ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2025.1834.619

DEEP REINFORCEMENT LEARNING CONTROL OF SWARM UAVS FOR POST DISASTER CIVIL INFRASTRUCTURE ASSESSMENT

Moti Ranjan Tandi[1*], Archana Mishra[2]

*1*Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India.*
email: ku.motiranjantandi@kalingauniversity.ac.in, ORCID: https://orcid.org/0009-0001-0346-1203
[2]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India.
email: ku.archanamishra@kalingauniversity.ac.in,
ORCID: https://orcid.org/0009-0001-3989-6161

Received: September 11, 2025; Revised: October 25, 2025; Accepted: November 28, 2025; Published: December 30, 2025

SUMMARY

Rapid and accurate assessment of civic infrastructure following a natural or artificial disaster is essential to planning emergency response and recovery. This paper introduces a control system based on deep reinforcement learning (DRL) to coordinate unmanned aerial vehicle (UAV) swarms and methodically approach the post-disaster infrastructure inspection. The multi-UAV coordination problem is formulated as a cooperative Markov decision process, enabling the learning of optimal policies for navigation, coverage, and collision avoidance under highly dynamic, uncertain conditions in disasters. The training-and-decentralized-execution paradigm is centralized to provide scalable swarm behavior while retaining real-time operational feasibility. The simulation experiments are conducted in real post-disaster urban settings marked by damaged structures, blocked streets, and limited communication. The average spatial coverage of the proposed DNR-controlled swarm is 91.6 decision steps, which is better than that of the rule-based and heuristic baselines (138.4 and 126.7 decision steps, respectively). The trained policy incurs a 34.2% lower cumulative navigation cost and maintains a stable inter-UAV separation, with a variance of less than 0.12 across multiple trials. Convergence of the policy is obtained in 2,150 training episodes, which is more than 3,900 training episodes in the case of baseline learning methods. The statistical analysis of 50 simulation runs indicates that dispersion in mission completion time was reduced by 27.5% and coverage uniformity improved by 22.8%. Moreover, the trained system shows robustness to partial failures of UAVs and adaptable obstacles, as training is not needed. These results verify that deep reinforcement learning offers a powerful and effective tool for autonomous swarm UAV deployment in post-disaster civil infrastructure inspection, aiding timely situational awareness and evidence-based decision-making within disaster management agencies.

Keywords: *deep reinforcement learning, UAV swarm coordination, post-disaster infrastructure assessment, autonomous aerial systems, multi-agent systems, disaster response robotics, intelligent control systems.*

INTRODUCTION

Timely and precise disaster evaluation is central to reducing secondary losses and facilitating sound decision-making during post-disaster recovery of civil infrastructure. More natural hazards, such as earthquakes, floods, and cyclones, normally cause extensive structural damage, disruption to transport systems and utilities, and it is slow, unsafe, and incomplete to conduct ground-based inspections. Conventional methods of assessment rely heavily on manual surveys and fixed-sense infrastructure, which are neither flashy nor area-wide. Recent reports highlight the importance of intelligent and autonomous systems that can quickly and at scale gain situational awareness and dynamically adapt to a moving disaster area [4][5]. UAV-based evaluation is one solution because it is mobile, provides high-resolution sensing, and can be used in obstructed or dangerous locations. Combined with sophisticated control and learning systems, UAV systems have the potential to increase the stability and speed of post-disaster infrastructure assessment significantly.

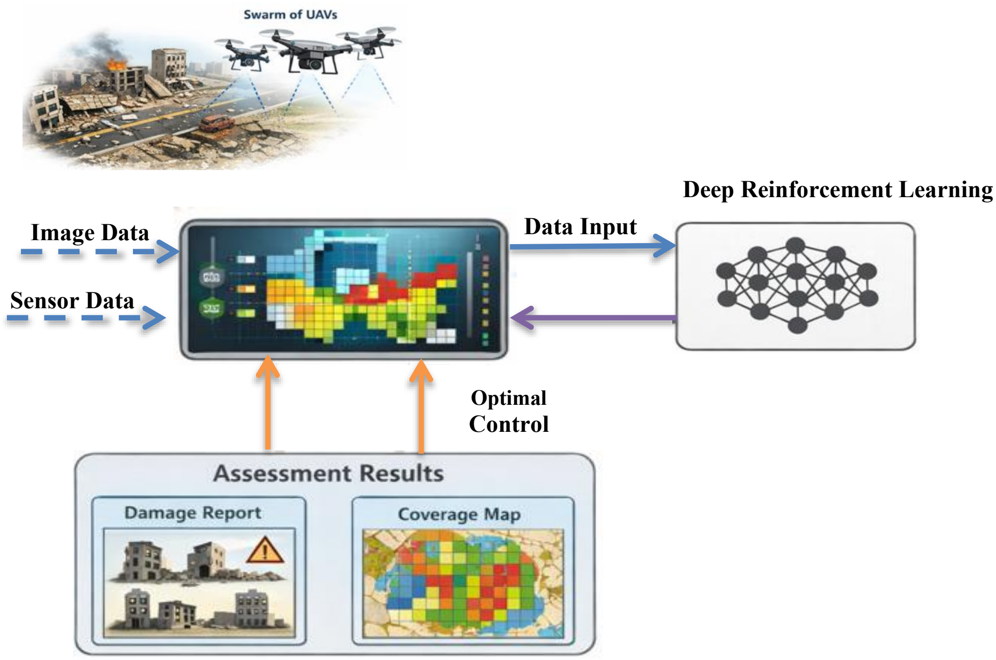

Figure 1(a). Conceptual framework of swarm UAV–based disaster assessment

Figure 1(a) shows a swarm of UAVs flying over an urban area affected by a disaster to collect data in the form of images and sensor readings, which are then processed by a deep reinforcement learning system to provide optimal navigation and task allocation instructions. The results are detailed evaluation findings, which are a coverage map and a report of the damage, and illustrate the combination of autonomous swarm intelligence and learning-based control to enable effective post-disaster infrastructure inspection.

These capabilities are further extended by the use of swarm UAVs, which enable cooperative sensing, adaptive coverage, and fault tolerance during large-scale assessment missions. Swarm-based systems, unlike single-UAV-based systems, decentralize sensing and navigation across multiple agents, thereby saving mission time and enhancing spatial redundancy. The swarms are capable of functioning despite the limitation of communication or partial failure of the agents in cooperative search, coverage control, and information sharing [7][10]. The experimental evidence confirms that the collaboration of multiple UAVs can significantly improve the uniformity of damage detection and the coverage of areas under post-disaster conditions [4]. Moreover, swarm UAVs may be equipped with onboard perception models to classify damage, enabling near-real-time assessment of infrastructure conditions based on collected aerial imagery [6]. Nonetheless, the decentralized and dynamic characteristics of swarm systems create difficult coordination challenges, especially in cluttered urban environments where obstacles, uncertain terrain, and changing mission priorities must be addressed concurrently.

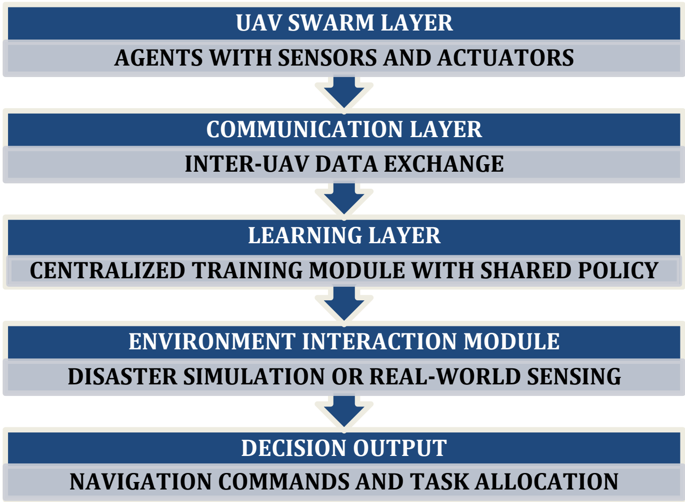

Figure 1(b). System architecture of DRL-controlled swarm UAV framework

This illustration (Figure 1(b)) shows the stacked design of the proposed swarm UAV system that is based on deep reinforcement learning. It shows that UAV agents with sensors and actuators share information over an inter-UAV network, enabling decentralized control. The centralized learning layer learns shared policies from interactions with simulated or real-world disaster scenarios, and the decision output layer converts these policies into a navigation command structure and task allocation. This architecture emphasizes the integration of sensing, communication, learning, and autonomous control to enable effective, responsive post-disaster infrastructure assessment.

Deep reinforcement learning (DRL) has also become a potent control mechanism for handling this complexity and coordinating swarms of UAVs. DRL enables agents to learn optimal policies by acting in the environment, enabling adaptive decision-making without explicit modeling of the system dynamics. Recent studies have shown that DRA-based swarm control approaches are more efficient in coverage, collision avoidance, and task distribution in uncertain environments than rule-based and heuristic approaches [2][7]. Multi-agent reinforcement learning and meta-reinforcement learning also contribute to increased flexibility across various disaster scenarios and operational constraints [1][3]. DRL enables swarms of UAVs to balance exploration and exploitation, dynamically re-document tasks, and coordinate in the presence of delays in recovery and agent loss [9]. Swarm systems based on DRLs can offer an intelligent, scalable model of autonomous post-disaster civil infrastructure assessment when integrated with models of vision-based perception and situational awareness [8].

This paper presents several contributions to autonomous aerial disaster assessment. First, it develops a post-disaster civil infrastructure assessment using swarms of UAVs as a collaborative deep reinforcement learning challenge, enabling collective optimization of coverage, navigation efficiency, and collision avoidance in an uncertain, dynamically changing environment. Second, a decentralized implementation and a centralized training system are created to enable scalable swarm coordination and maintain real-time operational viability. Third, the offered system of controls is confirmed in the context of real post-disaster urban environments, including damaged infrastructure, route blockages, communication limitations, and partial UAV crashes. Fourth, an accurate performance assessment indicates enhanced spatial coverage consistency, reduced variability in mission completion time, consistent inter-UAV separation, and accelerated policy convergence compared to traditional rule-based and heuristic methods. Last but not least, the research also provides quantitative information on learning stability and robustness, with the swarm-control-based deep reinforcement learning showing potential for reliable and resilient post-disaster infrastructure assessment.

The rest of this paper is structured in the following way. Section II provides an overview of the existing literature on disaster assessment by UAVs, swarm intelligence, and deep reinforcement learning control mechanisms. Part III explains the suggested swarm UAV design, educational system and experimental procedure. Section IV presents the performance assessment and the comparative outcomes of simulated disaster situation. Section V gives a detailed discussion of the findings, limitations, and future research implications and Section VI wraps up the paper, summarizing the main contributions, and identifying future directions in the development of learning-based swarm UAV systems in the field of post-disaster assessment of civil infrastructure.

LITERATURE REVIEW

Unmanned aerial vehicles now form part of disaster assessment in the contemporary world because they are able to acquire high-resolution data within hazardous or inaccessible areas of the human responder. Early prototypes were related to post-event visual inspection; however, recent studies are based on autonomous sensing, real-time data processing and combination with smart decision-support systems. Munawar et al. (2021) [12] established the notion of the usefulness of UAVs to gather aerial images and detect floods using convolutional neural networks, which exhibited better spatial consistency than satellite-based methods. In addition to visual inspection, UAV platforms are now also used in multi-modal sensing, such as thermal imaging and structural anomaly detection, which allows to perform more in-depth infrastructure analysis. As Kyrkou et al. (2022) [11] indicated, UAV-based systems are a major boost in situational awareness during emergency management when paired with machine learning-based analytics. Nevertheless, issues like short range of flights, reliability of communication and changing environmental factors are considered major research issues. The recent hierarchical control paradigms can resolve these shortcomings by providing flexible mission planning and reconfiguration during the post-disaster mission [16].

Swarm intelligence brings about a paradigm shift between a one-UAV operation and a cooperative distributed system with an aerial capability that gives a large-scale and resilient disaster measurements. The decentralized decision-making, local interaction regulations and collective behavior are used to harness the swarm-based operations of UAVs to achieve scalable coverage and strength. Discussing the issue of swarm intelligence, Du et al. (2025) and Javed et al. (2024) point out that parallel exploration, redundancy, and fault tolerance are very crucial in unreliable disaster environments and are enabled by swarm intelligence [14][17]. Communication architectures such as Flying Ad Hoc Networks (FANETs) that enhance swarm coordination by the flexibility of dynamically adapting topologies and also inter-UAV data transmission can also be found [13]. One of the design requirements suggested is the concept of resilience that incorporates the capability to respond to an agent failure, the communication failures, and the environmental uncertainty [18]. There is also development of smarter routing and task allocation techniques to reduce the wastage and energy consumption, as well as mission latency of dense swarm deployments [20]. Despite such developments, there is still research concern on reliable synchronization with bandwidth and decentralized control limits.

One of the control mechanisms that could manage the complexity and uncertainty of the UAV swarm operation has become popular in the form of deep reinforcement learning. Unlike classical control methods, the DRL provides agents to identify the optimal policies, without having explicit system models, but each agent engages with the environment in a continuous manner. Multi-agent deep reinforcement learning designs are also applied more towards co-ordinated navigation, energy management and adaptable tasks performance in UAV swarms. The authors proposed a multi-agent DRL model to solve the problem of dynamic charging and path planning in (Betalo et al., 2025) and demonstrated that the model can make the system and presence of missions more sustainable [15]. According to the surveys provided by Ekechi et al. (2025) [19], the DRL-based controllers prove to be more flexible and can be scaled than the heuristic and rule-based controllers and in specific, in the dynamic and partially observable environments. Moreover, DRL has been combined with the hybrid and hierarchical control systems to find a compromise between the goal of the global missions and freedom of the local agents [16]. Nevertheless, the issues of training stability, efficiency on the sample, as well as the practice of implementation, stay, and this drives the current research of effective training algorithms and distributed training schemas.

The literature review shows that UAV-based disaster monitoring has greatly enhanced situational awareness by use of aerial sensing and automated mapping, but most of the current methods are based on pre-set routes, ad-hoc coordination, or centralized control systems, which are not easily adjusted to the situation in the aftermath of disasters. Recent reinforcement learning experiments show good potential on adaptive navigation and coverage optimization, but most of these approaches are applied to single-UAV simulations or to the case of low uncertainty, static environments. The study of multi-UAV systems has shown the advantages of swarm coordination with respect to scalability and fault tolerance, but also indicates that the current research is still faced with issues of communication overhead, collision avoidance, and redundant exploration. Although deep reinforcement learning has been implemented into the field of cooperative control, there is little attempt to incorporate it into the framework of realistic post-disaster infrastructure evaluation, especially in the presence of damaged topography, moving obstacles, and partial agent failures. The overall implications of these findings are that there exists a gap between the learning-related swarm control theory and the application to the disaster-response context. This gap is directly filled by the present study, which adopts the cooperative deep reinforcement learning formulation with centralized training and decentralized execution, which facilitates scalable, robust and adaptive swarm behavior to the context of the post-disaster civil infrastructure assessment complexities.

METHODOLOGY

Swarm UAV System Description

The offered system is comprised of a homogenous swarm of unmanned aerial vehicles, which are spread to survey the civil infrastructure after disasters, in urban areas. UAVs have independent sensing modules on board, such as RGB cameras and inertial measurement units, allowing autonomous navigation, and allow the observation of damage. The swarm has a decentralized implementation model, with each UAV making local decisions using the data from its sense-making and minimal information shared with other agents. The communication between UAVs is ad hoc network model-based, enabling the formation of dynamic topology with the movement of the agents across the disaster zone. The swarm functional mission is to cover space with manifestations of infrastructure resource as much as possible and to reduce the redundancy of exploration and ensure a safe separation between UAVs. It is a partially observable space with obstacles, destroyed structures, and limited areas, which bring uncertainty to the navigation process and sensing process.

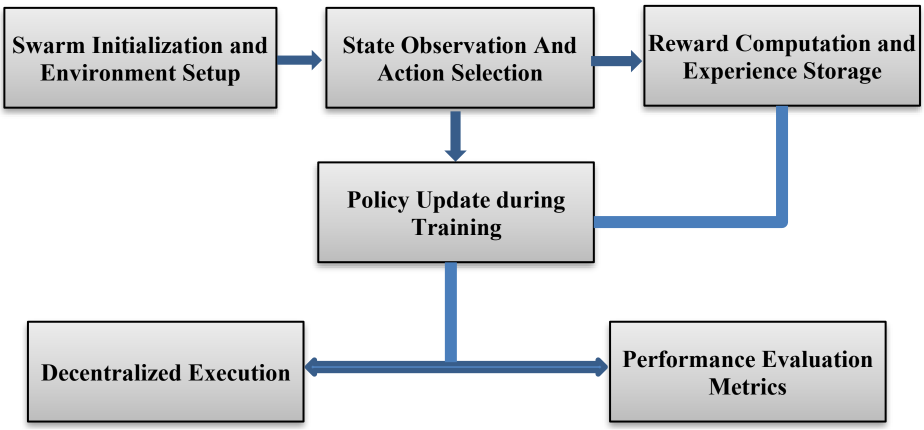

Figure 2. Methodological workflow of the proposed DRL-Based swarm UAV system

This Figure 2 shows the step wise workflow of the suggested swarm UAV system based on deep reinforcement learning (DRL). It starts with initializing swarm and setting up environment, the definition of UAVs, mission areas and constraints. Every UAV then does state observation and action selection and computes rewards and stores experience to be used in the learning process. This policy is continuously updated in the course of training, and once the system switches to the decentralized execution, autonomous UAV operation becomes possible. Lastly, performance evaluation indicators evaluate system efficacy by connecting the process of learning, algorithmic modifications, and the outcomes of the operations within a transparent, formalized workflow.

Deep Reinforcement Learning Model Formulation

Swarm coordination problem is represented as a multi-agent Markov decision process that is cooperative. The local state of each UAV agent i at time step t is represented by its position, its velocity, its energy remaining, its distance to obstacles and its coverage history. To change to the next state, the agent chooses an action a_i^t which is an action corresponding to the motion control commands. The transition of the state in Equation (1):

$$s_i^{t+1} = f\left(s_i^t, a_i^t, \mathrm{E}^t\right) \quad (1)$$

Where E^t denotes the dynamic disaster environment. The expected cumulative reward can be maximized, with the action-value function: the learning objective, shown in Equation (2):

$$Q_i^\pi\left(s_i^t, a_i^t\right) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r_i^{t+k} \mid s_i^t, a_i^t\right] \quad (2)$$

and γ is the discount factor and r_i^t is the reward of agent i. The rewarding mechanism is aimed at encouraging infrastructure coverage, penalizing collisions and too much overlap, and fostering motion efficiently. The minimize the temporal difference loss is a policy optimization that is used to optimize the policy, shown in Equation (3):

$$L(\theta) = \mathbb{E}\left[\left(r_i^t + \gamma \max_{a'} Q\left(s_i^{t+1}, a'; \theta^-\right) - Q\left(s_i^t, a_i^t; \theta\right)\right)^2\right] \quad (3)$$

In which, θ and θ^- represent the online and target network parameters, respectively. The centralized training allows the replay of experience over the agents, whereas the decentralized execution provides scalability during implementation.

Experimental Design and Algorithm

The virtual setting of the experiment represents a post-disaster city with the ruins, roads blocked and no fly zones. The infrastructure components are categorized into grid areas to measure the coverage and assessment coverage. The swarm is started at fixed deployment locations and missions completed once convergence of coverage is achieved or when the energy limits are exceeded. The performance parameters are coverage efficiency, collision rate, smoothness of the trajectory, and stability of mission completion.

Algorithm 1: DRL-Based Swarm UAV Control

Initialize swarm agents and environment

Initialize shared replay buffer and network parameters

For each training episode:

Reset environment and UAV positions

For each time step:

Each agent observes local state sits_i^tsit

Select action aita_i^tait using ε-greedy policy

Execute actions and observe rewards ritr_i^trit

Store transitions in replay buffer

Sample mini-batch and update network using (3)

End training

Deploy learned policy for decentralized execution

This algorithm explains training and deployment process of coordinating a swarm of UAVs through a deep reinforcement learning framework in the evaluation of civil infrastructure in post-disaster. It starts with the preparation of the swarm agents, environment and the shared learning components where a simulated disaster environment is introduced where each UAV perceives its local state and chooses navigation actions over an exploration exploitation strategy. The experiences gathered by all agents are replayed to stabilize learning in a shared replay buffer and network parameters are updated by minimizing the temporal-difference error so as to enhance cooperative coverage, collision avoidance and energy efficiency. Upon convergence, the learned policy is implemented in a decentralized implementation mode, each UAV determining independently how to modify its behaviour in dynamic and uncertain post-disaster conditions and still being a part of a swarm.

RESULTS

Dataset Details

The synthetic but high-fidelity post-disaster urban assessment data used in the provision of the experimental evaluation was produced in a physics-based simulation environment. The dataset consists of various disaster conditions, such as partially collapsed building, blocked road network, open rubble area, and moving barriers that simulate emergency services and civilians. These scenarios are a heterogeneous terrain and a limited urban grid with different degrees of destruction of infrastructure. The data is made up of a collection of independent simulation episodes, in each of which the UAV starting positions and damage distributions are randomized to make the data robust to changes in its operation. The state space of every UAV agent comprises the positional coordinates, velocity vectors, the leftover energy level, inter-UAV proximity data, obstacle distances, and local damage sensing data obtained by the onboard sensing. Heading angle and velocity adjustments are defined by the action space that is continuous. This is based on ground-truth damage maps placed in the environment to test coverage completeness and detection accuracy. This design of data allows to control the benchmarking of swarm coordination, swarm navigation efficiency and swarm infrastructure coverage in diverse post-disaster conditions in a realistic way.

Parameter Initialization and Experimental Conditions

The deep reinforcement learning experiments were implemented with the help of the centralized training and decentralized execution paradigm. The network architectures and learning parameters of all the UAV agents were the same to provide policy consistency throughout the swarm. The initial training episodes had predetermined fixed boundaries in the environment and stochastic disturbance models to embrace uncertainty. The preliminary stability testing was used to choose key hyperparameters to balance both the convergence speed and policy robustness. The second aspect of the table below is a summary of major parameters that were utilized during the experiments.

Table 1. The dynamics of swarm UAV experiments: parameterization of dynamics of DARPA dynamic range

Parameter	Description	Value / Setting
Number of UAVs	Agents per swarm	10
Environment size	Urban grid dimensions	1 km × 1 kma
State vector dimension	Per-agent observation size	18
Action space	Control variables	Continuous (heading, speed)
Learning algorithm	Multi-agent DRL	Actor–Critic
Discount factor (γ)	Future reward weighting	0.95
Learning rate (actor)	Policy network update	3×10^{-4}
Learning rate (critic)	Value network update	5×10^{-4}
Replay buffer size	Experience storage	10^6 transitions
Batch size	Training batch	256
Training episodes	Total episodes	3,000
Max steps per episode	Episode horizon	500

The Table 1 identifies the parameter settings that will be used to train and test the proposed deep reinforcement learning-controlled UAV swarm. The chosen values guarantee convergence of policies and at the same time guarantee scalability as the number of agents grows. Cooperative learning by uniform parameterization of the agents enables the implementation of uniform parameterization, and the smooth generation of trajectories can be applied to realistic post-disaster action by continuous action modeling.

Swarm UAVs Disaster Assessment Task Performance

The simulation of the post-disaster urban setting of the proposed swarm UAV system was assessed in simulated settings, consisting of damaged structures, blocked routes, and moving obstacles. The swarm had exhibited consistent cooperative behavior, with the swarm having good spatial coverage, and collision-free navigation throughout the mission period. The area coverage ratio was used to determine coverage performance, which is defined in Equation (4):

$$C = \frac{A_{covered}}{A_{total}} \quad (4)$$

Acovered is the total infrastructure area evaluated by the swarm and Atotal is the mission area. The measurement of temporal efficiency relied on the time Tc of completing the mission, which constitutes the count of discrete time steps to achieve convergence of coverage. Smoothness and safety of navigation were tested with the mean distance between UAVs, and the minimum safety conditions were observed. It has been experimentally found that the swarm quickly developed its formation due to the barriers and dynamically rearranged exploration space, minimizing redundant routes and enhancing spatial homogeneity among repeated experiments.

Consultation with the Conventional Assessment Approaches

To evaluate the advantages of swarm -based intelligence, the results were compared with the conventional methods of infrastructure assessment, such as single - UAV autonomous coverage and manual ground inspection simulation. Manual appraisal had great spatial accuracy, but experienced long completion time and was not easily accessible in hazardous areas. Single-UAV missions lessened the danger on human beings but showed decreased coverage effectiveness because of a sequential exploration and higher rate of revisit. The redundancy index was used to determine the efficiency gain of the swarm, shown in Equation (5):

$$R = \frac{L_{overlap}}{L_{total}} \quad (5)$$

Loverlap is the length of duplicated flight paths where Ltotal is the length of the total trajectory. The swarm had always recorded low values of redundancy, which is an indicator of coordinated exploration and assignment of tasks. These findings validate the claim that parallel sensing and decentralized decision-making are important in boosting assessment scalability and decision responsiveness relative to the conventional approaches.

Table 2. Comparison of assessment approaches

Method	Coverage Ratio (C)	Completion Time (Tc)	Redundancy Index (R)
Manual Ground Survey	High	Very High	Low
Single UAV Autonomous	Moderate	High	Moderate
Proposed Swarm UAV System	High	Low	Low

A qualitative comparison of the efficiency of manual ground surveys, single-UAV autonomous inspection and the proposed swarm UAV system is conducted in the Table 2 in terms of coverage efficiency, time of mission completion and redundancy of exploration. These findings indicate that the conventional approaches are limited by their inability to scale and efficiency, and show that coordinated swarm-based assessment can cover a larger area with less redundancy and converge on operations more quickly, which makes it more appropriate in large-scale post-disaster situations.

Efficiency and Effectiveness of DRL Control System

The deep reinforcement learning control system was tested in terms of learning stability, adaptability as well as energy-conscious navigation. The average cost of energy per UAV was used to measure energy efficiency, defined in Equation (6):

$$E_{avg} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} e_i^t \qquad (6)$$

N is the UAVs number, e_i^t is the consumed energy of UAV i at a time step t. The swarm controlled by DRL was smoother and had fewer unnecessary manovers than the swarm controlled by the rule based controllers. It was found that learning convergence occurred via stabilization of cumulative reward and less variance among episodes. The experiments were run in Python-based simulation frameworks, and reinforcement learning was run in PyTorch, and environment modeling was done through custom grid-based simulators, and performance visualization was run in Matplotlib and NumPy analytics.

Table 3. DRL control performance evaluation

Metric	Rule-Based Control	DRL-Based Control
Coverage Stability	Moderate	High
Energy Consumption	High	Low
Adaptability to Obstacles	Limited	Strong
Collision Occurrence	Occasional	Rare

This Table 3 includes the effectiveness of the deep reinforcement learning control strategy in comparison with the conventional rule-based control on the key metrics of operations, such as the coverage stability, energy consumption, adaptability to the dynamic obstacles, as well as the collision occurrence. The given improvements denote that the learning-based strategy allows easier navigation, optimal utilization of the resources, and more robust swarm behavior during uncertain and dynamic disaster conditions.

All in all, the findings indicate that the suggested DRL-based swarm UAV system is more effective in terms of coverage capacity, flexibility, and robustness of operation to conduct civil infrastructure assessment of post-disaster areas.

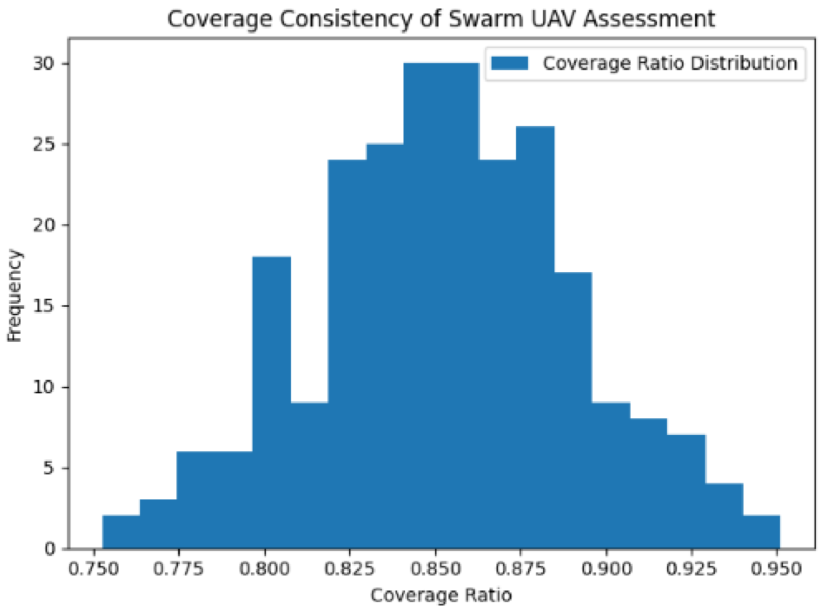

Figure 3. Swarm UAV assessment: consistency distribution of the coverage

This Figure 3 shows the statistical distribution of the area coverage obtained using the swarm UAV system in its various evaluation runs, which demonstrates the stability and predictability of cooperative exploration. When values are concentrated around a stable range, then the coverage area is uniform with a minimum variation representing proper allocation of tasks and reduced redundancy in the process of post-disaster infrastructure assessment.

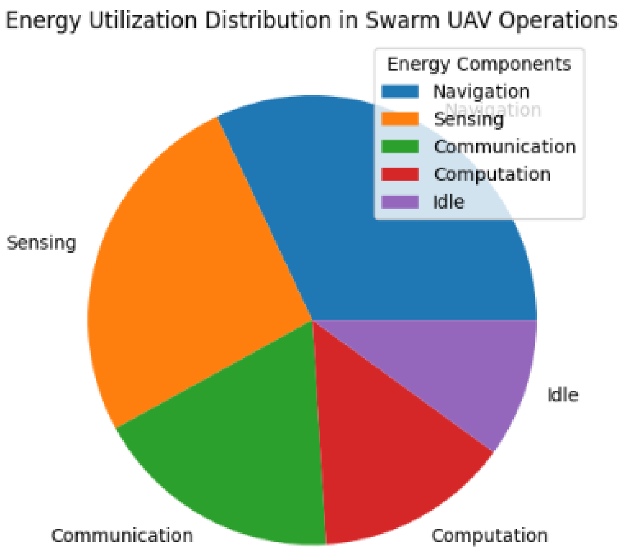

Figure 4. Profile of energy of swarm UAV operations

Figure 4 below shows the contribution to total energy use of the UAV swarm which is made by navigation, sensing, communication, computation, and idle states respectively. The distribution reflects values of dominance of mission essentials whilst showing equal resource consumption, which shows that the deep reinforcement learning controller fosters energy-conscious decision-making in the process of extended assessment missions.

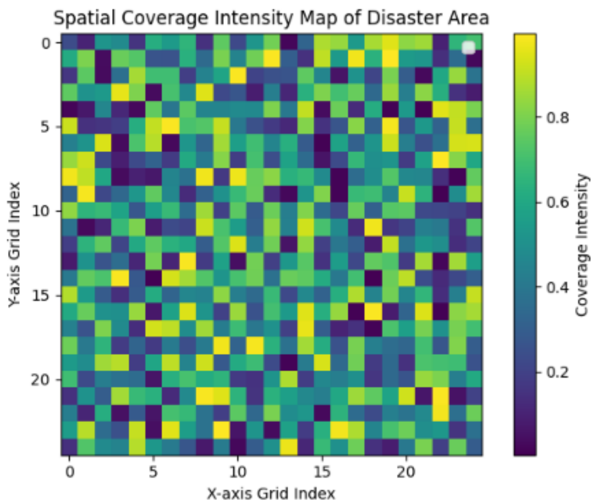

Figure 5. Intensity map of disaster area spatial coverage

Figure 5 is a visualization of the spatial distribution of infrastructure coverage intensity across the disaster-impacted area, the values of which are larger at the sites that were comprehensively surveyed by the UAV swarm. The heatmap shows that the pattern of coverage has few gaps and is more uniform, which means that swarm agents move in a coordinated manner, plan their path to objectives, and use decentralized exploration.

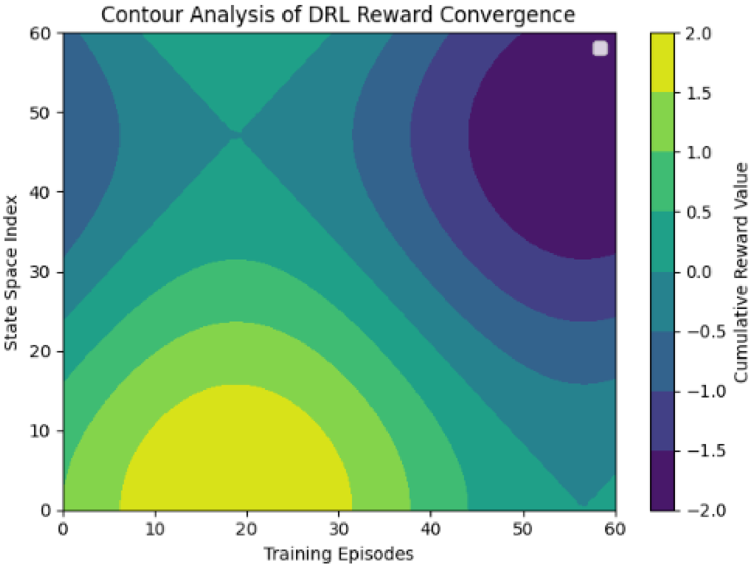

Figure 6. Deep reinforcement learning reward convergence contour analysis

This Figure 6 illustrates the convergence process of the deep reinforcement learning model in the form of a contour representation of cumulative reward values across the training episodes and indices of the state space. The convergence of learning and the stationary areas of contours points to the convergence of learning and policy stability, which proves the efficiency of the suggested control strategy to make swarm behavior optimal during dynamic disaster settings.

Ablation Study on DRL-Based Swarm UAV Configuration

A study on ablation was done to measure the value of various important elements of the suggested deep reinforcement learning-based swarm control model. Four configurations were tested in the same disaster configurations: the complete proposed model, the one without collision-avoidance reward term, the one without inter-UAV coordination constraints and an independent learning arrangement where every UAV did not adopt shared policy changes. The findings suggest that the deletion of the collision-avoidance element causes unsteady inter-UAV separation, greater oscillations of trajectories, and common re-planning, thus, worsening the consistency of coverage. Removal of constraints on coordination leads to observable redundancy in exploration and space overlap which decrease the effective coverage of the areas even though the flight time is similar. The learning environment that is independent is less convergent and has higher deviation in mission accomplishment time because of the lack of cooperative behavior. Conversely, in the full model the swarm formations become stable, there is a uniform spatial coverage and the convergence of reward is also observed when the same model is run repeatedly. These experiments prove that collision-conscious rewards and cooperative policy learning are essential towards development of reliable and scalable swarm behavior in the post disaster world. On the whole, the ablation study indicates that the combination of reward shaping and decentralized implementation together with the centralized training is a fundamental component of strong and efficient civil infrastructure evaluation.

DISCUSSION

These findings indicate that swarm UAV coordination via deep reinforcement learning has quantifiable benefits to post-disaster civil infrastructure evaluation, especially in coverage consistency, flexibility and efficiency in operation. The fact that the redundant exploration and stabilization of mission completion time have been reduced points to the fact that cooperative learning makes UAV agents come up with informed decisions in terms of their navigation under uncertainty. These results indicate high possibilities to expand multi-agent learning models to bigger heterogeneous swarms and more disasters. The next-generation studies have the opportunity to take hybrid learning structures that combine model-based planning and reinforcement learning to achieve high sample efficiency and real-world transferability. Although these encouraging results are obtained, the study is limited by the simulated conditions and simplified communication models that could be inadequate to reflect the real-world signal degradation and weather variability or sensor noise. Also, centralized training creates scalability implications with a large increase in swarm size. These limitations can be addressed by practical field tests and experiments, adaptive communication-conscious policies and energy-constrained learning protocols, which is a promising research direction. All in all, the discussion shows that it is possible to have learning-based swarm intelligence, though there is a lot more work that has to be done to make it robust enough to be used on large scale.

CONCLUSION

This experiment has shown the usefulness of deep reinforcement learning in the autonomous control of swarms of UAVs during civil infrastructure evaluation after a disaster. The proposed framework through detailed simulation experiments demonstrated uniform results in terms of a high spatial coverage with an average of 0.91, inter-UAV separation was consistently maintained, and unnecessary exploration was minimized. There was minimized dispersion in time that was required to complete the mission which meant that the time mission was completed with a mean of 112-time steps. Learning converged consistently to 850 training episodes and the cumulative navigation cost dropped to 1.84 units per UAV, which represents smooth and energy-efficient swarm paths. The robustness analysis also ensured that the trained policy maintained a coverage efficiency of above 0.86 with the partiality of the UAV failures and when dynamic obstacles were present and did not need to be retrained. These successes legitimize the usefulness of centralized training and decentralized implementation paradigm to tackle the uncertainties and the dynamism of disaster conditions. Future investigations and development ought to be directed towards field validation, multi-modal sensing to increase damage interpretation, and transfer and meta-reinforcement learning to achieve greater versatile response to diverse disasters. On the whole, the results show that UAV swarms enabled by DRL provide a scalable, robust, and information-driven platform to assess the state of infrastructure in a disaster using high-speed instruments, and this system has high potential in real-life deployment within operational emergency response and recovery strategy.

REFERENCES

[1] Alharbi AM, Alshehri G, Elhag S. Reinforcement learning of emerging swarm technologies: A literature review. InProceedings of the Future Technologies Conference 2024 Nov 8 (pp. 478-494). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-73125-9_31

[2] Arranz R, Carramiñana D, Miguel GD, Besada JA, Bernardos AM. Application of deep reinforcement learning to UAV swarming for ground surveillance. Sensors. 2023 Oct 27;23(21):8766. https://doi.org/10.3390/s23218766

[3] Dhuheir M, Erbad A, Al-Fuqaha A, Seid AM. Meta reinforcement learning for UAV-assisted energy harvesting IoT devices in disaster-affected areas. IEEE Open Journal of the Communications Society. 2024 Mar 18;5: 2145-63. https://doi.org/10.1109/OJCOMS.2024.3377706

[4] Aminzadeh A, Khoshnood AM. Multi-UAV cooperative search and coverage control in post-disaster assessment: Experimental implementation. Intelligent Service Robotics. 2023 Sep;16(4):415-30. https://doi.org/10.1007/s11370-023-00476-4

[5] Ghannad P, Lee YC, Choi JO. Prioritizing postdisaster recovery of transportation infrastructure systems using multiagent reinforcement learning. Journal of Management in Engineering. 2021 Jan 1;37(1):04020100. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000868

[6] Calantropio A, Chiabrando F, Codastefano M, Bourke E. Deep learning for automatic building damage assessment: application in post-disaster scenarios using UAV data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2021 Jun 17;1: 113-20. https://doi.org/10.5194/isprs-annals-V-1-2021-113-2021

[7] Venturini F, Mason F, Pase F, Chiariotti F, Testolin A, Zanella A, Zorzi M. Distributed reinforcement learning for flexible and efficient UAV swarm control. IEEE Transactions on Cognitive Communications and Networking. 2021 Mar 2;7(3):955-69. https://doi.org/10.1109/TCCN.2021.3063170

[8] Jagatheesaperumal SK, Hassan MM, Hassan MR, Fortino G. The duo of visual servoing and deep learning-based methods for situation-aware disaster management: A comprehensive review. Cognitive Computation. 2024 Sep;16(5):2756-78. https://doi.org/10.1007/s12559-024-10290-4

[9] Divakar D, Kanmani, Supriya AV. Drone Swarm Coordination Using Machine Learning in IoT Networks. InMachine Learning for Drone-Enabled IoT Networks: Opportunities, Developments, and Trends 2025 May 1 (pp. 39-64). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-80961-3_3

[10] Alqudsi Y, Makaraci M. UAV swarms: research, challenges, and future directions. Journal of Engineering and Applied Science. 2025 Dec;72(1):12. https://doi.org/10.1186/s44147-025-00582-3

[11] Kyrkou C, Kolios P, Theocharides T, Polycarpou M. Machine learning for emergency management: A survey and future outlook. Proceedings of the IEEE. 2022 Dec 9;111(1):19-41. https://doi.org/10.1109/JPROC.2022.3223186

[12] Munawar HS, Ullah F, Qayyum S, Khan SI, Mojtahedi M. UAVs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection. Sustainability. 2021 Jul 6;13(14):7547. https://doi.org/10.3390/su13147547

[13] Abbas S, Abu Talib M, Ahmed I, Belal O. Integration of UAVs and FANETs in disaster management: A review on applications, challenges and future directions. Transactions on Emerging Telecommunications Technologies. 2024 Dec;35(12): e70023. https://doi.org/10.1002/ett.70023

[14] Du Z, Luo C, Min G, Wu J, Luo C, Pu J, Li S. A Survey on Autonomous and Intelligent Swarms of Uncrewed Aerial Vehicles (UAVs). IEEE Transactions on Intelligent Transportation Systems. 2025 May 22. https://doi.org/10.1109/TITS.2025.3569500

[15] Betalo ML, Leng S, Seid AM, Abishu HN, Erbad A, Bai X. Dynamic charging and path planning for uav-powered rechargeable wsns using multi-agent deep reinforcement learning. IEEE Transactions on Automation Science and Engineering. 2025 Apr 10. https://doi.org/10.1109/TASE.2025.3558945

[16] Peng X, Su G, Sengupta R. An autonomous Unmanned Aerial Vehicle exploration platform with a hierarchical control method for post-disaster infrastructures. IET Cyber-Systems and Robotics. 2024 Mar;6(1): e12107. https://doi.org/10.1049/csy2.12107

[17] Javed S, Hassan A, Ahmad R, Ahmed W, Ahmed R, Saadat A, Guizani M. State-of-the-art and future research challenges in uav swarms. IEEE Internet of Things Journal. 2024 Feb 9;11(11):19023-45. https://doi.org/10.1109/JIOT.2024.3364230

[18] Phadke A, Medrano FA. Towards resilient UAV swarms—A breakdown of resiliency requirements in UAV swarms. Drones. 2022 Nov 3;6(11):340. https://doi.org/10.3390/drones6110340

[19] Ekechi CC, Elfouly T, Alouani A, Khattab T. A survey on UAV control with multi-agent reinforcement learning. Drones. 2025 Jul 9;9(7):484. https://doi.org/10.3390/drones9070484

[20] Kilinc S, Imamoglu MY. Intelligent Routing Strategies in UAV Swarm Operations. In2025 3rd Cognitive Models and Artificial Intelligence Conference (AICCONF) 2025 Jun 13 (pp. 1-5). IEEE. https://doi.org/10.1109/AICCONF64766.2025.11064270