IET Communications

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Adaptive Satellite Selection via Deep Reinforcement Learning for Dynamic Emergency Scenarios

Ke Chen | Li Zhang | Qihang Jiao | Samuel Krain

School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK

Correspondence: Li Zhang (l.x.zhang@leeds.ac.uk)

Received: 20 November 2025 | **Revised:** 13 May 2026 | **Accepted:** 21 May 2026

ABSTRACT

Disasters often lead to severe disruptions of terrestrial backbone networks, resulting in information isolation in affected areas. In such situations, very small aperture terminals (VSATs) can be rapidly deployed to establish terrestrial–satellite communication links, enabling the transmission of large volumes of situational awareness data, such as high-definition videos and three-dimensional maps. However, the high mobility of low earth orbit (LEO) satellites necessitates frequent satellite switching to maintain service continuity. This challenge is further exacerbated in emergency scenarios, where the timely and reliable transmission of post-disaster data is of critical importance, while unpredictable disturbances can severely degrade the stability of ground–satellite backhaul links through abrupt channel variations and resource fluctuations. Consequently, robust and efficient satellite selection is crucial for sustaining reliable and high-throughput satellite backhaul. To address these challenges, this paper proposes a disturbance-aware deep reinforcement learning (DRL) framework for adaptive satellite selection, aiming to support robust and high-capacity terrestrial–satellite backhaul for VSATs in disaster scenarios. Specifically, the proposed approach jointly accounts for link quality, handover (HO) frequency, and connection failures in the decision-making process. Moreover, the robustness and stability of the DRL model is improved by incorporating the stochastic disturbances encountered in LEO satellite networks into the training stage. These disturbances comprise weather-induced carrier-to-noise ratio (CNR) degradation and reductions in available satellite bandwidth caused by user contention. Simulation results demonstrate that the proposed approach achieves robust and stable performance under dynamic conditions, outperforming existing methods.

1 | Introduction

In disaster scenarios, terrestrial backbone networks are often severely damaged, leading to the emergence of information-isolated regions [1–4]. In this case, very small aperture terminals (VSATs) can be employed to establish terrestrial-satellite wireless backhaul links, facilitating the transmission of substantial volumes of data from these regions to the outside world [5]. Such data may include high-definition video, three-dimensional maps, and real-time disaster monitoring information [6]. However, the high mobility of low earth orbit (LEO) satellites and the dynamic propagation environment pose inherent challenges to satellite selection, as ground–satellite channels become highly time-variant [7, 8]. In particular, these challenges are further exacerbated by unpredictable disturbances within the satellite constellation, such as sudden reductions in available bandwidth due to traffic contention from users in other regions, as well as degradations in the carrier-to-noise ratio (CNR) of ground–satellite links caused by adverse weather conditions [5]. To address these challenges, VSATs must frequently switch connections to different satellites in the network in order to maintain communication links, that is, perform handovers (HOs). Collectively, these factors hinder the establishment of reliable and high-throughput terrestrial–satellite backhaul connections, consequently impeding the efficient execution of post-disaster search and rescue (SAR) missions.

1.1 | Related Work

Currently, satellite selection has been widely studied in LEO satellite networks. For instance, the authors in [9] proposed a multi-attribute decision-making (MADM)-based satellite selection scheme that reduced unnecessary HOs while ensuring balanced resource utilization. For long-term satellite selection optimization, a graph-based framework was proposed in [10], in which satellite HO was modeled as a shortest-path problem on visibility graphs, an approach which could accommodate multiple HO criteria simultaneously. Building on this, quality-of-service (QoS)-related utility functions (e.g., data rate and delay) were incorporated into edge weights in the graph-based framework in [11], which led to improved QoS compared to conventional elevation-threshold schemes. Similarly, [12] proposed a weighted graph-based satellite selection strategy for aeronautical traffic in LEO satellite networks. In this method, an improved MADM algorithm with a channel reservation mechanism was introduced to enhance throughput and reduce connection failures.

Alternatively, LEO-satellite-assisted unmanned aerial vehicle (UAV) networks were studied in [13], where a joint UAV trajectory and satellite selection optimization problem was formulated. In this framework, satellite selection was essential for ensuring reliable backhaul in delay-sensitive internet of remote things (IoRT) applications. In emergency scenarios, the study in [14] addressed joint UAV trajectory planning and LEO satellite selection by formulating a two-stage multi-armed bandit problem and developing lightweight online algorithms that significantly improved energy efficiency and reduced HOs. Additionally, a lightweight Dijkstra-based satellite selection method was proposed in [6] to facilitate the backhaul transmission of large volumes of data in emergency scenarios. In this method, the uplink data rate, latency, and HO frequency were jointly considered, and an effective trade-off among these key performance metrics was achieved.

In addition, the study in [15] proposed a dynamic time-based channel reservation (DTCR) algorithm tailored for LEO satellite selection. By reserving channels and dynamically adjusting the reservation timing, the DTCR scheme achieved an effective balance between blocking probability and forced termination probability, thereby enhancing the overall efficiency of channel utilization. Building upon this, a joint optimization-based satellite HO strategy was proposed in [16], aiming to simultaneously minimize HO frequency, balance network load, and perform adaptive power allocation. In this approach, graph-based handover path optimization was integrated with gradient-based power control, resulting in lower switching frequency, improved load distribution, and higher overall system capacity. To address interference during satellite HO, the authors in [17] investigated spectrum coexistence between gateway and user terminal links in non-geostationary satellite orbit (NGSO) systems. The proposed HO mechanisms effectively mitigated interference and enhanced spectral efficiency in shared-spectrum environments.

In addition to these static optimization methods, deep reinforcement learning (DRL) has emerged as a widely adopted approach for satellite selection. For example, the authors in [18] formulated the satellite selection problem as a Markov decision process, in which a deep Q-network (DQN) was employed to jointly account for CNR, interference, remaining service time, and satellite load, thereby reducing HO frequency. To improve service continuity in LEO satellite networks, a DQN was integrated with a conditional handover (CHO) mechanism in [19]. In this framework, the DQN was employed to select the target satellite by jointly considering satellite elevation angle, distance, and CNR, thereby improving mobility management of the satellites.

Moreover, architectural and replay-based variants of the DQN framework have been investigated in the context of satellite selection. In [20], a Dueling DQN architecture was employed for LEO satellite HO. The Q-function was decomposed into state-value and action-advantage components to improve value estimation when the CNR differences among candidate satellites were relatively small. This design resulted in higher long-term cumulative rewards. The study in [21] proposed a PER-DDQN-based satellite selection strategy that combines Double DQN with prioritized experience replay. In this approach, satellite load information was incorporated into the state representation and reward formulation to support load-aware and QoS-aware HO decisions under high user density conditions. For multi-user satellite selection, a multi-agent DQN-based selection mechanism was developed in [22]. Specifically, each user was modeled as an agent, and distributed decision-making was adopted to support QoS-aware and load-balanced HO in multi-user LEO satellite networks.

1.2 | Motivation

Despite recent advances in satellite selection strategies, several critical gaps remain unaddressed, particularly for emergency satellite backhaul scenarios that are highly sensitive to dynamic variations and unpredictable disturbances.

- **Limitations of static optimization methods in dynamic environments**: Satellite selection methods based on MADM and graph optimization are typically formulated in a static manner. When environmental conditions vary, particularly under CNR degradation or reductions in available bandwidth, frequent re-optimization is required. This leads to increased response latency and computational overhead, which limits their suitability for highly dynamic, resource-constrained, and disturbance-sensitive emergency scenarios.
- **Insufficient focus on VSAT-based emergency satellite backhaul**: Existing studies have devoted limited attention to VSAT-based satellite selection for emergency backhaul, where stable and high-throughput terrestrial–satellite links are required to support the transmission of situational awareness information.
- **Lack of disturbance-aware learning under emergency scenarios**: DRL-based satellite selection methods exhibit enhanced adaptability. However, existing approaches are typically trained under stable channel assumptions, with limited consideration of stochastic disturbances such as bandwidth contention or weather-induced CNR degradation. Emergency backhaul scenarios are highly sensitive to these disturbances, requiring adaptive and robust learning mechanisms.

These gaps motivate the development of a disturbance-aware learning framework that accounts for stochastic channel and resource variations during training, thereby supporting robust and efficient satellite selection for emergency LEO satellite backhaul in the presence of stochastic disturbance.

1.3 | Contribution

This work focuses on robust satellite selection for VSAT-based terrestrial–satellite backhaul in emergency scenarios, where unpredictable bandwidth contention and weather-induced CNR degradation are frequently encountered. Specifically, a disturbance-aware DRL framework is proposed based on a DQN. The framework adopts a reward function that jointly accounts for link quality, HO frequency, and connection failures. In addition, stochastic disturbances are incorporated into the training process across different satellites and time slots, improving robustness and adaptability in highly dynamic LEO satellite environments. Extensive numerical simulations are conducted to evaluate the effectiveness of the proposed approach. The main contributions of this work are summarized as follows:

- A realistic LEO satellite constellation is established using systems tool kit (STK). This facilitates the accurate evaluation of the proposed HO method under practical orbital dynamics and ground–satellite communication conditions.
- A DQN-based satellite selection framework is developed for VSAT-based backhaul, where ground-to-satellite link quality, HO frequency, and connection failures are jointly considered in the reward function.
- To address the disturbance-sensitive nature of emergency backhaul scenarios, stochastic impairments encountered in LEO satellite networks are incorporated into the training process. These impairments include weather-induced CNR degradation and fluctuations in available satellite bandwidth. By modeling these uncertainties during training, the robustness and adaptability of the learned policy are enhanced.
- Numerical simulations are conducted to evaluate the proposed method through comparisons with representative MADM, graph-based, and proximal policy optimization (PPO) approaches. An ablation study is also conducted to assess the impact of CNR degradation and bandwidth reduction, as well as the performance when these factors are excluded.

The remainder of this paper is organized as follows. In Section 2, the system model and problem formulation are presented. Section 3 is devoted to the detailed description of the proposed DQN-based satellite selection method. In Section 4, the simulation setup and performance results are discussed. Finally, Section 5 concludes the paper.

2 | System Model and Problem Formulation

As illustrated in Figure 1, we consider an emergency communication scenario in which the terrestrial backbone network is damaged. In this context, a VSAT establishes ground-to-satellite wireless backhaul links, enabling information-isolated regions in disaster-affected areas to connect with the outside world. Due to the high mobility of LEO satellites, the set of satellites visible to the VSAT varies over discrete time slots t, with $t \in \{1, 2, \dots, T\}$. Specifically, at each time slot t, the VSAT has access to a set of candidate satellites, denoted by $\mathcal{S}_t = \{1, 2, \dots, N_t\}$, where $N_t = |\mathcal{S}_t|$ represents the number of visible satellites at time t. Due to the heterogeneity of available resources across satellites in the constellation, each satellite $i \in \mathcal{S}_t$ has different available bandwidth and computing frequency. In addition, the CNR, propagation delay, and remaining service time (RST) of these satellites vary dynamically across different time slots t. These factors pose significant challenges to satellite selection for establishing reliable and high-throughput ground-to-satellite links, thereby ensuring robust data backhaul in emergency scenarios.

The quality of the ground-to-satellite wireless link between a LEO satellite i and the VSAT at time slot t is characterized by the CNR. Specifically, the received signal power can be expressed as

$$G_i^t = P_T^i + G_T - L_T - PL_i^t + G_R, \tag{1}$$

where P_T^i is the transmission power of satellite i, G_T and G_R are the transmitter and receiver antenna gains, respectively, L_T is the transmitter feeder loss, and PL_i^t is the propagation path loss between satellite i and the VSAT at time slot t. Notably, the term $P_T^i + G_T - L_T$ corresponds to the effective isotropic radiated power (EIRP).

Due to the line-of-sight conditions of ground-to-space links, the propagation loss PL_i^t is characterized by the free-space path loss (FSPL) model as

$$PL_i^t = 20 \log_{10}\left(\frac{4\pi d_i^t}{\lambda}\right), \tag{2}$$

where d_i^t denotes the distance between satellite i and the VSAT at time slot t, and $\lambda = c/f$ is the carrier wavelength corresponding to the satellite carrier frequency f.

Moreover, the noise power at the VSAT is given by

$$N_i^t = K + F_E + \mathrm{BW}_i^t, \tag{3}$$

where $K = -228.6$ dB is the Boltzmann constant, F_E denotes the equivalent noise temperature at the receiver, and BW_i^t is the bandwidth allocated to the VSAT by satellite i at time slot t. Accordingly, the CNR of the ground-to-satellite link is calculated as

$$\mathrm{CNR}_i^t = G_i^t - N_i^t. \tag{4}$$

In this work, five attributes are considered to characterize link quality in the satellite selection process. Hence, each satellite $i \in \mathcal{S}_t$ at time slot t is represented by a vector of time-varying attributes, as follows:

$$\boldsymbol{\phi}_i^t = \left(\mathrm{CNR}_i^t,\ \mathrm{BW}_i^t,\ \mathrm{CF}_i^t,\ d_i^t,\ \tau_i^t\right), \tag{5}$$

where BW_i^t and CF_i^t denote the available bandwidth and onboard computing frequency of satellite i, respectively. CNR_i^t and d_i^t

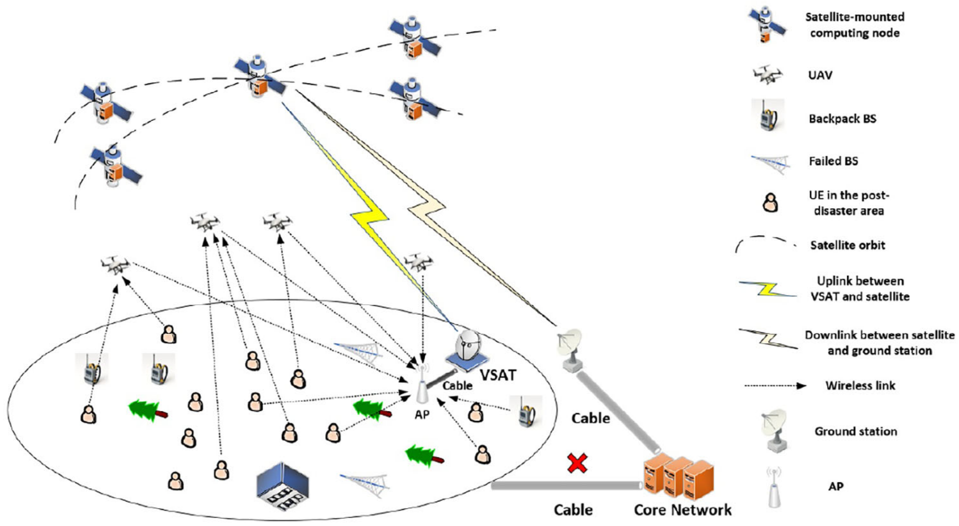

FIGURE 1 | Schematic diagram of satellite selection in the emergency scenario.

denote the carrier-to-noise ratio and the propagation delay if the VSAT selects satellite i for connection. Importantly, the CNR serves as a key indicator of the ground-to-satellite link condition, with larger values corresponding to better signal quality. Finally, τ_i^t corresponds to the RST before satellite i moves out of the VSAT's coverage.

At each time slot t, the data rate of the VSAT connection to satellite i is evaluated against a minimum threshold R_θ. If the data rate falls below this threshold, the connection is considered failed. Mathematically, this can be expressed as:

$$C_{i,t} = \begin{cases} 1, & \text{if } \mathrm{BW}_i^t \cdot \log_2\left(1 + \mathrm{CNR}_i^t\right) < R_\theta, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where $C_{i,t}$ is an indicator variable that equals 1 when a connection failure occurs, and 0 otherwise.

In emergency communication scenarios, frequent HOs can incur substantial signaling overhead, while excessive connection failures may severely undermine the reliability of transmitting critical data (e.g., high-definition video and medical information). In light of these considerations, the optimization problem is formulated to simultaneously optimize for three objectives: maximizing link quality, minimizing HO frequency, and reducing connection failures. It can be expressed as follows:

$$\begin{aligned} \max_{\{x_{i,t}\}} \quad & \frac{1}{T}\sum_{t=1}^{T}\sum_{i\in S_t} Q_{i,t}\, x_{i,t} \\ & - \frac{1}{T-1}\sum_{t=2}^{T}\sum_{i\in S_t}\sum_{j\in S_{t-1}} (1-\delta_{i,j})\, x_{i,t}\, x_{j,t-1} \qquad (7) \\ & - \frac{1}{T}\sum_{t=1}^{T}\sum_{i\in S_t} C_{i,t}\, x_{i,t} \\ \text{s.t.} \quad & \sum_{i\in S_t} x_{i,t} = 1, \quad \forall t \in \{1,2,\dots,T\}, \qquad (7.1) \\ & x_{i,t} \in \{0,1\}, \quad \forall i, t, \qquad (7.2) \end{aligned}$$

$$\delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \tag{7.3}$$

where the first term represents the average link quality, with $Q_{i,t} \in [0,1]$ denoting the normalized quality score of the VSAT's connection to satellite i at time t. The second term penalizes HOs, with a penalty incurred whenever the selected satellite changes between consecutive time slots. The third term penalizes connection failures. The binary decision variable $x_{i,t} \in \{0,1\}$ indicates whether satellite i is selected at time slot t (1 if selected, 0 otherwise), while the constraint in (7.1) guarantees that exactly one satellite is selected at each time slot.

More specifically, the link quality score $Q_{i,t}$ is derived from the attribute vector ϕ_i^t using min-max normalization followed by weighted aggregation, as shown below:

$$\begin{aligned} Q_{i,t} = {} & \lambda_1\, \mathrm{CNR}_{i,t}^{\mathrm{norm}} + \lambda_2\, \mathrm{BW}_{i,t}^{\mathrm{norm}} + \lambda_3\, \mathrm{CF}_{i,t}^{\mathrm{norm}} \\ & - \lambda_4\, d_{i,t}^{\mathrm{norm}} + \lambda_5\, \tau_{i,t}^{\mathrm{norm}}. \end{aligned} \tag{8}$$

Here, $(\cdot)^{\mathrm{norm}}$ denotes the normalized value of each attribute, and $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5]^{\mathrm{T}}$ represents the set of importance weights determined by the analytic hierarchy process (AHP) [23].

In particular, the relative importance of the five link attributes is quantified using Saaty's 1–9 scale [24]. Accordingly, a pairwise comparison matrix $\mathbf{A} \in \mathbb{R}^{5\times 5}$ is constructed as

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{15} \\ a_{21} & 1 & a_{23} & \cdots & a_{25} \\ a_{31} & a_{32} & 1 & \cdots & a_{35} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{51} & a_{52} & a_{53} & \cdots & 1 \end{bmatrix}, \tag{9}$$

where a_{mn} denotes the relative importance of the mth attribute over the nth attribute, with $a_{mn} = 1/a_{nm}$. Importantly, the elements of $\mathbf{A}$ are assigned based on Saaty's importance scale, where larger values indicate stronger preferences.

The weight vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5]^\top$ is obtained by solving the principal eigenvalue problem of $\mathbf{A}$, that is

$$\mathbf{A}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\\ \lambda_4\\ \lambda_5\end{bmatrix} = \lambda_{\max}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\\ \lambda_4\\ \lambda_5\end{bmatrix}, \tag{10}$$

where $\lambda_{\max}$ denotes the maximum eigenvalue of $\mathbf{A}$.

To verify the rationality of the pairwise comparisons, the consistency index (CI) and consistency ratio (CR) are computed as

$$\mathrm{CI} = \frac{\lambda_{\max} - N}{N - 1}, \quad \mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}}, \tag{11}$$

where $N = 5$ denotes the number of attributes and RI represents the random consistency index. When the CR is below 0.1, the constructed pairwise comparison matrix satisfies the consistency requirement, thereby yielding a reliable weight vector.

Since the ground-to-satellite wireless backhaul link is designed to replace the damaged terrestrial backbone network in emergency scenarios, stringent requirements are imposed on uplink throughput and latency. Consequently, attributes related to the CNR, available bandwidth, and propagation delay are assigned higher importance weights than other metrics. Notably, a larger value of $Q_{i,t}$ indicates better link quality, which in turn leads to a higher overall utility.

3 | The Proposed Deep Reinforcement Learning Based Satellite Selection Scheme

The optimization problem in (7) can be solved using conventional approaches, such as greedy heuristics and graph-based algorithms. Nevertheless, these methods typically require re-optimization whenever the satellite constellation experiences unpredictable disturbances, such as sudden bandwidth and CNR degradation. Furthermore, their strong dependence on up-to-date ephemeris and resource information severely limits their adaptability in highly dynamic and uncertain environments. To overcome these limitations, this paper adopts a DQN-based approach. Unlike policy-gradient algorithms under the Actor–Critic framework, such as PPO, which are typically designed for continuous action spaces [25], the satellite selection problem considered in this study involves a finite and discrete set of candidate satellites. This favours a DQN framework, which can learn the state–action value function to efficiently determine the optimal satellite under varying environmental conditions. Furthermore, the DQN framework offers improved training stability and reduced computational complexity, which are particularly advantageous in resource-constrained emergency scenarios. By leveraging its strong generalization capability, a DQN can effectively adapt to dynamic network conditions [26], thereby enhancing the reliability and robustness of terrestrial-satellite connectivity.

Importantly, the DQN is a value-based reinforcement learning algorithm that approximates the action–value function using a neural network [26]. To enhance training stability and efficiency, DQN incorporates experience replay and a target network. In this work, the agent (i.e., the VSAT) observes the environment state at each time step t, denoted by o_t, which is defined as the collection of attribute vectors of all visible satellites:

$$o_t = \{\phi_i^t \mid i \in \mathcal{S}_t\}. \tag{12}$$

The action a_t corresponds to selecting one satellite from the candidate set $\mathcal{S}_t$, that is, $a_t \in \mathcal{S}_t$. The agent adopts an ϵ-greedy exploration strategy and receives a reward, r, that reflects the performance of the selected terrestrial-satellite link. To maintain link quality, avoid excessive HOs, and penalize connection failures, the reward function is formulated as follows:

$$r_t = \begin{cases} -\zeta & \text{if } C_{a_t,t} = 1,\\ \alpha\, Q_{a_t,t} - \beta + \eta & \text{if } C_{a_t,t} = 0 \ \wedge\ a_t \neq a_{t-1},\\ \alpha\, Q_{a_t,t} + \eta & \text{if } C_{a_t,t} = 0 \ \wedge\ a_t = a_{t-1}. \end{cases} \tag{13}$$

Here, the coefficients ζ, β, and α, respectively, represent the penalty weights for connection failures and handovers, as well as the reward weight for link quality. These coefficients directly correspond to the three components of the optimization objective defined in (7). Note that η acts as an offset to mitigate the reduction in reward that may occur after HOs due to normalization. These parameters assist the agent in learning effective strategies, thereby enabling more efficient satellite selection.

The environment state is encoded as the collection of attribute vectors of all visible satellites. All satellite link attributes are normalized using min–max normalization before being input to the DQN. This process scales the features to a consistent numerical range, thereby improving training stability and efficiency.

The normalized state is then used as the input to the DQN. Each output neuron corresponds to a candidate satellite selection action in $\mathcal{S}_t$, and its output represents the estimated Q-value associated with selecting the corresponding satellite. Accordingly, the DQN produces the action-value estimates $\{Q(o_t, a; \theta) \mid a \in \mathcal{S}_t\}$.

Action selection follows the ε-greedy policy. During exploitation, the satellite associated with the maximum Q-value is selected, that is, $a_t = \arg\max_{a \in \mathcal{S}_t} Q(o_t, a; \theta)$. During exploration, a satellite is randomly selected from the candidate set $\mathcal{S}_t$. If $a_t \neq a_{t-1}$, a handover is triggered; otherwise, the current satellite connection is maintained.

The Q-network parameters, θ, are updated by minimizing the temporal-difference (TD) loss:

$$\mathcal{L}(\theta) = \mathbb{E}_{(o_t, a_t, r_t, o_{t+1}) \sim \mathcal{D}}\left[\left(r_t + \gamma \max_{a' \in \mathcal{S}_{t+1}} Q(o_{t+1}, a'; \theta^-) - Q(o_t, a_t; \theta)\right)^2\right], \tag{14}$$

where θ and θ^- denote the parameters of the online and target Q-networks, respectively, γ denotes the discount factor, and $\mathcal{D}$ represents the replay buffer. $Q(o_t, a_t; \theta)$ represents the

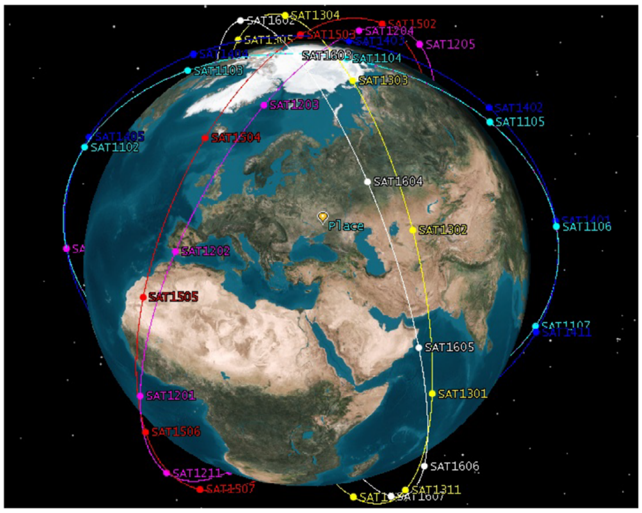

FIGURE 2 | LEO satellite constellation in STK for an emergency scenario.

action–value function estimating the expected cumulative reward of taking action a_t given the observation o_t.

In this implementation, a fully connected feedforward neural network with four hidden layers is employed. Specifically, the hidden layers contain [256,128,64,32] neurons, respectively, each followed by a ReLU activation function. The network is trained using the Adam optimizer with a discount factor of $\gamma = 0.95$. An experience replay buffer of $|\mathcal{D}| = 5000$ is employed, with a minimum buffer threshold of $|\mathcal{D}|_{\min} = 1000$, from which mini-batches of $\mathcal{B} = 128$ are sampled for parameter updates. ϵ is initialized at 1.0 and decays by $\epsilon_{\text{decay}} = 0.995$ per episode until a minimum of $\epsilon_{\min} = 0.05$ is reached.

To enhance robustness and adaptability in dynamic environments, disturbances are introduced during training. In each episode, 40% of satellites across 40% of the time slots are randomly subjected to a 20% reduction in both bandwidth and CNR, thereby simulating potential disturbances to the satellite constellation. The training procedure of the proposed DRL-based satellite selection method is shown in Algorithm 1.

In practical deployment scenarios, satellite operators can provide ephemeris data over a future time horizon T following a disaster, including orbital positions, available resources, and satellite visibility periods. Based on this information, the proposed method can be trained offline, after which the trained model is deployed on the VSAT. This enables the VSAT to select satellites at each time slot t to establish reliable ground–satellite backhaul links.

ALGORITHM 1 | Training Procedure of the Proposed DQN-Based Satellite Selection Method.

Environment: Satellite constellation with S satellites over T time slots
Inputs: Normalized matrices of CNR, Bandwidth, Delay, Computing Frequency, and RST
Initialize: Q-network Q_θ, target network $Q_{\theta^-} \leftarrow Q_\theta$
Initialize replay buffer $\mathcal{D} \leftarrow \emptyset$
Set training parameters as described in Section 3
Set disturbance parameters: $\rho_s = 0.4$, $\rho_t = 0.4$, $\delta = 0.2$
for episode $= 1$ to N_{episodes} **do**
  Randomly sample $\mathcal{S}_d \subseteq \{1, \ldots, S\}$ with $|\mathcal{S}_d| = \lceil \rho_s S \rceil$
  Randomly sample $\mathcal{T}_d \subseteq \{1, \ldots, T\}$ with $|\mathcal{T}_d| = \lceil \rho_t T \rceil$
  for all $i \in \mathcal{S}_d$ **do**
    for all $t \in \mathcal{T}_d$ **do**
      $\text{BW}_i^t \leftarrow (1-\delta)\,\text{BW}_i^t$
      $\text{CNR}_i^t \leftarrow (1-\delta)\,\text{CNR}_i^t$
    end for
  end for
  Select a random valid initial satellite as the current connection
  for time slot $t = 1$ to T **do**
    Construct state s_t from normalized features of the connected satellite
    Predict Q-values $\{Q_\theta(s_t, a)\}$ for candidate satellites
    Select action a_t using an ϵ-greedy policy
    Execute a_t and observe next state s_{t+1}
    Compute reward r_t according to (13)
    Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$
    if $|\mathcal{D}| \geq |\mathcal{D}|_{\min}$ **then**
      Sample a mini-batch of size B and update Q_θ using (14)
    end if
    $s_t \leftarrow s_{t+1}$
  end for
  if episode mod target update frequency $= 0$ **then**
    Update target network: $Q_{\theta^-} \leftarrow Q_\theta$
  end if
  $\epsilon \leftarrow \max(\epsilon \cdot \epsilon_{\text{decay}}, \epsilon_{\min})$
end for
Output: Trained Q-network Q_θ

4 | Simulation and Performance Evaluation

In this study, simulations are carried out in STK using the Iridium constellation [9], with parameters summarized in Table 1. Suppose that the VSAT is positioned at the Russia–Ukraine border, as shown in Figure 2. Additionally, the satellite ephemeris is illustrated in Figure 3, which presents the visibility windows of the satellites relative to the VSAT location. A time horizon of $T = 200$ slots is considered, with each slot lasting 1 min. Specifically, the link metrics of each candidate satellite $i \in \mathcal{S}_t$ at time slot t are obtained from STK, including CNR_i^t, τ_i^t, and the slant range D_i^t between the VSAT and satellite i. The corresponding propagation delay is computed as $d_i^t = D_i^t / c$, where c denotes the speed of light. Meanwhile, the bandwidth BW_i^t and computing frequency CF_i^t are randomly assigned within the ranges of 1–3 MHz and

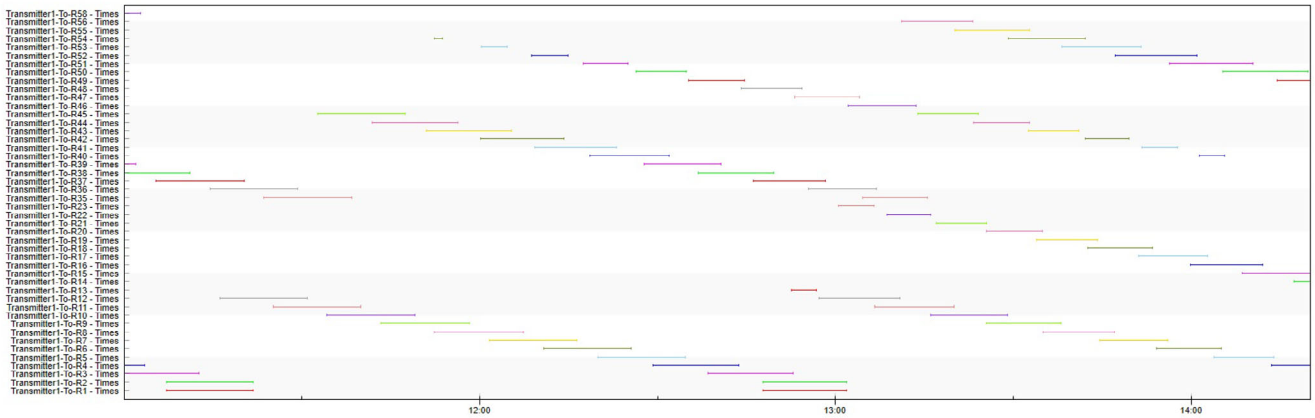

FIGURE 3 | Satellite ephemeris diagram in STK.

TABLE 1 | LEO constellation parameters.

Parameter	Value	Parameter	Value
Constellation type	Iridium	Number of orbital planes	6
Satellites per plane	11	Orbital height	780 km
Orbital inclination	86.4°	Min. elevation angle	10°
Carrier frequency	2 GHz	Modulation type	BPSK
Demodulator	BPSK	EIRP	10 dBW

200–2000 MHz, respectively. In addition, the reward function parameters are set as $\alpha = 10$, $\beta = 8$, $\zeta = 13$, and $\eta = 8$.

To assess the effectiveness of the proposed method, we conduct a comparative evaluation against several representative satellite selection approaches, including PPO [27], maximum CNR, maximum RST [28], technique for order preference by similarity to ideal solution (TOPSIS) [23], simple additive weighting (SAW) [9], and a graph-based approach [10].

4.1 | Reward Convergence and Random Seed Evaluation

As illustrated in Figure 4, the original and smoothed reward curves depict the training dynamics of the proposed method in comparison with PPO. The proposed method demonstrates a consistently faster convergence rate and attains a 16.67% higher total reward than PPO, thereby indicating its superior learning efficiency.

A statistical performance evaluation was conducted using five random seeds for both the proposed DQN-based method and the PPO baseline. The corresponding final and peak rewards are summarized in Table 2, where the final average reward is computed as the mean value over the final 10% of training episodes. The results indicate that the proposed approach achieves a higher mean reward and a smaller standard deviation,

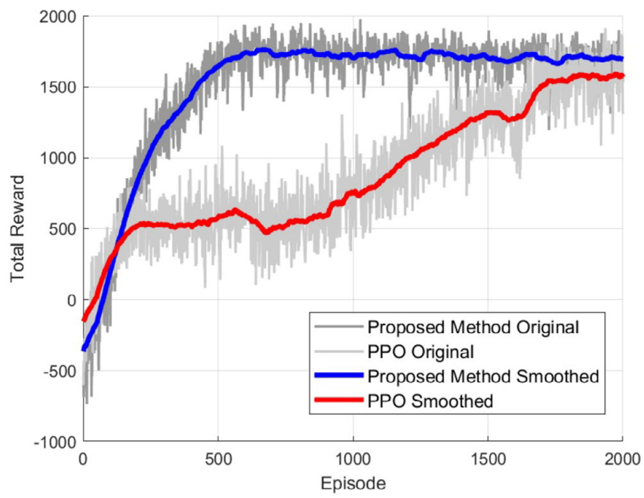

FIGURE 4 | Training rewards of the proposed method and PPO.

demonstrating superior robustness, convergence stability, and resistance to random fluctuations during training.

4.2 | HO Performance

Figure 5 presents cumulative HOs of each algorithm during the simulation. The results indicate that the proposed method yields a slightly higher number of HOs compared to the graph-based approach, which achieves the fewest. This difference is attributed to the distinct decision mechanisms. In detail, the graph-based algorithm selects satellites along a globally optimized connectivity path, inherently preserving continuity and minimizing HOs. In contrast, the proposed DRL method performs sequential per–time-slot decisions, and its adaptive policy updates may introduce minor action fluctuations, resulting in slightly more HOs. Nonetheless, the proposed method significantly outperforms PPO and other satellite selection methods in reducing the overall number of HOs. Importantly, excessive HOs result in substantial signaling overhead, which poses a challenge in resource-constrained emergency scenarios.

Figure 6 shows the number of connection failures, with the threshold data rate for link failure set to 10 Mbps. In emergency

TABLE 2 | Final and peak rewards over five random seeds.

Metric	Method	S1	S2	S3	S4	S5	Mean ± std
Final avg.	Proposed	1710	1682	1629	1730	1745	**1699.2 ± 44.6**
	PPO	1546	1352	1549	1482	1452	**1482.2 ± 84.5**
Peak	Proposed	1977	1948	1973	1978	1982	**1971.6 ± 13.8**
	PPO	1913	1708	1860	1776	1768	**1805.0 ± 80.8**

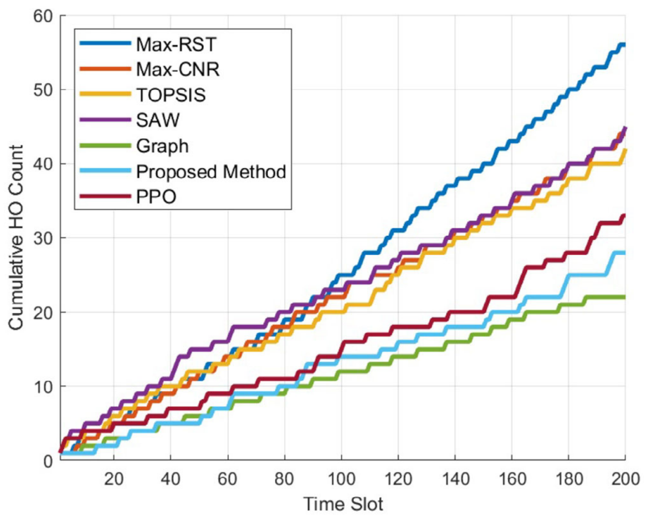

FIGURE 5 | Comparison of cumulative HOs across different methods.

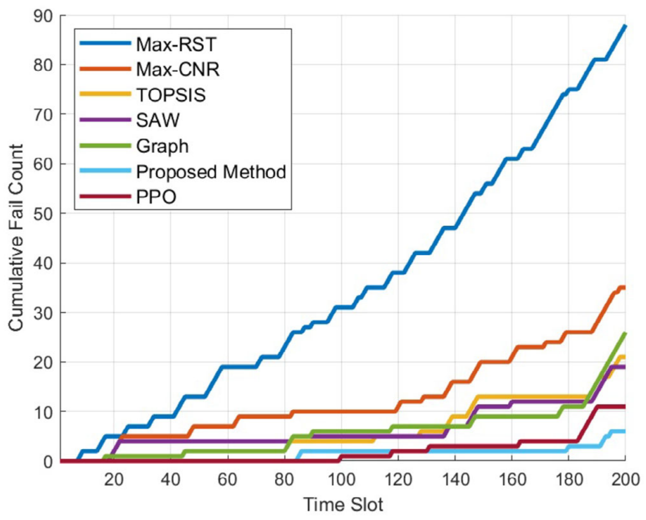

FIGURE 6 | Comparison of cumulative failures across different methods.

scenarios, the wireless backhaul link must support high data rates for applications such as high-definition video, 3D mapping, and disaster monitoring. As illustrated, the proposed method achieves the lowest number of connection failures among all approaches. This indicates that the method can consistently select satellites that enable the VSAT to maintain a high-throughput and reliable link over the entire time horizon T.

TABLE 3 | Comparison of average data rate, delay, and CF across satellite selection strategies.

Strategy	Data rate (Mbps)	Delay (ms)	CF (Hz)
Max RST	10.96	27.25	867
Max CNR	13.83	16.76	830
TOPSIS	14.12	17.13	963
SAW	14.57	16.79	910
Graph	14.47	19.18	894
PPO	14.81	19.20	795
Proposed method	**14.87**	**20.21**	**806**

TABLE 4 | Sensitivity analysis of weights on HO and failure counts.

Weight	Setting	HO	ΔHO (%)	Fail	Δ Fail (%)
Baseline	(10, 8, 13, 8)	**28**	—	**6**	—
α −20%	(8, 8, 13, 8)	26	−7.1	15	+150.0
α +20%	(12, 8, 13, 8)	25	−10.7	24	+300.0
β −20%	(10, 6.4, 13, 8)	25	−10.7	12	+100.0
β +20%	(10, 9.6, 13, 8)	28	0.0	12	+100.0
ζ −20%	(10, 8, 10.4, 8)	25	−10.7	18	+200.0
ζ +20%	(10, 8, 15.6, 8)	29	+3.6	10	+66.7
η −20%	(10, 8, 13, 6.4)	29	+3.6	14	+133.3
η +20%	(10, 8, 13, 9.6)	33	+17.9	12	+100.0

Additionally, average link-level metrics, including data rate, delay, and computing frequency (CF), are evaluated. As shown in Table 3, the proposed method achieves the highest average data rate, while incurring a modest increase in delay and a slight decrease in computing frequency compared to other strategies. This trade-off facilitates the establishment of a high-throughput ground–satellite backhaul link in emergency scenarios, where terrestrial infrastructure is often severely disrupted and satellite-based backhaul may serve as a crucial communication link to the outside world.

To evaluate the sensitivity of the reward function with respect to its weight parameters, a series of experiments was conducted. Table 4 summarizes the corresponding results. In detail, each parameter was perturbed by ±20% while keeping the others fixed, with the baseline weight setting defined as $(\alpha, \beta, \zeta, \eta) = (10, 8, 13, 8)$. As shown in the table, although adjusting certain

TABLE 5 | HO performance at Location 1.

Method	HO	Fail	Data rate (Mbps)	Delay (ms)	CF (Hz)
Proposed	23	36	12.15	23.21	1259.44
PPO	24	42	11.82	24.36	1170.00
TOPSIS	38	60	11.72	22.09	1243.33
SAW	37	68	11.73	22.09	1163.89
Graph	20	43	11.85	24.87	1190.00
Max CNR	37	68	11.67	21.26	1102.78
Max RST	35	106	9.14	30.69	1015.56

TABLE 6 | HO performance at Location 2.

Method	HO	Fail	Data rate (Mbps)	Delay (ms)	CF (Hz)
Proposed	21	1	14.85	19.73	951.50
PPO	23	4	14.60	19.86	1032.50
TOPSIS	41	18	14.29	15.54	1043.50
SAW	40	23	14.14	15.45	1106.50
Graph	16	5	14.46	19.28	970.50
Max CNR	43	34	14.01	14.88	939.00
Max RST	51	95	10.61	27.68	919.00

weights can slightly reduce the HO frequency, it leads to a considerable increase in connection failures, indicating that the chosen weight configuration provides an effective trade-off between HO frequency and connection reliability.

4.3 | HO Performance Under Different Latitude Scenarios

To evaluate the impact of geographical latitude on HO performance, two additional disaster-prone locations at different latitudes are considered. Location 1 corresponds to a low-latitude region within the Indonesian seismic belt, where satellite visibility is limited and link quality is constrained by lower CNR conditions. In contrast, Location 2 represents a high-latitude region within the Alaska seismic belt, where orbital convergence leads to increased satellite visibility and higher elevation angles, thereby yielding higher CNR.

Tables 5 and 6 summarize the HO performance at the low-latitude and high-latitude locations, respectively. At Location 1, the proposed method achieves fewer HOs, fewer connection failures, and a higher data rate compared with the baseline methods under constrained link conditions. Notably, the failure count is higher than that observed in the mid-latitude scenario (the Russia–Ukraine border region), primarily due to the reduced satellite visibility and lower CNR conditions at low latitudes.

At Location 2, where satellite visibility is higher and CNR conditions are more favorable, the proposed method significantly reduces the failure count while improving the data rate compared with the low-latitude case. Moreover, it achieves the lowest failure count and the highest data rate among all evaluated methods. These results indicate that the proposed method effectively supports high-throughput disaster information backhaul across diverse geographical conditions.

TABLE 7 | Deterministic disturbance scenarios.

Scenario	Type	Slots	Satellites	Disturbance
P_0	Baseline	0%	0%	None
P_1	BW-only	50%	50%	BW ↓ 30%
P_2	CNR-only	70%	70%	CNR ↓ 50%
P_3	Mixed	50%	50%	BW ↓ 40%, CNR ↓ 50%

4.4 | HO Performance Under Disturbance Scenarios

To evaluate the robustness of the proposed method under time-varying disturbances, two categories of disturbance are considered. The first category corresponds to deterministic scenarios, in which the disturbance conditions are predefined and remain constant throughout the evaluation. In contrast, the second category introduces a stochastic disturbance setting designed to emulate unpredictable environmental variations, wherein the affected time slots, satellites, and the degradation levels of bandwidth and CNR are randomly sampled within prescribed ranges.

4.4.1 | Deterministic Disturbance

Table 7 summarizes the deterministic disturbance scenarios considered in the evaluation. The baseline scenario P_0 introduces no degradation, while P_1, P_2, and P_3 correspond to bandwidth-only, CNR-only, and mixed-disturbance conditions, respectively. For example, in scenario P_1, 50% of the time slots and satellites experience a 30% reduction in available bandwidth, reflecting severe resource contention. Scenario P_2 represents CNR degradation induced by adverse weather conditions, whereas scenario P_3 imposes simultaneous reductions in both bandwidth and CNR to reflect more severe disturbance conditions.

Figure 7 compares the cumulative number of HOs obtained by the proposed method and PPO under the deterministic disturbance scenarios P_0–P_3. Across all scenarios, the proposed method consistently achieves fewer HOs than PPO. As the disturbance severity increases from P_0 to P_3, both methods experience an overall rise in HO occurrences; however, the PPO curves exhibit noticeably larger fluctuations, whereas the proposed method maintains substantially more stable switching behavior. This stability indicates a stronger capability to accommodate bandwidth, CNR, and mixed disturbances while suppressing unnecessary HOs. It is worth noting that excessive HOs incur significant signaling and power overhead, which is particularly undesirable in resource-constrained emergency environments.

Figure 8 presents the cumulative connection failures under the deterministic disturbance scenarios P_0–P_3. Consistent with the

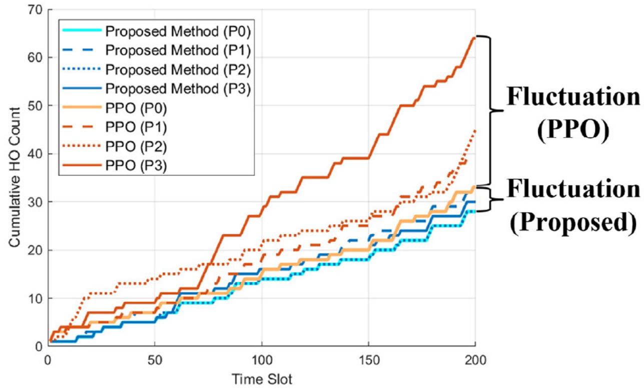

FIGURE 7 | Comparison of cumulative HOs under P_0–P_3.

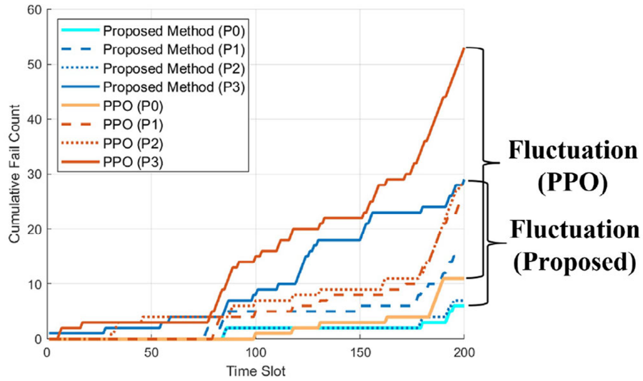

FIGURE 8 | Comparison of cumulative failures under P_0–P_3.

HO results shown in Figure 7, the proposed method incurs fewer failures than PPO across all disturbance levels, indicating a more reliable maintenance of satellite backhaul links. Furthermore, as the impairment severity increases, the proposed method exhibits noticeably smaller fluctuations, demonstrating a more stable response to bandwidth, CNR, and mixed degradations. This stability is particularly critical in emergency environments, as it helps ensure reliable data return from affected regions, especially for services with stringent continuity and throughput requirements, such as real-time video transmission and remote medical assistance.

Table 8 reports the average data rate, delay, and CF achieved under the deterministic disturbance scenarios P_0–P_3. The proposed method consistently achieves a higher data rate than PPO, indicating more effective satellite selection in degraded environments. Moreover, the proposed method exhibits smaller variations in both delay and CF, reflecting more stable link conditions. This performance trade-off enables reliable and high-throughput ground–satellite backhaul in emergency scenarios.

TABLE 8 | Comparison of average data rate, delay, and CF under deterministic disturbance scenarios.

Strategy	Data rate (Mbps)	Delay (ms)	CF (Hz)
Proposed method (P0)	14.87	20.21	806
Proposed method (P1)	14.38	20.23	808
Proposed method (P2)	14.12	20.33	803
Proposed method (P3)	13.77	20.21	808
PPO (P0)	14.81	19.20	795
PPO (P1)	13.69	20.05	779
PPO (P2)	13.13	20.11	952
PPO (P3)	12.93	20.86	793

TABLE 9 | HO performance under stochastic disturbances.

Method	HO mean	HO std	Fail mean	Fail std
Proposed method	30.57	±2.96	26.58	±14.75
PPO	52.75	±12.65	41.60	±17.03
TOPSIS	50.70	±5.17	48.43	±9.89
SAW	49.99	±3.79	47.87	±10.23
Graph-based	29.27	±1.86	48.36	±11.40

4.4.2 | Stochastic Disturbance

To evaluate the performance of the proposed model in the presence of unpredictable disturbances, a stochastic disturbance scenario is introduced. Specifically, the proportion of affected time slots and satellites is randomly sampled between 30% and 50%, while the degradation levels of bandwidth and CNR are randomly drawn from a 20%–50% range. In addition, the disturbance type at each sampling instant is randomly selected from bandwidth-only, CNR-only, and mixed impairment cases. This configuration reflects the intrinsic randomness of real-world satellite communication environments, where link quality can fluctuate as a result of user contention, weather dynamics, or the combined influence of multiple external factors. To obtain statistically reliable performance evaluations, the stochastic scenario is executed over 100 independent random seeds. The associated HO, failure, and link-quality metrics are then gathered, with their mean values and standard deviations reported below.

Based on the results in Table 9, the proposed method demonstrates improved HO performance under stochastic disturbances compared with different baseline algorithms. In particular, it achieves a lower HO frequency and a smaller average number of connection failures, indicating enhanced robustness. Compared with PPO, a DRL-based method, the proposed approach reduces the average HO count by 42.1% and the average failure count by 36.1%, while maintaining lower variability in both metrics. In addition, while the graph-based method attains the lowest HO count, it exhibits a significantly higher failure rate. This behavior is likely attributable to the global optimization characteristic of the graph-based approach, which emphasizes long-term optimality but exhibits limited responsiveness to rapid channel degradation induced by stochastic disturbances.

As shown in Table 10, the proposed method achieves the highest average data rate under stochastic disturbance scenarios. This result suggests that the proposed satellite selection strategy can effectively exploit available link resources to enhance

TABLE 10 | Comparison of average data rate, delay, and CF under stochastic disturbance scenarios.

Method	Data rate (Mbps)	Delay (ms)	CF (Hz)
Proposed method	13.81 ± 0.56	20.25 ± 0.11	806.92 ± 6.31
PPO	13.34 ± 0.72	20.24 ± 0.71	812.87 ± 46.28
TOPSIS	13.05 ± 0.49	18.19 ± 0.24	976.00 ± 28.55
SAW	13.08 ± 0.48	17.87 ± 0.20	958.28 ± 21.37
Graph-based	13.24 ± 0.55	19.06 ± 0.29	941.72 ± 32.83

TABLE 11 | Ablation study under stochastic disturbances.

Method	HO mean	HO std	Fail mean	Fail std
DQN-Mix (proposed)	30.57	±2.96	26.58	±14.75
DQN-NoDist	27.86	±3.22	39.31	±11.68
DQN-BW	31.92	±2.55	34.19	±16.84
DQN-CNR	36.40	±2.27	31.60	±16.66
PPO-Mix	52.75	±12.65	41.60	±17.03
PPO-NoDist	57.38	±12.29	49.60	±17.35

TABLE 12 | QoS performance ablation under stochastic disturbances.

Method	Data rate (Mbps)	Delay (ms)	CF (Hz)
DQN-Mix (proposed)	13.81 ± 0.56	20.25 ± 0.11	806.92 ± 6.31
DQN-NoDist	12.94 ± 1.21	20.61 ± 0.48	821.45 ± 35.27
DQN-BW	13.26 ± 0.88	20.43 ± 0.32	815.30 ± 18.64
DQN-CNR	13.08 ± 0.97	20.39 ± 0.29	813.84 ± 22.51
PPO-Mix	13.34 ± 0.72	20.24 ± 0.71	812.87 ± 46.28
PPO-NoDist	12.47 ± 1.35	20.88 ± 0.83	835.62 ± 51.74

the throughput of the ground–satellite backhaul link while improving robustness against disturbances, which is critical for disaster-response communications under dynamically varying environments.

Tables 11 and 12 present the ablation results under stochastic disturbance scenarios, where different disturbance-related modules incorporated in the DQN training are selectively removed to assess their impact on HO and link performance. Specifically, the proposed DQN-based framework is evaluated under several ablation settings, including mixed disturbance modeling that incorporates both CNR and bandwidth disturbances (DQN-Mix), no-disturbance training (DQN-NoDist), and partial disturbance modeling considering either bandwidth or CNR (DQN-BW and DQN-CNR).

TABLE 13 | Training time of the proposed method and PPO under five random seeds.

Seed	Proposed (min)	PPO (min)	Avg. time/ episode (s)
1	17.3	21.4	1.04/1.28
2	16.8	20.9	1.01/1.25
3	17.1	21.7	1.03/1.30
4	17.5	20.8	1.05/1.25
5	16.9	21.2	1.02/1.27

Notably, the proposed method (DQN-Mix) achieves a lower HO frequency and the lowest number of connection failures, while maintaining a high average data rate. By comparison, DQN-NoDist results in fewer HOs than DQN-Mix but exhibits a higher number of connection failures. This behavior is attributed to the training of DQN-NoDist under ideal channel conditions. As a result, when link quality degrades due to stochastic disturbances, the policy tends to maintain existing connections, leading to fewer handovers but more frequent connection failures. DQN-BW and DQN-CNR exhibit intermediate performance and remain less robust than DQN-Mix, indicating that jointly modeling both CNR and bandwidth disturbances during training is essential for reliable decision-making.

Both PPO-Mix and PPO-NoDist result in higher HO counts and connection failures than the DQN-based methods, with PPO-NoDist exhibiting the lowest overall performance. The observed performance gap stems from the inherent differences between value-based and policy-gradient methods. For the satellite selection problem considered in this work, the action space is discrete and relatively small, while the environment is characterized by abrupt variations in link quality and sparse reward signals. Under such conditions, a DQN is well suited to learning stable state–action value estimates by effectively leveraging long-term reward, thereby facilitating more effective HO decisions. By contrast, PPO relies on stochastic policy updates that are more sensitive to reward variance and fluctuations caused by channel and resource variations, which can hinder the learning of stable HO policies in discrete and highly dynamic environments.

4.5 | Computational Efficiency and Complexity Analysis

4.5.1 | Computational Efficiency Evaluation

To assess computational efficiency, both the proposed DQN-based method and the PPO algorithm were trained under five random seeds using identical simulation environments and hyperparameter configurations. All experiments were conducted on a laptop equipped with an Intel Core i7-1370P CPU (14 cores, 20 threads). The corresponding total training times were recorded and analyzed for comparative evaluation. As shown in Table 13, the proposed method consistently required less total training time than PPO across all seeds. Specifically, the proposed approach completed training in about 17 min, while PPO required approximately 21 min, reflecting a 19% reduction in training time. Together with the reward results shown in Table 2, the results indicate that the proposed DQN-based framework provides higher computational efficiency than PPO. This advantage makes it well suited for emergency communication scenarios with limited computational resources.

4.5.2 | Computational Complexity Analysis

The proposed method adopts a value-based DQN framework consisting of a Q-network and a target network. Let N_{ep} denote the number of training episodes, T the number of time slots per episode, B the mini-batch size used in experience replay, and P the number of learnable parameters in the neural network. During training, DQN selects actions through a single forward pass at each decision step, while network parameters are updated via mini-batch temporal-difference learning. Accordingly, the training complexity of DQN can be approximated as $O\left(N_{\mathrm{ep}}TBP\right)$.

In contrast, PPO adopts an actor–critic architecture, in which the actor and critic networks are optimized jointly. Let P_a and P_c denote the numbers of learnable parameters in the actor and critic networks, respectively, and M the number of policy update epochs per batch. Therefore, the training complexity of PPO can be approximated as $O\left(N_{\mathrm{ep}}TM(P_a + P_c)\right)$. Importantly, PPO requires the optimization of two networks and multiple policy update epochs, leading to higher computational complexity than the proposed DQN-based method under comparable settings.

5 | Conclusion

In this paper, we propose a DRL-based adaptive satellite selection method for VSATs to establish reliable and high-throughput ground–satellite backhaul links in dynamic emergency scenarios. Specifically, a reward function is designed to jointly balance link quality, HO frequency, and connection failures, thereby ensuring reliability and efficiency in terrestrial–satellite connectivity. Moreover, disturbances that may arise in the satellite constellation are incorporated into the training process, thereby significantly enhancing the robustness of the learned policy. Simulation results demonstrate that the proposed method not only reduces connection failures and HOs, but also maintains stable performance under unpredictable disturbance conditions, enabling reliable and high-throughput wireless backhaul for emergency communications.

Author Contributions

Ke Chen: conceptualization, investigation, methodology, software, validation, writing – original draft. **Li Zhang:** project administration, supervision, writing – review & editing. **Qihang Jiao:** data curation. **Samuel Krain:** writing – review & editing.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

No new data were generated or analysed in support of this research.

References

1. D. G. C., A. Ladas, Y. A. Sambo, H. Pervaiz, C. Politis, and M. A. Imran, "An Overview of Post-Disaster Emergency Communication Systems in Future Networks," *IEEE Wireless Communications* 26, no. 6 (2019): 132–139.

2. Y. Lin, Z. Xu, J. Li, J. Wang, and C. Li, "Deep Reinforcement Learning-Based Intelligent Resource Management in Multi-UAVs-Assisted MEC Emergency Communication System," *IET Communications* 19, no. 1 (2025): e70063.

3. M. Matracia, N. Saeed, M. A. Kishk, and M.-S. Alouini, "Post-Disaster Communications: Enabling Technologies, Architectures, and Open Challenges," *IEEE Open Journal of the Communications Society* 3 (2022): 1177–1205.

4. K. Chen, L. Zhang, X. Chen, S. Krain, and Y. Ge, "Reinforcement Learning-Based Network Selection for Emergency Satellite–Aerial Systems," in *Proceedings of the IEEE International Conference on Signal Processing Systems (ICSPS)* (IEEE, 2025), 1650–1655.

5. K. Chen, L. Zhang, and J. Zhong, "Space-Air-Ground Integrated Network (SAGIN) in Disaster Management: A Survey," *IEEE Transactions on Network and Service Management* 22, no. 5 (2025): 4021–4049.

6. K. Chen, L. Zhang, S. Krain, and Q. Jiao, "Graph-Based Lightweight Satellite Selection for Emergency Communication Scenarios," in *Proceedings of the Communication and Information Technologies (KIT)* (IEEE, 2025), 1–7.

7. S. R. Sabuj and H.-S. Jo, "A Complex Wavelet-Based MIMO-OTFS Scheme for Joint Sensing and Communication LEO Satellite," in *IEEE International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)* (IEEE, 2025), 84–90.

8. S. R. Sabuj and H.-S. Jo, "A Complex Wavelet-Based OTFS Scheme for LEO Satellite Communication," *IEEE Access* 13 (2025): 85807–85825.

9. L. Shi, F. Yang, W. Wu, A. Sun, Y. Sun, and T. Sun, "Load Balancing and Remaining Visible Time Based Handover Algorithm for LEO Satellite Networks," in *Proceedings IEEE 8th International Conference on Computer Communications and Networks (ICCC)* (IEEE, 2022), 391–395.

10. Z. Wu, F. Jin, J. Luo, Y. Fu, J. Shan, and G. Hu, "A Graph-Based Satellite Handover Framework for LEO Satellite Communication Networks," *IEEE Communications Letters* 20, no. 8 (2016): 1547–1550.

11. M. Hozayen, T. Darwish, G. K. Kurt, and H. Yanikomeroglu, "A Graph-Based Customizable Handover Framework for LEO Satellite Networks," in *Proceedings of the IEEE Globecom Workshops (GC Wkshps)* (IEEE, 2022), 868–873.

12. X. Lv, S. Wu, A. Li, J. Jiao, N. Zhang, and Q. Zhang, "A Weighted Graph-Based Handover Strategy for Aeronautical Traffic in LEO SATCOM Networks," *IEEE Networking Letters* 4, no. 3 (2022): 132–136.

13. Z. Jia, M. Sheng, J. Li, D. Niyato, and Z. Han, "LEO-Satellite-Assisted UAV: Joint Trajectory and Data Collection for Internet of Remote Things in 6G Aerial Access Networks," *IEEE Internet of Things Journal* 8, no. 12 (2021): 9814–9826.

14. E. M. Mohamed, M. A. Alnakhli, and M. M. Fouda, "Joint UAV Trajectory Planning and LEO-SAT Selection in SAGIN," *IEEE Open Journal of the Communications Society* 5 (2024): 1624–1638.

15. S. Karapantazis and F.-N. Pavlidou, "Dynamic Time-Based Handover Management in Leo Satellite Systems," *Electronics Letters* 43, no. 5 (2007): 291–293.

16. Y. Liu, L. Feng, L. Wu, et al., "Joint Optimization Based Satellite Handover Strategy for Low Earth Orbit Satellite Networks," *IET Communications* 15, no. 7 (2021): 935–947.

17. M. O. K. Mendonça, E. Lagunas, A. Perez-Neira, et al., "Managing User and Gateway Link Spectral Coexistence in LEO Satellite Systems

via Hand-Over," in *Proceedings of the 40th International Communications Satellite Systems Conference (ICSSC)* (IEEE, 2023), 17–22.

18. J. Wang, W. Mu, Y. Liu, L. Guo, S. Zhang, and G. Gui, "Deep Reinforcement Learning-Based Satellite Handover Scheme for Satellite Communications," in *2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)* (IEEE, 2021), 1–6.

19. H. Zhang and B. Li, "DQN-Based Conditional Handover Algorithm for LEO Satellite Networks," in *Proceedings of the IEEE International Conference on Electronic Technology. (ICET)* (IEEE, 2025), 203–208.

20. X. Li, H. Feng, Y. Weng, and X. Xu, "A LEO Satellite Handover Algorithm Based on Dueling DQN," in *Proceedings of the IEEE International Conference on Computer Science and Blockchain (CCSB)* (IEEE, 2024), 287–293.

21. G. Zhong and B. Li, "PER-DDQN-Based Load Balancing Handover Strategy for LEO Satellite Networks," in *Proceedings of the IEEE International Conference on Electronics Technology (ICET)* (IEEE, 2025), 1–6.

22. C. Liang, Y. Guo, Z. Wu, and R. Chai, "A QoS-Aware Handover Mechanism for LEO Satellite Networks Based on Multi-Agent DRL," in *IEEE Asia-Pacific Conference on Communications (APCC)* (IEEE, 2024), 108–113.

23. K. Chen, L. Zhang, and J. Zhong, "Vertical Handover Strategy in Satellite–Aerial Based Emergency Communication Networks," in *Proceedings of the IEEE International Conference on Wireless Networks and Mobile Communications (WINCOM)* (IEEE, 2024), 1–6.

24. Y. Zhong, H. Wang, H. Lv, and F. Guo, "A Vertical Handoff Decision Scheme Using Subjective-Objective Weighting and Grey Relational Analysis in Cognitive Heterogeneous Networks," *Ad Hoc Networks* 134 (Sep. 2022): 102924.

25. P. Hou and N. Cen, "DC-PPO for Joint User Association and Power Allocation in Dynamic Indoor Hybrid VLC/RF Networks," in *25th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)* (IEEE, 2024), 186–195.

26. N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," *IEEE Communications Surveys & Tutorials* 21, no. 4 (2019): 3133–3174.

27. S. Zhang, C. Sun, S. Hu, W. Ni, and X. Wang, "Proximal Policy Optimization With Adaptive Weight Adjustment for Airborne Integrated Sensing and Communication," *IEEE Wireless Communications Letters* 14, no. 6 (2025): 1703–1707.

28. P. K. Chowdhury, M. Atiquzzaman, and W. Ivancic, "Handover Schemes in Satellite Networks: State-of-the-Art and Future Research Directions," *IEEE Communications Surveys & Tutorials* 8, no. 4 (2006): 2–14.