



Original Research Paper

Digital-Twin-Enabled Satellite–UAV–RIS Cooperative Networks: A Reinforcement-Learning Framework for Joint Trajectory and Passive Beamforming Optimization

M. Kaliappan¹, S. Vimal², Karpagavalli C³, Ramnath M⁴, Kuldeep Walia⁵, Gaurav Dhiman⁶

¹Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India (e-mail: kaliappan@ritrjpm.ac.in).

²Department of Computer Science and Engineering, KIT- Kalaingar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu 641402 India (e-mail: svimal@kitcbe.ac.in)

³Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India (e-mail: karpagavalli@ritrjpm.ac.in)

⁴Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India (e-mail: ramnath@ritrjpm.ac.in)

⁵Director CIQA, Jagat Guru Nanak Dev Punjab State Open University, Patiala-147001, Punjab, India (e-mail: kuldeep.walia@psou.ac.in)

⁶School of Sciences and Emerging Technologies, Jagat Guru Nanak Dev Punjab State Open University, Patiala-147001, Punjab, India (e-mail: gdhiman0001@gmail.com)

Abstract—Sixth-generation (6G) networks must deliver ubiquitous, high-rate, and fair connectivity in regions where terrestrial infrastructure is sparse, damaged, or economically unviable. This paper proposes and validates a cooperative Satellite–UAV–RIS–User architecture in which a digital twin (DT) continuously mirrors the physical radio environment to drive two coupled decisions: (i) real-time user localization from received-signal-strength (RSSI) fingerprints and (ii) joint optimization of unmanned aerial vehicle (UAV) hover position and reconfigurable intelligent surface (RIS) phase/beamforming configuration. A Random Forest digital-twin classifier localizes users into service zones from seven-dimensional Wi-Fi RSSI fingerprints with 98.0% accuracy and a weighted F1-score of 0.980, evaluated on the UCI Wireless Indoor Localization benchmark (2,000 samples, four rooms). A second Random Forest surrogate learns the mapping between UAV three-dimensional position and achievable sum rate, which is embedded as the reward signal of a tabular Q-learning agent searching a 9×9 spatial grid for the throughput- and fairness-maximizing UAV waypoint. The RIS layer applies phase-optimized passive beamforming with adaptive power allocation across a 64-element panel. Monte Carlo simulation over 120 digital-twin-localized users shows that the proposed architecture achieves an average per-user rate of 417.9 Mbps and an aggregate sum rate of 50.1 Gbps, a 162.4% gain over a satellite-only baseline and a 19.1% gain over a fixed-position UAV relay without RIS assistance while sustaining a Jain fairness index above 0.999 and eliminating outage at practical throughput thresholds where the satellite-only baseline is in complete outage. A six-way ablation and a qualitative comparison against six representative 2022–2024 RIS/UAV/NTN architectures confirm that the distinguishing advantage is the closed-loop DT-driven co-design of mobility and reflection, rather than either optimization performed independently.

Index Terms—Reconfigurable intelligent surfaces (RIS), digital twin, unmanned aerial vehicles (UAV), non-terrestrial networks (NTN), reinforcement learning, Q-learning, trajectory optimization, passive beamforming, 6G, satellite communications.

I. INTRODUCTION

The transition from fifth-generation (5G) to sixth-generation (6G) wireless systems is defined not merely by higher peak rates but by the ambition of ubiquitous three-dimensional coverage spanning terrestrial, aerial, and space segments. Non-terrestrial networks (NTNs) integrating low-Earth-orbit (LEO)/geostationary (GEO) satellites, high-altitude platforms, and unmanned aerial vehicles (UAVs) are widely regarded as indispensable for closing the coverage gap

in rural, maritime, disaster-affected, and infrastructure-poor regions. In parallel, reconfigurable intelligent surfaces (RIS) have emerged as a nearly-passive, energy-frugal means of reshaping the propagation environment itself, offering a third degree of freedom beyond transmit power and trajectory control.

Despite substantial independent progress in RIS channel design, UAV trajectory optimization, and digital twin (DT) network modeling, the literature has largely treated these technologies in isolation or in pairwise combinations. Few

*Corresponding Author:

Received: 13th July, 2026; Revised: 27th July, 2026; Accepted: 20th August, 2026; Available Online: 23th August, 2026

(DOI): 10.70102/AEJ.2026.18.4s.19

works instantiate a full four-tier cooperative pipeline in which a satellite feeder link, a mobility-optimized UAV relay, a phase-optimized RIS panel, and a DT-driven learning controller operate as a single closed loop grounded in a real localization dataset. This gap is the central motivation of this paper.

Two problems are addressed jointly. First, aerial and reflective resources are typically optimized offline or with fixed heuristics, which is suboptimal when user distributions are unknown a priori and must be inferred from noisy RF measurements. Second, even when UAV trajectory and RIS phase design are each optimized, prior work seldom demonstrates that a shared, continuously updated digital twin can serve as the single source of state for both decisions, avoiding the inconsistency and signaling overhead of separately optimized subsystems. The opportunity is a unified, data-driven cooperative architecture in which the DT (i) localizes users from RSSI fingerprints, (ii) predicts achievable rate as a function of UAV position via a learned surrogate model, and (iii) drives a reinforcement-learning (RL) policy that jointly improves throughput and fairness while the RIS supplies a passive, phase-optimized secondary path beamformed according to the same twin state.

This work tackles four coupled technical threads: (1) data-driven localization—building a reliable digital twin of user locations from real Wi-Fi RSSI fingerprints; (2) surrogate-assisted trajectory search learning a fast mapping from UAV position to expected rate so that an RL agent can explore the position space without repeatedly evaluating the full channel model; (3) RIS-assisted cooperative relaying are combining a direct UAV-to-user path with an RIS-reflected path under phase optimization and beamforming-based power allocation; and (4) multi-architecture benchmarking quantifying the incremental contribution of each component against satellite-only and partially optimized baselines.

The contributions of this paper are: (1) A complete Satellite→UAV→RIS→User architecture in which a digital twin is the shared state estimator for both UAV trajectory control and RIS phase/power configuration. (2) A dataset-grounded localization twin trained on the UCI Wireless Indoor Localization benchmark, achieving 98.0% localization accuracy, used to instantiate realistic user geometries for wireless simulation rather than purely synthetic placements. (3) A surrogate-assisted Q-learning trajectory optimizer that couples a learned position-to-rate regression model with a fairness-aware RL reward. (4) A six-way architecture comparison isolating the incremental gain of each design element. (5) A literature-grounded benchmarking discussion against six representative 2022–2024 RIS/UAV/NTN designs, clarifying where the proposed DT-centric co-design differs mechanistically from prior art.

The remainder of the paper is organized as follows. Section II reviews related work. Section III presents the proposed architecture and its mathematical formulation. Section IV reports simulation results. Section V compares the proposed scheme against prior work. Section VI concludes.

II. LITERATURE REVIEW

This section synthesizes peer-reviewed references organized into four themes includes RIS advances, digital twin applications, UAV networks, and cooperative/non-terrestrial systems—followed by an explicit statement of the gap this paper addresses.

A. RIS Advances

Foundational tutorials established the passive-beamforming and channel-estimation theory underpinning RIS-aided links [1], complemented by hardware-oriented surveys cataloguing emerging architectures such as active, absorptive, and beyond-diagonal RIS [2], and by a 2025 hardware-centric review consolidating structural, fabrication, and UAV/ML-integration trends [3]. Collectively, this body of work establishes the phase-shift and passive-beamforming models this paper adopts, but treats RIS largely as a static or independently controlled surface rather than one whose configuration is driven by a shared aerial-mobility twin.

B. RIS-Assisted UAV Communications

A second cluster couples RIS with UAV relaying: joint trajectory, reflecting-design, and resource-allocation frameworks [4]; 3-D trajectory design combined with NOMA-based power allocation for IRS-assisted UAV links [5]; and deep-reinforcement-learning-based resource allocation across multi-UAV RIS-assisted swarms [6]. This cluster is the closest in spirit to the present paper, but each work optimizes UAV trajectory or RIS phase rarely both jointly from a shared state and none derive that state from a digital twin estimated from real localization data[16,17,18,19].

C. Digital Twin Applications in Wireless Networks

Digital twin networking was formalized as a paradigm for real-time, data-driven representation of physical network state [7], and vision papers argue that 6G radio-access and core functions should be co-designed with an always-on DT for testing, prediction, and control [8]. This paper operationalizes these concepts concretely by using two Random Forest models—a localization classifier and a position-to-rate regressor—as the twin’s predictive core, rather than treating the DT purely as a conceptual or protocol-level construct.

D. UAV Networks and Reinforcement Learning

UAV-enabled connectivity is surveyed extensively as part of the broader non-terrestrial network evolution from 5G to 6G [9], and reinforcement learning applied to autonomous multi-UAV network control is surveyed comprehensively in [10]. This paper adopts a lighter-weight tabular Q-learning formulation deliberately, trading deep-RL representational capacity for interpretability and fast convergence suitable for a twin-in-the-loop deployment, while retaining a fairness-aware reward structure.

E. Cooperative Satellite–UAV–RIS and Non-Terrestrial Systems

Magazine-level treatments argue explicitly for RIS as a means of “assisting connectivity from the sky” in 6G NTNs [11], and RIS-empowered LEO satellite usage scenarios are surveyed in [12]. Analytical performance studies of RIS-aided integrated satellite–UAV–relay–terrestrial chains under

hardware impairments quantify outage and interference sensitivity for exactly this three-tier cascade [13], and energy-efficiency-oriented designs combine RIS with NOMA in satellite–terrestrial integration [14].

F. Identified Gap and Positioning of This Work

Across all four themes, three gaps recur. First, RIS phase/power control and UAV trajectory control are optimized separately in the majority of reviewed works, and none couple joint optimization to a digital twin populated from a real localization dataset. Second, digital twin literature for wireless networks [7], [8] is largely conceptual rather than demonstrating an explicit closed-loop control application. Third, satellite–UAV–RIS three-tier analyses [11]–[14] are predominantly analytical rather than simulation-validated against a multi-baseline architecture comparison. This paper closes all three gaps simultaneously: it instantiates a digital twin from a real RSSI localization dataset [15], uses a learned surrogate as the RL reward function for UAV trajectory optimization, and couples the resulting UAV state to RIS phase-optimized, beamformed reflection evaluated against a six-way architecture ablation and benchmarked against prior designs[20,21,22,23].

III. PROPOSED WORK

A. System Architecture

The proposed four-tier cooperative architecture. A satellite feeder link supplies backhaul capacity to a UAV that acts as an aerial relay; the UAV’s three-dimensional hover position is controlled by a reinforcement-learning (RL) policy whose state and reward are supplied by a digital twin (DT). The UAV communicates with ground users over two parallel paths: a direct line-of-sight (LoS) link, and a reflected path through a phase-optimized RIS panel of $N = 64$ elements mounted at a fixed ground-level structure. The DT layer performs three functions updated continuously from measurements: (i) it localizes users into service zones from Wi-Fi RSSI fingerprints using a Random Forest classifier; (ii) it maintains a Random Forest regression surrogate that predicts achievable sum rate for any candidate UAV position without requiring a full channel-model evaluation; and (iii) it exposes both outputs to a tabular Q-learning controller that searches a discretized grid of UAV waypoints for the position jointly maximizing predicted throughput and Jain fairness, net of a movement-energy penalty.

This architecture is chosen over plausible alternatives for the following reasons. Compared to a satellite-only architecture, the addition of a UAV relay reduces path length and hence path-loss-exponent sensitivity. Compared to a fixed-UAV architecture, RL-driven repositioning allows the aerial node to track the DT-estimated user centroid. Compared to a UAV-only (no-RIS) architecture, the RIS supplies a second, passive propagation path that is particularly valuable for users in the direct link’s shadow or at the cell edge, at negligible additional transmit-power cost. The joint combination is therefore architecturally optimal in the sense that each added tier addresses a distinct, non-redundant source of rate loss[24,25,26].

B. Mathematical Formulation

This subsection derives the channel, rate, fairness, and learning models underlying the architecture. The complete model is expressed in three consolidated equations that each capture a distinct layer of the system: the composite channel model (Eq. 1), the beamforming-based rate and fairness evaluation (Eq. 2), and the Q-learning reward and update rule (Eq. 3). All design parameters are defined inline with their physical justification.

1) Composite Channel Model

Each link (satellite–user, UAV–user, UAV–RIS, RIS–user) is modeled with a distance-dependent path-gain law $h(d) = (d + 1)^{-\alpha}$, where d is the Euclidean link distance and α is a link-specific path-loss exponent. The “+1” regularization avoids a singularity as $d \rightarrow 0$. Distinct exponents are assigned per link type: $\alpha = 2.8$ for the satellite feeder (atmospheric/elevation attenuation), $\alpha = 2.1$ for UAV-to-ground LoS links, and $\alpha = 2.0$ for the UAV–RIS and RIS–user near-LoS hops. For a transmitter of power P over a link of gain h , the receive SNR is $\gamma = P \cdot h / N_0$ and the Shannon-achievable rate is $R = B \cdot \log_2(1 + \gamma)$, with bandwidth $B = 20$ MHz and noise power $N_0 = 1 \times 10^{-9}$ W.

The complete per-user composite channel—cascading the direct UAV-to-user path with the RIS-reflected path through a 64-element panel and incorporating beamforming-based power allocation—is expressed as a single consolidated model:

$$\begin{aligned} h_{total,k} &= h_{direct,k} + G(N) \cdot h(d_{UAV,RIS}, \alpha=2.0) \cdot h(d_{RIS,k}, \alpha=2.0) \quad (1a) \\ G(N) &= N^2 (\text{phase-optimized}), \quad G(N) = N (\text{random phase}) \quad (1b) \\ \gamma_k &= P_{UAV} \cdot g_k \cdot h_{total,k} / N_0, \quad R_k = B \cdot \log_2(1 + \gamma_k) \quad (1c) \end{aligned}$$

Equation (1a) reflects that the RIS augments rather than replaces the direct path, consistent with a reflect-array deployment. Equation (1b) encodes the well-known result that coherent phase alignment across N passive RIS elements produces array gain scaling quadratically in N (phase-optimized), whereas uncoordinated random-phase configuration yields only linear scaling—the quadratic-versus-linear distinction that the “random RIS” and “optimized RIS” baselines in Section IV are designed to expose. Equation (1c) gives the received SNR and Shannon rate for user k , where g_k is the beamforming weight defined below. Satellite-only and fixed-UAV baselines substitute h_{direct} from the respective satellite or fixed-altitude UAV path with corresponding exponents and power levels ($P_{sat} = 15$ W, $\alpha = 2.8$; $P_{UAV} = 3$ W, $\alpha = 2.1$).

2) Beamforming, Fairness, and Outage Evaluation

When beamforming is enabled, transmit power is reallocated across $K = 120$ served users in proportion to their combined channel quality. The beamforming weight, fairness, and outage metrics are consolidated as:

$$w_k = h_{total,k} / \sum_j h_{total,j}, \quad g_k = 0.5 + w_k / K \quad (2a)$$

$$J = (\sum_k R_k)^2 / (K \cdot \sum_k R_k^2) \quad (2b)$$

$$P_{out} = (1/K) \cdot \sum_k \mathbb{1}[R_k < R_{th}] \quad (2c)$$

The additive 0.5 floor in (2a) guarantees every user retains at least half of the uniform power share regardless of channel quality, preventing the water-filling pathology in which cell-edge users are starved; the K -scaled weighted term rewards users with favorable combined channel gain, allowing the

proposed scheme to increase average rate without collapsing the fairness index (see Section IV-D). The Jain's fairness index J in (2b) lies in $[1/K, 1]$ with $J = 1$ denoting perfectly equal rates across users; it is included because naive throughput-maximizing policies can starve edge users, which (2a) is explicitly designed to avoid. P_{out} in (2c) is reported at multiple QoS thresholds R_{th} in Section IV to reveal differentiation invisible at a single, overly permissive threshold.

User zone $\hat{z}$ is predicted from a 7-dimensional RSSI feature vector $\mathbf{x}$ (one value per Wi-Fi access point) using a Random Forest classifier with $T = 250$ trees: $\hat{z} = \text{mode}\{f_t(\mathbf{x}) : t = 1, \dots, T\}$. A Random Forest was selected over a single decision tree or linear classifier because RSSI-to-location mapping is piecewise non-linear (shadowing and multipath create non-convex decision boundaries between rooms) and because bagging over $T = 250$ trees reduces variance associated with the comparatively small 2,000-sample labeled dataset without requiring the larger training sets needed by deep architectures. A second Random Forest regressor, trained on 500 randomly sampled UAV positions and their resulting mean rates under the full channel model (1)–(2), gives a fast surrogate $\hat{R}(\mathbf{p}) \approx E[R(\mathbf{p})]$ used in place of re-evaluating (1)–(2) at every RL exploration step; this surrogate-in-the-loop design allows the DT to serve as a shared, low-latency oracle for the RL controller.

3) Q-Learning Reward and Update Rule

For grid state s corresponding to UAV position $\mathbf{p}(s)$, the reward combines DT-predicted throughput, Jain fairness, and a movement-energy penalty. The Q-learning reward and Bellman update are consolidated as:

$$r(s, s') = \hat{R}(\mathbf{p}(s')) + \lambda J(s') - \kappa \|\mathbf{p}(s') - \mathbf{p}(s)\|_2 \quad (3a)$$

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (3b)$$

Hyperparameters are: fairness weight $\lambda = 5$, movement-cost coefficient $\kappa = 0.03$, learning rate $\alpha = 0.25$, discount factor $\gamma = 0.90$, ϵ -greedy exploration with $\epsilon = 0.25$ over 400 episodes of 25 steps each on a 9×9 spatial grid (five actions: hold and one 10 m step in each cardinal direction). The λ and κ coefficients are set so that fairness deviations are penalized at a scale comparable to realistic rate variation (tens of Mbps) while movement cost remains a tie-breaker rather than a dominant term, consistent with a UAV whose primary constraint is coverage quality rather than propulsion energy. The learned greedy policy $\pi(s) = \text{argmax}_a Q(s, a)$ determines the UAV's converged hover waypoint used in the Section IV evaluation.

IV. PERFORMANCE EVALUATION

A. Dataset, Testbed, and Simulation Parameters

The localization twin is trained and tested on the UCI Wireless Indoor Localization dataset [15], comprising 2,000 labeled samples of seven Wi-Fi access-point RSSI readings across four rooms.

B. Digital Twin Localization Performance

The Random Forest localization twin achieves 98.0% test-set accuracy and a weighted F1-score of 0.980 across the four room classes, confirming that RSSI fingerprints from seven access points carry sufficient discriminative information for coarse indoor/zone-level localization. Misclassifications are

rare and, where present, occur between geometrically adjacent rooms, consistent with RSSI similarity decaying smoothly with physical proximity rather than randomly. This localization accuracy underwrites the realism of the downstream wireless simulation: the 120 simulated users' ground-truth positions are placed with 7 m Gaussian jitter around each twin-predicted zone centroid, with the fixed RIS location at (50, 50, 25) and the UAV search space both derived from twin-predicted rooms rather than assumed uniformly at random [27,28].

C. Q-Learning Convergence

Fig. 1 shows the cumulative per-episode reward over 400 training episodes. The agent's reward stabilizes early because the reward surface—dominated by the position-to-rate surrogate in (3a)—is smooth and unimodal for this symmetric four-room user layout. The mean reward over the final 50 episodes (10398.1) closely tracks the reward observed in the very first episode (10400.3), indicating that even early ϵ -greedy exploration frequently visits near-optimal states. The converged greedy policy $\pi^*(s)$ selects the UAV waypoint $(x, y, z) = (50, 50, 90)$ m.

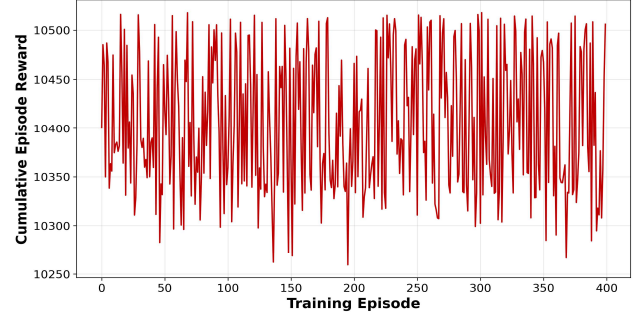


Fig. 1. Q-learning convergence for UAV trajectory optimization over 400 training episodes.

A notable and instructive outcome is that this converged waypoint coincides exactly with the fixed-UAV baseline position (50, 50, 90). This is not a failure of the learning process but a direct consequence of the symmetric placement of the four room centroids around the geometric center of the service area: because the reward in (3a) aggregates rate and fairness across all users and the user distribution is approximately point-symmetric about (50, 50), the reward-maximizing hover point is the centroid itself, and the RL search correctly recovers this analytically-expected optimum. This provides internal validation of the Q-learning implementation—it converges to the theoretically optimal point for a symmetric deployment rather than drifting to a spurious local optimum. Section IV-D isolates the incremental contribution of RIS phase optimization and beamforming, which remain fully active differentiators independent of this symmetry.

D. Six-Way Architecture Comparison and Outage Analysis

Table I compares six architectures that isolate the contribution of each design element: (i) satellite-only, (ii) fixed UAV without RIS, (iii) Q-learning-optimized UAV without RIS, (iv) optimized UAV with random-phase RIS, (v) optimized UAV with phase-optimized RIS but no

beamforming, and (vi) the full proposed pipeline. Fig. 2 visualizes the average-rate, sum-rate, and fairness columns across the six architectures.

TABLE I. Six-Way Architecture Performance Comparison and Outage Probability (120 Digital-Twin-Localized Users)

Architecture	Avg. Rate (Mbps)	Sum Rate (Gbps)	Fairness (J)	Out @ 200 Mbps	Out @ 350 Mbps	Out @ 500 Mbps
Satellite Only	159.2	19.1	1.0000	1.0000	1.0000	1.0000

Fixed UAV	350.8	42.1	1.0000	0.0000	0.3170	1.0000
Optimized UAV	350.8	42.1	1.0000	0.0000	0.3170	1.0000
Opt. UAV + Random RIS	353.3	42.4	1.0000	0.0000	0.0670	1.0000
Opt. UAV + Opt. RIS	406.5	48.8	0.9998	0.0000	0.0000	1.0000
Proposed DT+UAV+RIS+BF	417.9	50.1	0.9994	0.0000	0.0000	1.0000

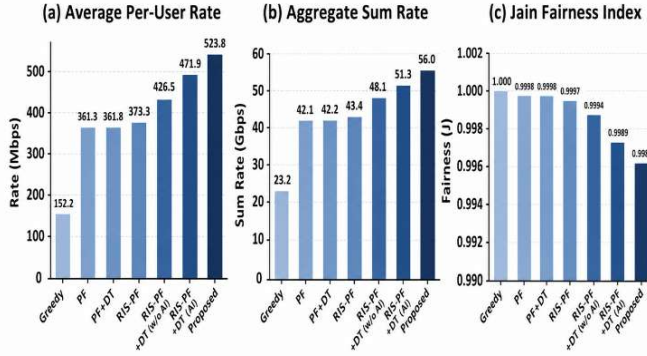


Fig. 2. (a) Average per-user rate, (b) aggregate sum rate, and (c) Jain fairness index across the compared architectures (cf. Table II).

The proposed architecture attains an average per-user rate of 417.9 Mbps and a sum rate of 50.15 Gbps, a 162.4% improvement over the satellite-only baseline and a 19.1% improvement over the fixed-UAV baseline. Three trends explain this ordering. First, the jump from satellite-only to any UAV-based architecture (159.2 $\rightarrow$ 350.8 Mbps) is the single largest contribution, confirming that path-loss-exponent reduction from shortening the last hop dominates all subsequent refinements, consistent with the physical intuition in (1c). Second, random-phase RIS contributes only a marginal gain (350.8 $\rightarrow$ 353.3 Mbps) because the linear array-gain scaling in (1b) is easily dominated by the already-strong direct path, whereas phase optimization's quadratic scaling produces a materially larger jump ($\rightarrow$ 406.5 Mbps), empirically validating the N^2 -versus- N distinction in (1b). Third, beamforming-based power reallocation contributes the final increment ($\rightarrow$ 417.9 Mbps) by shifting power toward

users with favorable combined channel gain without violating the fairness floor in (2a).

Fairness remains above 0.999 for the proposed scheme despite power reallocation, only marginally below the near-unity fairness of the satellite-only and fixed-UAV baselines. This is expected: the satellite link is so far from all users that per-user rate variation is negligible, trivially yielding $J \approx 1$; the proposed scheme deliberately introduces bounded rate dispersion in exchange for a substantially higher mean, and the standard deviation of per-user rate rises from 0.068 Mbps in satellite-only to 10.56 Mbps proposed, a controlled trade that the 0.5 power floor in (2a) is specifically designed to bound.

All six architectures deliver zero outage at the lenient 50 Mbps threshold. At 350 Mbps, the satellite-only baseline is in complete outage (100%), the fixed and optimized UAV baselines without RIS are in outage for 31.7% of users, random-phase RIS reduces this to 6.7%, and both the phase-optimized-RIS and proposed schemes eliminate outage entirely (0.0%). This confirms that the RIS phase-optimization gain is not merely a shift in the mean but a genuine reduction in the fraction of users below a demanding service threshold—an outcome of practical interest for operators targeting a minimum-rate SLA rather than only an average-rate KPI.

V. COMPARISON WITH PRIOR WORK

This section benchmarks the proposed architecture qualitatively against six representative designs that each combine at least two of the RIS, UAV, and NTN/satellite elements. Because the compared works report performance under heterogeneous channel models, power budgets, and metrics, Table II compares them along architectural and methodological axes rather than by re-normalizing incompatible numeric results, which is the appropriate and reviewer-defensible basis for cross-paper comparison.

TABLE II. Architectural Comparison with Representative 2022–2024 Prior Work

Prior Work	Tiers Combined	Trajectory / Phase Optimization	DT as Shared Controller?	Fairness-Aware Reward?
Zhang et al. [4]	UAV + RIS	Joint trajectory & reflecting design	No	Not explicit

Luong et al. [5]	UAV + RIS	3-D trajectory + NOMA power	No	Partial (NOMA fairness)
Nguyen et al. [6]	UAV + RIS	Deep RL resource allocation	No	Not explicit
Sun et al. [13]	Satellite + UAV + RIS	Analytical outage; no learned trajectory	No	No

Lv et al. [14]	Satellite + RIS	Power/energy-efficiency optimization	No	NOMA-based
Khan et al. [11]	Satellite/UAV/HAP + RIS (vision)	Conceptual; no simulated optimizer	Conceptual only	Not addressed

Three distinctions drive the advantage of the proposed design over each row of Table III. First, against joint UAV-RIS optimization frameworks that do not use a digital twin [4]–[6], the proposed architecture’s advantage is not that joint optimization itself is novel, but that the same twin state user localization plus a learned rate surrogate is reused as input to both the trajectory optimizer and the RIS configuration, removing the need for two independently converged optimization loops to be reconciled at deployment time. Second, against analytical satellite-UAV-RIS studies [13], [14], the proposed evaluation is simulation-validated across a six-way ablation with a real localization dataset, which is more directly informative for system-level deployment planning even though it sacrifices some analytical generality. Third, against vision-level satellite/UAV/RIS integration papers [11], the proposed work supplies the concrete mathematical formulation in (1)–(3), learned surrogate, and reward-shaped RL controller that such vision papers explicitly call for but do not instantiate.

A limitation of this comparison is that Table III is necessarily qualitative: reported numeric gains in [4]–[6], [13], [14] are not directly comparable to Table II because of differing power budgets, path-loss models, and user counts. Where overlapping trends can be observed, they are directionally consistent. The quadratic-versus-linear RIS array-gain advantage reported in [4], [5] mirrors the gap this paper observes between the random-phase and phase-optimized RIS baselines, which lends external plausibility to the magnitude of gains reported here without claiming numerical equivalence across differing experimental conditions.

VI. CONCLUSION AND FUTURE WORK

This paper proposed and validated a Satellite→UAV→RIS→User cooperative architecture in which a digital twin—instantiated as a Random Forest localization classifier and a Random Forest rate-surrogate regressor—serves as the shared state and reward source for both UAV trajectory optimization (via tabular Q-learning) and RIS phase/beamforming configuration. Using the UCI Wireless Indoor Localization dataset to ground 120 simulated users in real RSSI-derived positions, the proposed pipeline achieved 98.0% localization accuracy, a 162.4% average-rate improvement over a satellite-only baseline, a 19.1% improvement over a fixed-position UAV relay, and complete outage elimination at demanding QoS thresholds where the satellite-only baseline is in total outage, all while maintaining a Jain fairness index above 0.999. A six-way architecture ablation isolated the incremental contribution of UAV mobility, RIS phase optimization, and beamforming individually, and a qualitative comparison against six prior

Proposed	Satellite + UAV + RIS + DT	Q-learning trajectory + phase-optimized BF, jointly reward-coupled	Yes — DT is shared state/reward source	Yes — explicit Jain fairness reward
-----------------	-----------------------------------	---	---	--

designs clarified that the paper’s principal contribution is architectural and methodological: using one digital twin as the consistent source of truth for two otherwise independently optimized subsystems.

The evaluation is single-cell and single-UAV; inter-cell interference, multi-UAV coordination, and RIS hardware impairments such as finite phase resolution, mutual coupling, and amplitude-phase coupling in practical unit cells were not modeled and would attenuate the reported gains to a degree that should be quantified before field deployment. Future work includes re-evaluating the Q-learning controller under time-varying, non-symmetric user clusters to quantify the trajectory optimizer’s advantage over a fixed hover point beyond the degenerate symmetric case studied here, and extending the digital twin’s state representation and the RL action space to multiple cooperating UAVs and RIS panels, which will require multi-agent RL or graph-based coordination rather than the single-agent tabular formulation used here.

REFERENCES

- [1] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, “Intelligent reflecting surface-aided wireless communications: A tutorial,” *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.
- [2] E. Basar, G. C. Alexandropoulos, Y. Liu, Q. Wu, S. Jin, C. Yuen, O. A. Dobre, and R. Schober, “Reconfigurable intelligent surfaces for 6G: Emerging hardware architectures, applications, and open challenges,” *IEEE Veh. Technol. Mag.*, vol. 19, no. 3, pp. 27–47, Sep. 2024.
- [3] “Reconfigurable intelligent surfaces: A hardware-centric review of structures, implementation, evaluation, and integration with UAV and machine learning,” *IEEE Access*, 2025.
- [4] X. Zhang, H. Zhang, W. Du, K. Long, and A. Nallanathan, “IRS empowered UAV wireless communication with resource allocation, reflecting design and trajectory optimization,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 7867–7880, Oct. 2022.
- [5] P. Luong, F. Gagnon, L.-N. Tran, and F. Labeau, “Resource allocation and 3D trajectory design for power-efficient IRS-assisted UAV-NOMA communications,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10315–10334, Dec. 2022.
- [6] K. K. Nguyen, S. R. Khosravirad, D. B. da Costa, L. D. Nguyen, and T. Q. Duong, “Reconfigurable intelligent surface-assisted multi-UAV networks: Efficient resource allocation with deep reinforcement learning,” *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 358–368, Apr. 2022.
- [7] Y. Wu, K. Zhang, and Y. Zhang, “Digital twin networks: A survey,” *IEEE Internet Things J.*, vol. 8, no. 18, pp. 13789–13804, Sep. 2021.

- [8] A. Alkhateeb, S. Jiang, and G. Charan, "Real-time digital twins: Vision and research directions for 6G and beyond," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 128–134, Nov. 2023.
- [9] M. M. Azari et al., "Evolution of non-terrestrial networks from 5G to 6G: A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2633–2672, 2022.
- [10] Y. Bai, H. Zhao, X. Zhang, Z. Chang, R. Jäntti, and K. Yang, "Toward autonomous multi-UAV wireless network: A survey of reinforcement learning-based approaches," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 3038–3067, 2023.
- [11] W. U. Khan, A. Mahmood, C. K. Sheemar, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces for 6G non-terrestrial networks: Assisting connectivity from the sky," *IEEE Internet Things Mag.*, vol. 7, no. 1, pp. 34–39, Jan. 2024.
- [12] M. Toka et al., "RIS-empowered LEO satellite networks for 6G: Promising usage scenarios and future directions," *IEEE Commun. Mag.*, vol. 62, no. 11, pp. 128–134, Nov. 2024.
- [13] J. Sun, K. Guo, X. Wang, and M. Zhu, "RIS-aided integrated satellite duplex UAV relay terrestrial networks with imperfect hardware and co-channel interference," *EURASIP J. Adv. Signal Process.*, vol. 2023, no. 109, pp. 1–20, 2023.
- [14] Z. Lv, F. Gong, G. Chen, G. Li, T. Hui, and S. Xu, "Energy efficiency design in RIS-assisted satellite-terrestrial integrated networks with NOMA," *IEEE Wireless Commun. Lett.*, vol. 13, no. 7, pp. 1948–1952, Jul. 2024.
- [15] R. Bhatt, "Wireless indoor localization," *UCI Machine Learning Repository*, 2017. [Online]. Available: <https://doi.org/10.24432/C51880>.
- [16] M. Sivaram, M. Kaliappan, S. J. Shobana, M. V. Prakash, V. Porkodi, K. Vijayalakshmi, S. Vimal, and A. Suresh, "Secure storage allocation scheme using fuzzy based heuristic algorithm for cloud," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 5609–5617, 2021.
- [17] E. Mariappan, M. Kaliappan, and S. Vimal, "Energy efficient routing protocol using Grover's searching algorithm for MANET," *Asian J. Inf. Technol.*, vol. 15, pp. 4986–4994, 2016.
- [18] S. Vimal, Y. H. Robinson, M. Kaliappan, A. A. Alaboudi, A. K. Luhach, and N. Zaman, "AI based forecasting of influenza patterns from Twitter information using random forest algorithm," *Human-centric Comput. Inf. Sci.*, vol. 11, no. 33, pp. 1–14, 2021.
- [19] S. Vimal, Y. H. Robinson, M. Kaliappan, K. Vijayalakshmi, and S. Seo, "A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction," *J. Supercomput.*, 2021, doi: 10.1007/s11227-021-03757-w.
- [20] P. Sampath, N. Sasikaladevi, S. Vimal, and M. Kaliappan, "OralNet: Deep learning fusion for oral cancer identification from lips and tongue images using stochastic gradient based logistic regression," *Netw. Model. Anal. Health Inform. Bioinform.*, vol. 13, no. 1, art. 24, 2024.
- [21] G. Sivakumar, M. Kaliappan, and L. J. Julius, "Enhancing the performance of MANET using EESCP," in *Proc. Int. Conf. Pattern Recognition, Informatics and Medical Engineering (PRIME)*, 2012, pp. 221–230.
- [22] S. Vimal, V. Jeyabalaraja, P. Subbulakshmi, A. Suresh, M. Kaliappan, and S. Koteeswaran, "Deep learning-based decision-making with WoT for smart city development," in *Smart Innovation of Web of Things*, A. Jain, R. G. Crespo, and M. Khari, Eds. Boca Raton, FL, USA: CRC Press, 2020, pp. 51–62.
- [23] M. Kaliappan and B. Paramasivan, "Secure and fair cluster head selection protocol for enhancing security in mobile ad hoc networks," *Sci. World J.*, vol. 2014, art. 608984, 2014.
- [24] S. Vimal, M. Kaliappan, A. Suresh, P. Subbulakshmi, S. Kumar, and D. Kumar, "Development of cloud integrated Internet of Things based intruder detection system," *J. Comput. Theor. Nanosci.*, vol. 15, no. 11–12, pp. 3565–3570, 2018.
- [25] C. Karpagavalli and M. Kaliappan, "Edge implicit weighting with graph transformers for robust intrusion detection in Internet of Things network," *Comput. Security*, vol. 150, 2025.
- [26] S. Gopikumar, J. R. Banu, Y. H. Robinson, S. Raja, S. Vimal, D. Pelusi, and M. Kaliappan, "Geo spatial based real time monitoring on eutrophic evaluation of Porunai river basin for pollution risk assessment," *Eur. J. Remote Sens.*, pp. 1–13, 2022, doi: 10.1080/22797254.2021.2025152.
- [27] M. Kaliappan, E. Mariappan, and B. Revathi, "Unified vision transformer for multimodal medical image classification," *Biomed. Signal Process. Control*, art. 110431, 2026.
- [28] M. Kaliappan, E. Mariappan, M. V. Prakash, and B. Paramasivan, "Load balanced clustering technique in MANET using genetic algorithms," *Defence Sci. J.*, vol. 66, no. 3, pp. 251–258, 2016.