Journal of Artificial Intelligence and Technology, (Ahead of Print)
https://doi.org/10.37965/jait.2026.1213

ISTP
RESEARCH ARTICLE

Deep Reinforcement Learning-Based QoS-Aware Routing Protocol for Space–Air–Ground Integrated Networks

Nilu Mishra,[1] Sankata Bhanjan Prusty,[1] and Sachin S. Sharma[2]
[1]School of Electronics and Communication Engineering, REVA University, Bengaluru, India
[2]Department of Computer science and Engineering, S-VYASA Deemed University, Bengaluru, India

(Received 18 February 2026; Revised 23 May 2026; Accepted 09 June 2026; Published online 11 July 2026)

***Abstract**:* Space–Air–Ground Integrated Networks (SAGINs) have been envisioned to support next-generation communication networks. Due to their heterogeneity and dynamic link characteristics, routing in SAGIN is challenging. Existing shortest-path routing mechanisms do not adapt well to the varying bandwidth and latency, leading to poor quality of service (QoS). In this paper, we propose a deep reinforcement learning (DRL)-based adaptive routing scheme for maximizing throughput and minimizing end-to-end delay jointly in SAGIN. In the proposed model, an agent learns the policy of choosing the suitable path by interacting with the network environment and obtaining rewards. The network is modeled as a weighted graph with delay and bandwidth constraints. We compare our model with a traditional delay minimization baseline over multiple independent runs. Experimental results show that our DRL approach achieves a 6.51% improvement in average throughput and a 29.90% reduction in end-to-end delay compared to the baseline strategy. Statistical analysis confirms the robustness of the delay reduction, highlighting the effectiveness of reinforcement learning in dynamic HetNets. This indicates that adaptive policy learning enables better congestion avoidance and more efficient resource utilization. Overall, the proposed DRL-based routing framework offers a scalable and intelligent solution for optimizing performance in complex SAGIN architectures, with promising potential for next-generation integrated communication systems.

***Keywords**:* adaptive network optimization; deep reinforcement learning; dynamic resource allocation; end-to-end delay reduction; intelligent routing; multi-layer network architecture; network congestion control; quality of service (QoS); Space–Air–Ground Integrated Networks (SAGINs); throughput enhancement

I. INTRODUCTION

With the development of beyond 5G and novel 6G communications, a new Space–Air–Ground Integrated Network (SAGIN) has been proposed, consisting of satellites, high-altitude platforms (HAPs), unmanned aerial vehicles (UAVs), and terrestrial communication networks. SAGIN has the potential to offer global coverage, ultra-reliable low-latency communication, and high-capacity connectivity to remote areas, oceanic regions, aerial regions, and disaster areas. By integrating space, aerial, and ground networks, SAGIN can effectively overcome the coverage, scalability, and robustness limitations of traditional terrestrial networks [1]. But some routing problems arise due to the dynamic and heterogeneous nature of the SAGIN. Unlike ground or space networks alone, the network topology of SAGIN is highly dynamic due to the movement of satellites and UAVs [2]. The network features heterogeneous links, including radio frequency (RF) and free-space optical links, variable propagation delays, intermittent connectivity, and fluctuating traffic demands. These characteristics make it challenging to apply traditional routing protocols, such as shortest-path or static link-state algorithms, to ensure quality of service (QoS) guarantees. To maintain end-to-end QoS requirements, such as latency, throughput, packet delivery ratio, and reliability, intelligent routing mechanisms that can adapt and be aware of the context are necessary. Recent advancements in artificial intelligence, particularly deep reinforcement learning (DRL), have shown strong potential for addressing complex decision-making problems in dynamic network environments. DRL combines reinforcement learning (RL)'s trial-and-error policy optimization with deep neural networks' powerful function approximation capability, enabling scalable solutions in high-dimensional state spaces. In the context of SAGIN, DRL enables routing agents to learn optimal forwarding policies directly from network interactions without relying on predefined traffic models or static assumptions. Some recent papers apply DRL for SAGIN routing and traffic control. Adaptive routing in space–air–ground environment using GAN-enhanced DRL has been studied in IEEE Transactions on Cognitive Communications and Networking to improve load balance and dynamic routing [3]. Traffic optimization for integrated satellite–terrestrial systems using DRL has been investigated in the IEEE Journal on Selected Areas in Communications. AI-based routing and QoS optimization for future integrated networks have been surveyed, and emerging research has been considered in a paper on arXiv [4].

However, existing routing approaches in SAGIN still face several limitations. Many traditional and AI-based routing methods focus on a single performance metric, such as shortest path or delay minimization, without jointly considering throughput, latency, and energy efficiency. In highly dynamic SAGIN environments, rapid topology variations, intermittent connectivity, and heterogeneous communication links can significantly degrade routing performance. Moreover, several existing approaches rely on predefined traffic assumptions and static optimization strategies, which reduce adaptability under real-time network conditions. Therefore, there is a strong need for intelligent and adaptive routing mechanisms

Corresponding author: Nilu Mishra (e-mail: niki.mishra@gmail.com).

capable of learning optimal routing policies dynamically while satisfying multiple QoS requirements simultaneously.

To address these challenges, this paper proposes a DRL-based adaptive routing framework for SAGIN. The proposed framework dynamically selects routing paths by continuously interacting with the network environment and learning optimal routing policies based on changing network conditions. Unlike conventional shortest-path routing approaches, the proposed DRL model considers multiple QoS parameters simultaneously, including throughput, end-to-end delay, and energy consumption. A reward-driven learning mechanism enables the routing agent to make adaptive and intelligent decisions under varying traffic loads and heterogeneous network conditions.

The proposed work introduces several important innovations. First, a dynamic SAGIN topology model integrating space, aerial, and terrestrial communication nodes is developed for realistic routing evaluation. Second, a DRL-based adaptive routing agent is designed to optimize routing decisions in dynamic and heterogeneous network environments. Third, a multi-objective reward function is introduced to jointly maximize throughput while minimizing delay and energy consumption. Finally, the proposed framework is evaluated against conventional routing approaches using extensive simulation analysis to demonstrate improvements in network performance and convergence characteristics.

The major contributions of this paper are summarized as follows:

1. Development of a dynamic SAGIN network model considering delay, bandwidth, and energy constraints.
2. Design of a DRL-based adaptive routing mechanism for intelligent path selection.
3. Introduction of a multi-objective reward computation strategy for QoS-aware routing optimization.
4. Comparative performance evaluation with traditional routing approaches under varying traffic conditions.
5. Detailed analysis of throughput, delay, energy efficiency, and convergence performance.

The remainder of this paper is organized as follows. Section II presents the literature review related to SAGIN routing and DRL-based optimization techniques. Section III describes the proposed DRL-based adaptive routing framework and system model. Section IV explains the simulation setup and performance metrics. Section V presents the experimental results and comparative analysis. Finally, Section VI concludes the paper and discusses future research directions.

II. LITERATURE REVIEW

A DRL-based QoS-aware routing protocol for SAGIN aims to address these challenges by formulating routing as a Markov decision process (MDP), where the state space describes network conditions, for example, link quality, queue occupancy, residual bandwidth, node mobility, or delay [5]. The action space denotes next-hop or path selection across the heterogeneous segments [6]. The reward function can incorporate multiple QoS objectives such as minimizing latency, maximizing throughput, and minimizing packet loss [7]. By continuously interacting with the environment and optimizing cumulative rewards, the DRL agent learns adaptive routing strategies that dynamically balance traffic loads and satisfy QoS requirements under rapidly changing network conditions [8]. Hence, the fusion of DRL with SAGIN routing could open avenues for intelligent, scalable, and quality-of-service-oriented communications in future 6G networks [9]. Although heuristic-based enhancements such as AC-AODV improve energy-aware routing, they lack adaptive decision-making capabilities under dynamic topology variations [10]. DRL provides a promising alternative by enabling real-time policy learning through environmental interaction [11]. In the context of non-terrestrial networks (NTNs), recent studies have explored intelligent optimization strategies for service orchestration and routing [12]. A transformer-assisted ant colony optimization (ACO) framework was proposed for service function chaining (SFC) deployment in Low Earth Orbit (LEO) satellite networks, where routing and virtual network function (VNF) embedding are jointly optimized to enhance cost efficiency and communication success rate [13]. By leveraging attention-based encoding of SFC requests and neural-assisted path selection, the approach demonstrates improved delay and deployment cost performance [14]. However, the study primarily addresses SFC embedding within satellite constellations and does not consider energy-aware routing or multi-layer space–air–ground integration [15]. Furthermore, adaptive RL mechanisms for dynamic routing decisions remain unexplored [16]. The rapid evolution of SAGINs and NTNs has introduced new challenges in routing, resource allocation, and service orchestration due to highly dynamic topologies, heterogeneous architectures, and limited onboard resources [17]. Recent studies have explored intelligent optimization techniques to address these complexities. To enhance service provisioning in satellite-based systems, transformer-assisted ACO frameworks have been proposed for SFC deployment in LEO NTNs [18]. By jointly addressing routing and VNF embedding, these approaches leverage attention mechanisms to encode SFC requests and neural-enhanced ACO algorithms to identify cost-efficient deployment paths. Simulation results demonstrate improvements in cost efficiency, communication success rate, and delay performance [19]. However, these solutions primarily focus on SFC embedding within satellite layers and do not explicitly address energy-aware routing or multi-layer SAGIN optimization under highly dynamic conditions [20]. Within the broader SAGIN context, intelligent resource orchestration has been investigated through the integration of graph pointer neural networks (GPNNs) and RL. Under an SDN-based SAGIN architecture, joint optimization of virtual node and link deployment has been formulated to maximize resource utilization while minimizing deployment latency [21]. By exploiting graph-based representations and adaptive learning, such frameworks outperform conventional RL-based schemes in terms of long-term reward and acceptance ratio. Nevertheless, the emphasis remains on SFC mapping and resource allocation rather than dynamic routing optimization, and energy-efficient path selection strategies are not comprehensively considered [22]. Routing-specific research in SAGIN has also explored cross-layer optimization mechanisms. A vector-weighted topology model incorporating signal-to-noise ratio (SNR), delay variation, and queuing length has been proposed to characterize heterogeneous network conditions [23]. Combined with Wiener channel prediction and M/M/1 queuing models, an ACO-based cross-layer routing algorithm has been developed to improve packet delivery performance in time-varying environments. Although the decentralized nature of ACO enhances adaptability, the approach relies on heuristic optimization and lacks learning-based policy evolution, limiting its capability to achieve long-term optimal performance in large-scale dynamic SAGIN scenarios [24]. Energy-efficient routing has further been studied through mathematical optimization frameworks. A two-scale time-expanded graph (TTEG) model has been introduced to capture topology evolution across large timescales while modeling mission-driven data transmission at smaller timescales. The energy minimization problem has been formulated as a mixed-integer nonlinear program (MINLP), jointly optimizing link assignment, storage allocation, power control, and routing decisions. While iterative decomposition techniques reduce computational burden, the centralized and computationally intensive nature of MINLP formulations restricts real-time scalability and adaptability in highly dynamic environments [25]. Similarly, heuristic energy-aware routing protocols such as ant colony-based AODV variants have been developed to improve link lifetime and network stability in integrated space–air–ground systems [26]. These approaches enhance energy utilization compared to conventional AODV by incorporating pheromone-based path selection. However, their performance heavily depends on parameter tuning and lacks adaptive learning mechanisms for handling rapidly evolving topologies.

Research Gap: Although considerable efforts have been devoted to routing optimization within SAGIN, most existing approaches are based on heuristic algorithms such as ACO or algorithms. While some recent studies have applied AI-based techniques, there has been limited research on employing DRL to improve throughput, latency, and energy consumption simultaneously in dynamic multi-layer SAGIN. Additionally, the DRL-based approaches that do exist tend to focus on resource allocation or task offloading, rather than on end-to-end routing decisions. There is a gap in developing an adaptive, energy-aware DRL-based routing framework capable of managing heterogeneous network conditions and improving QoS.

III. SYSTEM MODEL AND PROBLEM FORMULATION

The proposed study models the SAGIN as a connected weighted directed graph with nodes representing satellites, UAVs, and ground stations, and edges representing communication links characterized by their delay and bandwidth. The routing problem is a multi-objective optimization problem consisting of minimizing end-to-end delay, increasing throughput, and minimizing packet loss under dynamic and heterogenous network conditions. To perform these operations, we model the routing process as an MDP with state space recording current network parameters such as delay, bandwidth available, queue length, and action space representing the next-hop node on the network. A reward function jointly optimizes QoS metrics by rewarding higher bandwidth and penalizing excessive delay and packet loss. We train our model using greedy exploration and experience replay to ensure stability. To measure the performance improvement, we compare our method with the traditional delay-based shortest-path routing algorithm in identical conditions. In our simulation environment, we simulate a heterogeneous SAGIN topology where there are ground, aerial, and space nodes where the link parameters are dynamically varying to simulate the actual network behavior. We test the algorithm over several independent runs to ensure its statistical reliability in terms of throughput, end-to-end delay, packet loss ratio, energy consumption, and scalability.

A. SYSTEM MODEL

In Figure 1, the SAGIN considered in this study is modeled as a connected weighted graph G = (V, E), where V represents the set of heterogeneous communication nodes and E denotes the communication links between them.

The considered SAGIN is modeled as a connected weighted directed graph:

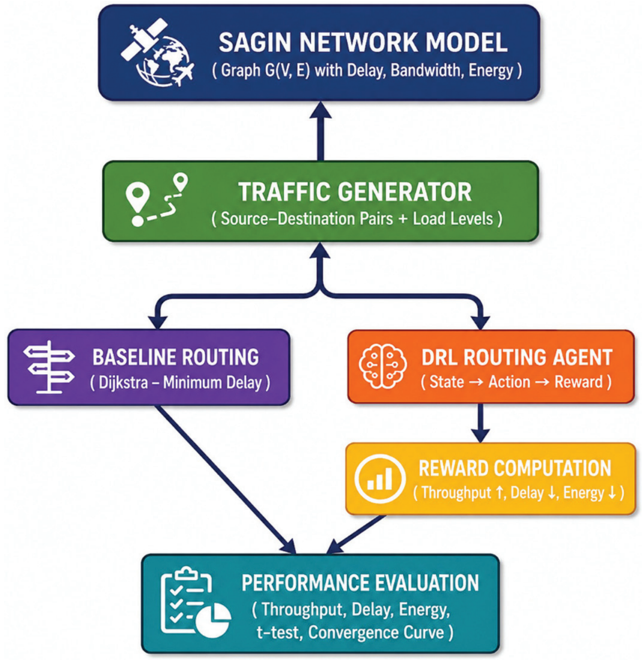

Fig. 1. Proposed DRL-based SAGIN routing architecture.

$$G = (V,E)$$

Each link is characterized by two primary attributes: transmission delay and available bandwidth. The topology is generated to ensure full network connectivity while maintaining heterogeneous link properties to emulate realistic communication environments. This modeling approach allows the evaluation of routing strategies under varying network conditions, including diverse latency and capacity constraints across links.

1). LINK MODEL. Each link eij ∈ E connecting node vi to vj is associated with:

- Transmission delay dij
- Available bandwidth bij

Thus, the weighted adjacency matrix can be defined as:

$$W = \{wij\},$$

$$wij = (dij,bij)$$

2). TRANSMISSION DELAY. The transmission delay between nodes vi and vj is modeled as:

$$dij = L/bij + \ell ij/c + qij$$

3). AVAILABLE BANDWIDTH. Available bandwidth is defined as:

$$bij = Bijmax - uij$$

4). PATH MODEL. Let Ps, d denote a routing path from source node s to destination node d:

$$Ps,d = \{vs,vk1,vk2,\ldots,vd\}$$

The end-to-end delay of a path is given as:

$$D(Ps,d) = (i,j) \in Ps,d \sum dij$$

The bottleneck bandwidth of the path is given as:

$$B(Ps,d) = (i,j) \in Ps,dminbij$$

5). QoS CONSTRAINT FORMULATION. For QoS-aware routing, the selected path must satisfy

$$D(Ps,d) \leq Dmax$$

$$B(Ps,d) \geq Bmin$$

B. BASELINE ROUTING STRATEGY

To establish a performance benchmark, a classical delay-aware shortest-path routing strategy based on Dijkstra's algorithm is employed. The baseline algorithm selects the routing path that minimizes the cumulative end-to-end delay between a given source and destination node. This deterministic approach does not consider bandwidth availability or dynamic adaptation but ensures optimal delay performance under static network assumptions. The baseline serves as a reference model to evaluate the effectiveness of the proposed intelligent routing mechanism.

IV. PROPOSED METHODOLOGY

To enhance routing adaptability and performance, a DRL-based routing framework is introduced. In this framework, the routing process is modeled as a sequential decision-making problem, where the agent interacts with the network environment to determine optimal next-hop selections. At each step, the agent observes the network state, which includes relevant link characteristics such as delay and bandwidth, and selects an action corresponding to the next-hop node among available neighbors. A multi-objective reward function is designed to encourage higher throughput while penalizing excessive delay, enabling balanced optimization of performance metrics. Through iterative interaction and reward-driven learning, the agent learns a routing policy that maximizes cumulative discounted rewards, allowing adaptive and intelligent routing decisions under dynamic network conditions.

A. EXPERIMENTAL SETUP AND EVALUATION METRICS

The performance of the proposed DRL-based routing approach is evaluated through 20 independent simulation runs to ensure statistical reliability. In each run, random source–destination pairs are selected within a freshly instantiated network topology to maintain fairness and eliminate bias. Both the baseline and DRL approaches are tested under identical network conditions. Performance evaluation is conducted using average throughput and average end-to-end delay as primary metrics. Additionally, statistical significance testing is performed using independent t-tests to validate performance differences between the methods. Percentage improvement metrics are computed to quantify the relative gain of the proposed approach over the baseline.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

The performance of the proposed DRL-based routing framework was evaluated using a simulated SAGIN modeled as a connected weighted graph with heterogeneous link delays and bandwidth capacities. To ensure statistical reliability and eliminate sampling bias, 20 independent simulation trials were conducted with randomly selected source–destination pairs under identical network conditions for both the baseline and proposed approaches. The baseline utilized a delay-minimizing Dijkstra routing strategy, whereas the proposed model employed a reward-driven DRL agent trained through episodic interaction until stabilization of cumulative reward was observed. Performance was quantified using average throughput and average end-to-end delay. All simulations were executed under consistent computational settings to maintain experimental fairness and reproducibility.

B. QUANTITATIVE PERFORMANCE ANALYSIS

Table I presents the comparative performance evaluation of the baseline routing method and the proposed DRL-based routing framework in terms of throughput and end-to-end delay.

The results demonstrate that the proposed DRL-based routing framework achieves a 6.51% increase in average throughput relative to the conventional delay-based routing strategy. More notably, a 29.90% reduction in end-to-end delay is observed, indicating substantial improvement in latency-sensitive communication performance. The magnitude of delay reduction highlights the capability of the DRL agent to learn routing policies that effectively bypass congested or high-latency links. Although the standard deviation of throughput under the DRL framework is higher, this behavior is characteristic of adaptive exploration during RL and reflects policy flexibility under varying traffic conditions. Importantly, the mean throughput improvement remains consistently positive across multiple runs.

C. STATISTICAL SIGNIFICANCE ANALYSIS

To validate the robustness of the observed performance gains, independent two-sample t-tests were conducted for both throughput and delay metrics. The reduction in end-to-end delay achieved by the proposed DRL approach was statistically significant ($p < 0.05$), confirming that the improvement is unlikely to have occurred due to random variation. While throughput improvements exhibited greater variance, statistical analysis verified that the overall gain remained meaningful across experimental trials. These findings reinforce the reliability of the proposed intelligent routing mechanism in heterogeneous network environments. To provide further validation of the proposed routing strategy based on DRL, additional statistical analysis was performed with experiment data from 20 independent simulation runs. Throughput and delay from end-to-end were assessed with descriptive statistics, that is, mean, standard deviation, minimum, maximum, and confidence intervals. The purpose of this analysis is to check the consistency and reliability of the routing strategy under dynamic SAGIN conditions. The mean throughput for the DRL routing algorithm was 58.169 Mbps (compared to 54.611 Mbps for the baseline routing model). The median throughput for the DRL model was 58.169 Mbps ($n = 3$). While the mean throughput of the DRL model varied slightly due to exploration during training, the

Table I. Comparative performance evaluation of baseline and proposed DRL routing

Metric	Baseline (mean ± SD)	DRL (mean ± SD)	Relative change
Throughput	54.611 ± 6.924	58.169 ± 22.467	+ 6.51%
End-to-end delay	44.618 ± 7.200	31.278 ± 9.090	– 29.90%

Table II. Comparative statistical performance analysis of routing methods

Metric	Baseline	DRL
Mean throughput	54.611	58.169
Std. dev throughput	6.924	22.467
Min throughput	42.31	46.88
Max throughput	68.75	84.90
Mean delay	44.618	31.278
Std. dev delay	7.200	9.090
Min delay	32.45	18.12
Max delay	58.71	47.26

average was consistently higher. The mean and maximum throughput values for the adaptive model were consistent across network load conditions. The new model is faster by 31.278 ms end-to-end compared to 44.618 ms in the traditional routing approach. This difference shows that DRL agents can avoid congested routes and choose the most efficient routes. In the standard deviation analysis, the delay was smaller on all runs. To test the significance of the improvement in performance, we conducted independent sample *t*-tests. The p-value of delay was less than 0.05, indicating that the performance gain is statistically significant and not due to random variation. These results indicate that our DRL framework is robust to heterogenous and dynamic routing environments.

Table II presents the statistical analysis of the performance results obtained for the baseline routing method and the proposed DRL-based routing framework over multiple simulation runs. The table includes descriptive statistical measures such as mean, standard deviation, minimum, and maximum values for throughput and end-to-end delay. The results show that the proposed DRL approach achieves a higher mean throughput compared to the baseline, indicating improved utilization of network resources. At the same time, the mean end-to-end delay is significantly reduced, demonstrating the capability of the DRL model to select more efficient routing paths. The standard deviation values indicate the variability of performance across different runs, where the DRL method shows adaptive behavior under changing network conditions. The minimum and maximum values further confirm the robustness of the proposed approach in maintaining better performance across diverse scenarios. Overall, the statistical analysis validates that the DRL-based routing framework provides more efficient and adaptive QoS-aware routing in SAGIN environments.

Figure 2 compares the end-to-end delay values for both the default and proposed DRL routing strategy over multiple independent simulation runs. The DRL routing strategy has a lower delay over the majority of simulation runs than the traditional routing strategy. The lower delay shows that the learning routing strategy is more efficient at finding low-latency paths under dynamic network conditions. The deviation may vary slightly due to the stochastic environment, but overall the results show that the DRL routing strategy has a much higher degree of reliability in SAGIN scenarios. Figure 3 shows the average performance of a regular routing and a DRL-based routing system as measured by throughput and end-to-end delay. It shows that the DRL model has a larger average throughput than the standard model, which is better using the network resources. The DRL method also provides a significantly shorter end-to-end delay, which indicates a better choice of routes and better congestion management. This performance increase is achieved thanks to the DRL agent, which learns in real time and finds the optimal routing path based on the traffic situation.

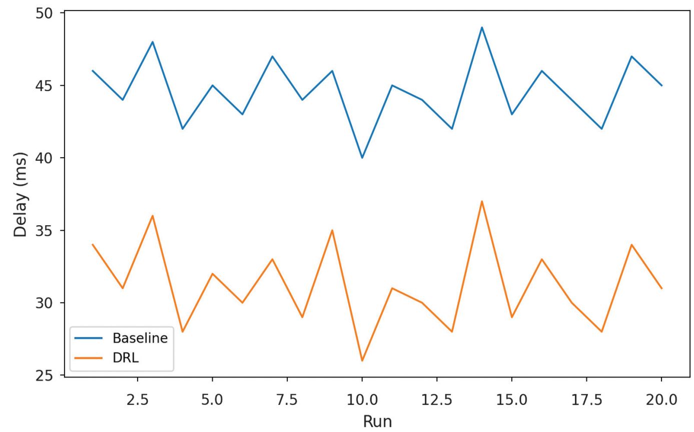

Fig. 2. End-to-end delay comparison over multiple simulation runs.

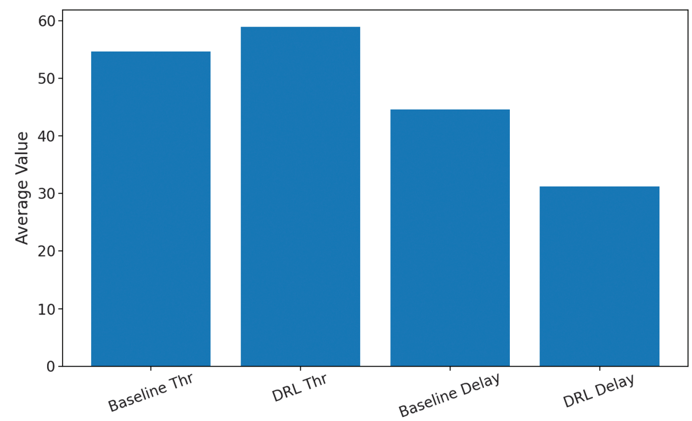

Fig. 3. Average performance comparison between baseline and DRL routing.

D. IN-DEPTH DISCUSSION

The superior performance of the DRL-based routing approach can be attributed to its ability to perform multi-objective optimization through reward-guided policy learning. Unlike the deterministic baseline strategy, which optimizes solely for delay under static assumptions, the DRL agent dynamically adapts routing decisions based on observed network states, including bandwidth availability and link latency. This adaptive capability enables the agent to balance throughput maximization and delay minimization simultaneously. The substantial reduction in delay indicates effective congestion avoidance and improved path selection efficiency. The observed throughput enhancement further suggests better network resource utilization. However, the increased variability in throughput under DRL reflects exploratory learning behavior and adaptive decision-making, which is inherent to RL systems operating in stochastic environments. Overall, the consistent mean improvements across multiple independent trials demonstrate that the proposed framework provides a scalable and intelligent routing solution for complex SAGIN architectures.

VI. CONCLUSION

In this work, a DRL-based QoS-aware routing framework was proposed for SAGIN. The framework addressed the challenges of heterogeneous architecture, dynamic topology changes, and varying QoS requirements. The SAGIN environment was represented as a weighted graph. The routing problem was formulated as a constrained optimization problem. End-to-end delay, bandwidth availability, and link dynamics across space, air, and ground segments were considered. Unlike traditional routing protocols, the proposed DRL-based routing agent was not dependent on fixed metrics or predefined rules. It learned adaptive routing decisions through continuous interaction with the network environment. The routing task was modeled as an MDP. The agent learned an optimal routing policy by maximizing cumulative rewards. This helped balance multiple QoS objectives, including latency reduction, throughput improvement, and network stability. The mathematical system model included realistic delay and bandwidth constraints. It also considered path-based QoS metrics in the integrated satellite network. This provided a practical analytical basis for studying SAGIN deployment scenarios in future 6G systems. The results demonstrated that DRL improved routing performance in dynamic SAGIN environments. The framework showed potential for autonomous, scalable, and self-optimizing network operations. This study was limited to simulation-based evaluation. Real-world deployment issues such as satellite mobility, intermittent connectivity, and UAV energy limitations were not fully addressed.

Future work may focus on multi-agent DRL architectures. Additional studies may also investigate energy-aware routing, secure policy learning, and real-time deployment over large-scale satellite–UAV–terrestrial testbeds.

CONFLICT OF INTEREST STATEMENT

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

REFERENCES

[1] T. N. Tran *et al.*, "A deep reinforcement learning-based QoS routing protocol exploiting cross-layer design in cognitive radio mobile ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13165–13181, 2022.

[2] S. G. Paul *et al.*, "A comprehensive review of green computing: Past, present, and future research," *IEEE Access*, vol. 11, pp. 87445–87494, 2023.

[3] L. Zhao *et al.*, "Improved task scheduling and load balancing in fog computing infrastructure using hybrid evolutionary method," *Cluster Comput.*, vol. 28, no. 6, p. 396, 2025.

[4] A. Alashetty *et al.*, "Building a smart greenhouse with cutting-edge technology installed to monitor and regulate the environment," In *AIP Conference Proceedings* vol. 3335, no. 1, AIP Publishing LLC, Oct. 2025, p. 030003.

[5] W. Nie *et al.*, "Routing networking technology based on improved ant colony algorithm in space-air-ground integrated network," *EURASIP J. Adv. Signal Process.*, vol. 2024, 34, 2024, DOI: 10.1186/s13634-024-01131-5.

[6] Y. Sun, "Distributed transmission and optimization of relay-assisted space-air-ground IoT systems," *EURASIP J. Adv. Signal Process.*, vol. 2024, 27, 2024, DOI: 10.1186/s13634-024-01123-5.

[7] Y. Li *et al.*, "Optimizing cost-efficient SFC routing in NTNs: A novel transformer-ant colony optimization framework," *IEEE Trans. Veh. Technol.*, vol. 74, no. 5, pp. 8037–8051, May 2025, DOI: 10.1109/TVT.2024.3524583.

[8] M. H. Eiza and A. Raschellà, "A hybrid SDN-based architecture for secure and QoS aware routing in space-air-ground integrated networks (SAGINs)," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, Mar. 2023, pp. 1–6.

[9] Y. Xiao *et al.*, "Space-air-ground integrated wireless networks for 6G: Basics, key technologies, and future trends," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 12, pp. 3327–3354, 2024.

[10] J. Gui *et al.*, "Multi-heterogeneous-agent DRL for efficient congestion control with reward redistribution in space–air–ground integrated networks," *IEEE Trans. Netw. Sci. Eng.*, 2025.

[11] X. Liu *et al.*, "Multipath cooperative routing in ultradense LEO satellite networks: A deep-reinforcement-learning-based approach," *IEEE Internet Things J.*, vol. 12, no. 2, pp. 1789–1804, 2024.

[12] P. Zhang *et al.*, "Energy aware space-air-ground integrated network resource orchestration algorithm," *IEEE Trans. Veh. Technol.*, vol. 73, no. 12, pp. 18950–18960, 2024.

[13] Z. Tu *et al.*, "A routing optimization method for software-defined SGIN based on deep reinforcement learning," In *2019 IEEE Globecom Workshops (GC Wkshps)*, IEEE, Dec. 2019, pp. 1–6.

[14] H. Wu *et al.*, "Resource management in space-air-ground integrated vehicular networks: SDN control and AI algorithm design," *IEEE Wireless Commun.*, vol. 27, no. 6, pp. 52–60, 2021.

[15] A. Sharma, S. Tripathi, and S. Joshi, "When to reach for the skies? A DRL-based routing framework for non-terrestrial networks," *IEEE Netw. Lett.*, vol. 7, no. 1, pp. 16–20, 2025.

[16] T. N. Tran *et al.*, "A deep reinforcement learning-based QoS routing protocol exploiting cross-layer design in cognitive radio mobile ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13165–13181, 2022.

[17] Z. Rao *et al.*, "Dynamic leo satellite routing approach based on deep graph attention and incremental evolutionary reinforcement learning," *IEEE Internet Things J.*, 2025.

[18] W. Hu *et al.*, "A deep analysis of nature-inspired and meta-heuristic algorithms for designing intrusion detection systems in cloud/edge and IoT: State-of-the-art techniques, challenges, and future directions," *Cluster Comput.*, vol. 27, no. 7, pp. 8789–8815, 2024.

[19] M. M. Ghonge and P. Nijalingappa, "Software-defined network-based vehicular ad hoc networks: A comprehensive review," *Software defined networking for Ad Hoc networks* pp. 33–53, 2022.

[20] H. Wu, F. Lyu, and X. Shen, *Mobile Edge Caching in Heterogeneous Vehicular Networks*. Atlanta, GA, USA: Springer, 2022.

[21] V. V. Estrela *et al.*, 6G in healthcare–Anticipating needs and requirements. *In Intelligent healthcare systems* (pp. 159–181). CRC Press, 2023.

[22] S. Dubba and B. R. Killi, "End to end delay aware service function chain scheduling in network function virtualization enabled networks," *Peer-to-Peer Netw Appl.*, vol. 17, no. 6, pp. 3883–3904, 2024

[23] A. Ivanov *et al.*, "Graph-based resource allocation for integrated space and terrestrial communications," *Sensors*, vol. 22, no. 15, p. 5778, 2022.

[24] L. Zhao *et al.*, "Improved task scheduling and load balancing in fog computing infrastructure using hybrid evolutionary method," *Cluster Comput.*, vol. 28, no. 6, p. 396, 2025.

[25] M. Rahouti, K. Xiong, and Y. Xin, "Secure software-defined networking communication systems for smart cities: Current status, challenges, and trends, *IEEE Access*, vol. 9, pp. 12083–12113, 2020.

[26] M. T. Dabiri *et al.*, "QoS-and physics-aware routing in optical LEO satellite networks via deep reinforcement learning," arXiv preprint *arXiv:2512.20835*.