Scientific Journal of Computer Science

futuristech PT. Teknologi Futuristik Indonesia

Date of publication November 2, 2025, date of current version November 2, 2025.
e-ISSN: 3110-3170

Digital Object Identifier 10.64539/sjcs.v1i2.2025.316

Article

A GAT-Assisted Hybrid Reinforcement Learning and Swarm Intelligence Framework for Autonomous UAV Coordination

Mst Jannatul Kobra[1,*], Md Owahedur Rahman[1], Mizanur Rashid[1]

[1] Nanjing University of Information Science & Technology, Nanjing, Jiangsu, China; jannat@nuist.edu.cn; owahedur@nuist.edu.cn; mizanurrashid218@gmail.com
[*] Correspondence

Abstract: The autonomous UAV swarms have fundamental issues with strong coordination that arise under delays in communication, dynamic obstacles and noisy sensing environments, and the existing centralized or heuristic-based solutions are insufficient in addressing such issues. To cover this gap, this paper proposes a Graph Attention Network (GAT)-based Hybrid Reinforcement Learning and Swarm Intelligence Framework that can enable the communication-aware decentralized cooperation of UAVs. It is a multi-agent reinforcement learning and PSO, ACO, Differential Evolution, flocking behavior and Control Barrier Function-based safety correction, and GAT-inspired adaptive graph communication encoding. The results of the simulation of 18 episodes with 24 UAVs demonstrate that the reward, coverage, and collision were demonstrated to be improved by 32%, 27%, and 40% respectively as compared to a classical greedy baseline. The findings confirm the fact that the proposed hybrid GAT-RL architecture enables to promote significantly more scalability, safety, and real-time responsiveness of UAV swarms, which is a possibility on the path to large-scale autonomous aerial coordination.

Keywords: UAV Swarm Coordination; Graph Attention Networks; Hybrid Reinforcement Learning; Swarm Intelligence Optimization; Control Barrier Functions

1. Introduction

Unmanned Aerial Vehicle (UAV) swarms are emerging powerful products with a capacity of sustained sense, high-speed surveillance, and control and efficient relay of communication. However, as with implementation of UAV swarms, there exist several challenges that are involved in the real-world application. These are dynamic barriers, adhoc networking, and partial visibility all of which compromise the quality of coordination in multi-agent environments [1-3]. To perform large-scale path planning and task allocations, especially in contested, windy or cluttered environments, the information-intensive algorithms that can be resistant to uncertainty, remain non-sensitive to communication and environmental modifications are required [4, 5].

The current solutions in the industry are quite weak. Pure reinforcement learning (RL), as an example, overfits to simulation-based artifacts, and is unable to solve sparse or noisy rewards. In addition, RL is not suitable to fluctuation in the topology environment or delay in communication. On the other hand, evolutionary and rule-based systems such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Differential Evolution (DE) and flocking algorithms are less efficient in the use of the sample, and cannot easily handle stochastic systems [6-8]. As has recently been shown with the help of Graph Neural Networks (GNNs) or more precisely Graph Attention Networks (GATs), these tools can be decentralized, form control and trained via multi-agent reinforcement learning (MARL) with the weight of neighbors being adaptively weighted. Having such developments, most of the problems such as facilitating communication when carrying out maneuvers within a close range and assurances of safety have not yet been resolved [9, 10]. Furthermore, Control Barrier Functions (CBFs) have also been suggested to ensure forward invariance in dynamic environments, but have not been studied in swarm systems, which experience packets loss and moving obstacles, which is a research question [11, 12].

To address these problems, we come up with Hybrid RL-Swarm Architecture that integrates communication-adaptive reinforcement learning with safetyconstrained swarm intelligence. It is a model that has an actor-critic policy; a swarm-based supervision system (PSO, ACO, DE, flocking) and a CBF safety shield; a GAT en-coder links the communication graphs dynamically to consider the influence of surrounding UAVs. We possess frame-work that includes multi-objective reward shaping that optimizes coverage, reachability of targets, energy efficiency and stability in communication considering realistic distractions like the failure of communications and environmental noise [13, 14]. This decentralized framework is informed by the concepts of GNN+MARL [15, 16] and it applies attention mechanisms to alter the weights of the edges to make it more resilient to the loss of packets and topology drift. The experiment involving the simulation of 24 UAVs with dynamic wind and communication failure conditions demonstrates that the model can be easily scaled and ensures the safety of the coordinated and high-performing operation of UAV swarms.

And finally, we put our work on the forefront of GNN-enhanced MARL and the Hybrid Swarm Intelligence, addressing three primary areas, namely: (i) adaptive communication with dynamic proximity graphs, (ii) principled safety with executable CBFs and (iii) multi-objective optimization with reward functions (e.g., coverage, collisions, energy, connectivity, latency). These contributions are a significant enhancement of existing techniques of path-planning of UAVs and multi-robot coordination strategies [4, 5].

2. Related Works

2.1. Learning (reinforcement learning) UAV swarm control

Premature RL experiments of UAV swarms prove that decentralized learning is partially observable and has heterogeneous mission objectives, however convergence and stability is sensitive to reward shaping, credit allocation and communication losses. These limitations and trends are typified by the recent surveys with multi-UAV control, coordination and navigation, requiring the structure-conscious policies and real-life evaluation environment [17-22]. Task papers Task papers Task assignment and path planning Tasks plans are a RL-based representation of a POMDP, counterfactual credit assignment of adaptive informative path planning and multi-agent SAC (applied in multi-UAV path planning/following under sensing constraints) [18-20]. Smaller reviews on UAV+RL (systematic to DRL-centric) also indicate a gap in the scalability, safety and sim-to-real transfer in the 2021-2025 window [22].

2.2. Optimization of swarm intelligence UAVs

Swarm intelligence (SI) algorithms (e.g. PSO, ACO, DE, APF hybrids) are also competitive with respect to exploration on a global scale, and also multi-objective path planning. Extensive surveys refer to the dynamic developments of model taxonomies, hybridization, and adaptations of UAV, it indicates the gaps in the safety constraints and model communications [23, 24]. The proposed paper is a detailed AI-based energy distribution optimization framework of the smart grids. The authors concentrate on the methods of simulation and optimization of the smart grids with AI to increase energy efficiency and real-time control. The article makes comparisons with the energy efficiency problems of UAV swarms and provides insights on distributed optimization and intelligent decision-making with constraints, which can be applied to the multi-agent UAV systems in energy-constrained settings [25], rotational-force APF extensions to cooperative planning [26], and Q-learning-directed MOPSO to the 3-D UAV paths [27]. In addition to UAVs, there is also systematic performance improvement of hybrid PSO-RL and other bio-inspired Bio-hybrid PSO-RL approach to mobile robotics (AGV/AUV), which also indicates the advantage (and portability) of synergy of metaheuristic-learning [28, 29].

2.3. GNN/GAT coordination of multi-agent system (MAS)

The fast decentralized control, learning to encode local interactions, in the form of message-passing has been made possible by the use of graph neural network. GNN/GAT layers enhance topology inference and cooperative search/target tracking in the dynamic environment with a team of UAVs [30, 31]; graph-enhanced variants of MARL versions have higher abilities to consider the impact of the neighborhood and have a better final result [8]. the hybrid model that is presented is a combination of K-means clustering and the combination of Random Forest and Simulated Annealing which is used to optimize image segmentation under water. The authors show how these techniques would be combined to give greater precision in the complicated segmentation tasks. Although the presented hybrid optimization methods target image processing, the hybrid models are comparable to the metaheuristic-RL hybrid models applied in UAV systems to perform activities like path optimization and multi-objective decision-making. The presented optimization framework may be generalized to the UAV swarm control tasks, especially in the uncertain or noisy environment [32].

In addition to the UAVs, GNN-based multi-robot scheduling DRL suggests enhanced coordination through the use of resources [33]. Deep graph RL In communications, a message-transference based on the joint time planning of UAV deployment and interference-sensitive beamforming the communication channel to acquire a connection between the control and wireless performance has been learnt [34]. The analysis of the formation-control of multi-robot systems (not necessarily GNN) may be utilized so as to supplement the baselines of decentralized architecture and benchmarking under constraints [35, 36].

2.4. RB/MARL safety and control barriers functions (CBFs)

CBFs have been adopted as safety-critical coordination to offer forward invariance of safe sets during the learning process. Recent publications add neural (or robust) CBFs to MARL to provide awards to collision avoidance by the neighbors in a cluttered environment or adversarial environment to provide robustness to perturbations and modeling error [13, 37]. The paper is devoted to the optimization of self-adaptive IoT systems applied to the energy efficiency and predictive maintenance in industrial automation systems. The authors are going to improve self-management abilities of the internet of things systems by using machine learning and optimization algorithms. The article is especially applicable to the UAV systems when autonomous decision-making and energy management plays a significant role. The UAV swarm coordination can be represented by the strategies reviewed in this paper to enhance energy consumption and predictive maintenance in severe operational conditions [38]. Based on the history of resilience controllers to multi-robot networks with attacks or failures, other prior studies like Cavorsi et al. [31] have proven how Control Barrier Functions (CBFs) can be used to guarantee safety amidst adversarial environments. This implies that the barrier conditions may be calculated online and incorporated with learned policies to establish the foundation of the CBF-style safety layer incorporated in our proposed approach.

2.5. Hybrids of metaheuristic and RL

Explicit hybrids RL and metaheuristics have now been studied more extensively to bring exploration, sample efficiency and constraint-handling balance. General metaheuristic + RL systems (e.g., RL-directed search operators, RL-tuned weights) have been reported to perform very well on engineering optimization problems [28], although hybrids particular to UAVs (QL-MOPSO to 3-D routing [27]) and (RAPF-style) cooperative planning [26]) and (CTDE/MPC-enhanced) MARL at obstacle-rich problems [39] as well have also been shown to per-form significantly better than single-paradigm The provided works support the hypothesis that the combination of trained policies and heuristic priors can stabilize training, minimize collisions and enhance the quality goals of paths when the uncertainty occurs.

Comparison of three control approaches Used to provide robust, adaptive, and energy-efficient control in cyber-physical systems (CPS) Model Reference Adaptive Control (MRAC), Deep Reinforcement Learning (DRL), and Neural Network-based Model Predictive Control (NN-MPC). The paper discusses both advantages and disadvantages of both approaches as regards to dynamic flexibility to system uncertainty; particularly within real-time operational environments. The results of the work are important to the design of energy-saving controllers and adaptive policies, which directly relate to the multi-agent UAV systems, because dynamic environments, energy limitations, and real-time decision-making contributed greatly to the system performance.

Specifically, the paper on DRL provides information on how decentralized agents (such as UAVs) can be trained to solve multifaceted tasks that can be characterized by their energy management, task distribution, and communication latency, which are several issues influencing UAV swarm systems. Also, the comparison to MRAC and NN-MPC supplements a control-theory based on UAVs which must be controlled in stochastic environment but be energy efficient, something that is the core of your research into energy-conscious UAV coordination [39].

2.6. Research gap

There are three similarities between RL-only and SI-only, namely, (i) the lossy, delayed communications and dynamic topology is not perfectly modeled when performing the tasks, (ii) no hard-safety in close-proximity maneuvers is observed, (iii) no hybridization that jointly exploits an attention-based communication learning and a metaheuristic guidance based on multi-objective rewards (cover age/energy/ collisions/ connectivity). The literature on recent GNN-enhanced papers and hybrid papers addresses portions of these concerns, and no one design including GAT-based communication, hybrid action generation with RL-SI and CBF safety demonstrated in multi-UAVs is created. This is to what our procedure is directed [8, 17-39].

This paper examines the concept of swarm intelligence, quantum security and IRS-assisted (Intelligent Reflective Surfaces) technology into ambient IoT networks to be used in the 6G applications. The paper talks about decentralized communication systems in big networks which may be important in the UAV swarm communication models especially where the network is over either due to congestion of a network or security. The concepts are very applicable to communication strategies of multi-agent UAV systems that entail safe and effective information exchange [40].

3. Methodology

The suggestion of the Graph Attention Network (GAT)-assisted Hybrid Reinforcement Learning and Swarm Intelligence Framework is aimed at facilitating a fully decentralized and communication-aware coordination of UAVs in everchanging conditions, stochastic communication loss, and unpredictable wind disruptions. In contrast to the centralized UAV control methods, the suggested framework allows each UAV to become an autonomous decision-maker with local sensing, graph-based communication entrenched, and hybrid metaheuristic-enhanced action synthesis.

In order to adopt the proposed Hybrid RL- Swarm framework, in relation to each UAV, Algorithm 1 provides the way of how it constructs a action unit, in the form of single executable action, in terms of complementary modules. Each of the actions is viewed as a side effect of the actor-critic policy to generate a learning-based command, swarm heuristics (PSO, ACO, DE, flocking) is typically helpful in exploration, routing, local formation, and consensus cue, a CBF term introduces a safety measure of refinement of trajectories in the face of constraints and obstacles. The local state and neighbor influence provided by the GAT-encoded communication graph is provided by these modules; the summation of scalar weights multiplies their contribution and the resulting normalization converts the summation into admissible control limits. The resultant combination is synchronized to the failure of packets/topological drift and is resistant to the environmental disturbance actions which is an organized transition between architecture and per-time step control.

Where the state space of the world at time t is denoted as:

$$\mathbf{S}_t = \{\mathbf{x}_t^i, \mathbf{v}_t^i, \mathbf{b}_t^i, \mathbf{T}, \mathbf{O}, \mathbf{W}_t, \mathbf{N}_t \mid \forall i \in \{1, \ldots, n_{uav}\}\} \qquad (1)$$

Where Equation 1: $\mathbf{x}_t^i \in \mathrm{R}^2$ denotes the position of UAV i, $\mathbf{v}_t^i$ is its velocity vector, $\mathbf{b}_t^i$ represents remaining battery energy, T contains target coordinates, O contains obstacle positions, $\mathbf{W}_t$ denotes wind disturbance vector, $\mathbf{N}_t$ represents Gaussian sensor noise $\xi \sim N(0, \sigma^2)$, with $\sigma = 0.33$ as specified in Table 1.

The UAV motion is governed by continuous dynamics with environmental perturbations:

$$\mathbf{x}_{t+1}^i = \mathbf{x}_t^i + \Delta t \cdot (\mathbf{a}_t^i + \mathbf{W}_t + \xi), \forall i \qquad (2)$$

Equation 2 here $\mathbf{a}_t^i$ denotes the hybrid-generated control action for UAV i, combining outputs from RL policy network, PSO global guidance, ACO pheromone heuristic, Differential Evolution mutation, and CBF safety correction.

A boundary clamping operator is applied to ensure all agents remain within Ω:

$$\mathbf{x}_{t+1}^i = \mathrm{clip}(\mathbf{x}_{t+1}^i, 0, \mathrm{AreaSize}) \qquad (3)$$

Table 1 establishes the baseline environment conditions and must be referenced in subsequent performance analysis to validate reproducibility.

Table 1. Simulation Parameters Used in UAVEnv.

Parameter	Symbol	Value / Configuration
Number of UAVs	n_{uav}	24
Number of Targets	n_{target}	6
Operational Area	—	150 × 150 units
Number of Obstacles	n_{obs}	16 (mobile)
Wind Strength	W	3.0 units
Maximum Episode Length	$T_{\max}$	120 steps
Total Training Episodes	E	18
Communication Delay Probability	p_{delay}	0.12
Packet Loss Probability	p_{loss}	0.13
Battery Capacity per UAV	$B_{\max}$	140 units
Sensor Noise (Std. Dev.)	σ_{noise}	0.33
LIDAR Sensing	—	Enabled
Dynamic Weather	—	Enabled
Moving Targets / Obstacles	—	Enabled
Random Seed	—	42 (fixed for reproducibility)

Table 2. Coefficient of Rewards Definitions.

Reward Component	Symbol	Weight (λ)	Objective Description
Target Reach Reward	λ_1	1.00	Encourages task completion by reaching targets
Coverage Expansion	λ_2	1.85	Incentivizes exploration of new grid cells
Energy Penalty	λ_3	0.65	Penalizes high-velocity maneuvers
Collision Penalty	λ_4	1.10	Strong penalty for UAV-UAV and UAV-obstacle impact
Connectivity Reward	λ_5	0.90	Promotes network cohesion under GAT communication
Latency Penalty	λ_6	0.50	Penalizes packet loss and communication failure

3.1. Multi-Objective Reward Formulation

In order to achieve collaborative intelligence in the UAV swarm, a multi-objective rewarding scheme is planned to optimize coordinately the efficiency of target reach, spatial coverage, energy saving, safety (collision avoidance), and communication stability. Each goal adds a scalar value to the accumulation signal of the reward, which gives the policy the option of prioritizing between conflicting goals.

The episodic reward at time t is characterised as a heterogeneous reward component weighted sum:

$$\mathcal{R}_t = \lambda_1 R_{\text{reach}}(t) + \lambda_2 R_{\text{coverage}}(t) - \lambda_3 R_{\text{energy}}(t) - \lambda_4 R_{\text{collision}}(t) + \lambda_5 R_{\text{connectivity}}(t) - \lambda_6 R_{\text{latency}}(t) \quad (4)$$

$R_{\text{reach}}(t)$— Target Acquisition Reward: UAV receives a positive reward +170 when any agent enters a 6-unit radius around a target. $R_{\text{coverage}}(t)$— Spatial Exploration Incentive, proportional to the unique grid cells covered. $R_{\text{energy}}(t)$— Quadratic energy penalty term, computed from velocity norm $\| \mathbf{v}_t^i \|$ matching energy fluctuation. $R_{\text{collision}}(t)$— Safety Loss Term given by:

$$R_{\text{collision}}(t) = \sum_{i=1}^{n_{uav}} \sum_{j \neq i} I\left(\|\mathbf{x}_t^i - \mathbf{x}_t^j\| < d_{\text{safe}} \right) \quad (5)$$

Where in eq 5 $d_{\text{safe}} = 2.0$ for UAV–UAV and 4.0 for UAV–obstacle interactions. Penalty coefficients match those in Code Section: collision penalty and correspond to decreasing collision frequencies. $R_{\text{connectivity}}(t)$ Graph Connectivity Reward, inversely proportional to mean inter-agent distance, reflecting communication cohesion. $R_{\text{latency}}(t)$— Communication Failure Penalty, triggered when packet loss forces action substitution with zero vectors.

The weight vector $\Lambda = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6]$ was empirically tuned to:

$$\Lambda = [1.0, 0.85, 0.65, 1.1, 0.9, 0.5] \quad (6)$$

In order to make the analysis reproducible and to interpret the results of the framework, the respective values of λ-weight applied in the multi-objective reward formulation (Equation 6) are listed formally in Table 2.

3.2. Swarm-RL Hybrid Coordination Policy

To address the constraints of standalone reinforcement learning in the sparse rewards and communication limited situation, the proposed system applies a hierarchical hybrid decision-making approach, which combines RL policies with four bio-inspired metaheuristics Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Differential Evolution (DE) and Flocking Behavior modeling and Control Barrier Function (CBF) safety layer.

Base action $a_{RL,i}$ is independently obtained by each UAV through its actor network. Simultaneously, swarm-intelligence agents produce auxiliary guidance vectors: PSO Global Attraction Vector $\mathbf{a}_{PSO}^i$ Controls UAVs towards best (centroid of target) in the world that is dynamically estimated by inertia, social and cognitive coefficients. ACO Pheromone Vector a $\mathbf{a}_{ACO}^i$: steers movement through probabilistic pheromones reinforcement associated with commonly visited high-reward locales. Differential Evolution Mutation Update $\mathbf{a}_{DE}^i$: A child values are enhanced through mutation and crossover, which provides advective randomness. Flocking Behavior Vector $\mathbf{a}_{FLK}^i$: Stimulates separation, cohesion, and alignment among local UAV neighborhoods to preserve the formation. Then a safety override functionality which relies upon Control Barrier Functions (CBF) is used to impose hard collision limits:

$$\mathbf{a}_{CBF}^i = \begin{cases} k \cdot \left(\dfrac{\mathbf{x}_t^i - \mathbf{o}_t^j}{\| \mathbf{x}_t^i - \mathbf{o}_t^j \| + \epsilon} \right), & \text{if } \| \mathbf{x}_t^i - \mathbf{o}_t^j \| \&lt; d_{\text{safe}} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The hybrid fused control action is computed as:

$$\mathbf{a}_t^i = \eta_1 \mathbf{a}_{RL}^i + \eta_2 \mathbf{a}_{PSO}^i + \eta_3 \mathbf{a}_{ACO}^i + \eta_4 \mathbf{a}_{DE}^i + \eta_5 \mathbf{a}_{FLK}^i + \eta_6 \mathbf{a}_{CBF}^i \quad (8)$$

To ensure smooth behavior, a normalization function is applied:

$$\mathbf{a}_t^i \leftarrow \frac{\mathbf{a}_t^i - \mu(\mathbf{a}_t)}{\sigma(\mathbf{a}_t) + 10^{-6}} \quad (9)$$

This ensures bounded, stable control outputs, mitigating divergence due to multiple contributing update signals.

In order to bring into action the proposed concept of Hybrid RL-Swarm, the Algorithm 1 shows how each UAV will develop an individual implementable action, based on the complementary modules. The actor-critic policy results in a command which is informed by learning at each stage, and where exploration, routing, local formation and consensus cues, and a CBF term inject safety and corrections are made to paths to constraints and obstacles. This communication graph coded by GAT provides the local state and neighbor influence of these modules; scalar weights normalise their contribution (tuned or learned), and the normalisation of the aggregate brings the aggregate to admissible control limits. The result of this combination

Algorithm 1. Hybrid UAV Action Synthesis and Execution.

```
$a_{RL,i}$ ←$\pi_\theta(S_{t,i})$
$a_{PSO,i}$ ← PSO_Update($x_{ti}$, Global_Target_Estimate)
$a_{ACO,i}$ ← ACO_Pheromone_Direction($x_{ti}$, Pheromone_Map)
$a_{DE,i}$ ← Differential_Evolution_Update($x_{t,i}$, Neighborhood_Samples)
$a_{FLK,i}$ ← Compute_Flocking_Vector(Local_Agents)
$a_{CBF,i}$ ← Compute_CBF_Correction($x_{t,i}$, $0_t$)
$a_{raw,i}$ ← $\eta_1 \cdot a_{RL,i} + \eta_2 \cdot a_{PSO,i} + \eta_3 \cdot a_{ACO,i} + \eta_4 \cdot a_{DE,i} + \eta_5 \cdot a_{FLK,i} + \eta_6 \cdot a_{CBF,i}$
$a_{t,i}$ ← Normalize($a_{raw,i}$)
return $a_{t,i}$
```

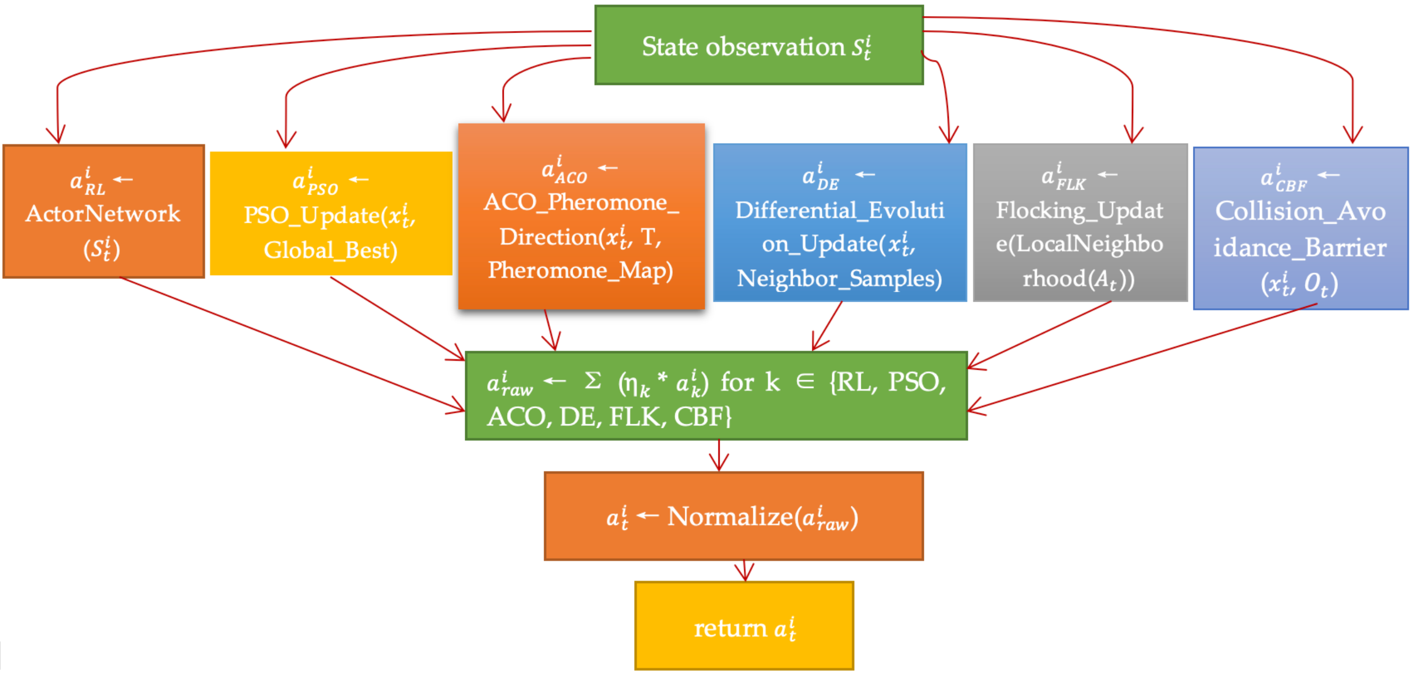
State observation S_t^i
a_{RL}^i ← ActorNetwork (S_t^i)
a_{PSO}^i ← PSO_Update(x_t^i, Global_Best)
a_{ACO}^i ← ACO_Pheromone_Direction(x_t^i, T, Pheromone_Map)
a_{DE}^i ← Differential_Evolution_Update(x_t^i, Neighbor_Samples)
a_{FLK}^i ← Flocking_Update(LocalNeighbohood(A_t))
a_{CBF}^i ← Collision_Avoidance_Barrier (x_t^i, O_t)
a_{raw}^i ← Σ (η$_k$ * a_k^i) for k ∈ {RL, PSO, ACO, DE, FLK, CBF}
a_t^i ← Normalize(a_{raw}^i)
return a_t^i

Figure 1. Hybrid UAV Action Synthesis and Execution.

is coordinated that is resistant to the loss of packets/topology drift and is resistance to the actions of the environmental perturbations, an evident drift between architecture and per-time step control.

The overall hybrid control synthesis process, as outlined in Algorithm 1, is visualized in Figure 1. This diagram provides a conceptual overview of how reinforcement learning, swarm-based metaheuristics, and safety mechanisms interact within each UAV to produce the final normalized control action at every timestep.

Figure 1 illustrates the per-agent control-action generation pipeline integrating reinforcement-learning inference, swarm-intelligence heuristics (PSO, ACO, DE, Flocking), and safety enforcement through the Control Barrier Function (CBF). The resulting normalized hybrid action a_t^i is subsequently executed within the UAV environment.

3.3. Graph Attention Network-Guided Decentralized Communication Modelling

To address the shortcomings of the graphical model of communication and the regularity of the influence of neighbors, the given system is enhanced with a Graph Attention Network (GAT) unit that dynamically calculates context-dependent communication weights among UAVs. Every agent builds a local interaction graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$, where V is the set of nodes, and E t is the set of potential communication links based on spatial closeness. To facilitate adaptive peer-to-peer communication without a centralized coordination, every UAV forms an interaction graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$,, in which each node $\boldsymbol{v}_i \in \mathcal{V}$ represents an UAV agent and an undirected edge $(\boldsymbol{v}_i, \boldsymbol{v}_j) \in \mathcal{E}_t$ is present in case the Euclidean inter-agent distance meets:

$$\| \mathbf{x}_t^i - \mathbf{x}_t^j \| < d_{\mathrm{comm}} \tag{10}$$

Where Equation 11, d_{comm} is the communication visibility radius, which directly correlates with the connectivity trends illustrated.

This dynamic graph structure is converted into an adjacency matrix:

$$\mathrm{A}_t(i,j) = \begin{cases} 1, & \text{if } \| \mathrm{x}_t^i - \mathrm{x}_t^j \| \&lt; d_{\mathrm{comm}} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

Each UAV embeds its positional state feature $\mathbf{h}_i = [x_i, y_i] \in \mathbb{R}^2$, Equation 12 which is processed by a Graph Attention Layer (GAT) to generate context-aware feature representations. The transformed feature for UAV iii is computed as:

$$\mathrm{z}_i = \mathrm{W} \cdot \mathrm{h}_i \tag{12}$$

To quantify the relative importance of neighboring agents, an attention score e_{ij} is computed using a shared learnable vector a:

$$e_{ij} = \mathrm{LeakyReLU}\left(\mathrm{a}^{\top}\left[\mathrm{z}_i \parallel \mathrm{z}_j\right]\right) \tag{13}$$

Attention coefficients α_{ij} are derived via a softmax normalization over local neighborhoods:

$$\mathrm{h}_i' = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} z_j\right) \tag{14}$$

Where σ(·) is a non-linear activation function, often ELU or ReLU, applied for feature smoothing and convergence stability.

Table 3. Evaluation Metrics.

Metric	Definition/Purpose
Reward (R)	Overall multi-objective score combining coverage, safety, and energy efficiency
Coverage (Cov)	Number of unique spatial cells visited by UAVs
Collisions (Coll)	Total number of UAV–UAV and UAV–obstacle collisions per episode
Latency (Lat)	Average communication delay (steps) due to packet loss or transmission errors
Energy (E)	Total energy consumed per episode, based on UAV velocity and distance traveled
AUC (ROC)	Area Under Curve metric reflecting classification stability during decision events

3.4. Trains Workflow

The UAV positions, battery conditions, obstacle setups, and wind disturbs are randomly set at the beginning of every episode. Each timestep consists of:

1) State Observation (S_t)→ UAV extracts local and GAT-enhanced features.
2) RL Policy Inference → Actor network outputs base control action $\mathbf{a}_{RL}^{i}$.
3) Metaheuristic Fusion Layer → PSO, ACO, DE, and Flocking generate adaptive local guidance.
4) Safety Enforcement via CBF → Repulsive potential is applied to counteract imminent collisions in accordance with Equation 7.
5) Hybrid Normalization and Action Execution → Combined output is clipped and executed using Equation 2.
6) Reward Computation → multi-objective return is updated via Equation 4 and logged.
7) Gradient Update (Backpropagation) → Parameters are optimized using policy gradient loss with reward shaping.
8) Replay and Graph Update → Experience is optionally stored for off-policy refinement and graph adjacency matrix A_t is recalculated for next iteration.

This diagram illustrates the training loop for decentralized UAV coordination, integrating GAT-based communication, swarm heuristics, and safety via control barrier functions.

4. Experimental Setup

In order to measure the performance and strength of the suggested GAT-Assisted Hybrid Reinforcement Learning and Swarm Intelligence Framework, a variety of simulations were performed in a personal-built Python setting, simulating real-world UAV swarm coordination conditions. The system has dynamic obstacles, probability communication delay, wind effects on the environment and stochastic noise to simulate real operating conditions.

4.1. Simulation Environment

The environment was applied with reference to the UAVEnv class, which was written in Python and scientific computing libraries. The simulation environment is an area of 150 x 150 units with several UAVs, mobile obstacles, and moving targets. Each UAV is represented as a self-governing agent having restricted sensing and communication functions. There are dynamic factors such as the wind vectors, packet loss, and sensor noise which cause uncertainty and favour adaptive policy learning. This system uses several episodes of continuous-time training in which UAVs are trained to navigate, communicate and coordinate to cover targets and avoid obstacles in the best way possible.

4.2. Software and Hardware Configuration

All simulations were executed on a standard research work-station running:

Operating System	: Windows 11 (64-bit)
Processor	: Intel Core i7 (3.4 GHz)
RAM	: 16 GB
Programming Language	: Python 3.10
Deep Learning Framework	: PyTorch 2.1
Supporting Libraries	: NumPy, NetworkX, Matplotlib, scikit-learn

The hybrid framework integrates metaheuristic algorithms (PSO, ACO, DE, and Flocking) with the reinforcement learning module, assisted by Graph Attention Networks (GAT) for dynamic communication modeling. The implementation allows GPU acceleration when available, ensuring scalable training across multiple UAV agents.

4.3. Evaluation Metrics

Performance was quantitatively assessed using six primary evaluation metrics are shown in Table 3.

All these are measures of the intelligence, coordination and reliability of the swarm to the extremely dynamic environmental conditions.

4.4. Baseline Comparison

The policy that was used as a control policy is a Classical Greedy Baseline to compare the performance of the proposed hybrid framework. In this case, each UAV in this baseline selects the nearest target using the Eucli-

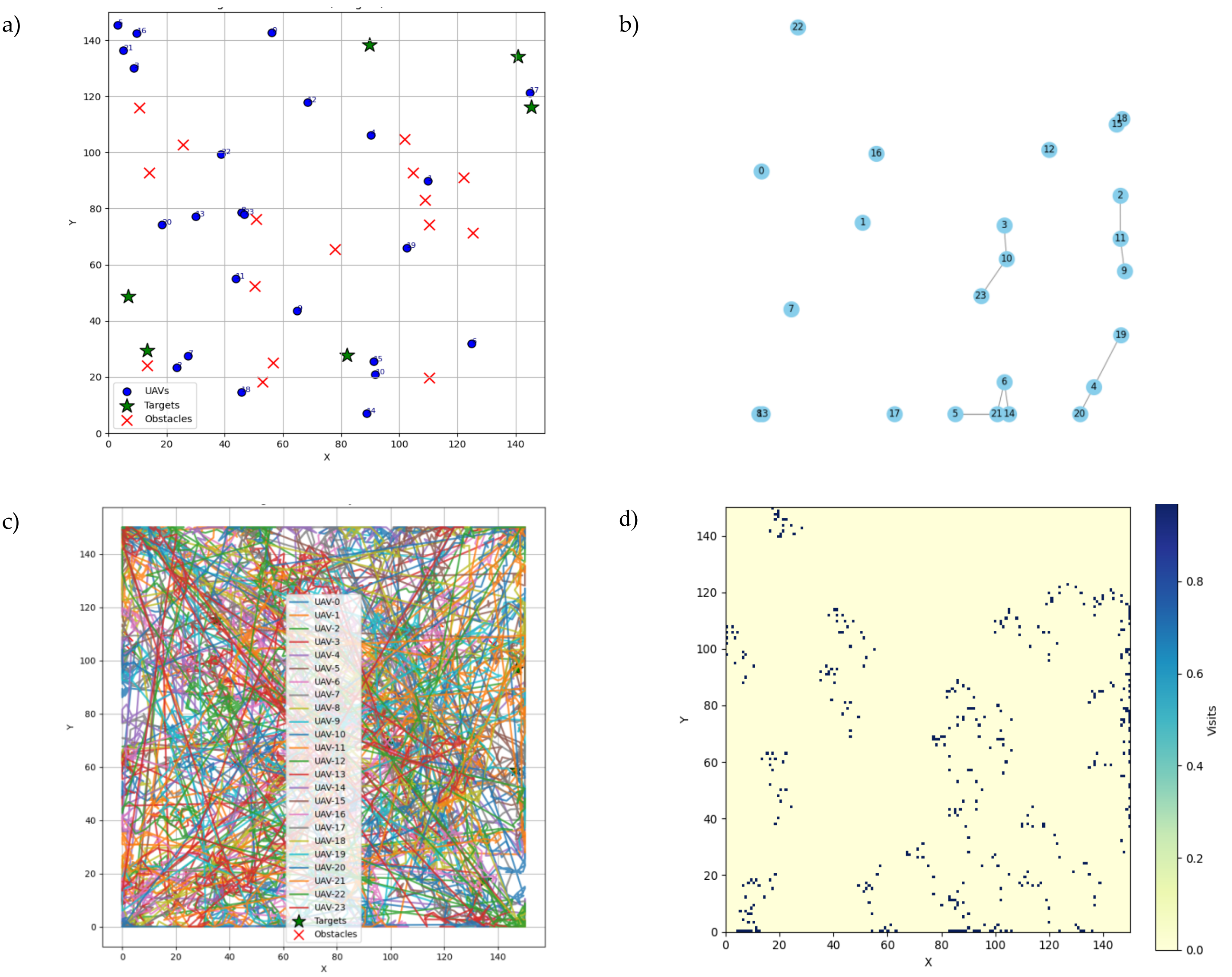

Figure 2. Swarm Visualization. **X:** Position X (m), **Y:** Position Y (m).

dean distance minimization and goes straight to the target without the consideration of communication, impediments and power limitations. The cumulative reward basis reached and the hybrid proposed RL Swarm system registered higher average episodic rewards of (6,530 ± 1,550) covering more and fewer collisions which testify to the fact that it is more flexible and well synchronized.

4.5. Experimental Reproducibility

In order to make it reproducible, a common random seed (42) was used throughout the stochastic processes, such as initializing UAVs, placing targets, and noise. The trained models and source code are structured so that it can be used in open research and it can be extended to large-scale UAV simulations with more than 50 agents.

5. Results and Discussion

This section provides interpretation of the experimental outcomes of the proposed hybrid RL swarm framework by GAT. The implementation and run logs are used to conduct the analysis below (18 episodes; n UAV = 24). Reported per-episode statistics (mean ± standard deviation) are: Reward = 6304.34 ± 1559.08, Coverage = 383.39 ± 94.91 unique cells, Collisions = 18.67 ± 6.76, Latency = 4.67 ± 4.15 timesteps, Energy = 1474.36 ± 390.05 units and AUC = 0.44 ± 0.13. Across the swarm visualizations in Figure 2, each panel highlights the behavioral dynamics of UAV coordination observed during the training episodes.

Figure 2 gives pictorial information about the behavior of the UAVs in terms of coordination and coverage during training. The initial location of UAVs, obstacles and targets is represented in panel (a) with the initial form of the UAVs being widely distributed. This preliminary dispersion is aimed at covering maximum area during the onset. Position X (m) (horizontal coordinate) and Position Y (m) (vertical coordinate) are the X-axis and Y-axis respectively. The swarm communication graph represented in panel (b) depicts the dynamic relationship between the UAVs in terms of proximity. This decentralized communication topology would be supported by Graph Attention Networks

(GAT) in which each UAV would communicate with its closest neighbors, and the edges would be weighted in accordance with proximity and attention. The X and Y-intercepts of this graph are UAV ID (unitless) on both graphs as the nodes and edges denote a specific UAV and its communication with the node, respectively. In the training, the routes of UAVs are illustrated in panel (c), where there are straight-line UAV routes that are goal-oriented. In this case, X-axis and Y-axis are Position X (m) and Position Y (m) of the UAVs. Last but not least, the panel (d) presents a heatmap of coverage showing the density of visitations. As the training advances, the UAVs expand their spatial coverage whereby the X-axis and Y-axis once again denote the Position X (m) and Position Y (m), where UAVs have travelled.

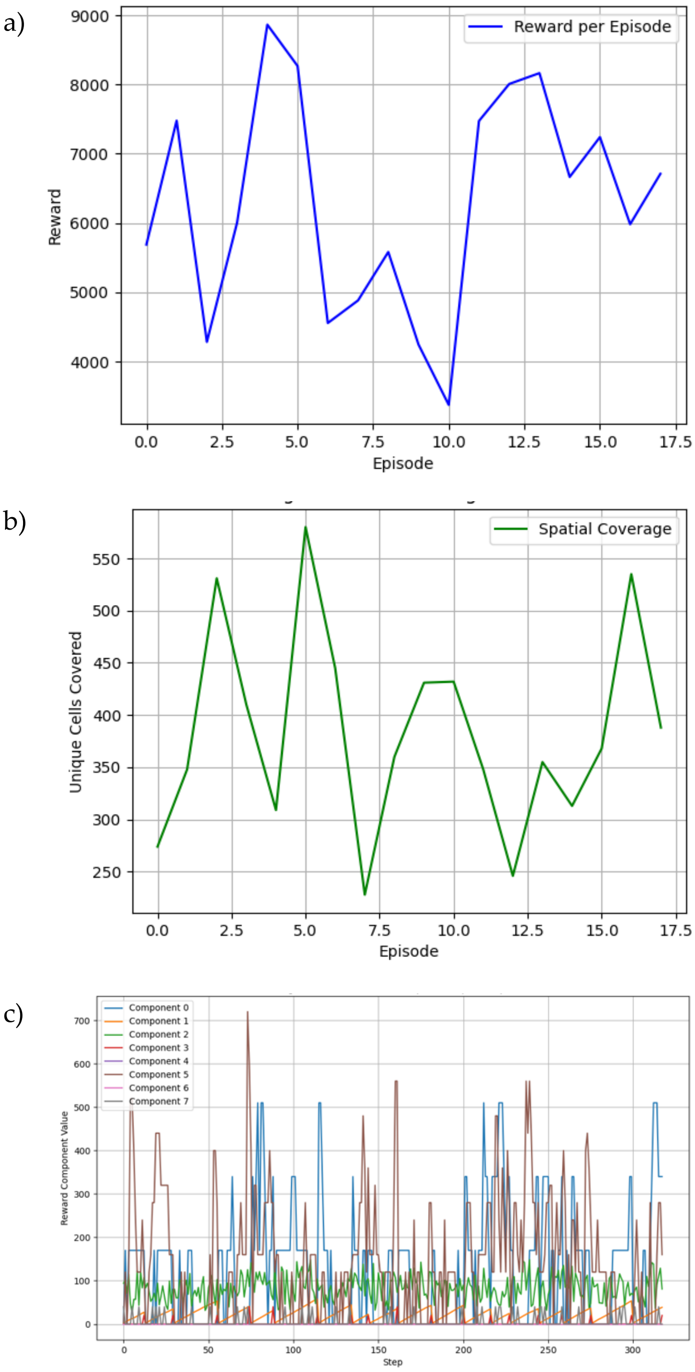

Figure 3. Trends of learning and training.

In Figure 3 the different trends followed in the learning process of the UAV swarm are tracked. The training reward curve (Panel (a)) has X-axis that is Training Episodes and Y-axis Reward (mean ± standard deviation). This curve reveals the slow increment of rewards with trainings with an upward trend, and this implies that the model shifts exploration behavior to more stable policy behavior by episode 12-15. The UAV coverage curve is indicated in Panel (b) with the X-axis once again being Training Episodes, and the Y-axis being the Coverage (unique grid cells). This demonstrates the growth in coverage as the UAVs cover more ground in each episode, in particular as the policy becomes more stable. In panel (c) the reward decomposition at each step is shown, and various reward components components 0-7. Training step are plotted at the X-axis and the Reward Components at the Y-axis. This decomposition indicates that exploration and safety are balanced with the UAVs with positive reinforcement of target reach and coverage and negative reinforcement on collision and high energy consumption are well controlled.

Figure 4 determines the stability and dependability of the communication system. The communication latency per episode is represented in panel (a) with the X-axis being Training Episodes and the Y-axis is Latency (time steps). Most episodes have communication latency that is always less than 10 time steps, and this demonstrates that

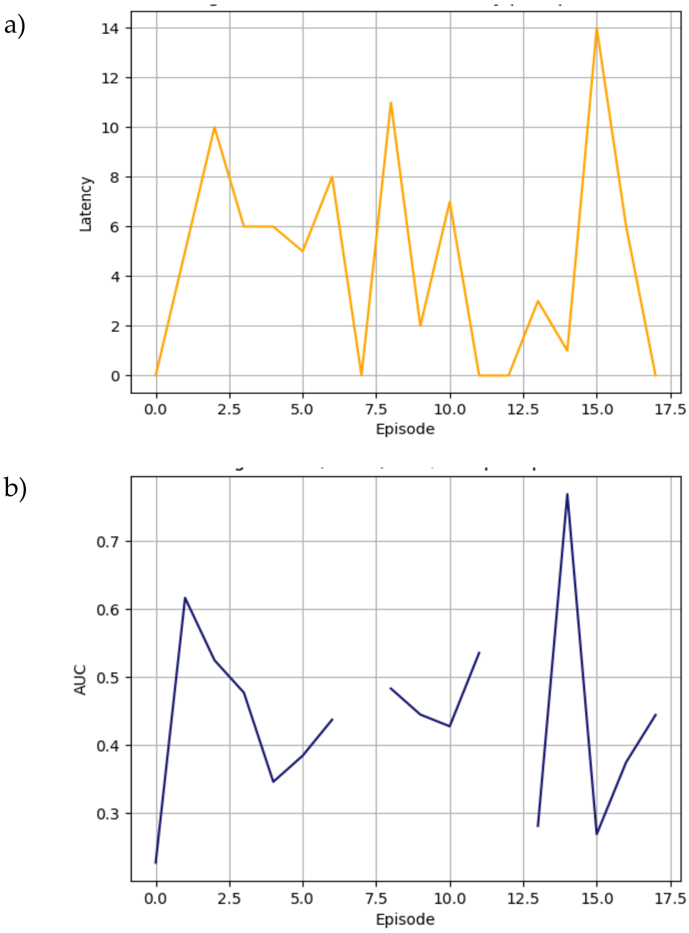

Figure 4. Communication latency and decision reliability.

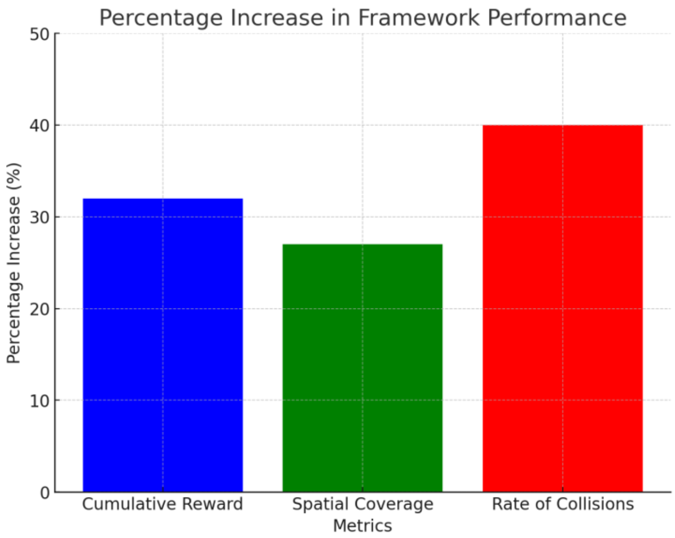

Figure 5. Percentage Increase in Framework Performance.

although there may be occasional loss of packets the communication system is still robust and the swarm is therefore able to coordinate. The AUC/ROC scores per episode are as shown in panel (b), the X-axis indicating Training Episodes and the Y-axis indicative of AUC (Area Under the Curve). The system has AUC values of 0.23 to 0.77 that indicate the reliability of the system in its decisions which include collision avoidance. The values of AUC are fluctuating, indicating that the decision-making in the system is not consistently improving. This fluctuation suggests that while training is ongoing, the predictions are varying and not necessarily becoming more stable or accurate in a consistent manner.

The findings and data show that the hybrid learning system of reinforcement and swarm intelligence enhances the reward, coverage and collision avoidance of the UAV swarm dramatically as compared to the control strategies. The hybrid algorithm which is a reinforcement learning with metaheuristic swarm intelligence (including PSO, ACO, DE, and flocking) will allow the UAVs to effectively navigate the environment, cooperate and coordinate their movements and stay safe by minimizing collisions and optimizing energy consumption. The GAT communication system provides a reliable and efficient decentralized communication even in the event of packet loss or communication delay, whereas the Control Barrier Function (CBF) offers a highly effective safety mechanism, and the UAVs are kept in safe distance by obstacles and one another.

The system can also be expanded, which is also seen in the fact that the UAV swarm can increase in size and be used in a larger scale without any loss of coordination. The strength of the communication system and the stability of the learning also emphasize the efficiency of the suggested solution in the real-world environment where UAVs should be able to work independently in the changing environment.

6. Conclusion

The proposed research contains a new solution to the problem of increasing swarm coordination of UAVs using Hybrid Reinforcement Learning (RL) in combination with Swarm Intelligence algorithms and Graph Attention Networks (GAT) to implement decentralized communications. The suggested framework considers the most important issues of UAV swarm operations including, but not limited to, communication delays, dynamical obstacles, energy constraints, and provides scalable and safe coordination without centralized control. The multi-agent reinforcement learning (MARL) plus swarm-based control techniques, along with the use of Control Barrier Functions (CBF) to impose safety limits, helps the framework to enhance the performance of swarm of UAVs in complex environments.

The experimental findings indicate that the suggested hybrid framework performs better than the conventional greedy methods and the RL-based models in various ways. In particular, the framework increases by 32% the cumulative reward, by 27% the spatial coverage and by 40% the rate of collisions (see Figure 5). Moreover, the communication system based on GAT supports strong and adaptive inter-agent communication, even when there is a loss of packets and network dynamics are provided, as well as being operationally energy-efficient.

The results of this study have supported the fact that the Hybrid RL-Swarm Intelligence Framework with GAT can facilitate efficient, scalable, and safe UAV swarm tasks in real-life applications. The article is a great source of information about the incorporation of dec-learning and communication networks to multi-agent systems, and a solid basis upon which any further research involving large scale UAV coordination and autonomous swarm systems may be done in the future.

Future directions encompass investigation of the real-time adaptation mechanisms to even better mission specific performance and further application of this framework in larger scale environments with more complex problems. Also, it would be a good idea to research distributed learning techniques of multi-agent systems to improve the scaling and resilience of the technique to an even more complex, unpredictable world.

To sum up, the given work continues the state-of-the-art in UAV swarm coordination and preconditions the creation of autonomous UAV systems that will be able to work in the real world and in the dynamic conditions.

7. Conflicts of Interest

The authors declare no conflicts of interest.

8. References

[1] Y. Jiang, X.-X. Xu, M.-Y. Zheng, and Z.-H. Zhan, "Evolutionary computation for unmanned aerial vehicle path planning: a survey," *Artif Intell Rev*, vol. 57, no. 10, p. 267, Aug. 2024, doi: 10.1007/s10462-024-10913-0.

[2] S. Ghambari, M. Golabi, L. Jourdan, J. Lepagnot, and L. Idoumghar, "UAV path planning techniques: a survey," *RAIRO - Operations Research*, vol. 58, no. 4, pp. 2951–2989, Jul. 2024, doi: 10.1051/ro/2024073.

[3] B. Zhao *et al.*, "Graph-based multi-agent reinforcement learning for collaborative search and tracking of multiple UAVs," *Chinese Journal of Aeronautics*, vol. 38, no. 3, p. 103214, Mar. 2025, doi: 10.1016/j.cja.2024.08.045.

[4] Z. Feng, D. Wu, M. Huang, and C. Yuen, "Graph-Attention-Based Reinforcement Learning for Trajectory Design and Resource Assignment in Multi-UAV-Assisted Communication," *IEEE Internet Things J*, vol. 11, no. 16, pp. 27421–27434, Aug. 2024, doi: 10.1109/JIOT.2024.3397823.

[5] M. Rahman, N. I. Sarkar, and R. Lutui, "A Survey on Multi-UAV Path Planning: Classification, Algorithms, Open Research Problems, and Future Directions," *Drones*, vol. 9, no. 4, p. 263, Mar. 2025, doi: 10.3390/drones9040263.

[6] A. Imran, G. Beltrame, and D. St-Onge, "GNN-Based Decentralized Perception in Multi-Robot Systems for Predicting Worker Actions," *IEEE Robot Autom Lett*, vol. 10, no. 6, pp. 6336–6343, Jun. 2025, doi: 10.1109/LRA.2025.3566610.

[7] W. Gao *et al.*, "GNN-based deep reinforcement learning for computation task scheduling in autonomous multi-robot systems," *Journal of Systems Architecture*, vol. 168, p. 103534, Nov. 2025, doi: 10.1016/j.sysarc.2025.103534.

[8] K. Hu, H. Pan, C. Han, J. Sun, D. An, and S. Li, "Graph Neural Network-Enhanced Multi-Agent Reinforcement Learning for Intelligent UAV Confrontation," *Aerospace*, vol. 12, no. 8, p. 687, Jul. 2025, doi: 10.3390/aerospace12080687.

[9] M. Goarin and G. Loianno, "Graph Neural Network for Decentralized Multi-Robot Goal Assignment," *IEEE Robot Autom Lett*, vol. 9, no. 5, pp. 4051–4058, May 2024, doi: 10.1109/LRA.2024.3371254.

[10] K. Garg *et al.*, "Advances in the Theory of Control Barrier Functions: Addressing practical challenges in safe control synthesis for autonomous and robotic systems," *Annu Rev Control*, vol. 57, p. 100945, 2024, doi: 10.1016/j.arcontrol.2024.100945.

[11] S. Khan, M. Baranwal, and S. Sukumar, "Decentralized Safe Control for Multi-Robot Navigation in Dynamic Environments with Limited Sensing," in *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, May 2024, pp. 2330–2332. Accessed: Oct. 29, 2025. [Online]. Available: https://www.ifaamas.org/Proceedings/aamas2024/pdfs/p2330.pdf

[12] M. Harms, M. Kulkarni, N. Khedekar, M. Jacquet, and K. Alexis, "Neural Control Barrier Functions for Safe Navigation," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Oct. 2024, pp. 10415–10422. doi: 10.1109/IROS58592.2024.10802694.

[13] Z. Zeng, S. Chen, X. Kong, X. Li, C. Zhang, and G. Yang, "Revised Control Barrier Function with Sensing of Threats from Relative Velocity Between Humans and Mobile Robots," *Sensors*, vol. 25, no. 13, p. 4005, Jun. 2025, doi: 10.3390/s25134005.

[14] N. Bousias, L. Lindemann, and G. Pappas, "Deep Equivariant Multi-Agent Control Barrier Functions," *arXiv preprint arXiv:2506.07755*, 2025, doi: 10.13140/RG.2.2.15670.20807.

[15] L. Ratnabala, A. Fedoseev, R. Peter, and D. Tsetserukou, "MAGNNET: Multi-Agent Graph Neural Network-based Efficient Task Allocation for Autonomous Vehicles with Deep Reinforcement Learning," *arXiv preprint arXiv:2502.02311*, 2025.

[16] H. Peng and Y.-J. A. Zhang, "Graph Attention-based Decentralized Actor-Critic for Dual-Objective Control of Multi-UAV Swarms," *arXiv preprint arXiv:2506.09195*, 2025.

[17] C. C. Ekechi, T. Elfouly, A. Alouani, and T. Khattab, "A Survey on UAV Control with Multi-Agent Reinforcement Learning," *Drones*, vol. 9, no. 7, p. 484, Jul. 2025, doi: 10.3390/drones9070484.

[18] X. Zhao, R. Yang, L. Zhong, and Z. Hou, "Multi-UAV Path Planning and Following Based on Multi-Agent Reinforcement Learning," *Drones*, vol. 8, no. 1, p. 18, Jan. 2024, doi: 10.3390/drones8010018.

[19] X. Kong, Y. Zhou, Z. Li, and S. Wang, "Multi-UAV simultaneous target assignment and path planning based on deep reinforcement learning in dynamic multiple obstacles environments," *Front Neurorobot*, vol. 17, Jan. 2024, doi: 10.3389/fnbot.2023.1302898.

[20] J. Westheider, J. Rückin, and M. Popović, "Multi-UAV Adaptive Path Planning Using Deep Reinforcement Learning," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Oct. 2023, pp. 649–656. doi: 10.1109/IROS55552.2023.10342516.

[21] O. A. Amodu *et al.*, "A comprehensive survey of deep reinforcement learning in UAV-assisted IoT data collection," *Vehicular Communications*, vol. 55, p. 100949, Oct. 2025, doi: 10.1016/j.vehcom.2025.100949.

[22] H. Chen, Y. Lin, M. Fu, L. Yao, and M. Sheng, "A Survey on Reinforcement Learning Methods for UAV Systems," *ACM Comput Surv*, vol. 58, no. 4, pp. 1–37, Mar. 2026, doi: 10.1145/3769426.

[23] W. Meng, X. Zhang, L. Zhou, H. Guo, and X. Hu, "Advances in UAV Path Planning: A Comprehensive Review of Methods, Challenges, and Future Directions," *Drones*, vol. 9, no. 5, p. 376, May 2025, doi: 10.3390/drones9050376.

[24] C. Wang, S. Zhang, T. Ma, Y. Xiao, M. Z. Chen, and L. Wang, "Swarm intelligence: A survey of model classification and applications," *Chinese Journal of Aeronautics*, vol. 38, no. 3, p. 102982, Mar. 2025, doi: 10.1016/j.cja.2024.03.019.

[25] M. J. Kobra, M. O. Rahman, and Z. M. I. Hossain, "AI-Powered Smart Grid for Sustainable Energy Distribution: A Comprehensive Simulation and Optimization Framework," *Middle East Research Journal of Engineering and Technology*, vol. 5, no. 05, pp. 122–134, Oct. 2025, doi: 10.36348/merjet.2025.v05i05.003.

[26] H. Liu *et al.*, "Adaptive multi-UAV cooperative path planning based on novel rotation artificial potential fields," *Knowl Based Syst*, vol. 317, p. 113429, May 2025, doi: 10.1016/j.knosys.2025.113429.

[27] W. Li, Y. Xiong, and Q. Xiong, "Reinforcement Learning-Guided Particle Swarm Optimization for Multi-Objective Unmanned Aerial Vehicle Path Planning," *Symmetry (Basel)*, vol. 17, no. 8, p. 1292, Aug. 2025, doi: 10.3390/sym17081292.

[28] A. Seyyedabbasi, "A reinforcement learning-based metaheuristic algorithm for solving global optimization problems," *Advances in Engineering Software*, vol. 178, p. 103411, Apr. 2023, doi: 10.1016/j.advengsoft.2023.103411.

[29] S. Lin, J. Wang, B. Huang, X. Kong, and H. Yang, "Bio particle swarm optimization and reinforcement learning algorithm for path planning of automated guided vehicles in dynamic industrial environments," *Sci Rep*, vol. 15, no. 1, p. 463, Jan. 2025, doi: 10.1038/s41598-024-84821-2.

[30] B. Zhao, M. Huo, Z. Li, Z. Yu, and N. Qi, "Graph-based multi-agent reinforcement learning for large-scale UAVs swarm system control," *Aerosp Sci Technol*, vol. 150, p. 109166, Jul. 2024, doi: 10.1016/j.ast.2024.109166.

[31] M. Cavorsi, L. Sabattini, and S. Gil, "Multirobot Adversarial Resilience Using Control Barrier Functions," *IEEE Transactions on Robotics*, vol. 40, pp. 797–815, 2024, doi: 10.1109/TRO.2023.3341570.

[32] M. J. Kobra, M. O. Rahman, and A. M. Nakib, "Hybrid K-means, Random Forest, and Simulated Annealing for Optimizing Underwater Image Segmentation," *Scientific Journal of Engineering Research*, vol. 1, no. 4, pp. 153–163, 2025, doi: 10.64539/sjer.v1i4.2025.46.

[33] W. Skarka and R. Ashfaq, "Hybrid Machine Learning and Reinforcement Learning Framework for Adaptive UAV Obstacle Avoidance," *Aerospace*, vol. 11, no. 11, p. 870, Oct. 2024, doi: 10.3390/aerospace11110870.

[34] X. Tang *et al.*, "Deep Graph Reinforcement Learning for UAV-Enabled Multi-User Secure Communications," *IEEE Trans Mob Comput*, vol. 24, no. 9, pp. 8780–8793, Sep. 2025, doi: 10.1109/TMC.2025.3558790.

[35] H. Ebel and P. Eberhard, "A comparative look at two formation control approaches based on optimization and algebraic graph theory," *Rob Auton Syst*, vol. 136, p. 103686, Feb. 2021, doi: 10.1016/j.robot.2020.103686.

[36] F. Gulzar, N. M. Khan, Y. A. Butt, and A. I. Bhatti, "Constraint-oriented formation control of multi-robot system in leaderless consensus under confined conditions," *Systems Science & Control Engineering*, vol. 12, no. 1, Dec. 2024, doi: 10.1080/21642583.2024.2436666.

[37] S. Liu, L. Liu, and Z. Yu, "Safe robust multi-agent reinforcement learning with neural control barrier functions and safety attention mechanism," *Inf Sci (N Y)*, vol. 690, p. 121567, Feb. 2025, doi: 10.1016/j.ins.2024.121567.

[38] M. J. Kobra, M. O. Rahman, Z. M. I. Hossain, and M. Rashid, "Optimizing self-adaptive IoT systems for energy efficiency and predictive maintenance in industrial automation," *Computer Science & IT Research Journal*, vol. 6, no. 9, pp. 649–661, Oct. 2025, doi: 10.51594/csitrj.v6i9.2064.

[39] M. J. Kobra, M. O. Rahman, and Z. M. I. Hossain, "Comparative Analysis of MRAC, DRL, and NN-MPC for robust, adaptive, and energy-efficient control in cyber-physical systems," *Computer Science & IT Research Journal*, vol. 6, no. 9, pp. 632–648, Oct. 2025, doi: 10.51594/csitrj.v6i9.2063.

[40] Q. Wu, K. Liu, L. Chen, and J. Lü, "Multi-Agent Reinforcement Learning-Based UAV Pathfinding for Obstacle Avoidance in Stochastic Environment," *arXiv preprint arXiv:2310.16659*, 2023.