Pages: 54-73 / Vol. 19 / Issue 1 / Winter - Spring 2026

Journal of Aerospace Science and Technology

Journal Homepage: https://.jast.ias.ir

Original Research Article

Deep Reinforcement Learning-Based Optimization of an Innovative Fuzzy Inference System Structure for Quadrotor UAV Control

Fatemeh khajemohammadi*, Jafar Roshanian, and Ali Moltajaei Farid

Institute of Intelligent Control System, K. N. Toosi University of Technology, Tehran, Iran

ARTICLE INFO

Received: 25 May 2025
Revised: 26 August 2025
Accepted: 14 September 2025
Available online: 23 September 2025

KEYWORDS:

Fuzzy Inference System
Deep Reinforcement Learning
Parameter Optimization
Adaptive Control
Nonlinear Systems Artificial Intelligence
System Tuning

DOI:
https://doi.org/10.22034/JAST.2025.525997.1229

ABSTRACT

The optimization of membership function parameters in Fuzzy-PID controllers presents significant computational and design challenges, often relying on time-consuming and heuristic-based methods that lack generalizability. In this paper, we present a novel and efficient technique to enhance the performance of Fuzzy Inference Systems (FIS) by employing Deep Reinforcement Learning (DRL) for automatic and intelligent parameter tuning. Fuzzy logic offers significant advantages in handling nonlinear, uncertain, and highly dynamic systems due to its rule-based reasoning and flexibility. Meanwhile, DRL has demonstrated outstanding capabilities in learning optimal control strategies in complex environments through trial-and-error interactions and policy optimization. By combining the strengths of both paradigms, our proposed method enables the automatic adjustment of membership function parameters without manual tuning, leading to improved control accuracy and system adaptability. The proposed DRL-Fuzzy PID framework is specifically applied to the control of Unmanned Aerial Vehicles (UAVs), which are highly nonlinear systems that demand robust, adaptive, and precise trajectory tracking. Additionally, by strategically reducing the number of parameters involved in the optimization process, we significantly shorten the learning time and reduce computational overhead. Extensive UAV simulation studies confirm the robustness and efficiency of the proposed approach, showing substantial improvements in tracking performance and overall flight stability compared to traditional fuzzy controllers. This hybrid DRL-Fuzzy strategy offers a promising solution for advanced UAV control applications where adaptability and precision are critical.

1 INTRODUCTION

The Unmanned Aerial Vehicles (UAVs) are widely recognized for their versatility, which enable them to perform high-risk and labor-intensive tasks in both military and civilian sectors [1-2]. The integration of intelligent controllers powered by advanced learning algorithms has garnered significant attention in contemporary research. Nevertheless, the control of UAVs remains a formidable challenge, primarily due to their intricate and nonlinear dynamics. To address these complexities, a variety of control strategies have been explored, including Proportional-Integral-Derivative (PID) controllers, Linear Quadratic Regulator methods, H∞ control, feedback linearization, backstepping control, sliding mode control, and Model Predictive Control .

*. Corresponding Author's Email, fatemehkhmohamad@gmail.com

Print ISSN:1735-2134 Online ISSN: 2345-3648

Traditional control strategies, such as PID controllers, are widely used due to their simplicity and effectiveness in eliminating steady-state errors. However, they encounter limitations when dealing with system constraints, noise, disturbances, and multi-input multi-output configurations [3–8]. The Linear Quadratic Regulator is capable of handling multi-input multi-output systems but may fall short in addressing steady-state performance issues [9–10]. H∞ control demonstrates strong robustness in the presence of channel cross-coupling in multivariable systems, yet it depends heavily on an accurate and structured mathematical model [11]. Feedback linearization provides a systematic control framework, but its efficacy deteriorates when dealing with highly nonlinear dynamics or hard constraints [12–13]. Backstepping control is advantageous due to its recursive structure and ability to manage mismatched uncertainties, although its tuning complexity increases with the number of parameters involved [14–15]. Sliding Mode Control offers robustness against uncertainties and performs reliably under severe nonlinearity. While chattering effects—caused by the ideal switching function—can be mitigated using boundary layer techniques, saturation functions, or higher-order sliding modes, these solutions introduce additional design complexity and require careful parameter selection, which may limit practical deployment in time-sensitive UAV applications. Model Predictive Control excels in forecasting system trajectories and managing constraints, yet it demands precise dynamic models, and any inaccuracies in modeling can lead to significant performance degradation. Moreover, solving real-time optimization problems in nonlinear Model Predictive Control is computationally intensive and may slow down the response [16–17].

To address the aforementioned limitations of traditional control approaches, adaptive control techniques have been developed, offering dynamic adjustment to time-varying parameters and external perturbations [18–20]. Despite their adaptability, maintaining robustness in environments characterized by high uncertainty remains a significant challenge. Consequently, data-driven and learning-based control paradigms—particularly reinforcement learning (RL) —have attracted considerable attention in recent years, owing to their ability to operate without an explicit model and adapt to nonstationary dynamics [21–23]. Alongside RL, the integration of artificial intelligence tools such as fuzzy logic systems and neural networks has also been widely explored. While neural networks are capable of approximating complex nonlinear mappings, their reliance on extensive labeled datasets for training and generalization poses practical challenges [24–25].

Recent studies have focused on advanced control design for nonlinear and complex dynamical systems, including approaches for fault-tolerant attitude tracking, vibration suppression in flexible spacecraft, and fuzzy adaptive tracking in stochastic environments. These works highlight the effectiveness of intelligent and adaptive methods in dealing with system uncertainties and nonlinearities. Inspired by such advancements, our study extends these ideas to UAV control using a fuzzy deep Q-learning framework [26-28].

Fuzzy PID controllers have emerged as a promising alternative for handling nonlinearities and uncertainties in complex systems. Their interpretability, combined with their model-free reasoning capabilities, makes them an attractive choice in many UAV applications [29–33]. One of the primary obstacles in deploying fuzzy PID controllers, however, lies in the manual tuning of parameters, which is often labor-intensive and requires expert domain knowledge [34–36]. To mitigate this issue, researchers have advocated for reinforcement learning-based optimization to autonomously tune the parameters of fuzzy systems. Through continuous interaction with the environment, RL agents can iteratively optimize fuzzy control gains, enhancing overall system adaptability and performance. Various hybrid approaches, including fuzzy actor-critic methods [37], fuzzy Q-learning, Takagi–Sugeno–Kang fuzzy inference systems [38], and policy gradient reinforcement learning algorithms [39], have been proposed to facilitate online tuning of fuzzy PID controllers [40-41].

Despite the widespread application of reinforcement learning in control systems, its computational inefficiency—mainly due to the high-dimensional data processing—remains a persistent obstacle. This challenge becomes even more critical in time-sensitive and resource-limited UAV applications. To overcome this bottleneck, we introduce a novel deep reinforcement learning-based framework that reduces computational complexity by limiting the number of variables involved in the learning loop. Instead of using fixed mathematical relationships for designing membership functions, the proposed method

randomly samples parameter values within a predefined range. Through deep Q-learning, the agent evaluates each candidate configuration and selects the set of parameters that yields the maximum Q-value. These selected parameters are then used to construct an optimized Fuzzy Inference System . Simulation results confirm that the proposed DRL-based fuzzy tuning strategy achieves significant performance improvements with lower control effort.

Recent years (2022–2025) have witnessed significant research on hybrid fuzzy and reinforcement learning-based control strategies for UAVs and related nonlinear systems. Notable works include: Robust Fuzzy Q-Learning-Based Strictly Negative Imaginary Tracking Controllers for quadrotor stabilization [42]; a hierarchical Fuzzy Actor–Critic combined with MPC framework for pursuit–evasion control [43]; an adaptive fuzzy RL cooperative method for flock system control under dynamic uncertainties [44]; DDPG-based reinforcement learning for online tuning of PID gains in quadrotors [45]; deep reinforcement learning with corrective feedback for autonomous UAV landing on mobile platforms [46]; and RL-enhanced PID tuning for CrazyFlie 2.X quadrotors [47]. While these approaches demonstrate promising improvements in adaptability and performance, many rely on extensive computational resources, require expert-designed membership functions, or do not fully exploit interpretable fuzzy structures. In contrast, the present study introduces a novel framework that constrains the membership function parameters via an upper-triangular sampling structure and utilizes deep Q-learning to efficiently select the optimal fuzzy PID gains. This method simultaneously achieves reduced computational complexity, enhanced trajectory tracking, and robust control performance, distinguishing it from prior research

This paper is organized into six sections to provide a clear and systematic presentation of the work. Section 2 defines the problem and highlights the challenges in UAV control. Section 3 details the methodology, encompassing the fuzzy logic system (3.1) with its control applications and PID gain adjustment (3.1.1–3.1.2), the reinforcement learning framework (3.2), the neural network including a comprehensive description of the critic model (3.3–3.3.1), the dynamic model of the aerial robot (3.4), and the proposed fuzzy logic control algorithm enhanced by Deep Q-learning (3.5). Section 4 presents extensive simulation results, demonstrating the effectiveness and advantages of the proposed approach. Finally, Section 5 concludes the paper by summarizing the key findings and emphasizing the contributions of this study. This structured layout guides the reader smoothly from theoretical foundations to practical implementation and results, ensuring a coherent and comprehensive understanding of the proposed methodology.

2 PROBLEM DEFINITION

Achieving precise and robust control in nonlinear systems such as UAVs demands the design of well-tuned fuzzy inference systems . The effectiveness of an FIS heavily depends on both its rule base and the accurate selection of membership function parameters. Traditionally, these parameters are often determined through trial and error, which is time-consuming and inefficient.

In this study, we propose a novel approach in which the membership function parameters are optimized using a deep reinforcement learning algorithm, allowing the agent to autonomously discover the optimal parameter configuration that maximizes control performance. By reducing the dimensionality of the input space and limiting the number of candidate parameters entering the learning process, we mitigate the computational burden typically associated with reinforcement learning. As a result, the constructed FIS operates with fewer states, while maintaining or even enhancing control quality.

To clarify the implementation framework, the design procedure is explicitly described in subsequent sections. It is also important to note that the UAV considered in this work has a cross-shaped configuration (┼), and its control inputs are limited solely to rotor speed modulation. The full specifications of the UAV platform are detailed in Section 3.3.

3 METHODOLOGY

This study presents a hybrid control tuning strategy that combines fuzzy inference systems (FIS) with deep reinforcement learning to optimize UAV control gains. The process starts with an exploration phase. Membership function parameters are randomly generated within predefined logical bounds. These parameters define candidate PID gains ($\boldsymbol{K_p}$, $\boldsymbol{K_d}$ and $\boldsymbol{K_i}$), which are applied to the UAV model. System performance is measured using a cumulative tracking error–based reward. Q-values are then updated for each configuration. This stage

enables the efficient identification of high-performing gain sets without exhaustive searches or fixed analytical formulations.

After sufficient exploration data are obtained, the framework enters the exploitation phase. A neural network–based critic is introduced and trained on the state–action–reward history from the fuzzy phase. The critic predicts Q-values for newly generated parameter sets. The configuration with the highest predicted Q-value is chosen as optimal. These parameters are embedded into the K_p.FIS, K_d.FIS and K_i.FIS structures. The system then uses these structures for real-time adaptive gain adjustment in UAV control. The complete workflow, as illustrated in Figure 1, offers a structured methodology that balances exploration efficiency, learning stability, and control performance.

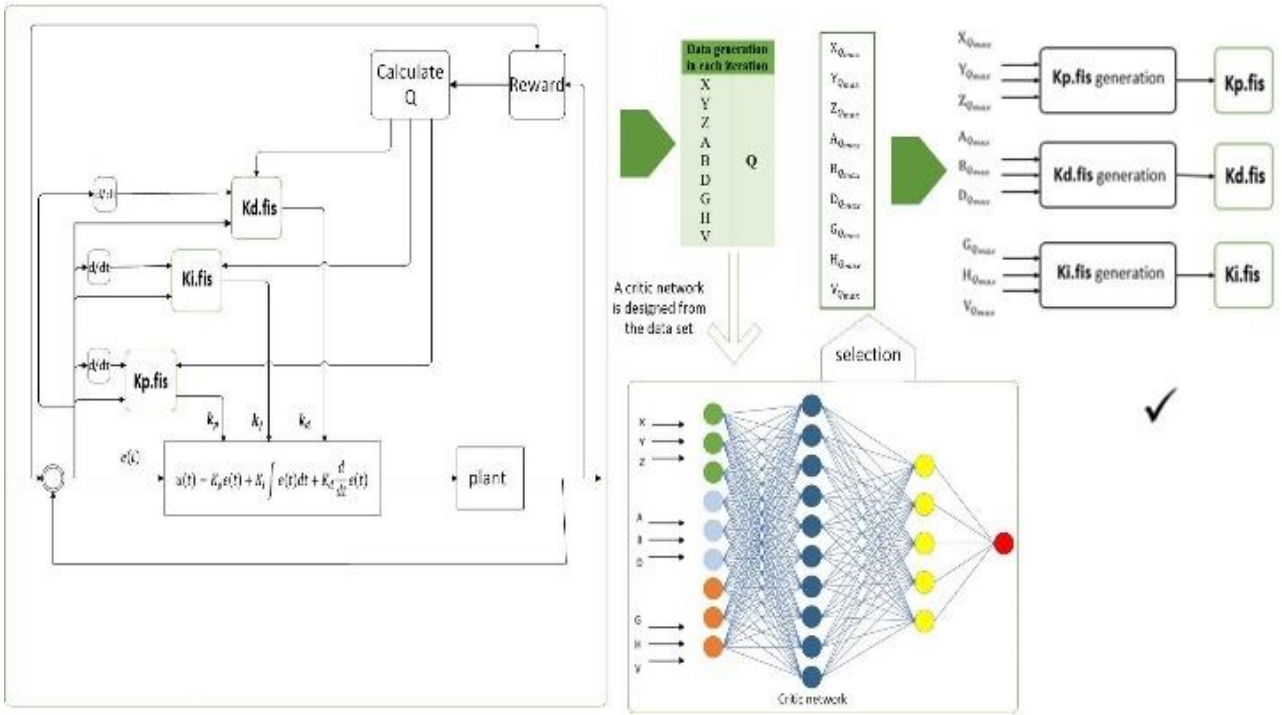


Fig. 1. Overall framework of the propose approach

3.1 Fuzzy logic

3.1.1 Control based on fuzzy logic

Fuzzy logic control has demonstrated considerable effectiveness in real-world applications, particularly in systems where precise mathematical modeling is difficult or impractical. This type of control relies primarily on qualitative and heuristic knowledge derived from human expertise, rather than detailed system equations. In a fuzzy control system, the available inputs often comprising sensor measurements, approximate system dynamics, and experiential knowledge are encoded into a set of linguistic rules formulated by domain experts. These rules describe the control behavior in the form of "if-then" statements and are implemented using fuzzy inference mechanisms.

Unlike traditional control strategies that typically require an accurate analytical model of the system, fuzzy control operates independently of such models. This model-free nature allows fuzzy controllers to handle uncertainty, imprecision, and partial knowledge more effectively. Therefore, fuzzy logic is a valuable tool for controlling systems that require human-like reasoning and adaptability, particularly in cases where conventional model-based methods are inadequate.

3.1.2 Gain adjustment of PID controller by fuzzy method

Figure 3 illustrates the structure of a control system that integrates a classical PID controller with a fuzzy-based gain tuner [48]. In the proposed approach, fuzzy inference systems (FIS) and a set of

heuristic rules are utilized to dynamically compute and adjust the PID gains. This configuration enables adaptive tuning of the controller parameters in response to changes in the system's behavior and environmental conditions, as demonstrated schematically in Figure 3.

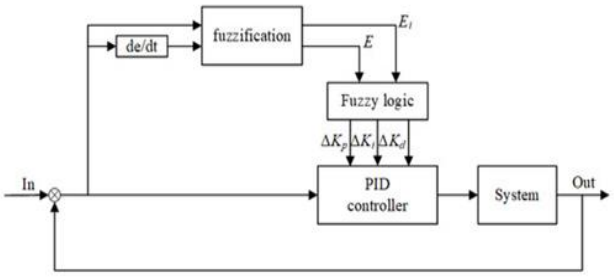


Fig. 3. Schematic of fuzzy PID algorithm [48]

3.2 Reinforcement learning

Reinforcement Learning is a dynamic area within machine learning that enables an autonomous agent to discover an optimal control strategy by interacting with its environment. The agent learns through trial and error, aiming to maximize a cumulative reward over time, without requiring an explicit model of the system dynamics.

Unlike traditional control approaches that are largely model-dependent, RL systems improve performance by continuously collecting and analyzing data generated through environmental interaction. One of the defining aspects of RL is its temporal structure learning unfolds over successive time steps.

One key characteristic distinguishing RL from other machine learning approaches is its explicit consideration of sequential decision-making over time. The RL agent interacts with the environment sequentially over time and at each time step t, this happens: the agent receives a state s(t) from the state space S, then according to a policy, an action A(t) It chooses from the action space (A). It is a mapping from the state s(t) (system state) to the required actions (control commands). In Figure 2, you can see a schematic of this Reinforcement learning algorithm [49].

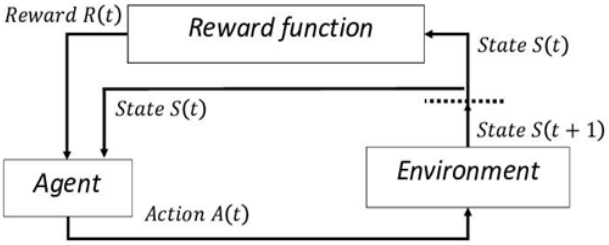


Fig. 2. Schematic of the Reinforcement Learning algorithm [49]

3.3 Neural Network

To elevate the performance and adaptability of the fuzzy reinforcement learning controller, a neural network is integrated into the architecture, transforming the approach into a deep reinforcement learning (DRL) framework. Neural networks possess powerful function approximation capabilities, making them ideal for modeling complex, nonlinear relationships in dynamic environments. In this work, the neural network is employed as a value function approximator that estimates the Q-values—expected cumulative rewards—for state-action pairs. This replaces the traditional Q-table, enabling the system to generalize across a continuous or high-dimensional state space. The inclusion of the neural network significantly enhances learning accuracy and scalability, allowing the controller to handle more intricate system behaviors without an exponential increase in computational complexity. As a result, the fuzzy logic controller gains the advantages of deep learning: improved precision, faster convergence, and the ability to operate effectively in environments with partial observability or noise. This crucial addition transitions the control paradigm from fuzzy reinforcement learning to deep fuzzy reinforcement learning, yielding a more robust and intelligent control system.

3.3.1 Critic Model Description

To approximate the Q-values corresponding to selected fuzzy PID parameters, a feedforward neural network was trained and employed as a function approximator, effectively replacing the Q-table used in conventional reinforcement learning. The input to the network consists of nine key parameters extracted from the fuzzy PID controller's action vector (i.e., the proportional, integral, and derivative gains from three main control loops), while the output is the estimated Q-value representing the quality of the corresponding action.

The neural network comprises two hidden layers with ten and five neurons respectively, using the Levenberg-Marquardt (trainlm) backpropagation algorithm for training. Figure 4 is a representation of this neural network.

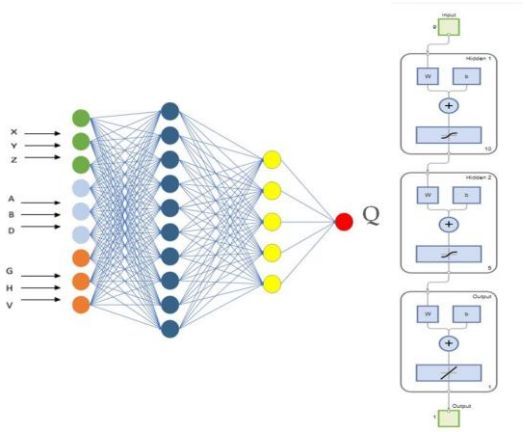


Fig. 4. schematic diagram of a designed neural network.(critic)

Input and target datasets were randomly split into training and testing subsets, with 60% of the data used for training. The loss function minimized during training was the mean squared error , and the training was configured with early stopping criteria, a gradient tolerance of 10^{-18} and a goal performance of 10^{-50}. The trained model was validated by comparing its Q-value predictions with the actual values from simulation data, yielding a low estimation error, as shown in Figure 5.

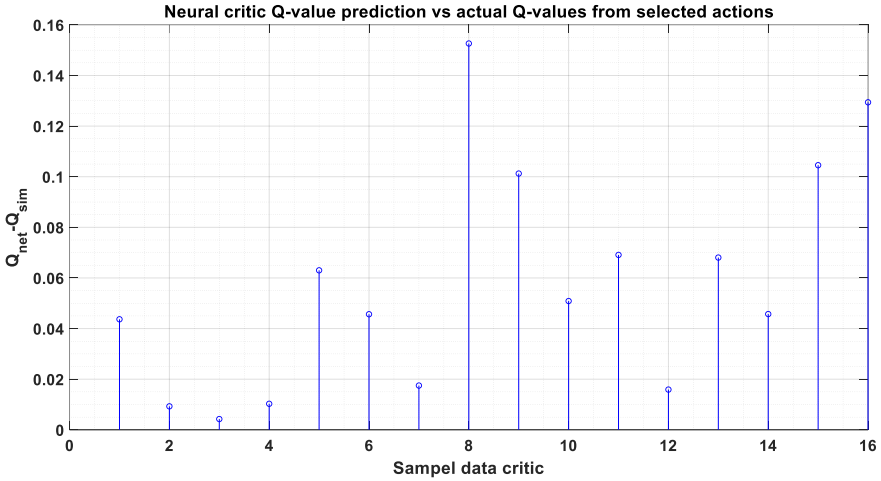


Fig 5. Neural critic Q-value prediction vs actual Q-values from selected actions

This neural critic model enables generalization across unseen states and actions, improving stability and learning efficiency in the fuzzy deep reinforcement learning framework.

3.4 Dynamic model of aerial robot

A typical quadrotor aerial robot features an cross-shaped structural frame equipped with four rotors, each mounted at one of the four corners. Each rotor is coupled with a propeller that generates thrust. The configuration and rotational directions of the propellers are designed to maintain balance and generate the desired motion. Figure 6 illustrates the rotational directions of the propellers, as well as the roll, pitch, and yaw movements of the quadrotor with respect to both the body-fixed and inertial reference frames [50].

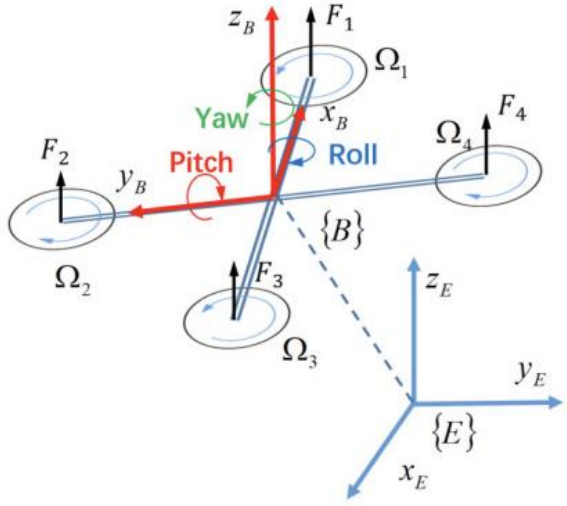


Fig 6. A view of roll, pitch and yaw movement in body coordinate system and inertial coordinate system [50]

According to the plant, a set of control signals is obtained as follows:

$$U = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \Omega_1^2 \\ \Omega_1^2 \\ \Omega_1^2 \\ \Omega_1^2 \end{bmatrix} \quad (1)$$

$$\Omega_R = (-\Omega_1 + \Omega_2 - \Omega_3 + \Omega_4) \quad (2)$$

Where u_i are the system inputs and Ω_i are the angular velocities of the four rotors, which by inverting the matrix of coefficients in the above relations, the angular velocity of the rotors according to the control inputs is obtained as follows:

$$\begin{bmatrix} \Omega_1^2 \\ \Omega_1^2 \\ \Omega_1^2 \\ \Omega_1^2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \quad (3)$$

Also, the equations of angular motion are obtained by the Newton-Euler method as follows: Where ϕ is the rotation angle around the x axis, θ is the pitch angle around the y axis and ψ is the yaw angle around the Z axis. Also, I_X , I_Y and I_Z are the moment of inertia of the body around 3 axes. I_R, b, d, and l are respectively the moment of inertia of rotor around the z axis, constant thrust coefficient and constant drag coefficient and the length of each moment arm. Also, Ω_R is the sum of the angular velocities of the rotors and is obtained as follows:

$$\Omega_R = (-\Omega_1 + \Omega_2 - \Omega_3 + \Omega_4) \quad (8)$$

Equations of translational motion are obtained by Newton-Euler method as follows:

$$\ddot{X} = sin(\theta)\, cos(\phi)\frac{b}{m}u_4 \quad (9)$$

$$\ddot{Y} = -sin(\phi)\frac{b}{m}u_4 \quad (10)$$

$$\ddot{Z} = -g + cos(\theta)\, cos(\phi)\frac{b}{m}u_4 \quad (11)$$

In the above relationship, (X, Y, Z) represents the position vector of the device in the inertial coordinate system. m is the mass of the device (including the equipment), g is the acceleration of the earth's gravity, b is equal to the thrust coefficient. Now we have the following values for other coefficients:

$$\begin{array}{ll} & J_x = 6.228e^{-3} \\ m = 0.52 & J_y = 6.225e^{-3} \\ L = o.232 & J_z = 1.121e^{-3} \\ D = 7.5e^{-7} & J_r = 6e^{-5} \end{array} \quad (12)$$

$$g = 9.81 \qquad b = 3.13e^{-5}$$

The desired path is a cylindrical path that has the following equations:

$$\begin{array}{c} X_d = sin(0.3t) \\ Y_d = 3\cos(0.3t) \\ Z_d = 0.1t \end{array} \quad (13)$$

3.5 Fuzzy logic control algorithm with the benefit of Deep Q-learning

In the proposed methodology, a hybrid fuzzy-neural framework is developed to enhance the parameter optimization process of fuzzy inference systems using Deep Q-learning (DQL). Unlike traditional fuzzy Q-learning approaches, where the Q-value is directly estimated based on fuzzy rules and actions, the current approach incorporates a neural critic network to evaluate and refine the fuzzy system's outputs.

$$\ddot{\phi} = \dot{\theta}\dot{\psi}\frac{I_Y - I_Z}{I_X} - \frac{I_R}{I_X}\dot{\theta}\Omega_R + \frac{lb}{I_X}u_1 \quad (5)$$

$$\ddot{\theta} = \dot{\phi}\dot{\psi}\frac{I_Z - I_X}{I_Y} + \frac{I_R}{I_Y}\dot{\phi}\Omega_R + \frac{lb}{I_Y}u_2 \quad (6)$$

$$\ddot{\psi} = \dot{\theta}\dot{\phi}\frac{I_X - I_Y}{I_Z} + \frac{d}{I_Z}u_3 \quad (7)$$

In the proposed fuzzy logic control algorithm, the membership functions are triangular, and the fuzzy rules for computing the control gains $\boldsymbol{K_p}$, $\boldsymbol{K_d}$ and $\boldsymbol{K_i}$are presented in Figures 9–11, with the rule tables positioned in the upper right corner of each Figure for clarity.

Initially, a set of membership function parameters is randomly generated within a predefined interval, forming upper-triangular parameter matrices for the input and output membership functions. These parameters are used to construct fuzzy inference systems for each control gain—namely K_p, K_d and K_i.

The corresponding control performance is evaluated by simulating the UAV dynamics, and a reward function is computed based on the average absolute position error in the x, y, and z directions.

The key novelty lies in the two-level evaluation of Q-values. First, the Q-values are estimated based on fuzzy system behavior via Bellman's recursive equation. Then, these Q-values, along with their associated state-action pairs, are used to train a neural critic network. After a certain number of iterations, the neural network replaces the traditional Q-table, providing more generalized and data-driven Q-value estimations.

By comparing the Q-values predicted by the neural critic with those obtained from the fuzzy Q-learning module, the action (i.e., set of membership function parameters) yielding the highest Q-value is selected. The corresponding random variables that produced these optimal actions are then stored and used to generate the final FIS files (i.e., K_p.FIS, K_d.FIS and K_i.FIS), which are subsequently applied to compute the fuzzy PID control gains.

This deep reinforcement learning-enhanced fuzzy control scheme offers a substantial improvement over previous Q-learning fuzzy control methods, both in terms of control performance and computational efficiency. By leveraging a neural critic, the system achieves better generalization in estimating the optimal actions, especially in complex and high-dimensional state-action spaces typical of UAV control problems.

In the following, the proposed algorithm is explained in detail. The developed control algorithm is based on a hybrid integration of fuzzy logic and deep reinforcement learning , aiming to optimize the fuzzy membership function parameters that generate the PID controller gains. The overall procedure consists of both an exploration phase using fuzzy logic and an exploitation phase enhanced by a critic neural network. The process of this algorithm is illustrated in Figure 7:

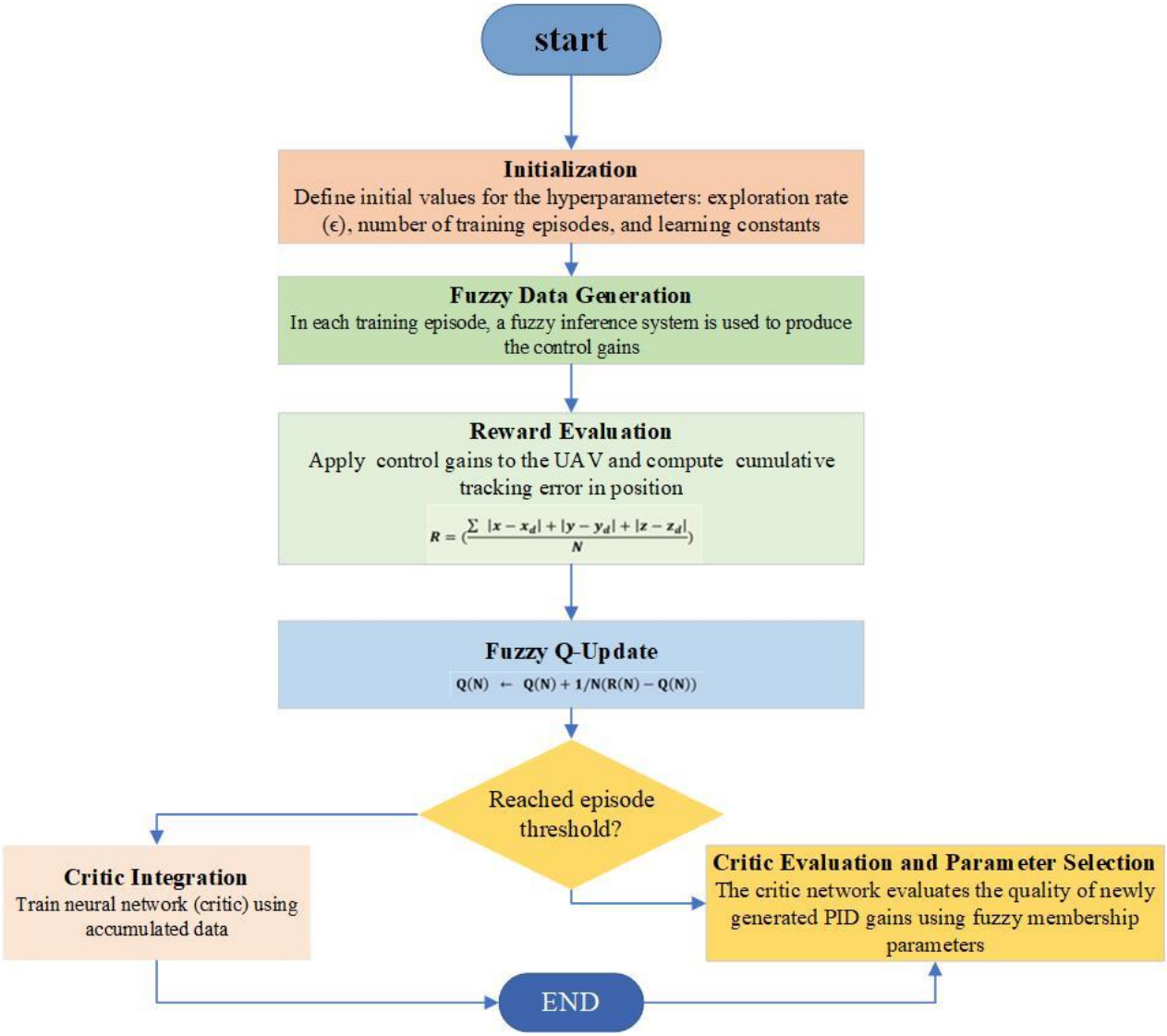


Fig 7. Flowchart of the Proposed Deep Q-Learning with Fuzzy Logic Algorithm

In the Figure 8 the performance of the intelligent fuzzy controller or the developed fuzzy Deep Q- learning are shown:

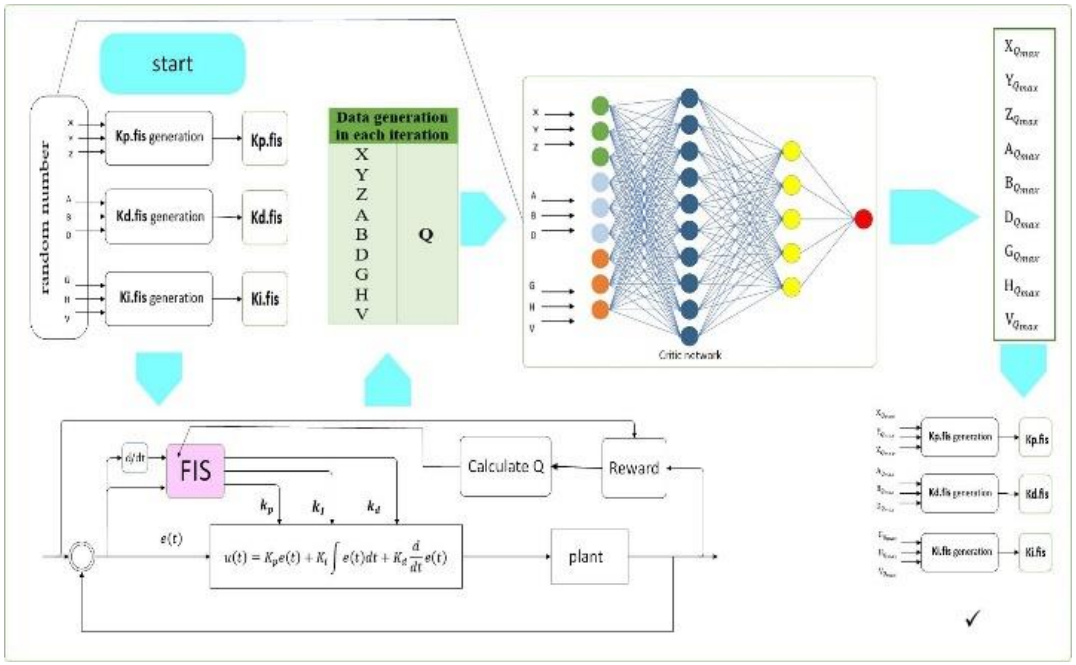


Fig 8. General flowchart of system performance and the role of Fuzzy Deep Q learning

Deep Reinforcement learning affects the fuzzy inference system. To be more precise, we have three fuzzy inference systems: K_p.FIS, K_d.FIS and K_i.FIS, which in the previous Figure shows its role. For better understanding of the designed algorithm, the process of producing these system FIS is shown in the Figures 9-11.

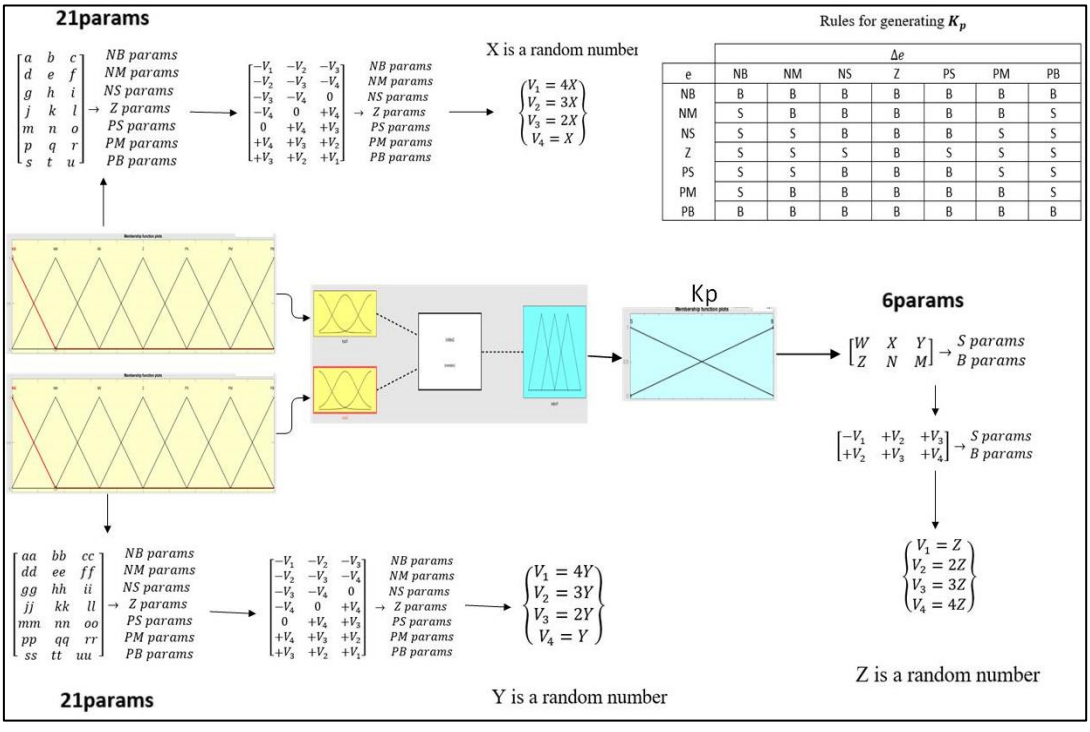


Fig 9. Flowchart of reducing the K_p.FIS membership parameters variable

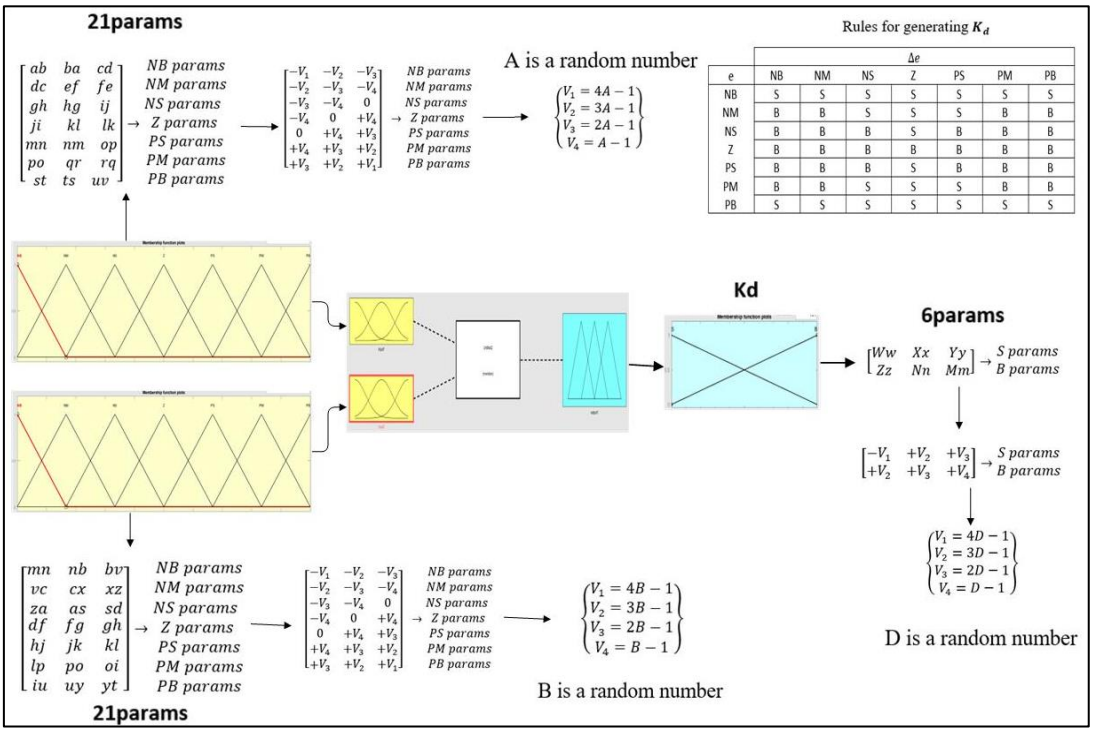


Fig 10. Flowchart of the variable reduction of K_d.FIS membership parameters

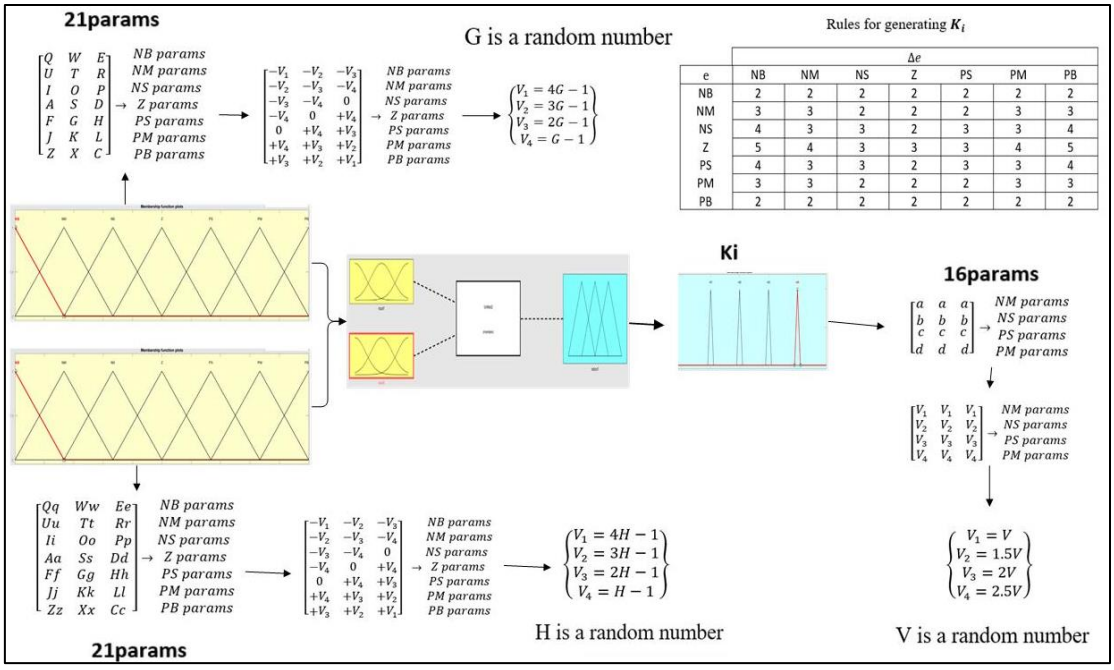


Fig 11. Flowchart of variable reduction of membership parameters K_i.FIS

we have two inputs and one output to produce each fuzzy inference system. Each input has seven membership functions. Each membership function has three starting, middle, and end points. the type of member function is trimf. The relationships between the parameters are in the form of a triangular upper matrix. According to the proposed algorithm, by defining four parameters

(V_1, V_2, V_3, V_4), twenty one parameter (for example, in Figure 8, the parameters a, b, c, d, e, f, g, h, I, j, k, l, m, n, o, p, q, r, s, t, u) values can be obtained.

It is possible to reduce the variables by generating 1 random number and finding the relationship between the four existing variables. So, the fuzzy inference system can be obtained for each value of K_p, K_d and K_i. We have shown all the cases mentioned in the above text in Figures 6-8.

According to the process illustrated in the flowchart (shown in Figure 8), a set of nine parameters (X, Y, Z, A, B, D, G, H, V) is initially generated through a randomized mechanism. These parameters define the structure of the fuzzy inference system , which is subsequently used to compute the corresponding control gains. Using the predefined reward function, a reward value is calculated for each generated FIS configuration. Thereafter, the Q-value for the corresponding state-action pair is updated based on the Bellman equation.

This procedure is repeated over multiple iterations to construct a comprehensive training dataset. During each iteration, a database entry consisting of the tuple (Q, X, Y, Z, A, B, D, G, H, V) is stored. After accumulating sufficient training data, a neural network—referred to as the critic network—is designed and trained to approximate the Q-value function. From this point onward, newly generated random parameter sets are evaluated by the neural network to estimate their associated action-value (Q) functions.

Ultimately, the parameter set that yields the highest Q-value is selected and stored. The corresponding fuzzy inference systems derived from this optimal configuration are then finalized and integrated into the fuzzy PID controller. This methodology enables an intelligent selection of control parameters by combining the exploratory power of fuzzy logic with the evaluative strength of deep reinforcement learning.

In general in the proposed hybrid fuzzy-deep Q-learning methodology, the PID gains (kp, kd and ki) are calculated through a two-phase procedure combining fuzzy logic-based exploration and deep Q-learning-based exploitation. Initially, candidate PID gains are generated by constructing fuzzy inference systems with triangular membership functions (trimf), where each input has seven membership functions and each membership function is defined by three points (start, middle, end). The parameters are organized in an upper-triangular structure, and a small set of random variables is used to generate all necessary FIS parameters. For each generated FIS, the UAV dynamics are simulated and a reward function based on the cumulative absolute position error in x, y, and z axes is calculated. The Q-values corresponding to each state-action pair are updated using the Bellman equation.

After multiple iterations and the accumulation of sufficient state-action-reward data, a neural critic network is trained to predict Q-values for newly generated random parameter sets. The parameter set producing the highest predicted Q-value is selected as optimal, and the corresponding fuzzy inference systems (K_p.FIS, K_d.FIS, K_i.FIS) are finalized. These optimized FISs are then used for real-time calculation of PID gains in the intelligent control of the quadrotor UAV. This approach combines the exploratory power of fuzzy logic with the evaluative capability of deep reinforcement learning, enabling automatic, adaptive, and optimal selection of control parameters.

To clarify the optimization results, the optimal parameters of the fuzzy membership functions are explicitly provided here. As described in the methodology, the random variables (X,Y,Z,A,B,D,G,H,V) define the structure of the fuzzy inference systems. After training with deep Q-learning, the optimal set of these variables is obtained. From this set, the start, middle, and end points of each triangular membership function are calculated, thereby defining the complete fuzzy inference systems for kp, kd and ki.

The optimized membership functions are shown in Figures 8–10, where each curve corresponds to one of the seven triangular functions for the respective input variables. These Figures illustrate how the optimal parameters generate the final fuzzy partitions, ensuring both smooth coverage of the input space and effective adaptation of the PID gains. Thus, the final fuzzy inference systems—optimized by the proposed hybrid fuzzy–deep Q-learning algorithm—provide the adaptive values of kp, kd and ki in real time during UAV control.

4 SIMULATION

Simulations are performed in MathWorks MATLAB R2019a on a desktop computer with an Intel® Core™ i7 processor and 8 GB of RAM. The proposed fuzzy–PID controller, enhanced via deep reinforcement learning–based tuning of membership function parameters, is compared against a baseline fuzzy–PID controller under identical conditions.

The results, demonstrating precise path tracking and rapid convergence to the desired trajectory, are summarized in Figures 12–15. These Figures illustrate the UAV's close adherence to the predefined flight path with minimal overshoot and steady-state error, highlighting the high accuracy and robustness of the control strategy. Key performance indicators such as attitude, position, control effort, and the three-dimensional trajectory are also depicted.

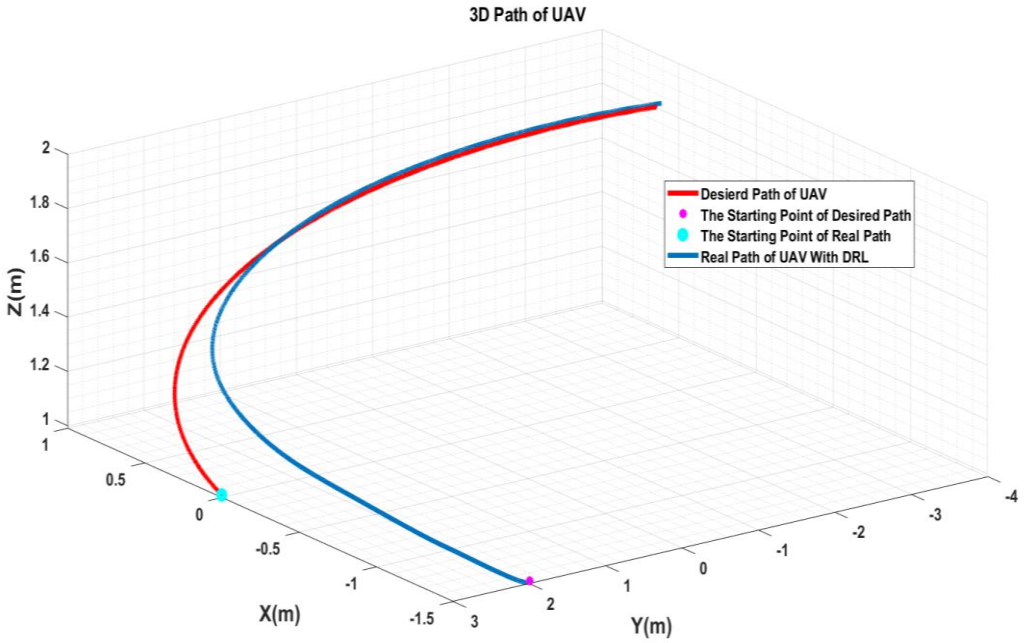


Fig 12. Three-Dimensional UAV Trajectory Relative to the Reference Path

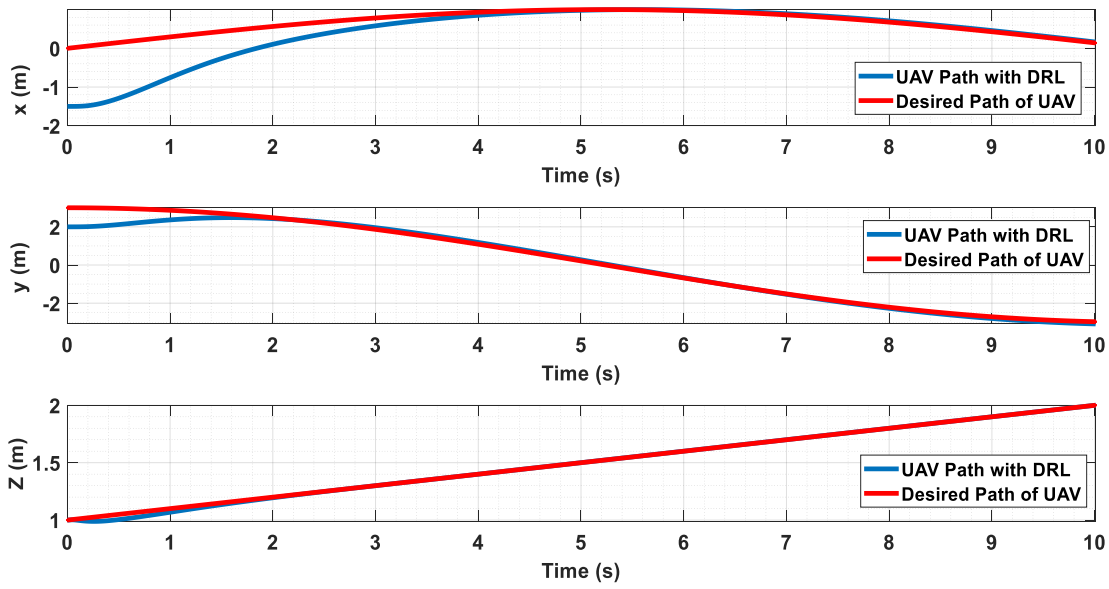


Fig 13. UAV Trajectory in x, y, and z over Time Relative to the Reference Path

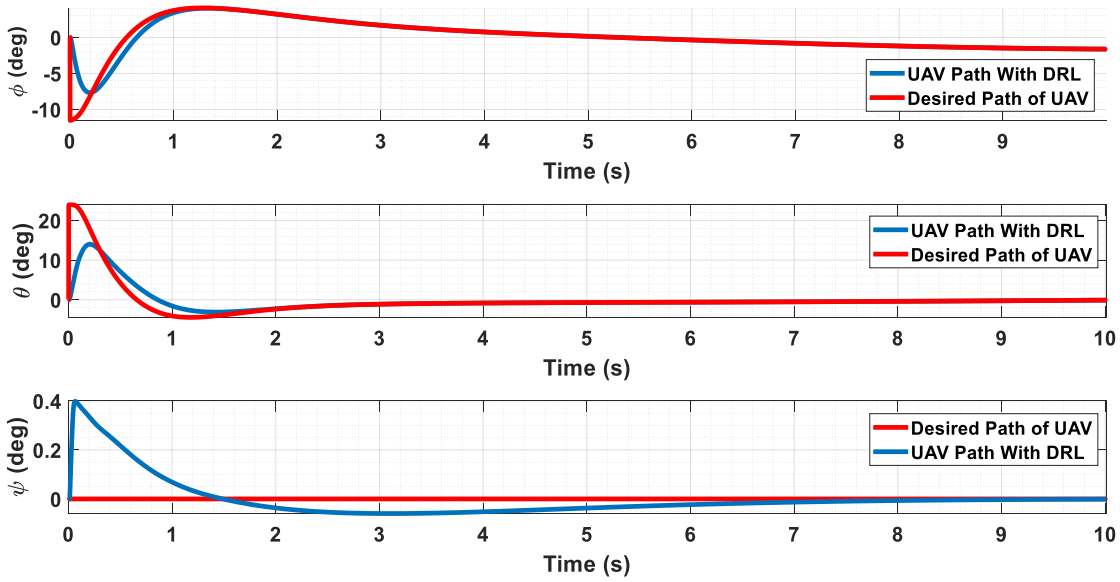


Fig 14. Angular Deviations of UAV Motion Relative to the Reference Path

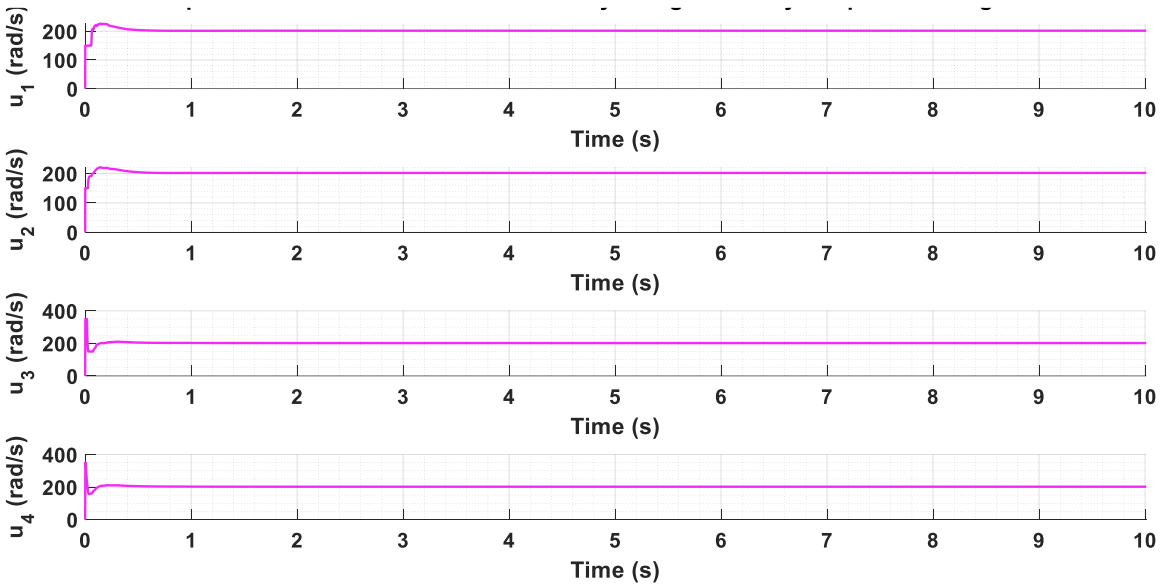


Fig 15. The amount of control effort required to follow the reference path by the UAV

Figures 16 through 19 illustrate a comparative analysis of the control effort exerted by the UAV under three distinct control strategies: the conventional fuzzy-PID controller, the fuzzy-PID controller enhanced with reinforcement learning, and the fuzzy-PID controller further optimized via deep reinforcement learning. The results clearly demonstrate a progressive reduction in control effort with the incorporation of learning-based enhancements. Notably, the application of deep reinforcement learning yields the most substantial improvement, indicating its effectiveness in optimizing control performance

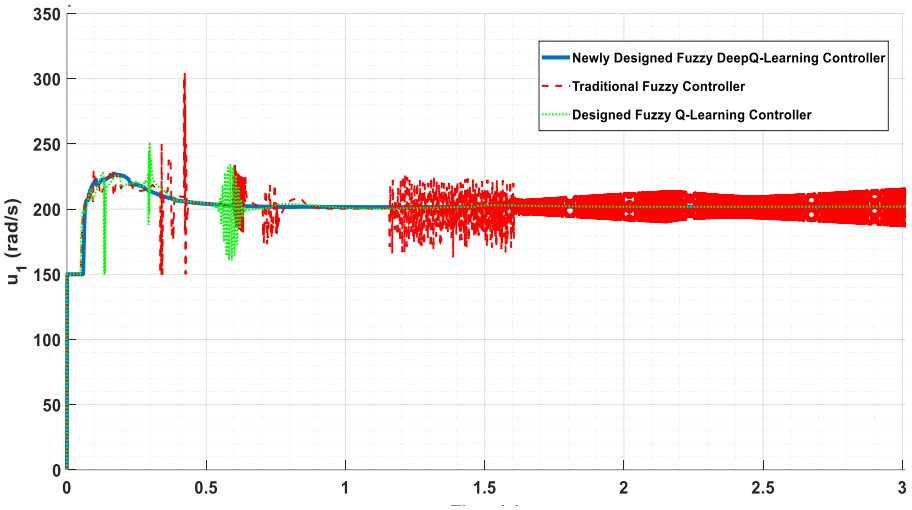


Fig 16. Comparison of Control Effort for Rotor 1 with and without Reinforcement Learning in the Fuzzy-PID Controller

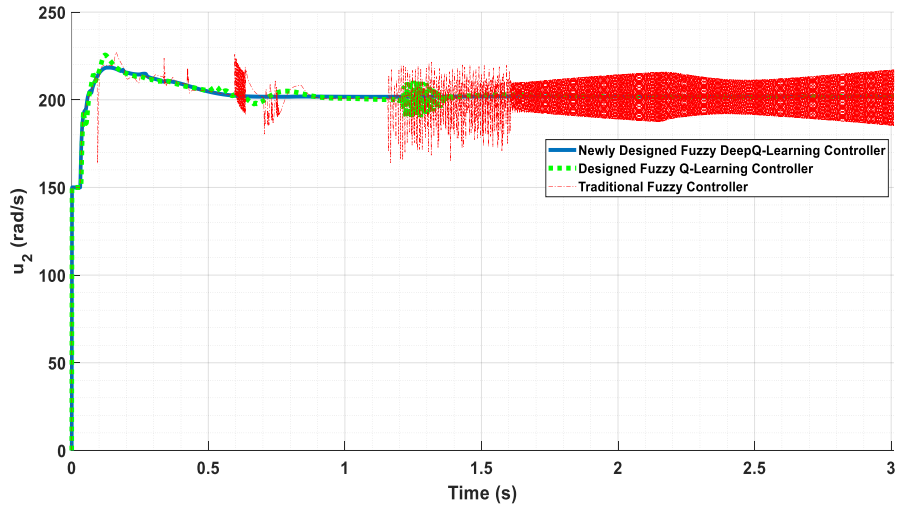


Fig 17. Comparison of Control Effort for Rotor 2 with and without Reinforcement Learning in the Fuzzy-PID Controller.

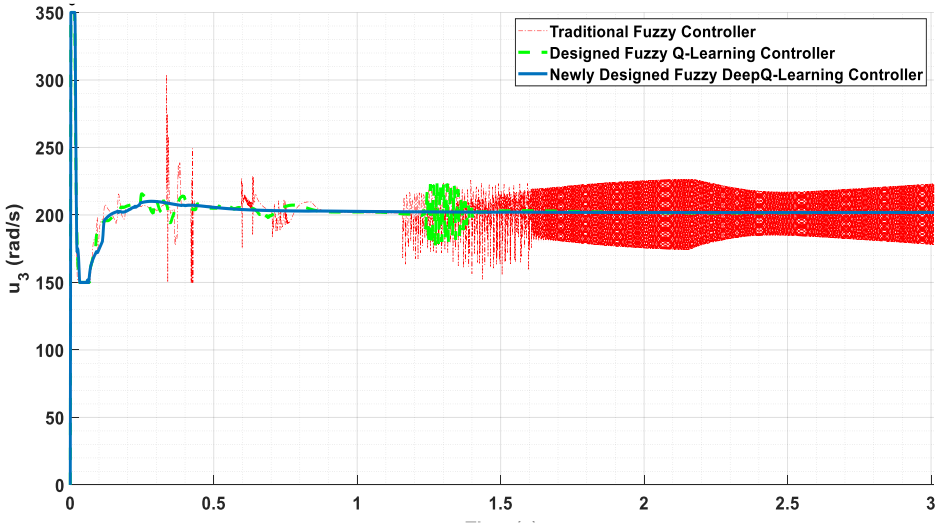


Fig 18. Comparison of Control Effort for Rotor 3 with and without Reinforcement Learning in the Fuzzy-PID Controller

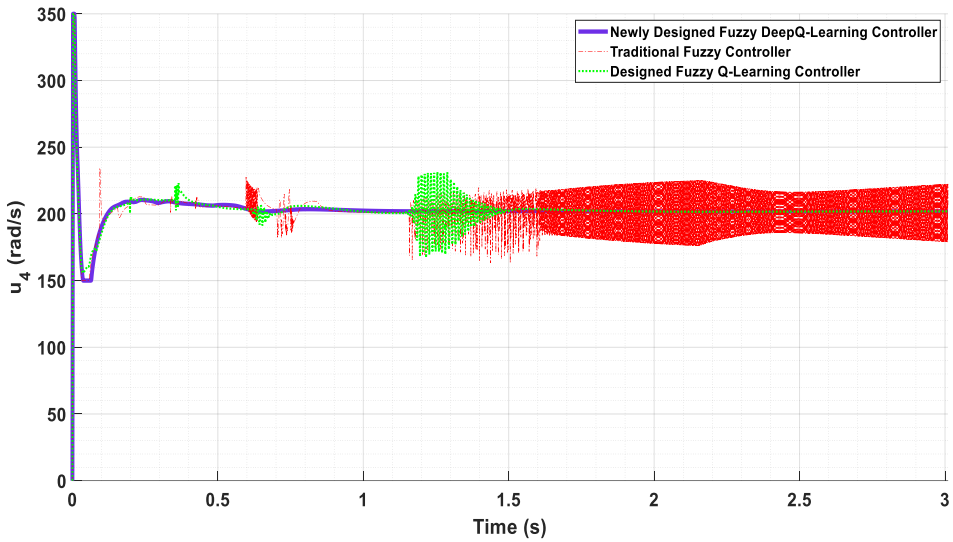


Fig 19. Comparison of Control Effort for Rotor 4 with and without Reinforcement Learning in the Fuzzy-PID Controller

As illustrated in Figures 16 to 19, the primary objective of the proposed approach is to enhance control effort. The results clearly show a significant improvement in control effort compared to the baseline.

To vividly demonstrate the impact of the learning process on the fuzzy inference system, Figures 20 and 21 present the 3D fuzzy rule surfaces for the membership function parameters. For illustration, only one fuzzy inference system—specifically Kd.FIS—is shown. Figure 20 depicts the scenario with randomly generated input membership function parameters, while Figure 21 illustrates the same system after applying deep reinforcement learning, where the parameters have been optimized. These visualizations not only highlight how the learning process refines and adjusts the membership function parameters, but also provide direct insight into the mapping from input variables to control gains. The numerical values of the input membership parameters are clearly observable in both Figures, offering a precise and intuitive understanding of the optimization achieved, thereby emphasizing the value and effectiveness of the proposed approach.

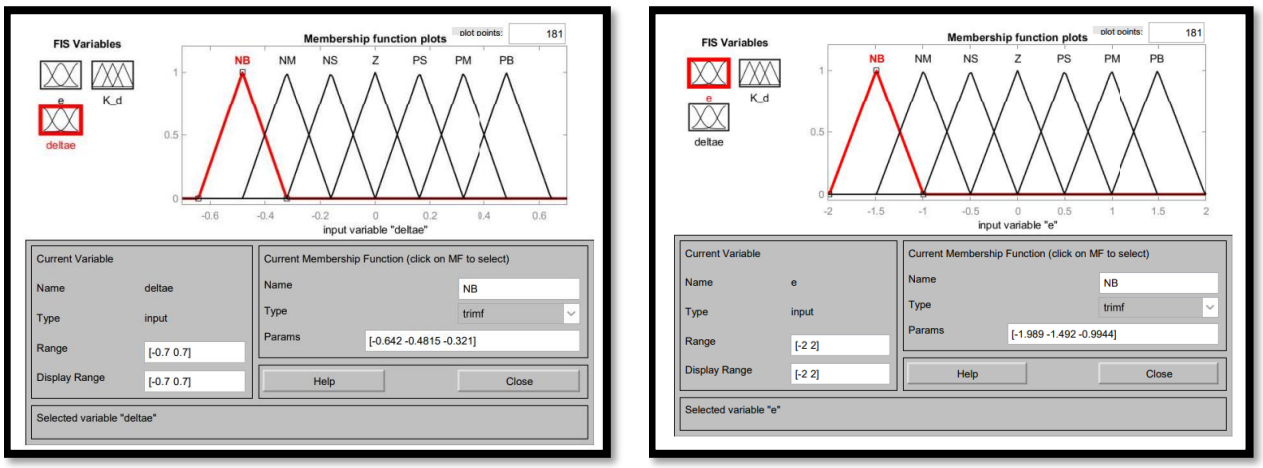


Fig 20-a. Fuzzy Membership function plots with Randomly Generated Input Membership Function Parameters

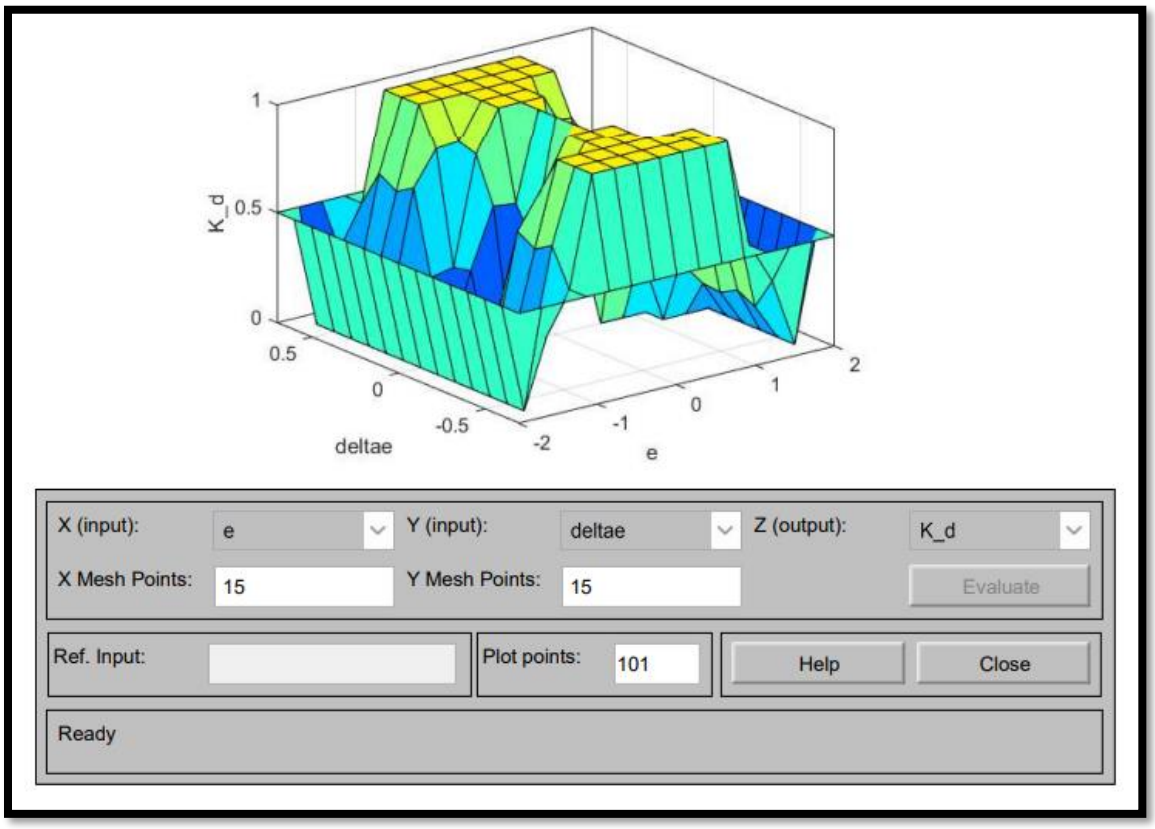


Fig 20-b. 3D Fuzzy Rule Surface with Randomly Generated Input Membership Function Parameters

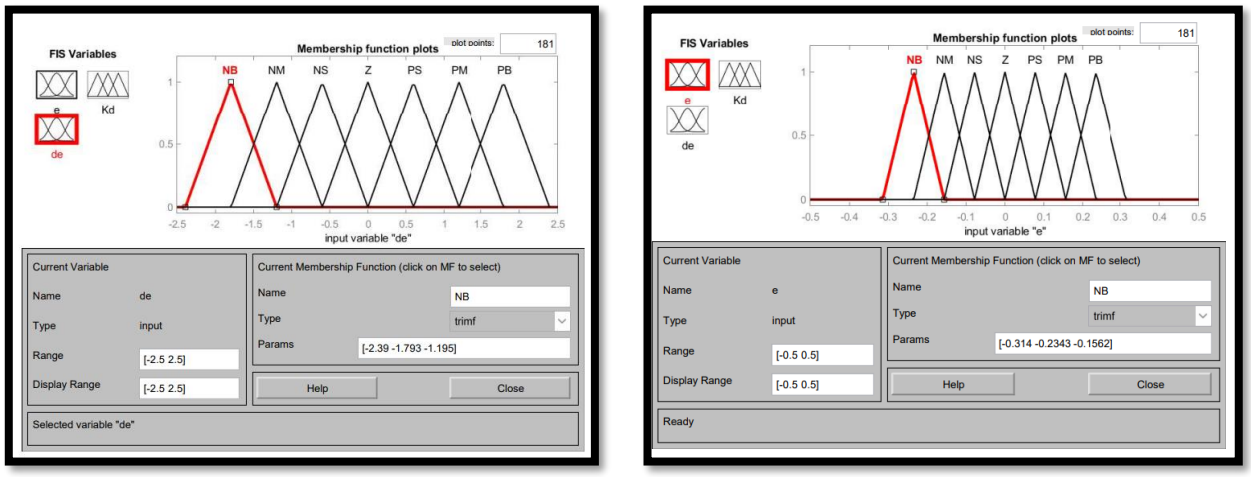


Fig 21-a. Fuzzy Membership function plots with Randomly Generated Input Membership Function Parameters

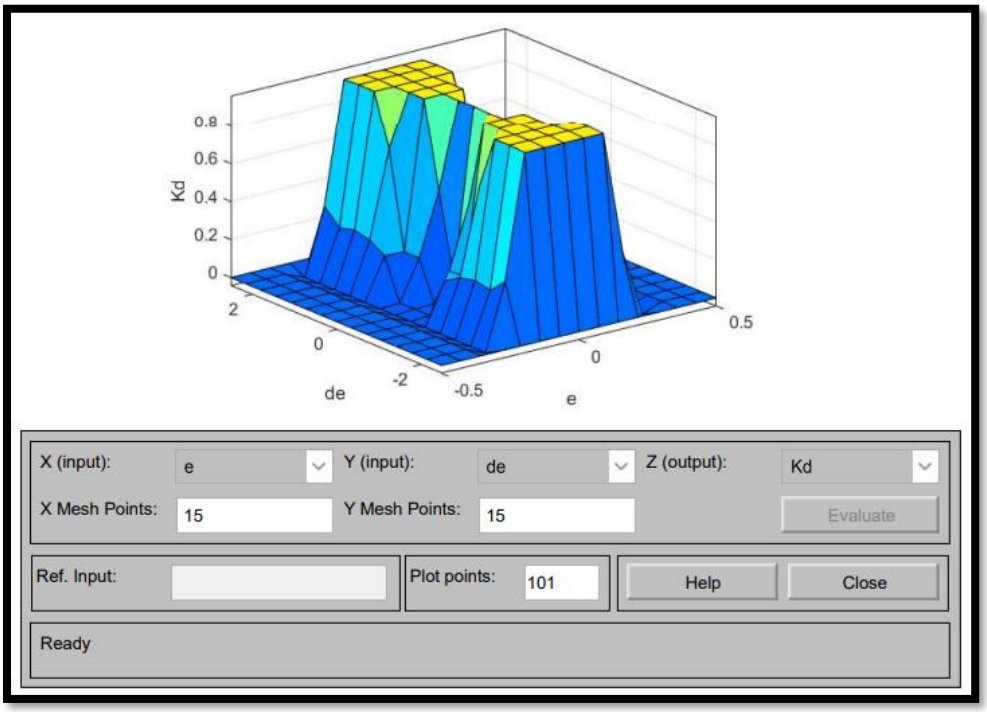


Fig 21-b. 3D Fuzzy Rule Surface with Optimized Membership Function Parameters via Deep Reinforcement Learning

5 CONCLUSION

This study has presented a novel method for determining the membership function parameters in fuzzy inference systems. This method significantly enhances the performance of fuzzy intelligent controllers by integrating deep reinforcement learning, which optimizes the parameter selection process. Traditionally, the parameters were determined through experimental methods, where researchers would iteratively adjust rules and consider all variables without reducing their scope. However, in our approach, we have reduced the number of variables involved, which leads to faster computations and fewer repetitions. As a result, the learning algorithm's speed improves significantly.

This novel approach not only accelerates the learning process but also enhances the control behavior of the fuzzy intelligent system. By optimizing the control gains, this method provides superior performance compared to previous methods, resulting in a substantial reduction in control effort while maintaining precise movement towards the desired reference path. Furthermore, the introduction of deep reinforcement learning in our algorithm adds a layer of complexity and flexibility, allowing the system to learn more efficiently and adapt to dynamic environments.

In the future, we could also involve modifying the reward functions, membership functions, and inference rules within the fuzzy system to further optimize the control strategy and adapt to a broader range of complex scenarios.

CONFLICTS OF INTEREST

No potential conflict of interest was reported by the authors.

REFERENCES

[1] A.D. Boursianis, M. S. Papadopoulou, P. Diamantoulakis, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, G. Karagiannidis, S. Wan, and S. K. Goudos, "Internet of things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review," Internet of Things, Vol. 18, 2022, p. 100187, https://doi.org/10.1016/j.iot.2020.100187.

[2] R. Gupta, A. Shukla, P. Mehta, P. Bhattacharya, S. Tanwar, S. Tyagi, and N. Kumar, "Vahak: A blockchain-based outdoor delivery scheme using uav for healthcare 4.0 services," *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE, 2020, pp. 255–260, https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9162738

[3] M. Mahfouz, and et al. "PID tuning approaches for UAVs unmanned aerial vehicles." *IOP Conference Series: Materials Science and Engineering*. vol. 1172. no. 1. IOP Publishing, https://doi.org/2021.10.1088/1757-899X/1172/1/012040

[4] K. Jinho, S. A. Wilkerson, and S. Andrew Gadsden, "Comparison of gradient methods for gain tuning of a pd controller applied on a UAV system." *Unmanned Systems Technology XVIII*. Vol. 9837. SPIE, 2016, https://doi.org/10.1117/12.2223329

[5] M. N. Duc, T. N. Trong, Y. S. Xuan, "The MAV system using PID control." *In Proceedings of the 2015 IEEE International Conference on Mechatronics and Automation (ICMA)*, Beijing, China, 2–5 August 2015; pp. 506–510, https://doi.org/10.1109/ICMA.2015.7237537.

[6] J. Li, and Y. Li. "Dynamic analysis and PID control for a UAV." *2011 IEEE International Conference on Mechatronics and Automation. IEEE*, 2011, https://doi.org/10.1109/INTECH.2017.8102445.

[7] I. M. Khairuddin, A. P. Majeed, A. Lim, M Jizat, J. A. Jaafar, A.A. "Modelling and PID Control of a UAV Aerial Robot." *In Advanced Materials Research; Trans Tech Publications Ltd.*: Bäch, Switzerland, 2014; Vol. 903, pp. 327–331, https://doi.org/10.1007/978-981-97-5337-6_18.

[8] A.R. Al Tahtawi, M. Yusuf, Low-cost quadrotor hardware design with PID control system as flight controller. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. no. 4, 2019, pp.1923–1930, http://doi.org/10.12928/telkomnika.v17i4.9529

[9] Sh. Khatoon, D. Gupta, L. Das, "PID and LQR control for a UAV: Modeling and simulation." *In Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Delhi, India, 24–27 September 2014; pp. 796–802, https:doi.org/10.1109/ICACCI.2014.6968232.

[10] H. Jafari, M. Zareh, J. Roshanian, A. Nikkhah, "An optimal guidance law applied to UAV using LQR method." *Trans. Jpn.Soc. Aeronaut. Space Sci*. vol. 53, 2010, pp. 32–39, https://doi.org/10.2322/TJSASS.53.32.

[11] X. C. Méndez Cubillos, de Souza, L.C.G. "Using of H-infinity control method in attitude control system of rigid-flexible satellite." *Math. Probl. Eng*. vol. 2009, Article ID 173145, 9 pages, https://doi.org/10.1155/2009/173145.

[12] C. I. Pop, E.H. Dulf, A. Mueller, "Robust feedback linearization control for reference tracking and disturbance rejection in nonlinear systems." *In Recent Advances in Robust Control-Novel Approaches and Design Methods*, Technical University of Cluj-Napoca, Romania, 2011, pp. 273–290.

[13] M. J. Kurtz and Henson, M.A. "Feedback linearizing control of discrete-time nonlinear systems with input constraints." *Int. J. Control*, vol. 70, 1998, pp. 603–616, https://doi.org/10.1109/SSST.1996.493465.

[14] C.W. Chung and Y. Chang, "Design of Adaptive Backstepping Controller for Systems with Mismatched Perturbations to Achieve Asymptotical Stability." Available online: https://digital-library.theiet.org/content/journals/10.1049/iet-cta_20050383 (accessed on 30 April 2021).

[15] M.A. Mohd Basri, K.A. Danapalasingam, A.R. Husain, "Design and optimization of backstepping controller for an underactuated autonomous UAV unmanned aerial vehicle." *Trans. FAMENA 2014*, vol. 38, 2014, pp, 27–44.

[16] G.V. Raffo, M.G. Ortega, F.R. Rubio, "MPC with nonlinear H∞ control for path tracking of a quad-rotor helicopter." *IFAC Proc.* vol. 41, Issue 2, 2008, 8564–8569, https://doi.org/10.3182/20080706-5-KR-1001.01448.

[17] P. Bouffard, "On-Board Model Predictive Control of a UAV Helicopter: Design, Implementation, and Experiments;" Technical ReportCalifornia Univ. Berkeley Dept. of Computer Sciences: Berkeley, CA, USA, 2012.

[18] K.J. Åström, B. Wittenmark, Adaptive Control; Courier Corporation: Chelmsford, MA, USA, 2013.

[19] H. Boudjedir, O. Bouhali, and N. Rizoug, Adaptive neural network control based on neural observer for UAV unmanned aerial vehicle. *Adv. Robot. 2014*, vol. 28, 1151–1164, https://doi.org/10.1080/01691864.2014.913498

[20] J.P. Hespanha, D. Liberzon, A.S. Morse, "Overcoming the limitations of adaptive control by means of logic-based switching." *Syst. Control Lett.*, vol. 49, 200349–65, https://doi.org/10.1016/S0167-6911(02)00342-0.

[21] R. S. Sutton, and A. G. Barto, *Reinforcement learning: An introduction*, 1.1, Cambridge: MIT press, 1998.

[22] L. P. Kaelbling, M. L. Littman and A. W. Moore, Reinforcement learning: A survey, *Journal of artificial intelligence research*, vol. 4, 1996, pp. 237-285, https://doi.org/10.1613/jair.301

[23] W. D. Smart and L. P. Kaelbling, "Effective reinforcement learning for mobile robots," *IEEE International Conference on Robotics and Automation*, Proceedings ICRA'02, vol. 4, 2002.

[24] J. Dunfied, M. Tarbouchi, and G. Labonte, "Neural network based control of a four rotor helicopter." *In Proceedings of the 2004 IEEE International Conference on Industrial Technology, 2004, IEEE ICIT'04, Hammamet, Tunisia, 8–10 December 2004*; Volume 3, pp. 1543–1548, https://doi.org/10.1109/ICIT.2004.1490796

[25] F. Khajeh Mohammadi, J. Roshanian, and A. Moltajaei Farid, "A novel approach to tune Fuzzy-PID membership function based on reinforcement learning." *Proceedings of the*

Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, *sage journal*, Vol. 240, Issue 6, 2021, https://doi.org/10.1177/09544100251404995.

[26] M, Zarourati, M, Mirshams, M. Taefi, "Integrated vibration suppression-based attitude tracking control of a flexible satellite in rapid imaging maneuver", *Aerospace Systems*, *Springer*, 2025, pp. 555–571 https://doi.org/10.1007/s42401-025-00388-4.

[27] M, Zarourati, M. Mirshams, and M. Taefi, "A Review on Underactuated Attitude Control of Reaction Wheel Actuated Satellites", *Journal of Space Science and Technology*, vol 18, pp. 83-119, 2025, https://doi.org/10.22034/jsst.2025.1528.

[28] M. Zarourati, M. Mirshams, M. Taefi, "Active underactuation fault-tolerant backstepping attitude tracking control of a satellite with interval error constraints", *Advanced control for applications*, vol. 6, 2024, https://doi.org/10.1002/adc2.215.

[29] G. Feng, "A survey on analysis and design of model-based fuzzy control systems." *IEEE Trans. Fuzzy Syst*. vol. 14, 2006, pp. 676–697, https:doi.org/10.1109/TFUZZ.2006.883415.

[30] X. Huang, A.L. Ralescu, H. Gao, H. Huang, "A survey on the application of fuzzy systems for underactuated systems." *Proc. Inst.Mech. Eng. Part I J. Syst. Control Eng*., vol. 233, 2019, pp. 217–244, https://doi.org/10.1177/0959651818791027.

[31] A. Rabhi, M. Chadli, C. Pégard, "Robust fuzzy control for stabilization of a UAV." I*n Proceedings of the 2011 15th International Conference on Advanced Robotics (ICAR)*, Tallinn, Estonia, 20–23 June 2011, pp. 471–475, https://doi.org/10.1109/ICAR.2011.6088629.

[32] N.M. Raharja, O. Wahyunggoro, A.I. Cahyadi, "Altitude control for UAV with mamdani fuzzy model." *In Proceedings of the 2015 International Conference on Science in Information Technology (ICSITech)*, Yogyakarta, Indonesia, 27–28 October 2015, pp. 309-314, https://doi.org/10.1109/ICSITech.2015.7407823

[33] X. Chen, D. Li, Z. Xu, Y. Bai, "Gain adaptive sliding mode controller based on interval type-II fuzzy neural network designed for attitude control for micro aircraft vehicle." *Int. J. Intell. Comput. Cybern*. vol. 7, 2014, pp. 209–226, https://doi.org/10.1108/IJICC-03-2012-0010.

[34] H.J. Lee, J.B. Park, and G. Chen, "Robust fuzzy control of nonlinear systems with parametric uncertainties." *IEEE Trans. Fuzzy Syst.*, vol.9, 2001, pp. 369–379, https://doi.org/10.1109/91.919258.

[35] C.L. Chen and M.H. Chang, "Optimal design of fuzzy sliding-mode control: A comparative study." *Fuzzy Sets Syst.*, vol. 93, 1998, pp. 37–48, https:doi.org/10.1016/S0165-0114(96)00221-7.

[36] H. Mo, G. Farid, "Nonlinear and adaptive intelligent control techniques for UAV uav–a survey. *Asian J. Control*, vol. 21, 2019, pp. 989–1008, https://doi.org/10.1002/asjc.1758.

[37] Papa, N., et al. "Autotuning of fuzzy inference system with RL." 2007 American Control Conference, New York City, USA, July 11-13, 2007, https://doi.org/10.1109/ACC.2007.4282156.

[38] Dai, X., et al, "An Approach to Fuzzy Controller Based on Reinforcement Learning for Autonomous Vehicle Control.", *IEEE*, Vol. 6, No. 3, September 2005. https://doi.org/10.1109/TITS.2005.853698

[39] V. Tavakol Aghaei, et al, "Tuning Scaling Factors of Fuzzy Logic Controllers via Reinforcement Learning Policy Gradient Algorithms.", *ICMRE 2017: Proceedings of the 3rd International Conference on Mechatronics and Robotics Engineering*, February 2017, Paris, France, pp. 146–151, http://dx.doi.org/10.1145/3068796.3068827.

[40] P., Kofinas, et al. "Fuzzy Q-Learning Agent for Online Tuning of PID Controller for DC Motor Speed Control.", *2018 MDPI*, vol. 11, Issue 10, 2018, https://doi.org/10.3390/a11100148.

[41] D. Park, et al, "Online Tuning of PID Controller Using a Multilayer Fuzzy Neural Network Design for Quadcopter Attitude Tracking Control." *Hypothesis and Theory article*, vol. 14, 2021, https://doi.org/10.3389/fnbot.2020.619350

[42] V. Tran, "Robust Fuzzy Q-Learning-Based Strictly Negative Imaginary Tracking Controllers for the Uncertain Quadrotor Systems." *IEEE*, vol. 53, Issue 8, 2023, https://doi.org/10.1109/TCYB.2022.3175366.

[43] P. Hu, "A Hierarchical Control Algorithm for a Pursuit–Evasion Game Based on Fuzzy Actor–Critic Learning and Model Predictive Control.", *MDPI*, vol. 9, no. 3, *Drones* 2025, 184; https://doi.org/10.3390/drones9030184.

[44] S. Qu, "An Adaptive Fuzzy Reinforcement Learning Cooperative Approach for the Autonomous Control of Flock Systems.", *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2022, https://doi.org/10.1109/ICRA48506.2021.9561204.

[45] S. Sönmez, L. Montecchio, S. Martini, et al, "Reinforcement Learning Based Prediction of PID Controller Gains for Quadrotor UAVs", *MDPI*, vol. 9, no. 8, Dornes 2025, https://doi.org/10.3390/drones9080581.

[46] Wu, L., Ch. Wang, P. Zhang, and Ch. Wei "Deep Reinforcement Learning with Corrective Feedback for Autonomous UAV Landing on a Mobile Platform," *MDPI* vol. 6, Issue 9, 238, Drones 2022, https://doi.org/10.3390/drones6090238.

[47] A. Javeed, V. López Jiménez, "Reinforcement Learning-Based Control of CrazyFlie 2.X Quadrotor", *arXiv:2306.0395*, 2023, https://doi.org/10.48550/arXiv.2306.03951.

[48] Zh. Fang Zhu, Y. Liu, Y. Ling He, W. Wu, H. Wang, Ch. Huang, B. Ye, "Fuzzy PID Control of the

Three-Degree-of-Freedom Parallel Mechanism Based on Genetic Algorithm," *Appl. Sci.* vol. 12, no. 21 2022, https://doi.org/10.3390/app122111128.

[49] S. Troia, et al. "On Deep Reinforcement Learning for Traffic Engineering in SD-WAN", *IEEE*, vol. 39, Issu 7, December 2020, pp. 2198-2212, http://dx.doi.org/10.1109/JSAC.2020.3041385.

[50] R, Ma, J. Han, and L. Ding, "Finite-time trajectory tracking control of UAV UAV via adaptive RBF neural network with lumped uncertainties", *Mathematical Biosciences and Engineering*, vol. 20, Issue 2, 2023, pp. 1841-1855, https://doi.org/10.3934/mbe.2023084.

COPYRIGHTS

Authors retain the copyright and full publishing rights.
Published by Iranian Aerospace Society. This article is an open access article licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0).

HOW TO SITE THIS ATRICLE

F. Author, S. Author, and T. Author, " Manuscript Template for the Journal of Space Science and Technology," *Journal of Aerospace Science and Technology*, Vol. 19, Issue 1, 2026, pp. 54-73.

DOI: https:doi.org/doi.org/10.22034/jast.2025.525997.1229

URL: https://jast.ias.ir/article_230230.html