International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 3, June 2026, pp. 1175~1187
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i3.pp1175-1187

Residual reinforcement learning for disturbance-resilient control under modeling uncertainties

Abolanle Adetifa, Rexcharles Enyinna Donatus, Daniel Udekwe
Department of Aerospace Engineering, Faculty of Air Engineering, Air Force Institute of Technology, Kaduna, Nigeria

Article Info

Article history:

Received Mar 4, 2026
Revised Apr 2, 2026
Accepted Apr 26, 2026

Keywords:

Deep deterministic policy gradient
Disturbance rejection
Flight control
Pitch-rate tracking
Residual reinforcement learning

ABSTRACT

Modern control systems must operate reliably in the presence of modeling uncertainties and external disturbances, conditions under which conventional fixed-gain controllers often exhibit performance degradation. This paper proposes a residual reinforcement learning framework for disturbance-resilient pitch-rate control of an aircraft longitudinal model. A classical proportional-integral-derivative (PID) controller is employed as a stabilizing baseline, while a deep deterministic policy gradient (DDPG) agent learns a bounded residual control signal to compensate for unmodeled dynamics and external perturbations. To promote favorable transient behavior, the learning process incorporates transient-aware and reference-model-based reward shaping, while actuator constraints are enforced within the environment dynamics. Simulation results demonstrate that the proposed residual controller achieves a superior balance between response speed, overshoot, and tracking accuracy compared with both the standalone PID controller and a pure DDPG-based controller. In particular, the residual architecture significantly reduces overshoot and tracking error while preserving fast transient response and providing robust disturbance rejection under large pitching moment disturbances. These results indicate that residual reinforcement learning offers a practical and effective approach for enhancing robustness and performance in safety-critical flight control applications.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Daniel Udekwe
Department of Aerospace Engineering, Faculty of Air Engineering, Air Force Institute of Technology
Nigerian Air Force Base, Mando, Kaduna, Nigeria
Email: daudekwe@afit.edu

1. INTRODUCTION

The increasing complexity and uncertainty of modern control systems have exposed limitations in traditional model-based control strategies [1]. Classical approaches such as proportional-integral-derivative (PID) control, linear quadratic regulation (LQR), and model predictive control (MPC) perform reliably when system dynamics are accurately modeled and disturbances are predictable. In practice, however, many systems are nonlinear, time varying, and affected by modeling uncertainties or external disturbances [2], [3]. Under such conditions, conventional controllers may experience degraded robustness and limited adaptability [4].

Reinforcement learning (RL) has emerged as a promising alternative because it enables control policies to be learned directly from interaction with the environment. RL does not require an explicit model and can adapt to complex or partially unknown dynamics [5]. Despite this flexibility, real-world deployment remains challenging. Deep RL algorithms such as deep deterministic policy gradient (DDPG) and proximal policy optimization (PPO) often suffer from high sample complexity, unstable convergence, and unsafe

exploration [6]. These limitations restrict their applicability in safety-critical domains where stability and constraint satisfaction are essential.

To overcome these drawbacks, recent research has focused on hybrid control frameworks that combine RL with established controllers. Residual reinforcement learning (RRL) is a prominent example, in which the RL agent learns a corrective signal that augments a stabilizing baseline controller [7]. The residual action is typically constrained to preserve nominal stability while improving tracking or disturbance rejection. Experimental results in robotic systems have shown that residual policies can enhance performance while maintaining structured and interpretable control behavior [8]. Broader analyses also indicate that hybrid architectures often outperform standalone RL agents in real-time and safety-sensitive applications [9].

Additional contributions have further reinforced this direction by integrating RL with structure-preserving filters and theoretical safety guarantees. Piccinelli *et al.* [10] proposed a passive RL framework that ensures safe policy convergence using Lipschitz continuity constraints and optimal control guidance. In a similar effort, Kostelac *et al.* [11] introduced a Lipschitz-filtered RL scheme guided by MPC that balances adaptability and constraint adherence in nonlinear settings. These approaches emphasize the value of embedding prior control knowledge and constraint handling directly into the learning architecture.

Additional studies have strengthened this approach by embedding theoretical safety guarantees and structural constraints into the learning process. For example, passive and Lipschitz-constrained RL formulations have been proposed to promote safe policy convergence and bounded control updates [12]. Similarly, RL schemes guided by MPC incorporate prior control knowledge to balance adaptability with constraint adherence in nonlinear systems [13]. In disturbance-rich environments, recent contributions demonstrate that safety-aware RL can improve robustness and disturbance rejection when appropriate filtering and constraint mechanisms are integrated into the control loop [14], [15]. These developments support the view that learning-based controllers are most effective when combined with established control principles [16].

Motivated by this line of research, we propose a safe residual reinforcement learning framework for disturbance-resilient control. A conventional PID controller provides baseline stability, while a DDPG-trained residual policy compensates for external disturbances and modeling inaccuracies. Training is conducted in a constrained and disturbance-intensive environment to promote robust and safe behavior. The framework is evaluated using a longitudinal aircraft dynamics model, demonstrating improved tracking performance and disturbance rejection compared to both classical control and standalone RL approaches, while maintaining smooth and bounded control actions.

The remainder of this manuscript is organized as follows. Section 2 describes the proposed methodology, including the mathematical model of the aircraft longitudinal dynamics and the controller design procedure. Section 3 presents the simulation results and discussion, focusing on time-domain responses, pitch rate tracking performance, and disturbance rejection. Section 4 concludes the paper and outlines the main limitations of the study.

2. METHOD

This section presents the modeling of the aircraft longitudinal dynamics, followed by the derivation of the linearized model and a description of the proposed controllers.

2.1. Modeling of the longitudinal dynamics

The longitudinal motion of a fixed-wing aircraft describes its translational and rotational behavior in the vertical plane and is governed by the forward velocity, angle of attack, pitch angle, and pitch rate. In this study, the longitudinal dynamics of an aircraft are adopted as the plant model for controller development, following standard aerodynamic formulations reported in the literature [17].

The aerodynamic forces and moments acting on the aircraft can be expressed in terms of nondimensional aerodynamic coefficients obtained from wind-tunnel testing. These forces and moments are given by (1) and (2).

$$X = C_x\overline{q}S, Y = C_y\overline{q}S, Z = C_z\overline{q}S, \tag{1}$$

$$L = C_l\overline{q}Sb, M = C_m\overline{q}S\overline{c}, N = C_n\overline{q}Sb, \tag{2}$$

where $\bar{q}$ denotes the dynamic pressure, S is the wing reference area, b is the wingspan, and $\bar{c}$ is the mean aerodynamic chord. The coefficients C_x, C_y, C_z, C_l, C_m, and C_n represent the dimensionless aerodynamic force and moment coefficients.

Assuming symmetric flight conditions and neglecting lateral-directional coupling and thrust vectoring effects, the nonlinear longitudinal equations of motion of the aircraft can be written as [18]:

$$\dot{V} = = \frac{\overline{q}S\overline{c}}{2mV}\left[C_{x_q}(\alpha)\cos\alpha + C_{z_q}(\alpha)\sin\alpha\right] - g\sin(\theta-\alpha) + \frac{\overline{q}S}{m}[C_x(\alpha,\delta_e)\cos\alpha + C_z(\alpha,\delta_e)\sin\alpha] + \frac{T}{m}\cos\alpha \quad (3)$$

$$\dot{\alpha} = q\left[1+\frac{\overline{q}S\overline{c}}{2mV^2}\left(C_{z_q}(\alpha)\cos\alpha - C_{x_q}(\alpha)\sin\alpha\right)\right] + \frac{\overline{q}S}{mV}[C_z(\alpha,\delta_e)\cos\alpha - C_x(\alpha,\delta_e)\sin\alpha] + \frac{g}{V}\cos(\theta-\alpha) - \frac{T}{mV}\sin\alpha \quad (4)$$

$$\dot{\theta} = q \quad (5)$$

$$\dot{q} = \frac{\overline{q}S\overline{c}}{2I_yV}\left[\overline{c}C_{m_q}(\alpha) + \Delta C_{z_q}(\alpha)\right] + \frac{\overline{q}S\overline{c}}{I_y}[C_m(\alpha,\delta_e) + \Delta C_z(\alpha,\delta_e)] \quad (6)$$

Here, V denotes the airspeed, α is the angle of attack, θ is the pitch angle, q is the pitch rate, δ_e represents the elevator deflection, T is the thrust force, m is the aircraft mass, g is the gravitational acceleration, and I_y is the moment of inertia about the pitch axis. These nonlinear equations capture the coupled aerodynamic and inertial effects governing the longitudinal motion of the aircraft.

2.1.1. Linearized longitudinal model

For controller synthesis and stability analysis, the nonlinear longitudinal dynamics are linearized about a steady, wings-level trim condition using a first-order Taylor series expansion while neglecting higher-order terms [17]. Defining the state vector as shown in (7) and the control input as the elevator deflection δ_e, the linearized state-space representation is given by (8) and (9).

$$\boldsymbol{x} = [V \quad \alpha \quad \theta \quad q]^T \quad (7)$$

$$\begin{bmatrix}\dot{\Delta V}\\ \dot{\Delta\alpha}\\ \dot{\Delta\theta}\\ \dot{\Delta q}\end{bmatrix} = \begin{bmatrix}-0.022 & -0.002 & 0 & 3\times10^{-7}\\ -1.395 & -0.582 & 0 & 0.324\\ -9.828 & 0 & 0 & 0\\ -0.672 & 0.908 & 1 & -0.708\end{bmatrix}\begin{bmatrix}V\\ \alpha\\ \theta\\ q\end{bmatrix} + \begin{bmatrix}-1.139\\ -0.072\\ 0\\ -4.301\end{bmatrix}\delta_e \quad (8)$$

$$\begin{bmatrix}q\\ \theta\end{bmatrix} = \begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}V\\ \alpha\\ \theta\\ q\end{bmatrix} \quad (9)$$

Eigenvalue analysis of the resulting linearized system reveals the presence of unstable modes in the open-loop dynamics, indicating that the aircraft is not inherently stable in the longitudinal axis. This motivates the need for a control augmentation system capable of stabilizing the aircraft while ensuring accurate pitch-rate tracking and disturbance rejection, which is addressed in the subsequent sections.

2.1.2. Control surface actuation model

Accurate modeling of actuator dynamics and constraints is essential in-flight control system design to ensure realistic simulation and safe command generation. The aircraft's elevator control surface, denoted δ_e, is modeled with both dynamic response and nonlinear physical limits to capture real-world behavior as shown in Figure 1.

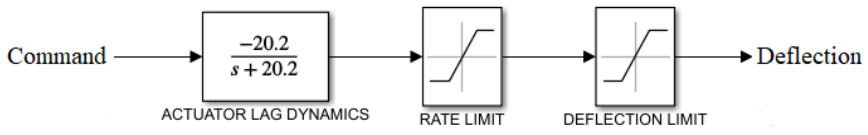


Figure 1. Elevator actuation model including lag dynamics, rate limiting and deflection saturation

The actuation model includes a first-order lag, a rate limiter, and a deflection limiter, applied sequentially to the commanded input. The elevator actuator dynamics are represented by a first-order lag system with the transfer function

$$\frac{1}{\tau s+1} \tag{10}$$

where $\tau = 49.5 \times 10^{-3} s$ is the actuator time constant. This models the internal response delay of the actuator to a control input. The output of the lag block is then passed through a rate limiter to enforce the maximum allowable rate of change in the elevator deflection:

$$\left|\frac{d\delta_e}{dt}\right| \leq 60^\circ/\text{s} \tag{11}$$

After the rate limit is applied, the signal is constrained by the elevator's deflection limits:

$$-25^\circ \leq \delta_e \leq +25^\circ \tag{12}$$

The sign convention used defines a positive δ_e as a trailing edge down deflection. This produces a negative pitching moment, causing the aircraft to pitch nose-down. These constraints ensure that the elevator deflection remains within the physical and mechanical bounds of the actuator, while the lag dynamics smooth the input response. This composite model plays an essential role in ensuring that generated control inputs are both feasible and reflective of realistic actuator behavior.

The complete control augmentation system (CAS) integrates the inner-loop pitch rate controller with the nonlinear aircraft dynamics and actuator model in a closed-loop configuration. As illustrated in Figure 2, the system is designed to regulate the pitch rate q about a specified trim condition q_{trim}. The error signal Δ_q is computed by subtracting the trim rate from the measured pitch rate and is then fed into the controller $C(s)$ which generates a desired elevator deflection increment $\Delta\delta_{ec}$.

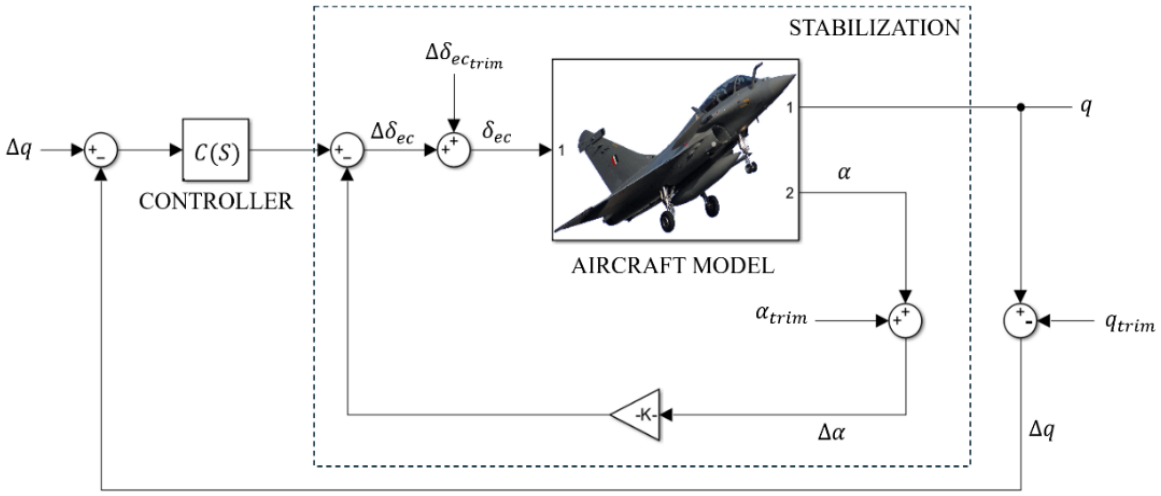


Figure 2. Block diagram of the control augmentation system for pitch rate stabilization

This control signal is summed with a trim bias $\Delta\delta_{ec,trim}$ to form the total elevator command δ_{ec}, which then passes through the actuator model described earlier. The aircraft's dynamic response is observed through both the pitch rate q and angle of attack α. The angle of attack is further fed back through a proportional gain K to provide an additional stabilizing influence on the controller input.

This augmented feedback loop enhances pitch axis stability and improves dynamic response by shaping the closed-loop behavior around the trimmed operating condition. Moreover, the actuator constraints embedded in the model ensure that all control inputs remain within physical limits. The architecture supports modular design and enables the integration of learning-based augmentation modules or disturbance rejection mechanisms within the existing control framework.

2.2. Controller design

This section presents the control architecture developed for stabilizing and regulating the longitudinal dynamics of the aircraft. Three controllers are considered: a classical PID controller, a DDPG-based controller, and a residual learning architecture that combines PID control with DDPG. The objective in all cases is to ensure stable pitch-rate regulation, accurate reference tracking, and robustness against external disturbances.

2.2.1. PID control design

The proportional-integral-derivative (PID) controller is adopted as a baseline control strategy due to its simplicity and widespread use in flight control applications [19]. In this study, the PID controller is designed to regulate the pitch rate by generating the elevator deflection command based on the tracking error between the desired and measured pitch rates [20].

Let the tracking error be defined as (13).

$$e(t) = q_{\text{ref}}(t) - q(t), \tag{13}$$

where $q_{ref}(t)$ denotes the reference pitch-rate command and $q(t)$ represents the measured pitch rate. The PID control law is expressed as (14).

$$u_{\text{PID}}(t) = K_p e(t) + K_i \int_0^t e(\tau)\, d\tau + K_d \frac{de(t)}{dt} \tag{14}$$

where K_p, K_i, and K_d denote the proportional, integral, and derivative gains, respectively. The controller gains are tuned using the MATLAB PID tuning tool to achieve a satisfactory trade-off between rise time, overshoot, and steady-state error.

Although the PID controller offers acceptable performance under nominal operating conditions, its fixed-gain structure limits adaptability when the aircraft is subjected to nonlinearities, modeling uncertainties, or external disturbances. These limitations motivate the use of learning-based control strategies.

2.2.2. DDPG-based deep reinforcement learning control

Deep deterministic policy gradient (DDPG) is a model-free, off-policy reinforcement learning algorithm designed for continuous action spaces, making it a suitable choice for aircraft control applications [19], [20]. It combines the actor-critic architecture with deterministic policy updates, where the actor network proposes control actions and the critic evaluates their expected performance [21]. By leveraging deep neural networks to approximate both the policy and value functions, DDPG enables the learning of complex, nonlinear control strategies directly from interaction with the environment [22], [23]. Its ability to handle high-dimensional state spaces and produce smooth, real-valued actions makes it particularly effective for systems such as fixed-wing aircraft, where continuous and precise actuation is required. The training framework of the DDPG agent showing the actor-critic networks, experience replay, and interaction with the environment is shown in Figure 3 according to Afzali *et al.* [24].

To formalize the control problem, the learning task is framed as a Markov decision process (MDP), characterized by a defined state space, action space, environment dynamics, and transition structure [25]. Within this framework, the agent observes the current state of the system and selects an appropriate control action to maximize cumulative future rewards. The action, corresponding to the elevator deflection command, is then applied to the aircraft model to influence its longitudinal motion.

a. State representation: The state vector is defined as (15):

$$s_t = [q(t), \theta(t), \alpha(t), V(t), e(t), \dot{e}(t), u(t-1)] \tag{15}$$

where q is the pitch rate, θ is the pitch angle, α is the angle of attack, V is the forward velocity, $e(t) = q_{ref}(t) - q(t)$ is the tracking error, $\dot{e}(t)$ is the error derivative, $u(t-1)$ is the previous control input.

b. Action space and control signal: The action generated by the agent corresponds to the elevator control command,

$$a_t = u(t) \tag{16}$$

which is applied directly to the longitudinal aircraft model. During training, exploration noise is added to encourage sufficient exploration:

$$a_t = \mu(s_t \mid \theta^{\mu}) + \mathcal{N}_t, \tag{17}$$

where $\mu(\cdot)$ denotes the actor network and $\mathcal{N}_t$ is Gaussian noise.

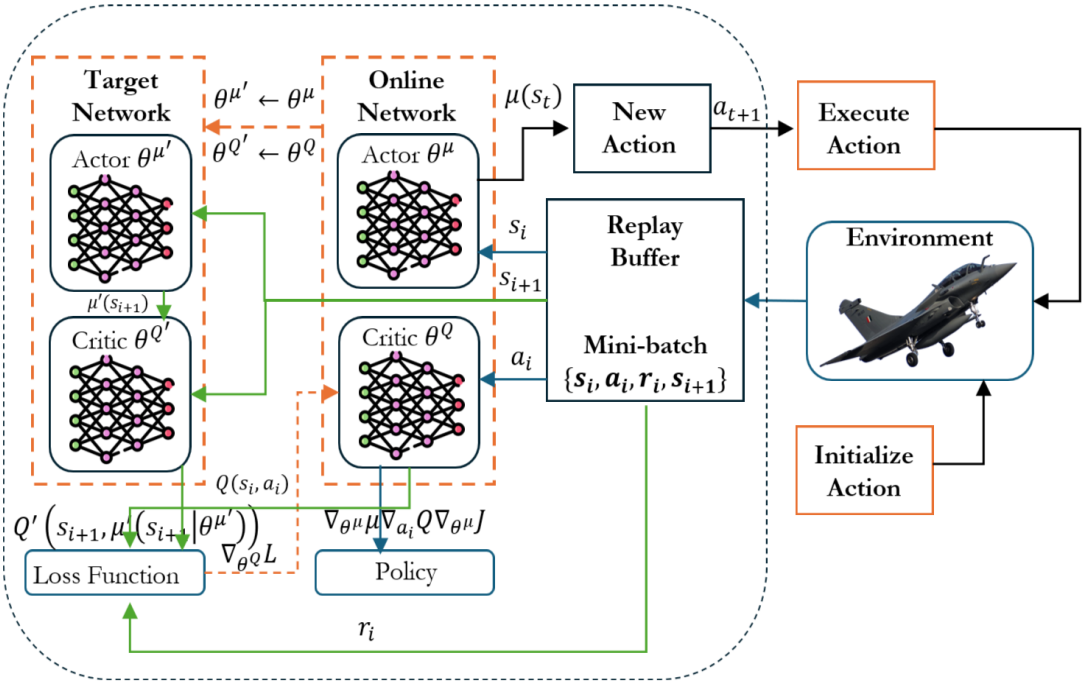


Figure 3. Architecture and training workflow of the DDPG based controller

c. Reward design and transient aware shaping: To mitigate a slow transient response, the reward function is designed to explicitly encode both transient and steady-state performance objectives. Instead of emphasizing only steady-state error minimization, the proposed reward incorporates penalties related to rise time, settling behavior, and excessive control activity.

The instantaneous reward is defined as (18):

$$R_t = -\left(w_1 e^2(t) + w_2 \mathrm{e}^2(t) + w_3 \Delta u^2(t) + w_4 |u(t)| + w_5 \mathbb{I}_{\text{out}}(t)\right) \tag{18}$$

where $e(t)$ is the tracking error, $\dot{e(t)}$ penalizes slow or oscillatory transient behavior, $\Delta u(t) = u(t) - u(t-1)$ discourages abrupt actuator motion, $|u(t)|$ limits excessive control magnitude, $I_{out}(t)$ is an indicator that penalizes time spent outside a predefined tolerance band around the reference.

d. Reference model-based reward shaping: To further guide the learning process toward desirable transient characteristics, a reference-model tracking formulation is introduced. A second-order reference model with desired natural frequency and damping ratio is defined as (19):

$$\ddot{q}_r + 2\zeta\omega_n \dot{q}_r + \omega_n^2 q_r = \omega_n^2 q_{\text{cmd}} \tag{19}$$

where q_r represents the ideal pitch-rate response. The tracking error used in the reward can then be redefined as (20):

$$e(t) = q(t) - q_r(t) \tag{20}$$

which explicitly encourages the learned policy to mimic a well-damped second-order response commonly used in flight control design. This reference-model formulation helps reconcile fast transient response with stability and smoothness.

e. Critic and actor updates: The critic network estimates the action–value function

$$Q(s_t, a_t \mid \theta^Q) \tag{21}$$

and is trained by minimizing the temporal-difference loss with target values given in (22) and (23).

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - Q(s_i, a_i \mid \theta^Q))^2 \tag{22}$$

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1})) \tag{23}$$

The actor parameters are updated via the deterministic policy gradient and soft target updates are applied as shown in (24) and (25)-(26).

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_{i}^{N} \nabla_a \, Q(s, a \mid \theta^Q)|_{a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s_i) \tag{24}$$

$$\theta^{Q'} \leftarrow (1-\tau)\theta^{Q'} + \tau\theta^Q \tag{25}$$

$$\theta^{\mu'} \leftarrow (1-\tau)\theta^{\mu'} + \tau\theta^\mu \tag{26}$$

To further improve training stability and transient response, a residual learning architecture is adopted. Instead of directly learning the full control signal, the DDPG agent learns a corrective residual on top of a baseline PID controller. The total control input is defined as:

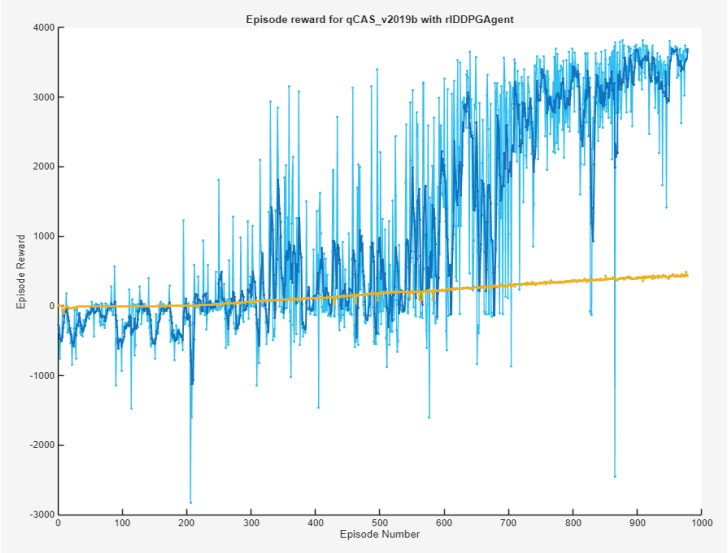
$$u(t) = u_{PID}(t) + u_{RL}(t) \tag{27}$$

where $u_{PID}(t)$ provides baseline stabilization and $u_{RL}(t)$ is the learned residual. In addition to reward shaping, physical and operational constraints are enforced directly within the environment dynamics rather than being handled solely through penalties. These include actuator saturation limits: $|\delta_e| \leq \delta_{e,max}$, actuator rate limits: $|\Delta\delta_e| \leq \dot{\delta}_e$ and safety envelopes on α, θ and q. When these limits are violated, the environment clips the control signal or terminates the episode. This separation of safety constraints from reward shaping improves learning stability and ensures physically realistic control behavior.

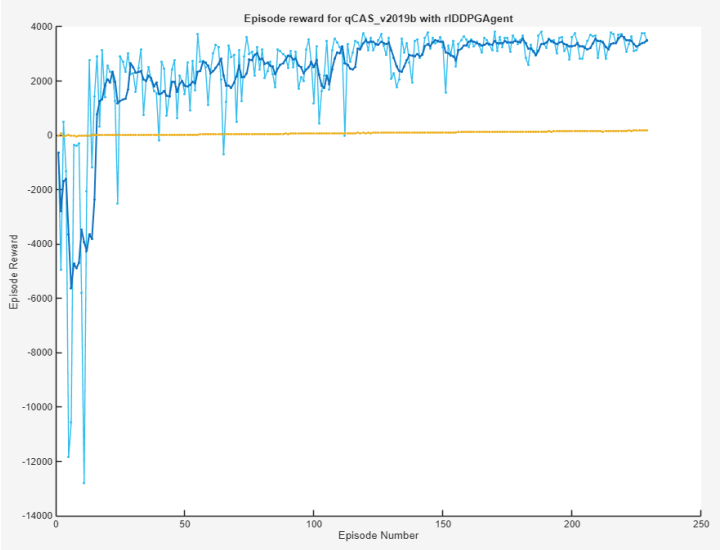
Overall, the proposed residual DDPG framework combines the stability of classical control with the adaptability of data-driven learning, resulting in improved transient performance, enhanced disturbance rejection, and better generalization across operating conditions compared with both standalone PID and pure DDPG controllers. The training parameters used for the reinforcement learning agents are summarized in Table 1. The learning curves shown in Figure 4 illustrate the training progression of the agents, where Figure 4(a) presents the learning behavior of the standalone DDPG-qCAS controller and Figure 4(b) shows the training performance of the Res-DDPG-qCAS agent. The residual agent demonstrates a more stable and consistent improvement in cumulative reward during training, indicating more efficient learning due to the stabilizing influence of the PID baseline.

Table 1. DDPG training parameters

Parameter	Value
Maximum episodes	1000
Episode duration	40s
Sample time	0.1s
Discount factor	0.99
Mini-batch size	64
Replay buffer size	1×10^6
Target smoothing factor	0.001
Actor learning rate	1×10^{-4}
Critic learning rate	1×10^{-3}
Initial noise std.dev	0.6
Noise decay rate	1×10^{-5}
Hidden layers	3
Neurons per layer	350
Activation function	ReLU



(a)



(b)

Figure 4. Training curves of deep reinforcement learning agents, (a) DDPG-qCAS and (b) Res-DDPG-qCAS

3. RESULTS AND DISCUSSION

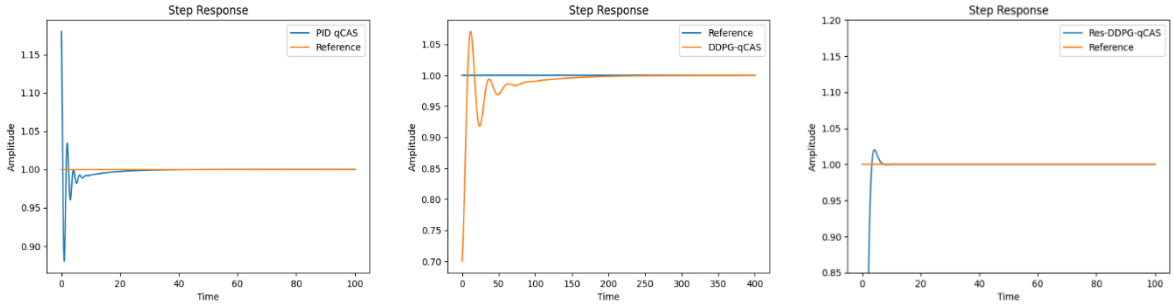
This section presents the results of the time response analysis, as well as the pitch rate tracking and disturbance rejection performance of both the classical controller and the deep reinforcement learning controller.

3.1. Time response

Figure 5 presents the step responses of the three controllers, while Table 2 summarizes their corresponding time-domain performance indices. As illustrated in Figure 5(a), the PID-qCAS exhibits the fastest rise time ($t_r = 0.2390\ s$)but suffers from a very large overshoot $(M_p = 18.9541\%)$and noticeable oscillatory transient behavior. These characteristics result in relatively high error metrics (MAE, ISE, and MSSE), indicating an aggressive yet poorly damped response.

As illustrated in Figure 5(b), the DDPG-qCAS significantly reduces overshoot $(M_p = 0.6580\%)$ and produces a smooth response; however, this improvement comes at the expense of very slow dynamics, with a rise time of $t_r = 9.4061\ s$and a settling time of $t_s = 17.7502\ s$. The prolonged transient response explains the large ISE value (0.8838), despite moderate MAE and MSSE values. In contrast, Figure 5(c) shows that the proposed Res-DDPG-qCAS achieves a more favorable balance between response speed and

accuracy. It maintains a fast rise time ($t_r = 0.4020\ s$), the shortest settling time ($t_s = 0.9825\ s$), and minimal overshoot ($M_p = 0.2731\%$). Furthermore, it achieves the lowest error metrics, with MAE $= 0.0043$, ISE $= 0.0212$, and MSSE $= 0.0061$, demonstrating superior transient performance and tracking accuracy compared with both PID-qCAS and DDPG-qCAS. These results confirm that residual learning effectively combines the fast transient response of PID control with the smooth and robust behavior of reinforcement learning.



(a) (b) (c)

Figure 5. Step response for (a) PID-qCAS, (b) DDPG-qCAS, and (c) Res-DDPG-qCAS controllers

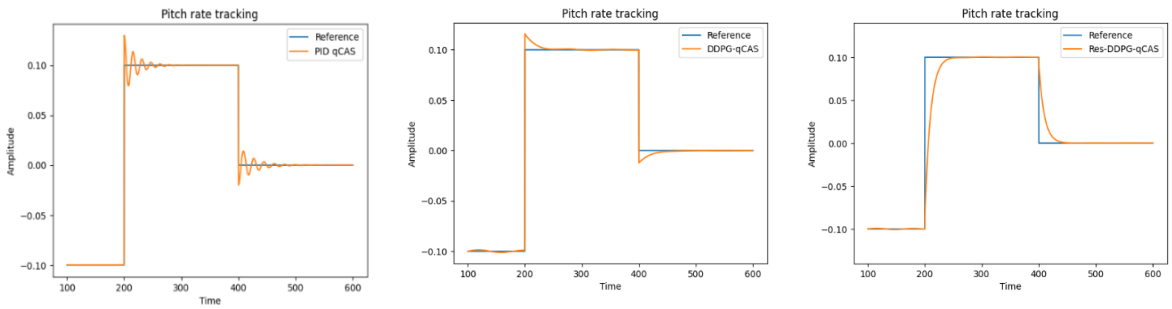
Table 2. Comparison of time-domain performance metrics for PID-qCAS, DDPG-qCAS, and Res-DDPG-qCAS Controllers

Method	t_r(s)	$t_s(s)$	$M_p(\%)$	MAE	ISE	$MSSE$
PID-qCAS	0.2390	1.4167	18.9541	0.1025	0.1106	0.0879
DDPG-qCAS	9.4061	17.7502	0.6580	0.0100	0.8838	0.0138
Res-DDPG-qCAS	0.4020	0.9825	0.2713	0.0043	0.0212	0.0061

3.2. Pitch rate command tracking

Figure 6 illustrates the pitch-rate command tracking performance of the PID-qCAS, DDPG-qCAS, and Res-DDPG-qCAS controllers under successive step changes in the reference signal. As shown in Figure 6(a), the PID-qCAS responds rapidly to command changes but exhibits noticeable overshoot and oscillations, particularly at the rising and falling edges of the command. These oscillations indicate limited damping and result in increased transient tracking error. In contrast, Figure 6(b) shows that the DDPG-qCAS produces a much smoother response with minimal oscillation; however, its convergence to the commanded value is relatively slow, leading to a sluggish transient during both the step-up and step-down transitions.

The proposed Res-DDPG-qCAS achieves a more favorable trade-off between speed and stability. As shown in Figure 6(c), it closely follows the reference with negligible overshoot and without oscillatory behavior, while maintaining a significantly faster response than the standalone DDPG controller. The smooth transitions and rapid settling demonstrate that the residual learning architecture effectively combines the fast response of the PID baseline with the damping and robustness provided by the learned residual policy, resulting in superior pitch-rate command tracking performance.



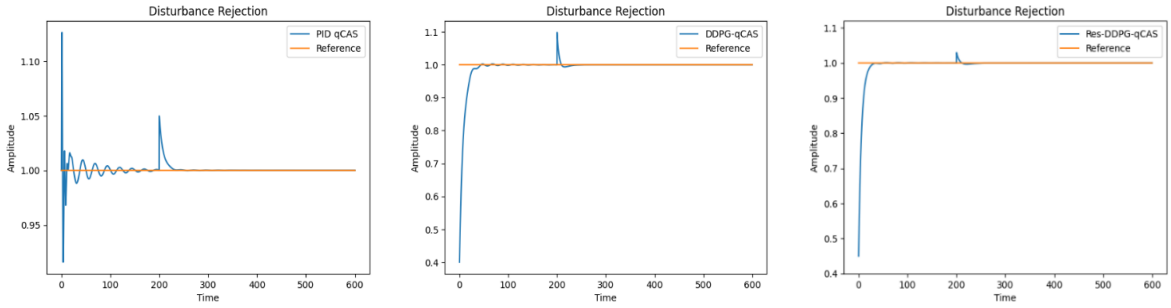
(a) (b) (c)

Figure 6. Pitch rate tracking for (a) PID-qCAS, (b) DDPG-qCAS, and (c) Res-DDPG-qCAS controllers

3.3. Disturbance rejection

In this test, the aircraft is required to maintain zero pitch-rate deviation in the presence of a large pitching moment disturbance of 30 kNm applied at $t = 200\,s$. Figure 7 compares the disturbance rejection performance of the PID-qCAS, DDPG-qCAS, and Res-DDPG-qCAS controllers. As shown in Figure 7(a), the PID-qCAS exhibits a rapid initial response but shows noticeable oscillations both during the initial transient and following the disturbance injection, indicating limited damping and sensitivity to the external perturbation. Although the pitch rate eventually returns to the reference, the oscillatory recovery implies increased transient error and reduced robustness under severe disturbance conditions.

As shown in Figure 7(b), the DDPG-qCAS demonstrates smoother behavior with reduced oscillations; however, its recovery from the disturbance is relatively slow, and a noticeable deviation from the reference persists for a longer duration. In contrast, Figure 7(c) shows that the proposed Res-DDPG-qCAS achieves the most effective disturbance rejection, with only a small transient deviation at the disturbance onset and rapid restoration of the pitch rate to the reference value. The response remains well damped and free of sustained oscillations, confirming that the residual learning architecture enhances robustness to large external disturbances while preserving stable and accurate regulation of the pitch rate.



(a) (b) (c)

Figure 7. Disturbance rejection performance for (a) PID-qCAS, (b) DDPG-qCAS, and (c) Res-DDPG-qCAS controllers

4. CONCLUSION

This study presented a disturbance-resilient control framework for aircraft pitch-rate regulation based on residual reinforcement learning integrated within a classical control architecture. The proposed approach combines a stabilizing PID controller with a DDPG agent that learns a bounded residual control signal. By augmenting rather than replacing the conventional controller, the framework preserves the stability and reliability of classical control while introducing adaptive, data-driven compensation for modeling uncertainties and external disturbances.

The objective of this work was to investigate whether a hybrid control strategy that integrates classical feedback control with deep reinforcement learning can improve pitch-rate regulation under uncertain and disturbance-rich operating conditions. To achieve this, the aircraft longitudinal dynamics were modeled and linearized around a trimmed operating point, and a control augmentation system incorporating actuator dynamics and constraints was developed. Three control strategies were evaluated within this framework: a classical PID controller, a standalone DDPG controller, and the proposed residual DDPG controller. The reinforcement learning formulation employed a Markov Decision Process representation, transient-aware reward shaping, and a reference-model-based reward design to guide the learning process toward desirable flight-control dynamics.

Simulation results demonstrated that the residual learning architecture significantly improves overall control performance. While the PID controller provided a fast response with considerable overshoot and oscillatory behavior, the standalone DDPG controller produced smoother responses at the expense of slower transient dynamics. In contrast, the proposed Res-DDPG-qCAS achieved a balanced performance, combining fast response with minimal overshoot and the lowest tracking error metrics. Command tracking and disturbance rejection experiments further confirmed that the residual controller provides smooth, well-damped responses and rapid recovery from large external disturbances.

Despite these promising results, several limitations should be acknowledged. First, the evaluation was conducted entirely in simulation, and additional robustness studies under varying disturbance profiles, parameter uncertainties, and sensor noise conditions are required to fully assess the controller's reliability.

Second, the study did not include comparisons with other hybrid reinforcement learning approaches such as PPO-based residual policies or model predictive control–guided reinforcement learning, which could provide further insight into the relative advantages of the proposed method. In addition, the work does not present a formal Lyapunov stability analysis of the combined PID–DDPG system. While the residual architecture helps preserve baseline stability, theoretical guarantees are important for safety-critical aerospace applications and should be addressed in future studies. Practical deployment considerations also remain, including the computational overhead associated with neural network inference and its impact on high-frequency flight control loops in embedded avionics systems.

Future research will therefore focus on extending the framework through robustness evaluation under broader disturbance conditions, incorporation of formal stability constraints, and comparison with alternative hybrid reinforcement learning strategies. Additional work will also investigate transfer learning and domain randomization techniques to improve policy generalization, as well as hardware-in-the-loop and experimental validation on real flight control platforms. Beyond aerospace applications, the proposed residual reinforcement learning framework may also be applicable to other safety-critical control domains such as autonomous drones, robotics, and intelligent automotive systems.

More broadly, this work contributes to the growing body of research exploring how reinforcement learning can be integrated safely into conventional control architectures. The findings support the emerging perspective that hybrid control frameworks, in which classical controllers ensure baseline stability while learning agents provide adaptive correction, represent a promising direction for next-generation intelligent control systems. Continued research in this area will be essential for addressing challenges related to verification, certification, and safe integration of learning-enabled controllers within existing aerospace control and avionics infrastructures.

FUNDING INFORMATION

Authors state no funding was involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Abolanle Adetifa	✓	✓		✓						✓				
Rexcharles Enyinna Donatus	✓	✓		✓						✓				
Daniel Udekwe	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] K. Ter, O.-O. Ajayi, and D. Udekwe, "Taxonomy and trends in reinforcement learning for robotics and control systems: a structured review," *arXiv preprint arXiv:2510.21758*, 2025.
[2] D. Udekwe, "Development of a deep reinforcement learning controller for trajectory tracking of classical control systems," *SSRN Electronic Journal*, 2025, doi: 10.2139/ssrn.5386478.

[3] M. Sadraey, "Command augmentation systems," in *Automatic Flight Control Systems*, Springer, 2020, pp. 133–153.
[4] L. Hsu and J. P. V. S. Cunha, "Chattering is a persistent problem in classical and modern sliding mode control," in *Proceedings of IEEE International Workshop on Variable Structure Systems*, 2022, vol. 2022-Septe, pp. 101–108, doi: 10.1109/VSS57184.2022.9902100.
[5] C. Wu, W. Pan, R. Staa, J. Liu, G. Sun, and L. Wu, "Deep reinforcement learning control approach to mitigating actuator attacks," *Automatica*, vol. 152, p. 110999, 2023, doi: 10.1016/j.automatica.2023.110999.
[6] D. Coraci, S. Brandi, T. Hong, and A. Capozzoli, "Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings," *Applied Energy*, vol. 333, p. 120598, 2023, doi: 10.1016/j.apenergy.2022.120598.
[7] Y. Yan, E. V. Mascaro, T. Egle, and D. Lee, "I-ctrl: Imitation to control humanoid robots through bounded residual reinforcement learning," *IEEE Robotics and Automation Magazine*, vol. 32, no. 1, pp. 59–67, 2025, doi: 10.1109/MRA.2025.3527284.
[8] L. Bu, T. Sc, and T. U. Lecture, "Residual reinforcement learning for robot control," *arXiv preprint arXiv:1812.03201*, no. March, 2011.
[9] T. Davchev, K. S. Luck, M. Burke, F. Meier, S. Schaal, and S. Ramamoorthy, "Residual learning from demonstration: Adapting DMPs for contact-rich manipulation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4488–4495, 2022, doi: 10.1109/LRA.2022.3150024.
[10] N. Piccinelli, D. Meli, E. Bonoldi, and R. Muradore, "Passive reinforcement learning with optimal control for safe convergence in cyber–physical systems," *Robotics and Autonomous Systems*, vol. 197, p. 105293, 2026, doi: 10.1016/j.robot.2025.105293.
[11] P. Kostelac, X. Wang, and A. Jamshidnejad, "MPC-guided safe reinforcement learning and Lipschitz-based filtering for structured nonlinear systems," *arXiv preprint arXiv:2512.12855*, 2025.
[12] K. Yelemessov *et al.*, "Algorithmic optimal control of screw compressors for energy-efficient operation in smart power systems," *Algorithms*, vol. 18, no. 9, p. 583, 2025, doi: 10.3390/a18090583.
[13] A. O. Adetifa, P. P. Okonkwo, B. B. Muhammed, and D. A. Udekwe, "Deep reinforcement learning for aircraft longitudinal control augmentation system," *Nigerian Journal of Technology*, vol. 42, no. 1, pp. 144–151, 2023, doi: 10.4314/njt.v42i1.18.
[14] Z. Zhao and J. Quan, "Research on SCR denitrification control strategy based on deep reinforcement learning," *Fuel*, vol. 411, p. 137998, 2026, doi: 10.1016/j.fuel.2025.137998.
[15] R. Donatus, K. Ter, O.-O. Ajayi, and D. Udekwe, "Multi-agent reinforcement learning in intelligent transportation systems: A comprehensive survey," *arXiv preprint arXiv:2508.20315*, 2025, [Online]. Available: http://arxiv.org/abs/2508.20315.
[16] H. M. Nguyen-Khac *et al.*, "Advanced feedforward control techniques: Comprehensive review and a real-time industrial application," *Annual Reviews in Control*, vol. 61, p. 101044, 2026, doi: 10.1016/j.arcontrol.2025.101044.
[17] I. Gumusboga and A. Iftar, "Aircraft trim analysis by particle swarm optimization," *Journal of Aeronautics and Space Technologies*, vol. 12, no. 2, pp. 185–196, 2019.
[18] E. Promtun and S. Seshagiri, "Sliding mode control of pitch-rate of an F-16 aircraft," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 1099–1104, 2008.
[19] E. Okafor, D. Udekwe, Y. Ibrahim, M. Bashir Mu'azu, and E. G. Okafor, "Heuristic and deep reinforcement learning-based PID control of trajectory tracking in a ball-and-plate system," *Journal of Information and Telecommunication*, vol. 5, no. 2, pp. 179–196, 2021, doi: 10.1080/24751839.2020.1833137.
[20] D. Udekwe, O. ofe Ajayi, O. Ubadike, K. Ter, and E. Okafor, "Comparing actor-critic deep reinforcement learning controllers for enhanced performance on a ball-and-plate system," *Expert Systems with Applications*, vol. 245, p. 123055, 2024, doi: 10.1016/j.eswa.2023.123055.
[21] E. G. Okafor, D. Udekwe, O. C. Ubadike, E. Okafor, P. O. Jemitola, and M. T. Abba, "Photovoltaic system MPPT evaluation using classical, meta-heuristics, and reinforcement learning-based controllers: A comparative study," *Journal of Southwest Jiaotong University*, vol. 56, no. 3, pp. 1–17, 2021, doi: 10.35741/issn.0258-2724.56.3.1.
[22] E. Okafor, D. Udekwe, M. Muhammad, O. Ubadike, and E. Okafor, "Solar system maximum power point tracking evaluation using reinforcement learning," in *Proceedings of 2021 Sustainable Engineering and Industrial Technology Conference*, 2021, pp. F10-1-F10-4.
[23] D. Udekwe, "Evaluating a DDPG reinforcement learning agent on a ball-and-plate system: A comparative study of intelligent control approaches," *Nigerian Journal of Technology*, vol. 44, no. 2, pp. 338–346, 2025, doi: 10.4314/njt.v44i2.16.
[24] S. R. Afzali, M. Shoaran, and G. Karimian, "A modified convergence DDPG algorithm for robotic manipulation," *Neural Processing Letters*, vol. 55, no. 8, pp. 11637–11652, 2023, doi: 10.1007/s11063-023-11393-z.
[25] O.-O. Ajayi, R. Donatus, K. Ter, and D. Udekwe, "Integrating reinforcement learning with virtual reality: a survey of techniques and use cases," *SSRN Electronic Journal*, 2025.

BIOGRAPHIES OF AUTHORS

Abolanle Adetifa received her MSc in aerospace vehicle design from the Air Force Institute of Technology (AFIT), Nigeria, where she specialized in aircraft performance analysis, stability and control, and aerospace systems engineering. Her research interests include flight dynamics, advanced control strategies, and the application of intelligent algorithms to aerospace vehicle design and optimization. She is particularly focused on developing robust and efficient control solutions for complex and safety-critical aerospace systems. She can be contacted at: adetifabibi@gmail.com.

Rexcharles Enyinna Donatus holds dual master's degrees in aerospace vehicle design from the Air Force Institute of Technology (AFIT), Nigeria, and Information Technology from the National Open University of Nigeria (NOUN). Since 2020, he has served as a lecturer in the Department of Aerospace Engineering at AFIT, where he leads research initiatives at the intersection of artificial intelligence and aerospace applications. His scholarly contributions span multiple peer-reviewed journals, with current research interests focused on deep learning, computer vision, affective computing and designing intelligent control systems for autonomous aircraft operations. He can be contacted at email: rdonatus@afit.edu.ng.

Daniel Udekwe received his MSc in aerospace vehicle design from the Air Force Institute of Technology (AFIT), Nigeria, where he developed a strong foundation in aircraft design, flight dynamics, and control systems. His research interests include flight control engineering, disturbance-resilient control, and the application of machine learning techniques to aerospace systems. He is particularly interested in integrating reinforcement learning with classical control architectures to enhance robustness and performance in safety-critical environments. He can be contacted at email: daudekwe@aft.edu.ng.