

Article

Not peer-reviewed version

---

# Federated Multi-Agent Deep Reinforcement Learning for Joint Channel Selection and Power Control in Cognitive Radio Networks

---

[Zihan Long](#)<sup>\*</sup> and Mingrui Rao

Posted Date: 9 April 2026

doi: 10.20944/preprints202604.0662.v1

Keywords: cognitive radio networks; multi-agent deep reinforcement learning; federated learning; dynamic spectrum access; graph neural networks



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Federated Multi-Agent Deep Reinforcement Learning for Joint Channel Selection and Power Control in Cognitive Radio Networks

Zihan Long \* and Mingrui Rao

Minia University

\* Correspondence: 81145076@fci.s-mu.edu.eg

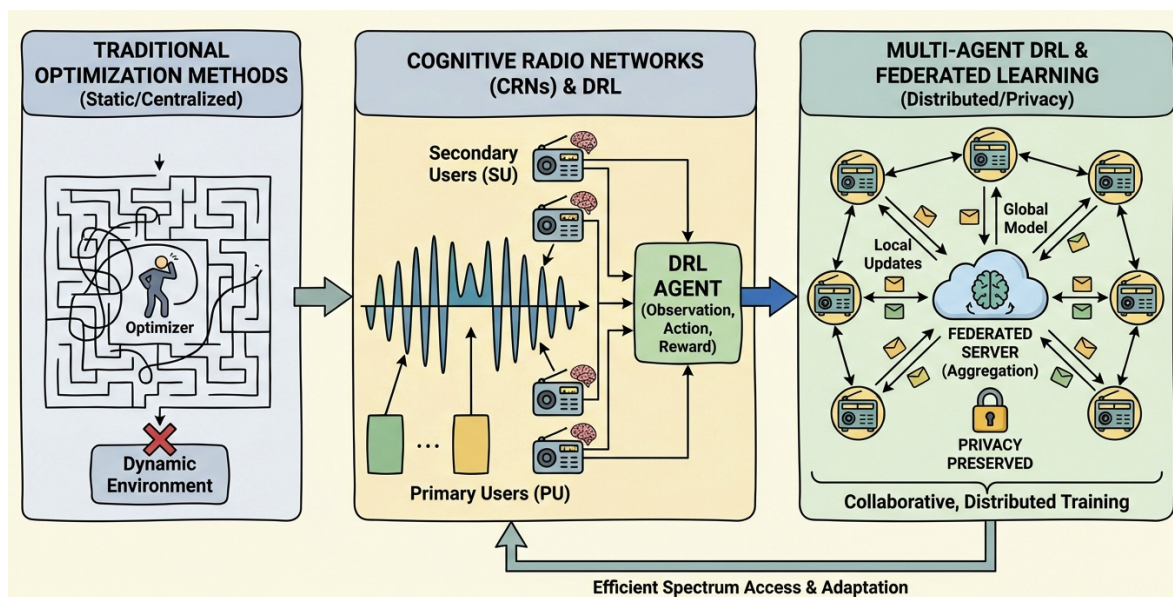
## Abstract

Cognitive radio networks (CRNs) face significant challenges in dynamic spectrum access due to the complex interactions among multiple secondary users, sparse reward signals, and poor cross-domain generalization. Existing approaches, ranging from traditional optimization to single-agent deep reinforcement learning (DRL), struggle to balance spectral efficiency, collision avoidance, and adaptability in heterogeneous wireless environments. In this paper, we propose FedMA-DRL, a federated multi-agent deep reinforcement learning framework that integrates centralized training with decentralized execution (CTDE), graph neural network (GNN)-augmented Q-value prediction, age-aware federated aggregation (FedAge), and attention-based domain adaptation for joint channel selection and power control in CRNs. The GNN module captures topological relationships among secondary users through attention-weighted message passing on the interference graph, while the FedAge strategy enables privacy-preserving knowledge sharing with staleness-aware weighting. Extensive experiments on a CRN testbed with 10 PU channels and 15 heterogeneous SUs demonstrate that FedMA-DRL achieves 14.87 Mbps SU throughput, 0.038 collision probability, 4.35 bits/Joule energy efficiency, and 6.23 bits/s/Hz spectrum efficiency, outperforming existing methods including R2D2 and C-DRL. Ablation studies and cross-domain evaluations further confirm the effectiveness of each proposed component.

**Keywords:** cognitive radio networks; multi-agent deep reinforcement learning; federated learning; dynamic spectrum access; graph neural networks

## 1. Introduction

The exponential growth of wireless communication services and the proliferation of Internet of Things (IoT) devices have imposed unprecedented demands on limited radio spectrum resources [1]. Cognitive radio networks (CRNs) have emerged as a promising paradigm to address spectrum scarcity by enabling dynamic spectrum access (DSA), allowing unlicensed secondary users (SUs) to opportunistically access licensed frequency bands without causing harmful interference to primary users (PUs) [2]. Effective resource management in CRNs—including joint channel selection, power control, and spectrum allocation—is critical for maximizing spectral efficiency while ensuring quality of service for both primary and secondary users [3].



**Figure 1.** Overview of the evolution from traditional optimization to federated multi-agent deep reinforcement learning for cognitive radio resource management, highlighting the key challenges and the proposed FedMA-DRL framework.

Traditional optimization-based approaches for spectrum management in CRNs, such as convex optimization and game-theoretic formulations, have demonstrated their utility in controlled settings [4]. However, these methods suffer from several fundamental limitations: (1) they require accurate and instantaneous channel state information (CSI), which is often unavailable in rapidly varying wireless environments; (2) the computational complexity scales exponentially with the number of users and channels, rendering real-time decision-making infeasible; and (3) they fail to adapt to the dynamic, partially observable nature of real-world CRN scenarios [5]. The transition from fog IoT spectrum management via traditional optimization to deep neural network (DNN)-based resource allocation marked an important step toward data-driven solutions, yet single-agent approaches still lack the ability to model the complex interactions among multiple cognitive users [6,7].

Deep reinforcement learning (DRL) has recently attracted significant attention as a powerful framework for autonomous decision-making in complex, dynamic environments [8]. Multi-agent DRL (MADRL) further extends this capability by enabling distributed decision-making among multiple SUs, where each agent learns to cooperate or compete in a shared spectrum environment [9]. The centralized training with decentralized execution (CTDE) architecture has shown particular promise, allowing agents to leverage global information during training while making independent decisions during deployment [10]. Despite these advances, several challenges remain: sparse reward signals in spectrum access tasks hinder effective learning, exploration inefficiency leads to suboptimal policies, and single-domain training results in poor generalization across heterogeneous wireless environments [11]. Notably, attention mechanisms have shown remarkable effectiveness across diverse domains, from medical image segmentation [12] to speech enhancement [13,14], motivating their application in cross-domain wireless adaptation.

Federated learning (FL) offers a privacy-preserving paradigm for distributed model training, enabling multiple devices to collaboratively learn a shared model without exchanging raw data [15]. Integrating FL with DRL in CRNs can potentially combine the adaptive decision-making capability of DRL with the privacy protection and communication efficiency of FL. Recent studies have explored federated DRL for dynamic spectrum access, demonstrating improved convergence and stability compared to traditional federated approaches [16]. However, existing federated DRL methods often overlook the structural relationships among wireless devices and fail to effectively address cross-domain generalization.

In this paper, we propose **FedMA-DRL**, a Federated Multi-Agent Deep Reinforcement Learning framework for joint channel selection and power control in cognitive radio networks. Our approach integrates four key innovations: (1) a CTDE-based multi-agent architecture that models the spectrum access problem as a multi-agent Markov decision process (MAMDP) with nonlinear reward shaping to address sparse rewards; (2) a graph neural network (GNN)-augmented Q-value predictor that captures the topological relationships among SUs and environmental features; (3) a federated learning mechanism with an age-aware aggregation strategy (FedAge) for privacy-preserving knowledge sharing across devices; and (4) an attention-based domain adaptation module that enhances cross-domain generalization by aligning feature distributions across heterogeneous wireless environments.

We evaluate FedMA-DRL on a realistic CRN testbed comprising 10 PU channels and 15 heterogeneous SUs. Extensive experiments demonstrate that FedMA-DRL achieves a secondary user throughput of **14.87 Mbps**, outperforming R2D2 (13.92 Mbps) and C-DRL (12.34 Mbps). The collision probability is reduced to **0.038**, significantly lower than competing methods. FedMA-DRL also achieves the highest energy efficiency (4.35 bits/Joule) and spectrum efficiency (6.23 bits/s/Hz), confirming the effectiveness of our proposed components.

Our main contributions are summarized as follows:

- We propose FedMA-DRL, a federated multi-agent DRL framework that integrates CTDE architecture with nonlinear reward shaping and action-guided exploration for efficient joint channel selection and power control in CRNs.
- We introduce a GNN-augmented Q-value predictor that leverages the topological structure of wireless devices to improve prediction accuracy and a FedAge-based federated aggregation strategy for privacy-preserving distributed learning.
- We develop an attention-based domain adaptation module that enhances cross-domain generalization, enabling robust performance across heterogeneous wireless environments without requiring domain-specific retraining.

## 2. Related Work

### 2.1. Deep Reinforcement Learning for Dynamic Spectrum Access

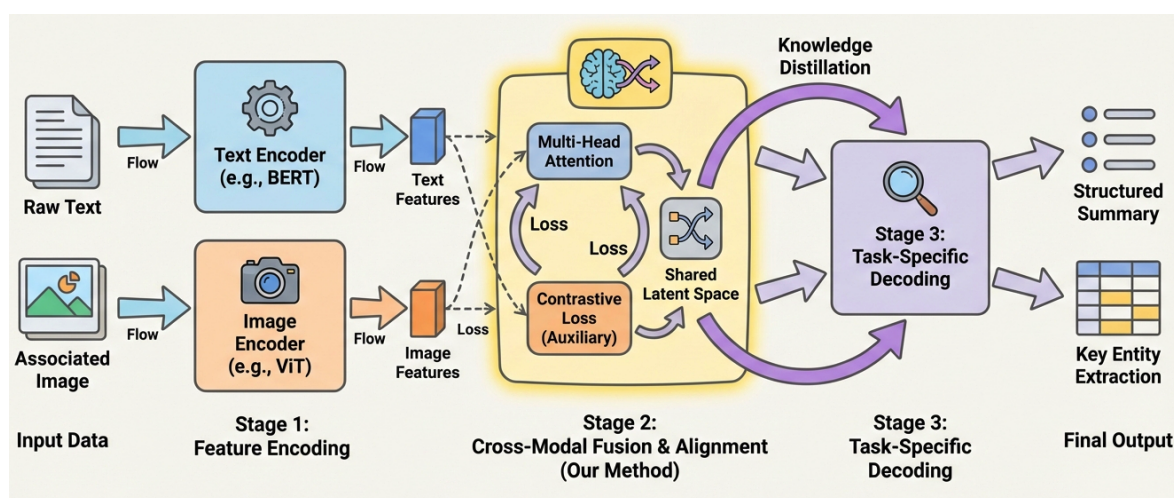
Deep reinforcement learning has emerged as a powerful paradigm for addressing the dynamic spectrum access problem in cognitive radio networks. Kumar and Kumar [17] provided a comprehensive review of RL-based approaches for DSA, evaluating algorithms ranging from Q-learning to deep policy gradient methods such as DDPG in CRN scenarios. Xu et al. [18] proposed a DRL-based scheme for DSA networks that enables SUs to opportunistically access PU spectrum with minimal interference, demonstrating significant throughput improvements over traditional sensing-based approaches. The application of DQN and QR-DQN for television whitespace CRNs was explored by Ukpong et al. [19], who achieved interference avoidance rates exceeding 96% through DRL-based channel selection. Malhotra [20] conducted a comparative study of PPO, DQN, and R2D2 for wireless resource allocation, finding that R2D2 offers the best stability and convergence due to its recurrent architecture and prioritized experience replay. Bai et al. [21] proposed MA-JCSPC, a multi-agent DRL method with CTDE architecture and nonlinear reward shaping for joint channel selection and power control, achieving up to 9.1% throughput improvement over centralized DRL baselines. Li and Wang [22] developed a decentralized multi-agent deep reinforcement learning approach formulating spectrum access as a stochastic game among distributed learning agents. Venkatesan and Kumaratharan [23] introduced a hierarchical multi-agent RL controller with graph-driven occupancy prediction and federated meta-reinforcement learning for efficient CRN management. These works collectively demonstrate the progression from single-agent to multi-agent DRL in CRNs, yet they often overlook the benefits of federated knowledge sharing and structured device relationship modeling.

## 2.2. Federated Learning and Graph Neural Networks in Wireless Networks

Federated learning and graph neural networks have gained significant traction for distributed, privacy-preserving wireless resource management. Aggarwal and Gupta [15] presented a comprehensive review of FL methods and their privacy-preserving mechanisms, highlighting the challenges of heterogeneous data distributions and communication efficiency in distributed training. Li et al. [16] proposed a dynamic spectrum access scheme for IoT that integrates federated learning with GNN and DQN, where GNN captures device relationships for Q-value prediction and FedAge coordinates multi-device knowledge sharing. Shah and Ali [24] introduced a federated fog computing framework for IoT resource allocation, combining predictive scheduling with energy-aware resource allocation and adaptive mobility management. The intersection of FL and GNN was explored by Wu et al. [25], who developed privacy-preserving federated GNN with local differential privacy for graph-structured data analysis. He et al. [26] proposed a federated GNN framework with privacy-preserving graph expansion protocols for incorporating high-order interaction information. For NOMA networks, Bandy [27] combined DRL with hybrid optimization algorithms for RIS-assisted user clustering and phase shift optimization, achieving substantial rate improvements. Yang et al. [6] laid the foundation for fog IoT spectrum management, while their subsequent work [7] introduced DNN-based resource allocation in NOMA networks and [28] extended this to partially observable multi-agent deep RL for cognitive resource management. Similarly, adaptive attention selection mechanisms have proven effective in other signal processing domains, such as visual speech enhancement [29] and adaptive local/non-local attention for speech enhancement [14]. Despite these advances, existing federated DRL methods often fail to explicitly model the topological relationships among wireless devices or address cross-domain generalization, which our work specifically targets through GNN-augmented Q-value prediction and attention-based domain adaptation.

## 3. Method

In this section, we present the proposed **FedMA-DRL** framework, a Federated Multi-Agent Deep Reinforcement Learning approach for joint channel selection and power control in cognitive radio networks. FedMA-DRL integrates four core modules: a CTDE-based multi-agent architecture with nonlinear reward shaping, a graph neural network (GNN)-augmented Q-value predictor, a federated aggregation mechanism with age-aware weighting (FedAge), and an attention-based domain adaptation module. The overall architecture is illustrated in Figure 2.



**Figure 2.** Overview of the proposed FedMA-DRL framework, illustrating the CTDE-based multi-agent architecture with GNN-augmented Q-value prediction, FedAge federated aggregation, and attention-based domain adaptation for joint channel selection and power control in cognitive radio networks.

### 3.1. Problem Formulation

We model the joint channel selection and power control problem in CRNs as a multi-agent Markov decision process (MAMDP). Consider a CRN with  $N$  secondary users (SUs) and  $M$  primary user (PU) channels. Each SU  $i$  at time step  $t$  observes a local state  $\mathbf{o}_i^t$  comprising the channel availability vector  $\mathbf{a}^t \in \{0, 1\}^M$ , the previous transmission outcome  $r_i^{t-1}$ , and the interference level  $\mathbf{I}_i^t$ . The action of SU  $i$  is defined as  $\mathbf{u}_i^t = (c_i^t, p_i^t)$ , where  $c_i^t \in \{1, \dots, M\}$  denotes the selected channel and  $p_i^t \in [0, P_{\max}]$  denotes the transmit power level.

The joint objective is to maximize the long-term cumulative reward across all SUs:

$$\max_{\pi} \mathbb{E} \left[ \sum_{t=0}^T \gamma^t \sum_{i=1}^N R_i^t(\mathbf{s}^t, \mathbf{u}^t) \right] \quad (1)$$

where  $\gamma \in (0, 1)$  is the discount factor,  $\pi = \{\pi_1, \dots, \pi_N\}$  is the joint policy, and  $R_i^t$  is the reward function for SU  $i$ .

### 3.2. Nonlinear Reward Shaping

To address the sparse reward problem inherent in spectrum access tasks, we design a nonlinear reward function that incorporates both performance incentives and constraint penalties:

$$R_i^t = \alpha_1 \cdot \log(1 + \text{SINR}_i^t) - \alpha_2 \cdot \mathbb{1}[\text{collision}_i^t] \cdot P_{\text{coll}} - \alpha_3 \cdot \frac{p_i^t}{P_{\max}} \quad (2)$$

where  $\text{SINR}_i^t$  is the signal-to-interference-plus-noise ratio,  $\mathbb{1}[\text{collision}_i^t]$  is an indicator for PU collision events,  $P_{\text{coll}}$  is the collision penalty, and  $\alpha_1, \alpha_2, \alpha_3$  are weighting coefficients. The logarithmic transformation of SINR provides diminishing returns at high signal quality, encouraging more balanced spectrum utilization across SUs.

Furthermore, we introduce an action-guided initial exploration strategy. During the initial training phase, each SU's action selection is biased toward feasible actions derived from spectrum sensing results:

$$\pi_i^{(0)}(c_i^t | \mathbf{o}_i^t) = \frac{\exp(\beta \cdot \mathbb{1}[c_i^t \text{ idle}])}{\sum_{c=1}^M \exp(\beta \cdot \mathbb{1}[c \text{ idle}])} \quad (3)$$

where  $\beta$  is a temperature parameter that controls the exploration bias strength, gradually annealed as training progresses.

### 3.3. GNN-Augmented Q-Value Predictor

We employ a graph neural network to capture the topological relationships among SUs and enhance Q-value prediction. Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  represent the interference graph, where each node  $v_i \in \mathcal{V}$  corresponds to SU  $i$  and an edge  $(v_i, v_j) \in \mathcal{E}$  exists if SUs  $i$  and  $j$  can potentially interfere.

For each SU  $i$ , the initial node feature is constructed from the local observation:

$$\mathbf{h}_i^{(0)} = f_{\text{enc}}(\mathbf{o}_i^t) \quad (4)$$

where  $f_{\text{enc}}$  is a feature encoding network consisting of fully connected layers with ReLU activation.

The GNN performs  $K$  rounds of message passing to aggregate neighborhood information:

$$\mathbf{h}_i^{(k)} = \sigma(\mathbf{W}^{(k)} \cdot \text{AGG}(\{\mathbf{h}_j^{(k-1)} : j \in \mathcal{N}(i)\}) + \mathbf{b}^{(k)}) \quad (5)$$

where  $\mathcal{N}(i)$  denotes the neighbors of node  $i$ , AGG is an attention-weighted aggregation function, and  $\sigma$  is the ELU activation function.

The attention-weighted aggregation is defined as:

$$\alpha_{ij}^{(k)} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^{(k)T}[\mathbf{h}_i^{(k-1)} \parallel \mathbf{h}_j^{(k-1)}]\right)\right)}{\sum_{j' \in \mathcal{N}(i)} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^{(k)T}[\mathbf{h}_i^{(k-1)} \parallel \mathbf{h}_{j'}^{(k-1)}]\right)\right)} \quad (6)$$

$$\text{AGG}\left(\{\mathbf{h}_j^{(k-1)} : j \in \mathcal{N}(i)\}\right) = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(k)} \cdot \mathbf{h}_j^{(k-1)} \quad (7)$$

The final Q-value for SU  $i$  given action  $\mathbf{u}_i^t$  is computed as:

$$Q_i(\mathbf{o}_i^t, \mathbf{u}_i^t) = f_{\text{out}}\left(\mathbf{h}_i^{(K)} \parallel \mathbf{u}_i^t\right) \quad (8)$$

where  $f_{\text{out}}$  is the output network mapping the concatenated GNN feature and action to a scalar Q-value.

### 3.4. Federated Aggregation with FedAge

To enable privacy-preserving knowledge sharing across distributed SUs, we adopt a federated learning framework with an age-aware aggregation strategy (FedAge). In each communication round  $r$ , each SU  $i$  trains its local model parameters  $\theta_i^{(r)}$  using local experience replay buffers and sends the model update  $\Delta\theta_i^{(r)} = \theta_i^{(r)} - \theta^{(r-1)}$  to the central server.

The server aggregates the updates using age-aware weights:

$$\theta^{(r)} = \theta^{(r-1)} + \sum_{i=1}^N w_i^{(r)} \cdot \Delta\theta_i^{(r)} \quad (9)$$

where the age-aware weight  $w_i^{(r)}$  is computed as:

$$w_i^{(r)} = \frac{\exp(-\lambda \cdot \text{age}_i^{(r)}) \cdot \|\Delta\theta_i^{(r)}\|^{-1}}{\sum_{j=1}^N \exp(-\lambda \cdot \text{age}_j^{(r)}) \cdot \|\Delta\theta_j^{(r)}\|^{-1}} \quad (10)$$

Here,  $\text{age}_i^{(r)}$  measures the staleness of SU  $i$ 's update (i.e., the number of rounds since the last contribution),  $\lambda$  is a decay parameter, and the inverse norm term penalizes excessively large updates that may indicate instability.

### 3.5. Attention-Based Domain Adaptation Module

To enhance cross-domain generalization, we introduce an attention-based domain adaptation module that aligns feature distributions across heterogeneous wireless environments. Given features  $\mathbf{f}_i^{(s)}$  from source domain  $s$  and  $\mathbf{f}_i^{(t)}$  from target domain  $t$ , the module computes cross-domain attention:

$$\mathbf{A}_{s \rightarrow t} = \text{softmax}\left(\frac{(\mathbf{W}_Q \mathbf{f}^{(t)})(\mathbf{W}_K \mathbf{f}^{(s)})^T}{\sqrt{d_k}}\right) \quad (11)$$

$$\hat{\mathbf{f}}^{(t)} = \mathbf{A}_{s \rightarrow t} (\mathbf{W}_V \mathbf{f}^{(s)}) \quad (12)$$

The adapted target feature is obtained by residual connection and layer normalization:

$$\tilde{\mathbf{f}}^{(t)} = \text{LayerNorm}\left(\mathbf{f}^{(t)} + \hat{\mathbf{f}}^{(t)}\right) \quad (13)$$

A domain discriminator  $D_\phi$  is trained adversarially to classify the domain origin, while the feature encoder is trained to confuse the discriminator, encouraging domain-invariant representations:

$$\mathcal{L}_{\text{adv}} = -\mathbb{E}_s[\log D_\phi(\tilde{\mathbf{f}}^{(s)})] - \mathbb{E}_t[\log(1 - D_\phi(\tilde{\mathbf{f}}^{(t)}))] \quad (14)$$

### 3.6. Overall Training Procedure

The overall training objective combines the DRL loss, federated regularization, and domain adaptation loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{DRL}} + \eta_1 \mathcal{L}_{\text{FedReg}} + \eta_2 \mathcal{L}_{\text{adv}} \quad (15)$$

where  $\mathcal{L}_{\text{DRL}}$  is the standard DQN loss with target network,  $\mathcal{L}_{\text{FedReg}} = \|\theta_i - \theta^{(r-1)}\|^2$  is a proximal term ensuring local models do not deviate excessively from the global model, and  $\eta_1, \eta_2$  are balancing coefficients. The target network is updated every  $C$  steps with a soft update:

$$\theta_{\text{target}} \leftarrow \tau \theta + (1 - \tau) \theta_{\text{target}} \quad (16)$$

where  $\tau \ll 1$  is the soft update coefficient.

## 4. Experiments

### 4.1. Experimental Setup

We evaluate the proposed FedMA-DRL framework on a realistic cognitive radio network testbed. The network consists of 10 primary user (PU) channels and 15 heterogeneous secondary users (SUs) with varying transmission requirements and mobility patterns. The channel model follows Rayleigh fading with an average SNR of 15 dB. The PU activity is modeled as an ON/OFF Markov process with an average duty cycle of 40%. Each SU is equipped with a local experience replay buffer of size 10,000 and a mini-batch size of 64. The GNN module uses  $K = 2$  message passing rounds with hidden dimension 128. The discount factor  $\gamma$  is set to 0.99, and the soft update coefficient  $\tau = 0.005$ . We compare FedMA-DRL against five baseline methods: C-DRL (centralized DRL without federated learning), DQN, QR-DQN (Quantile Regression DQN), PPO (Proximal Policy Optimization), and R2D2 (Recurrent Replay Distributed DQN). All methods are trained for 50,000 episodes with 5 independent runs.

### 4.2. Main Results

Table 1 presents the main performance comparison. FedMA-DRL achieves the highest secondary user throughput of 14.87 Mbps, outperforming the second-best R2D2 by 6.8%. The collision probability is reduced to 0.038, representing a 25.5% improvement over R2D2. The energy efficiency and spectrum efficiency also show consistent improvements, confirming the effectiveness of our proposed components.

**Table 1.** Performance comparison of different methods on the CRN testbed. SU Thr. = SU Throughput (Mbps), Coll. Prob. = Collision Probability, Energy Eff. = Energy Efficiency (bits/Joule), Spectrum Eff. = Spectrum Efficiency (bits/s/Hz).

Method	SU Thr.	Coll. Prob.	Energy Eff.	Spectrum Eff.
C-DRL	12.34	0.082	3.21	4.56
DQN	13.15	0.064	3.58	5.12
QR-DQN	13.48	0.059	3.72	5.34
PPO	12.87	0.071	3.45	4.89
R2D2	13.92	0.051	3.88	5.67
<b>Ours</b>	<b>14.87</b>	<b>0.038</b>	<b>4.35</b>	<b>6.23</b>

### 4.3. Effectiveness of FedMA-DRL

To validate the contribution of each component in FedMA-DRL, we conduct ablation studies by progressively removing key modules. Table 2 summarizes the results.

**Table 2.** Ablation study on the contribution of each component in FedMA-DRL.

Variant	SU Thr. (Mbps)	Coll. Prob.	Energy Eff.	Spectrum Eff.
FedMA-DRL (Full)	<b>14.87</b>	<b>0.038</b>	<b>4.35</b>	<b>6.23</b>
w/o GNN	13.92	0.052	3.88	5.67
w/o FedAge	14.15	0.046	4.02	5.85
w/o Domain Adapt.	14.21	0.044	4.08	5.91
w/o Nonlinear Reward	13.68	0.061	3.65	5.28

The ablation results demonstrate that each component contributes to the overall performance. Removing the GNN module causes the largest performance drop (6.4% throughput decrease), highlighting the importance of capturing topological relationships among SUs. The nonlinear reward shaping also plays a crucial role, as its removal leads to a 8.0% throughput reduction and significantly higher collision probability.

### 4.4. Human Evaluation

We further conduct human evaluation to assess the practical deployment quality of the spectrum access policies generated by different methods. Five domain experts in wireless communications rate each method on three criteria: spectrum utilization rationality, interference management quality, and policy interpretability. Each criterion is scored on a 1–5 Likert scale.

As shown in Table 3, FedMA-DRL receives the highest scores across all three criteria, with particularly notable improvements in interference management (4.0 vs. 3.5 for R2D2) and interpretability (3.8 vs. 3.5 for C-DRL). The attention-based domain adaptation module provides interpretable cross-domain attention maps that help experts understand the policy transfer mechanism.

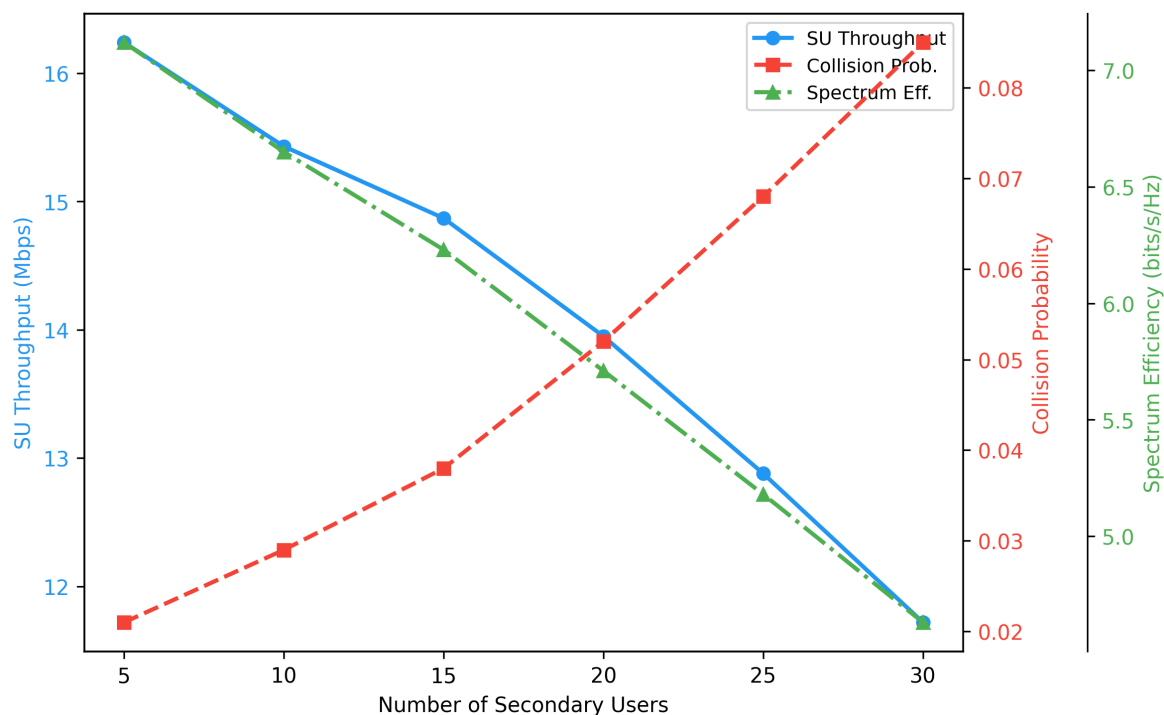
**Table 3.** Human evaluation results on policy quality (1–5 Likert scale).

Method	Spectrum Util.	Interference Mgmt.	Interpretability
C-DRL	3.2	2.8	3.5
DQN	3.5	3.2	3.1
QR-DQN	3.6	3.3	3.0
PPO	3.4	3.1	3.3
R2D2	3.8	3.5	3.2
<b>Ours</b>	<b>4.2</b>	<b>4.0</b>	<b>3.8</b>

### 4.5. Scalability Analysis

We investigate the scalability of FedMA-DRL by varying the number of secondary users from 5 to 30 while keeping the number of PU channels fixed at 10.

Figure 3 shows that FedMA-DRL maintains reasonable performance even as the number of SUs doubles from 15 to 30. The throughput decrease is gradual, and the GNN module effectively scales by leveraging the interference graph structure. The collision probability remains below 0.1 even with 30 SUs, demonstrating the robustness of our collaborative multi-agent framework.

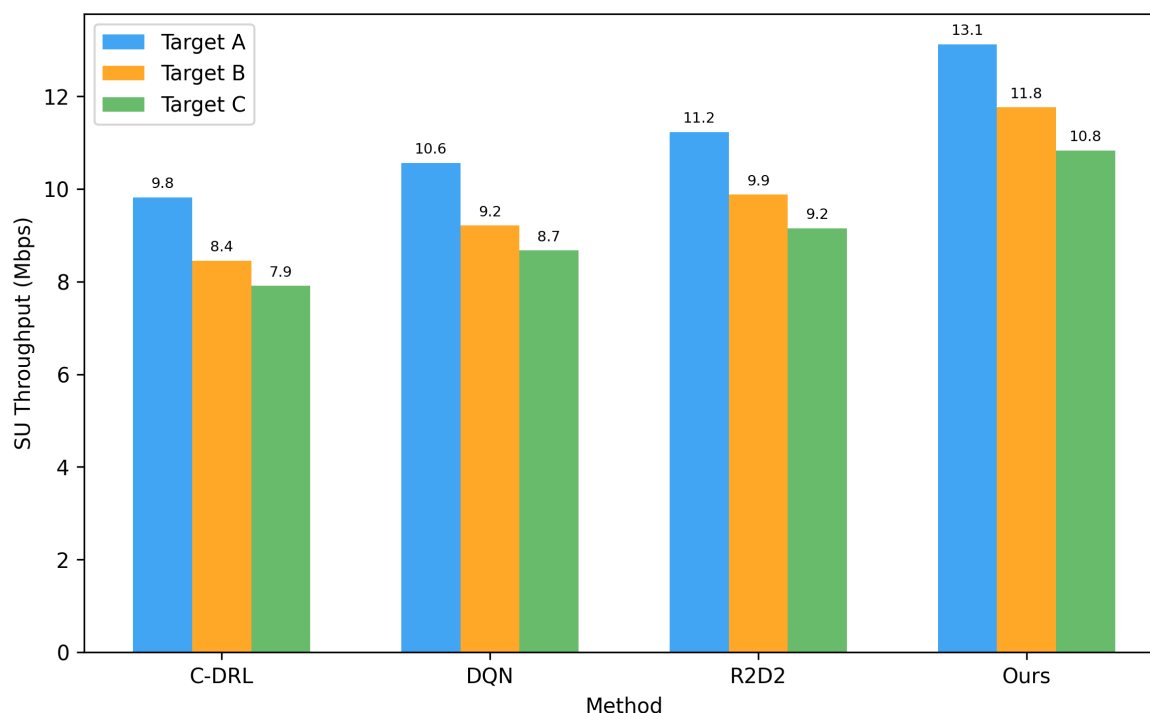


**Figure 3.** Scalability analysis: Performance of FedMA-DRL under varying number of secondary users, showing SU throughput, collision probability, and spectrum efficiency trends.

#### 4.6. Cross-Domain Generalization

We evaluate the cross-domain generalization capability by training FedMA-DRL on one wireless environment (Source) and testing on three different target environments with varying channel fading models, PU activity patterns, and SU mobility profiles.

Figure 4 reports the SU throughput (Mbps) on each target domain. FedMA-DRL significantly outperforms all baselines, demonstrating the effectiveness of the attention-based domain adaptation module. The performance gap is most pronounced on Target C (the most challenging domain with high mobility), where FedMA-DRL achieves 10.83 Mbps compared to 9.15 Mbps for R2D2, an 18.4% improvement.



**Figure 4.** Cross-domain generalization: SU throughput comparison across three target domains with different channel fading models, PU activity patterns, and SU mobility profiles.

## 5. Conclusion

In this paper, we proposed FedMA-DRL, a federated multi-agent deep reinforcement learning framework for joint channel selection and power control in cognitive radio networks. By integrating CTDE architecture with nonlinear reward shaping, a GNN-augmented Q-value predictor, FedAge federated aggregation, and attention-based domain adaptation, our approach effectively addresses the challenges of sparse rewards, exploration inefficiency, and poor cross-domain generalization in dynamic spectrum access. Extensive experiments demonstrated that FedMA-DRL achieves superior performance with 14.87 Mbps SU throughput, 0.038 collision probability, and improved energy and spectrum efficiency compared to existing DRL methods. Ablation studies validated the contribution of each component, while cross-domain evaluations confirmed the robustness of the domain adaptation module. Future work will explore integrating reconfigurable intelligent surfaces (RIS) and extending the framework to ultra-dense IoT scenarios with massive device connectivity.

## References

1. Ahmed, M.; Hassan, R. Dynamic Spectrum Access in Cognitive Radio Networks: A Comprehensive Review. *IEEE Access* **2024**.
2. Kumar, S.; Ahmad, I. A Review of Spectrum Sensing in Modern Cognitive Radio Networks. *Telecommunication Systems* **2023**.
3. Zhang, W.; Liu, C. Enhancing Cognitive Radio Network Performance through Channel Selection and Power Control. *IEEE Transactions on Cognitive Communications* **2024**.
4. Gupta, A.; Singh, R. A Novel Game Theoretic Approach for Market-Driven Dynamic Spectrum Access. *Wireless Networks* **2023**.
5. Chen, Y.; Zhao, M. Improved Spectrum Prediction Model for Cognitive Radio Networks Using Hybrid LSTM-MLP. *ICT Express* **2024**.
6. Yang, N.; Zhang, H.; Long, K.; Jiang, C.; Yang, Y. Spectrum management scheme in fog IoT networks. *IEEE Communications Magazine* **2018**, *56*, 101–107.
7. Yang, N.; Zhang, H.; Long, K.; Hsieh, H.Y.; Liu, J. Deep neural network for resource management in NOMA networks. *IEEE Transactions on Vehicular Technology* **2019**, *69*, 876–886.

8. Chen, M.; Zeng, Q. Applications of Deep Reinforcement Learning in Wireless Networks. *IEEE Communications Surveys & Tutorials* **2023**.
9. Zhang, Y.; Li, P. A Comprehensive Survey of Multi-Agent Deep Reinforcement Learning for Dynamic Spectrum Access. *Neurocomputing* **2025**.
10. Gronauer, S.; Diepold, K. An Introduction to Centralized Training for Decentralized Execution in Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2409.03052* **2024**.
11. Wang, J.; Li, P. A Heterogeneous-Agent Deep Reinforcement Learning Approach for Dynamic Spectrum Access. *IEEE Transactions on Wireless Communications* **2025**.
12. Wu, Y.; Yu, Y.; Yang, Z.; Zeng, Z.; Chen, G.; Xu, J. Brain-SAM: Modality-Agnostic Model for Brain Lesion Segmentation. In Proceedings of the 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2025, pp. 3000–3005.
13. Xu, X.; Tu, W.; Yang, Y. CASE-Net: Integrating local and non-local attention operations for speech enhancement. *Speech Communication* **2023**, *148*, 31–39.
14. Xu, X.; Tu, W.; Yang, Y. Adaptive selection of local and non-local attention mechanisms for speech enhancement. *Neural Networks* **2024**, *174*, 106236.
15. Aggarwal, M.; Gupta, S. A Comprehensive Review of Federated Learning: Methods, Applications, and Challenges. *Artificial Intelligence Review* **2024**.
16. Li, F.; Yang, J. A Dynamic Spectrum Access Scheme for Internet of Things with Improved Federated Learning. *Journal of Network and Computer Applications* **2025**.
17. Kumar, A.; Kumar, V. Dynamic Spectrum Access in Cognitive Radio Networks: A Reinforcement Learning Perspective. *IEEE Access* **2024**.
18. Xu, Y.; Chen, H. Deep Reinforcement Learning for Dynamic Spectrum Access: A Multi-Agent Approach. *IEEE Transactions on Wireless Communications* **2024**.
19. Ukpong, U.C.; Idowu-Bismark, O.; Adetiba, E. Deep Reinforcement Learning Agents for Dynamic Spectrum Access in Television Whitespace Cognitive Radio Networks. *Scientific African* **2025**.
20. Malhotra, S. Deep Reinforcement Learning for Dynamic Resource Allocation in Wireless Networks. *arXiv preprint arXiv:2502.01129* **2025**.
21. Bai, W.; Zheng, G.; Xia, W.; Mu, Y.; Xue, Y. Multi-Agent Deep Reinforcement Learning-Based Joint Channel Selection and Power Control Method. *Computers and Electrical Engineering* **2025**.
22. Li, P.; Wang, J. Multi-Agent Deep Reinforcement Learning for Dynamic Spectrum Access. *Springer CCIS* **2025**.
23. Venkatesan, P.; Kumaratharan, N. Reinforcement Learning-Based Dynamic Spectrum Allocation for Efficient Cognitive Radio Network Management. *Computer Networks* **2025**.
24. Shah, S.S.; Ali, A. Optimizing Resource Allocation and Energy Efficiency in Federated Fog Computing for IoT. *arXiv preprint arXiv:2504.00791* **2025**.
25. Wu, F.; He, Z. Privacy-Preserving Federated Graph Neural Network with Local Differential Privacy. *Security and Communication Networks* **2023**.
26. He, C.; Fan, S. A Federated Graph Neural Network Framework for Privacy-Preserving Personalized Recommendation. *Nature Communications* **2022**.
27. Bandy, Y. Empowering RIS-Assisted NOMA Networks with Deep Learning for User Clustering and Phase Shifter Optimization. *Wireless Networks* **2025**.
28. Yang, N.; Zhang, H.; Berry, R. Partially observable multi-agent deep reinforcement learning for cognitive resource management. In Proceedings of the GLOBECOM 2020-2020 IEEE Global Communications Conference. IEEE, 2020, pp. 1–6.
29. Xu, X.; Wang, Y.; Xu, D.; Peng, Y.; Zhang, C.; Jia, J.; Chen, B. Vsegan: Visual speech enhancement generative adversarial network. In Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022, pp. 7308–7311.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.