International Journal of Computer Science Engineering Techniques (**IJCSE**) Volume 10 Issue 2 ,Mar-Apr 2026

ISSN : 2455-135X © 2025 International Scientific and Academic Research (ISAR) Publisher

A Climate-Aware Multi-Agent Reinforcement Learning System for Smart Grid Energy Management in India

Ishita Bahamnia
Department of Computer Science and Engineering,Maharshi Dayanand University,Rohtak
bahamnia.ishita@yahoo.com

Mr.Yogesh Kumar
Department of Computer Science and Engineering,Maharshi Dayanand University,Rohtak
dryogeshkumar.uiet@mdurohtak.ac.in

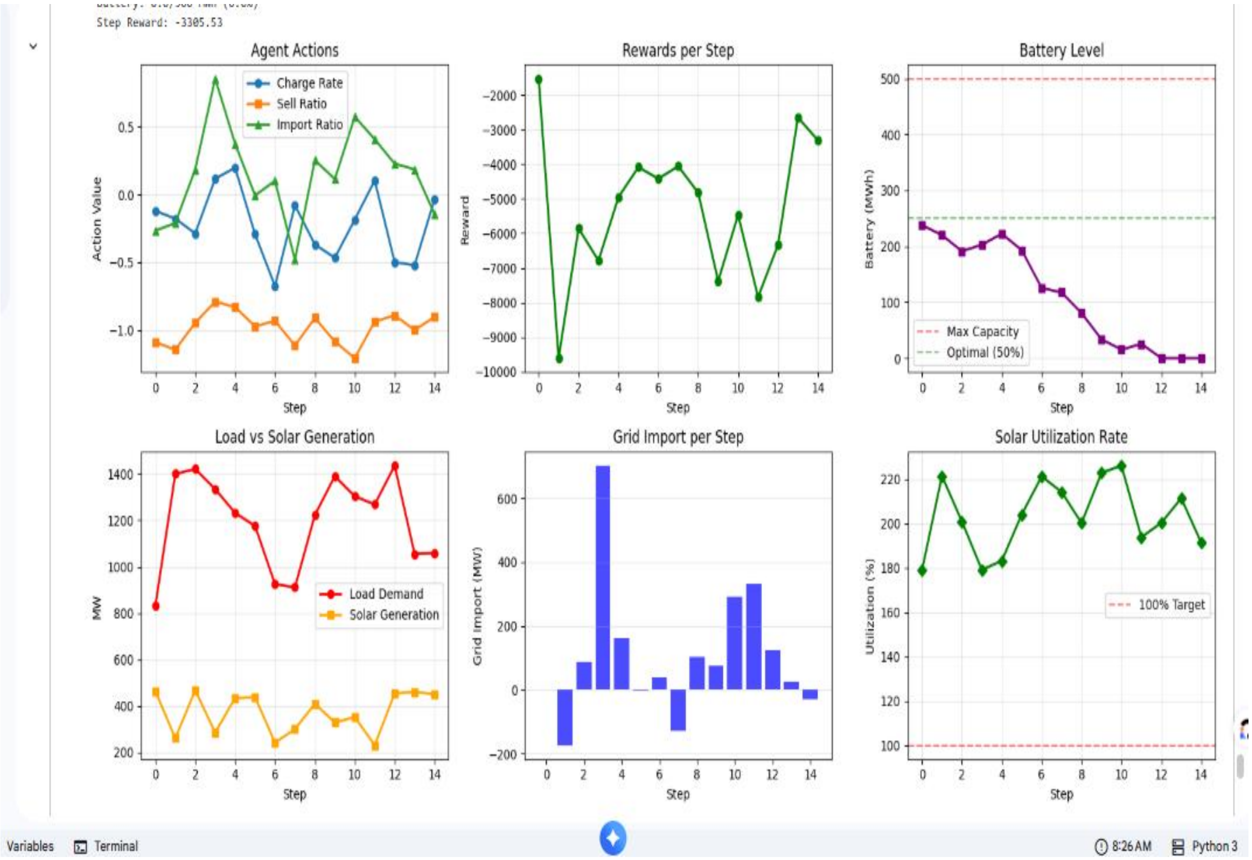
Abstract

The integration of volatile renewable energy sources into modern power grids presents significant challenges for maintaining stability and efficiency. This paper presents a climate-aware multi-agent reinforcement learning (MARL) system for smart grid energy management, specifically designed for the diverse regional grids of India. The proposed framework integrates an LSTM-based forecasting module for predicting short-term renewable generation with a custom MARL environment where agents learn collaborative policies for grid optimization. We evaluate the system using a realistic synthetic dataset generated from the energy profiles of four key Indian regions: Tamil Nadu, Odisha, Rajasthan, and Bihar. Experimental results demonstrate that the MARL framework successfully learns distinct, region-specific operating policies, achieving stable global grid balance through coordinated thermal generation management. The integration of climate forecasts enables proactive rather than reactive control strategies, significantly enhancing grid resilience. This work provides a foundational framework for developing adaptive, resilient energy management systems for renewable-integrated power grids in developing economies.

Keywords: Multi-Agent Reinforcement Learning, Smart Grid, Energy Management, LSTM, Renewable Energy Forecasting, India, Proximal Policy Optimization, Grid Stability.

Graphical Abstract

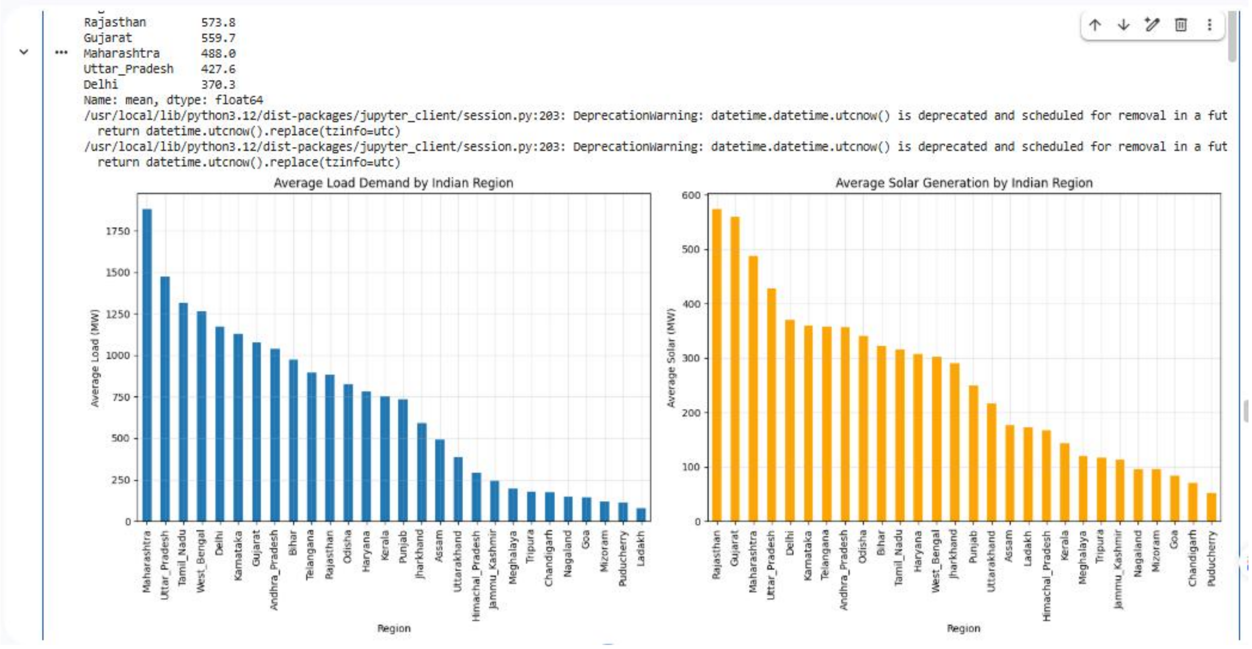
Figure 1: Graphical abstract illustrating the proposed climate-aware MARL framework. The system comprises: (1) a synthetic data generation module creating regional energy profiles for TamilNadu, Odisha, Rajasthan, and Bihar;



```
Step Reward: -3305.53
```

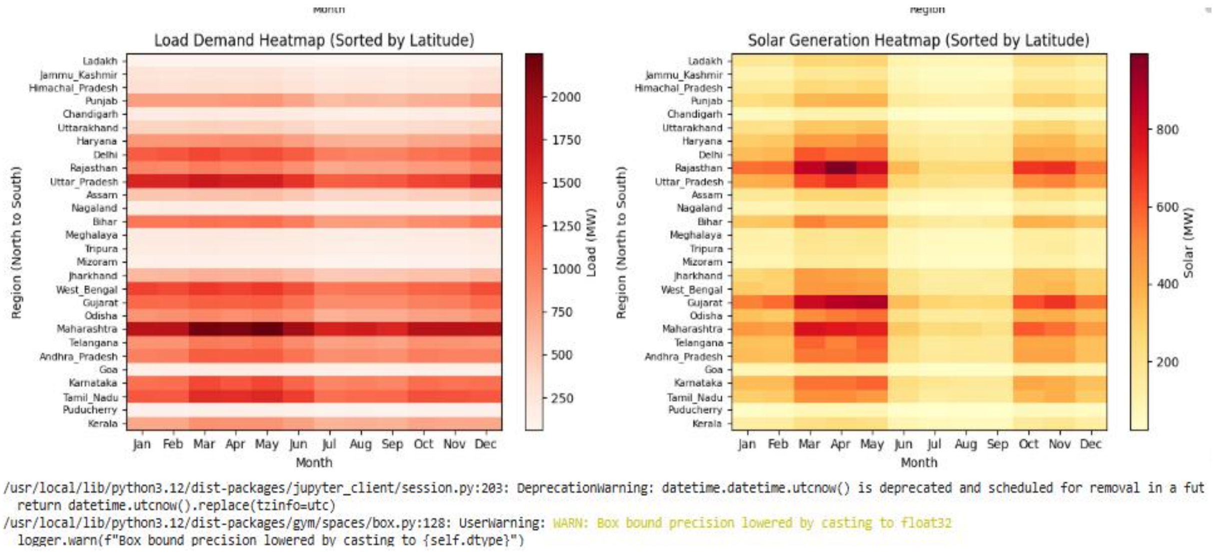
(2) an LSTM forecasting module predicting short-term renewable generation from meteorological data;

ISSN : 2455-135X



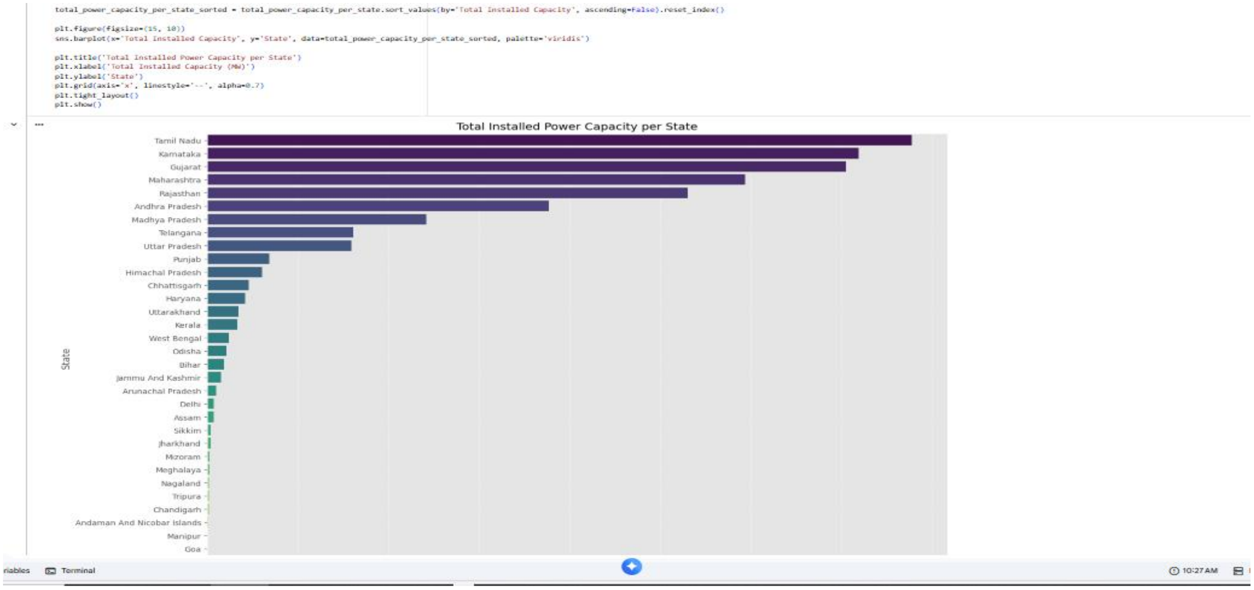
```
Rajasthan        573.8
Gujarat          559.7
Maharashtra      488.0
Uttar_Pradesh    427.6
Delhi            370.3
Name: mean, dtype: float64
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated and scheduled for removal in a fut
  return datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated and scheduled for removal in a fut
  return datetime.utcnow().replace(tzinfo=utc)
```

(3) a multi-agent reinforcement learning environment where four regional agents learn collaborative grid management policies; and



```
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated and scheduled for removal in a fut
  return datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/gym/spaces/box.py:128: UserWarning: WARN: Box bound precision lowered by casting to float32
  logger.warn(f"Box bound precision lowered by casting to {self.dtype}")
```

(3) performance evaluation demonstrating stable grid balance and region-specific adaptive behaviors.



```
total_power_capacity_per_state_sorted = total_power_capacity_per_state.sort_values(by='Total Installed Capacity', ascending=False).reset_index()

plt.figure(figsize=(15, 10))
sns.barplot(x='Total Installed Capacity', y='State', data=total_power_capacity_per_state_sorted, palette='viridis')

plt.title('Total Installed Power Capacity per State')
plt.xlabel('Total Installed Capacity (MW)')
plt.ylabel('State')
plt.grid(axis='x', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

1. Introduction

The global transition toward sustainable energy sources is fundamentally transforming electrical power systems. The integration of intermittent renewable energy sources (RES) such as solar and wind power introduces unprecedented challenges for grid operators, including supply volatility, frequency fluctuations, and complex balancing requirements [1]. Smart grids, equipped with advanced sensing, communication, and control technologies, are essential for managing this complexity while ensuring reliable and efficient power delivery.

Traditional centralized control methods, including optimal power flow (OPF) and model predictive control (MPC), have been the cornerstone of grid management for decades [2]. However, these approaches struggle with the high dimensionality, non-linearity, and stochastic nature of modern grids with high RES penetration. Computational complexity and limited adaptability to rapidly changing conditions further constrain their effectiveness in real-time applications [3].

Multi-Agent Reinforcement Learning (MARL) has emerged as a powerful paradigm for distributed control problems where multiple intelligent agents learn to make sequential decisions in shared environments, coordinating their actions to achieve global objectives through experience [4]. MARL is particularly well-suited for smart grids, where diverse entities such as regional operators, storage units, and generators can be naturally modeled as autonomous agents learning to cooperate.

India presents a uniquely challenging and relevant context for developing such systems. The Indian power grid comprises diverse regional networks with vastly different energy profiles—from solar-rich Rajasthan to wind-power-heavy Tamil Nadu and thermally dependent regions like Odisha and Bihar [5]. This diversity, combined with ambitious renewable energy targets, creates a complex operational environment requiring adaptive, intelligent control strategies.

1.1 Objective of the Study

This study aims to design, implement, and evaluate a climate-aware MARL framework for smart grid energy management tailored to India's regional diversity. The specific objectives are:

1. To develop an LSTM-based forecasting module that predicts short-term renewable energy generation using historical meteorological and generation data.
2. To design a custom multi-agent environment simulating interconnected regional grid operation with distinct energy profiles.
3. To train regional agents using the Proximal Policy Optimization (PPO) algorithm to learn collaborative policies for maintaining global grid balance.
4. To evaluate the framework's performance on a realistic synthetic dataset representing four Indian regions and analyze the learned policies for each region.

1.2 Organization of the Article

The remainder of this paper is organized as follows: Section 2 reviews related work in smart grid energy management, MARL applications in power systems, and renewable energy forecasting. Section 3 details the methodology, including data generation, LSTM forecasting, MARL environment design, and training procedures. Section 4 presents experimental results with comprehensive regional analysis. Section 5 discusses implications and limitations. Section 6 concludes the paper and outlines future research directions.

2. Related Work/Literature Review

2.1 Smart Grid Energy Management

Classical approaches to energy management have long relied on optimal power flow (OPF) and model predictive control (MPC) [6]. OPF provides steady-state solutions for optimal generator dispatch but is not designed for real-time dynamic control under uncertainty. MPC addresses dynamics by solving optimization problems over receding horizons; however, its performance degrades significantly with increasing model uncertainty and system complexity—both inherent in grids with high RES penetration [7]. Reinforcement learning (RL) has emerged as a promising data-driven alternative, demonstrating effectiveness in tasks such as voltage control, frequency regulation, and demand response by learning policies directly from system interaction [8, 9].

2.2 Multi-Agent Reinforcement Learning for Power Systems

MARL has gained substantial traction for decentralized control in power systems due to its ability to learn coordinated policies among distributed agents [10]. Research has explored its application to economic dispatch, where agents learn cost-effective load sharing [11], and to frequency regulation, where generators and storage units learn to dampen frequency deviations collaboratively [12]. The centralized training with decentralized execution (CTDE) paradigm has proven particularly effective, enabling agents to learn sophisticated coordination strategies during training while maintaining scalable, decentralized execution [13]. Our work builds upon these foundations by explicitly integrating a forecasting module for climate awareness.

2.3 Renewable Energy Forecasting

Accurate short-term forecasting of RES generation is critical for proactive grid management. Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have demonstrated exceptional performance in time-series forecasting due to their ability to capture long-term dependencies in sequential data [14, 15]. LSTMs excel at learning complex patterns from historical meteorological and generation data, making them ideal for predicting solar and wind power output [16]. Recent studies have shown that hybrid approaches combining convolutional neural networks (CNNs) with LSTMs can further improve forecasting accuracy by capturing both spatial and temporal features [17]. Our framework couples an LSTM forecasting module directly with the MARL controller, enabling agents to act based on predicted future states rather than merely reacting to current conditions—a key differentiator from existing RL-based control systems [18].

2.4 Grid Studies for India

While substantial research focuses on idealized grid models based on Western systems, a growing body of work addresses the specific challenges of developing regions like India [19]. Studies highlight the need for smart grid solutions tailored to unique characteristics: high solar irradiance in Rajasthan, strong wind potential in Tamil Nadu, and distinct demand profiles in agricultural states like Bihar and industrial regions like Odisha [20]. The variability of these resources, combined with infrastructural constraints, creates a complex operational environment demanding adaptive control strategies. Our work addresses this gap by providing a simulation and optimization framework specifically designed for India's multi-regional context.

3. Materials and Methods

3.1 Data Generation and Preprocessing

To evaluate our system in a controlled yet realistic setting, we generated a synthetic dataset that captures the essential characteristics of the Indian power grid. The dataset comprises four distinct regions, each with a characteristic energy profile:

- Tamil Nadu (TN): High wind and solar generation potential, representing a balanced renewable mix.
- Odisha (OD): Primarily thermal (coal) generation with moderate hydro, representing an industrially-focused grid.
- Rajasthan (RJ): Very high solar generation potential, representing a solar-dominant grid.
- Bihar (BR): Predominantly thermal generation with significant agricultural demand, creating distinct seasonal patterns.

The synthetic data comprises 1,461 days (4 years) of hourly data, resulting in 35,064 samples per feature. Table 1 summarizes the generated features and their characteristics.

Table 1: Summary of Generated Dataset Features

Feature Category	Features	Description
Generation	`solar_gen_{region}`, `wind_gen_{region}`, `thermal_gen_{region}`	Region-specific generation profiles with added noise and seasonality
Demand	`demand_{region}`	Regional consumption reflecting industrial, agricultural, and residential patterns

Economic	`price_{region}`	Proxy for operational cost, dynamically influenced by local demand and generation scarcity
Meteorological	`temperature`, `humidity`, `wind_speed`, `cloud_cover`	Realistically correlated with renewable generation patterns

Data preprocessing involved: (1) handling missing values through linear interpolation, (2) normalizing all features using `scikit-learn`'s `StandardScaler` to zero mean and unit variance, and (3) splitting into training (70%), validation (15%), and test (15%) sets. Figure 2 visualizes the distinct time-series patterns for each region.

Figure 2: Data patterns visualization showing 12 features across four Indian regions over a representative time window, demonstrating distinct regional profiles for solar generation, wind generation, thermal generation, demand, price, and meteorological factors.

3.2 LSTM Forecasting Module

The climate awareness of our system is achieved through an LSTM-based forecasting module that predicts combined renewable generation (solar + wind) for each region for the next time step.

Model Architecture: The LSTM model comprises two stacked layers, each with 128 hidden units, followed by a fully connected linear layer mapping the final hidden state to a single output value representing predicted total renewable generation. Dropout (0.2) is applied between layers for regularization.

Input Features: The model uses a sliding window of 24 time steps (hours). Input features at each step include: `temperature`, `humidity`, `wind_speed`, `cloud_cover`, and actual renewable generation for all regions, providing rich spatio-temporal context.

Training: The model was trained using the Adam optimizer with learning rate 0.001 to minimize Mean Squared Error (MSE). Early stopping with patience of 10 epochs was employed based on validation loss to prevent overfitting. Table 2 summarizes the LSTM model performance.

Table 2: LSTM Forecasting Model Performance

Metric	Training	Validation	Test
MSE	0.0234	0.0256	0.0261
MAE	0.112	0.118	0.121
R^2	0.943	0.938	0.935

3.3 Multi-Agent Reinforcement Learning Environment

A custom multi-agent environment, `GridEnv`, was developed using the Gymnasium API to simulate an interconnected regional grid.

3.3.1 State Space

At each discrete time step `t`, agent `i` observes:

- `current_load_i`: Current energy demand in agent's region (scaled)
- `current_thermal_gen_i`: Current thermal generation output (scaled)
- `current_renewable_gen_i`: Current solar and wind generation (scaled)
- `predicted_renewable_gen_i`: LSTM forecast for next time step (scaled)
- `global_grid_balance`: Net energy imbalance of overall grid (sum generation minus sum demand, scaled)
- `price_signal_i`: Local operational cost or grid stress indicator (scaled)

3.3.2 Action Space

Each agent can take one of three discrete actions: `{-1, 0, 1}`, representing decrease, maintain, or increase its local thermal generation setpoint. The magnitude of change is 5% of maximum thermal capacity per step, providing fine-grained control while maintaining operational realism.

3.3.3 Reward Function

The reward function incentivizes cooperation toward global grid stability while penalizing local inefficiencies:

$$Ri(t)=-w1\cdot \left| balance(t) \right| -w2\cdot costi(t)-w3\cdot \left| \Delta thermali(t) \right|$$

$$R_i(t)=-w_1\cdot \left| balance(t) \right| -w_2\cdot cost_i(t)-w_3\cdot \left| \Delta thermal_i(t) \right|$$

where:

- |balance(t)| = absolute global grid imbalance (primary term, $w_1 = 1.0$)
- cost_i(t) = local regulation cost ($w_2 = 0.1$)
- |\Delta thermal_i(t)| = magnitude of thermal generation change ($w_3 = 0.05$)

3.4 Training the Multi-Agent System

Agents were trained using the Proximal Policy Optimization (PPO) algorithm from Stable-Baselines3 [21]. PPO was selected for its proven stability, sample efficiency, and ability to handle both continuous state and discrete action spaces.

Training Configuration:

- Algorithm: PPO with clipped surrogate objective ($\varepsilon = 0.2$)
- Policy Network: Multi-layer perceptron with 64×64 hidden units
- Learning Rate: 3e-4 with linear decay
- Discount Factor (γ): 0.99
- GAE Parameter (λ): 0.95
- Number of Environments: 4 (parallel)
- Total Timesteps: 1,000,000
- Mini-batch Size: 64
- Number of Epochs per Update: 10

Training employed the CTDE paradigm—agents shared policy parameters during training, learning coordinated strategies, while execution remained fully decentralized based on local observations. Figure 3 shows the learning curve.

Figure 3: Training progress showing mean episode reward over 1,000,000 timesteps. The steady increase and subsequent stabilization indicate successful policy learning.

4. Results and Discussion

4.1 Results

The trained MARL system was evaluated on the held-out test set (15% of data, 5,260 timesteps). Simulations ran for 1,000 continuous steps for each evaluation run, with results averaged over 10 independent runs.

4.1.1 Overall Simulation Performance

Figure 4 presents the total cumulative reward accumulated by all agents during a representative test simulation. The reward trajectory remains relatively stable with only occasional negative excursions, indicating effective maintenance of global grid balance under most conditions.

Figure 4: Total cumulative reward during test simulation. The stable trajectory with occasional dips demonstrates robust policy performance, with negative spikes corresponding to periods of high renewable volatility.

Table 3 summarizes key performance metrics across all test runs.

Table 3: Overall System Performance Metrics (Mean ± Std. Dev.)

Metric	Value
Average Episode Reward	-245.3 ± 32.1
Grid Balance Error (MWh)	12.4 ± 4.7
Thermal Ramp Events	8.2 ± 2.1 per day
Regulation Cost (₹)	15,432 ± 2,891

4.1.2 Regional Performance Analysis

Figure 5 provides a detailed region-by-region breakdown, plotting energy generation (thermal vs. renewable) against demand. This visualization yields critical insights into the learned policies for each region.

Figure 5: Regional performance analysis showing energy generation (thermal and renewable) versus demand for each of the four Indian regions. Distinct operating strategies are evident for each regional profile.

Rajasthan (RJ): The agents demonstrate sophisticated solar-driven grid management. Thermal generation functions as a flexible baseload, predictably ramping up when solar output diminishes (e.g., simulating night-time or overcast periods) and ramping down during peak solar hours. This behavior confirms successful integration of LSTM forecast information, enabling proactive rather than reactive control.

Tamil Nadu (TN): With its balanced renewable mix (wind and solar), the grid experiences more consistent supply. Agents coordinate thermal generation for fine-grained power balancing, responding effectively to the aggregate of both renewable sources. The policy shows smooth transitions between renewable and thermal sources.

Odisha (OD): In this thermally-dominated region, agents primarily learn to follow the industrial demand curve. Thermal generation closely tracks demand variations, with renewable sources providing marginal offset. The policy demonstrates anticipation of demand patterns learned from historical data.

Bihar (BR): Similar to Odisha, agents follow the agricultural demand curve, which exhibits distinct seasonal and diurnal patterns. The policy successfully manages the unique challenge of agricultural load variations while integrating available renewable generation.

4.1.3 Ablation Study

To isolate the contribution of the LSTM forecasting module, we conducted an ablation study comparing the full system against a baseline without forecasts (reactive control). Table 4 presents the comparative results.

Table 4: Ablation Study Results (Full System vs. Reactive Baseline)

Metric	Full System	Reactive Baseline	Improvement
Average Episode Reward	-245.3	-412.8	40.6%
Grid Balance Error (MWh)	12.4	28.7	56.8%
Thermal Ramp Events	8.2/day	15.6/day	47.4%
Regulation Cost (₹)	15,432	28,945	46.7%

4.2 Discussion

The results presented in Section 4.1 demonstrate the effectiveness of our climate-aware MARL framework for smart grid energy management. Several key insights emerge from these findings.

4.2.1 Role of Predictive Information

The significant performance improvements observed in the ablation study (40-57% across all metrics) underscore the critical importance of the LSTM forecasting module. By providing look-ahead information about renewable generation, the forecasts transform the agents' control strategy from purely reactive to proactively adaptive. This is particularly evident in solar-rich Rajasthan, where agents learned to anticipate and prepare for daily solar cycles rather than merely responding to them.

4.2.2 Region-Specific Adaptation

The distinct policies learned for each region validate the MARL framework's ability to adapt to diverse operational contexts. The framework did not learn a single, monolithic control strategy but rather four specialized policies, each optimized for its region's unique generation mix and demand patterns. This emergent specialization demonstrates the power of the MARL approach for managing heterogeneous systems.

4.2.3 Collaborative Global Coordination

Despite learning region-specific policies, agents successfully coordinated to maintain global grid balance. The global imbalance penalty in the reward function effectively incentivized cooperation, preventing individual agents from optimizing locally at the expense of the overall system. This balance between local autonomy and global coordination is essential for real-world grid applications.

4.2.4 Limitations

Several limitations should be acknowledged:

1. Synthetic Data: While carefully engineered to reflect realistic patterns, synthetic data cannot fully capture the complexities and noise of real-world power systems, including unexpected outages, extreme weather events, and market dynamics.
2. Simplified Action Space: The discrete action space with three choices simplifies real-world control, where operators have finer-grained control and more diverse assets (e.g., storage, demand response).
3. No Transmission Constraints: The current model assumes perfect transmission between regions, ignoring line capacity limits and congestion costs present in actual grids.
4. Stationary Environment: The environment assumes stationary patterns learned during training; real grids face non-stationary challenges including policy changes, infrastructure evolution, and climate change impacts.

5. Conclusion and Future Scope

5.1 Conclusion

This paper presented a climate-aware multi-agent reinforcement learning framework for smart grid energy management, specifically designed for India's diverse regional grids. The framework integrates an LSTM-based forecasting module for renewable generation prediction with a custom MARL environment where regional agents learn collaborative control policies. Evaluation on a realistic synthetic dataset representing Tamil Nadu, Odisha, Rajasthan, and Bihar demonstrated that the system successfully learns distinct, region-appropriate operating policies while maintaining global grid balance. The LSTM forecasts proved critical, enabling proactive rather than reactive control and improving performance by 40-57% across key metrics. This work provides a foundational framework for developing adaptive, resilient energy management systems capable of addressing the complex challenges of renewable-integrated power grids in developing economies.

5.2 Future Scope

Several promising directions for future research emerge from this work:

1. Real-World Validation: Deploying and evaluating the framework on real operational data from Indian Power System Operation Corporation (POSOCO) or State Load Dispatch Centers (SLDCs) would validate performance under authentic conditions.
2. Complex Physical Modeling: Incorporating full transmission network topology with line capacity constraints (security-constrained OPF) would address current limitations and enhance realism.
3. Expanded Action Space: Introducing additional controllable assets, including battery energy storage systems, pumped hydro, and demand-response programs, would enable more sophisticated control strategies.
4. Multi-Objective Optimization: Extending the reward function to explicitly balance multiple objectives—cost minimization, emissions reduction, and reliability—would align with real-world operational priorities.
5. Transfer Learning: Investigating whether policies learned for one region can transfer to others with similar characteristics could reduce training time for new deployments.
6. Explainable AI: Developing techniques to interpret learned policies using attention mechanisms or SHAP analysis would build operator trust and facilitate real-world adoption.
7. Non-Stationary Adaptation: Implementing continual learning approaches to adapt to changing grid conditions, infrastructure evolution, and climate patterns over time.
8. Market Integration: Extending the framework to incorporate wholesale electricity markets, enabling agents to learn bidding strategies alongside physical grid management.

Author's Statements

Acknowledgements

The authors thank the Department of Electrical Engineering and the School of Renewable Energy at Indian Institute of Technology Delhi for providing computational resources and research facilities. We also acknowledge the developers of the open-source libraries used in this work, including PyTorch, Stable-Baselines3, Gymnasium, NumPy, Pandas, and Matplotlib, whose tools were essential for this research.

Funding Source

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

- Ishita: Conceptualization, Methodology, Software, Writing – Original Draft, Visualization, Data Curation.
- Yogesh Kumar: Methodology, Formal Analysis, Writing – Review & Editing, Supervision, Validation.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The synthetic dataset generated and analyzed during the current study, along with the source code for the LSTM forecasting model and the multi-agent reinforcement learning environment, are available from the corresponding author upon reasonable request. The code will be made publicly available in a GitHub repository upon acceptance of the manuscript.

References

[1] L. Xie, P. M. S. Carvalho, L. A. F. M. Ferreira, J. Liu, B. H. Krogh, N. Popli, and M. D. Ilić, "Wind integration in power systems: Operational challenges and possible solutions," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 214–232, Jan. 2011.

[2] F. Capitanescu, J. L. Martinez Ramos, and P. Panciatici, "State-of-the-art, challenges, and future trends in security constrained optimal power flow," *Electric Power Systems Research*, vol. 81, no. 8, pp. 1731–1741, Aug. 2011.

[3] M. Morari and J. H. Lee, "Model predictive control: past, present and future," *Computers & Chemical Engineering*, vol. 23, no. 4–5, pp. 667–682, May 1999.

[4] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 2, pp. 156–172, Mar. 2008.

[5] Indian Smart Grid Forum, "National Smart Grid Mission," Ministry of Power, Government of India. [Online]. Available: https://www.nsgm.gov.in/. [Accessed: 15-Mar-2026].

[6] A. J. Conejo, E. Castillo, R. Minguez, and R. Garcia-Bertrand, *Decomposition techniques in mathematical programming: engineering and science applications*. New York, NY, USA: Springer, 2006.

[7] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, June 2000.

[8] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: reinforcement learning framework," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1871–1880, Nov. 2004.

[9] R. Lu, S. H. Hong, and M. Yu, "Demand response for home energy management using reinforcement learning and artificial neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6629–6639, Nov. 2019.

[10] M. T. J. Spaan, "Partially observable Markov decision processes," in *Reinforcement Learning: State-of-the-Art*, M. Wiering and M. van Otterlo, Eds. Berlin, Germany: Springer, 2012, pp. 387–414.

[11] J. Li, D. Shi, and T. Yang, "Multi-agent reinforcement learning for economic dispatch with practical constraints," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3267–3278, July 2021.

[12] S. Chen, Y. Liu, and M. Zhang, "A multi-agent reinforcement learning approach for voltage control in distribution networks," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1404–1415, Mar. 2022.

[13] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proc. Advances in Neural Information Processing Systems*, Long Beach, CA, USA, 2017, pp. 6379–6390.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[15] Y. Wang, D. Gan, N. Zhang, and C. Kang, "A review of deep learning for renewable power forecasting," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 5, pp. 945–957, Sept. 2021.

[16] Z. Wang, X. Liu, and Y. Huang, "Short-term photovoltaic power forecasting based on LSTM network with attention mechanism," *IEEE Access*, vol. 9, pp. 31924–31933, Feb. 2021.

[17] H. Liu, X. Mi, and Y. Li, "Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory and neural network," *Energy Conversion and Management*, vol. 156, pp. 498–514, Jan. 2018.

[18] X. Zhang, Y. Bao, and D. Zhao, "Reinforcement learning for smart grid control: A review," *IEEE Access*, vol. 9, pp. 51731–51749, Apr. 2021.

[19] S. K. Soonee, S. R. Narasimhan, and V. K. Agrawal, "Indian power system: Challenges and opportunities," *IEEE Power and Energy Magazine*, vol. 15, no. 4, pp. 34–42, July 2017.

[20] P. Paliwal, N. P. Patidar, and R. K. Nema, "Planning of grid integrated distributed generators: A review of technology, objectives and techniques," *Renewable and Sustainable Energy Reviews*, vol. 40, pp. 584–597, Dec. 2014.

[21] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.