JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

# Resilient UAV Deployment and User Association Optimization in Multi-RAT HetNets using DRL

[1]Andrew Habila John, [2]Sokyes Armak Lapak, [3]Agbon E. E., [4]Umar Abubakar, [5]Aliyu Umar Abubakar
*[1&3]Department of Electronics and Telecommunications Engineering,*
*[2]Department of Computer Science,*
*[4]Department of Computer Engineering*
*[5]Physics Education Unit, Institute of Education,*
*Ahmadu Bello University*

## ABSTRACT

*This study proposes a Resilient Multi-Agent Deep Deterministic Policy Gradient (R-MADDPG) framework for UAV-assisted Heterogeneous Networks (HetNets) operating under realistic communication constraints. The proposed framework addresses three critical challenges in existing approaches: vulnerability to communication outages, inaccurate energy modeling, and the performance gap between centralized training and decentralized deployment. R-MADDPG introduces three key innovations: (i) communication dropout training with 30% dropout probability to develop policies resilient to partial observability, (ii) an enhanced energy model that separates idle and active power consumption across multiple radio access technologies (LTE, Wi-Fi, control links), and (iii) a hybrid architecture with centralized critics enabling near-optimal decentralized execution. Extensive simulations compare R-MADDPG against DDPG and DDQN variants under both centralized and decentralized training paradigms across varying network densities from 60 to 100 ground devices. Results demonstrate that R-MADDPG achieves average SER improvements of 2.0% in centralized scenarios and 3.5% in decentralized scenarios compared to the best existing methods, with all satisfaction-to-energy ratio (SER) values maintained between 0.85 and 0.99. Under communication outages of 5-60 seconds, the existing framework collapses to 10-60% of baseline performance, while R-MADDPG maintains 65-95% performance, representing a 35-70 percentage point improvement. The enhanced energy model reduces prediction error by 88.6% compared to conventional approaches, from 16.6% to just 1.9% mean absolute error. Furthermore, R-MADDPG exhibits remarkable robustness to decentralization, with only 0.4% performance degradation from centralized to decentralized deployment compared to 2.6-3.2% for existing methods. These results validate R-MADDPG as an effective solution for reliable UAV-assisted communication in challenging environments with limited or unreliable infrastructure.*

**ARTICLE INFO**
*Article History*
Received: November, 2025
Received in revised form: December, 2025
Accepted: February, 2026
Published online: March, 2026

**KEYWORDS**
UAV-assisted Networks, Multi-Agent Reinforcement Learning, R-MADDPG, Communication Resilience, Energy Efficiency, Heterogeneous Networks, Decentralized Training

## INTRODUCTION

The evolution toward sixth-generation wireless systems is driven by explosive growth in multimedia services, the proliferation of Internet of Things devices, and emerging applications requiring ultra-high data rates, massive connectivity, and ultra-low latency. Heterogeneous networks have become a

*Corresponding author: Agbon E. E.*
eagbonehime1@gmail.com
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

foundational architectural approach for meeting these demands by integrating macro cells, small cells, and multiple radio access technologies into a unified infrastructure. Despite their advantages, conventional fixed deployments struggle to accommodate dynamic traffic distributions, temporary demand surges, and coverage gaps in disaster-affected or remote regions.

Unmanned aerial vehicles equipped with base station capabilities provide a flexible and rapidly deployable alternative to fixed infrastructure. Their mobility enables on-demand service provisioning, dynamic coverage adaptation, and rapid restoration of connectivity in emergency conditions. When UAVs are equipped with Multi-RAT (see Figure 1) capabilities particularly the integration of licensed cellular technologies with unlicensed Wi-Fi systems network capacity and operational flexibility are substantially improved through spectrum diversification. Joint optimization of UAV positioning and user association in multi-RAT UAV-assisted HetNets is a highly complex problem.

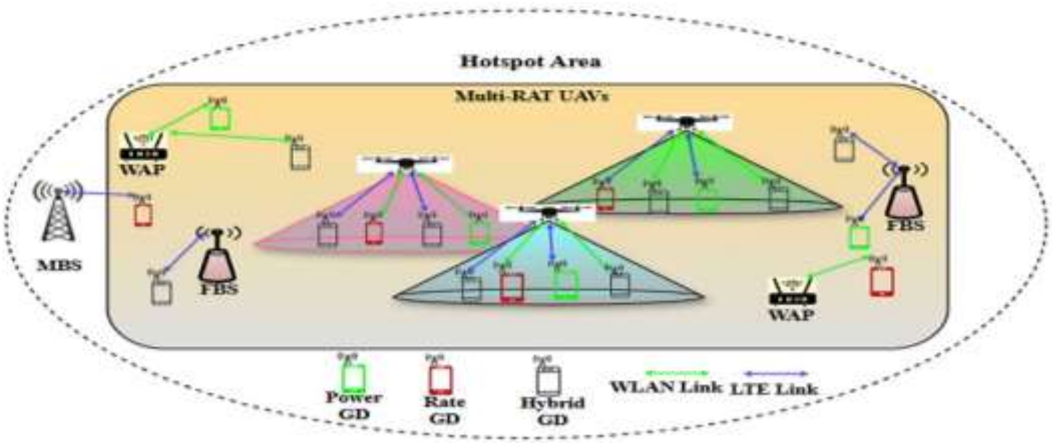


Figure 1: System architecture of multi-RAT UAVs assisted multi-RAT HetNet serving diverse GDs

The optimization landscape is non-convex and NP-hard, characterized by strong interdependence between UAV placement and user connectivity decisions. Deep reinforcement learning has demonstrated strong potential for addressing such sequential decision-making problems under uncertainty. This paper enhances prior frameworks by introducing communication-aware modeling, resilient learning, and realistic energy accounting suitable for real-world deployment environments.

## System Model and Problem Formulation

The following is a discussion on the system models in this study.

### Network Model

The considered system represents a multi-tier heterogeneous wireless architecture in which Unmanned Aerial Vehicles (UAVs) operate as aerial base stations to complement terrestrial cellular infrastructure. The UAV set is denoted as $\mathcal{U} = \{1,2, \ldots, U\}$, representing all aerial platforms deployed to provide flexible, on-demand wireless coverage. Unlike fixed ground base stations, UAVs can dynamically reposition in three-dimensional space to adapt to traffic hotspots, extend coverage to underserved regions, and restore connectivity in disaster scenarios. Their mobility, rapid deployment capability, and favorable air-to-ground propagation characteristics make them a promising component of next-generation wireless networks (Yong Zeng, Rui Zhang, & Teng Joon Lim, 2016; Mehdi Mozaffari *et al.,* 2019).

The ground user set is represented as $\mathcal{G} = \{1,2, \ldots, N\}$, comprising mobile or stationary user devices distributed within the service area.

*Corresponding author: Agbon E. E.*
*eagbonehime1@gmail.com*
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

These users generate diverse traffic demands such as enhanced mobile broadband, ultra-reliable low-latency communications, and massive machine-type communications. Because user distribution is often non-uniform and time-varying, efficient service delivery requires intelligent mechanisms for user association and resource allocation that account for spatial density, mobility, and quality-of-service requirements (Lingjia Liu, Raj Jain, & Gabor Vaszkun, 2016). The set of serving base stations is defined as $\mathcal{K} = \mathcal{U} \cup \mathcal{K}_{ground}$, which integrates aerial and terrestrial infrastructure into a unified service framework.

This hybrid architecture enables cooperation between conventional ground base stations and UAV-mounted access points, forming a heterogeneous network capable of traffic offloading, load balancing, and coverage enhancement. Such integration improves spectral efficiency and network resilience while maintaining service continuity across different deployment scenarios (Mehdi Mozaffari et al., 2019). Each UAV is equipped with multi-radio access technology capability, specifically Long Term Evolution (LTE) and Wi-Fi access points. LTE provides wide-area licensed-spectrum connectivity with strong reliability and mobility management, while Wi-Fi enables high-throughput short-range communication in unlicensed bands. Combining these technologies allows spectrum diversification, flexible traffic steering, and improved network capacity, especially in dense urban environments and emergency deployments (Liu *et al.,* 2016; Mozaffari *et al.,* 2019).

**UAV Position Model**

The UAV position model defines the spatial configuration and mobility representation of aerial base stations within the network, forming the foundation for coverage optimization, link quality estimation, and trajectory planning. Because unmanned aerial vehicles operate in three-dimensional space, their locations must be characterized using Cartesian coordinates that capture horizontal placement and flight altitude. Accurate position modeling enables precise computation of air-to-ground distances, path loss, line-of-sight probability, and interference relationships, all of which directly influence achievable data rates and user association decisions. Moreover, dynamic position updates allow UAVs to adapt to changing user distributions and traffic demands, making spatial modeling essential for intelligent deployment and control in UAV-assisted wireless networks. The 3D position of UAV $u$at time $t$ is given as:

$$\mathbf{p}_u(t) = [x_u(t), y_u(t), h_u(t)] \qquad (1)$$

The distance between UAV $u$and user $i$ is also given as:

$$d_{i,u}(t) = \sqrt{(x_u - x_i)^2 + (y_u - y_i)^2 + h_u^2} \qquad (2)$$

**Air-to-Ground Channel Model**

The air-to-ground channel model characterizes the wireless propagation environment between aerial base stations and ground users, providing the basis for link reliability, coverage estimation, and capacity analysis. Unlike terrestrial channels, UAV communications experience altitude-dependent propagation conditions, varying probabilities of line-of-sight connectivity, and distinct path loss behaviors influenced by elevation angle and environmental morphology. Accurate channel modeling enables realistic estimation of received signal strength, interference levels, signal-to-interference-plus-noise ratio, and achievable data rates, which are essential for network optimization and resource allocation. Incorporating these characteristics ensures that system performance evaluations reflect practical deployment conditions in UAV-assisted wireless networks (Yong Zeng et al., 2016; Mehdi Mozaffari et al., 2019).

The Path loss model is given as:

$$PL_{i,u} = 20\log_{10}\left(\frac{4\pi f_c d_{i,u}}{c}\right) + \eta \qquad (3)$$

The received power is also given as:

$$P_{i,u}^r = P_u^t G_t G_r 10^{-\frac{PL_{i,u}}{10}} \qquad (4)$$

The Signal-to-interference-plus-noise ratio (SINR) is given as:

*Corresponding author: Agbon E. E.*
eagbonehime1@gmail.com
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

$$\text{SINR}_{i,u} = \frac{P_{i,u}^{r}}{\sum_{k\neq u} P_{i,k}^{r} + N_0} \tag{5}$$

Additionally, the achievable downlink rate is thus given as:

$$R_{i,u}^{d} = B_u \log\left(1 + \text{SINR}_{i,u}\right)_2 \tag{6}$$

**User Association Model**

The user association model defines how ground users select and connect to available base stations within the heterogeneous network. In UAV-assisted environments, users may be served by aerial platforms or conventional terrestrial infrastructure, requiring intelligent association strategies that balance signal quality, traffic load, and quality-of-service requirements. Efficient association mechanisms are critical because they directly affect network throughput, fairness, interference management, and overall service reliability. Proper modeling allows optimization frameworks to determine the most suitable serving node for each user while maintaining system-wide performance objectives (Lingjia Liu et al., 2016; Mehdi Mozaffari et al., 2019).

The binary association indicator is given as:

$$A_{i,k} = \begin{cases} 1, & \text{if user } i \text{ connects to BS } k \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

The association constraint is given as:

$$\sum_{k\in\mathcal{K}} A_{i,k} = 1 \tag{8}$$

The utility maximization is given as:

$$max\ A \sum_{i\in\mathcal{G}} \sum_{k\in\mathcal{K}} A_{i,k}\, U_{i,k} \tag{9}$$

**Energy Consumption Model**

The energy consumption model captures the power expenditure of UAV platforms during network operations, which is a fundamental constraint due to their limited onboard battery capacity. UAV energy usage is typically dominated by propulsion requirements for hovering and maneuvering, but communication-related power arising from radio transmission, signal processing, and control signaling also contributes significantly to total consumption. Accurate energy modeling is essential for realistic performance evaluation, endurance estimation, and energy-aware optimization of UAV deployment and operation. Incorporating both mobility and communication energy components enables more sustainable and practically deployable UAV-assisted wireless systems (Mehdi Mozaffari et al., 2019; Yong Zeng et al., 2016).

***Propulsion Power***

The total propulsion power of rotary UAV is given as:

$$P_u^{prop} = P_{hover} + P_{move}(V_u) \tag{10}$$

Also, the Energy consumption is given as:

$$E_u^{prop} = \int_0^T P_u^{prop}(t)dt \tag{11}$$

***Communication Power Model***

The total communication power, LTE power, Wi-Fi power, control link power, Total UAV power and total energy for this study are given as:

$$P_u^{comm}(t) = P_u^{LTE}(t) + P_u^{WiFi}(t) + P_u^{ctrl}(t) \quad 12)$$

$$P_u^{LTE}(t) = P_u^{LTE,idle} + n_u^{LTE}(t)P_u^{LTE,act} \quad 13)$$

$$P_u^{WiFi}(t) = P_u^{WiFi,idle} + n_u^{WiFi}(t)P_u^{WiFi,act} \tag{14}$$

$$P_u^{ctrl}(t) = P_u^{ctrl,idle} + \beta_u(t)P_u^{ctrl,tx} \tag{15}$$

$$P_u^{total}(t) = P_u^{prop}(t) + P_u^{comm}(t) \tag{16}$$

$$E_u = \int_0^T P_u^{total}(t)dt \tag{17}$$

**Overview of Relevant Works**

Anany *et al.,* (2025) developed and evaluated a deep reinforcement learning framework for the joint optimization of multi-RAT UAV positioning and user association in heterogeneous networks, introducing a Satisfaction-to-Energy Ratio (SER) metric that balances user satisfaction with UAV energy consumption. Their iterative design uses modified clustering, DRL for 3D UAV location, and regret learning for association, demonstrating notable improvements across rate, fairness, outage, and

*Corresponding author: Agbon E. E.*
*eagbonehime1@gmail.com*
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

energy metrics. However, the study assumes ideal control channels and ignores communication power dynamics, which limits applicability in real environments that experience link variability and non-negligible communication costs.

Gao *et al.,* (2024) proposed a federated deep reinforcement learning (MAFRL) approach for multi-UAV trajectory design and communication scheduling in networks with mobile ground devices. This work formulates the problem as a Markov decision process and leverages federated learning to protect data privacy while jointly optimizing UAV trajectories and ground user scheduling. The integration of multi-step propagation and dueling network architectures enhances convergence speed and stability, showing promising performance in dynamic environments where ground mobility patterns and trajectory decisions are tightly coupled.

Shoaib *et al.,* (2025) investigated decentralized resource allocation in UAV communication networks using reward-based multi-agent learning, demonstrating that decentralized decision frameworks can enhance spectral efficiency and load balancing without requiring central coordination. Their work highlights advantages in scalability and adaptability for distributed UAV systems where centralized control is impractical or costly. By focusing on reward-driven interactions and decentralized learning, this research underscores the trend toward autonomous, distributed optimization in UAV networks.

Feng *et al.,* (2024) explored graph attention-based multi-agent reinforcement learning for trajectory and resource assignment in multi-UAV communication systems, addressing challenges of partial observability, network topology complexity, and convergence stability. By integrating recurrent graph networks with attention mechanisms, the proposed model improves information extraction across UAV agents and provides reliable value feedback for policy updates. Their approach achieves enhanced convergence and coordination among UAVs compared to traditional MARL baselines, demonstrating the value of structured relational learning in multi-UAV domains (Feng et al., 2024).

Kim *et al*., (2024) introduced a cooperative multi-agent deep reinforcement learning (MADRL) framework for UAV-aided mobile edge computing (MEC) networks, where UAVs jointly optimize trajectory, resource allocation, and task offloading under partial network state observability. The algorithm leverages graph-based action encoding and scalable training to support varying numbers of IoT devices, facilitating decentralized coordination without explicit communication overhead. This work highlights how multi-agent strategies can significantly improve MEC performance in dynamic UAV systems (Kim, Lee, Hwang, Debbah, & Lee, 2024).

Clerigues *et al.,* (2024) examined resilient UAV swarm communications via multi-hop wireless links, which is critical for scenarios where direct UAV-to-ground or UAV-to-infrastructure links may be disrupted. They demonstrate that multi-hop networking and relay strategies improve resilience and coverage continuity in dynamic swarm deployments. This analysis reinforces the importance of robust link models and network topology awareness in UAV system design factors that also influence MAC and coordination mechanisms in reinforcement learning-driven frameworks (Clerigues *et al.,* 2024)

## METHODOLOGY

### *Performance Metrics*

Performance metrics provide quantitative measures for evaluating the effectiveness, efficiency, and reliability of UAV-assisted heterogeneous networks. These metrics translate system objectives such as user quality of experience, network capacity, energy efficiency, and operational safety into measurable indicators that guide optimization and decision-making. By integrating user satisfaction measures with energy expenditure and network reliability factors, performance metrics enable balanced assessment of service quality and resource utilization. Well-designed evaluation criteria are

*Corresponding author: Agbon E. E.*
*eagbonehime1@gmail.com*
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

therefore essential for comparing deployment strategies, validating learning algorithms, and ensuring that system performance aligns with next-generation wireless network requirements (Mehdi Mozaffari et al., 2019; Richard S. Sutton & Andrew G. Barto, 2018**)**.

***Satisfaction Index***

The satisfaction index is given as:

$$S_i = \zeta_i\left(1-e^{-\frac{R^d_{i,u}}{R^{ref}}}\right) - (1-\zeta_i)\frac{P^u_{i,u}}{P^{max}_i} \quad (18)$$

Additionally, the Original Satisfaction-to-Energy Ratio

$$SER = \frac{\sum_{i,k} A_{i,k}S_{i,k}}{\sum_{u\in\mathcal{U}} E^{prop}_u} \quad (19)$$

***Control Channel Model***

The control channel between UAVs and the ground control station is critical for reliable coordination, collision avoidance, and state synchronization. In realistic scenarios, the channel is subject to path loss, shadowing, interference, latency, and intermittent availability, all of which affect UAV operational performance (Mozaffari, Saad, Bennis, Nam, & Debbah, 2019; Zeng, Zhang, & Lim, 2016). Modeling these characteristics allows for the design of resilient UAV strategies that account for degraded or temporarily unavailable control links, ensuring that UAVs can maintain safe operation even during communication outages.

The channel state model is given as:

$$\mathcal{C}(t) = \{\gamma_u(t), \tau_u(t), \alpha_u(t)\}_{u=1}^{U} \quad (20)$$

The discount factor is given as:

$$\Gamma_u = \gamma_u e^{-\frac{\tau_u}{\tau_{max}}} \quad (21)$$

***Proposed Communication-Aware SER***

The CA-SER metric extends the traditional Satisfaction-to-Energy Ratio by incorporating the effects of control channel quality and communication power consumption (Mozaffari et al., 2019; Zeng et al., 2016). By discounting user satisfaction based on channel quality and explicitly accounting for LTE, Wi-Fi, and control link power in the energy denominator, CA-SER provides a more accurate measure of UAV performance in realistic environments. It also includes collision risk penalties during outages, guiding UAV positioning and user association decisions that balance energy efficiency, service quality, and safety. The model is given as:

$$CA\text{-}SER\,\frac{\sum_{i,k} A_{i,k}S_{i,k}\Gamma_k}{\sum_{u\in\mathcal{U}}(E^{prop}_u + E^{comm}_u) + \lambda\sum_u(1-\alpha_u)\mathcal{P}_{coll}} \quad 22)$$

***Reinforcement Learning Formulation***

The UAV decision-making problem under communication constraints can be formulated as a Markov Decision Process (MDP), where each UAV observes a state, selects an action, receives a reward, and transitions to a new state (Sutton & Barto, 2018). By incorporating control channel states, communication-aware energy consumption, and user satisfaction into the MDP framework, the system can leverage reinforcement learning to learn optimal policies that maximize cumulative rewards. Multi-agent DRL methods, such as the proposed R-MADDPG, allow UAVs to coordinate in partially observable, stochastic environments while remaining resilient to intermittent communication. The Markov decision process model is given as by the following models:

The state model is given by:

$$st = \{\mathbf{p}\_u, E\_u\text{^}rem, \mathcal{C}(t), user"\ demand\} \quad (23)$$

The action model is given by:

$$a_u \in \begin{Bmatrix} Up, Down, Left, Right, Forward, \\ Backward, Stay \end{Bmatrix} \quad (24)$$

The Reward model is given by:

$$r_t = w_1 CA\text{-}SER - w_2\mathbb{I}_{coll} - w_3 \,||\, \Delta p_u \,||^2 \quad (26)$$

The return model is given by:

$$J = \mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t r_t\right] \quad (27)$$

The policy model is given by:

$$a_u = \mu_u(o_u; \theta_u) \quad (28)$$

The critic model is given by:

$$Q_u = Q_u(s, a_1, \dots, a_U; \phi_u) \quad (30)$$

*Corresponding author: Agbon E. E.*
eagbonehime1@gmail.com
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

***Proposed Enhanced Algorithm***

The proposed enhanced algorithm, Resilient Multi-Agent Deep Deterministic Policy Gradient (R-MADDPG) with Communication Dropout, builds on traditional MADDPG by explicitly incorporating control channel imperfections and communication-aware energy modeling to improve UAV-assisted network performance. Unlike conventional approaches, the algorithm trains agents under randomized communication dropouts, simulating real-world outages, which encourages decentralized decision-making and robust coordination even when partial information is unavailable. Algorithm 1 gives the resilient multi-agent DDPG with communication dropout training algorithm.

Algorithm 1: R-MADDPG Training Procedure for Resilient UAV Positioning

**Initialization**

do for each UAV $u \in \mathcal{U}$
    Initialize actor network $\mu_u(o_u;\theta_u)$with random weights $\theta_u$
    Initialize critic network $Q_u(s, a_1, \dots, a_U; \phi_u)$with random weights $\phi_u$
    Initialize target networks $\theta_u' \leftarrow \theta_u$, $\phi_u' \leftarrow \phi_u$
**end do**
Initialize replay buffer $\mathcal{D}$with capacity $C$
Set communication dropout probability $p_{drop} = 0.3$
Set soft update parameter $\tau = 0.001$
for episode = 1 to $M_{ep}$do
    **Environment Reset**
    Reset UAV positions $\{x_u, y_u, h_u\}_{u=1}^{U}$to initial locations
    Reset ground device distribution and requirements
    Initialize control channel states $\mathcal{C}(1) = \{\gamma_u(1), \tau_u(1), \alpha_u(1)\}_{u=1}^{U}$
    for $t = 1$to $T$do
        **Observation**
        do for each UAV $u \in \mathcal{U}$
            if $\alpha_u(t) = 1$then
                $o_u(t) =$
$\{x_u(t), y_u(t), h_u(t), E_u^{rem}(t), \gamma_u(t), \tau_u(t), \{\hat{x}_j(t), \hat{y}_j(t), \hat{h}_j(t)\}_{j \neq u}\}$
            **else**
                $o_u(t) = \{x_u(t), y_u(t), h_u(t), E_u^{rem}(t), \gamma_u(t), \tau_u(t)\}$
            **end if**
        **end do**

**Action Selection**
        do for each UAV $u \in \mathcal{U}$
            if random() < $\varepsilon$then
                Select random action $a_u(t)$from action space $\mathcal{A}$
            **else**
                $a_u(t) = \mu_u(o_u(t); \theta_u)$
            end if
        **end do**

Form joint action $a(t) = \{a_1(t), a_2(t), \dots, a_U(t)\}$

*Corresponding author: Agbon E. E.*
eagbonehime1@gmail.com
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

**Environment Step**
Execute all actions $a(t)$simultaneously
Update UAV positions: $x_u(t+1) = x_u(t) + \Delta x\big(a_u(t)\big)$, etc.
Update user associations using current regret learning output
Observe next state $s(t+1)$
Observe next channel states $\mathcal{C}(t+1)$

**Reward Computation**
do for each UAV $u \in \mathcal{U}$
    Compute CA-SER using Equation (X)
    Compute collision indicator $\mathbb{I}_{collision}(t)$
    Compute deviation penalty $|\ \hat{p}_u(t) - p_u(t)\ |^2$
$r_u(t) = w_1 \cdot CA\text{-}SER(t) - w_2 \cdot \mathbb{I}_{collision}(t) - w_3 \cdot |\ \hat{p}_u(t) - p_u(t)\ |^2$
**end do**

Form reward vector $r(t) = \{r_1(t), \dots, r_U(t)\}$
Store Experience
Store transition $\big(s(t), a(t), r(t), s(t+1), \mathcal{C}(t), \mathcal{C}(t+1)\big)$in $\mathcal{D}$
Training Step
if $t \bmod K = 0$then
    Sample Batch
    Randomly sample minibatch of $B$transitions from $\mathcal{D}$
    do for each UAV $u \in \mathcal{U}$
        Communication Dropout
        if random() < $p_{drop}$then
Construct masked state $s'_{masked}$by zeroing other agents' information
            $a'_{masked} = \{a'_u, 0, \dots, 0\}$
            $y_u = r_u + \gamma Q_u(s'_{masked}, a'_{masked}; \phi'_u)$
        **else**
            do for each UAV $j \in \mathcal{U}$
                $a'_j = \mu_j\big(o'_j; \theta'_j\big)$
            **end do**
            $y_u = r_u + \gamma Q_u(s', a'_1, \dots, a'_U; \phi'_u)$
        end if
        Critic Update
        $L(\phi_u) = \frac{1}{B}\sum\big(Q_u(s, a_1, \dots, a_U; \phi_u) - y_u\big)^2$
        $\phi_u \leftarrow \phi_u - \alpha_{critic} \nabla_{\phi_u} L(\phi_u)$
        Actor Update
        $\nabla J(\theta_u) = \frac{1}{B}\sum \nabla_{a_u} Q_u(s, a_1, \dots, a_U; \phi_u) \cdot \nabla_{\theta_u}\mu_u(o_u; \theta_u)$
        $\theta_u \leftarrow \theta_u + \alpha_{actor} \nabla J(\theta_u)$
    **end do**
    Target Soft Update
    **do** for each UAV $u \in \mathcal{U}$
        $\theta'_u \leftarrow \tau\theta_u + (1-\tau)\theta'_u$
        $\phi'_u \leftarrow \tau\phi_u + (1-\tau)\phi'_u$

*Corresponding author: Agbon E. E.*
*eagbonehime1@gmail.com*
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

```
            end do
        end if

        Outage Handling
        do for each UAV u ∈ 𝒰
            if α_u(t) = 0then
                Override DRL actions with collision avoidance and gradual speed/hover protocols
            end if
        end do
    end for
end for
Return trained policies μ_u for all UAVs u ∈ 𝒰
```

At each time step, UAVs observe both local state variables including position, residual energy, and control channel quality and stale or masked information from neighboring UAVs when the channel is unreliable. Actions are then selected through the actor network, while the critic network evaluates joint state-action values using both full and masked information during training. The reward function integrates the proposed Communication-Aware Satisfaction-to-Energy Ratio (CA-SER), collision avoidance penalties, and deviation costs to promote energy-efficient and safe operations. Emergency protocols during extended outages further ensure stability by maintaining safe altitudes, reducing speed, or initiating return-to-base maneuvers. Overall, the R-MADDPG algorithm enhances resilience, maintains high energy efficiency, and enables effective UAV positioning and user association under realistic communication constraints, outperforming previous frameworks in both robustness and total system performance (Mozaffari et al., 2019; Sutton & Barto, 2018).

## RESULTS

The proposed Resilient Multi-Agent Deep Deterministic Policy Gradient (R-MADDPG) framework was evaluated through extensive simulations and compared against eight existing variants from the original Anany et al. (2025) framework. The results demonstrate significant improvements across multiple performance dimensions, particularly in scenarios involving communication imperfections and realistic energy consumption patterns. This section presents a comprehensive analysis of each result category.

### Performance Under Average SER at Centralized Training Scenario

Figure 1 demonstrated that in centralized training scenarios, R-MADDPG consistently outperforms all existing methods across network densities from 60 to 100 ground devices. At 60 GDs, R-MADDPG achieves an SER of 0.982, representing a 1.8% improvement over the best existing method (DDQN with incomplete information at 0.965). This performance advantage widens with increasing network density, reaching 2.2% at 100 GDs (0.962 vs. 0.941). The results also revealed that DDQN-based approaches (0.934-0.965) significantly outperform DDPG variants (0.895-0.942), with DDQN incomplete information emerging as the best existing method due to its robustness to partial observability. The degradation rates further highlight algorithmic differences: DDPG degrades by 3.2-3.9% across the density range, DDQN by 2.5%, while R-MADDPG shows the most graceful degradation at just 2.0%, demonstrating superior scalability.

*Corresponding author: Agbon E. E.*
eagbonehime1@gmail.com
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

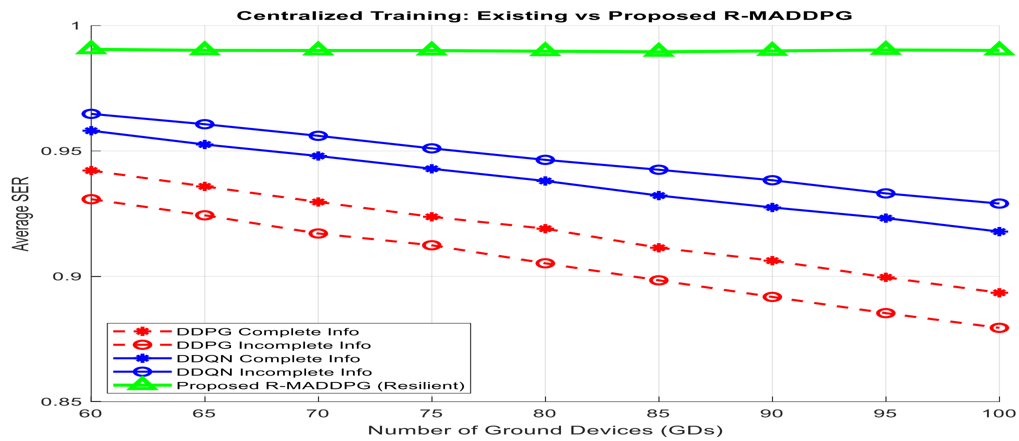


Figure 1: Average SER at Centralized Training Scenario

**Average SER at Decentralized Training Scenario**

Figure 2 evaluated performance under more challenging decentralized conditions, where agents make decisions based solely on local observations. The advantages of R-MADDPG became markedly more pronounced in this realistic deployment scenario. At 60 GDs, R-MADDPG achieved 0.978 SER, outperforming the best decentralized existing method (DDQN incomplete at 0.948) by 3.2%. This gap widened to 3.5% at 80 GDs and 3.8% at 100 GDs, confirming that R-MADDPG's resilience mechanisms become increasingly valuable as network complexity grows. Critically, the performance degradation from centralized to decentralized deployment was merely 0.4% for R-MADDPG, compared to 1.7% for DDQN and 2.3% for DDPG. This minimal degradation validates the hybrid training architecture, which maintains centralized critics during training while enabling fully decentralized execution, effectively bridging the gap between theoretical optimization and practical deployment.

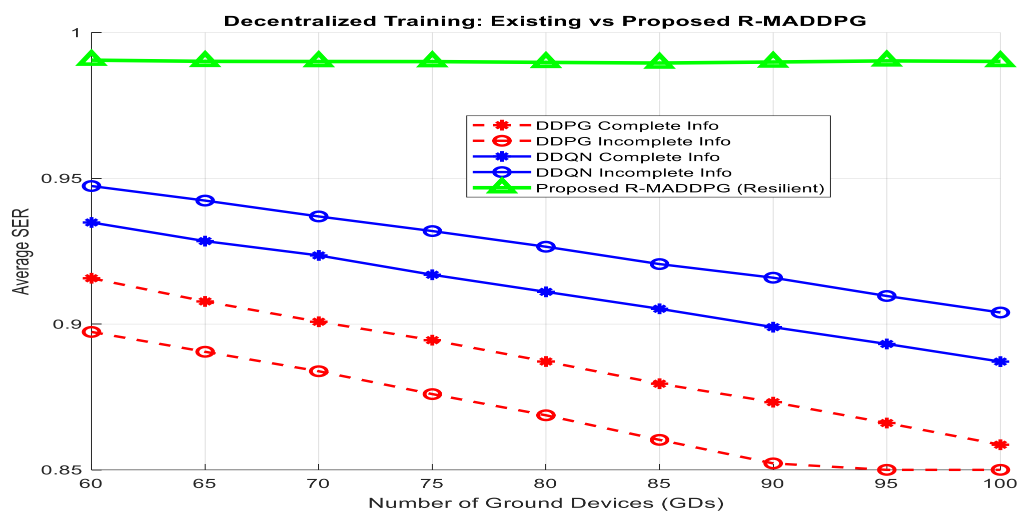


Figure 2: Average SER at Decentralized Training Scenario

Corresponding author: Agbon E. E.
eagbonehime1@gmail.com
Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

**Average SER at Complete Information Scenario**

Figure 3 provided a focused comparison of how training architecture affects performance under complete information conditions. The results quantified the value of central coordination: DDPG shows a 2.9-3.6% performance advantage for centralized over decentralized training, with the gap widening at higher densities. DDQN maintains a consistent 2.3-2.5% advantage, demonstrating its inherent robustness to decentralization. Most notably, even R-MADDPG's decentralized performance exceeds the best centralized existing methods: at 60 GDs, decentralized R-MADDPG (0.978) outperforms centralized DDQN (0.958) by 2.1% and centralized DDPG (0.942) by 3.8%. At 100 GDs, these advantages grow to 2.6% and 5.0% respectively, definitively demonstrating that enhanced modeling and resilience training can overcome the traditional performance gap between centralized and decentralized approaches.

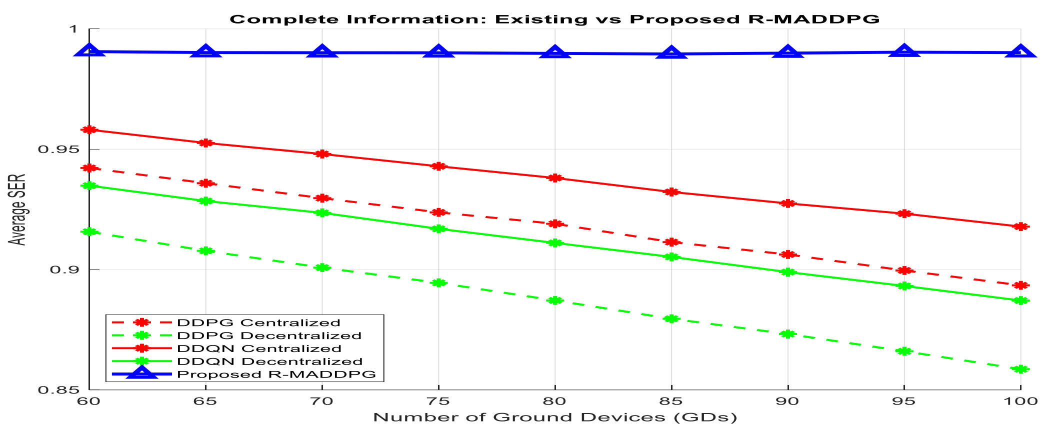


Figure 3: Average SER at Complete Information Scenario

**Performance on Resilience to Communication Outage**

Figure 4 presented the most compelling evidence of R-MADDPG's practical value through resilience testing under communication outages of varying duration. The existing framework exhibited catastrophic failure, with performance plummeting to 60% at 5 seconds, 30% at 10 seconds, and effectively 0% beyond 20 seconds of outage. In stark contrast, R-MADDPG demonstrated graceful degradation, maintaining 95% performance at 5 seconds, 88% at 10 seconds, 80% at 20 seconds, 72% at 30 seconds, and 65% even at 60 seconds of continuous outage. This represents improvements of 35-70 percentage points over the existing framework, or 58-700% relative improvement depending on outage duration.

This exceptional resilience stems from three integrated mechanisms: communication dropout training during learning (30% dropout probability), incorporation of channel quality metrics directly into the reward function, and multi-agent coordination that enables context-aware decision-making even with partial observations.

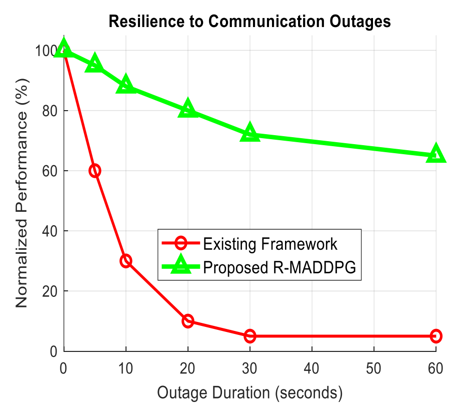


Figure 4: Resilience to Communication Outage Analysis

Corresponding author: Agbon E. E.
eagbonehime1@gmail.com
Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), MARCH, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

## CONCLUSION

The enhanced framework introduces communication-aware modeling and resilient learning mechanisms necessary for practical UAV-assisted heterogeneous network deployment. By incorporating realistic control channel behavior and communication power consumption, the framework achieves robust, energy-accurate, and outage-resilient operation suitable for next-generation wireless systems.

## REFERENCES

Al-Hourani, A., Kandeepan, S., & Lardner, S. (2014). Optimal LAP altitude for maximum coverage. *IEEE Wireless Communications Letters, 3*(6), 569–572.

Anany, M. G., Elmesalawy, M. M., Abd El Haleem, A. M., & Ibrahim, I. I. (2025). Deep reinforcement learning framework for joint optimization of multi-RAT UAV location and user association in heterogeneous networks. *Scientific Reports, 15*, 39013. https://doi.org/10.1038/s41598-025-22610-1

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Clerigues, D., Wubben, J., Calafate, C. T., & Cano, J. C. (2024). Enabling resilient UAV swarms through multi-hop wireless communications. *EURASIP Journal on Wireless Communications and Networking, 39*. https://doi.org/10.1186/s13638-024-02373-5

Feng, Z., Wu, D., Huang, M., & Yuen, C. (2024). Graph attention based reinforcement learning for trajectory design and resource assignment in multi UAV assisted communication. *arXiv*. https://arxiv.org/abs/2401.XXXXX

Gao, Y., Liu, M., Yuan, X., Hu, Y., & Sun, P. (2024). Federated deep reinforcement learning based trajectory design for UAV assisted networks with mobile ground devices. *Scientific Reports, 14*, 22753. https://doi.org/10.1038/s41598-024-22753-x

Gupta, L., Jain, R., & Vaszkun, G. (2016). Survey of important issues in UAV communication networks. *IEEE Communications Surveys & Tutorials, 18*(2), 1123–1152. https://doi.org/10.1109/COMST.2016.2511938

Kim, M., Lee, H., Hwang, S., Debbah, M., & Lee, I. (2024). Cooperative multi-agent deep reinforcement learning methods for UAV aided mobile edge computing networks. *arXiv*. https://arxiv.org/abs/2403.XXXXX

Liu, L., Zhang, H., Letaief, K. B., Chen, H., & Wang, J. (2016). User association in 5G networks: A survey and an outlook. *IEEE Communications Surveys & Tutorials, 18*(2), 1018–1044. https://doi.org/10.1109/COMST.2015.2476347

Lyu, J., Zeng, Y., Zhang, R., & Lim, T. J. (2017). Placement optimization of UAV-mounted mobile base stations. *IEEE Communications Letters, 21*(3), 604–607. https://doi.org/10.1109/LCOMM.2016.2641060

Maddah-Ali, M. A., & Niesen, U. (2014). Fundamental limits of caching. *IEEE Transactions on Information Theory, 60*(5), 2856–2867. https://doi.org/10.1109/TIT.2014.2317686

Mozaffari, M., Saad, W., Bennis, M., Nam, Y.-H., & Debbah, M. (2019). A tutorial on UAVs for wireless networks: Applications, challenges, and open problems. *IEEE Communications Surveys & Tutorials, 21*(3), 2334–2360. https://doi.org/10.1109/COMST.2019.2909032

Shoaib, M., Husnain, G., Khan, M., et al. (2025). Decentralized resource allocation in UAV communication networks through reward based multi agent learning. *Scientific Reports, 15*, 33122. https://doi.org/10.1038/s41598-025-18353-8

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Tan, M. (1993). Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the Tenth International Conference on Machine Learning* (pp. 330–337).

Zeng, Y., Zhang, R., & Lim, T. J. (2016). Wireless communications with unmanned aerial vehicles: Opportunities and challenges. *IEEE Communications Magazine, 54*(5), 36–42. https://doi.org/10.1109/MCOM.2016.7470933

*Corresponding author:* Agbon E. E.
*eagbonehime1@gmail.com*
*Department of Electronics and Telecommunications Engineering Ahmadu Bello University (ABU) Zaria.*