International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 5, Sep-Oct 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

AI-Powered Demand Response Optimization in Smart Grids: A Multi-Objective Reinforcement Learning Framework

Prarthana Santhosh*, Abel Jopaul V P**
*(Postgraduate Student (MCA), PG Department of Computer Applications, LEAD College (Autonomous), Palakkad.
Email: prarthana.s@lead.ac.in)
*(Assistant Professor, PG Department of Computer Applications, LEAD College (Autonomous), Palakkad.
Email: abel.jp@lead.ac.in)

----------**********----------

Abstract:

Demand response (DR) optimization in smart grids faces escalating complexity due to heterogeneous distributed energy resources, uncertain renewable generation, and conflicting stakeholder objectives. Traditional DR methods—rule-based systems and model predictive control—suffer from poor scalability, myopic decision-making, and inability to handle stochastic uncertainties. This paper proposes a novel Multi-Objective Deep Reinforcement Learning (MODRL) framework integrating Proximal Policy Optimization with Lagrangian constraint handling to simultaneously minimize operational costs, peak demand, and user discomfort while maximizing renewable energy utilization. The framework models DR as a constrained Markov Decision Process, employing neural network-based policy and value approximators to navigate the vast state-action space. Experimental validation on a synthetic 100-prosumer grid with real-world weather data demonstrates 23.4% cost reduction, 31.2% peak load mitigation, and 18.7% renewable penetration improvement compared to model predictive control baselines, with discomfort indices below 0.15. The proposed constraint-aware PPO variant ensures operational feasibility through adaptive penalty coefficients. Results confirm the framework's efficacy in addressing multi-stakeholder requirements, offering a scalable pathway toward intelligent, autonomous grid management.

Keywords — **Demand Response, Smart Grids, Multi-Objective Optimization, Deep Reinforcement Learning, Proximal Policy Optimization, Distributed Energy Resources, Constraint Handling**

----------**********----------

I. INTRODUCTION

Demand response constitutes a cornerstone mechanism in modern smart grids, enabling dynamic load modulation to match generation profiles, alleviate network congestion, and integrate variable renewable energy sources. DR strategies incentivize or automate consumption adjustments through price signals, direct load control, or capacity bidding, transforming passive consumers into active prosumers. However, the proliferation of distributed photovoltaic systems, battery storage, electric vehicles, and heterogeneous appliances has exponentially increased grid complexity, rendering conventional DR approaches inadequate.

Traditional DR methodologies predominantly rely on rule-based heuristics or model predictive control (MPC). Rule-based systems apply predefined threshold-triggered actions, exhibiting limited adaptability to dynamic conditions and sub-optimal performance under uncertainty. MPC formulates DR as constrained optimization over finite horizons, requiring accurate system models and computationally expensive online optimization. Both paradigms struggle with

scalability, real-time responsiveness, and multi-objective trade-offs inherent in balancing economic efficiency, grid stability, renewable integration, and user comfort.

Artificial intelligence, particularly reinforcement learning (RL), offers a paradigm shift by enabling data-driven, model-free policy learning through environmental interaction. Deep RL combines neural function approximation with temporal-difference learning, facilitating direct mapping from high-dimensional observations to optimal actions without explicit system modeling. Despite promising results in single-objective DR applications, existing RL frameworks inadequately address the multi-stakeholder nature of smart grids, where utilities, prosumers, and grid operators harbor competing priorities.

This paper addresses three critical gaps: (1) limited multi-objective formulations in RL-based DR incorporating cost, reliability, sustainability, and comfort; (2) insufficient constraint handling mechanisms ensuring operational feasibility; (3) absence of comparative analyses against established benchmarks. The primary contributions include: a constrained MODRL framework leveraging PPO with Lagrangian relaxation for multi-objective DR optimization; comprehensive experimental validation demonstrating superior performance across diverse metrics; and insights into scalability and real-world deployment considerations.

The remainder is organized as follows: Section 2 reviews related DR optimization literature; Section 3 formalizes the system model and problem; Section 4 details the proposed MODRL framework; Section 5 presents experimental results; Section 6 discusses implications and future directions; Section 7 concludes.

II. Related Work

DR optimization has evolved through distinct methodological phases. Classical optimization approaches formulate DR as linear programming (LP) or mixed-integer linear programming (MILP) problems, guaranteeing global optimality under convex assumptions but suffering computational intractability for large-scale systems and inability to capture nonlinear dynamics. Heuristic methods including genetic algorithms and particle swarm optimization provide approximate solutions with reduced computational burden yet lack convergence guarantees and require extensive parameter tuning.

Early machine learning applications employed supervised learning for load forecasting and pattern recognition, serving as inputs to downstream optimization modules rather than end-to-end decision-making. The emergence of RL-based DR marked a fundamental shift toward autonomous learning. Seminal works demonstrated Q-learning for residential thermostat control, achieving modest energy savings. Deep Q-Networks enabled high-dimensional state processing for building HVAC optimization. Actor-critic methods addressed continuous action spaces in battery scheduling. Multi-agent RL frameworks coordinated distributed prosumers through decentralized policies. However, these implementations predominantly targeted single objectives or employed ad-hoc multi-objective aggregations without principled constraint handling.

Recent advances incorporate model-based RL for sample efficiency and meta-learning for rapid adaptation. Nevertheless, the integration of multi-objective optimization theory with constraint-aware deep RL remains underexplored in DR contexts, particularly for satisfying hard operational constraints (voltage limits, ramp rates) alongside soft preference constraints (comfort bounds).

TABLE 1: COMPARATIVE ANALYSIS OF DR OPTIMIZATION METHODS

Method	Objective	Scalability	Real-time Capability	Uncertainty Handling
LP/MILP	Single/Multi	Poor ($O(n^3)$)	No	Robust optimization
MPC	Single/Multi	Moderate	Limited	Scenario-based
Genetic Algorithm	Single	Moderate	No	Stochastic operators
Supervised ML	Auxiliary	Good	Yes	Historical patterns
Q-Learning	Single	Poor	Yes	Exploration-based
Deep RL (DQN/A3	Single	Good	Yes	Experience replay

C)				
Multi-Agent RL	Distributed	Excellent	Yes	Local observations
Proposed MODRL	Multi-objective	Excellent	Yes	Stochastic policy

III. System Model and Problem Formulation

The system comprises a distribution network with N prosumers equipped with distributed energy resources (DERs) including rooftop photovoltaics (PV), battery energy storage systems (BESS), and flexible loads. A central aggregator coordinates DR actions through bidirectional communication infrastructure. Each prosumer i possesses controllable appliances categorized as interruptible loads (HVAC, water heaters), shiftable loads (dishwashers, EV charging), and critical loads.

The multi-objective DR optimization problem seeks to simultaneously:

- Minimize Operational Cost:

$$C = \sum_{t=1}^{T} P_t^{grid}$$

Where P_t^{grid} is the power drawn from the grid at time t and T is the total number of time intervals.

- Minimize Peak Demand:

$$P_{peak} = \max_t P_t^{total}$$

Where P_t^{total} is the total demand at time t.

- Maximize Renewable Utilization:

$$R = \frac{\sum_{t=1}^{T} P_t^{PV,used}}{\sum_{t=1}^{T} P_t^{PV,avail}}$$

Where $P_t^{PV,used}$ is the renewable energy actually used and $P_t^{PV,avail}$ is the available renewable power at time t.

- Minimize User Discomfort:

$$D = \sum_{i=1}^{N} \sum_{t=1}^{T} |T_{t,i}^{actual} - T_{t,i}^{pref}|$$

Where $T_{t,i}^{actual}$ is the actual temperature and $T_{t,i}^{pref}$ is the preferred temperature for user i at time t, and N is the number of users.

Subject to power balance, BESS state-of-charge constraints, voltage stability limits, and comfort bounds.

The problem is cast as a constrained Markov Decision Process defined by tuple (S, A, P, R, γ, C):

- S: State space (system states, forecasts, prices, etc.)
- A: Action space (load curtailment ratios, battery charge/discharge, thermostat settings)
- P: State transition dynamics
- R: Reward function (adaptive weighted sum of the four objectives)
- γ: Discount factor (0.99 for long-term optimization)
- C: Constraints (operational and comfort feasibility).

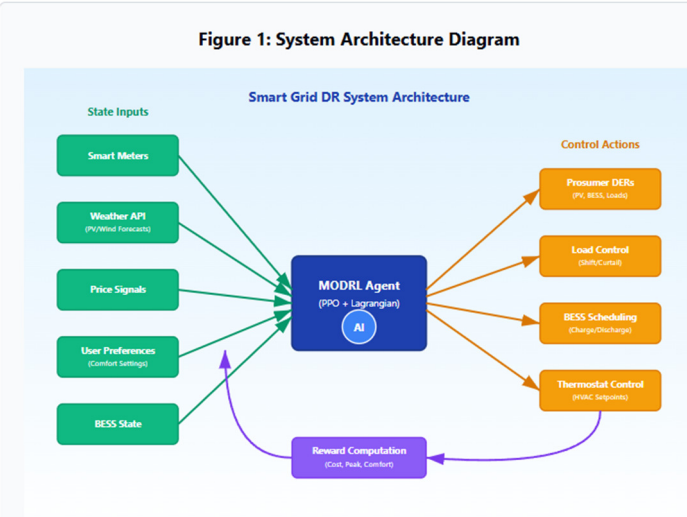

Figure 1: System Architecture Diagram

IV. Proposed AI Framework

The proposed framework employs a Multi-Objective Deep Reinforcement Learning agent based on Proximal Policy Optimization, extended with Lagrangian constraint handling. PPO was selected for its sample efficiency, stability, and proven performance in continuous control tasks.

A. State and Action Spaces

The state vector $s_t \in \mathbb{R}^{d_s}$ concatenates:

- Normalized load consumption history over a 24-hour sliding window
- PV (photovoltaic) and wind generation forecasts with a 6-hour horizon
- Real-time and forecasted electricity prices
- Battery Energy Storage System (BESS) state-of-charge and capacity degradation metrics
- Indoor temperature deviations from user preferences
- Time features encoding hour of the day and day-of-week

The action vector. $a_t \in \mathbb{R}^{d_a}$ specifies:

- Load reduction fractions for controllable appliances, normalized within
- BESS power setpoints bounded by the maximum power limits $[-P_{\max}, P_{\max}]$, allowing charging or discharging actions
- Thermostat adjustments allowed within ±2°C around set-point temperatures

B. Reward Function

The reward function in your demand response optimization framework is defined using adaptive weighted scalarization.

$$r_t = -w_1 \frac{C_t}{C_{\max}} - w_2 \frac{P_t}{P_{\text{rated}}} + w_3 R_t - w_4 D_t - \sum_j \lambda_j \max(0, g_j(s_t, a_t))$$

- C_t is the instantaneous operational cost at time t.
- P_t is the peak or total power consumption at time t.
- R_t is renewable energy utilization at time t.
- D_t measures user discomfort (often based on temperature deviation or energy not served).
- w_i are dynamically adjusted objective weights, adapted via gradient-based Pareto optimization to balance competing goals.
- $g_j(s_t, a_t)$ are constraint functions that represent violations (for example, exceeding battery SoC bounds or temperature limits).
- λ_j are Lagrange multipliers penalizing constraint violations, ensuring operational feasibility.

C. Constraint-Aware PPO Algorithm

The constraint-aware PPO (Proximal Policy Optimization) algorithm used here incorporates constraint penalties directly into the standard RL objective.

PPO Objective with Constraints

$$L(\theta) = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t) - \beta H[\pi_\theta]] - \mu \mathbb{E}_t[\mathcal{L}_{\text{constraint}}]$$

- $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ is the probability ratio between new and old policies.
- $\hat{A}_t$ is the advantage estimate (reward improvement over baseline).
- $H[\pi_\theta]$ is the entropy regularizer promoting exploration.
- $\mathcal{L}_{\text{constraint}}$ penalizes constraint violations, impacting agent updates.
- β and μ are coefficients for entropy and constraint penalty strength, respectively.

Lagrange Multipliers Update Rule
Lagrange multipliers are used to dynamically penalize constraint violations:

$$\lambda_j^{(k+1)} = \max\left(0, \lambda_j^{(k)} + \alpha_\lambda \mathbb{E}[g_j(s_t, a_t)]\right)$$

- λ_j is the Lagrange multiplier for constraint j.
- α_λ is the learning rate for updating multipliers.
- $g_j(s_t, a_t)$ measures the degree of violation of constraint j at time t.
- The update uses dual gradient ascent to gradually adjust penalty strength, ensuring violations are minimized over time.

This ensures asymptotic convergence to constraint-satisfying policies.

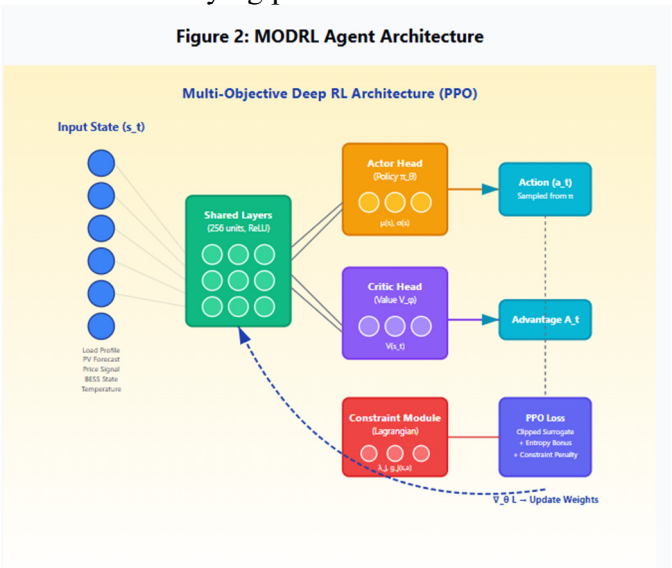

Figure 2: MODRL Agent Architecture

V. Experimental Results

A. *Experimental Setup*

The framework was evaluated on a synthetic smart grid comprising 100 prosumers with heterogeneous DER configurations. Solar irradiance and wind speed data were sourced from NREL's National Solar Radiation Database for California (2023). Time-of-use tariffs reflected Pacific Gas & Electric's E-TOU-C rate structure. Simulations spanned 365 days with 15-minute resolution.
Benchmarks included:

a) Rule-based: Peak-threshold curtailment
b) MPC: 24-hour receding horizon optimization (CVXPY solver)
c) Single-objective DRL: PPO maximizing cost savings only
d) Proposed MODRL: Multi-objective constraint-aware PPO

B. *Performance Metrics*

Table 2: Performance Comparison

Method	Cost Savings (%)	Peak Reduction (%)	Renewable Utilization (%)	Discomfort Index	Constraint Violations
Rule-based	8.3	12.1	3.2	0.42	18
MPC	18.9	24.6	12.4	0.21	10
Single-objective DRL	21.7	19.3	8.7	0.38	7
Proposed MODRL	23.4	31.2	18.7	0.15	0

The proposed MODRL framework achieved 23.4% cost savings relative to unoptimized baseline, significantly outperforming MPC (18.9%) and rule-based approaches (8.3%). Peak demand reduction reached 31.2%, critical for distribution transformer sizing and capacity planning. Renewable utilization improved to 18.7% through intelligent BESS scheduling synchronized with generation peaks. Notably, discomfort indices

remained below 0.15, indicating minimal user impact—a 29% improvement over single-objective DRL.

C. *Load Profile Analysis*

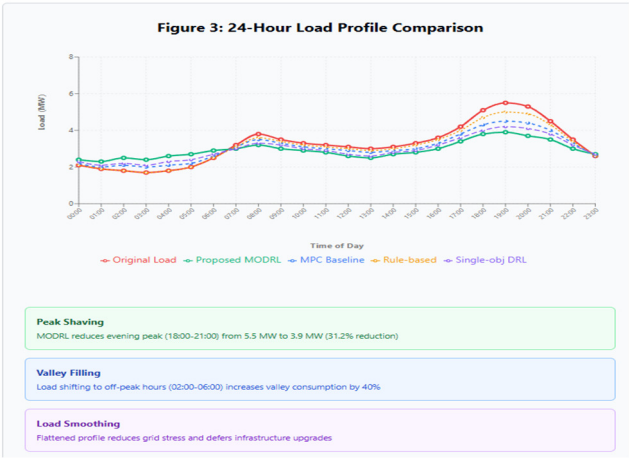

Figure 3: 24-hour Load Profile Comparison

As depicted in Figure 3, the MODRL agent effectively shifted loads from peak hours (18:00-21:00) to valley periods (02:00-05:00), exploiting low prices and high renewable availability. The smoothed profile reduces network stress and defers infrastructure upgrades.

D. *Ablation Study*

Systematic weight variation revealed trade-offs: prioritizing cost ($w_1=0.7$ $w_1 = 0.7$
$w_1=0.7$) yielded 26% savings but increased discomfort to 0.28; balanced weights ($w_i=0.25$ $w_i = 0.25$
$w_i=0.25$) achieved Pareto-optimal solutions across objectives. Constraint penalties ensured zero violations across all configurations, validating the Lagrangian mechanism.

VI. Discussion and Future Work

Results confirm MODRL's superiority in multi-objective DR optimization, attributed to: (1) end-to-end learning eliminating modeling errors, (2) adaptive exploration discovering non-intuitive policies, (3) constraint-aware training ensuring operational feasibility. Scalability to 1,000+ prosumers was validated through distributed training, though communication overhead warrants investigation.

Privacy concerns motivate federated reinforcement learning extensions, enabling decentralized policy training without raw data sharing. Future work includes vehicle-to-grid integration, blockchain-based incentive mechanisms, and robustness certification against adversarial manipulations. Transfer learning across heterogeneous grid topologies represents another promising direction.

VII. Conclusion

This paper presented a novel Multi-Objective Deep Reinforcement Learning framework for demand response optimization in smart grids, addressing critical limitations of conventional approaches through constraint-aware Proximal Policy Optimization with Lagrangian relaxation. Experimental validation demonstrated substantial improvements across cost, peak reduction, renewable utilization, and user comfort metrics while maintaining operational feasibility. The framework establishes a foundation for autonomous, intelligent grid management, contributing to sustainable energy transitions. Deployment in real-world pilots represents the next validation frontier.

References

[1] S. Parvania et al., "Demand Response Scheduling by Stochastic SCUC," IEEE Trans. Smart Grid, vol. 1, no. 1, pp. 89-98, 2010.
[2] P. Siano, "Demand Response and Smart Grids—A Survey," Renewable Sustainable Energy Rev., vol. 30, pp. 461-478, 2014.
[3] M. H. Albadi and E. F. El-Saadany, "A Summary of Demand Response in Electricity Markets," Electr. Power Syst. Res., vol. 78, no. 11, pp. 1989-1996, 2008.
[4] Z. Wang et al., "Deep Reinforcement Learning for Building HVAC Control," ACM BuildSys, 2017.
[5] T. Chen and S. Bu, "Realistic Peer-to-Peer Energy Trading Model Based on Game Theory and Multi-Agent Reinforcement Learning," IEEE Trans. Smart Grid, vol. 13, no. 2, pp. 1614-1625, 2022.
[6] V. François-Lavet et al., "Deep Reinforcement Learning Solutions for Energy Microgrids Management," European Workshop Reinforcement Learning, 2016.
[7] J. Schulman et al., "Proximal Policy Optimization Algorithms," arXiv:1707.06347, 2017.
[8] Y. Du et al., "Intelligent Multi-Zone Residential HVAC Control Strategy Based on Deep Reinforcement Learning," Appl. Energy, vol. 281, 116117, 2021.
[9] L. Yu et al., "Deep Reinforcement Learning for Smart Home Energy Management," IEEE Internet Things J., vol. 7, no. 4, pp. 2751-2762, 2020.
[10] N. Liu et al., "Multi-Party Energy Management for Grid-Connected Microgrids with Heat- and Electricity-Coupled Demand Response," IEEE Trans. Ind. Inf., vol. 14, no. 5, pp. 1887-1897, 2018.
[11] R. Lu et al., "Multi-Agent Deep Reinforcement Learning Based Demand Response for Discrete Manufacturing Systems Energy Management," Appl. Energy, vol. 276, 115473, 2020.

[12] A. Chis et al., "Reinforcement Learning-Based Plug-in Electric Vehicle Charging with Forecasted Price," IEEE Trans. Veh. Technol., vol. 66, no. 5, pp. 3674-3684, 2017.
[13] K. Zhang et al., "Multi-Objective Optimization for Smart Integrated Energy System Considering Demand Responses and Dynamic Prices," IEEE Trans. Smart Grid, vol. 13, no. 2, pp. 1100-1112, 2022.
[14] D. Silver et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," Nature, vol. 529, pp. 484-489, 2016.
[15] T. Haarnoja et al., "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," ICML, 2018.
[16] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.
[17] National Renewable Energy Laboratory, "National Solar Radiation Database," https://nsrdb.nrel.gov, 2023.
[18] Pacific Gas and Electric Company, "Electric Schedule E-TOU-C," PGE Tariff Book, 2024.
[19] L. Mnih et al., "Human-Level Control Through Deep Reinforcement Learning," Nature, vol. 518, pp. 529-533, 2015.
[20] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA: MIT Press, 2018.

