Received: 5 August 2022 | Revised: 23 November 2022 | Accepted: 12 December 2022
DOI: 10.1111/exsy.13224

ORIGINAL ARTICLE

Expert Systems WILEY

Synergetic fusion of Reinforcement Learning, Grey Wolf, and Archimedes optimization algorithms for efficient health emergency response via unmanned aerial vehicle

Himanshu Gupta [1] | K. Sreelakshmy [1] | Om Prakash Verma [1] | Tarun Kumar Sharma [2] | Chang Wook Ahn [3] | Kapil Kumar Goyal [4]

[1]Department of Instrumentation and Control Engineering, Dr. B R Ambedkar National Institute of Technology Jalandhar, Punjab, India

[2]Department of Computer Science and Engineering, Shobhit Institute of Engineering and Technology (Deemed-to-be University), Meerut, Uttar Pradesh, India

[3]AI Graduate School, Gwangju Institute of Science and Technology, Gwangju, South Korea

[4]Department of Industrial and Production Engineering, Dr. B R Ambedkar National Institute of Technology Jalandhar, Punjab, India

Correspondence
Om Prakash Verma, Department of Instrumentation and Control Engineering, Dr. B R Ambedkar National Institute of Technology Jalandhar, Punjab, India.
Email: vermaop@nitj.ac.in

Funding information
Jagadish Chandra Bose Research Organisation

Abstract

Owing to the recent technological innovations, unmanned aerial vehicles (UAVs) are progressively employed in various civil and military applications, including healthcare. This requires estimating an optimum route under various real-world complexities, such as non-uniform obstacles. However, most of the reported work considers only uniform obstacles as an object, which limit their practical applicability. Hence, Archimedes optimization algorithm (AOA) is examined to overcome this limitation. Further, it is observed that many a time, AOA over-exploits the search space, resulting in higher computational time. Therefore, the present work fuses AOA with grey wolf optimizer (GWO) to improve the convergence capability. Also, reinforcement learning (RL) is employed to intelligently switch between the exploration and exploitation phases. The efficacy of the developed algorithm is statistically analysed and validated against various metaheuristics on several benchmark functions. The simulated results verified that the developed RLGA provides optimal or near-optimal solutions more efficiently relative to other metaheuristics. Moreover, it also affirms the hypothesis that the proposed modifications significantly improve the convergence speed of AOA. Finally, the appropriateness of RLGA is tested and validated by rigorous experimentation on real-world 3D-route estimation problems for UAVs. The simulated results reveal that RLGA produces a flyable path with 51.46%, 62.06%, and 70.42% lesser cost than RLGWO, AOA, and GWO, respectively. This ensures the employability of RLGA for efficient medical assistance in minimum time-, energy-, and transportation-cost with safe and smooth UAV auto-navigation for developing drone doctors.

KEYWORDS
auto-navigation, healthcare, metaheuristic, path planning, unmanned aerial vehicles

1 | INTRODUCTION

Unmanned Aerial Vehicles (UAVs), popularly known as drones, are receiving significant recognition because of their manoeuvrability, cost-effectiveness, and auto-navigation capability (Gigante et al., 2018; Gomaa et al., 2020). Therefore, UAVs touch almost all aspects of commercial and scientific applications, such as disaster prediction and management (Yu et al., 2020), geomatics (Giordan et al., 2020), construction (Dupont et al., 2017; Tatum & Liu, 2017), security and surveillance (Ucgun et al., 2021), search and rescue operations (Sierra-Garcia & Santos, 2022), transport and delivery (Gupta & Verma, 2021), and so forth. In addition, the speed and versatility of UAVs open unlimited opportunities to the healthcare sector as well by delivering life-saving medications and life-saving medications and enhancing patient-hospital connectivity, especially to unreached areas (Damoah et al., 2021). The potential of UAVs is already witnessed during COVID-19 for the rapid last-mile delivery of hazardous and environmentally susceptible medical products, such as vaccines, blood samples, and medical kits hazardous and environmentally susceptible medical products, such as vaccines, blood samples, and medical kits (Subhan et al., 2019; Sun et al., 2021). Such developments integrated with internet-of-things (IoT) and internet-of-vehicles (IoV) may prove to be a boon for future smart healthcare systems (Ullah et al., 2019). This smart system may be employed to prevent life-critical incidents, such as heart failure and chronic attacks. These advancements reduce the burden on the existing healthcare systems, which offer time- and cost-efficient follow-ups by medical practitioners. Further, the effectiveness of these coordinated health solutions depends upon the auto-navigation capability, payload capacity, and flying range of UAVs. The current study investigates the auto-navigation capability with minimum travel time and energy consumption in a real-world environment with non-uniform obstacles. The auto-navigation ability of UAVs, particularly in a 3D environment, is considered an NP-hard optimization problem. It refers to the problem of estimating the optimum path from source to destination, where length of the traversed path, fuel consumption, altitude variations, and collision avoidance are the key constraints. Also, the variable shape and size of threats, such as mountains, buildings, trees, etc., increases the complexity associated with the trajectory estimation of UAVs which leads to the inability in accomplishing any task autonomously. Therefore, route estimation approaches are becoming indispensable for generating a predefined flyable path to achieve feasible and efficient solutions.

The reported literature reveals many computationally intelligent algorithms (CIAs) for estimating the flyable path of UAVs, such as graph-based (A star (Primatesta et al., 2018), Rapidly-Exploring Random Trees (RRT) (Shan et al., 2009), etc.), potential field-based (Artificial Potential Field (APF) (Chen et al., 2016)), and nature-physics inspired metaheuristics (Particle Swarm Optimization (PSO) (Kennedy & Russell, 1995), ant colony optimization (ACO) (Dorigo & Di Caro, 1999), grey wolf optimization (GWO) (Mirjalili et al., 2014), gravitational search algorithm (GSA) (Rashedi et al., 2009), Archimedes optimisation algorithm (AOA) (Sreelakshmy et al., 2022), etc.). However, the inefficiency of obtaining a reliable path in complex environmental conditions limits the widespread acceptance of earlier graph and potential field-based CIAs. On the contrary, metaheuristics efficiently obtain feasible solutions for many real-world problems (Gupta et al., 2021). Therefore, the path optimization problem is exhaustively investigated by employing metaheuristic algorithms, such as PSO (Kennedy & Russell, 1995), ACO (Dorigo & Di Caro, 1999), whale optimization algorithm (WOA) (Mirjalili & Lewis, 2016), GWO (Mirjalili et al., 2014), GSA (Rashedi et al., 2009), and water cycle algorithm (WCA) (Eskandar et al., 2012). Recently, reinforcement learning (RL) based GWO (RLGWO) algorithm was proposed to generate a reliable 3D trajectory of UAVs (Qu, Gai, Zhong, & Zhang, 2020). Subsequently, an improved bat lagorithm (IBA) offered a feasible path in a complex battlefield environment (Zhou et al., 2021). Meanwhile, improved PSO with multiple processing steps was utilized for path planning of swarm UAVs under dynamic threats, such as surface-to-air missiles (Shin & Bang, 2020). Furthermore, Spherical vector-based PSO (SPSO) incorporates spherical vectors to enhance the path quality (Phung & Ha, 2021). In addition, the solution quality of PSO was improved by employing the Metropolis criterion and RTS smoother (Wu et al., 2018). Similarly, PSO with global best path competition (GBPSO) introduces the competition strategy among the particles during evolution to achieve better performance (Huang & Fei, 2018). Later, GWO and symbiotic organisms search (SOS) were fused to ensure a higher convergence rate and exploration capability (Qu, Gai, Zhang, & Zhong, 2020). Similarly, a hybrid PSO-SOS was utilized for the cooperative route planning of swarm UAVs (He et al., 2021). Additionally, the employability of metaheuristics was also examined for dynamic route planning of UAVs (Liu et al., 2019). GWO was employed to acquire the feasible path for unmanned combat aerial vehicles in military applications (Zhang et al., 2016). Further, the waypoint-based evaluation function approach was examined to improve the quality of the generated path (Yang et al., 2015). Similarly, UAVs were employed to deliver medical essentials by avoiding traffic congestion using metaheuristics (PSO, ACO, and genetic algorithm [GA]) (Khan et al., 2021). In addition, physics-inspired optimizers such as equilibrium optimizer (EO) (Tang et al., 2021), quantum-entanglement pigeon-inspired optimization (QEPO) (Li & Deng, 2019), multiverse optimizer (MVO) (Jain et al., 2019), and GSA (Xu et al., 2021) were successfully employed for optimal flight path generation. Recently, other physics-based algorithms, such as AOA (Hashim et al., 2021), flow direction algorithm (FDA) (Karami et al., 2020), and artificial electric field algorithm (AEFA) (Yadav, 2019) were also introduced and examined on benchmark functions. However, their potential in optimal path planning needs to be explored.

The earlier path-planning algorithms are time-consuming and slow because they employ graph-based methods, limiting their implementation to real-time problems (Zhao et al., 2018). Compared with graph-based methods, metaheuristics are observed to perform faster. However, they often get traped into local minima and produce slow convergence rates. Also, the No Free Lunch (NFL) theorem opens the opportunity to investigate the possibilities of new metaheuristics. According to NFL, an optimization algorithm that results in global minima and has good convergence rates on one class of optimization problem does not guarantee similar results on other optimization problems (Wolpert & Macready, 1997). This motivates to explore other metaheuristics for real-world problems, such as path planning in a 3D environment. Also, previously reported literature reveals the estimation of a feasible path with similar obstacles, which is impractical and infeasible in a real-world scenario.

The present work is an attempt to develop a hybrid algorithm for planning a safe and flyable path in the presence of various obstacles of different shapes and sizes in a 3D environment. The RL-based GWO-AOA (RLGA) is developed by improving the exploration stage of AOA using GWO. Furthermore, RL is proven effective in regulating the exploration-exploitation balance in metaheuristics (Seyyedabbasi et al., 2021).

Therefore, RL is employed to select the exploration-exploitation stages intelligently. The similarities and differences between GWO and AOA, based on which the hybrid algorithm is developed along with the advantages of incorporating RL, are enlisted here:

1. Both are population-based algorithms.
2. AOA is physics-based, whereas GWO is a swarm intelligence (SI)-based optimization algorithm.
3. GWO depends on the best three individuals for updating their position during all the iterations. In contrast, AOA updates the position based on a randomly selected individual during one-third of the iterations and, afterward, updated based on the best individual.
4. In GWO, exploration and exploitation balance is maintained by parameter a_t which decreases linearly from 2 to 0. Depending on a_t, the parameter A varies in the range $[-2a_t, 2a_t]$. $|A| < 1$ indicates exploitation and $|A| > 1$ represents exploration. However, AOA performs global search in 30% of the iterations and local search in rest of the iterations, which is maintained by transfer operator (TO).
5. The convergence speed of AOA is less since it is in the global search for one-third of the iterations. During this phase, the acceleration and position are updated based on random selection of the individuals. On the contrary, the convergence rate of GWO is better since it depends only on the leader group for updating the positions. However, the dependency on the best three wolves limits diversity of the population and might lead to local minima solutions. Further, the population update based on the best three wolves results in good exploitation capacity.
6. To enhance the local search abilities of AOA, GWO is incorporated during the exploration stage of AOA. Further, RL (Q learning) agent is employed to balance the exploitation-exploration stage. This intelligent selection between global and local search may improve the convergence rate of AOA.
7. Further, Q learning agent selects the action (exploration or exploitation) based on the Q table values. A positive or negative reward is assigned depending on the obtained cost function values during each iteration. The agent tries to maximize the reward signal by selecting appropriate action based on the Q table values.

In a nutshell, GWO is employed to enhance the exploration phase of AOA. The slow convergence rate of AOA results from the 30:70 global–local search ratio. RL is exhaustively investigated in the proposed algorithm to adaptively select the exploration-exploitation stages, resulting in higher convergence rates. The exhaustive evaluation of RLGA is initially carried out on four sets of benchmark functions. The first and second set contains unconstrained and constrained single objective real parameter optimization problems from CEC 2017, the third set includes single objective real parameter optimization problems from CEC 2019, whereas the last set consists of 22 selected benchmark functions. Subsequently, the effectiveness of RLGA is investigated for 3D path estimation of UAVs. The major contributions of the present work are listed below:

1. Analysis of main bottlenecks for slow convergence in AOA.
2. Development of RLGA with improvement in the exploration phase of AOA using GWO.
3. Enhancement of the convergence rates of RLGA by utilizing RL to control the exploration-exploitation balance.
4. Exhaustive performance comparison of RLGA with other metaheuristics based on benchmark functions.
5. Employment and validation of RLGA for 3D path estimation of UAVs to endorse its successful applicability in real-world healthcare tasks.

Thus, the present study aims to develop and validate a hybrid algorithm based on metaheuristics and RL to provide an efficient and obstacle-free path for UAVs. Section 2 briefly describes the RL, GWO, and AOA. The development process of RLGA is presented in section 3. Section 4 comprises the benchmark function evaluation. The route estimation in the 3D environment is discussed in section 5, along with the problem formulation, cost function, and simulated results. Finally, concluding remarks and the scope of future research are presented in the last section.

2 | BASIC KNOWLEDGE

2.1 | Reinforcement learning

Reinforcement learning (RL) is a branch of Machine Learning (ML) and recently gained significant interest in the research fraternity. In RL, an agent automatically unleashes an optimal policy to maximize the reward by interacting with the environment. The RL can be categorized into policy-based and value-based approaches. Popularly, policy gradients indicate the policy-based RL in which the policy function directly maps a state into action. In contrast, the RL approach, popularly known as Q learning, represents the value-based RL in which the agent performs an action based on the highest values in the Q-table to learn the optimal policy during training. It is an off-policy Temporal Difference (TD) learning algorithm where Q represents the action-value function. The update policy of the Q table is mathematically formulated as in Equation (1),

$$Q_{it+1}(s_{it}, a_{it}) = Q(s_{it}, a_{it}) + \mu\left(R_{it+1} + \gamma \max_{a} Q(s_{it+1}, a_{it}) - Q(s_{it}, a_{it})\right), \quad (1)$$

where, s and a denote states and actions, respectively. μ depicts the learning rate, R is the reward, γ is the discount factor, and it represents the current iteration. Pseudocode 1 illustrates the informal implementation of the Q learning algorithm.

2.2 | Grey wolf optimization

Grey wolf optimization (GWO) is one of the most popular community-hunting behaviour-based SI algorithms. Grey wolf groups follow a strict social hierarchy that can be categorized as the α, β, δ, and ω wolves. α (the best), β(the second best), and δ (the third best) wolves represent the leader groups with the best information on the prey position (optimal solution). Therefore, all the individuals update their position according to the position of leader wolves during position updates. Equations (2)–(9) represents the position (Z_{oi}) update equations for individual wolves during each iteration (it).

$$Z_{oi}(it+1)=(Z_1+Z_2+Z_3)/3, \tag{2}$$

$$Z_1=Z_\alpha(it)-A_1D_\alpha \tag{3}$$

$$Z_2=Z_\beta(it)-A_2D_\beta \tag{4}$$

$$Z_3=Z_\delta(it)-A_3D_\delta, \tag{5}$$

$$D_\alpha=|C_1Z_\alpha(it)-Z_{oi}(it)| \tag{6}$$

$$D_\beta=|\,C_2Z_\beta(it)-Z_{oi}(it)\,| \tag{7}$$

$$D_\delta=|\,C_3Z_\delta(it)-Z_{oi}(it)\,| \tag{8}$$

$$\begin{gathered} C_k=2r_{1k} \\ A_k=2a_t.r_{2k}-a_t \end{gathered} \tag{9}$$

where, a_t decreases linearly during each iteration $\forall a_t\in[0,2]$, r_{1k} and r_{2k} represent random numbers used to calculate C_k and A_k, respectively, $k=1,2,3$. The updated position is represented by the mean of Z_1, Z_2, and Z_3 which depends on leader group positions, Z_α, Z_β, and Z_δ, respectively.

2.3 | Archimedes optimization algorithm

The physics-inspired optimization algorithm, Archimedes optimization algorithm (AOA) is emanated from Archimedes' Principle. AOA assumes that the initial population is constituted based on different objects with diverse densities and volumes immersed in a particular fluid. The objective is to achieve neutral buoyancy for all the objects.

If liq and obj represent the liquid and object floating in the liquid, $F_{(.)}$ indicates the force exerted, then at equilibrium Equation (10) is satisfied, which can be re-written as Equation (11),

Pseudocode 1 The Q learning

Initialize $Q(s, a)$ arbitrarily
while the terminal condition is not satisfied
  Choose an initial state
  Choose an action for the initialized state
  Take action, return the reward, and next state
  Update Q table using Equation (1)
End **while**

$$F_{\text{liq}} = F_{\text{obj}}, \tag{10}$$

$$\rho_{\text{liq}} V_{\text{liq}} a_{\text{liq}} = \rho_{\text{obj}} V_{\text{obj}} a_{\text{obj}}, \tag{11}$$

where, $\rho_{(.)}$, $V_{(.)}$, and $a_{(.)}$ depict the density, volume, and acceleration, respectively. Acceleration is obtained by rearranging Equation (11) as shown in Equation (12),

$$a_{\text{obj}} = \frac{\rho_{\text{liq}} V_{\text{liq}} a_{\text{liq}}}{\rho_{\text{obj}} V_{\text{obj}}}, \tag{12}$$

Further, for externally influencing forces, such as collision with other objects, Equations (10) and (11) are modified to Equations (13) and (14), respectively.

$$W_{\text{liq}} - W_{eo} = F_{\text{obj}}, \tag{13}$$

$$\rho_{\text{liq}} V_{\text{liq}} a_{\text{liq}} - \rho_{eo} V_{eo} a_{eo} = \rho_{\text{obj}} V_{\text{obj}} a_{\text{obj}}, \tag{14}$$

where, eo corresponds to the external object that is interacting and $W_{(.)}$ depicts the weight.

Positions of the objects, Z_{oi} are initialized using Equation (15) where lb_o and ub_o indicate the lower and upper bounds, respectively, and r_1 is a random number in the range [0, 1]. Also, the densities and volumes are initialized randomly in the range [0, 1].

$$Z_{oi} = lb_o + r_1 (ub_o - lb_o) \tag{15}$$

Acceleration is initialized as shown in Equation (16), acl_{oi} and r_2 indicate an acceleration of the ith object and random number, respectively.

$$acl_{oi} = lb_o + r_2 (ub_o - lb_o) \tag{16}$$

From the initial population, the object with the best position, Z_α is computed and correspondingly acl_α, ρ_α, and V_α are assigned. The TO (Equation 17) defines the search from exploration to exploitation. Additionally, the density factor (df) indicates the exploration to exploitation shift, which is calculated using Equation (18). Here, it and Max_it depict the current iteration and the maximum number of iterations, respectively.

$$\text{TO} = \exp\left(\frac{it - \text{Max}\_it}{\text{Max}\_it}\right), \tag{17}$$

$$\text{df} = \exp\left(\frac{it - \text{Max}\_it}{\text{Max}\_it}\right) - \left(\frac{it}{\text{Max}_{it}}\right), \tag{18}$$

Further, the densities and volumes are updated using Equations (19), (20), r_3 and r_4 represent random numbers and $\rho_{oi}(it+1)$ and $V_{oi}(it+1)$ indicate the updated densities and volumes during the iteration it, respectively.

$$\rho_{oi}(it+1) = \rho_{oi}(it) + r_3(\rho_\alpha - \rho_{oi}(it)), \tag{19}$$

$$V_{oi}(it+1) = V_{oi}(it) + r_4(V_\alpha - V_{oi}(it)), \tag{20}$$

During the exploration phase that is, when TO ≤ 0.5, the acceleration and position are updated using Equations (21) and (23), respectively. Before updating the position, the acceleration is normalized using Equation (22).

$$\text{acl}_{oi}(it+1) = \frac{\rho_{or} + V_{or}\text{acl}_{or}}{\rho_{oi}(it+1) V_{oi}(it+1)}, \tag{21}$$

$$\text{norm\_acl}_{oi}(it+1) = u \times \frac{\text{acl}_{oi}(it+1) - \min(\text{acl})}{\max(\text{acl}) - \min(\text{acl})} + l, \tag{22}$$

$$Z_{oi}(it+1) = Z_{oi}(it) + c_1 \times r_5 \times \text{norm\_acl}_{oi}(it+1) \times df \times (Z_{or} - Z_{oi}(it)), \tag{23}$$

where, c_1 is a constant, r_5 is a random number and *or* represents the random object.

when $\text{TO} > 0.5$, the search is in exploitation, and acceleration is updated using Equation (24). After normalizing the acceleration using Equation (22), the position is updated using Equation (25).

$$\text{acl}_{oi}(it+1) = \frac{\rho_\alpha + V_\alpha \text{acl}_\alpha}{\rho_{oi}(it+1) V_{oi}(it+1)}, \tag{24}$$

$$Z_{oi}(it+1) = Z_\alpha(it) + \text{Flag} \times c_2 \times r_6 \times \text{norm\_acl}_{oi}(it+1) \times df \times (\text{TF} \times Z_\alpha - Z_{oi}(it)), \tag{25}$$

where, c_2 is a constant, r_6 is a random number, TF is given using Equation (26) and Flag by Equation (27).

$$\text{TF} = c_3 \times TO, \tag{26}$$

$$\text{Flag} = \begin{cases} +1, \text{ if } F \leq 0.5 \\ -1, \text{ otherwise} \end{cases}, \tag{27}$$

where, c_3 is another constant, and F is given using Equation (28).

$$F = 2 \times r_7 - c_4, \tag{28}$$

where, c_4 is a constant, and r_7 is a random number.

3 | THE PROPOSED RLGA

The above discussions enlighten that the AOA performs exploration during the initial 30% of the iterations and updates the parameters based on a random selection of objects. Therefore, the algorithm is considered to search the unexplored regions of the search space, which yields slow convergence. On the contrary, the GWO offers good convergence speed because of selection of the best three wolves. Still, the unified search behaviour may not always guarantee a global solution; therefore, the present work proposes to enhance the exploration capability of AOA derived from GWO. Moreover, in AOA, TO controls the switching between exploration-exploitation and forces the algorithm to continue in exploration for one-third of the iterations. Therefore, many times, the AOA may over-exploit the search space, resulting in a large computational time. This necessitates the employment of intelligent techniques for optimum switching, and the embedding of *Q* learning-based RL may accomplish this task efficiently. Hence, the proposed RLGA enhances the exploration of AOA by employing GWO and obtains optimum switching by incorporating *Q* learning, which may uplift the overall solution quality and reduces the computational time of AOA.

As discussed earlier, *Q* learning is an RL where *Q* indicates the action-value function. During each operation, the *Q* table is updated by Equation (1). This *Q* table in the proposed RLGA consists of two actions and states: exploration and exploitation. The present study employs the ε greedy policy to balance the exploration-exploitation in RL (Liu et al., 2021; Tapia et al., 2021). The probability of exploitation or local search is ε, and exploration or global search is $1-\varepsilon$. Here, the learning rate (μ), as mentioned in Equation (1), also updates using Equation (29), which controls the rate of learning during each iteration.

$$\mu = \frac{\mu_{\text{initial}} + \mu_{\text{final}}}{2} - \frac{\mu_{\text{initial}} - \mu_{\text{final}}}{2 \times \cos\left(\pi\left(1 - \frac{it}{\text{Max}_{it}}\right)\right)}, \tag{29}$$

where, μ_{initial} and μ_{final} depict the initial and final values of μ, respectively. The value of μ decides the dependency on the current and new information. When the value of μ is decreased adaptively during each iteration, the *Q* table will become more influenced by the current *Q* values. Further, the reward (*R*) (Equation 1) is assigned to the learning agent based on the attained cost function value, as shown in Equation (30).

$$R = \begin{cases} +1, & \text{if cost function is improved} \\ -1, & \text{otherwise} \end{cases}. \tag{30}$$

Inspired by GWO, the acceleration of the particle during exploration is calculated using the densities and volumes of the best three objects as denoted in Equation (31).

$$\mathrm{acl}_{oi}(it+1) = \frac{\rho_{\mathrm{epl}} + V_{\mathrm{epl}\times} acl_{\mathrm{epl}}}{\rho_{\mathrm{oi}}(it+1) \times V_{\mathrm{oi}}(it+1)}, \tag{31}$$

where, ρ_{oi} and V_{oi} are obtained using Equations (19), (20), ρ_{epl}, V_{epl}, and acl_{epl} are computed using Equations (32)–(34) based on the best three objects α, β, and δ, respectively.

$$\rho_{\mathrm{epl}} = \frac{\rho_\alpha + \rho_\beta + \rho_\delta}{3}, \tag{32}$$

$$V_{\mathrm{epl}} = \frac{V_\alpha + V_\beta + V_\delta}{3}, \tag{33}$$

$$\mathrm{acl}_{\mathrm{epl}} = \frac{acl_\alpha + acl_\beta + acl_\delta}{3}, \tag{34}$$

Then, the acceleration obtained using Equation (31) is normalized using Equation (22). Consequently, the position is getting updated using GWO as presented in Equation (35).

$$Z_{oi}(it+1) = Z_{\mathrm{epl}}(it) + c_1 \times r_5 \times \mathrm{norm}_{\mathrm{acl\,oi}}(it+1) \times df \times \left(Z_{\mathrm{epl}}(it) - Z_{oi}(it)\right), \tag{35}$$

where, c_1 is a constant, r_5 is a random number, and Z_{epl} is calculated based on Equations (36)–(44).

$$Z_{\mathrm{epl}} = \frac{x_1 + x_2 + x_3}{3}, \tag{36}$$

$$x_1 = Z_\alpha(it) - A_1 \times D_\alpha, \tag{37}$$

$$x_2 = Z_\beta(it) - A_2 \times D_\beta, \tag{38}$$

$$x_3 = Z_\delta(it) - A_3 \times D_\delta, \tag{39}$$

$$D_\alpha = |C_1 \times Z_\alpha(it) - Z_{oi}(it)|, \tag{40}$$

$$D_\beta = |C_2 \times Z_\beta(it) - Z_{oi}(it)|, \tag{41}$$

$$D_\delta = |C_3 \times Z_\delta(it) - Z_{oi}(it)|, \tag{42}$$

$$A_k = 2 \times a_t \times r_{1k} - a_t, \tag{43}$$

$$C_k = 2 \times r_{2k}. \tag{44}$$

Finally, the step-by-step methodology of the proposed RLGA is depicted using the self-explanatory flowchart illustrated in Figure 1 and Pseudocode 2.

The Q table is designed as a 2×2 matrix, and its update mechanism is depicted in Figure 2. Each cell in the Q table represents the Q value estimated for the provided state in the row during the corresponding action in the column. If the state is selected as exploration before the update, then the action continues to be in exploration (highest Q value). However, if after updating this table using ε greedy policy, the state is assigned as exploitation, then the highest Q value corresponding to this state governs the respective action. Therefore, the agent maximizes the reward by selecting the action with the highest Q value. In addition, the independent learning process is ensured by individual Q tables for the entire population.

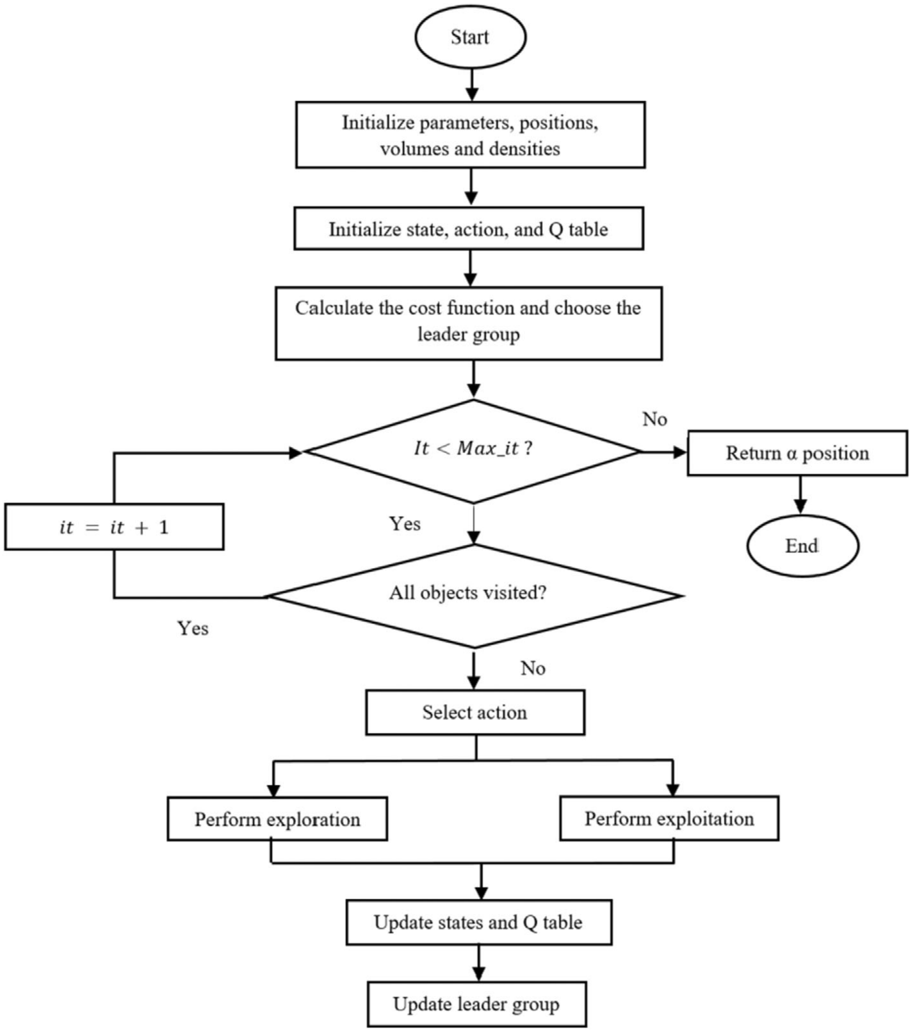

FIGURE 1 Flowchart of the reinforcement learning-based grey wolf optimization-Archimedes optimization algorithm

4 | BENCHMARK FUNCTION EVALUATION

4.1 | The algorithm setup

Initially, the effectiveness of RLGA is evaluated in terms of both accuracy and convergence speeds on four sets of benchmark functions. For this purpose, the optimum values of the cost function are evaluated for 1000 iterations with a population size of 30 during all the tests. The mean values are calculated using obtained optimized cost function values over 20 runs with the parameters listed in Table 1.

4.2 | Benchmark function analysis

The RLGA is used to evaluate the 22 benchmark functions as depicted in Appendix S1, single (Awad et al., 2017) and constrained single (Wu et al., 2016) objective real parameter optimization (CEC 2017), and Single objective real parameter optimization, 100 Digit Challenge (Price et al., 2018) (CEC 2019). The obtained results are compared and analysed with PSO (Kennedy & Russell, 1995), GWO (Mirjalili et al., 2014), AOA (Hashim et al., 2021), EAOA (Desuky et al., 2021), and RLGWO (Qu, Gai, Zhong, & Zhang, 2020). Table 2 indicates the number of estimated optimum values and their rank for all the compared algorithms for the benchmark functions mentioned in Appendix S1. The detailed results are provided in Appendix S1. Further, the average ranks of all the compared algorithms are computed on the considered benchmark functions by employing the Friedman test. From Table 2, it is quite evident that on evaluating the average values, the developed RLGA performs superior to

Pseudocode 2 The proposed RLGA

Set the parameters
Set the states and actions
Set the Q table
Initialize the population
Initialize density, volume, and acceleration
Calculate the cost function
Choose the leader group positions volumes, densities, and acceleration
Set $it = 0$
while it < Max_it**do**
update a, μ, TF, and df
for every object in population **do**
ε greedy policy derived from Q selects the action
switch action
    Case 1: exploration
        Calculate x_1, x_2, and x_3 using Equations (37)–(49)
        Calculate ρ_{epl}, V_{epl}, and acl_{epl} using Equations (32)–(34)
        Update acl_{oi} as mentioned in Equation (31)
        Update Z_{oi} as mentioned in Equation (35)
    Case 2: exploitation
        Update acl_{oi} as mentioned in Equation (24)
        Update Z_{oi} as mentioned in Equation (25)
end switch
end for
  Update cost function
  Get reward using Equation (30)
  Update leader group
  Update the Q table
  $it = it + 1$
end while
Return position

Actions

States		Exploration	Exploitation
	Exploration	10.8	5.4
	Exploitation	1.6	-2.3

Before Q table update

Actions

States		Exploration	Exploitation
	Exploration	9.6	8.3
	Exploitation	2.5	0.09

After Q table update

FIGURE 2 The Q table update

the other compared metaheuristics. For 12 out of 22 considered functions, the RLGA results in optimal or near-optimal values, indicating outstanding performance on 54.54% of the functions followed by GWO.

Table 3 indicates the summary of the statistical results obtained for CEC 2017 single objective real parameter optimization, whereas the detailed analysis is provided in Appendix S1. From analysing the mean values presented in Table 3, it is noticed that the RLGA is able to obtain optimal values for 14 functions among the 29 benchmark functions. In addition, RLGA obtains an average rank of 2.66, which is 2.21% less than its closest competitor (GWO).

The performance of RLGA is observed to be significantly higher for CEC 2017 constrained single objective real parameter optimization. The summary of estimated statistical results is represented in Table 4, whereas the detailed results are tabulated in Appendix S1. Among the total 28 functions, RLGA gives optimal mean values for 19 functions when compared with other discussed algorithms. The obtained results indicate the overwhelming performance of RLGA on 67.87% for the CEC 2017 constrained benchmark functions.

Finally, the summary of computed statistical results for CEC 2019 single objective real parameter optimization (100 Digit Challenge) is depicted in Table 5, and the detailed results are provided in Appendix S1. Although the simulated results reveal that none of the compared algorithms perform efficiently because of the very high complexities in this dataset, but still the Friedman test estimates RLGA as rank 1. Therefore, the proposed RLGA is assessed as the best-performing metaheuristic algorithm among the compared ones.

The convergence graphs are also acquired using the best optimal values from 20 runs for all the benchmark functions. The patterns of sample functions (F8, F14, F17, F20, F21, and F22) are illustrated in Figure 3a–f, respectively. Similarly, the convergence curves of sample functions in CEC 2017 constrained (F2, F3, F6, F8, F10, F12, F24, F25, and F26) and unconstrained (F3, F4, F5, F10, F15, F19, F21, F28, and F29) are deliberated in Figures 4a–f and 5a–f, respectively. Also, for CEC 2019 sample functions (F1, F2, and F9), these are depicted in Figure 6a–c. Evidently, RLGA is observed most suitable among the compared metaheuristics. Specifically, Figure 3 indicates that RLGA is most prominent for the 22 selected benchmark functions (Table 2). It not only achieves the optimum or near-optimum cost values but also reduces the computational burden. Moreover, Figures 4, 5, and 6 verifies that the developed RLGA quickly obtains the optimum cost under various complexities and

TABLE 1 The parameters required for each algorithm

Algorithm	Parameters
AOA	$c_1=2,\ c_2=6,\ c_3=2,\ c_4=1$
GWO	$a=[0, 2]$
PSO	$w=1.09, c_1=c_2=2.2345$
EAOA	$c_1=2,\ c_2=6,\ c_3=2,\ c_4=1$
RLGWO	$a=[0, 2]$
RLGA	$c_1=2,\ c_2=6,\ c_3=2,\ c_4=1,\ a=[0, 2]$

TABLE 2 Summary of results obtained for benchmark functions mentioned in Appendix S1

Parameter	GWO (Mirjalili et al., 2014)	AOA (Hashim et al., 2021)	PSO (Kennedy & Russell, 1995)	EAOA (Desuky et al., 2021)	RLGWO (Qu, Gai, Zhong, & Zhang, 2020)	RLGA (proposed)
Best	9	10	11	4	10	11
Worst	11	4	5	2	6	9
SD	10	0	3	1	3	5
Mean	10	3	6	1	8	12
Average rank	2.55	3.91	4.32	4.86	3.18	2.18
Total rank	2	4	5	6	3	1

TABLE 3 Summary of results obtained for CEC 2017 single objective real parameter optimization

Parameter	GWO (Mirjalili et al., 2014)	AOA (Hashim et al., 2021)	PSO (Kennedy & Russell, 1995)	EAOA (Desuky et al., 2021)	RLGWO (Qu, Gai, Zhong, & Zhang, 2020)	RLGA (proposed)
Best	8	6	8	0	5	14
Worst	9	1	4	0	8	7
SD	10	1	2	1	5	9
Mean	11	1	3	0	3	14
Average rank	2.72	3.72	2.79	6.00	3.10	2.66
Total rank	2	5	3	6	4	1

TABLE 4 Summary of results obtained for CEC 2017 constrained single objective real parameter optimization

Parameter	GWO (Mirjalili et al., 2014)	AOA (Hashim et al., 2021)	PSO (Kennedy & Russell, 1995)	EAOA (Desuky et al., 2021)	RLGWO (Qu, Gai, Zhong, & Zhang, 2020)	RLGA (proposed)
Best	7	1	0	1	1	18
Worst	9	0	0	2	0	18
SD	5	0	1	4	1	17
Mean	6	1	0	2	0	19
Average rank	1.93	3.07	5.07	5.64	3.82	1.46
Total rank	2	3	5	6	4	1

TABLE 5 Summary of results obtained for CEC 2019 single objective real parameter optimization

Parameter	GWO (Mirjalili et al., 2014)	AOA (Hashim et al., 2021)	PSO (Kennedy & Russell, 1995)	EAOA (Desuky et al., 2021)	RLGWO (Qu, Gai, Zhong, & Zhang, 2020)	RLGA (proposed)
Best	1	1	4	1	0	3
Worst	1	2	3	1	0	3
SD	1	0	3	1	2	3
Mean	2	1	3	0	0	4
Average rank	2.72	3.72	2.79	3.10	6.00	2.65
Total rank	2	5	3	4	6	1

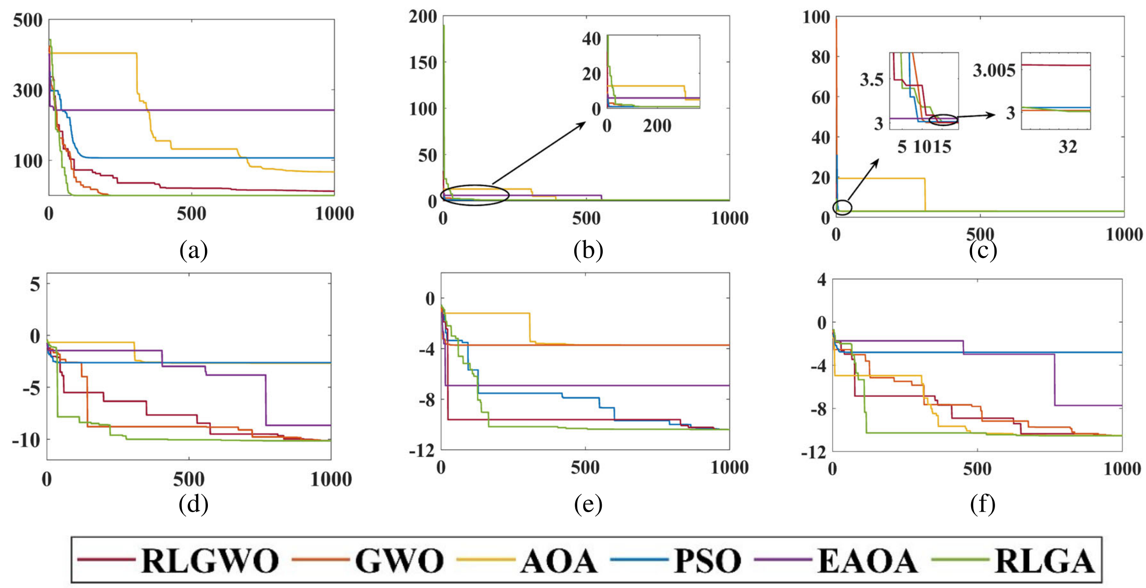

FIGURE 3 Convergence patterns for benchmark functions in Appendix S1. (a) F8, (b) F14, (c) F17, (d) F20, (e) F21 and (f) F22

difficulty levels. However, in some cases (Figures 5 and 6), it is observed that the RLGA is not the quickest yet produces optimum cost in a reasonable time relative to the compared algorithms.

5 | PATH PLANNING

5.1 | Problem formulation

The present work transformed the problem of estimating the optimum flyable route for UAVs in a 3D-constrained environment (varying shapes and sizes) into a non-linear, single objective constrained optimization problem, whose cost function is minimized by employing metaheuristic

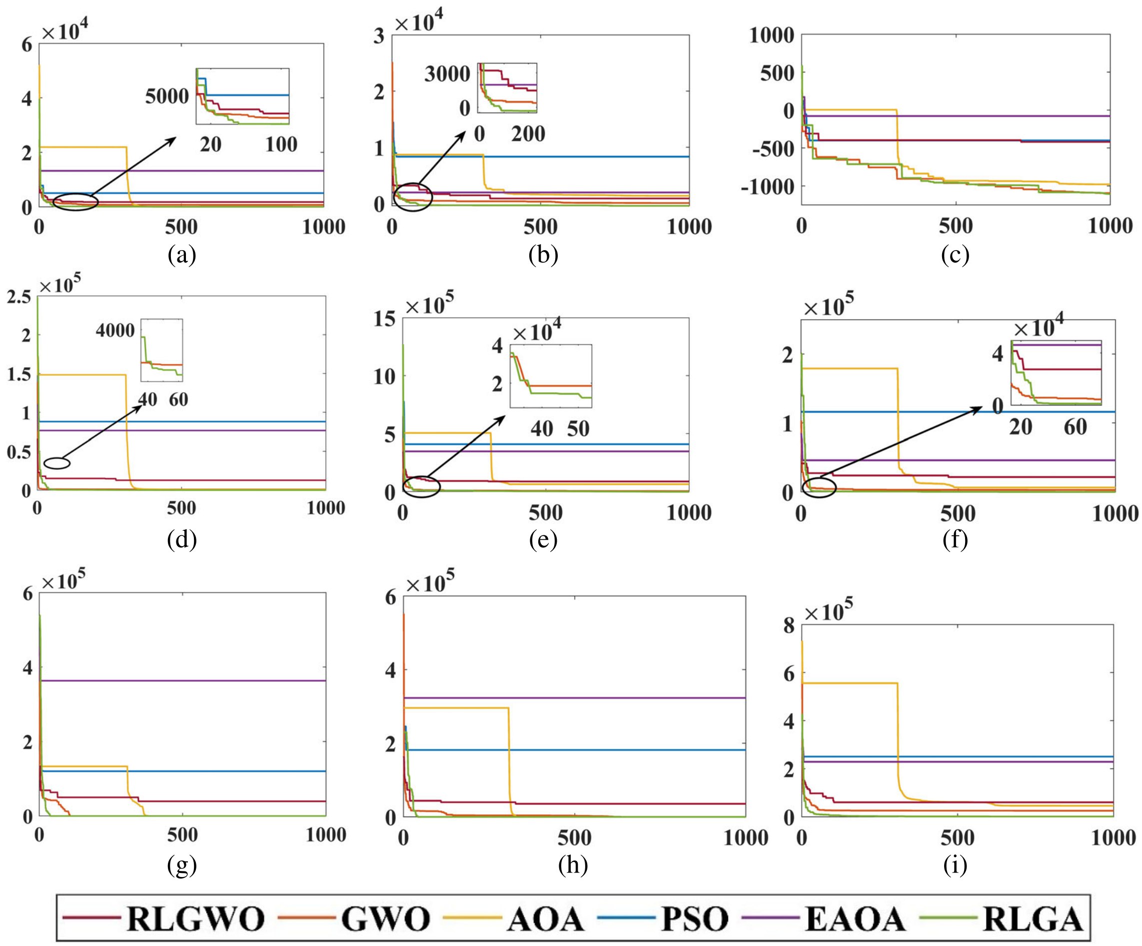

FIGURE 4 Convergence patterns for CEC 2017 constrained benchmark functions. (a) F2, (b) F3, (c) F6, (d) F8, (e) F10, (f) F12, (g) F24, (h) F25 and (i) F26

algorithms. For this purpose, the population is initialized with the candidate paths in order to avoid obstacles with minimum fuel consumption. The metaheuristics update the path and return the best possible route for the UAV to travel with minimum costs from the initial population of candidate paths. This ensures the employability of the UAV for effective and quick medical assistance in a real-world scenario. Further, the path is divided into *N* number of waypoints to calculate the cost function and to estimate the unhindered path.

5.1.1 | Cost function

The cost function for finding the best possible route through the 3D environment is formulated to produce a collision-free route, along with other constraints such as finding the shortest possible path, altitude, and angle constraints. In this work, the cost function is framed to estimate a safe trajectory from the initial position to the target with minimum fuel consumption and the slightest deviation from the shortest trajectory.

To develop the 3D simulating environment, multiple obstacles with various shapes and sizes are considered. The cost function (C_{cost}) is calculated as the weighted sum of fuel cost (C_{fuel}), obstacle collision cost ($C_{obstacle}$), and the deviations cost ($C_{deviation}$) as indicated in Equation (45). Here, m_1, m_2, m_3 and m_4 denote the weights associated with fuel, obstacle-collision, and deviations cost (y and z), respectively. Considering optimum and safe flyable path as the main requirement of the path planning problems, this work gives higher weightage to m_1 and m_2 relative to m_3 and m_4. These values are empirically selected as 55, 25, 10, and 10, respectively.

$$C_{cost} = m_1 C_{fuel} + m_2 C_{obstacle} + m_3 C_{ydeviation} + m_4 C_{zdeviation}, \tag{45}$$

The fuel cost (C_{fuel}) indicates the fuel consumed for the travel. Considering the path length as directly related to fuel consumed, C_{fuel} is calculated as the aggregated length of the candidate path. The path between the source and target is divided into N waypoints, and then the obstacle-collision cost is computed by dividing each path segment into five points, and each point is checked if it is inside the obstacle. If the path segment

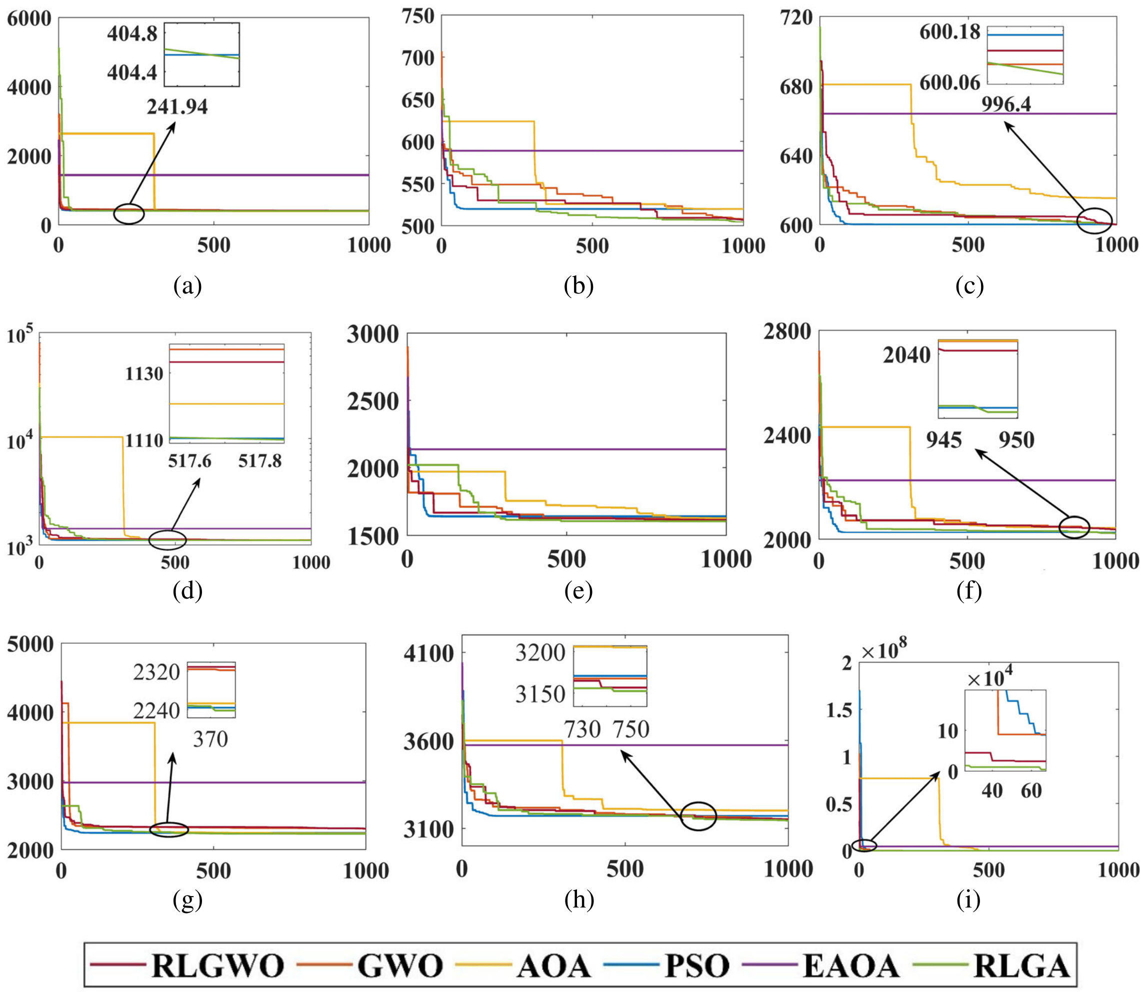

FIGURE 5 Convergence patterns for CEC 2017 unconstrained benchmark functions. (a) F3, (b) F4, (c) F5, (d) F10, (e) F15, (f) F19, (g) F21, (h) F28 and (i) F29

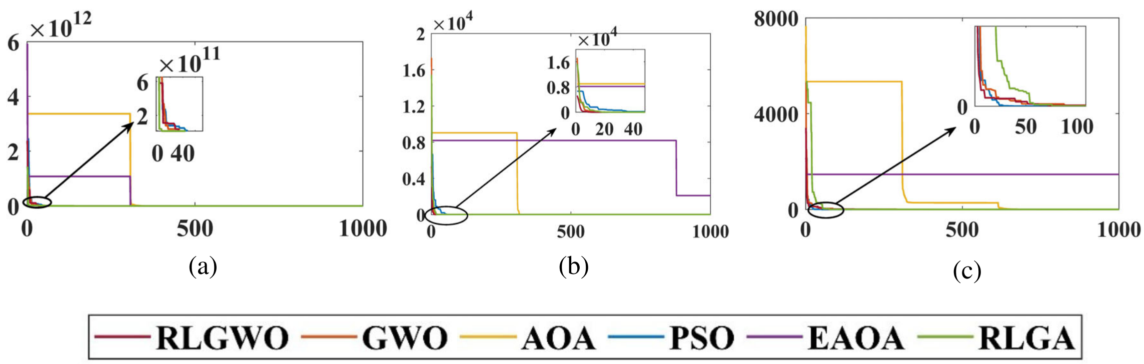

FIGURE 6 Convergence patterns for CEC 2019 constrained benchmark functions. (a) F3, (b), F4 and (c) F5

is inside the obstacle, $C_{obstacle}$ is calculated; otherwise, it is assigned as zero. The computational process of $C_{obstacle}$ is presented by Equations (46), (47). For an obstacle j, assume a path segment of length (l_p) of the candidate path passing through it, as illustrated in Figure 7. Here, $d^{p}_{0.25,j}$ denotes the distance from the center of the j^{th} obstacle to the second point on the path segment l_m under consideration and r_j indicates the base radius of the obstacle.

$$C_{obstacle} = \int_0^l c_{obstacle} dl, \tag{46}$$

$$c_{\text{obstacle},l_m} = \frac{l_p}{5}\sum_{j=1}^{N}\frac{1}{5}\left(|d^p_{0,j} - r_j| + |d^p_{0.25,j} - r_j| + |d^p_{0.5,j} - r_j| + |d^p_{0.75,j} - r_j| + |d^p_{1,j} - r_j|\right). \tag{47}$$

It should be guaranteed that the created flight routes must avoid collisions and should require the minimum energy. The deviations cost are evaluated using Equations (48)–(51), which signifies the deviation from consecutive waypoints or the oscillations from the straight path. Here, y_m and z_m indicate the coordinates of the mth waypoint on the estimated path that is being considered. These deviations cost can eliminate unnecessary oscillations that may arise in the generated paths. Further, cubic spline curves are used after path optimization to smoothen the obtained path (Mahdi et al., 2019; Ravankar et al., 2018).

$$C_{\text{ydeviation}} = \int_0^l c_{\text{ydeviation}}, \tag{48}$$

$$c_{\text{ydeviation}} = \sqrt{(y_m - y_{m-1})^2}, \tag{49}$$

$$C_{\text{zdeviation}} = \int_0^l c_{\text{zdeviation}}, \tag{50}$$

$$c_{\text{zdeviation}} = \sqrt{(z_m - z_{m-1})^2}. \tag{51}$$

5.1.2 | Flight environment

Being motivated by the efficacy of the developed RLGA in optimizing the benchmark functions with various difficulty levels, it is employed for route estimation in a complex 3D environment. The simulated 3D environment contains obstacles of different configurations, as explained in Table 6. The simulated obstacles include Gaussian shapes, cones, hemispheres, and cylinders. The obstacles are also merged to ensure a more realistic representation of the simulated environment. Different start and target points are selected for generating the path, and the developed algorithm is utilized to produce optimized paths between the mentioned start and target points. In addition, the path estimated by RLGA is compared against the aforementioned metaheuristics to endorse the effectiveness of RLGA in real-world problems. Further, all the compared algorithms are initialized with 100 candidate paths, and the simulation is carried out for 2000 iterations.

Based on the above discussion, the step-by-step procedure to generate the optimal flyable path for UAVs is illustrated by a self-explanatory flowchart in Figure 8.

5.2 | Path planning results

The performance of the developed RLGA on several benchmark functions served as the motivation to employ it for estimating the flyable path and corresponding cost function values in a complex 3D environment. The simulated results using RLGA are compared with the GWO, AOA,

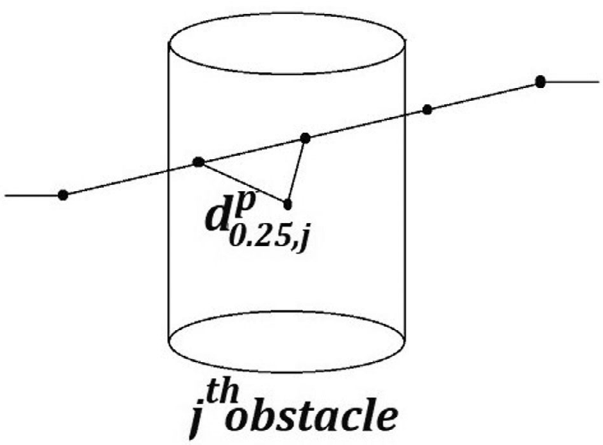

FIGURE 7 Calculating the obstacle-collision cost

PSO, EAOA, and RLGWO. For this purpose, the flight trajectory for UAV in the simulated 3D environment is estimated between the start position (−15, −15, 3) and the goal position (35, 35, 9). Further, the number of waypoints (N) is empirically selected as 15. The estimated trajectory for 2D (top view) and 3D by these metaheuristics are illustrated in Figures 9 and 10, respectively. The obtained results reveal that RLGA dominates all

TABLE 6 Obstacle configuration

Obstacle	Base center	Radius (m)	Height (m)
Gaussian	(0,0,0)	–	7
Cone	(−17,7,0)	2.5	10
	(−10,−7,0)	5	4
Hemispheres	(6,7,0)	5	–
	(20,16,0)	5	–
	(0,25,0)	6	–
	(30,10,0)	3.5	–
Cylinders	(22,0,0)	5	12
	(10,16,0)	3	8
	(12,−0.5,0)	4	8
	(15,8,0)	3	10
	(25,25,0)	2.7	9
	(15,25,0)	2.7	7.5
	(7.5,25,0)	1	15
	(−10,10,0)	5	6

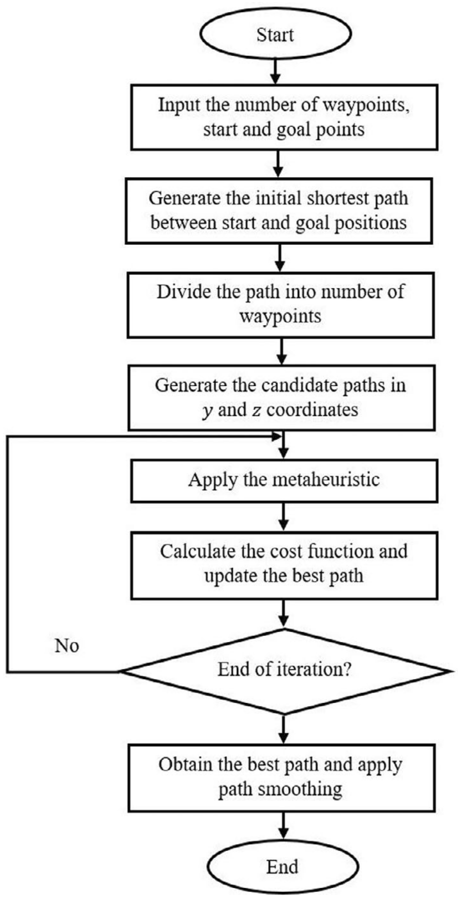

FIGURE 8 Flowchart of the reinforcement learning-based grey wolf optimization-Archimedes optimization algorithm

the compared algorithms and offers the optimum path. EAOA is able to deliver a collision-free path but at the cost of higher fuel requirements by producing a longer flight route. GWO also generates a collision-free path in the simulated environment but produces a longer path with oscillations about the z-axis because of the incorporation of the fuel and oscillation costs. The paths resulting from other algorithms, such as RLGWO, AOA, and PSO, are unable to provide a collision-free path in the considered environment.

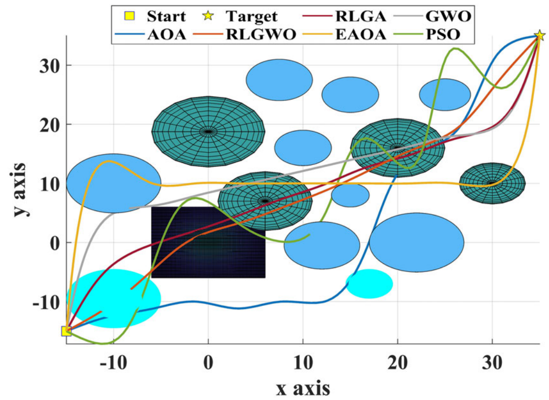

FIGURE 9 Top view of the acquired paths

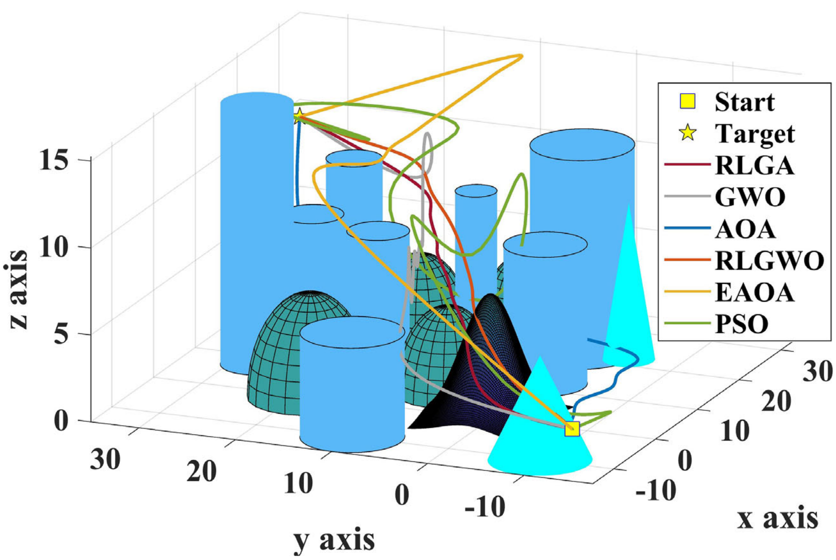

FIGURE 10 3D view of the generated paths

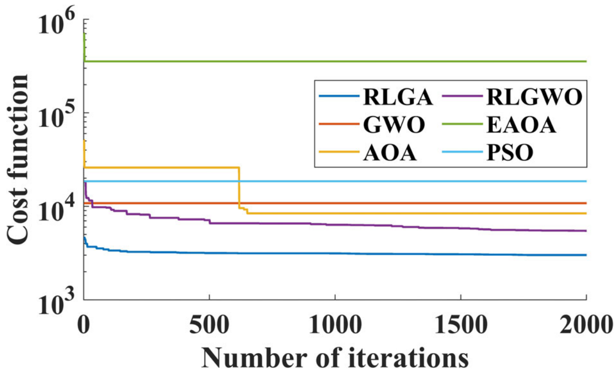

FIGURE 11 Convergence curve

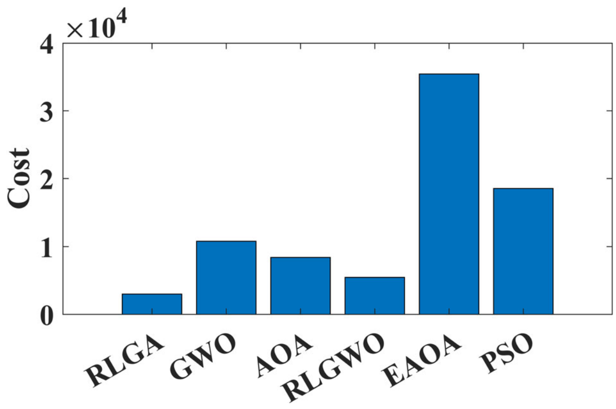

FIGURE 12 Obtained cost function values

TABLE 7 Cost function values for unmanned aerial vehicle path planning

S. no.	Algorithm	Best cost function value	Converging at iteration
1	PSO (Kennedy & Russell, 1995)	18547.07	50
2	GWO (Mirjalili et al., 2014)	10799.13	30
3	AOA (Hashim et al., 2021)	8421.02	650
4	EAOA (Desuky et al., 2021)	354243.95	10
5	RLGWO (Qu, Gai, Zhong, & Zhang, 2020)	6581.27	505
6	RLGA (Proposed)	3194.44	382

Furthermore, the convergence curve for 3D path planning is depicted in Figure 11 to evaluate the computational time required by considered metaheuristics. The RLGA converges faster and delivers optimum values compared with other algorithms, whereas EAOA, PSO, and GWO settle for a higher but constant value from the very first iteration itself. Additionally, the cost function values obtained by the considered metaheuristics are illustrated in Figure 12, which discloses the dominance of RLGA over other compared algorithms. Further, the obtained cost function and the required number of iterations to converge are tabulated in Table 7. It is observed that the RLGA provides 82.77%, 70.42%, 62.06%, 99.10%, and 51.46% lesser cost compared with PSO, GWO, AOA, EAOA, and RLGWO, respectively. However, it is noticed that only RLGA, EAOA, and GWO are able to deliver collision-free paths, therefore, are suitable for this task under the considered environmental conditions, whereas others collided with the obstacle boundaries. Furthermore, the proposed RLGA requires 41.23% and 24.35% lesser iterations than AOA and RLGWO, respectively. Nonetheless, it requires higher iterations compared with PSO, GWO, and EAOA to achieve the optimum values, but considering the complexities involved in the path planning and the dominance of RLGA over these metaheuristics for optimum path generation, this may be neglected.

6 | CONCLUSIONS

In critical medical scenarios, response time is very crucial, and many times, ambulances fail to reach the required destination because of traffic congestion. Thus, the present work focuses on emergency medical intervention, which reduces fatality and lethality rates by providing the medical facility with UAVs in a limited time window. However, it requires an efficient algorithm to estimate the optimal and collision-free trajectory in a complex 3D environment with several dissimilar obstacles. In view of this, a novel RLGA is developed to estimate a safe and flyable path for UAVs in a complex 3D environment under consideration of various constraints. The efficacy of the proposed RLGA is first validated on benchmark functions with different difficulty levels, including CEC 2017 and CEC 2019. The acquired results are compared with the most popular metaheuristics (PSO, GWO, AOA, RLGWO, and EAOA), which indicates the superior convergence capability of RLGA over other algorithms under discussion. Based on the Friedman test, the rank of RLGA is estimated as 1 for all the benchmark functions, which validates the efficacy of RLGA against the compared CIAs. The convergence patterns of RLGA indicate that it is able to significantly improve the convergence rates of AOA because *Q* learning enables intelligent switching between exploration and exploitation. Further, it is observed that the RLGA obtains a safer path with 51.46% and 70.42% lesser cost than RLGWO and GWO, respectively. In addition, compared with AOA, it is found that RLGA delivers a safe trajectory with 62.06% lesser cost. This validates the dominance of RLGA over other algorithms in providing global or near-optimal solutions. The convergence curves also affirm the superiority of RLGA over other CIAs to yield better cost function values at faster convergence rates. The employability of RLGA over benchmark functions and path planning demonstrates the promising capabilities of the developed algorithm. However, while investigating the convergence pattern on benchmark functions, it is evident that the convergence rate of RLGA is slower than PSO and GWO. Nevertheless, the high convergence rates, flying range, and payload capacity are required to overcome the time pressure of the healthcare system using UAVs which would be very interesting and provide new directions for future research.

FUNDING INFORMATION

This research was funded by Jagadish Chandra Bose Research Organisation (JCBRO) (File No. Research/sponsor/ADHAAR_2021).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest to report regarding the present study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Om Prakash Verma https://orcid.org/0000-0002-7421-295X

REFERENCES

Awad, N. H., Ali, M. Z., Suganthan, P. N., Liang, J. J., & Qu, B. Y. (2017). Special session & competitions on real-parameter single objective optimization. In *Proceedings of the IEEE congress on evolutionary computation (CEC)*. IEEE.

Chen, Y. B., Luo, G. C., Mei, Y. S., Yu, J. Q., & Su, X. L. (2016). UAV path planning using artificial potential field method updated by optimal control theory. *International Journal of Systems Science*, *47*(6), 1407–1420. https://doi.org/10.1080/00207721.2014.929191

Damoah, I. S., Ayakwah, A., & Tingbani, I. (2021). Artificial intelligence (AI)-enhanced medical drones in the healthcare supply chain (HSC) for sustainability development: A case study. *Journal of Cleaner Production*, *328*, 129598. https://doi.org/10.1016/J.JCLEPRO.2021.129598

Desuky, A. S., Hussain, S., Kausar, S., Islam, A., & El Bakrawy, L. M. (2021). EAOA: An enhanced Archimedes optimization algorithm for feature selection in classification. *IEEE Access*, *9*, 120795–120814. https://doi.org/10.1109/ACCESS.2021.3108533

Dorigo, M., & Di Caro, G. (1999). Ant colony optimization: A new meta-heuristic. In *Proceedings of the 1999 congress on evolutionary computation-CEC99* (Cat. No. 99TH8406) (Vol. 2, pp. 1470–1477). IEEE.

Dupont, Q. F. M., Chua, D. K. H., Tashrif, A., & Abbott, E. L. S. (2017). Potential applications of UAV along the construction's value chain. *Procedia Engineering*, *182*(3), 165–173. https://doi.org/10.1016/j.proeng.2017.03.155

Eskandar, H., Sadollah, A., Bahreininejad, A., & Hamdi, M. (2012). Water cycle algorithm - a novel metaheuristic optimization method for solving constrained engineering optimization problems. *Computers and Structures*, *110–111*, 151–166. https://doi.org/10.1016/j.compstruc.2012.07.010

Gigante, G., Pascarella, D., Luongo, S., Di Benedetto, C., Vozzella, A., & Persechino, G. (2018). Game-theoretic approach for the optimal configuration computing of an interoperable fleet of unmanned vehicles. *Expert Systems*, *35*(5), e12293. https://doi.org/10.1111/EXSY.12293

Giordan, D., Adams, M. S., Aicardi, I., Alicandro, M., Allasia, P., Baldo, M., de Berardinis, P., Dominici, D., Godone, D., Hobbs, P., Lechner, V., Niedzielski, T., Piras, M., Rotilio, M., Salvini, R., Segor, V., Sotier, B., & Troilo, F. (2020). The use of unmanned aerial vehicles (UAVs) for engineering geology applications. *Bulletin of Engineering Geology and the Environment*, *79*(7), 3437–3481. https://doi.org/10.1007/s10064-020-01766-2

Gomaa, A., Abdelwahab, M. M., & Abo-Zahhad, M. (2020). Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis. *Multimedia Tools and Applications*, *79*(35), 26023–26043. https://doi.org/10.1007/s11042-020-09242-5

Gupta, H., Kumar, S., Yadav, D., Verma, O. P., Sharma, T. K., Ahn, C. W., & Lee, J. H. (2021). Data analytics and mathematical modeling for simulating the dynamics of COVID-19 epidemic—A case study of India. *Electronics*, *10*(2), 127. https://doi.org/10.3390/electronics10020127

Gupta, H., & Verma, O. P. (2021). Monitoring and surveillance of urban road traffic using low altitude drone images: A deep learning approach. *Multimedia Tools and Applications*, *81*, 1–21. https://doi.org/10.1007/s11042-021-11146-x

Hashim, F. A., Hussain, K., Houssein, E. H., Mabrouk, M. S., & Al-Atabany, W. (2021). Archimedes optimization algorithm: A new metaheuristic algorithm for solving optimization problems. *Applied Intelligence*, *51*(3), 1531–1551. https://doi.org/10.1007/s10489-020-01893-z

He, W., Qi, X., & Liu, L. (2021). A novel hybrid particle swarm optimization for multi-UAV cooperate path planning. *Applied Intelligence*, *2021*, 1–15. https://doi.org/10.1007/S10489-020-02082-8

Huang, C., & Fei, J. (2018). UAV path planning based on particle swarm optimization with global best path competition. *International Journal of Pattern Recognition and Artificial Intelligence*, *32*(6), 1859008. https://doi.org/10.1142/S0218001418590085

Jain, G., Yadav, G., Prakash, D., Shukla, A., & Tiwari, R. (2019). MVO-based path planning scheme with coordination of UAVs in 3-D environment. *Journal of Computer Science*, *37*, 101016. https://doi.org/10.1016/j.jocs.2019.07.003

Karami, H., Valikhan, M., Farzin, S., & Mirjalili, S. (2020). Computers & industrial engineering flow direction algorithm (FDA): A novel optimization approach for solving optimization problems. *Computers and Industrial Engineering*, *156*, 2021.

Kennedy, J., & Russell, E. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4). IEEE.

Khan, S. I., Qadir, Z., Munawar, H. S., Nayak, S. R., Budati, A. K., Verma, K. D., & Prakash, D. (2021). UAVs path planning architecture for effective medical emergency response in future networks. *Physics Communication*, *47*, 101337. https://doi.org/10.1016/J.PHYCOM.2021.101337

Li, S., & Deng, Y. (2019). Quantum-entanglement pigeon-inspired optimization for unmanned aerial vehicle path planning. *Aircraft Engineering and Aerospace Technology*, *91*(1), 171–181. https://doi.org/10.1108/AEAT-03-2018-0107

Liu, X., Du, X., Zhang, X., Zhu, Q., & Guizani, M. (2019). Evolution-algorithm-based unmanned aerial vehicles path planning in complex environment. *Computers and Electrical Engineering*, *80*, 106493. https://doi.org/10.1016/j.compeleceng.2019.106493

Liu, Y., Cao, B., & Li, H. (2021). Improving ant colony optimization algorithm with epsilon greedy and levy flight. *Complex & Intelligent Systems*, *7*(4), 1711–1722. https://doi.org/10.1007/s40747-020-00138-3

Mahdi, A. A. H., Sahib, M., & Altaei, M. (2019). Cubic spline based path planning for UAV. *International Journal of Computer Science and Mobile Computing*, *8*, 80–95.

Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, *95*, 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, *69*, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

Phung, M. D., & Ha, Q. P. (2021). Safety-enhanced UAV path planning with spherical vector-based particle swarm optimization. *Applied Soft Computing*, *107*(April), 107376. https://doi.org/10.1016/j.asoc.2021.107376

Price, K. V., Awad, N. H., Ali, M. Z., & Suganthan, P. N. (2018). *The 100-digit challenge: Problem definitions and evaluation criteria for the 100-digit challenge special session and competition on single objective numerical optimization*. Nanyang Technological University.

Primatesta, S., Guglieri, G., & Rizzo, A. (2018). A risk-aware path planning strategy for UAVs in urban environments. *Journal of Intelligent & Robotic Systems*, *95*(2), 629–643. https://doi.org/10.1007/S10846-018-0924-3

Qu, C., Gai, W., Zhang, J., & Zhong, M. (2020). A novel hybrid grey wolf optimizer algorithm for unmanned aerial vehicle (UAV) path planning. *Knowledge-Based Systems*, *194*, 105530. https://doi.org/10.1016/j.knosys.2020.105530

Qu, C., Gai, W., Zhong, M., & Zhang, J. (2020). A novel reinforcement learning based grey wolf optimizer algorithm for unmanned aerial vehicles (UAVs) path planning. *Applied Soft Computing Journal*, *89*, 106099. https://doi.org/10.1016/j.asoc.2020.106099

Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algorithm. *Information Sciences*, *179*(13), 2232–2248. https://doi.org/10.1016/j.ins.2009.03.004

Ravankar, A., Ravankar, A. A., Kobayashi, Y., Hoshino, Y., & Peng, C. C. (2018). Path smoothing techniques in robot navigation: State-of-the-art, current and future challenges. *Sensors*, *18*(9), 1–30. https://doi.org/10.3390/s18093170

Seyyedabbasi, A., Aliyev, R., Kiani, F., Gulle, M. U., Basyildiz, H., & Shah, M. A. (2021). Hybrid algorithms based on combining reinforcement learning and metaheuristic methods to solve global optimization problems. *Knowledge-Based Systems*, *223*, 107044. https://doi.org/10.1016/j.knosys.2021.107044

Shan, E., Dai, B., Song, J., & Sun, Z. (2009). An dynamic RRT path planning algorithm based on B-spline. In *Proceedings of the international IEEE international symposium on computational intelligence and design* (Vol. 2, pp. 25–29). IEEE. https://doi.org/10.1109/ISCID.2009.155

Shin, J. J., & Bang, H. (2020). UAV path planning under dynamic threats using an improved PSO algorithm. *International Journal of Aerospace Engineering*, *2020*, 1–17. https://doi.org/10.1155/2020/8820284

Sierra-Garcia, J. E., & Santos, M. (2022). Combining reinforcement learning and conventional control to improve automatic guided vehicles tracking of complex trajectories. *Expert Systems*, e13076. https://doi.org/10.1111/EXSY.13076

Sreelakshmy, K., Gupta, H., Ansari, I. A., Sharma, S., Goyal, K. K., & Verma, O. P. (2022). Metaheuristic optimization for three dimensional path planning of UAV. *Lecture Notes in Networks and Systems*, *425*, 791–802. https://doi.org/10.1007/978-981-19-0707-4_71/COVER

Subhan, I., Ghazi, S. S., & Nabi, S. (2019). Use of drones (unmanned aerial vehicles) for supporting emergency medical services in India. *Apollo Medicine*, *16*(1), 61. https://doi.org/10.4103/AM.AM_79_18

Sun, X., Andoh, E. A., & Yu, H. (2021). A simulation-based analysis for effective distribution of COVID-19 vaccines: A case study in Norway. *Transportation Research Interdisciplinary Perspectives*, *11*, 100453. https://doi.org/10.1016/J.TRIP.2021.100453

Tang, A.-D., Han, T., Zhou, H., & Xie, L. (2021). An improved equilibrium optimizer with application in unmanned aerial vehicle path planning. *Sensors*, *21*(5), 1814. https://doi.org/10.3390/s21051814

Tapia, D., Soto, R., Palma, W., Lemus-Romani, J., Cisternas-Caneo, F., Castillo, M., Becerra-Rozas, M., Paredes, F., & Misra, S. (2021). Embedding Q-learning in the selection of metaheuristic operators: The enhanced binary grey wolf optimizer case. In *Proceedings od the IEEE international conference on automation/XXIV congress of the chilean association of automatic control* (Vol. 2021). IEEE. https://doi.org/10.1109/ICAACCA51523.2021.9465259

Tatum, M. C., & Liu, J. (2017). Unmanned aircraft system applications in construction. *Procedia Engineering*, *196*(June), 167–175. https://doi.org/10.1016/j.proeng.2017.07.187

Ucgun, H., Yuzgec, U., & Bayilmis, C. (2021). A review on applications of rotary-wing unmanned aerial vehicle charging stations. *International Journal of Advanced Robotic Systems*, *18*(3), 1–20. https://doi.org/10.1177/17298814211015863

Ullah, S., Kim, K. I., Kim, K. H., Imran, M., Khan, P., Tovar, E., & Ali, F. (2019). UAV-enabled healthcare architecture: Issues and challenges. *Future Generation Computer Systems*, *97*, 425–432. https://doi.org/10.1016/J.FUTURE.2019.01.028

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1109/4235.585893

Wu, G., Mallipeddi, R., & Suganthan, P. N. (2016). Problem definitions and evaluation criteria for the CEC 2017 competition and special session on constrained single objective real-parameter optimization. Technical report. http://www.ntu.edu.sg/home/epnsugan/

Wu, X., Bai, W., Xie, Y., Sun, X., Deng, C., & Cui, H. (2018). A hybrid algorithm of particle swarm optimization, metropolis criterion and RTS smoother for path planning of UAVs. *Applied Soft Computing*, *73*, 735–747. https://doi.org/10.1016/J.ASOC.2018.09.011

Xu, H., Jiang, S., & Zhang, A. (2021). Path planning for unmanned aerial vehicle using a mix-strategy-based gravitational search algorithm. *IEEE Access*, *9*, 57033–57045. https://doi.org/10.1109/ACCESS.2021.3072796

Yadav, A. (2019). AEFA: Artificial electric field algorithm for global optimization. *Swarm and Evolutionary Computation*, *48*(March), 93–108. https://doi.org/10.1016/j.swevo.2019.03.013

Yang, P., Tang, K., Lozano, J. A., & Cao, X. (2015). Path planning for single unmanned aerial vehicle by separately evolving waypoints. *IEEE Transactions on Robotics*, *31*(5), 1130–1146. https://doi.org/10.1109/TRO.2015.2459812

Yu, X., Li, C., & Zhou, J. F. (2020). A constrained differential evolution algorithm to solve UAV path planning in disaster scenarios. *Knowledge-Based Systems*, *204*, 106209. https://doi.org/10.1016/j.knosys.2020.106209

Zhang, S., Zhou, Y., Li, Z., & Pan, W. (2016). Grey wolf optimizer for unmanned combat aerial vehicle path planning. *Advances in Engineering Software*, *99*, 121–136. https://doi.org/10.1016/j.advengsoft.2016.05.015

Zhao, Y., Zheng, Z., & Liu, Y. (2018). Survey on computational-intelligence-based UAV path planning. *Knowledge-Based Systems*, *158*, 54–64. https://doi.org/10.1016/J.KNOSYS.2018.05.033

Zhou, X., Gao, F., Fang, X., & Lan, Z. (2021). Improved bat algorithm for UAV path planning in three-dimensional space. *IEEE Access*, *9*, 20100–20116. https://doi.org/10.1109/ACCESS.2021.3054179

AUTHOR BIOGRAPHIES

Himanshu Gupta received his B. Tech degree in Electronics and Instrumentation Engineering from Uttar Pradesh Technical University, India, in 2006 and M. Tech degree in Control and Instrumentation from Dr B R Ambedkar National Institute of Technology Jalandhar, India in 2012. He is currently working towards the Ph.D. degree from Dr B R Ambedkar National Institute of Technology Jalandhar, India. His research interest includes soft-computing, optimization, and computer vision.

K. Sreelakshmy completed B. Tech degree in Applied Electronics and Instrumentation Engineering from Government Engineering College Kozhikode, India in 2019 and M. Tech degree in Control and Instrumentation Engineering from Dr. B R Ambedkar National Institute of Technology, Jalandhar, India in 2022. She is currently employed in Research and Development wing of Chetak Technology Limited, India. Her current research interests are electric vehicles and optimization algorithms.

Om Prakash Verma is currently serving as an Assistant Professor in the Department of Instrumentation and Control Engineering, Dr. B R Ambedkar National Institute of Technology, Jalandhar, India. His research interests includes Machine/Deep/Quantum Learning, Machine/Computer Vision, and UAV Autonomous Navigation System. He is a senior member of IEEE and member of IEEE Computational Intelligence Society, IEEE Control Systems Society, and Automatic Control and Dynamic Optimization Society (ACDOS). He is life time member of other international renowned societies such as Instrument Society of India (ISOI) and STEM Research Society (STEM-RS). He has credit for publishing more than 90+ research publications including international peer-reviewed SCI journals, patent applications, edited books, conferences, and book chapters. He has credit to author the book titled Butterfly Optimization Algorithm: Theory and Engineering Applications, published by Springer and edited seven different renowned book of international conference proceedings which are published by Springer Nature. He has associated with five research projects as PI and Co-PI funded by various funding agencies of cumulative amount 89.26 Crore. He has supervised one Ph.D and currently supervising two Ph.D (one regular and one part time).

Tarun Kumar Sharma holds Ph.D. in Computational Intelligence from Indian Institute of Technology, Roorkee, India. Since April 2022, he has been associated with Shobhit Deemed University, Meerut as a Professor and Head of CSE and Dean, School of Engineering and Technology. Earlier he worked with Amity University, Rajasthan as an Associate Professor and head, Department of Computer Science and Engineering as well as an Alternate Director—outcome. He has supervised four Ph.D. candidates, three in the submission process, nine M.Tech Dissertations, several MCA projects and B.Tech projects. He has over 100 research publications in international journals, chapters in edited books, conferences of repute (SCI and SCOPUS) and has over 1100+ Google citations to his research work with H-index of 19. He had been to Amity Institute of Higher Education Mauritius on deputation. He has availed travel grants from Microsoft Research India, CSIR, New Delhi and DST-New Delhi to visit Australia, Singapore, and Malaysia, respectively to present the research papers in the conference of high repute. He is a founding member of the International Conference on Soft Computing: Theories and Applications (SoCTA Series) and Congress on Advances in Materials Science and Engineering (CAMSE). He has edited six volumes of Conference Proceedings published by AISC series of Springer and two volumes with LNNS series of Springer (SCOPUS) Publication, two edited books with Asset Analytics, Springer and one authored book with Springer Briefs. He is an Associate Editor of International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM—SCOPUS Indexed). At present, he is editing a book with Springer, which is indexed in Scopus, and four thematic issues in SCI indexed Journals.

Chang Wook Ahn received Ph.D. degree from the Department of Information and Communications, Gwangju Institute of Science and Technology (GIST), South Korea, in 2005. From 2005 to 2007, he worked with the Samsung Advanced Institute of Technology, South Korea. From 2007 to 2008, he was a Research Professor with the GIST. From 2008 to 2016, he was an Assistant/Associate Professor with the Department of Computer Engineering, Sungkyunkwan University (SKKU), South Korea. He is currently working as a Professor with the School of Electrical Engineering and Computer Science, GIST. His research interests include genetic algorithms/programming, multiobjective optimization, neural networks, and quantum machine learning.

Kapil Kumar Goyal has obtained his BE in 1999 from NIT Kurukshetra and did M.Tech from NITTTR Chandigarh, India. He accomplished his doctorate from IIT Roorkee in 2012. Presently he is working as an Assistant Professor in Department of Industrial and Production Engineering at Dr. B R Ambedkar National Institute of Technology, Jalandhar. He has 18 years of teaching experience along with 4 years of public sector and industrial experience in NHPC, ISGEC Yamunanagar and Jindal Steel Strips, Hisar. He has more than 40 publications to his credit and his research articles appeared in the renowned publishers like Elsevier, Taylor & Francis, Springer and Inderscience. He has been awarded with the Best Paper Award in the international conference and he has also been awarded with the best employee award from MM University. His research interests include evolutionary algorithms, multiple objective optimization, neural networks, decision sciences, manufacturing systems and processes planning and optimization. He has delivered expert lectures in the various premier institutes including IITs and NITs. He is also serving as editor on the board of various international journals and also acting as reviewer in various reputed international journals.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gupta, H., Sreelakshmy, K., Verma, O. P., Sharma, T. K., Ahn, C. W., & Goyal, K. K. (2022). Synergetic fusion of Reinforcement Learning, Grey Wolf, and Archimedes optimization algorithms for efficient health emergency response via unmanned aerial vehicle. *Expert Systems*, e13224. https://doi.org/10.1111/exsy.13224