dispatching optimization strategy for hybrid renewable energy system with low-carbon and sustainability in ancillary service market. Renew Energy 207:647-659. https://doi.org/10.1016/j.renene.2023.03.050

[16] Liu J, Sun XY, Bo R, Wang S, Ou M (2022) Economic dispatch for electricity merchant with energy storage and wind plant: State of charge based decision making considering market impact and uncertainties. J Energy Storage 53:104816. https://doi.org/10.1016/j.est.2022.104816

[17] Biggar DR, Hesamzadeh MR (2022) An integrated theory of dispatch and hedging in wholesale electric power markets. Energy Econ 112:106055. https://doi.org/10.1016/j.eneco.2022.106055

[18] Ouyang T, Li Y, Xie S, Wang C, Mo C (2024) Low-carbon economic dispatch strategy for integrated power system based on the substitution effect of carbon tax and carbon trading. Energy 294:130960. https://doi.org/10.1016/j.energy.2024.130960

[19] Huang Y, Wang Y, Liu N (2022) Low-carbon economic dispatch and energy sharing method of multiple Integrated Energy Systems from the perspective of System of Systems. Energy 244:122717. https://doi.org/10.1016/j.energy.2021.122717

[20] Zhang S, Hu W, Du J, Bai C, Liu W, Chen Z (2023) Low-carbon optimal operation of distributed energy systems in the context of electricity supply restriction and carbon tax policy: A fully decentralized energy dispatch strategy. J Clean Prod 396:136511. https://doi.org/10.1016/j.jclepro.2023.136511

[21] Seif A (2025). Deep and Reinforcement Learning-Based EMS for Optimal P2p Energy Trading in Smart Grids. Available at SSRN 5244857. https://dx.doi.org/10.2139/ssrn.5244857

[22] Zhai S, Li W, Qiu Z, Zhang X, & Hou S (2023). An improved deep reinforcement learning method for dispatch optimization strategy of modern power systems. Entropy, 25(3), 546. https://doi.org/10.3390/e25030546