SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

September 2025, Vol. 68, Iss. 9, 190303:1–190303:14
https://doi.org/10.1007/s11432-024-4479-0

Special Topic: Cohesive Clustered Satellites System for 5GA and 6G Networks

Federated learning-based ISAC network in cohesive clustered satellite: resource optimization in heterogeneous datasets and systems

Hongbo ZHAO[1], Menghan WU[1], Tiannuo LIU[1], Deyou ZHANG[1*], Dawei WANG[2*] & Rongke LIU[1]

[1]*School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*
[2]*School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China*

Received 30 November 2024/Revised 20 March 2025/Accepted 4 June 2025/Published online 15 August 2025

Abstract With the continuous advancement of space technology and the rapid increase in the number of low Earth orbit satellites, cohesive clustered satellites (CCS) are thriving. Additionally, the growing computational capabilities of onboard satellite equipment have enhanced constellations' data processing power. These developments have made federated learning (FL)-based CCS (CCSFL), a feasible and promising approach. Therefore, this paper proposes a CCSFL-based integrated sensing and communication (ISAC) network, where FL convergence, transmission latency, and energy consumption are optimized using a deep reinforcement learning (DRL) approach under heterogeneous datasets and system conditions. To further enhance communication performance, we adopt intra-orbit inter-satellite link (ISL) multi-hop routing, inter-orbit ISL neighbor forwarding, and sparse gradient compression techniques. Specifically, we introduce a utility function based on the sensing signal-to-noise ratio (SNR) as a reward for the double deep Q-network (DDQN) algorithm, addressing the optimal client selection problem under heterogeneous datasets and systems. Additionally, we employ the deep deterministic policy gradient (DDPG) algorithm to optimize system-wide latency and energy consumption. Simulation results show that the proposed algorithm outperforms the benchmark in both FL accuracy and resource utilization.

Keywords cohesive clustered satellites (CCS), federated learning (FL), integrated sensing and communications (ISAC), heterogeneous datasets and systems, deep reinforcement learning (DRL)

Citation Zhao H B, Wu M H, Liu T N, et al. Federated learning-based ISAC network in cohesive clustered satellite: resource optimization in heterogeneous datasets and systems. Sci China Inf Sci, 2025, 68(9): 190303, https://doi.org/10.1007/s11432-024-4479-0

1 Introduction

Owing to their low latency, extensive coverage, and strong robustness, dense low Earth orbit (LEO) constellations are widely utilized for global Internet access, remote sensing, navigation, and positioning, providing enhanced solutions for international communications and data services [1–3]. For example, the Blackjack program, led by the Defense Advanced Research Projects Agency (DARPA), provides reliable communication and detection support in military conflicts, thereby enhancing the military's command and control capabilities [4]. Additionally, the Starlink program, developed by SpaceX, aims to achieve seamless global civilian network coverage and deliver high-speed, low-latency Internet access services [5]. In China, the "Thousand Sails Project", led by the China Aerospace Science and Industry Corporation (CASIC), seeks to deploy a constellation of 10000 low Earth orbit broadband multimedia satellites to enhance and expand the coverage and capabilities of global satellite Internet services.

With enhanced computing capabilities, satellites can now collect and process ground data in real time, quickly extracting key information. Furthermore, the development of inter-satellite links (ISLs) has facilitated the formation of satellite clusters [6]. These cohesive clustered satellites (CCS) not only enhance the observational capacity of individual satellites but also significantly improve the overall efficiency and stability of multi-target observation missions [7]. Compared to traditional centralized processing models, federated learning (FL)-based CCS (CCSFL) systems have garnered increasing attention from researchers

* Corresponding author (email: deyou@buaa.edu.cn, wangdw@nwpu.edu.cn)

due to their ability to reduce communication overhead while preserving data privacy [8]. This approach works by sending only the updated model parameters to the central parameter server, avoiding the transmission of raw data [9].

The authors in [10] proposed a terrestrial-satellite collaborative FL approach to manage machine learning services in remote areas, focusing on minimizing training latency by optimizing offloading volumes and computation speeds. Similarly, Ref. [11] proposed a routing strategy combined with resource allocation to minimize delays in transmitting FL model parameters, leveraging ISLs to expedite aggregation at ground stations. Furthermore, Ref. [12] proposed a novel scheduling method to address intermittent connectivity and optimize the use of available time by utilizing the predictable visibility patterns between satellites and ground stations.

While the aforementioned studies enhance the training speed and energy efficiency of CCSFL systems through resource management, time scheduling, and inter-satellite routing, they overlook a critical practical challenge: the heterogeneity of datasets and systems. On one hand, local datasets gathered by satellites are often non-independent and non-identically distributed (non-IID) due to variations in observation missions, sensor equipment, and spatial locations. On the other hand, differences in hardware result in varying computing and communication capacities across satellites, leading to inconsistent system latency.

To mitigate the heterogeneity caused by non-IID datasets, an alternating contrastive training algorithm was proposed in [13], where local models were divided into public and private, with only public models being uploaded to the server station. Unlike [13], the authors in [14] introduced a satellite grouping scheme to address data heterogeneity. This approach used a clustering algorithm based on data distribution and communication latency, which served as constraints for power optimization. Additionally, an experience-driven control framework named "FAVOR" was proposed in [15], which addressed the bias introduced by heterogeneous datasets by intelligently selecting clients. The authors further compared the FL training rounds required by the clustering algorithm and the client selection algorithm, concluding that a well-designed client selection strategy can significantly accelerate FL convergence.

Although the above studies explored the application of federated learning in CCS systems [10–12] and implemented effective strategies to address data heterogeneity [13–15], limited attention has been given to the heterogeneity of satellite systems. Furthermore, most existing research treats satellite local datasets as fixed conditions, overlooking the impact of satellite sensing accuracy on the data quality.

Recent advancements in integrated sensing and communication (ISAC) technology offer promising solutions. ISAC enhances satellite sensing accuracy, reducing the impact of non-IID datasets and hardware disparities. Its dynamic resource allocation capabilities can integrate with federated learning strategies, optimizing communication and computation efficiency in heterogeneous satellite systems. For example, the authors in [16] proposed a dual-function LEO satellite constellation framework. This system enables simultaneous information communication for multiple user equipment and location sensing for specific targets using the same hardware and spectrum. In [17], the authors adopted rate-splitting multiple access (RSMA) to manage interference while improving the communication-sensing trade-off. These studies highlight ISAC's potential in optimizing CCS networks, particularly in scenarios where data quality and system efficiency are critical.

Building upon these insights, this paper proposes a CCSFL framework within an ISAC network, aiming to optimize the latency and energy consumption of CCSFL systems while considering constraints such as sensing signal-to-noise ratio (SNR), ISL routing, and heterogeneity in both datasets and systems. Unlike previous approaches, this work specifically addresses both dataset and system heterogeneity, with a particular focus on variations in satellite sensing accuracy, which have often been overlooked. The specific contributions of this work are summarized as follows.

- This paper proposes a novel FL-based ISAC network for cohesive clustered satellites. Each CCS orbit consists of a leader satellite (LS) and several follower satellites (FSs) functioning as FL clients. To enhance communication performance, intra-orbit ISL multi-hop routing, inter-orbit neighbor forwarding, and sparse gradient compression methods are employed. Additionally, during each FL training round, a specific LS visible to the ground is selected to establish a star-ground link (SGL) with the ground station, ensuring consistent communication.
- To tackle both dataset and system heterogeneity while maintaining sensing accuracy, this paper proposes a double deep Q-network (DDQN)-based client selection algorithm. A utility function is designed using the sensing SNR as a reward for deep reinforcement learning (DRL), facilitating the evaluation of

FL accuracy under non-IID datasets and diverse hardware conditions. By dynamically selecting different FS sets as FL clients in each training round, this utility function effectively accelerates FL convergence.

• To reduce system latency and energy consumption, we employ the deep deterministic policy gradient (DDPG) algorithm to optimize transmission power and CPU frequency allocation for each FS. Simulation results demonstrate that our proposed algorithms achieve superior convergence speed and resource utilization compared to baseline methods. For instance, the DDQN algorithm reduces FL communication rounds by 60%, while the DDPG algorithm enhances convergence speed by a factor of ten compared to the soft actor-critic (SAC) approach.

The structure of this paper is as follows. Section 2 presents the CCSFL framework within the ISAC network and defines the optimization problem. In Section 3, the DDQN method is used for FL client selection, while the DDPG algorithm addresses network latency and energy consumption. Section 4 provides the simulation results, and Section 5 concludes the paper.

2 System model and problem formulation

The proposed system is illustrated in Figure 1. A total of N follower satellites $s_{F,i}, i \in \mathbb{N} = \{1, 2, \ldots, N\}$ are distributed across different orbital planes, with each orbit containing a leader satellite $s_{L,p}$. Each follower satellite $s_{F,i}$ acts as an FL client, equipped with a synthetic aperture radar (SAR) for detecting targets via reflected echo signals. Subsequently, Earth images are added to the dataset D_i for local FL training. The ground station (GS) serves as the FL server, aggregating and updating local model parameters. Given the rapid movement of the LEO constellation and the highly dynamic nature of the SGL, leader satellites $s_{L,p}$ are introduced as inter-satellite relays. To maintain the stability of the SGL, only one leader satellite, $s_{L,v}$, which is visible to the ground station, communicates with the GS during each round of FL training. The specific satellite routing methods are discussed in detail in Subsection 2.3.1. Additionally, to minimize communication and computation overhead, only a subset of M ($M < N$) follower satellites participate in federated learning each round, with their local models compressed by the leader satellites. In summary, the main framework proposed in this paper is categorized into the following three models.

2.1 Sensing model

Radar sensing is widely utilized to identify target locations by analyzing the received echo signals. With its exceptional all-day operation, all-weather adaptability, and global coverage capabilities, SAR has considerable potential for industrial and military uses. In this paper, each SAR-equipped FS transmits penetrating electromagnetic waves toward Earth targets, capturing structural features in the resulting images. The pulse signal transmitted from the FS $s_{F,i}$ is defined as $s_i(\tau)$, and the echo signal received at time j can be expressed as [18]

$$r_i(\tau, j) = \left\{ s_i(\tau) W_a \left(\frac{j}{K_{\text{syn}}} \right) \right\} \otimes h(\tau, j), \tag{1}$$

where $W_a(x)$ represents the ideal rectangular window function, K_{syn} denotes the synthetic aperture duration, and $h(\tau, j)$ describes the SAR system function formed by target scattering. Subsequently, the Earth target can be imaged by back-convolution of the system function with the echo signal, and the resulting image is added to the dataset D_i of the follower satellite $s_{F,i}$.

According to the radar equation [19], the sensing SNR can be written as[1)]

$$\gamma_i = \frac{P_i^S G_t^S G_r^S \lambda^2 \sigma}{(4\pi)^3 R_i^4 k T B_{\text{SAR}} L_s L_a}, \tag{2}$$

where P_i^S is the transmit power, G_t^S, G_r^S are the antenna gains of the transmitter and receiver, respectively. λ is the signal wavelength and σ is the radar cross section (RCS) of the target. R_i^k, k, T, B_{SAR}, L_s, and L_a represent distance, Boltzmann constant, system noise temperature, bandwith, system loss, and atmospheric attenuation, respectively.

1) The sensing SNR is further used in the final reputation-utility reward function.

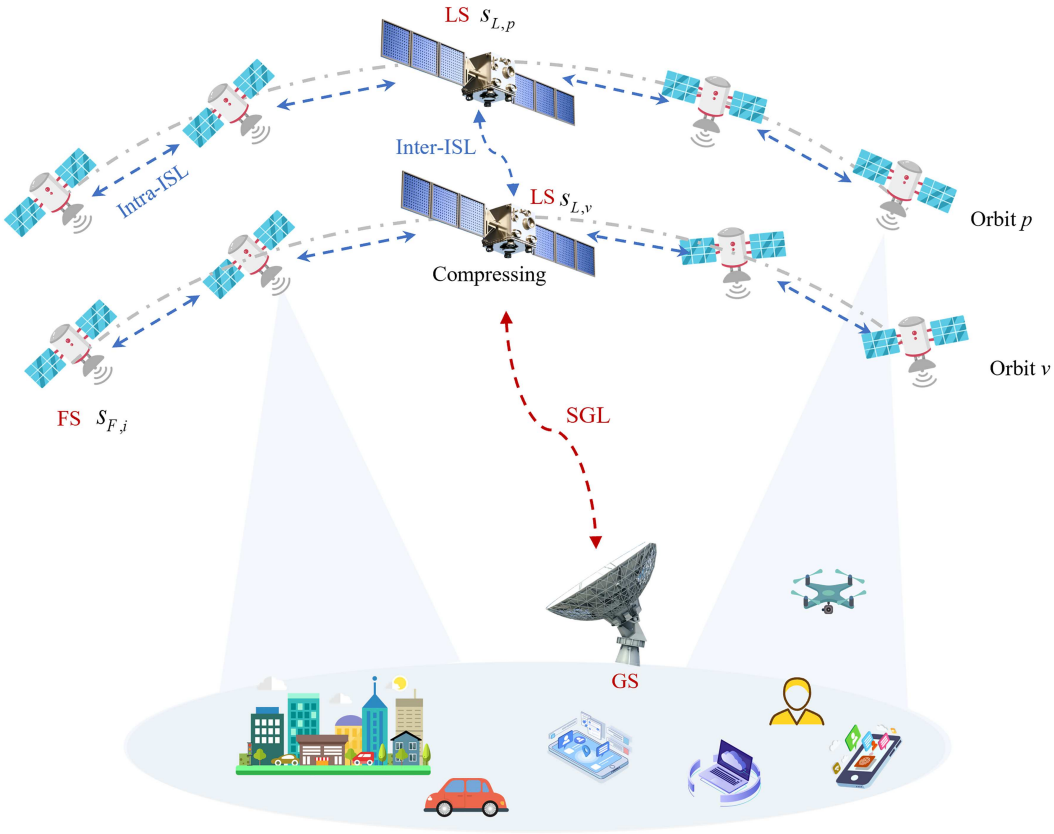

Figure 1 (Color online) FL-based ISAC system under cohesive clustered satellites.

2.2 Computation model

The computation model consists of three segments: (1) local computation, (2) parameter compression, and (3) central aggregation.

2.2.1 *Local computation*

In this stage, the selected FSs $s_{F,i}, i \in \mathbb{M} = 1, 2, \ldots, M$ perform local training on their respective datasets D_i to minimize the local loss function $g_i(x)$. Specifically, $D_i = (x_{i,1}, y_{i,1}), \ldots, (x_{i,q}, y_{i,q}), \ldots, (x_{i,Q_i}, y_{i,Q_i})$, where $x_{i,q}$ and $y_{i,q}$ denote the input features and corresponding output labels of the q-th sample in $s_{F,i}$, respectively, and $Q_i = |D_i|$ represents the total number of samples. For image classification tasks, a commonly used loss function is the cross-entropy function, which is defined as [20]

$$g_i(x_{i,n}, y_{i,n}) = -\sum_{q=1}^{Q_i} y_{i,q} \times \log(p_{i,q}), \tag{3}$$

where $y_{i,n}$ is the correct label, $p_{i,q}$ denotes the probability that sample q belongs to $y_{i,n}$. Therefore, the total loss function in dataset D_i can be written as

$$F(\omega_i^k) = \frac{1}{Q} \sum_{q \in Q_i} g_i(\omega_i^k; x_{i,q}, y_{i,q}), \tag{4}$$

where ω_i^k is the local training model parameter vector for the k-th global iteration of $s_{F,i}$. Therefore, the updated model parameter based on the stochastic gradient descent (SGD) method is written as

$$\omega_i^{k+1} = \omega_G^k - \eta \nabla F(\omega_i^k), \tag{5}$$

Algorithm 1 Sparse compression of local training parameter gradients in s_L.

Initialize model parameter ω_i^k, learning rate l_r, momentum factor m_f, sparsity factor P, momentum buffer V, and gradient residual R;
repeat
  Compute the local gradient $g_i(\omega_i^k) = \omega_i^k - \omega_i^{k-1}$;
  Add the accumulated residuals R to the current gradient;
  Calculate the amplitude of $g_i(\omega_i^k)$;
  Find the top Q largest elements of the gradient vector and set all other elements to zero, denoted as $g_i^{\mathrm{TOP_P}}(\omega_i^k)$;
  Update the momentum buffer $V = m_f \times V + g_i^{\mathrm{TOP_P}}(\omega_i^k)$;
  Update the model parameter $\omega_i^k = \omega_i^k - l_r \times V$;
  Calculate the new gradient residual $R = g_i(\omega_i^k) - g_i^{\mathrm{TOP_P}}(\omega_i^k)$;
until no parameters.

where ω_G^k is the global training model parameter vector in the k-th iteration. Then the computation time in the local training under I iterations is

$$T_{i,k}^{\mathrm{CP}} = \frac{IC_iQ_i}{f_i^k}, \tag{6}$$

where f_i^k is the computing capability allocated by $s_{F,i}$, C_i represents the CPU cycles needed to process a single bit of data. And the corresponding energy consumption is $E_{i,k}^{\mathrm{CP}} = I(\kappa(f_i^k)^2 C_iQ_i)$, where κ is the energy factor.

2.2.2 *Parameter compression*

Although federated learning avoids uploading raw data and reduces communication overhead, the massive transmission of parameters still introduces non-negligible latency. To tackle this issue, leader satellites utilize deep gradient compression (DGC) to substantially lower the required communication bandwidth while preserving model convergence [21]. First, the local parameter gradient vector $g_i(\omega_i^k)$ is computed to enable subsequent compression. Then, $g_i(\omega_i^k)$ is sparsified to $g_i^{\mathrm{TOP_P}}(\omega_i^k)$, transmitting only the top Q elements of the gradient vector with the largest magnitudes, while setting the remaining elements to zero. Gradients not transmitted are aggregated after a certain number of rounds and sparsified at that point, a process known as gradient accumulation. To ensure optimization stability, momentum-based algorithms, such as SGD with Momentum, are used to retain historical momentum information and correct it during sparsification. The detailed gradient compression process is outlined in Algorithm 1. Notably, the latency and energy consumption during this phase are negligible, as the primary objective is to reduce SGL latency.

2.2.3 *Central aggregation*

After collecting the local model parameters' gradient, the federated averaging (FEDAVG) algorithm is used to update the global model, written as

$$g(\omega_G^k) = \sum_{i=1}^{M} \beta_i g(\omega_i^k), \tag{7}$$

and then

$$\omega_G^{k+1} \leftarrow \omega_G^k + g(\omega_G^k). \tag{8}$$

2.3 Communication model

The communication model (as shown in Figure 2) consists of intra-orbit ISL, inter-orbit ISL, and SGL between GS and the visible leader satellite $s_{L,v}$. In the intra-orbit ISL, follower satellites transmit their local models to the leader satellite in the same orbit using a multi-hop routing protocol. The inter-orbit ISL connects leader satellites across different orbital planes, transmitting compressed data using neighbor forwarding techniques. Finally, the SGL establishes communication between the ground station and a selected leader satellite, which is visible to the ground station during the current FL round.

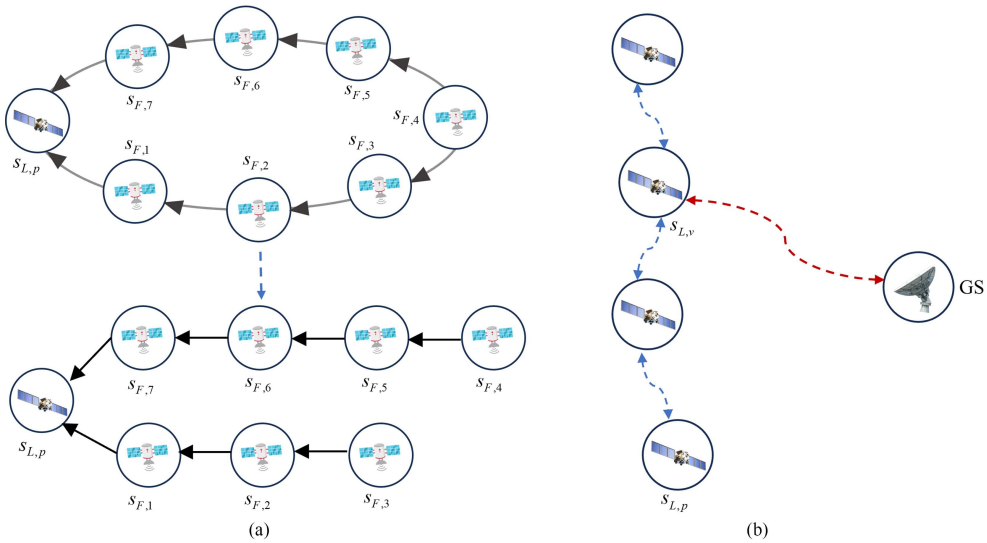

Figure 2 (Color online) (a) Intra-orbit ISL, (b) inter-orbit ISL and SGL models in cohesive clustered satellites.

2.3.1 *ISL*

The intra-orbit ISLs utilize a multi-hop routing protocol [22], as illustrated in Figure 2(a). Given that each satellite is restricted to communication with a maximum of two co-orbital counterparts, ISLs are established exclusively with immediate left and right neighboring satellites. The ISL model parameter transmission is assigned the highest priority, allocating 80% of the total bandwidth to ensure aggregation efficiency. Additionally, quadrature phase-shift keying (QPSK) modulation is employed to balance spectral efficiency and the robustness of federated learning with limited ISL resources. Based on this architecture, each FS $s_{F,i}$ identifies the shortest transmission path to the LS $s_{L,p}$ to facilitate local model updates. In scenarios where the number of FSs is odd, exemplified in Figure 2(a), $s_{F,4}$ exhibits two potential shortest paths. To resolve this ambiguity, a unique path is established by designating $s_{F,i+1}$ as the parent node of $s_{F,i}$, as demonstrated by the hierarchical relationship between $s_{F,5}$ and $s_{F,4}$. Intra-orbit ISLs are established contingent upon line-of-sight visibility, with the communication rate between $s_{F,i}$ and its adjacent satellites $s_{F,i\pm1}$ defined as

$$R_i^k = B_i \log_2 \left(1 + \frac{P_i^{C,k} G_{s_{F,i}} G_{s_{F,\pm1}} |h_i|^2}{\sigma_1^2}\right), \tag{9}$$

where B_i, $P_i^{C,k}$, $G_{s_{F,i}}$ are the bandwidth, power, and antenna gain of $s_{F,i}$, respectively. $G_{s_{F,i\pm1}}$ is the receiving antenna gain of $s_{F,i\pm1}$ and σ_1^2 is the noise power. Therefore, the maximum latency of $s_{F,i}$ in the k-th global round is expressed as

$$T_{i,k}^{\mathrm{CM}_1} = m_i \times \frac{|\omega_i^k|}{R_i^k} + \frac{d}{\mathrm{c}}, \tag{10}$$

where m_i is the multi-hop count, d is the distance between $s_{F,i}$ and $s_{L,p}$, and c is the speed of light. Accordingly, the transmission energy consumption is

$$E_k^{\mathrm{CM}_1} = P_i^{C,k} T_{i,k}^{\mathrm{CM}_1}. \tag{11}$$

Since SGL is feasible only if it satisfies

$$\alpha_{sg}(t) = \pi/2 - \angle(r_g(t), r_s(t) - r_g(t)) \geqslant \alpha_{\min}, \tag{12}$$

where $\alpha_{sg}(t)$ is the elevation angle of the satellite, $\alpha_{\min}$ is the minimum thresholds, $r_g(t)$ and $r_s(t)$ is the position of GS and satellite, respectively. To enhance the transmission efficiency and stability of CCSFL, only one satellite, $s_{L,v}$, that satisfies Eq. (12) is selected to establish the SGL with the ground station. The selected satellite $s_{L,v}$ aggregates the compressed local parameters from the leader satellite $s_{L,p}$ in

each orbit p via inter-orbit ISL. Similar to intra-orbit ISL, the leader satellite $s_{L,p}$ communicates only with its neighboring leader satellites $s_{L,p\pm1}$ in adjacent orbits. Since compressed data are transmitted and laser communication is commonly used in ISLs, the energy consumption and latency at this stage are negligible.

2.3.2 *SGL*

The satellite-ground link follows the Shade-Rice distribution, whose channel coefficient is

$$h_L = h_{\text{shadowed}} \cdot \left(\sqrt{\frac{K_r}{K_r+1}} \cdot h_{\text{LOS}} + \sqrt{\frac{1}{K_r+1}} \cdot h_{\text{NLOS}} \right), \tag{13}$$

where h_{shadowed} is the shadow factor, K_r is the Rice factor, indicating the power ratio of the LOS (line of sight) component (h_{LOS}) to the NLOS (non-line of sight) component (h_{NLOS}). The communication rate is

$$R_2 = B_L \log_2 \left(1 + \frac{P_{s_{L,v}} G_G G_{L,t} |h_L|^2}{\sigma_2^2} \right), \tag{14}$$

where B_L, $P_{s_{L,v}}$, G_G, $G_{L,t}$, and σ_2^2 have similar meanings as in (9). Consequently, the transmission latency and energy consumption are

$$T_k^{\text{CM}_2} = \frac{\sum_{i=1}^{M} \beta_i g(\omega_i^k)}{R_2}, \tag{15}$$

and

$$E_k^{\text{CM}_2} = P_L T_k^{\text{CM}_2}. \tag{16}$$

2.4 Problem formulation

This paper focuses on improving system efficiency by reducing the total latency T_k and energy consumption E_k, which are expressed as

$$T_k = \sum_{i=1}^{M} T_{i,k}^{\text{CP}} + 2 \times \sum_{i=1}^{M} T_{i,k}^{\text{CM}_1} + T_k^{\text{CM}_2} \tag{17}$$

and

$$E_k = \sum_{i=1}^{M} E_{i,k}^{\text{CP}} + 2 \times \sum_{i=1}^{M} E_{i,k}^{\text{CM}_1} + E_k^{\text{CM}_2}, \tag{18}$$

respectively. Notice that the multiplication factor "2" represents an approximation, as the round-trip from $s_{F,i}$ to s_L is considered equivalent. Additionally, an essential factor influencing the FL process is the scheduling of client satellites $s_{F,i}, i \in 1, 2, \ldots, M, M < N$, which can significantly impact FL convergence. Besides, the non-IID nature of datasets on each $s_{F,i}$, stemming from differences in locations and sensing tasks, affects the overall FL iteration count due to the bias introduced by non-IID data. Furthermore, the sensing SNR and system heterogeneity caused by differences in hardware devices can severely affect the operational efficiency of CCSFL. Therefore, this paper aims to address the following problems:

$$\text{(P0)}: \text{minimize} \sum_{k=1}^{K} \alpha_t T_k + \alpha_e E_k, \tag{19a}$$

$$\text{s.t. } \alpha_t + \alpha_e = 1, \tag{19b}$$

$$\text{Maximizing sensing SNR}, \tag{19c}$$

$$\text{Reducing the impact of data and system heterogeneity}, \tag{19d}$$

$$\text{Satisfying power and energy constraints}, \tag{19e}$$

where Eq. (19b) represents the weighting factor of latency and energy consumption. To address (19c) and (19d), a DRL-based client selection scheme is reviewed in Subsection 3.1, where the sensing SNR, data, and system heterogeneity are considered. Therefore, problem P0 can be further visualized as

$$\text{(P1)}: \text{minimize} \sum_{k=1}^{K} \alpha_t T_k + \alpha_e E_k, \tag{20a}$$

$$\text{s.t. } \alpha_t + \alpha_e = 1, \tag{20b}$$

$$\sum_{i=1}^{N} a_i^k = M, \tag{20c}$$

$$0 < f_i^k \leqslant F, \tag{20d}$$

$$0 < P_i^{C,k} \leqslant P, \tag{20e}$$

$$T_k \leqslant T_{k,\max}, \tag{20f}$$

where Eqs. (20d) and (20e) are the computing and power capability constraint, respectively. And finally, Eq. (20f) means the total latency in the k-th interaction cannot exceed the thresholds $T_{k,\max}$. Eq. (20c) indicates the client selection method, where

$$a_i^k = \begin{cases} 1, \ s_{F,i} \text{ has been selected as the FL client,} \\ 0, \ \text{else.} \end{cases} \tag{21}$$

It can be observed that increasing computational capacity or power reduces latency but results in higher energy consumption. Moreover, Eq. (20c) is a binary constraint but Eqs. (20d) and (20e) are continuous constraint. As a result, the proposed problem is categorized as mixed-integer non-linear programming (MINLP), making it difficult to tackle with conventional mathematical approaches. Hence, this paper adopts a DRL algorithm to resolve the challenge.

3 DRL method for the proposed problem

3.1 Optimal satellites client selection

This section uses the DDQN algorithm to select the best M client FSs in each FL's round. Details of each step are described below.

3.1.1 *Markov decision process*

The corresponding Markov decision process consists of a state space, an action space, and a reward function based on sensing SNR.

The state space vector in the k-th FL round can be expressed as

$$s^k = \left\{ s_1^k, \ldots, s_i^k, \ldots, s_N^k \right\}, \tag{22}$$

where s_i^k represents FS $s_{F,i}$'s state at global round k, written as

$$s_i^k = \left\{ \omega_i^k, Q_i^k, c_i, \gamma_i \right\}, \tag{23}$$

where c_i is the number of cores of $s_{F,i}$.

In general, the size of the action space for selecting M clients from N clients is C_N^M, which will cause a large computational burden. Therefore, we introduce an action value function $Q(s^k, a^k)$ to select clients and learn multiple actions simultaneously, where a^k is the action space and expressed as

$$a^k = a_i^k \times \mathbb{N}, \ \sum_{i=1}^{N} a_i^k = M. \tag{24}$$

Ultimately, client selection in each round is determined by the actions corresponding to the top M highest Q values. To further mitigate the risk of overlooking important characteristics of certain clients, the ε-greedy exploration strategy is employed.

In this paper, a reputation function based on sensing SNR is utilized as a reward. The average difference $d\left(\omega_i^k, \omega_G^k\right)$ is defined as the discrepancy between client i's model weights and the global model during the k-th training iteration,

$$d\left(\omega_i^k, \omega_G^k\right) = \frac{1}{\left|\omega_G^k\right|} \sum_{j=1}^{\left|\omega_G^k\right|} \left| \frac{\omega_i^{k,j} - \omega_G^{k,j}}{\omega_G^{k,j}} \right|. \tag{25}$$

We use the sigmoid function as the utility function for its stability regardless of smaller or larger distances. Based on sensing SNR, the utility function is expressed as

$$U_i^k = \frac{\gamma_i}{1 + \exp\left(a \cdot d\left(\omega_i^k, \omega_G^k\right)\right)}, \tag{26}$$

where a is the slope parameter, a higher value of a makes the utility more sensitive to the discrepancy, whereas a lower value reduces its impact. This function ensures higher utility when the local model aligns better with the global model and the sensing SNR is higher. The final reputation-utility reward function based on sensing SNR is expressed as

$$r_i^k = \mu U_i^k + (1 - \mu)\, r_i^{k-1}, \tag{27}$$

where μ adjusts the balance between current and past contributions. It means that we not only value the current round's utility but also consider the historical model's implications.

3.1.2 *DDQN algorithm*

Unlike traditional DQN, this paper adopts the DDQN algorithm to implement the client selection scheme[2)]. This approach reduces the overestimation of Q-values, thereby improving the stability and convergence of the reinforcement learning strategy [23]. The DDQN framework comprises two neural networks: the main network, which is utilized for training, and the target network, responsible for evaluating actions in subsequent states. To stabilize the learning process, the target network is updated every P rounds. The DDQN update rule is given by

$$Q(s^k, a^k) \leftarrow Q(s^k, a^k) + \alpha \left(r_i^k + \beta Q_{\text{target}} \left(\hat{s}^k, \arg\max_{\hat{a}_i^k} Q_{\text{main}}\left(\hat{s}^k, \hat{a}_i^k\right) \right) - Q(s^k, a_i^k) \right), \tag{28}$$

where α is learning rate and β is discount rate. $\tilde{a}^k = \arg\max_{\hat{a}^k} Q_{\text{main}}\left(\hat{s}^k, \hat{a}^k\right)$ is used to select the optimal action in the next state $\hat{s}^k$. And $Q_{\text{target}}(\hat{s}^k, \tilde{a}^k)$ is the Q value estimation procedure.

To mitigate the correlation between consecutive training samples and to improve the utilization of historical experiences, this paper employs an experience replay mechanism. Specifically, the transition tuple $\left(s^k, a^k, r_i^k, \hat{s}^k\right)$, observed during round k, is stored in a experience pool $\mathcal{D}_1$. A small batch of transitions is randomly selected from $\mathcal{D}_1$ during each Q-network update. Subsequently, Q_{main} is used to select the optimal action $\tilde{a}_i^k$ for a given state, while Q_{target} is utilized to approximate the optimal action-value function. Furthermore, the ε-greedy strategy is employed to prevent convergence to suboptimal solutions by encouraging exploration through random action choices with a certain probability.

3.2 Latency and energy consumption minimization

In the previous section, the optimal selection of M client satellites to participate in federated learning during each round was determined. In this section, we employ the DDPG algorithm to minimize the latency and energy consumption of the CCSFL system by optimizing power allocation and computational frequency. DDPG is a powerful deep reinforcement learning algorithm specifically designed to handle continuous action spaces [24]. By utilizing the Actor-Critic framework alongside experience replay and target networks, DDPG achieves notable stability and convergence in reinforcement learning tasks.

2) DDQN is well-suited for discrete action spaces, where client selection is a discrete decision problem.

Algorithm 2 DDPG algorithm for latency and energy consumption minimization.

Initialize $\mu(s_2^k|\theta^\mu)$, $Q(s_2^k, a_2^k|\theta^Q)$, $\mu'(s_2^k|\theta^{\mu'})$, $Q'(s_2^k, a_2^k|\theta^{Q'})$, and $\mathcal{D}_2$;
for episode $= 1, \ldots, K$ **do**
  Obtain initial state s_2^k;
  for $j = 1, \ldots, J$ **do**
    Choose a power and frequency action $a_2^k = \mu(s_2^k|\theta^\mu) + \mathcal{N}_t$ under state s_2^k;
    Execute action a_2^k, observing the reward r_2^k and the new state $\hat{s}_2^k$;
    Store transition (s_2^k, a_2^k , r_2^k, $\hat{s}_2^k$) into $\mathcal{D}_2$;
    Randomly sampling small batches ($s_2^k(j)$, $a_2^k(j)$, $r_2^k(j)$, $\hat{s}_2^k(j)$) from the experience pool for training;
    Calculating target Q value by $y = r_2^k + \eta Q'(\hat{s}_2^k, \mu'(\hat{s}_2^k \mid \theta^{\mu'}) \mid \theta^{Q'})$;
    Minimizing the loss function $L = E[(y - Q(s_2^k, a_2^k|\theta^Q))^2]$;
    Update actor network by $\nabla_{\theta^\mu} J \approx \frac{1}{|\mathcal{D}_2|} \sum_j \nabla_{a_2^k} Q(s_2^k, a_2^k|\theta^Q)|_{a_2^k = \mu(s_2^k)} \nabla_{\theta^\mu} \mu(s_2^k|\theta^\mu)$;
    Update the parameters of the target network with rate τ: $\theta^{\mu'} = \tau\theta^\mu + (1-\tau)\,\theta^{\mu'}$, $\theta^{Q'} = \tau\theta^Q + (1-\tau)\,\theta^{Q'}$;
  end for
end for
Return result.

3.2.1 *Markov decision process*

The state space s_2^k can be written as

$$s_2^k = \left\{ f_i^k, \ldots, f_M^k, P_1^{C,k}, \ldots, P_M^{C,K}, g\left(\omega_1^k\right), \ldots, g\left(\omega_M^k\right) \right\}. \tag{29}$$

The continuous action space is represented as

$$a_2^k = a_2^k = \left\{ f_i^k \in [0, F], P_i^{C,k} \in [0, P] \right\}, \tag{30}$$

and the reward function is $r_2^k = -(\alpha_t T_k + \alpha_e E_k)$.

3.2.2 *DDPG algorithm*

The main network of the DDPG contains an actor network $\mu(s_2^k|\theta^\mu)$ and a critic network $Q(s_2^k, a_2^k|\theta^Q)$ [25]. Correspondingly, its target network also includes an actor network $\mu'(s_2^k|\theta^{\mu'})$ and a critic network $Q'(s_2^k, a_2^k|\theta^{Q'})$. The action a_2^k is obtained by $a_2^k = \mu\left(s_2^k|\theta^\mu\right) + \mathcal{N}_t$, where $\mathcal{N}_t$ is the noise used for exploration. The target Q value is calculated by

$$y = r_2^k + \eta Q'\left(\hat{s}_2^k, \mu'\left(\hat{s}_2^k \mid \theta^{\mu'}\right) \mid \theta^{Q'}\right), \tag{31}$$

where $\hat{s}_2^k$ is the next state of s_2^k. Then, the critic network undergoes an update process by minimizing the loss function defined as follows:

$$L = E\left[\left(y - Q\left(s_2^k, a_2^k\left|\theta^Q\right.\right)\right)^2\right], \tag{32}$$

where $E[\cdot]$ is the expectation. Finally, the actor network is updated using the deterministic policy gradient theorem, written as

$$\nabla_{\theta^\mu} J \approx \frac{1}{|\mathcal{D}_2|} \sum_j \nabla_{a_2^k} Q(s_2^k, a_2^k|\theta^Q)|_{a_2^k = \mu(s_2^k)} \nabla_{\theta^\mu} \mu(s_2^k|\theta^\mu), \tag{33}$$

where $|\mathcal{D}_2|$ is the size of experience pool $\mathcal{D}_2$. The DDPG algorithm for the latency and energy consumption minimization is described in Algorithm 2.

4 Simulation results

In this section, we perform numerical simulations to evaluate the effectiveness of the proposed algorithm. The MNIST and CIFAR-10 datasets are utilized for federated learning validation, with non-IID data distributions simulated using FedLab [26]. FedLab is an open-source Python library designed for FL research and experimentation, which facilitates the implementation of communication, model updates, and data partitioning.

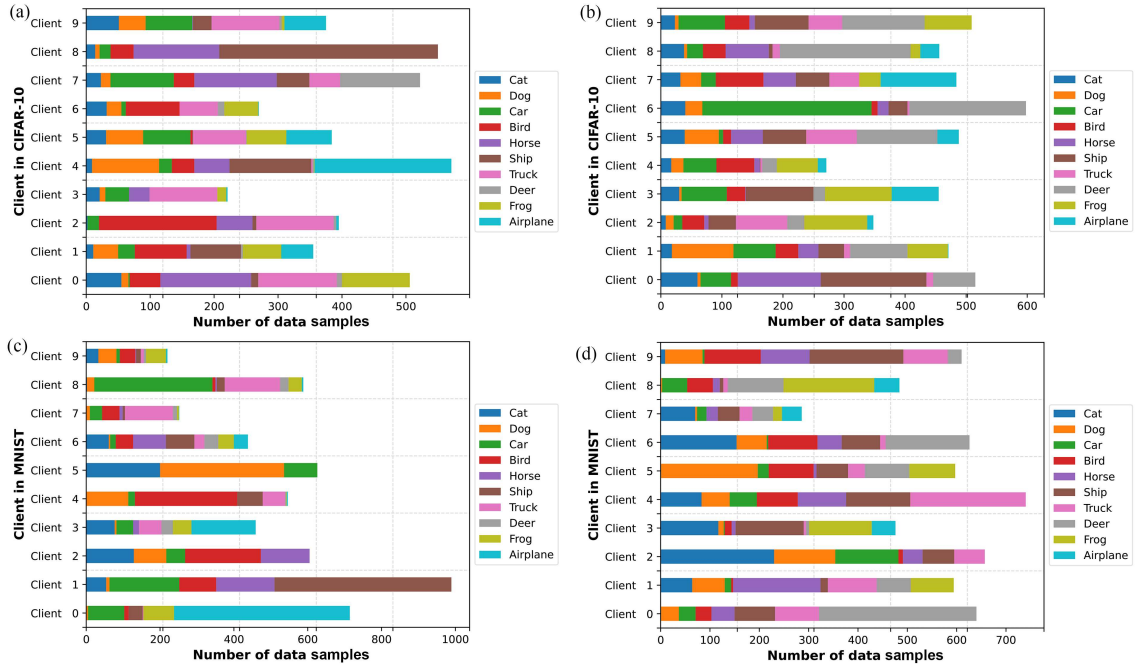

Figure 3 (Color online) Heterogeneous datasets: non-IID CIFAR-10 and MNIST datasets with different partition parameters. (a) Non-IID CIFAR-10 datasets with $\alpha_{\mathrm{d}} = 0.5$; (b) non-IID CIFAR-10 datasets with $\alpha_{\mathrm{d}} = 0.9$; (c) non-IID MNIST datasets with $\alpha_{\mathrm{d}} = 0.5$; (d) non-IID MNIST datasets with $\alpha_{\mathrm{d}} = 0.9$.

Table 1 Simulation settings.

Parameter in system	Value in system	Parameter in optimization	Value in optimization
Number of FS N	10	Weighting factor α_t and α_e	0.5 and 0.5
Selected client FS M	5	FL training round	100
RCS of the earth targets σ	10 m^2	Size of experience pool	$\lvert\mathcal{D}_1\rvert = \lvert\mathcal{D}_2\rvert = 10000$
Intra ISL bandwith B_i	20 Mbps	Learning rate	0.01
Intra ISL transmission capacity P	5 W	Reputationparameter μ	0.8
Local computation capacity F	3 GHz	$\varepsilon_{\mathrm{init}}$ and $\varepsilon_{\mathrm{end}}$ for greedy exploration	0.9 and 0.2
Distance of SGL R_2	500 km	Batchsize	50

The non-IID distribution of the CIFAR-10 and MNIST datasets, partitioned by FedLab, is illustrated in Figure 3. MNIST includes 70000 28×28 grayscale images of handwritten digits (0−9) [27], while CIFAR-10 contains 60000 32×32 color images across 10 categories like airplanes, cars, and birds [28]. Both datasets are widely used for image classification and serve as benchmarks for machine learning algorithms. We use the Hetero Dirichlet partitioning method to assign labels to the MNIST and CIFAR-10 datasets, with the data distribution governed by the Dirichlet distribution parameter α_{d}. A larger value of α_{d} (e.g., $\alpha_{\mathrm{d}} = 0.9$) results in a more balanced distribution of data categories among clients, whereas a smaller value (e.g., $\alpha_{\mathrm{d}} = 0.5$) leads to fewer data categories per client, resulting in a more non-IID distribution.

Next, we compare the DDQN scheme proposed in Subsection 3.1 with "FAVOR" and federated proximal optimization (FedProx) schemes, and the DDPG scheme proposed in Subsection 3.2 with the SAC and baseline schemes. This comparison aims to further validate the advantages of the proposed system in terms of latency, energy consumption, and convergence speed. The specific parameters are listed in Table 1.

Figure 4 illustrates the convergence of FL accuracy on the MNIST and CIFAR-10 datasets with different partitioning factors, $\alpha_{\mathrm{d}} = 0.8$ and $\alpha_{\mathrm{d}} = 0.5$. The results show that $\alpha_{\mathrm{d}} = 0.8$ achieves higher accuracy and faster convergence on both MNIST and CIFAR-10, indicating that non-IID data significantly degrade federated learning performance. Furthermore, the MNIST dataset demonstrates faster convergence and higher training accuracy compared with CIFAR-10 when using the same neural network model. This is attributed to the fact that MNIST has lower resolution and simpler features, which allow the model to fit the data effectively without requiring extensive regularization. However, despite MNIST's superior

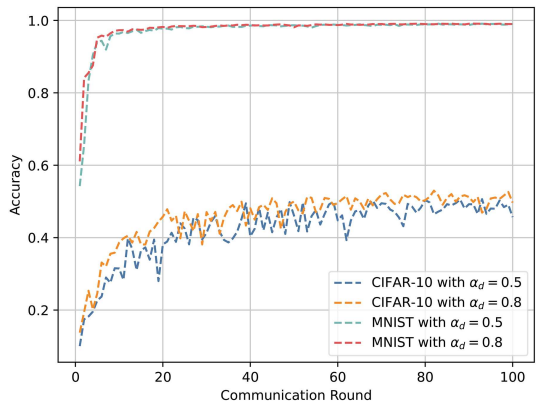

Figure 4 (Color online) Convergence of FL accuracy for MNIST and CIFAR-10 datasets with different partitioning factors $\alpha_{\mathrm{d}} = 0.8$ vs. $\alpha_{\mathrm{d}} = 0.5$.

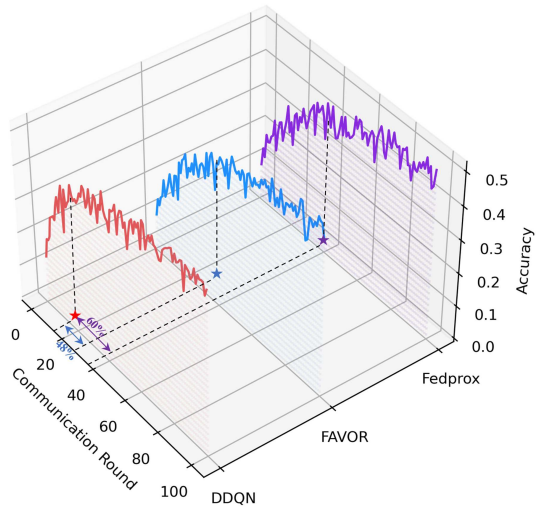

Figure 5 (Color online) Heterogeneous datasets and system: convergence of DDQN, FAVOR, and FedProx algorithms under the partitioning factor $\alpha_{\mathrm{d}} = 0.5$.

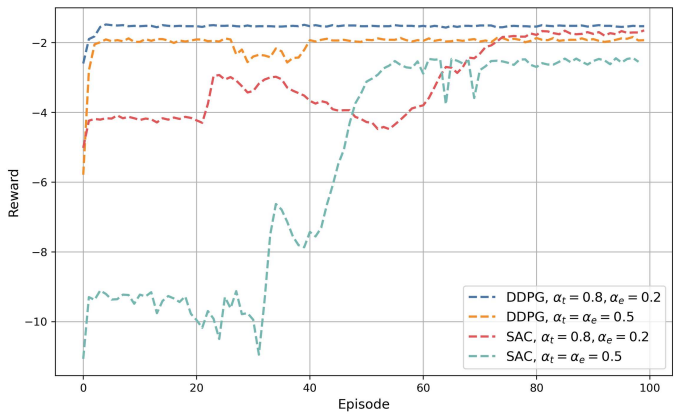

Figure 6 (Color online) Reward comparison under DDPG vs. SAC algorithm with different weighting factors α_t and α_e.

accuracy and convergence speed, its simulation curves have poor visibility. Therefore, the following validation of the proposed DDQN scheme focuses on the CIFAR-10 dataset.

Figure 5 compares the convergence of the proposed DDQN algorithm, "FAVOR" in [15], and the FedProx algorithm with partitioning factors $\alpha_{\mathrm{d}} = 0.5$, respectively. The "FAVOR" method employs the DQN approach to address non-IID datasets but overlooks system heterogeneity and sensing accuracy. FedProx, an improved federated learning algorithm, enhances stability by introducing a proximal term to align local updates with the global model [29]. It addresses training instability and convergence challenges caused by client heterogeneity, such as differences in computing power and non-IID data. The results show that the DDQN algorithm achieves higher accuracy, converging 48% faster than 'FAVOR' and 60% faster than FedProx. It is because DDQN employs a dual Q-network architecture to mitigate overestimation bias and dynamically adjusts its learning strategy based on client capabilities, enhancing robustness with non-IID data and overall performance.

Figure 6 compares the rewards of the proposed DDPG algorithm with the SAC algorithm under different latency and energy weighting factors. SAC is an off-policy reinforcement learning algorithm designed to maximize both reward and entropy, encouraging efficient exploration [30]. Using an actor-critic framework with stochastic policies, it is particularly well-suited for continuous action spaces, effectively balancing exploration and exploitation. The results clearly show that the DDPG algorithm achieves significantly higher rewards and faster convergence compared with SAC, regardless of α_t and α_e. Specifically, when $\alpha_t = 0.8$ and $\alpha_e = 0.2$, DDPG converges by the 8th round, whereas SAC requires 80 rounds, making

DDPG nearly 10 times faster in terms of convergence and improving 22% latency-energy reward. It is because SAC uses stochastic policies for exploration, while DDPG's deterministic policy enables faster learning with less exploration noise. DDPG's soft target update mechanism reduces training instability, leading to quicker policy refinement and avoiding SAC's entropy trade-off.

5 Conclusion

This paper proposes an FL-based ISAC network for cohesive clustered satellites, optimizing FL convergence, satellite transmission power, and computational frequency under heterogeneous datasets and systems. Additionally, intra-orbit ISL multi-hop routing, inter-orbit neighbor forwarding, and SGL gradient sparse compression techniques are employed to enhance the system's communication efficiency. Specifically, to address the heterogeneity of satellite detection data and systems, we introduce a reputation-utility function based on the sensing SNR as a reward for the DDQN algorithm, enabling optimal client selection and reducing the number of FL iterations. To further optimize latency and energy consumption in the CCSFL system, we utilize the DDPG algorithm to refine satellite transmission power and computational frequency. Simulation results validate the effectiveness of the proposed algorithms, where the FL convergence speed and the latency-energy reward outperform benchmarks. In future work, we will explore hierarchical FL architectures for mega-constellations alongside real-time adaptation to orbital dynamics, such as Doppler shift compensation in inter-satellite links.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 62271399) and National Key Research and Development Program of China (Grant Nos. 2022YFB1807102, 2024YFB3910100).

References

1 Lu K, Liu H, Zeng L, et al. Applications and prospects of artificial intelligence in covert satellite communication: a review. Sci China Inf Sci, 2023, 66: 121301
2 Zeng G, Zhan Y, Xie H. Channel allocation for mega LEO satellite constellations in the MEO-LEO networked telemetry system. IEEE Internet Things J, 2023, 10: 2545–2556
3 Guo N, Liu L, Zhong X. Task-aware distributed inter-layer topology optimization method in resource-limited LEO-LEO satellite networks. IEEE Trans Wireless Commun, 2024, 23: 3572–3585
4 Jansson G. Telesat lightspeed^{TM}-enabling mesh network solutions for managed data service flexibility across the globe. In: Proceedings of the IEEE International Conference on Space Optical Systems and Applications (ICSOS), Kyoto City, 2022. 232–235
5 Neinavaie M, Khalife J, Kassas Z M. Acquisition, Doppler tracking, and positioning with Starlink LEO satellites: first results. IEEE Trans Aerosp Electron Syst, 2022, 58: 2606–2610
6 Deng R, Di B, Song L. Ultra-dense LEO satellite based formation flying. IEEE Trans Commun, 2021, 69: 3091–3105
7 Xu L, Jiao J, Jiang S Y, et al. Semantic-aware coordinated transmission in cohesive clustered satellites: utility of information perspective. Sci China Inf Sci, 2024, 67: 199301
8 Sheng M, Zhou D, Bai W G, et al. Coverage enhancement for 6G satellite-terrestrial integrated networks: performance metrics, constellation configuration and resource allocation. Sci China Inf Sci, 2023, 66: 130303
9 Cao B, Zhao J W, Liu X, et al. Adaptive 5G-and-beyond network-enabled interpretable federated learning enhanced by neuroevolution. Sci China Inf Sci, 2024, 67: 170306
10 Han D J, Hosseinalipour S, Love D J, et al. Cooperative federated learning over ground-to-satellite integrated networks: joint local computation and data offloading. IEEE J Sel Areas Commun, 2024, 42: 1080–1096
11 Wang P, Li H, Chen B. FL-task-aware routing and resource reservation over satellite networks. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM), Rio de Janeiro, 2022. 2382–2387
12 Razmi N, Matthiesen B, Dekorsy A, et al. Scheduling for on-board federated learning with satellite clusters. In: Proceedings of the IEEE Globecom Workshops (GC Wkshps), Kuala Lumpur, 2023. 426–431
13 Duan C, Li Q, Chen S. FedAC: satellite-terrestrial collaborative federated learning with alternating contrastive training. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM), Kuala Lumpur, 2023. 5733–5738
14 Xiong T, Xu X, Du P, et al. Energy-efficient federated learning for earth observation in LEO satellite systems. In: Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Dubai, 2024. 1–6
15 Wang H, Kaplan Z, Niu D, et al. Optimizing federated learning on non-IID data with reinforcement learning. In: Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Toronto, 2020. 1698–1707
16 Li C, Jiang J, Chen W, et al. Realtime multiuser multicarrier communications. IEEE Trans Wireless Commun, 2025, 24: 1234–1251
17 Yin L, Liu Z, Shankar M R B, et al. Integrated sensing and communications enabled low Earth orbit satellite systems. IEEE Netw, 2024, 38: 252–258

18 Sommervold O, Gazzea M, Arghandeh R. A survey on SAR and optical satellite image registration. Remote Sens, 2023, 15: 850

19 Zhou R, Chen J X, Huang Y, et al. A compact MIMO automotive radar using phase-aligned daisy-chain cascading topology and elevation compensation for 2D angle estimation. Sci China Inf Sci, 2023, 66: 162305

20 Zhang X M, Fan S T, Song Z H. Reinforcement learning-based cost-sensitive classifier for imbalanced fault classification. Sci China Inf Sci, 2023, 66: 212201

21 Yan G, Li T, Huang S L, et al. AC-SGD: adaptively compressed SGD for communication-efficient distributed learning. IEEE J Sel Areas Commun, 2022, 40: 2678–2693

22 Razmi N, Matthiesen B, Dekorsy A, et al. On-board federated learning for satellite clusters with inter-satellite links. IEEE Trans Commun, 2024, 72: 3408–3424

23 Zhang R, Xiong K, Lu Y, et al. Joint coordinated beamforming and power splitting ratio optimization in MU-MISO SWIPT-enabled HetNets: a multi-agent DDQN-based approach. IEEE J Sel Areas Commun, 2022, 40: 677–693

24 Zheng K, Luo R, Liu X, et al. Distributed DDPG-based resource allocation for age of information minimization in mobile wireless-powered Internet of Things. IEEE Internet Things J, 2024, 11: 29102–29115

25 Wang J, Wang Y, Cheng P, et al. DDPG-based joint resource management for latency minimization in NOMA-MEC networks. IEEE Commun Lett, 2023, 27: 1814–1818

26 Zeng D, Liang S, Hu X, et al. FedLab: a flexible federated learning framework. J Mach Learn Res, 2023, 24: 1–7

27 Liu L, Zhang J, Song S, et al. Binary federated learning with client-level differential privacy. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM), Kuala Lumpur, 2023. 3849–3854

28 Yu Y, Shin S, Lee S, et al. Block selection method for using feature norm in out-of-distribution detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, 2023. 15701–15711

29 Bouaziz S, Benmeziane H, Imine Y, et al. FLASH-RL: federated learning addressing system and static heterogeneity using reinforcement learning. In: Proceedings of the IEEE International Conference on Computer Design (ICCD), Washington, 2023. 444–447

30 Li M, Ma S, Si P, et al. Relay selection and resource allocation for Ad Hoc networks-assisted train-to-train communications: a federated soft actor-critic approach. IEEE Trans Veh Technol, 2024, 73: 15359–15371