

# Deep Reinforcement Learning with Local Interpretability for Transparent Microgrid Resilience Energy Management

Mohammad Hossein  
Nejati Amiri  
College of Engineering  
Birmingham City University  
Birmingham, UK  
mohammadhossein.nejatiamiri  
@mail.bcu.ac.uk

Fawaz Annaz  
College of Engineering  
Birmingham City University  
Birmingham, UK  
fawaz.annaz@bcu.ac.uk

Mario De Oliveira  
College of Engineering  
Birmingham City University  
Birmingham, UK  
mario.deoliveira@bcu.ac.uk

Florimond Gueniat\*  
College of Engineering  
Birmingham City University  
Birmingham, UK  
florimond.gueniat@bcu.ac.uk

**Abstract**—Renewable energy integration into microgrids has become a key approach to addressing global energy issues such as climate change and resource scarcity. However, the variability of renewable sources and the rising occurrence of High Impact Low Probability (HILP) events require innovative strategies for reliable and resilient energy management. This study introduces a practical approach to managing microgrid resilience through Explainable Deep Reinforcement Learning (XDRL). It combines the Proximal Policy Optimization (PPO) algorithm for decision-making with the Local Interpretable Model-agnostic Explanations (LIME) method to improve the transparency of the actor network’s decisions. A case study in Ongole, India, examines a microgrid with wind, solar, and battery components to validate the proposed approach. The microgrid is simulated under extreme weather conditions during the Layla cyclone. LIME is used to analyse scenarios, showing the impact of key factors such as renewable generation, state of charge, and load prioritization on decision-making. The results demonstrate a Resilience Index (RI) of 0.9736 and an estimated battery lifespan of 15.11 years. LIME analysis reveals the rationale behind the agent’s actions in idle, charging, and discharging modes, with renewable generation identified as the most influential feature. This study shows the effectiveness of integrating advanced DRL algorithms with interpretable AI techniques to achieve reliable and transparent energy management in microgrids.

**Index Terms**—Interpretable and Explainable AI, Microgrid, Deep Reinforcement Learning, Resilient Energy Management, Smart Grid

## I. INTRODUCTION

The electricity grid has witnessed major changes in recent years, driven by advancements such as bidirectional communication between suppliers and consumers, enabling smart grids’ development. Smart grids incorporate concepts like microgrids, demand response, prosumers, self-healing systems, and local electricity markets to meet evolving energy needs [1]. Renewable energy sources have received significant attention as a response to global warming, the push for net-zero emissions, and the challenges of fossil fuel depletion and pollution

[2]. Microgrids, with their ability to function in both isolated and grid-connected modes, serve as a practical framework for leveraging renewable energy despite its intermittent nature [3].

Modern power grids face growing vulnerabilities from natural disasters and cyber threats, intensified by digitalisation.

These rare but high-impact events underscore the importance of enhancing resilience in smart grids [4]. At the same time, advances in Artificial Intelligence (AI) have provided new tools to tackle challenges in power systems, such as real-time operations and uncertainties like variable loads (e.g., electric vehicles) and energy production [5]. Reinforcement learning (RL) stands out among AI approaches for its ability to quickly adapt to dynamic environments and effectively manage systems without the need for detailed models. This makes RL particularly suitable for addressing various planning and operational challenges in power systems, such as optimal dispatch and control [6].

Recent advancements in Deep Reinforcement Learning (DRL), which combines RL with Deep Neural Networks (DNN), have shown outstanding performance in solving complex problems. Despite its potential, DRL’s application in critical industries such as power systems is still limited due to concerns over its black-box nature and the lack of transparency in decision-making [7]. To tackle concerns about the black-box nature of DNN, along with ethical issues and regulations like the General Data Protection Regulation (GDPR), eXplainable Artificial Intelligence (XAI) has become a prominent research area in computer science [8]. In DRL, eXplainable Deep Reinforcement Learning (XDRL) is classified into three approaches: Interpretable Agents (IA), Intrinsic Explainability (IE), and Post-hoc Explainability (PHE) [9].

Interpretable Agents (IA) are inherently designed to be easily understood by humans, relying on rule-based or linear models. However, this simplicity often comes at the cost of performance. Intrinsic Explainability (IE) focuses on enhanc-

ing explainability by modifying the RL agent or its model, such as reward and transition functions, as part of its design. Post-hoc Explainability (PHE) generates explanations for the decisions of pre-trained models through external techniques without altering their internal structure.

Among these approaches, PHE offers high performance despite its limited inherent explainability, which makes it well-suited for smart grid applications where performance is a priority. Since it operates after model training, PHE can also be used to analyse pre-existing models. Methods like Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are commonly used, with LIME offering local insights and SHAP providing both local and global explanations [10], [11]. This study uses LIME for its simplicity and focus on specific decision analysis.

The main contributions of this work are as follows:

- A real-world case study in Ongole (India), utilizing actual geographical and load data for microgrid design. The microgrid comprises wind turbines, solar panels, batteries, and loads with different priority levels. Component sizing is performed using HOMER Pro to achieve an optimal design tailored to the location's specific needs.
- Renewable generation simulated under the impact of the Layla cyclone that occurred in this region in mid-May 2010 to capture the challenges posed by extreme weather events.
- Use of the DRL Proximal Policy Optimization (PPO) algorithm within an Actor-Critic framework, a well-established and efficient method, enhances the resilience of energy management of the microgrid.
- Application of the LIME method to explain specific decisions made by the DRL agent. This strategy enhanced transparency and built stakeholder trust in the system's operations.

By integrating these features, the study aims to present an explainable approach to microgrid resilience management that aligns with real-world scenarios and stakeholder expectations.

The structure of this article is as follows: Section II describes the microgrid modelling approach, the LIME method, and the mathematical formulation used in the study. Section III presents and analyses the simulation results. Finally, Section IV concludes the study and outlines potential future research directions.

## II. MICROGRID MODELLING

We used HOMER Pro to size an appropriate microgrid for the considered site, consisting of 140 kW of solar power, 80 kW of wind power, a 780 kWh battery, and a 52 kW converter. Figure 1 illustrates the resilient energy management of the microgrid using XDRL. The total load consumption and the renewable power generated based on the weather data are depicted in Figure 2.

The following section discusses modelling the RL environment.

### A. Microgrid Environment Design

The design of the microgrid environment is crucial for the DRL agent to interact effectively and learn optimal policies. The environment encapsulates the physical and operational characteristics of a microgrid, including battery storage, renewable energy generation, and load demands. The custom environment is developed using the OpenAI Gymnasium framework, which offers a standardised interface for DRL agents. It allows libraries like Stable-Baselines3 to seamlessly interact with the environment. The environment simulates battery operations, renewable energy inputs, and load demands over discrete time steps. Renewable energy generation and load demands are deterministic, based on historical data, while the initial State Of Charge (SOC) is randomized to introduce variability in initial conditions.

The state space represents all possible states of the microgrid at any given time step. It includes variables essential for the agent to make informed decisions. The state vector at time step  $t$  is defined as:

$$\mathbf{s}_t = [\text{SOC}_t, L_{1,t}, L_{2,t}, L_{3,t}, P_{\text{RE},t}, P_{\text{net},t}]. \quad (1)$$

The SOC of the battery at time  $t$ , denoted as  $\text{SOC}_t$ , is constrained between  $\text{SOC}_{\min} = 0.2$  and  $\text{SOC}_{\max} = 0.9$  to consider the safety and longevity of the battery. The load demand at time  $t$  is represented by  $L_{i,t}$ , where  $i = 1, 2, 3$  corresponds to different priority levels. Essential load ( $L_{1,t}$ ) holds the highest priority with a weight 3.5 times that of business load ( $L_{2,t}$ ), which itself has 2 times higher priority than agricultural load ( $L_{3,t}$ ), the lowest priority.

Renewable energy generation ( $P_{\text{RE},t}$ ) at time  $t$  is determined from historical weather data, and the net energy ( $P_{\text{net},t}$ ) is calculated as:

$$P_{\text{net},t} = P_{\text{RE},t} - \sum_{i=1}^3 L_{i,t} \quad (2)$$

The action space defines the set of all possible actions that the agent can take at each time step. It is a continuous space consisting of five variables, represented as:

$$\mathbf{a}_t = [a_{\text{ch},t}, a_{\text{dis},t}, w_{1,t}, w_{2,t}, w_{3,t}], \quad (3)$$

where  $a_{\text{ch},t}$  and  $a_{\text{dis},t}$  are the normalized charging and discharging power actions, respectively, both constrained to the range  $[-1, 1]$ . The variables  $w_{i,t}$ , for  $i = 1, 2, 3$ , represent the raw weights for distributing power among the three loads, and each is also bounded within  $[-1, 1]$ . This action space enables the agent to determine charging and discharging actions while allocating power to different loads effectively.

The weights for power distribution among the loads are normalised using a softmax function to ensure they sum to one:

$$\hat{w}_{i,t} = \frac{\exp(w_{i,t})}{\sum_{j=1}^3 \exp(w_{j,t})}, \quad \text{for } i = 1, 2, 3. \quad (4)$$

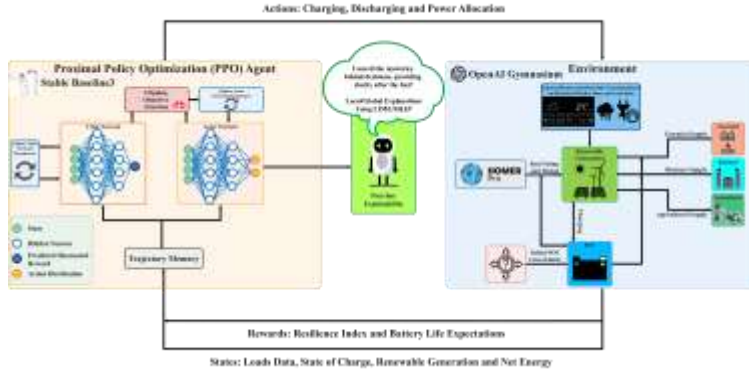


Fig. 1: Proposed XDRL Framework for Microgrid Resilience Energy Management.

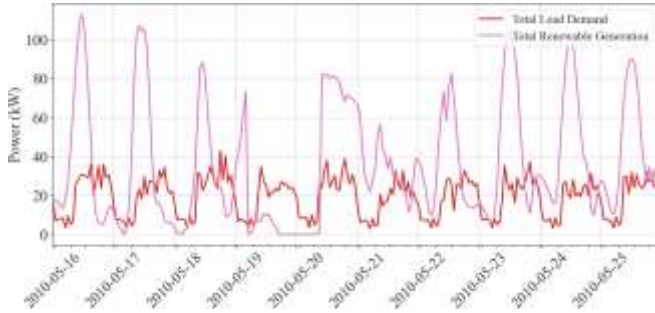


Fig. 2: Total Load and Total Renewable Power Generation Profiles Based on Weather Data.

The SOC of the battery is updated based on charging and discharging actions while considering efficiency losses:

$$\text{SOC}_{t+1} = \text{SOC}_t + \frac{\eta_{\text{ch}} P_{\text{ch},t} - \frac{P_{\text{dis},t}}{\eta_{\text{dis}}}}{E_{\text{max}}}, \quad (5)$$

subject to the constraint:

$$\text{SOC}_{\text{min}} \leq \text{SOC}_{t+1} \leq \text{SOC}_{\text{max}}, \quad (6)$$

where  $\eta_{\text{ch}} = 0.90$  is the charging efficiency,  $\eta_{\text{dis}} = 0.95$  is the discharging efficiency, and  $E_{\text{max}} = 780$  kWh is the maximum battery capacity. The available capacities for charging and discharging are defined as:

$$E_{\text{avail,ch},t} = (\text{SOC}_{\text{max}} - \text{SOC}_t) E_{\text{max}}, \quad (7)$$

$$E_{\text{avail,dis},t} = (\text{SOC}_t - \text{SOC}_{\text{min}}) E_{\text{max}}. \quad (8)$$

The charging and discharging powers are further constrained by the available capacities:

$$P_{\text{ch},t} = \min(P_{\text{ch},t}, E_{\text{avail,ch},t}), \quad (9)$$

$$P_{\text{dis},t} = \min(P_{\text{dis},t}, E_{\text{avail,dis},t}). \quad (10)$$

Additionally, charging and discharging cannot occur simultaneously. Depending on the net energy, either charging or discharging is permitted:

$$\begin{cases} P_{\text{dis},t} = 0, & \text{if } P_{\text{net},t} \geq 0, \\ P_{\text{ch},t} = 0, & \text{if } P_{\text{net},t} < 0. \end{cases} \quad (11)$$

The net power supply after battery operation ( $P_{s,t}$ ) is calculated as:

$$P_{s,t} = P_{\text{RE},t} + P_{\text{dis},t} - P_{\text{ch},t}. \quad (12)$$

This power is allocated to the loads based on the normalized weights:

$$P_{s,i,t} = \hat{w}_{i,t} P_{s,t}, \quad \text{for } i = 1, 2, 3. \quad (13)$$

The power imbalance for each load is given by:

$$P_{\text{imb},i,t} = P_{s,i,t} - L_{i,t}, \quad \text{for } i = 1, 2, 3. \quad (14)$$

Negative imbalances indicate shortages, which are critical for resilience calculations. These shortages are defined as:

$$P_{\text{sh},i,t} = -\min(0, P_{\text{imb},i,t}), \quad \text{for } i = 1, 2, 3. \quad (15)$$

The reward function guides the agent towards actions that enhance system resilience. At each time step, the reward  $r_t$  calculates the Resiliency Index (RI) reward.

The RI reward incentivizes the agent to minimize power shortages, especially for high-priority loads. It is calculated as:

$$r_t = 1 - \frac{7P_{\text{sh},1,t} + 2P_{\text{sh},2,t} + 1P_{\text{sh},3,t}}{7L_{1,t} + 2L_{2,t} + 1L_{3,t}}. \quad (16)$$

The cumulative reward for the entire episode is obtained by summing over all time steps  $t$  from the start to the termination of the episode:

$$R_{\text{episode}} = \sum_{t=1}^T r_t. \quad (17)$$

The overall RI for the episode is calculated as:

$$\text{RI} = 1 - \frac{\sum_{t=1}^T P_{\text{sh},1,t+2} \sum_{t=1}^T P_{\text{sh},2,t+1} \sum_{t=1}^T P_{\text{sh},3,t}}{7 \sum_{t=1}^T L_{1,t+2} \sum_{t=1}^T L_{2,t+1} \sum_{t=1}^T L_{3,t}}. \quad (18)$$

The total reward for the episode is then adjusted by adding the contributions from the overall RI at the termination of the episode:

$$R_{\text{final}} = R_{\text{episode}} + \text{RI}. \quad (19)$$

Finally, the total reward for the episode is normalized by dividing it by the maximum possible total reward  $r_{\text{max}}$ :

$$R_{\text{final}}^{\text{normalized}} = \frac{R_{\text{final}}}{r_{\text{max}}}. \quad (20)$$

### B. Conceptual Overview: PPO and LIME

PPO is a DRL algorithm designed to improve policy-gradient methods by ensuring stable and efficient training. PPO simplifies Trust Region Policy Optimization (TRPO) by avoiding complex second-order optimization while maintaining stable updates. The core idea of PPO is to use a clipped surrogate objective that limits excessively large policy updates [12]. The policy update is formulated as:

$$L^{\text{CLIP}}(\vartheta) = E_t \min r_t(\vartheta) \hat{A}_t, \text{clip}(r_t(\vartheta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t, \quad (21)$$

where  $r_t(\vartheta) = \frac{\pi_{\vartheta}(a_t|s_t)}{\pi_{\vartheta_{\text{old}}}(a_t|s_t)}$  is the probability ratio between the new policy  $\pi_{\vartheta}$  and the old policy  $\pi_{\vartheta_{\text{old}}}$ ,  $\hat{A}_t$  is the advantage function, and  $\epsilon$  is the clipping parameter (e.g., 0.2). The agent follows a policy  $\pi_{\vartheta}$  that determines its actions to optimize rewards over time, with identifying the policy equivalent to determining its parameters  $\vartheta$ . This objective ensures that policy updates remain within a trust region, avoiding performance degradation from overly large updates.

The PPO algorithm alternates between collecting experience through interaction with the environment and optimizing the policy using gradient ascent. The combined objective includes terms for the value function  $L^{\text{VF}}$  and an entropy bonus  $L^{\text{entropy}}$ , which encourage exploration. The overall loss is given by:

$$L_t^{\text{PPO}}(\vartheta, \phi) = E_t [L_t^{\text{CLIP}}(\vartheta) - c_1 L_t^{\text{VF}}(\phi) + c_2 L_t^{\text{entropy}}(\vartheta)], \quad (22)$$

where  $c_1$  and  $c_2$  are coefficients balancing the contributions of each term.

LIME is a model-agnostic interpretability method that explains individual predictions by approximating the behaviour of a complex model locally with a simpler, interpretable surrogate model [10]. Given an input instance  $x$ , LIME generates perturbed samples around  $x$ , calculates the corresponding outputs using the original model, and fits a simple model  $g$  to these outputs. The optimization problem is:

$$\xi(x) = \arg \min_{g \in G} (L(f, g, \pi_x) + \Omega(g)), \quad (23)$$

where  $L(f, g, \pi_x)$  measures the fidelity of  $g$  in approximating the original model  $f$  locally,  $\pi_x$  is a proximity measure that assigns higher weights to samples closer to  $x$ , and  $\Omega(g)$  is a complexity penalty ensuring the model has a low number of parameters. The proximity function is defined as:

$$\pi_x(z) = e^{-\frac{D(x,z)^2}{\sigma^2}}, \quad (24)$$

where  $D(x, z)$  is the distance between  $x$  and a perturbed sample  $z$ , and  $\sigma$  controls the locality. LIME prioritizes samples closer to the instance  $x$ , ensuring the explanation model  $g$  focuses on local behavior.

In this work, LIME is applied to interpret the decisions made by the actor in the PPO framework. Specifically, the actor's policy  $\pi_{\vartheta}(a_t|s_t)$ , which determines the action probabilities given a state, is analysed locally using LIME. By generating explanations for the actor's decisions, LIME helps uncover the factors influencing the policy's behavior. This interpretability is crucial for understanding and debugging reinforcement learning agents, especially in critical infrastructures like microgrid energy management. By combining PPO's robust policy optimization with LIME's interpretability, this approach balances performance and transparency, enabling the DRL agent's decisions to be both effective and explainable.

### III. SIMULATION RESULTS

This section presents the simulation results. Figure 3 illustrates the battery's SOC, which indicates charging and discharging patterns over time. Figure 4 shows the load supply

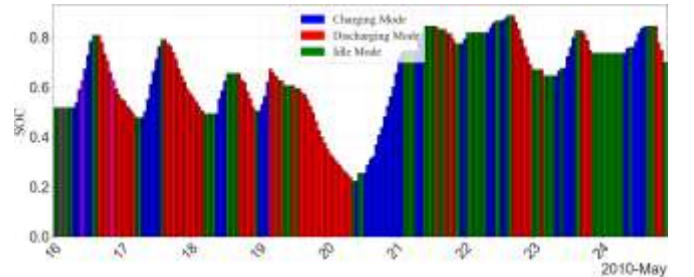


Fig. 3: SOC plot during the cyclone.

for the three prioritized loads. The total resilience index achieved was 0.9736, a reasonable value for an event like a cyclone. Additionally, the expected battery life was estimated to be 15.11 years, aligning with the design parameters from HOMER Pro. Additionally, the reward convergence curve is shown in Figure 5, where it can be observed that the reward stabilizes and converges after approximately 40,000 episodes and approaches the oracle strategy. This article focuses on analysing the factors driving the DRL agent's decision-making process for battery charging and discharging. In Figure 3, three specific scenarios are shown in pink to represent distinct operational modes on day 16: idle at hour 4, charging at hour 11, and discharging at hour 21. The agent's choices are analysed in these scenarios to explore the factors influencing its decisions under different scenarios.

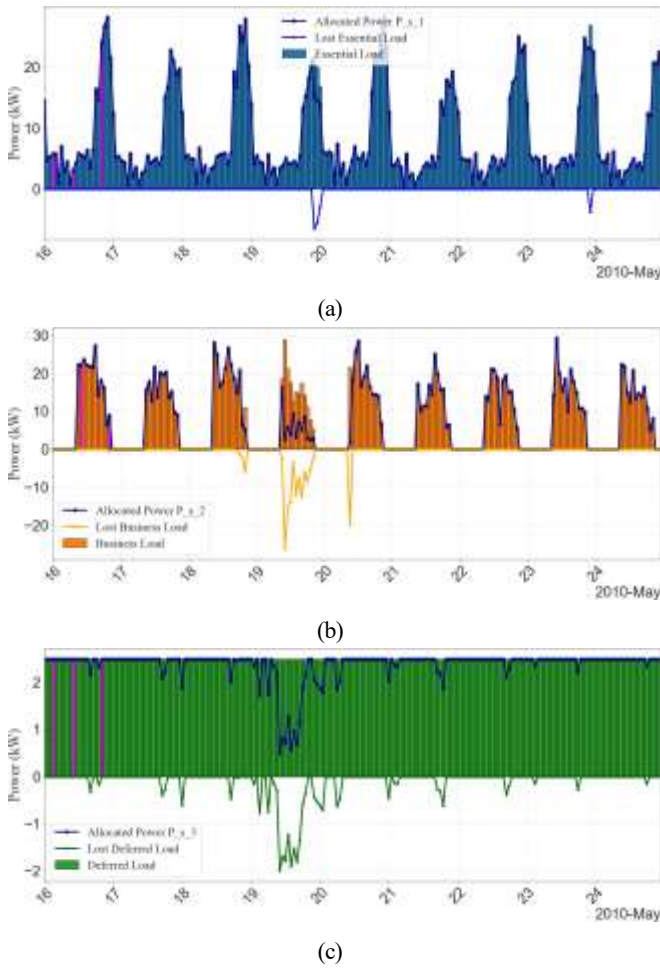


Fig. 4: a) Essential Load, b) Business Load, and c) Agricultural Load Demand, Supply, and Imbalances.

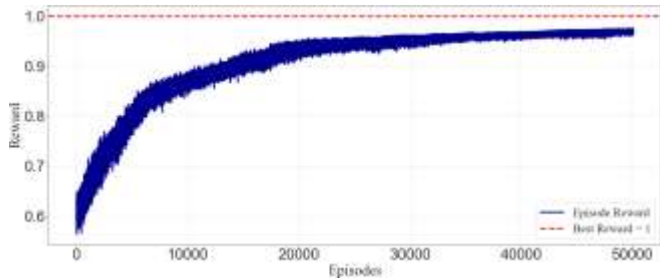


Fig. 5: Reward Convergence Over episodes.

The LIME results for the idle mode, presented in Figures 6b and 6a, indicate that the DRL actor's decision to neither charge nor discharge is driven by conflicting impacts of various features. For the charging action (Figure 6a), all features discourage charging, with significant negative contributions from SOC, renewable generation, and Load<sub>2</sub>, suggesting an unfavourable condition for charging. In contrast, for the discharging action (Figure 6b), SOC, renewable generation, and net energy positively influence discharging, but these are nearly counterbalanced by the high negative impact of

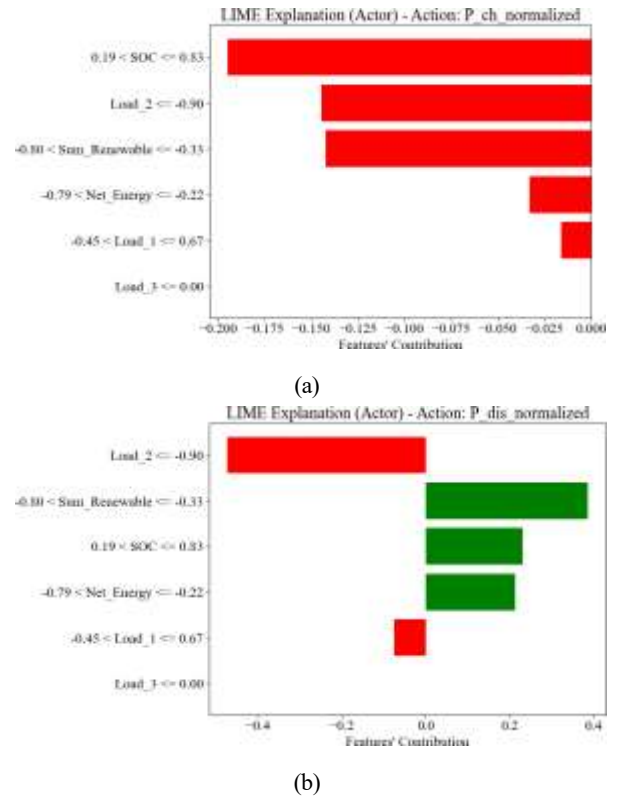
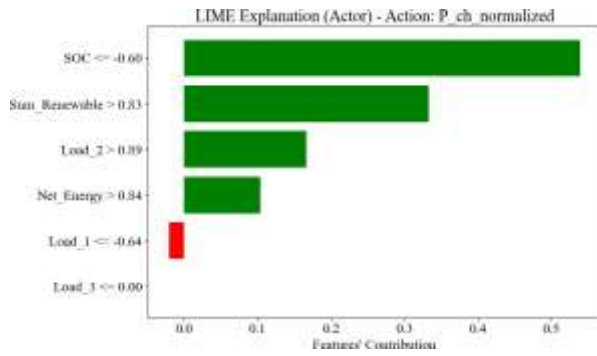


Fig. 6: LIME-based Explanations in IDLE Mode: a) Charging, and b) Discharging Action.

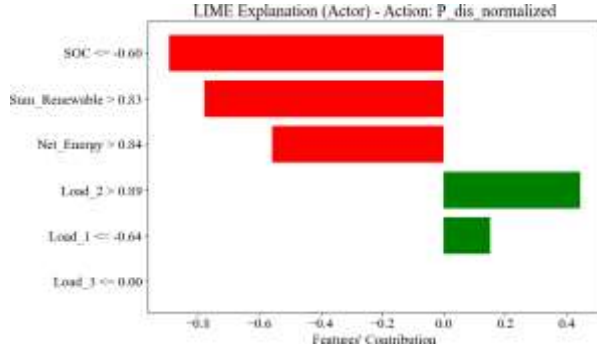
Load<sub>2</sub>, resulting in no clear encouragement for discharging. This balance of influences leads the actor to select the idle action. Additionally, Load<sub>3</sub> remains insignificant across all scenarios due to its fixed nature and low priority, while Load<sub>2</sub> emerges as the most impactful feature, highlighting the actor's sensitivity to intermediate-priority loads in decision-making.

In the charging scenario, illustrated in Figures 7a and 7b, the actor's decision to charge is strongly influenced by multiple encouraging features. In the charging action (Figure 7a), SOC emerges as the most impactful positive factor, supported by contributions from renewable generation, Load<sub>2</sub>, and net energy, all favouring the charging action. Conversely, in the discharging action (Figure 7b), SOC, renewable generation, and net energy significantly discourage discharging, while Load<sub>2</sub> and Load<sub>1</sub> do not provide sufficient encouragement. This leads the actor to select the charging action as the optimal choice in this instance.

In the discharging scenario, shown in Figures 8a and 8b, the actor's decisions are shaped by contrasting feature impacts. In the charging action (Figure 8a), factors such as renewable generation, SOC, and net energy have significant negative impacts, discouraging the actor from charging. However, in the discharging action (Figure 8b), these same features contribute positively, strongly encouraging discharging. This alignment of positive impacts on discharging explains the actor's choice to discharge in this scenario.



(a)



(b)

Fig. 7: LIME-based Explanations in Charging Mode: a) Charging, and b) Discharging Action.

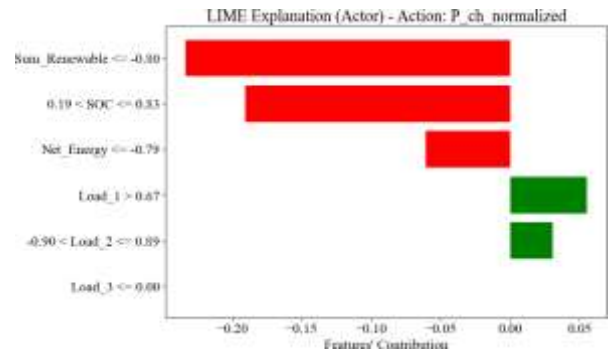
#### IV. CONCLUSION

This study introduced a transparent framework for microgrid resilience management by combining PPO and LIME. The microgrid was designed for the coastal city of Ongole in India, aligning with local load and weather conditions under the Layla cyclone scenario. The approach achieved a high Resilience Index (0.9736) while maintaining an estimated battery lifespan of over 15 years. LIME identified key factors influencing decisions, including renewable generation, load priorities, and battery SOC, enhancing stakeholder confidence and trust in the agent's charging and discharging actions.

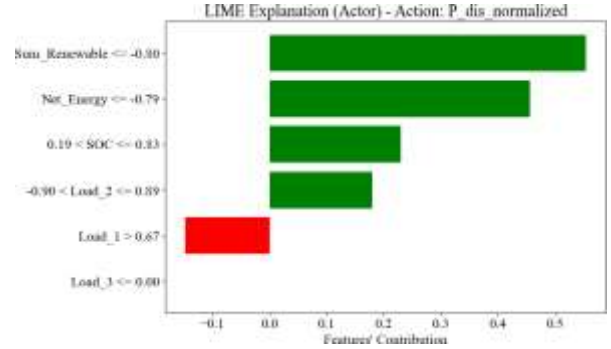
Future research can explore alternative explainability methods (e.g., SHAP) for a broader perspective and global explanations of the DRL agent's decisions. Investigations into multi-microgrid coordination, real-time constraints, and cybersecurity considerations would further strengthen overall resiliency. Additionally, incorporating economic factors and battery life into the reward function could balance resilience with profitability.

#### REFERENCES

- [1] M. H. Nejadi Amiri, M. Mehdinejad, A. Mohammadpour Shotorbani, and H. Shayanfar, "Heuristic retailer's day-ahead pricing based on online-learning of prosumer's optimal energy management model," *Energies*, vol. 16, no. 3, p. 1182, 2023.
- [2] Y. Han, M. Bao, Y. Niu, and J. ur Rehman, "Driving towards net zero emissions: The role of natural resources, government debt and political stability," *Resources Policy*, vol. 88, p. 104479, 2024.



(a)



(b)

Fig. 8: LIME-based Explanations in Discharging Mode: a) Charging, and b) Discharging Action.

- [3] M. H. N. Amiri, F. Annaz, M. De Oliveira, and F. Gueniat, "Strategies for resilience and battery life extension in the face of communication losses for isolated microgrids," in *2024 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, pp. 1–5, IEEE, 2024.
- [4] M. H. N. Amiri and F. Gueniat, "Towards a framework for measurements of power systems resiliency: Comprehensive review and development of graph and vector-based resiliency metrics," *Sustainable Cities and Society*, p. 105517, 2024.
- [5] P. Arevalo and F. Jurado, "Impact of artificial intelligence on the planning and operation of distributed energy systems in smart grids," *Energies*, vol. 17, no. 17, p. 4501, 2024.
- [6] P. Mohammadi, R. Darshi, S. Shamaghdari, and P. Siano, "Comparative analysis of control strategies for microgrid energy management with a focus on reinforcement learning," *IEEE Access*, 2024.
- [7] S. Stavrev and D. Ginchev, "Reinforcement learning techniques in optimizing energy systems," *Electronics*, vol. 13, no. 8, p. 1459, 2024.
- [8] V. Chamola, V. Hassija, A. R. Sulthana, D. Ghosh, D. Dhingra, and B. Sikdar, "A review of trustworthy and explainable artificial intelligence (xai)," *IEEE Access*, 2023.
- [9] Y. Bekkemoen, "Explainable reinforcement learning (xrl): a systematic literature review and taxonomy," *Machine Learning*, vol. 113, no. 1, pp. 355–441, 2024.
- [10] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.
- [11] S. Lundberg, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.
- [12] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.