2026 15th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP) | 979-8-3315-8089-6/26/$31.00 ©2026 IEEE | DOI: 10.1109/CSNDSP68462.2026.11654398

Explainable AI-based Feature Selection Method for Energy Harvesting-assisted IoT Network

Huijun Tang*, Wang Zeng†, Min Xue‡, Junhui Du§, Zhizhou He¶, Pengfei Jiao†, Huaming Wu§ and Hongjian Sun*
*Department of Engineering, Durham University, Durham DH1 3LE, United Kingdom
†School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China
‡Institute of Computer Science, Heidelberg University, Germany
§Center for Applied Mathematics, Tianjin University, Tianjin 300072, China
¶Institute for Communication Systems, University of Surrey, Guildford GU2 7XH, United Kingdom
Emails: huijun.tang@durham.ac.uk, 242270069@hdu.edu.cn, min.xue@uni-heidelberg.de, dujunhui0325@tju.edu.cn
zhizhou.he@surrey.ac.uk, pjiao@hdu.edu.cn, whming@tju.edu.cn, hongjian.sun@durham.ac.uk

***Abstract*—Energy Harvesting (EH)-assisted Backscatter IoT networks offer a promising solution for sustainable connectivity but face complex optimization challenges. While Deep Reinforcement Learning (DRL) provides a viable model-free solution, its black-box nature often obscures the underlying decision mechanisms, leading to redundant state representations and inefficient resource usage. To address these challenges, we propose XAI-LyDRL, a novel framework that integrates Lyapunov optimization with explainable artificial intelligence (XAI)–based feature selection. We derive three key mechanistic insights: (1) we empirically verify that the agent implicitly internalizes physical path-loss models, rendering explicit geometric distance features redundant; (2) we unveil a regime-dependent attention mechanism where the policy dynamically shifts focus from queue stability to channel quality based on resource constraints, confirming theoretical consistency with Lyapunov optimization; and (3) we identify effective state sparsity where optimal actions are primarily driven by bottleneck users. Guided by these insights, we construct a streamlined policy that prunes 40% of the state space. Simulation results demonstrate that this lightweight design achieves comparable stability and throughput to full-scale models while providing an empirical foundation for scalable deployment in resource-constrained edge devices.**

***Index Terms*—Internet-of-Things, Energy Harvesting, Explainable AI, Deep Reinforcement Learning**

I. Introduction

The Internet of Things (IoT) ecosystem is undergoing an unprecedented expansion and transformation [1]. Mobile Edge Computing (MEC) pushes resources to the network edge, essential for meeting the low-latency and high-reliability demands of mission-critical applications [2]. However, this densification of edge intelligence faces a severe sustainability bottleneck [3]. Conventional batteries are untenable for maintaining massive distributed devices, especially in remote or implantable applications where replacement is infeasible [4].

To address this, wireless power transfer (WPT) and backscatter communication (BackCom) have emerged as promising technologies [5]. WPT enables devices to harvest energy from radio-frequency (RF) signals transmitted by dedicated power beacons (PBs) or hybrid access points (APs), thus creating a self-sustaining energy supply. Meanwhile, BackCom allows devices to transmit data by modulating and reflecting incident RF signals without generating an energy-intensive active carrier, reducing communication power consumption to the μW level [6], [7]. By harvesting energy from ambient RF signals and transmitting data via reflection modulation, the integration of Energy Harvesting (EH) and BackCom effectively extends the lifespan of IoT networks.

To enhance spectral efficiency in these resource-constrained environments, technologies such as Non-Orthogonal Multiple Access (NOMA) are often employed, allowing multiple devices to share the same time-frequency resources. However, resource management in such networks is mathematically intractable due to the stochastic nature of wireless channels, energy arrivals, and task requests. Deep Reinforcement Learning (DRL) offers a viable solution due to its model-free ability to learn policies directly from environmental interactions. Mazhar *et al.* [8] proposed a DRL architecture integrating cognitive NOMA and backscatter technology to minimize the Age of Information. Hsu *et al.* [9] propose a DRL-based EH scheduling scheme for NOMA-based networks to optimize average binary scale satisfaction. However, DRL struggle to enforce the strict stability constraints required by EH systems. To address this, Lyapunov optimization theory has been integrated with DRL [10], [11], giving rise to Lyapunov-guided DRL (LyDRL) frameworks that provide theoretical stability guarantees while leveraging the adaptability of deep learning. Wu *et al.* [12] introduced a LyDRL approach that effectively handles long-term energy constraints by stabilizing energy queues while maximizing system throughput.

Nevertheless, it is often unclear which features actually drive the agent's decision-making process and which are mere noise or redundant data, leading to extra computational overhead. While in the context of EH, redundant input features not only waste limited energy and computational resources, but model lightweighting becomes as critical as performance maximization. Recently, Explainable AI (XAI) has gained traction as a powerful tool to enhance the transparency and computational efficiency [13]–[15]. In this paper, we propose

Corresponding author: Min Xue

979-8-3315-8089-6/26/$31.00 ©2026 IEEE

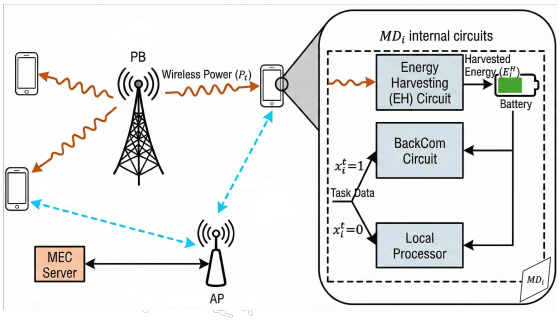


Fig. 1. The system model.

an XAI-aided feature selection method for EH-assisted MEC networks. The main contributions are as follows:

- We empirically verify that the DRL agent effectively internalizes the physical mapping between geometric distance and channel quality through training. Our analysis reveals that explicit distance features contribute negligibly to decision-making when Channel State Information (CSI) is available. Then, we construct a lightweight state space that reduces input redundancy by 40%.
- We unveil the adaptive nature of the Lyapunov-guided policy through feature sensitivity analysis across different resource regimes. We observe a distinct attention shift: the agent prioritizes queue length information to ensure stability under resource-constrained regimes, while shifting focus to channel quality to maximize utility under resource-abundant regimes. This result confirms that the black-box DRL model has successfully learned the theoretical trade-off logic between the drift and penalty terms inherent in Lyapunov optimization.
- Through trajectory-level attribution analysis, we identify the intrinsic sparsity in multi-user concurrent decision-making. We demonstrate that the agent's optimal actions are primarily driven by bottleneck users rather than the uniform aggregation of all users. This discovery provides an empirical foundation for scalability in massive IoT deployments, suggesting that future decision-making frameworks can potentially decouple decision complexity from network size.

II. System Model and Problem Formulation

As illustrated in Fig. 1, the system consists of one PB, one AP integrated with an MEC server, and K Mobile Devices (MDs), denoted by the set $\mathcal{K} = \{1,2,...,K\}$. The system operates in time slots $t \in \{1,2,...,T\}$, each with duration τ.

A. Channel Model and Energy Harvesting

Let $h_{i,B}^t$ and $h_{i,A}^t$ denote the channel power gains from the PB to the i-th MD and from the i-th MD to the AP in time slot t, respectively. We assume a quasi-static channel model where the channel conditions remain constant within each time slot but vary independently across slots. We define a compound channel gain H_i^t for the backscatter link as:

$$H_i^t = h_{i,B}^t \times h_{i,A}^t, \tag{1}$$

where the indices are sorted such that the channel gains are arranged in descending order without loss of generality. This ordering is assumed to facilitate the Successive Interference Cancellation (SIC) decoding order at the AP.

The MDs employ a dynamic mode selection policy denoted by $x_i^t \in \{0,1\}$. When $x_i^t = 1$, the MD performs backscatter communication. To enhance spectrum efficiency, the system employs NOMA. The i-th MD splits the received RF energy based on a reflection coefficient $\alpha_t \in [0,1]$. A portion $\alpha_t P_t$ is reflected to backscatter data, while the remaining portion $(1-\alpha_t)P_t$ is directed to the energy harvesting circuit.

Adopting a practical non-linear energy harvesting model, the amount of energy harvested by the i-th MD in time slot t is explicitly coupled with the reflection coefficient:

$$E_{i,t}^H = \left(\frac{c_i(1-\alpha_t)P_t h_{i,B}^t + d_i}{(1-\alpha_t)P_t h_{i,B}^t + v_i} - \frac{d_i}{v_i} \right) \tau, \tag{2}$$

where P_t is the transmit power, and c_i, d_i, v_i are circuit-specific constants. This formulation explicitly captures the trade-off where increasing the reflection coefficient for communication directly reduces the energy available for harvesting.

To ensure long-term energy sustainability, we model the energy dynamics using a virtual energy queue Q_i^t:

$$Q_i^{t+1} = \max(Q_i^t - E_{i,t}^H + E_{i,t}^{cons}, 0), \tag{3}$$

where $E_{i,t}^{cons}$ accounts for the circuit energy consumption depending on the mode selection:

$$E_{i,t}^{cons} = x_i^t P_s \tau + (1-x_i^t)\kappa (f_i^t)^3 \tau, \tag{4}$$

where P_s is the circuit power consumption during backscattering, and κ represents the effective switched capacitance of the local processor. In practice, the queue state Q_i^t is maintained computationally by recursively tracking the energy harvesting and consumption at each time slot via Eq. (3).

B. Problem Formulation

The achievable total data throughput, summing both offloading and local computing data, is given by:

$$R_{total}^t = \sum_{x_i^t=1} \tau B \log_2 \left(1 + \frac{\alpha_t P_t H_i^t}{\sigma^2 + I_{inter}} \right) + \sum_{x_i^t=0} \frac{f_i^t \tau}{\phi}, \tag{5}$$

where ϕ denotes the number of CPU cycles required per bit, B is the bandwidth, and σ^2 is the noise power. Our objective is to maximize the long-term average system throughput while ensuring the stability of all energy queues. This optimization problem is formulated as:

$$\mathscr{P}_1 : \max_{\{x^t, f^t, \alpha_t\}} \quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} R_{total}^t \tag{6a}$$

$$\text{s.t.} \quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[Q_i^t] < \infty, \quad \forall i \in \mathcal{K} \tag{6b}$$

$$x_i^t \in \{0,1\}, \quad \forall i \in \mathcal{K} \tag{6c}$$

$$0 \le \alpha_t \le 1, \tag{6d}$$

$$0 \le f_i^t \le f_i^{max}, \tag{6e}$$

where (6b) represents the queue stability constraint ensuring that the average energy consumption does not exceed the harvested energy over time.

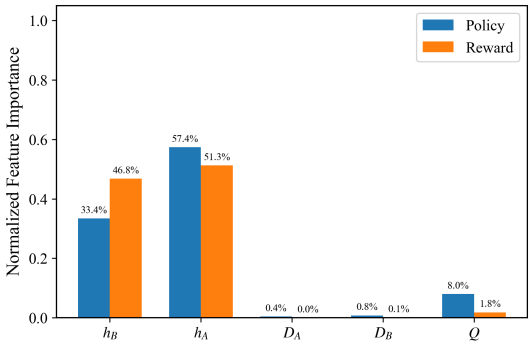


Fig. 2. Shapley Proportion.

III. XAI-Driven Lightweight LyDRL

The optimization problem $\mathscr{P}_1$ is a mixed-integer non-linear programming (MINLP) problem. To solve it efficiently in real-time, we propose a Lightweight framework, which integrates Lyapunov optimization to handle stability constraints and XAI to optimize the DRL state space representation.

A. Lyapunov-Guided DRL Framework

Using Lyapunov optimization theory, we transform the long-term constraint (6b) into a per-slot drift-plus-penalty minimization problem. We define the Lyapunov function as $L(\mathbf{Q}^t) = \frac{1}{2}\sum_{i=1}^{K}(Q_i^t)^2$. The objective in each time slot becomes minimizing the upper bound of the drift-plus-penalty term:

$$\min_{\mathbf{A_t}} \sum_{i=1}^{K} Q_i^t(E_{i,t}^{cons} - E_{i,t}^{H}) - V \cdot R_{total}^t, \tag{7}$$

where V is a control parameter balancing queue stability and throughput. We employ a DRL agent to solve Eq.(7). In the standard LyDRL approach, the agent observes a comprehensive state space $\mathscr{S}_{full}$ to make decisions. The state at time t typically includes CSI, queue lengths, and device geometries:

$$\mathscr{S}_{full}^t = \{\mathbf{h}_B^t, \mathbf{h}_A^t, \mathbf{D}_B^t, \mathbf{D}_A^t, \mathbf{Q}^t\}, \tag{8}$$

where $\mathbf{h}$ represents channel gains and $\mathbf{D}$ represents the physical distances of MDs to the PB and AP.

B. XAI-Driven Redundancy Diagnosis

To identify potential redundancies, we employ *SHapley Additive exPlanations* (SHAP) [16], a game-theoretic approach to interpret model predictions. We pre-train a LyDRL agent using $\mathscr{S}_{full}$ and calculate the global importance value ϕ_j for each feature j. As shown in Fig.2, $0 \approx \phi_{distance} < \phi_{queue} < \phi_{channel}$, which means the explicit distance inputs provide negligible marginal information gain.

C. Proposed Lightweight-LyDRL Algorithm

We prune the state space by removing the redundant distance features, resulting in a compact state representation:

$$\mathscr{S}_{pruned}^t = \{\mathbf{h}_B^t, \mathbf{h}_A^t, \mathbf{Q}^t\}, \tag{9}$$

This reduction decreases the input dimension from $5K$ to $3K$, leading to a significant reduction in the number of parameters in the input layer of the Policy Network (Actor) and Value Network (Critic). The detailed execution flow operating on $\mathscr{S}_{pruned}$ is summarized below:

1) **Action:** The policy network generates continuous offloading probabilities $\hat{\mathbf{x}}^t = \pi_\theta(\mathscr{S}_{pruned}^t)$. These are mapped to binary decisions $\mathbf{x}^t \in \{0,1\}^K$.
2) **Resource Optimization [12]:** For MDs assigned to local computing ($x_i^t = 0$), the optimal CPU frequency is derived analytically:
$$f_i^t = \min\left(\sqrt{\frac{V}{3\kappa\phi Q_i^t}}, f_i^{max}\right), \tag{10}$$
Simultaneously, for offloading MDs ($x_i^t = 1$), the optimal reflection coefficient α_t^* is obtained by solving the following convex sub-problem via binary search:
$$\alpha_t^* = \arg\min_{0\le\alpha\le 1}\left(\sum_{i\in\mathscr{K}_{off}} Q_i^t(P_s\tau - E_{i,t}^H(\alpha)) - V\cdot R_{off}^t(\alpha)\right), \tag{11}$$
where $\mathscr{K}_{off} = \{i | x_i^t = 1\}$ denotes the set of offloading users, and $R_{off}^t(\alpha)$ represents the NOMA backscatter throughput dependent on α.
3) **Reward Calculation:** The reward r_t is defined as:
$$r_t = V \cdot R_{total}^t - \sum_{i=1}^{K} Q_i^t(E_{i,t}^{cons} - E_{i,t}^H). \tag{12}$$
4) **Training:** The policy is updated using experience tuples $(\mathscr{S}_{pruned}^t, \hat{\mathbf{x}}^t, r_t, \mathscr{S}_{pruned}^{t+1})$ stored in the replay buffer.

By removing redundant geometric inputs, the Lightweight-LyDRL reduces computational overhead and inference time, making it highly suitable for deployment on edge devices with limited processing capabilities, without compromising the stability guarantees provided by the Lyapunov optimization. Comprising 4,394 parameters in 32-bit floating-point format (4 bytes/parameter), the 3f model utilizes merely 17.1 KB.

IV. Simulation Results and Discussions

A. Parameter Settings

The scenario is similar with [12], including $N = 10$ MDs, a AP located at $P_A = (0,0,10)$, and a PB positioned at $P_B = (10,10,5)$. MDs move randomly within a bounded region $[-2.5,2.5]\times[-2.5,2.5]$. The total simulation consists of 500 time frames, divided into episodes of length 50 frames each. At the beginning of each episode, MD positions are randomly reinitialized and virtual energy queues are reset to 0.

The wireless channel follows a distance-dependent path loss model with $\beta_0 = 0.01$ and path loss exponent 2.5. The nonlinear energy harvesting model parameters are set to $c = 2.463$, $d = 1.635$, and $v = 0.826$ [17]. The transmit power of PB is $P_t = 10$ mW, and the circuit power during offloading is $P_s = 0.01$ mW. For local computation, the maximum CPU frequency is $f_{\max} = 300$ MHz with computation intensity $\phi = 100$ cycles/bit and effective capacitance $\kappa = 10^{-26}$ [18]. The communication bandwidth is $B = 2$ MHz with noise power $N_0 = 10^{-17.4}$ W. The Lyapunov parameter is set to $V = 10$ to investigate the trade-off between queue stability and computation rate.

The memory-augmented DNN adopts a three-layer architecture $[50, 64, 32, 10]$, using Adam optimizer with learning rate $\eta = 0.01$, batch size 128, and experience replay buffer capacity 1024. The training update interval is set to 1 frame.

To quantify how different state features influence both the policy decisions and the optimization objective, we first conduct a global interpretability study under a fixed system configuration. Specifically, we set the transmit power to $P_t = 5$ mW and the maximum local computing frequency to $f_{\max} = 300$ MHz. In this experiment, we adopt a five-feature state representation $\mathscr{S}^t_{full}$. We compute Shapley values from two complementary perspectives: (i) *policy-based Shapley*, which attributes the contribution of each input feature to the policy output (i.e., offloading probability) and (ii) *reward-based Shapley*, which attributes the contribution to the Lyapunov-driven objective used for action selection. For each feature type, we aggregate Shapley attributions over time using the mean absolute value, and further normalize them to obtain global contribution proportions.

Fig. 2 reports the resulting global feature importance under the fixed configuration. The distance-related features $\mathbf{D}^t_A$ and $\mathbf{D}^t_B$ contribute negligibly in both the policy and reward views: their normalized proportions are below 0.5%. Channel gain-related features $\mathbf{h}^t_B$ and $\mathbf{h}^t_A$ dominate the feature importance, accounting for 33.4% and 57.4% in the policy-based analysis, and 46.8% and 51.3% in the reward-based analysis, respectively. The queue feature $\mathbf{Q}^t$ contributes 8.0% in the policy-based view and 1.8% in the reward-based view, reflecting its auxiliary role in balancing energy queue stability within the Lyapunov optimization framework.

To validate the generalizability of our interpretability findings and examine the stability of feature importance under varying operational conditions, we conduct a systematic parameter sensitivity study. Specifically, we vary the transmit power $P_t \in \{1, 2, \ldots, 10\}$ mW and the maximum local computing frequency $f_{\max} \in \{200, 250, \ldots, 500\}$ MHz, which jointly define the system's energy harvesting capacity and local computation capability. For each configuration, we train an independent LyDRL policy and compute the normalized Shapley-based feature importance from both policy and reward perspectives. Fig. 3 and Fig. 4 reveal that across all parameter configurations, the distance features $\mathbf{D}_A$ and $\mathbf{D}_B$ remain around or below 1% in most cases and are consistently much smaller than the channel and queue features in both policy decisions and reward optimization. This pattern across diverse operating regimes—ranging from resource-constrained (low P_t, low $f_{\max}$) to resource-abundant (high P_t, high $f_{\max}$) scenarios—supports that distance inputs are largely redundant given the presence of channel gain features. The learned policies universally prioritize channel measurements over explicit distance information, confirming that the strong functional dependency induced by the path loss model enables implicit spatial reasoning through channel gains alone.

In addition, as Fig. 3 and Fig. 4 shown, in resource-constrained regimes (e.g., $P_t \leq 3$ mW or $f_{\max} \leq 300$ MHz), the queue feature $\mathbf{Q}$ maintains substantial importance (10-

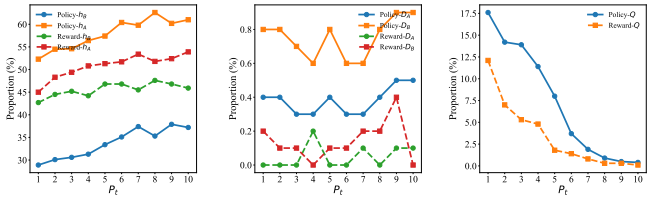


Fig. 3. Feature importance across transmit power P_t. Distance features remain negligible across all regimes, while queue importance decays as energy constraints relax.

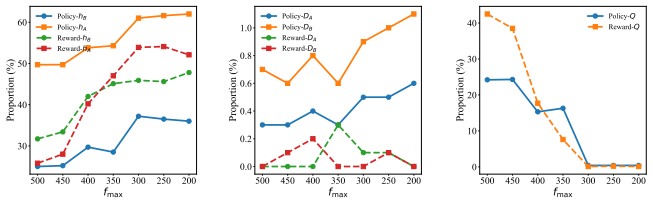


Fig. 4. Feature importance across maximum frequency $f_{\max}$. Queue feature exhibits sharp decay and convergence toward channel-dominated attribution at high computing capacity.

40%), reflecting the policy's need to balance immediate performance against long-term energy queue stability. As constraints relax with increasing P_t or $f_{\max}$, the attribution progressively converges toward a stable distribution dominated by channel gains $\mathbf{h}_B$ and $\mathbf{h}_A$ (together exceeding 90%), while $\mathbf{Q}$ diminishes to near-zero. This convergence pattern indicates that the learned policy adaptively transitions from a queue-aware energy management strategy in constrained regimes to a channel-quality-driven offloading strategy in abundant regimes. Notably, the rate of convergence differs across parameters: queue importance decays more sharply with respect to $f_{\max}$ than P_t, suggesting that local computing capacity has a more pronounced impact on decoupling queue dynamics from decision-making. The stabilization of feature importance distributions at high resource levels implies that the policy reaches a regime where wireless link quality becomes the sole bottleneck, and further resource augmentation yields diminishing returns in altering decision strategies.

To empirically validate the redundancy of distance features identified through Shapley analysis, we remove $\mathbf{D}_A$ and $\mathbf{D}_B$ from the state representation. Specifically, we compare the original five-feature state $\mathscr{S}_{full}$ (denoted as 5f, based on [12]) with a reduced three-feature state $\mathscr{S}_{pruned}$ (denoted as 3f), while maintaining identical network architecture, training protocol, and evaluation settings. This experiment serves to verify whether the negligible Shapley attributions of distance features translate into negligible performance degradation when these features are entirely removed. Fig. 5 presents the global Shapley proportions under the 3f configuration at $P_t = 5$ mW and $f_{\max} = 300$ MHz. The attribution distribution remains qualitatively consistent with the 5f case: channel gains $\mathbf{h}_B$ and $\mathbf{h}_A$ jointly dominate both policy-based (90.9%) and reward-based (96.9%) attributions, while the queue feature $\mathbf{Q}$ contributes moderately (9.1% and 3.1%, respectively). Moreover, the sensitivity analysis in Fig. 6 demonstrates that the feature im-

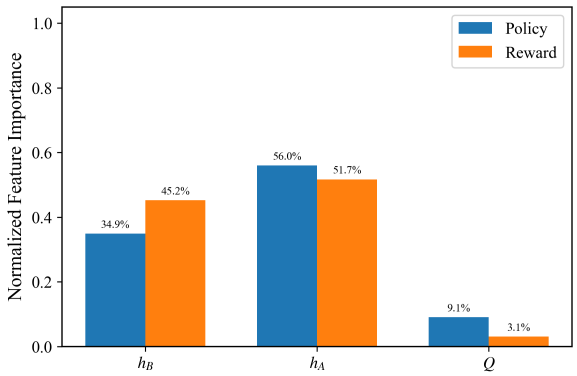


Fig. 5. Global Shapley proportions under 3f configuration. Channel gains dominate attribution, consistent with the 5f case despite the absence of distance inputs.

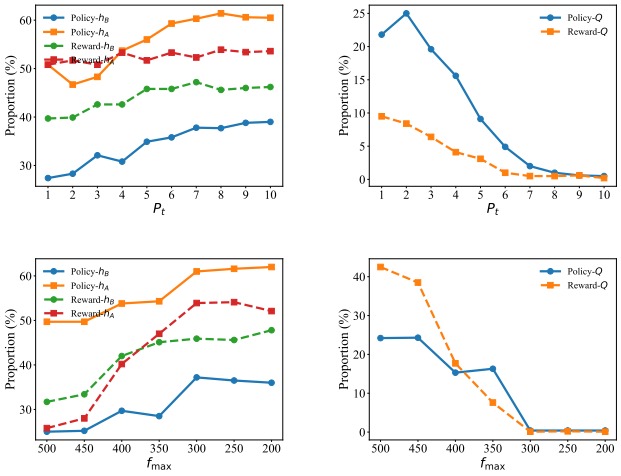


Fig. 6. Feature importance sensitivity under 3f configuration across P_t and $f_{\max}$. Attribution trends mirror the 5f case, confirming preserved decision mechanisms.

portance trends under 3f mirror those observed under 5f across varying P_t and $f_{\max}$: queue importance decays as constraints relax, while channel gains progressively concentrate attribution mass. This structural consistency confirms that the learned decision mechanism—characterized by channel-quality-driven offloading in resource-abundant regimes and queue-aware energy management in constrained regimes—remains intact after removing distance inputs.

Table. I quantifies the performance comparison between 3f and 5f configurations. The reduced state achieves a computation rate of 5.039 Mbps, which is very close to the 5.061 Mbps achieved by the full-state model, corresponding to only a marginal 0.4% difference. Energy-related metrics exhibit similar parity: harvested energy and consumed energy remain close across both configurations. These results empirically substantiate that explicit distance features provide no measurable performance benefit, corroborating the Shapley-based redundancy diagnosis.

While the preceding global analysis aggregates feature importance across all users and time frames, it obscures the heterogeneity in how individual users contribute to decision-

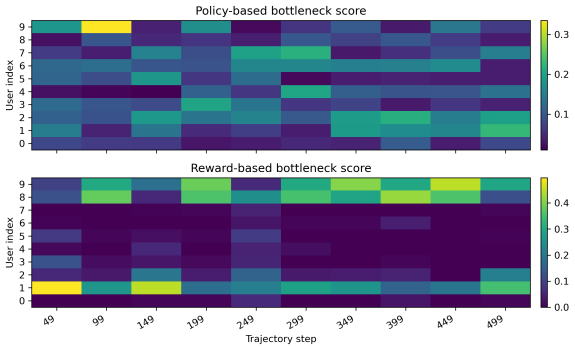
TABLE I
COMPARISON OF METRICS FOR 3F AND 5F MODELS

Method	Rate (Mbps)	EH (J)	Energy (J)
3f	5.039	0.160	0.192
5f	**5.061**	0.151	0.192

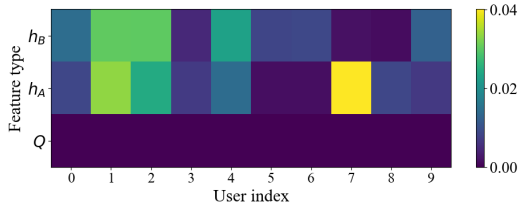
making at specific time instances. To unveil this user-level heterogeneity, we conduct a trajectory-level Shapley analysis on a representative episode segment. Specifically, we select a stable trajectory window and compute the mean absolute Shapley values for each user-feature pair under both policy-based and reward-based perspectives. To further verify that the bottleneck-user observation is not limited to a single selected segment, we extend the analysis to multiple trajectory windows. Specifically, we use the saved Shapley values at steps 49, 99, 149, 199, 249, 299, 349, 399, 449, and 499, and compute a user-level bottleneck score by aggregating the mean absolute Shapley contributions of $\mathbf{h}_B$, $\mathbf{h}_A$, and $\mathbf{Q}$. Fig. 7 presents (a) the full-trajectory bottleneck-user heatmap over multiple trajectory windows and (b) the corresponding per-user attribution heatmaps averaged over the segment.

The local attributions reveal a pronounced user-level heterogeneity in feature importance. As illustrated in the heatmaps, only a subset of users (e.g., users 1, 7, 8, and 9) exhibit strong channel-gain attributions ($\mathbf{h}_B$ and $\mathbf{h}_A$), while the majority of users display near-zero contributions during this segment. Comparing the representative segment-level heatmaps with the full-trajectory bottleneck map, the high-attribution users are not limited to a single selected segment. Instead, several users repeatedly exhibit large bottleneck scores across different trajectory windows, indicating that the learned policy consistently focuses on a small subset of influential users. This user-level selectivity indicates that, within a local decision context, the learned policy prioritizes link-quality information from a few bottleneck users whose channel states are most informative for optimizing the global offloading strategy, rather than treating all users uniformly. In contrast, users with stable or favorable channel conditions contribute minimally to the marginal decision value, as their offloading outcomes are relatively predictable.

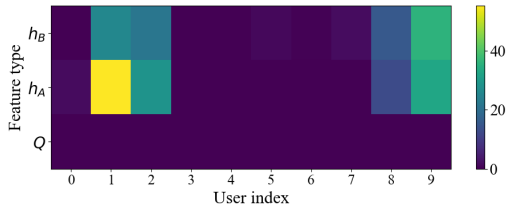
Moreover, the queue feature $\mathbf{Q}$ exhibits uniformly low attribution across all users in both policy and reward views, consistent with the global findings that queue dynamics exert limited influence under resource-relaxed regimes. This suggests that, in the examined trajectory segment where energy constraints are not binding, the policy's attention shifts predominantly toward wireless link quality rather than energy queue management. Taken together, this trajectory-level analysis complements the global interpretability results by elucidating *which users* and *when along the trajectory* channel features dominate the decision process, thereby providing actionable insights into the temporally adaptive nature of the learned offloading policy.



(a) Full-trajectory bottleneck-user heatmap



(i) Policy Shapley Values



(ii) Reward Shapley Values

(b) Shapley heatmaps

Fig. 7. Full-trajectory bottleneck-user analysis and representative segment-level Shapley heatmaps for policy and reward functions.

V. Conclusion

In this paper, we developed an XAI-driven LyDRL framework for EH-assisted Backscatter MEC networks. By integrating LyDRL with Shapley value analysis, we provided a transparent view into the decision-making logic of DRL agents, yielding three critical insights for future edge AI design. First, we empirically verified that DRL agents implicitly internalize path-loss physics, rendering explicit geometric distance features redundant. Pruning these features reduced the state space by 40% without compromising stability or throughput. Second, we revealed a regime-dependent attention mechanism where the policy dynamically shifts its focus from queue stability to channel quality as resource constraints relax, confirming the agent's alignment with theoretical Lyapunov drift-plus-penalty logic. Third, our trajectory-level analysis uncovered an intrinsic state sparsity in multi-user scenarios, where optimal decisions are primarily driven by bottleneck users. These findings demonstrate that XAI can serve as a powerful tool for model compression and logic validation, paving the way for deploying scalable, explainable, and energy-efficient intelligence in IoT edge devices.

Acknowledgment

This work is supported by Engineering and Physical Sciences Research Council grant (No. EP/X040518/1 and EP/Y037421/1).

References

[1] K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6g: Vision, enabling technologies, and applications," *IEEE journal on selected areas in communications*, vol. 40, no. 1, pp. 5–36, 2021.

[2] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, 2019.

[3] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE transactions on communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[4] Y. Wang, H. Wu, R. H. Jhaveri, and Y. Djenouri, "Drl-based urllc-constraint and energy-efficient task offloading for internet of health things," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 6, pp. 3305–3316, 2024.

[5] Y. Xu, R. Xu, D. Li, G. Yang, G. Wang, C. Yuen, and J. Zhou, "Robust resource allocation for wireless-powered backscatter communication systems with noma," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 12 288–12 299, 2023.

[6] M. Ahmed, M. Shahwar, F. Khan, W. U. Khan, A. Ihsan, U. S. Khan, F. Xu, and S. Chatzinotas, "Noma-based backscatter communications: Fundamentals, applications, and advancements," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19 303–19 327, 2024.

[7] S. Mondal, D. Bepari, A. Chandra, K. Singh, C.-P. Li, and Z. Ding, "A comprehensive survey on noma-based backscatter communication for iot applications," *IEEE Internet of Things Journal*, 2025.

[8] N. Mazhar, S. A. Ullah, S. Basheer, H. Jung, M. S. J. Solaija, A. Mahmood, M. Gidlund, and S. A. Hassan, "Towards trustworthy and fresh data delivery in 6g iot: A drl-aided cognitive noma and backscatter framework," *IEEE Internet of Things Journal*, pp. 1–1, 2025.

[9] Y.-H. Hsu, J.-I. Lee, and C.-H. Lee, "A drl-based high-altitude platform transmission and energy harvesting scheduling scheme for 6g noma sagins," *IEEE Transactions on Cognitive Communications and Networking*, vol. 12, pp. 3036–3046, 2026.

[10] H. Wu, J. Chen, T. N. Nguyen, and H. Tang, "Lyapunov-guided delay-aware energy efficient offloading in iiot-mec systems," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2117–2128, 2023.

[11] Y. Liang, H. Tang, H. Wu, Y. Wang, and P. Jiao, "Lyapunov-guided offloading optimization based on soft actor-critic for isac-aided internet of vehicles," *IEEE Transactions on Mobile Computing*, vol. 23, no. 12, pp. 14 708–14 721, 2024.

[12] H. Wu, J. Du, H. Tang, P. Jiao, and R. Li, "A lyapunov-guided task offloading approach for backscatter communication assisted edge computing," in *IEEE INFOCOM 2025 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2025, pp. 1–6.

[13] A. K. Gizzini, Y. Medjahdi, A. J. Ghandour, and L. Clavier, "Towards explainable ai for channel estimation in wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 5, pp. 7389–7394, 2024.

[14] W. Guo, "Explainable artificial intelligence for 6g: Improving trust between human and machine," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 39–45, 2020.

[15] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan *et al.*, "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM computing surveys*, vol. 55, no. 9, pp. 1–33, 2023.

[16] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 4768–4777.

[17] Y. Chen, N. Zhao, and M.-S. Alouini, "Wireless energy harvesting using signals from multiple fading channels," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 5027–5039, 2017.

[18] X. Hu, K.-K. Wong, and K. Yang, "Wireless powered cooperation-assisted mobile edge computing," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2375–2388, 2018.