Integrating Deep Reinforcement Learning for Initialization and Adaptive Pheromone Updates in Ant Colony Optimization for UAV Pathing

Mohamed A.damos[1], Wenbo Xu[2], Abdolraheem Khader[3],
Ali Ahmed[4], Mohammed Al-Mahbashi[5], Almuhannad S.Alorfi[6]
School of Resources and Environment, University of Electronic Science and Technology of China,
Chengdu 610054, China[1,2]
School of Computer Science and Engineering, Nanjing University of Science and Technology,
200 Xiaolingwei, Xuanwu District, Nanjing 210094, China[3]
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia[4]
School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, Shaanxi, P. R. China[5]
Faculty of Computing and Information Technology-Department of Information Systems,
Rabigh King Abdulaziz University, P.O. Box 344, Rabigh, 21911, Saudi Arabia[6]

***Abstract*—Unmanned Aerial Vehicles (UAVs) are indispensable assets for missions in dynamic and complex environments, requiring highly efficient path planning that simultaneously optimizes the often-conflicting objectives of minimizing flight distance, energy consumption, and mission time. While Ant Colony Optimization (ACO) is a recognized and effective metaheuristic for this domain, its performance is significantly constrained by a static, empirically-derived pheromone update mechanism, which prevents the algorithm from adaptively learning or optimally managing the search process. To overcome this critical limitation, this study introduces a novel DRL-Assisted ACO framework where a Deep Reinforcement Learning (DRL) agent is seamlessly integrated with the ACO to strategically determine the optimal paths under multi-objective constraints. This intelligent agent is tasked with learning the optimal, mission-specific pheromone update strategy. It achieves this by observing the performance of generated paths and receiving a sophisticated reward signal meticulously derived from the Analytic Hierarchy Process (AHP), which systematically weights the mission objectives. Validated through a simulated case study conducted in Khartoum State, Sudan, the DRL-Assisted ACO approach has demonstrably achieved superior performance, exhibiting marked gains in convergence speed and generating paths with a significantly higher overall multi-objective utility score, thereby delivering a robust and adaptive solution essential for high-stakes autonomous UAV operations.**

Keywords*—Deep Reinforcement Learning; Ant Colony Optimization; adaptive pheromone update; UAV pathing

I. Introduction

Unmanned Aerial Vehicles (UAVs) have become indispensable tools across diverse sectors, including disaster response, surveillance, and autonomous logistics [1], [2]. The successful execution of any UAV mission is fundamentally dependent on efficient path planning, which seeks to identify an optimal flight trajectory. In realistic scenarios, this constitutes a complex multi-objective optimization problem where the desired path must simultaneously minimize conflicting criteria, primarily flight distance, energy consumption, and total mission time [3].

Traditional approaches often rely on metaheuristic approaches, such as ACO and Genetic Algorithms (GA), which have demonstrated effectiveness in navigating complex search landscapes [4], [5]. However, traditional ACO is inherently limited by its fixed, empirical pheromone update rules. This static mechanism prevents the algorithm from adaptively learning the best search strategy, often leading to sluggish convergence rates and a suboptimal performance when attempting to balance the conflicting objectives of a UAV mission [6], [7]. While recent studies have explored integrating DRL to enhance initialization processes, thereby providing an initial population of high-quality paths, a significant research gap remains in leveraging DRL to dynamically control and optimize the core evolutionary mechanism of the ACO algorithm itself [8], [9].

Enhancing algorithms with Deep Reinforcement Learning (DRL) is critical for improving path-planning efficiency. For instance, Binghui Jin [10] proposed an effective method known as DDQN-ACO to address the complex challenge of coordinating multiple Unmanned Ground Vehicles (UGVs) in rugged 3D environments. This approach divides the problem into two key components: the Multi-UGV Path Planning (MUPP-DDQN) algorithm, which uses DRL to find the minimum time-cost path between tasks by considering terrain slope and actual driving speed, thereby generating an accurate cost matrix; and the Multi-UGV Task Assignment (MUTA-ACO) algorithm, which employs Ant Colony Optimization (ACO) to determine the optimal task sequence for each UGV, minimizing the total system time cost and achieving superior global optimization compared to existing methods. In a similar context, You Caihong [11], [12] proposed an RL-ACO hybrid model to address the limitations of static path planning in dynamic, real-time logistics environments. This framework constructs a dynamic directed graph using real-time traffic data, applies ACO for global search and high-quality initial path generation, and integrates a Deep Q-Network (DQN) agent for dynamic mid-route re-optimization in response to real-time events like congestion or road closures. In contrast, our proposed algorithm builds upon and refines the core parameters of these

models, offering enhanced efficiency and improved overall performance.

This study presents a novel approach combining DRL with ACO to address significant limitations in existing ACO algorithms [13]. The key innovation lies in the incorporation of a DRL agent that serves as an intelligent control system, tasked with the continuous challenge of identifying the optimal pheromone update strategy specific to the mission at hand. The agent learns by observing the performance of the paths generated by the ant colony and subsequently receives a reward signal [14]. Importantly, this reward is not arbitrary, but is carefully calculated using the AHP, which systematically integrates and weighs the various competing mission objectives into a single, cohesive score. The main contributions of this study are as follows:

- A hybrid DRL-Ant ACO approach in which a DRL agent is employed to control the pheromone update mechanism, moving beyond conventional static rules.
- The integration of the AHP to systematically formulate a multi-objective cost function, which drives the reward signal for the DRL agent.
- An empirical evaluation of the proposed methodology within a defined operational context, demonstrating its practical advantages in achieving faster, more optimal, and adaptive path planning solutions.

The structure of this study is organized as follows: Section I provides the introduction and motivation behind the research. Section II presents the proposed methodology, including the formulation of the multi-objective problem, the DRL-based initialization process, and the adaptive pheromone control model. Section III discusses the experimental setup and presents the results of the evaluation. Finally, Section IV concludes the study by summarizing the key findings and offering suggestions for future research directions.

A. Related Work

The optimization of UAV path planning, particularly in multi-objective optimization, has been an active area of research. This has led to the development of various hybrid algorithms combining metaheuristic techniques with machine learning to overcome challenges related to computational efficiency and robust multi-objective optimization.

The core advancement of our proposed DRL-ACO method remains its dual-integration of DRL to both generate a high-quality initial path set and, most critically, to learn the optimal, adaptive pheromone update value [15]. This comprehensive and deep integration contrasts sharply with other adaptive and hybrid ACO variants [16]. For instance, in the domain of adaptive parameter tuning, methods such as PF3SACO leverage classical techniques such as Particle Swarm Optimization (PSO) and Fuzzy Logic to adjust ACO parameters such as 1α and 2ρ.3 While adaptive, these methods rely on predefined metaheuristic interactions or empirical, rule-based inference systems. Our approach, by contrast, uses DRL to learn the optimal pheromone update directly from the search state, allowing the system to derive complex, non-linear control policies that are superior to any manually engineered or statically constrained adaptation mechanism [17], [18].

In contrast to recent hierarchical DRL-ACO models, such as ADACO-ATD3, which separate path planning (handled by ACO) and local motion control (managed by a dedicated DRL agent, such as ATD3 for real-time obstacle avoidance, this approach fundamentally optimizes the ACO engine itself [19]. In these hierarchical models, the ACO core remains static, with the DRL agent primarily refining local control mechanisms. In this approach, however, the DRL agent directly influences the global search process by dynamically controlling the pheromone update mechanism, an essential memory component of the ACO algorithm [20]. This DRL-driven adjustment significantly improves the global search quality and accelerates convergence by impacting the behavior of all subsequent ants across iterations [21].

Furthermore, when compared to other multi-objective DRL-ACO hybrids, such as NSACOWDRL which uses a Double Deep Q-Network (DDQN) agent for local search refinement or to modify ant selection probabilities, this DRL agent targets the pheromone update mechanism itself [14]. This allows our approach to function as a true meta-controller of the ACO's collective memory, rather than merely enhancing local path construction or solution selection heuristics. The ability to holistically and intelligently control the most influential parameter of the ACO, the pheromone update, is the key feature that enables superior and adaptively managed search performance [22].

In summary, the proposed DRL-ACO approach introduces a deeply integrated learning mechanism in which a DRL agent dynamically optimizes the pheromone update process, enhancing the global search capability and convergence efficiency of the ACO algorithm. Unlike previous adaptive or hierarchical hybrids, this approach functions as a true meta-controller of ACO's collective memory, enabling intelligent and adaptive management of its most influential parameter [23], [24]. Consequently, the framework achieves superior performance in multi-objective UAV path planning through a holistic and self-adaptive optimization strategy [25].

II. Methodology

UAVs are vital assets in surveillance, delivery, and mapping, especially in complex and dynamic environments. Efficient path planning for UAVs is a crucial, non-linear optimization challenge that requires balancing multiple, often conflicting, objectives, such as minimizing flight distance, mission time, and energy consumption. The DRL-APC methodology is structured into three phases: Problem Formulation, DRL-Assisted Initialization, and Adaptive ACO Optimization.This hybrid DRL-ACO approach, as shown in Fig. 1, aims to harness DRL learning and decision-making capabilities to provide a robust starting point for the ACO global search, thereby leading to more efficient convergence and higher-quality solutions for complex multi-objective UAV path planning problems.

A. Problem Formulation and Multi-Objective

1) Path representation: The mission space is modeled as a weighted graph $G = (V, E)$, where V are waypoints and E are possible flight segments. The UAV path P is a sequence of connected waypoints.

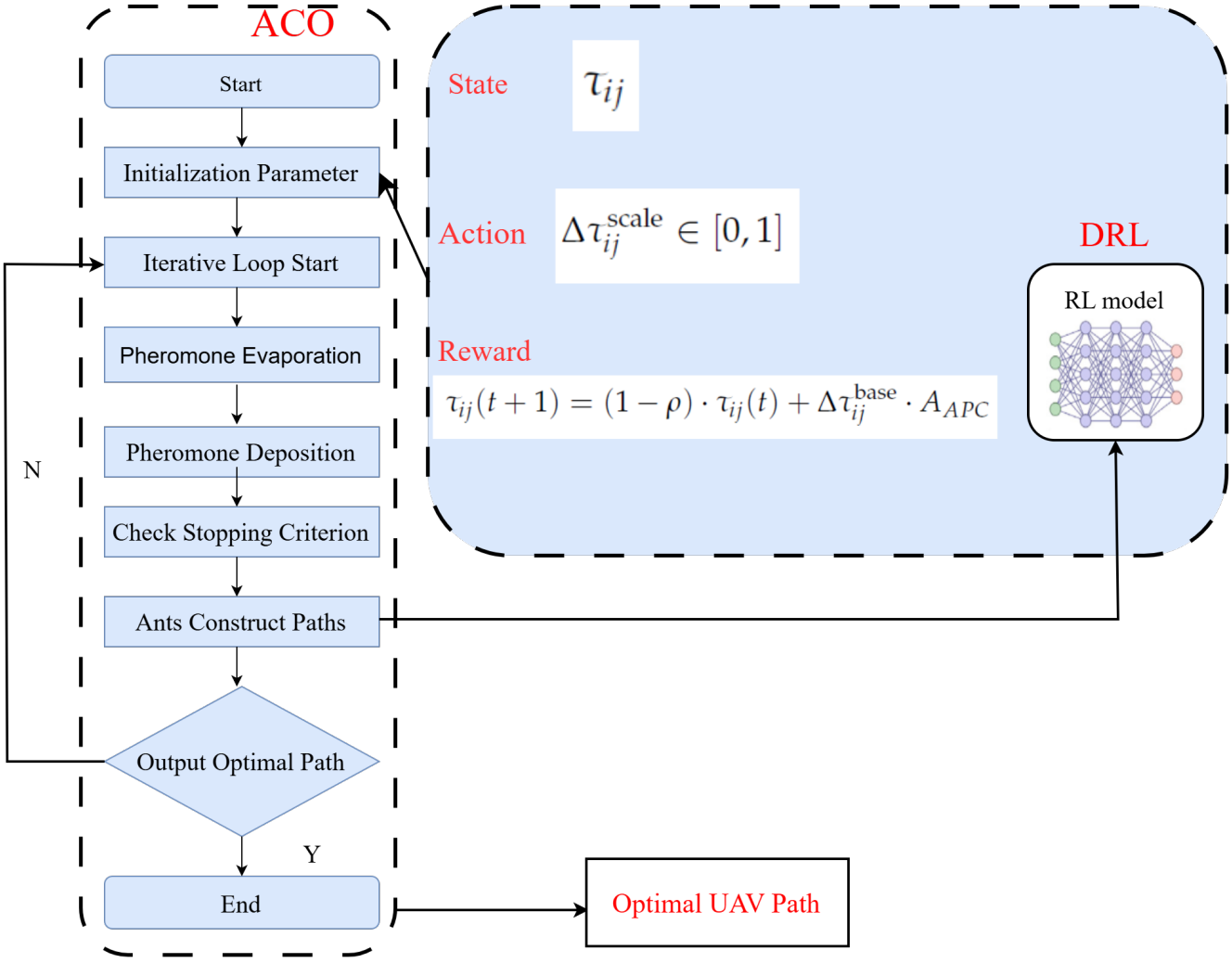

Fig. 1. Approach of the process for UAV path planning using RDL and ACO.

2) Multi-objective cost function C_{MOC} : The optimization objectives are normalized and combined using AHP-derived weights to form a single cost function for a path P:

$$C_{MOC}(P) = w_D \cdot C_D(P) + w_T \cdot C_T(P) + w_B \cdot C_B(P). \quad (1)$$

where,

C_D, C_T, C_B: Normalized costs (Distance, Time, Battery Consumption).

W_D, W_T, W_B: AHP-determined priority weights (summing to 1). The objective is to $\min C_{MOC}$.

B. Ant Colony Optimization (ACO)

ACO is a computational metaheuristic primarily employed for finding near-optimal solutions to discrete optimization problems, particularly those related to pathfinding. As a form of swarm intelligence, ACO is inspired by the cooperative foraging behavior observed in ant species.

The effectiveness of ant colonies in locating the shortest routes between their nests and food sources stem from their system of indirect communication, known as stigmergy. Ants deposit chemical signals, or pheromones, along their paths, which serve as odor trails. Subsequent ants follow these trails, using the scent concentration as a probabilistic guide to determine the most efficient route. Paths that are frequently traversed become reinforced with higher pheromone concentrations, thereby attracting more followers and establishing a positive feedback loop for optimal solutions.

The fundamental operation of the Ant Colony Algorithm can be broadly categorized into two iterative stages:

- Solution Construction: Artificial ants sequentially build solutions (a path) by making probabilistic choices. These choices are governed by the existing pheromone levels on the network edges and a local heuristic value.
- Pheromone Update: After a construction cycle, the pheromones along the utilized paths are modified. Pheromone levels are increased on components be-

longing to high-quality solutions (such as reinforcement) and simultaneously decreased across the entire network via evaporation to prevent premature convergence and encourage exploration. In essence, ACO harnesses the collective, decentralized decision-making of a swarm to select new path segments, assigning probabilities based on the accumulated, self-reinforcing knowledge encoded in the pheromone trails.

C. Deep Reinforcement Learning (DRL)

In response to the identified shortcomings of fixed meta-heuristic methods and the growing complexity of the optimization landscape, DRL presents a sophisticated and highly effective framework for complex decision-making. DRL is a powerful subset of machine learning that integrates the trial-and-error methodology of Reinforcement Learning (RL), where a computational agent learns an optimal policy through interaction with an environment to maximize cumulative reward—with the robust data processing capabilities of Deep Neural Networks.These networks enable the agent to effectively process and synthesize high-dimensional, complex state information, such as large graph structures and real-time sensor inputs, mapping these observations directly to an optimal action sequence.

DRL is particularly well-suited for solving the sequential decision-making and routing problems inherent in path planning. Furthermore, DRL is employed in a meta-learning capacity to enhance the adaptability of the optimization process itself. Instead of merely solving the primary path problem, a DRL agent is trained to learn the optimal control policies for the underlying search algorithm. This powerful capability allows the DRL agent to dynamically adjust internal ACO parameters, such as the pheromone update magnitude, based on the real-time feedback regarding environmental changes and the current state of the colony's search. This integration of DRL for both initial guidance and dynamic control is the key for creating a path planning framework that is both fast-converging and highly robust to environmental variability.

D. DRL-Assisted Pheromone Initialization

This phase addresses the fundamental drawback of traditional ACO: starting with a random or uniform search space. We utilize DRL to pre-train an agent that quickly learns optimal pathing strategies, which are then used to intelligently seed the ACO pheromone map. This represents a direct application of the DRL-initialization adapted for an ACO approach.

A Deep DQN is chosen for this task due to its effectiveness in solving sequence-selection problems TSP , which is analogous to our path-finding task. The objective of this training is to teach the agent the optimal action (next waypoint) to take from any given state to minimize the Multi-Objective Cost (C_{MOC}).

1) DRL Approach Setup:

- Agent: The DQN architecture typically consists of an input layer, several hidden layers, and an output layer equal to the size of the action space.
- Experience Replay Buffer: A mechanism is used to store and randomly sample past experiences $(\text{State}, \text{Action}, \text{Reward}, \text{Next State})$ to stabilize the learning process and break the correlation between consecutive samples.
- Target Network: A separate, delayed-update target network is used to compute the target Q-values, further enhancing stability during training.

E. State Space Definition (S)

The state S captures all necessary information for the agent to make an informed decision. For an agent at waypoint v_i, the state is defined as:

$$S = \{v_i, H_t, C_{t,objective}\} \tag{2}$$

where,

- Current Waypoint (v_i): The index of the UAV's current location.
- History Vector ($\mathbf{H}_t$): A binary vector representing all waypoints visited up to time t. This is crucial for preventing redundant visits and ensuring a complete path is generated.
- objectives Cost Vector ($\mathbf{C}_{t,\text{objective}}$): The aggregated normalized objectives costs (distance, time, battery).

F. Action Space Definition (A)

The action A is the agent's choice of the next waypoint v_j to move to from v_i:

$$A_t \in \{v_j \mid (v_i, v_j) \in E, and, v_j \notin Visitted\} \tag{3}$$

The action space is dynamic: the set of available actions changes as the path is constructed, always excluding waypoints already visited.

G. Reward Function (R)

The reward function is the critical link between the agent's actions and the optimization goal. Since the goal is to minimize the C_{MOC}, the reward is inversely proportional to this cost.

$$R(P) = \frac{R_{base}}{\epsilon + C_{MOC}(P)} \tag{4}$$

(where, ϵ is a small constant to prevent division by zero). A negative reward is given for illegal moves.

H. Initial Path Generation

The trained DQN is used to greedily generate N_{init} high-quality, initial paths: $P_{init} = \{P_1, P_2, \ldots, P_{N_{init}}\}$.

1) Pheromone map initialization: These paths are used to initialize the ACO pheromone matrix τ_{ij}. Instead of starting with uniform pheromones, the initial pheromone level $\tau_{ij}(0)$ on each edge (i, j) is set proportional to the inverse cost of the DRL-generated paths that use that edge:

$$\tau_{ij}(0) = \tau_{min} + \sum_{k=1}^{N_{init}} \mathbb{I}_{ij}(P_k) \cdot \frac{Q_{DRL}}{\sum_{m=1}^{N_{init}} C_{MOC}(P_m)} \quad (5)$$

where,

τ_{min} is the minimal pheromone level (evaporation lower bound).

$\mathbb{I}_{ij}(P_k)$ is an indicator function (1 if path P_k uses edge (i, j), 0 otherwise).

Q_{DRL} is a fixed constant representing the total initial pheromone mass.

2) Ant movement (path construction): Ants construct paths stochastically based on pheromone levels and heuristic information, as in traditional ACO.

$$P_{ij}^k = \frac{\tau_{ij}^{\alpha}.\eta_{ij}^{\beta}}{\sum_{l \in \text{Allowed}_i} \tau_{il}^{\alpha} \cdot \eta_{il}^{\beta}} \quad (6)$$

where,

τ_{ij} is the pheromone level on edge (i, j).

η_{ij} is the heuristic value (inverse of distance).

α and β are control parameters (importance of pheromone vs. heuristic).

3) DRL-Learned Adaptive Pheromone Update (APC): This is the core novelty, replacing fixed update rules with a learned policy. A separate, smaller DRL agent (a simple Q-Learning) is used to determine the optimal pheromone update magnitude.

The pheromone update rule becomes:

$$\tau_{ij}(t + 1) = (1 - \rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij} \quad (7)$$

The new pheromone deposit $\Delta\tau_{ij}$ is determined by the DRL-APC Agent:

- APC Agent State (S_{APC}): A vector including the path segment (i, j) , the current global path quality, and the local pheromone density τ_{ij}.

$$S_{APC} = \tau_{ij} \quad (8)$$

- APC Agent Action (A_{APC}): The action is the determination of the pheromone scaling factor, $\Delta\tau_{ij}^{\text{scale}} \in [0, 1]$.

$$A_{APC} = \Delta\tau_{ij}^{\text{scale}} \in [0, 1] \quad (9)$$

- APC Agent Reward (R_{APC}): A reward function that encourages convergence and diversity: a high reward for finding a new global best solution, and a penalty for stagnation or over-prioritizing short-term, low-quality paths local optima.

The resulting pheromone update is:

$$\tau_{ij}(t + 1) = (1 - \rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}^{\text{base}} \cdot A_{APC} \quad (10)$$

where,

$\Delta\tau_{ij}^{\text{base}}$ is the standard deposit for the best path, and A_{APC} is the scaling action learned by the DRL agent.

I. Path Output

The path calculation in this approach is a two-stage process: first, the probabilistic selection of discrete waypoints based on the environment's pheromone map, and second, the physical translation of those waypoints into a navigable 3D trajectory for the UAV.

- Probabilistic Movement Rule (Node Selection). The UAV (represented as an "ant" in the ACO framework) starts at a designated launch coordinate S. At each time step t, the agent must select the next node j from a set of reachable neighboring nodes N_i. This selection is governed by the transition probability P_{ij}, which balances exploration and exploitation.
- Multi-Objective Cost Constraints such as standard pathfinding, the UAV path calculation incorporates a cost function C derived via the Analytic Hierarchy Process (AHP). Every potential move is evaluated against:
  - Distance (C_D): Euclidean distance between nodes.
  - Time (C_T): Calculated based on the UAV's average velocity across the terrain.
  - Battery (C_B): Energy consumption model considering climb rate and air resistance.
- Path Refinement and DRL Influence. The "Deep Reinforcement Learning" component intervenes during the calculation by adjusting the pheromone scaling factor.

The final output is the path P_{best} corresponding to the lowest C_{MOC} found across all iterations.

J. Analytical Hierarchy Process

The AHP is one of the most widely applied techniques in multi-criteria decision analysis (MCDA), particularly within natural resource management and environmental studies. In this study, AHP is employed to determine the weights of the resource layers [26]. The method operates by assigning priority values based on relative importance, typically using a 1–9 scale [27]. Each criterion is evaluated individually through pairwise comparisons, after which the eigenvalues of the resulting comparison matrix are computed and utilized as the corresponding criterion weights.

This study follows the same methodological framework established in the referenced work by [16].

Table I Presents the objective weights assigned based on their relative importance in this approach.

TABLE I. THE OPTIMAL PATH

Objectives	D	T	B	Weight Score
Distance (D)	1	2	2	0.5
Time (T)	0.5	1	1	0.25
UAV battery (B)	0.5	0.5	1	0.25

During the training phase, the DRL agent learns optimal policies aimed at minimizing the AHP-weighted cost function, thereby generating paths that effectively balance the prioritized objectives. Consequently, the AHP framework not only provides objective weighting for the ACO fitness evaluation but also actively guides the DRL exploration process, ensuring that the initially generated paths capture mission-critical trade-offs prior to the ACO-based refinement stage.

III. STUDY AREA AND EXPERIMENTAL ANALYSIS

A. Study Area

The study area is located in the Red Sea Mountains in eastern Sudan, a striking geological formation that runs parallel to the Red Sea coast. Known for its rugged terrain and diverse ecosystems, the region offers a unique landscape shaped by the Arabian-Nubian Shield. The mountains are predominantly made up of granite and volcanic rock formations, with elevations ranging from 600 to over 2,500 meters above sea level. The ecological variety in this area includes dry forests, scrublands, and semi-desert habitats, providing a rich environment for a wide array of flora and fauna adapted to harsh conditions. Given the challenging terrain of the area, traditional methods of transportation between different points are inefficient and impractical. To overcome this issue, the present study employs a drone for logistical transport between the selected points. The specific locations of these points are provided in the accompanying Table II, while the layout of the terrain is depicted in Fig. 2.

B. Results

This study employs a multi-objective optimization approach to develop and determine the optimal UAV flight paths by integrating three key objectives: distance, time, and battery consumption. The process begins with the application of DRL to estimate a set of promising initial parameters for the optimization. These parameters are then refined through the ACO algorithm.

- ACO Parameters (α and β). Pheromone Influence ($\alpha = 1.0$). This value was chosen to ensure that the "swarm intelligence" (historical data) has a significant impact on path selection without causing premature convergence to suboptimal paths. Heuristic Influence ($\beta = 2.0$) we assigned a higher weight to the heuristic information (distance to goal). This is critical in UAV pathing to ensure that ants remain "goal-oriented".
- Evaporation Rate ($\rho = 0.1$). The evaporation rate was set to 0.1 to provide a "long memory" for the pheromone trails. In a complex 3D environment, rapid evaporation (> 0.5) can lead to the loss of good global solutions, while too low a rate (< 0.05) prevents the algorithm from discarding poor paths.
- Learning Rate ($\gamma = 0.001$). A conservative learning rate was selected for the Deep Q-Network (DQN) to ensure stable convergence of the pheromone scaling factor.
- Discount Factor ($\gamma = 0.95$). This high value ensures the DRL agent considers the long-term cost of the entire path (Total Battery/Time) rather than just the immediate next step.
- Exploration-Exploitation (ϵ-greedy). Utilized a decaying epsilon starting at 1.0 and ending at 0.01. This allows the DRL agent to explore a wide variety of pheromone update strategies in early iterations before settling on an optimized policy.

The effectiveness of the proposed DRL-ACO hybrid depends on the calibration of several key hyperparameters. These values were selected based on a combination of established literature and empirical tuning during the preliminary simulation phase.

In this context, the DRL agent functions as an intelligent control system, continuously tasked with identifying the optimal pheromone update strategy tailored to the specific mission objectives. This intelligent control mechanism is crucial for determining the optimal flight trajectory by balancing the mission's multi-objective criteria. To systematically prioritize and combine the three objectives (distance, time, and battery consumption), the AHP is employed. The AHP ensures a balanced optimization process by assigning relative importance to each objective based on mission-specific constraints and application needs.

Following the weight assignment and construction of the multi-objective decision matrix, an initial population of candidate solutions is generated through DRL. This approach enhances the performance of ACO by providing an optimal pheromone update strategy value, set at 200, which improves the search efficiency.

Finally, the optimal UAV path is determined by applying the ACO, which has been enhanced through integration with the multi-objective DRL framework. The DRL component guides the initialization and search process by providing a more informed starting population, allowing the ACO to converge more effectively toward solutions that balance the three objectives: minimizing travel distance, energy consumption, and mission time.

The results demonstrate the superiority of our proposed algorithm in determining the optimal logistic trajectory length for the UAV, which was found to be 750 meters. In contrast, alternative algorithms aimed for trajectory lengths of 800 and 850 meters. This outcome underscores the reliability of our approach, as evidenced in Table III and Fig. 4.

Fig. 4 shows the optimal path using our approach, and Fig. 3 shows the path using traditional ACO and GA.

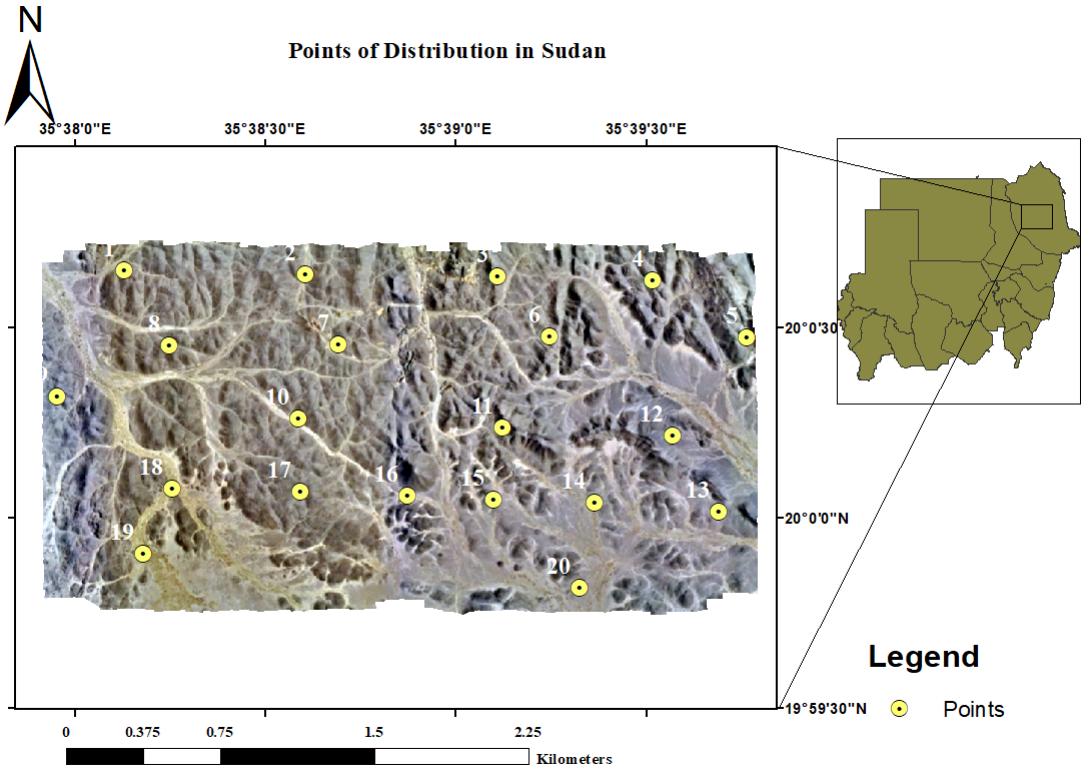

Fig. 2. Points of distribution in Sudan.

TABLE II. POINTS SELECTED FOR OPTIMAL PATH IMPLEMENTATION USING THE UAV

NO	X	Y	H	NOTE
1	777724	2213582	567.89	P1
2	778531	2213530	560.30	P2
3	777866	2214318	570.12	P3
4	778447	2214406	568.81	P4
5	775792	2213514	559.35	P5
6	775778	2214369	555.72	P6
7	776970	2214391	573.14	P7
8	778181	2215498	573.32	P8
9	777720	2216367	562.56	P9
10	776577	2216587	556.08	P10
11	776284	2216169	561.366	P11
12	776111	2214297	557.59	P12
13	776127	2213926	567.58	P13
14	776477	2214631	572.83	P14
15	776570	2215327	565.57	P15
16	775662	2216103	559.32	P16
17	776869	2217447	556.15	P17
18	776484	2218677	553.27	P18
19	776418	2219481	537.95	P19
20	775577	2220404	543.84	P20

C. Discussion

The integration of DRL with ACO significantly enhances UAV path planning by improving convergence speed and optimizing multi-objective criteria, such as distance, time, and energy consumption. The DRL-ACO framework not only overcomes the limitations of traditional ACO, which relies on static pheromone update rules, but also provides adaptability through dynamic pheromone management, leading to faster convergence towards optimal solutions [28], [18]. In the experimental results, the DRL-ACO approach demonstrated superior performance, finding shorter paths 750 meters compared to traditional ACO and GA 850 meters, thus ensuring greater efficiency in UAV operations. The use of the AHP further allowed the DRL agent to balance the conflicting objectives by assigning appropriate weights, ensuring that the solution met mission-specific requirements effectively [23].

This approach proved to be particularly robust in a complex environment, as demonstrated in the study area in Sudan, where the challenging terrain required adaptability in path planning. Unlike traditional approaches, the DRL-ACO approach could handle the unpredictability of real-world conditions, making it highly suitable for mission-critical applica-

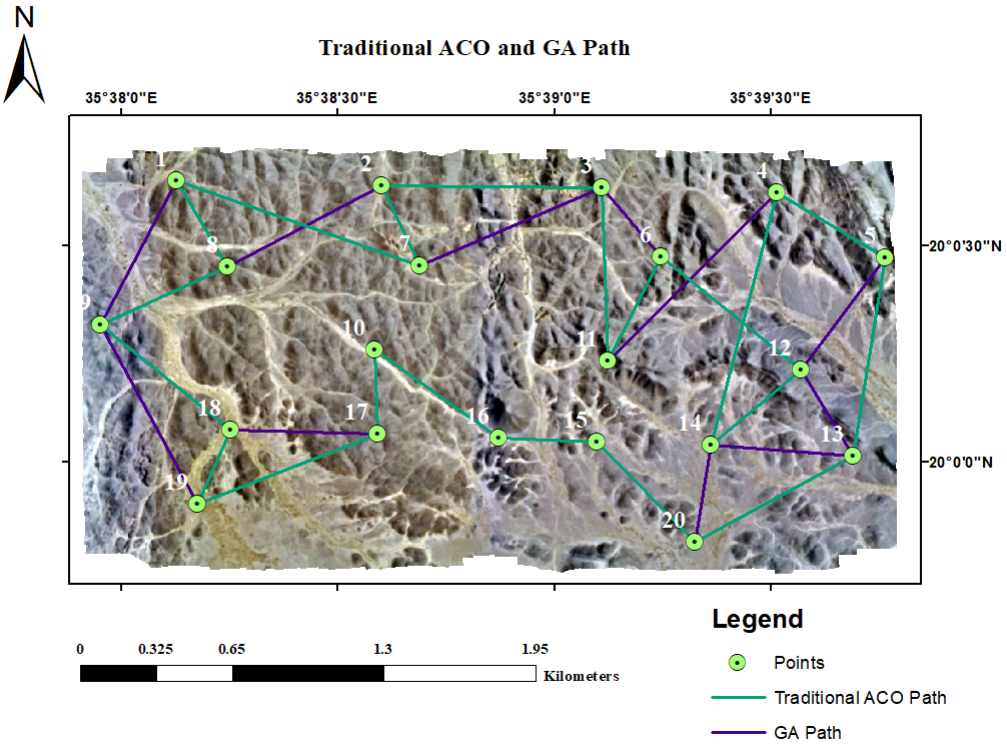

Fig. 3. Traditional ACO and GA path.

TABLE III. OPTIMAL LOGISTIC TRAJECTORY LENGTH

NO	Name	Points	Length
1	DRL-ACO Path	1→8→ 7→2→11→3→6→14→4→5→12→20→13→15→16→10→17→19→9→18→1	750
2	ACO Path	9→8→ 1→7→2→5→11→6→12→14→4→5→13→20→15→16→10→17→19→18→9	850
3	GA Path	17→18→ 19→9→1→8→2→7→3→6→11→4→5→12→13→14→20→15→16→10→17	850

TABLE IV. COMPARISON BETWEEN DRL-ACO AND TRADITIONAL ACO AND GA

NO	Length(m)	Time Implementation(s)
DRL-ACO	750	0.27
Traditional ACO	800	0.35
GA	850	0.55

tions, such as disaster response and surveillance. While the framework showed significant improvements, there are still opportunities for future enhancement, particularly in incorporating additional factors like obstacle avoidance or real-time re-planning, and improving computational efficiency. Despite its success, the DRL-ACO approach requires substantial computational resources for training, suggesting that further optimization of the DRL model could improve its applicability in real-time scenarios.

In comparison to other path planning algorithms, the DRL-ACO approach demonstrated clear advantages in reducing path length and optimizing mission objectives. This outcome not only highlights the potential of the framework for real-world applications but also sets a foundation for future advancements, where the system can adapt to even more complex and dynamic environments, improving its effectiveness in various UAV mission scenarios.

The proposed algorithm was tested against two established methods: the traditional ACO and the GA algorithm. The results of this comparison are presented in Table IV, assessed performance based on three key metrics: path length 750 m, and execution time 0.27 s. The results demonstrate that the proposed algorithm consistently outperforms both the ACO and GA, achieving more optimal path lengths in fewer iterations and with significantly lower computational overhead.

IV. CONCLUSION

This study successfully introduced a novel DRL-Assisted Ant Colony Optimization (DRL-ACO) approach to address the limitations of traditional ACO in multi-objective UAV path planning. The critical innovation involves the deep integration of a DRL agent to serve as an intelligent control system, replacing static rules with an adaptive pheromone update (APC) mechanism. The problem was formalized using the AHP to

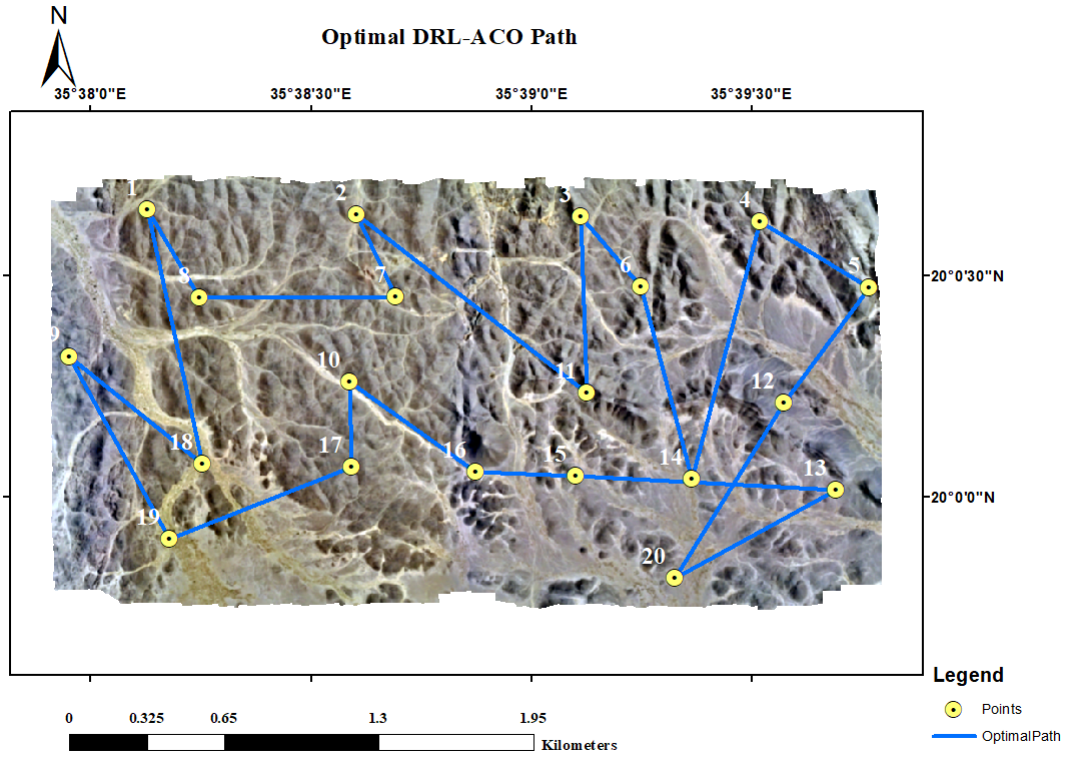

Fig. 4. Optimal DRL-ACO path.

systematically weigh and combine the conflicting objectives of minimizing flight distance, energy consumption, and mission time into a single multi-objective cost function (C_{MOC}). Empirical validation, conducted in a challenging, rugged terrain environment in Sudan, confirmed the framework's superiority. The DRL-ACO approach found the optimal trajectory length of 750 meters, significantly outperforming the 800-850 meters found by traditional ACO and GA, while also demonstrating a lower computational time of 0.27 seconds. This outcome confirms that the DRL-ACO framework provides a robust and adaptive solution that achieves faster convergence and superior optimization of multi-objective criteria for autonomous UAV operations.

Despite the demonstrated success, the DRL-ACO framework presents several opportunities for future enhancement to increase its real-time applicability. Firstly, future research should focus on incorporating additional real-time constraints, such as obstacle avoidance and dynamic path re-planning, to enhance the system's robustness in highly unpredictable environments. Secondly, the current approach requires substantial computational resources for DRL model training. Finally, researchers could explore more advanced DRL algorithms or hierarchical DRL structures to manage local control mechanisms, potentially yielding even more sophisticated, non-linear control policies for the adaptive ACO parameters.

FUNDING

This study was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia, under grant no (GPIP:882-830-2024). The authors, therefore, gratefully acknowledge and thank DSR for its technical and financial support.

AUTHORS' CONTRIBUTIONS

Mohamed A. Damos: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. Wenbo Xu: Methodology (Genetic Algorithm design), Software, Validation, Writing – review editing. Ali Ahmed: Methodology (AHP integration), Validation, Writing – review. Funding acquisition, Resources, Project administration .Abdolraheem Khader: Software (simulation environment), Validation, Formal analy- sis. Data curation, Visualization, Writing – review editing.

AVAILABILITY OF DATA AND MATERIALS

The data and materials supporting the findings of this study are available upon request from the corresponding author. Simulation code, algorithm parameters, and waypoint coordinates used for UAV path planning are fully shareable.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Deanship of Scientific Research at King Abdulaziz University, Jeddah, for its technical and financial support under Grant No. (GPIP:882-830-2024).

REFERENCES

[1] H. Hildmann and E. Kovacs, "Using unmanned aerial vehicles (uavs) as mobile sensing platforms (msps) for disaster response, civil security and public safety," *Drones*, vol. 3, no. 3, p. 59, 2019.

[2] M. Damos, J. Zhu, W. Li, A. Hassan, and E. Khalifa, "A novel urban tourism path planning approach based on a multiobjective genetic algorithm. isprs int., geo-inf 2021, 10.530 https," *doi. org/https://doi. org/10.3390/ijgi10080530*, 2021.

[3] X. Zhang, X. Yu, and X. Chen, "D-aco-based path planning for auv in uwsn: A hybrid approach combining dqn and aco," *IEEE Sensors Journal*, 2025.

[4] R. Priyadarshi and R. R. Kumar, "Evolution of swarm intelligence: a systematic review of particle swarm and ant colony optimization approaches in modern research," *Archives of Computational Methods in Engineering*, pp. 1–42, 2025.

[5] M. A. Damos, W. Xu, J. Zhu, A. Ahmed, and A. Khader, "An efficient tourism path approach based on improved ant colony optimization in hilly areas," *ISPRS International Journal of Geo-Information*, vol. 14, no. 1, p. 34, 2025.

[6] C. Chen, L. Cao, Y. Chen, B. Chen, and Y. Yue, "A comprehensive survey of convergence analysis of beetle antennae search algorithm and its applications," *Artificial Intelligence Review*, vol. 57, no. 6, p. 141, 2024.

[7] R. S. Othman and I. M. Ibrahim, "A review of exploring recent advances in ant colony optimiza-tion: applications and improvements."

[8] P. Li, J. Hao, H. Tang, X. Fu, Y. Zhen, and K. Tang, "Bridging evolutionary algorithms and reinforcement learning: A comprehensive survey on hybrid algorithms," *IEEE Transactions on evolutionary computation*, 2024.

[9] S. Nimmala, M. Ramchander, M. Mahendar, P. Manasa, D. D. Bhavani, and K. Raghavendar, "Dynamic rl-aco: Reinforcement learning-based ant colony optimization for load balancing in cloud networks," in *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE, 2024, pp. 475–480.

[10] B. Jin, Y. Sun, W. Wu, Q. Gao, and P. Si, "Deep reinforcement learning and ant colony optimization supporting multi-ugv path planning and task assignment in 3d environments," *IET Intelligent Transport Systems*, vol. 18, no. 9, pp. 1652–1664, 2024.

[11] Y. Caihong, "Dynamic logistics path optimization via integrated ant colony optimization and reinforcement learning," *Informatica*, vol. 49, no. 6, 2025.

[12] A. Ahmed, H. Ju, Y. Yang, and H. Xu, "An improved unit quaternion for attitude alignment and inverse kinematic solution of the robot arm wrist," *Machines*, vol. 11, no. 7, p. 669, 2023.

[13] Y. Wang, J. Liu, Y. Qian, and W. Yi, "Path planning for multi-uav in a complex environment based on reinforcement-learning-driven continuous ant colony optimization," *Drones*, vol. 9, no. 9, p. 638, 2025.

[14] Z. Guo, Y. Xia, J. Liu, J. Gao, P. Wan, and K. Xu, "Path planning design and experiment for a recirculating aquaculture agv based on hybrid nrbo-aco with dueling dqn," *Drones*, vol. 9, no. 7, p. 476, 2025.

[15] M. M. Hamdi, B. S. Abdulhakeem, and A. A. Nafea, "Psoa-crl: A hybrid multi-objective routing mechanism using particle swarm optimization and actor-critic reinforcement learning for vanets," *Mesopotamian Journal of Big Data*, vol. 2025, pp. 241–260, 2025.

[16] M. A. Damos, J. Zhu, W. Li, E. Khalifa, A. Hassan, R. Elhabob, A. Hm, and E. Ei, "Enhancing the k-means algorithm through a genetic algorithm based on survey and social media tourism objectives for tourism path recommendations," *ISPRS International Journal of Geo-Information*, vol. 13, no. 2, p. 40, 2024.

[17] X. Zhou, H. Ma, J. Gu, H. Chen, and W. Deng, "Parameter adaptation-based ant colony optimization with dynamic hybrid mechanism," *Engineering Applications of Artificial Intelligence*, vol. 114, p. 105139, 2022.

[18] B. Wang, D.-T. Duan, Q. Yang, X.-Y. Zhao, T. Li, D. Liu, and J. Zhang, "Adaptive ant selection for pheromone update in ant colony optimization," in *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2024, pp. 667–672.

[19] A. Gharbi, "A dynamic reward-enhanced q-learning approach for efficient path planning and obstacle avoidance in mobile robotics," *Applied Computing and Informatics*, 2024.

[20] S. Goyal, L. K. Awasthi, V. Garg, and G. Kumar, "Multi-resource aware virtual machine consolidation approach for modern cloud data centers," *Computing*, vol. 107, no. 11, pp. 1–32, 2025.

[21] A. Sezgin and A. Boyacı, "Optimizing drone deployment for network coverage using a hybrid ant colony optimization and reinforcement learning approach: Karınca kolonisi optimizasyonu ve pekiştirmeli öğrenme yaklaşımı kullanarak ağ kapsaması için drone konuşlandırmasının optimizasyonu," *Journal of Aeronautics and Space Technologies*, vol. 18, no. 2, pp. 268–285, 2025.

[22] J. Zhang, H. Xu, D. Liu, and Q. Yu, "Advancing dynamic emergency route optimization with a composite network deep reinforcement learning model." *Systems*, vol. 13, no. 2, 2025.

[23] S. Fang, Z. Deng, P. Li, and D. Long, "Improved strategy of ant colony optimization for path planning via stochastic pheromone updating and cyclic initialization," *Journal of Mechanical Science and Technology*, pp. 1–12, 2025.

[24] G. Vachtsevanos, L. Tang, G. Drozeski, and L. Gutierrez, "From mission planning to flight control of unmanned aerial vehicles: Strategies and implementation tools," *Annual Reviews in Control*, vol. 29, no. 1, pp. 101–115, 2005.

[25] X. Li, Z. Ruan, Y. Ou, D. Ban, Y. Sun, T. Qin, and Y. Cai, "Adaptive deep ant colony optimization–asymmetric strategy network twin delayed deep deterministic policy gradient algorithm: Path planning for mobile robots in dynamic environments," *Electronics*, vol. 13, no. 20, p. 4071, 2024.

[26] R. Kumar and R. Anbalagan, "Landslide susceptibility mapping using analytical hierarchy process (ahp) in tehri reservoir rim region, uttarakhand," *Journal of the Geological Society of India*, vol. 87, pp. 271–286, 2016.

[27] Y. Wind and T. L. Saaty, "Marketing applications of the analytic hierarchy process," *Management science*, vol. 26, no. 7, pp. 641–658, 1980.

[28] B. Abdulghani and M. Abdulghani, "A comprehensive review of ant colony optimization in swarm intelligence for complex problem solving," *Acadlore Transactions on Machine Learning*, vol. 3, no. 4, pp. 214–224, 2024.