Proceedings of the 4th International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2026.34791

Federated Learning for Space-Air-Ground Integrated Networks: Communication-Efficient Techniques and Research Challenges

Hongyi Liu

School of Intelligent Manufacturing, Guangzhou College of Applied Science and Technology, Guangdong, China
hy1840881186@gmail.com

Abstract. The Space-Air-Ground Integrated Network (SAGIN), a key 6G architecture for global ubiquitous coverage, faces intelligent processing from distributed, large-scale, and heterogeneous data. By keeping data local and only transferring model updates, federated learning provides a viable approach to distributed machine learning within SAGIN. However, the dynamic topology, severe communication constraints, and heterogeneity of satellite networks impose formidable challenges on generic FL algorithms. This paper systematically reviews communication-efficient federated learning technologies for SAGIN. It analyzes the SAGIN architecture and satellite network characteristics, revealing the fundamental constraints they place on distributed learning. A taxonomy of communication optimization techniques is established, and the performance trade-offs among gradient quantization, sparsification, and error compensation mechanisms are compared. The technological evolution is examined along three dimensions: dynamic topology adaptation, communication efficiency optimization, and statistical heterogeneity handling,identifying core development challenges and proposing future research directions. This study provides a systematic reference for research on space-air-ground intelligent networks.

Keywords: Space-Air-Ground Integrated Network, Federated learning, Communication efficiency, Gradient compression, Satellite networks

1. Introduction

The sixth-generation (6G) communication network aims to create an intelligent, ubiquitous connectivity architecture that deeply integrates terrestrial, aerial, and space-based resources [1]. The Space-Air-Ground Integrated Network (SAGIN) coordinates low-, medium-, and high-Earth-orbit satellites, high-altitude platforms, unmanned aerial vehicles, and terrestrial base stations and edge nodes, and has emerged as a key 6G infrastructure [2]. As modern satellite systems evolve from simple communication relays into complex sensing platforms, the data generated by an entire large-scale constellation can reach the petabyte (PB) level per day [3]. Given the exponential growth of massive and privacy-sensitive data, the conventional "transmit-then-process" paradigm has encountered insurmountable structural bottlenecks. For instance, a typical Low Earth Orbit (LEO)

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.34791

satellite possesses a cumulative daily downlink capacity of less than 10 GB, which is orders of magnitude below the requirements for real-time raw data offloading [4]. In the face of such massive and often sensitive data, the traditional "download first, process later" paradigm encounters fundamental bottlenecks. Taking a typical LEO satellite as an example, its total daily downlink capacity is less than 10 GB, far from sufficient for real-time transmission of raw data [4]. Federated Learning (FL) offers a transformative solution by keeping data local while moving models. Under server coordination, multiple clients train on local data and upload only parameter updates, while raw data remain stored at the source [5]. This design substantially reduces communication overhead and privacy risks, and its distributed architecture aligns well withSAGIN's intrinsic characteristics.

However, deploying FL in SAGIN environments introduces fundamental challenges distinct from terrestrial scenarios. Dynamic network topology leads to highly uncertain client availability. Severe communication constraints force the per-round communication volume to be compressed to the megabyte (MB) or even kilobyte (KB) level. Cross-regional shifts in data distribution induce severe statistical heterogeneity, and disparities in on-board computation and energy resources exacerbate training instability.

Although existing studies explore FL applications, a systematic review is still lacking, and no unified technical taxonomy has been developed to adapt algorithm design to the physical characteristics of SAGIN. This paper systematically reviews communication-efficient FL technologies for SAGIN. It analyzes the constraints unique to SAGIN and surveys current taxonomies for communication optimization. By investigating typical FL frameworks, this study aims to provide theoretical guidance.

2. Fundamentals and core challenges of Federated Learning

2.1. Basic principles of Federated Learning

The objective of Federated Learning is to collaboratively train a global model without aggregating raw data. Considering a system consisting of K clients, each client K possesses a local dataset D_k of size $|D_k|$. The goal is to learn the global model parameter w by minimizing the weighted empirical loss:

$$\min_{w} \sum_{k=1}^{K} \frac{d_k}{d} f_k\left(w\right) \tag{1}$$

$$f_k\left(w\right) = \frac{1}{d_k} \sum_{x \in D_k} f\left(w, x\right) \tag{2}$$

where $f(w, x)$ denotes the sample loss function. The classical Federated Averaging (FedAvg) algorithm is a representative implementation of this paradigm [5]. After the server initializes the global model, a subset of clients is selected in each round to receive the current model. Each client performs multi-step stochastic gradient descent (SGD) on its local data and uploads the resulting updates, which the server then weights and averages to form the global model for the next round. Although FedAvg reduces the number of communication rounds through local multi-step updates, its convergence performance remains constrained by data heterogeneity. To address this, several improvements have been proposed. For example, FedProx introduces a proximal term to prevent local updates from deviating excessively from the global model, and SCAFFOLD employs control variates to estimate the global gradient direction, thereby correcting bias in client updates [6, 7].

DOI: 10.54254/2755-2721/2026.34791

2.2. Core challenges in the SAGIN environment

The SAGIN environment imposes three fundamental challenges on federated learning: dynamic network topology, severe communication constraints, and data and system heterogeneity. The highly dynamic nature of SAGIN topologies results in stochastic client availability. Low Earth orbit (LEO) satellites travel at approximately 7.8 km/s, and their visibility windows with ground stations typically last only 5 to 15 minutes, though intervals between windows can extend to several hours. Under the synchronous update mode assumed by most FL algorithms, the set of available clients varies significantly from round to round; some satellites may miss participation due to expired windows, leading to stale or lost updates [5].

Severe communication constraints further limit the amount of data that can be exchanged per round. Satellite-to-ground links typically offer bandwidths of only 10 to 100 Mbps, and atmospheric attenuation further degrades these already scarce communication resources. For example, a single ResNet50 model in 32-bit floating-point format requires approximately 100 MB. Transmission over a 10 Mbps link takes nearly 80 seconds, whereas a typical visibility window lasts only about 300 seconds. Consequently, at most three communication rounds can be completed within a single visibility window, far fewer than the tens to hundreds of rounds generally needed for FL convergence. The per-round communication volume must therefore be compressed to the MB or even KB level, corresponding to compression ratios of tens to hundreds.

Data and system heterogeneity further complicate convergence at the algorithmic level. The distribution of satellite observation data differs markedly across orbits and geographical locations. For instance, polar satellites primarily observe ice- and snow-covered regions, whereas equatorial satellites monitor tropical rainforests and oceans. Such Non-IID data distributions lead to conflicting local update directions, causing the global model to converge slowly or even diverge [8]. Simultaneously, substantial differences exist in computational capabilities and energy budgets among satellites; the computing power of some legacy satellites may be only one-tenth that of newer models. As a result, the assumption of uniform local iteration steps in conventional FL no longer holds, and system heterogeneity amplifies the straggler effect.

2.3. Communication optimization techniques for Federated Learning

To address the severe communication constraints of SAGIN, researchers have proposed techniques that reduce per-round communication volume. These techniques primarily fall into the categories of quantization and sparsification, while error compensation mechanisms are critical to maintaining convergence under compression.

Radient quantization reduces communication overhead by lowering the representation precision of each gradient value. QSGD stochastically quantizes gradients into a small number of discrete levels, improving efficiency at the cost of precision while guaranteeing unbiasedness [9]. SignSGD transmits only the sign of each gradient (1 bit), reducing overhead by a factor of 32, though its convergence is relatively slow and sensitive to noise [10]. TernGrad lies between these two, quantizing gradients into ternary values $\{-\gamma, 0, \gamma\}$ to achieve a relatively high compression ratio while maintaining computational simplicity [11].

Gradient sparsification transmits only the most important gradient components, setting most small gradients to zero. Top-K sparsification is a typical representative; it transmits only the K gradients with the largest absolute values and can achieve compression ratios of several hundred times [12]. However, directly discarding small gradients results in cumulative information errors that, if left unaddressed, cause convergence failure. To solve this problem, Deep Gradient

DOI: 10.54254/2755-2721/2026.34791

Compression (DGC) introduces an error-compensation mechanism. The client retains the untransmitted gradient residuals after each round and accumulates them with the current gradients before sparsification in the next round, ensuring that all gradient information is eventually transmitted. This mechanism enables DGC to achieve compression ratios of 270×-600× with almost lossless accuracy at a 300× setting [12]. Table 1 provides a detailed comparison of these methods in terms of mechanism, performance, and applicability.

Table 1. Comparison of typical gradient compression methods

Method	Compression Mechanism	Typical Compression Ratio	Advantages	Disadvantages	Application Scenarios
QSGD-4bit	Stochastic Quantization	8×	Sparsification with Error Feedback	Noise grows as bit-width decreases	Moderate bandwidth scenarios
SignSGD	1-bit Sign Transmission	32×	Extremely low communication cost	Slow convergence, noise-sensitive	Robustness-tolerant tasks
DGC (Top-0.1%)	Sparsification with Error Feedback	300×	High compression, near-lossless convergence	Extra computation and memory for residual accumulation	Bandwidth-starved satellite downlinks
TernGrad	Ternary Quantization	32×	Computationally lightweight	Sensitive to gradient distribution	Scenarios with bounded gradient distributions

3. Federated Learning research for SAGIN

Research focused on three key directions: hierarchical asynchronous aggregation architectures for dynamic topology and intermittent connectivity; personalized and domain adaptation methods for data and system heterogeneity; and semantic and hybrid compression techniques for severe communication constraints.

3.1 Hierarchical asynchronous aggregation architecture for dynamic topology and intermittent connectivity

Hierarchical asynchronous aggregation has proven effective for addressing short satellite visibility windows, intermittent connectivity, and dynamic client sets in LEO constellations. The core idea is to introduce an intermediate aggregation layer that groups satellites by orbital plane or geographic region. Within each group, high-frequency aggregation uses high-speed inter-satellite links (ISLs), while low-frequency global synchronization across groups occurs via ground stations or GEO satellites. This design decomposes the global synchronization wait time into multiple local synchronization processes, significantly mitigating the training stagnation caused by waiting for all satellites.

3.1.1 RAFL: Reputation-aware hierarchical Federated Learning

RAFL (Reputation-Aware Federated Learning) introduces a hierarchical FL architecture integrated with a reputation mechanism [13]. Satellites are divided into groups by orbital region, and the satellite with the highest reputation value in each group is elected as the edge aggregator. The server

DOI: 10.54254/2755-2721/2026.34791

evaluates the consistency between each satellite's model updates and the global model using cross-validation, dynamically adjusting reputation scores. During intra-group aggregation, the edge aggregator collects updates from satellites within the same group and generates a local model via weighted averaging, with weights proportional to each satellite's reputation. For inter-group aggregation, the local models are transmitted to the global server via ISLs or ground stations. Experimental results show that RAFL effectively resists Byzantine attacks from up to 30% of malicious satellites, and its hierarchical design reduces the risk of single-point failures and long-distance communication latency [13].

3.1.2 HiSatFL: A three-layer heterogeneous collaborative architecture

HiSatFL proposes a three-layer architecture comprising a LEO sensing layer, a MEO aggregation layer, and a GEO coordination layer [8]. LEO satellites serve as data-acquisition and local-training nodes. MEO satellites act as regional aggregators, collecting model updates from multiple LEO satellites and applying domain adaptation to handle cross-orbital data distribution disparities. GEO satellites provide continuous global coverage and manage final model aggregation and broadcasting. This architecture exploits the complementary strengths of different orbital altitudes, improving global model convergence speed by approximately 40% in simulations with four orbital planes [8].

Hierarchical asynchronous aggregation effectively mitigates dynamic topology issues, but it relies on fast ISLs. ISL deployment is constrained by satellite positions and pointing accuracy, and stable connections cannot be guaranteed across all orbital planes. Moreover, the hierarchical structure adds system complexity, requiring extra scheduling and synchronization to maintain cross-layer consistency.

3.2. Adaptation methods for data and system heterogeneity

Given data heterogeneity arising from differing orbital positions and observation targets, as well as system heterogeneity in computing and storage capabilities, forcing all satellites to learn a single global model is clearly suboptimal. Consequently, enabling satellites to retain personalized components and employing domain adaptation to align feature distributions across orbits have become mainstream research directions.

3.2.1. Domain adaptation mechanism in HiSatFL

HiSatFL employs domain adaptation techniques to address cross-orbital data distribution shifts [8]. LEO satellites share a common feature extractor (encoder) while maintaining their respective personalized classification heads. At the MEO aggregation layer, domain adaptation methods such as Maximum Mean Discrepancy (MMD) or adversarial training align the feature distributions extracted by satellites in different orbits, enabling the shared feature extractor to generalize across multiple orbits. Meanwhile, each satellite's classification head is fine-tuned on local data to adapt to its region's class distribution. On the EuroSAT dataset (10 land cover classes), this method improved accuracy from 76.3% (FedAvg) to 84.5% [8].

3.2.2. EMSFL: Expert Model Splitting

EMSFL (Expert Model Splitting) offers another effective personalization strategy [14]. This framework decomposes the model into a shared feature extractor and multiple expert classification heads. Based on a satellite's data distribution characteristics (assessed from metadata or a small

DOI: 10.54254/2755-2721/2026.34791

sample set), the system dynamically assigns the most appropriate expert head. Expert heads are synchronized through ISLs, sharing knowledge while preserving parameter independence. It enables global knowledge sharing while satisfying local personalization requirements. It is especially suitable for scenarios with large differences in data distribution but limited category sets. When simulating different orbits' data distributions on the CIFAR-10 dataset, EMSFL achieved approximately 6% higher accuracy than FedAvg and reduced communication rounds by 20% [14].

Personalization and domain adaptation methods effectively alleviate convergence oscillation and accuracy degradation caused by data heterogeneity. However, these gains come at the cost of increased storage and computational overhead. Satellites must maintain their own model components (e.g., classification heads), and additional mechanisms are needed to coordinate updates between shared and personalized parts. For resource-constrained microsatellites, the feasibility of such methods still requires verification.

3.3. Semantic and hybrid compression for severe communication constraints

On satellite-to-ground links with extremely limited bandwidth, general-purpose gradient compression alone may still prove insufficient. Therefore, some studies have explored more aggressive compression methods or integration with emerging semantic communication technologies, aiming to transmit more effective information within short communication windows.

3.3.1. SemSpaceFL: Semantic communication-driven compression

SemSpaceFL introduces semantic communication into the hierarchical FL framework. Instead of transmitting complete gradient vectors, each satellite first extracts semantic features from its model update, such as attention maps and significant gradient distributions [15]. After local training, the client computes the update, maps it into a low-dimensional semantic vector via a pre-trained semantic encoder (e.g., an autoencoder), and uploads the quantized vector. Upon receiving the semantic vector, the server reconstructs an approximate update through a decoder for global aggregation. A semantics-based approach to data drastically compresses communication volume; simulation results show a compression ratio of up to 500 × with an accuracy loss of only about 2% [15]. Semantic compression is better suited to tasks with clear objectives (e.g., fixed classification), whereas general gradient compression is more suitable when model architectures change frequently or tasks diversify.

3.3.2. Hybrid compression methods

Hybrid compression has become a common choice in practical deployment. For example, DGC is first applied for sparsification (e.g., Top 1%) to preserve critical gradient information; then 8-bit quantization (QSGD) is applied to the nonzero gradients; and finally, Huffman coding is used for lossless compression. This multi-stage strategy maximizes communication efficiency while maintaining convergence performance. When training ResNet50 on ImageNet, the combination of DGC and 8-bit quantization achieved a compression ratio of approximately 600× with less than than 1% loss in accuracy [12].

While semantic compression offers significant advantages in extremely low-bandwidth scenarios, its performance depends heavily on the stability and extractability of task-specific semantics. For multi-task or dynamically changing scenarios, the generalizability of semantic encoding poses substantial challenges. In contrast, hybrid compression is more robust but requires meticulous tuning

DOI: 10.54254/2755-2721/2026.34791

of compression parameters; moreover, the cumulative error from multi-stage compression may adversely affect convergence.

4. Future directions

Future research can be pursued along the following three directions, which correspond explicitly to the aforementioned challenges and aim to construct a FL framework better adapted to the SAGIN environment.

4.1. Adaptive compression and hierarchical asynchronous aggregation

Current compression and aggregation strategies are largely static and do not fully exploit the predictability of satellite orbits. Future work can develop orbit-aware adaptive compression that dynamically adjusts the compression ratio based on the remaining visibility window length and channel quality. When the visibility window is long, high-precision transmission strategies can be used. When the window is short, a high-compression mode can be activated. Meanwhile, more efficient hierarchical asynchronous aggregation algorithms can optimize the hierarchical strategy and the aggregation frequency based on network conditions.

4.2. Construction of personalized Federated Learning

Balancing global knowledge sharing with local personalization is a core issue in addressing data heterogeneity. Future research may explore more sophisticated model decoupling methods by distinguishing between universal base models and task-specific models, and combine meta-learning or transfer learning to enable satellite networks to rapidly adapt to new scenarios or tasks. System heterogeneity, which includes disparities in computing power and energy, calls for resource-aware local iteration strategies. Satellites with higher capacity perform more local updates, while those with lower capacity reduce the number of updates or employ model pruning.

4.3. Lightweight security and privacy protection mechanisms

The severe resource constraints of satellite nodes make it difficult to apply conventional encryption methods, such as homomorphic encryption and secure multi-party computation, directly. Secure aggregation and privacy protection are the newly pathways. Lightweight secure aggregation and trust evaluation mechanisms can be designed. For example, integrating blockchain technology can be introduced to a distributed reputation system that conducts trusted evaluations of satellite update quality. By incorporating differential privacy and similar techniques, malicious attacks and data theft can be effectively resisted without substantially increasing computational burdens.

5. Conclusion

This paper systematically reviews communication-efficient federated learning technologies for Space-Air-Ground Integrated Networks. It analyzes the SAGIN architecture and identifies the fundamental challenges posed by dynamic topology, communication bottlenecks, and heterogeneity in distributed learning, quantifying the severity of communication constraints through a specific numerical example. The paper further surveys existing research on satellite federated learning, summarizes representative architectures and characteristics from the perspectives of hierarchical asynchronous aggregation, semantic compression, and personalized domain adaptation, and assesses

DOI: 10.54254/2755-2721/2026.34791

the applicability and limitations of each category. Finally, it proposed future research directions, orbit-aware adaptation, personalized learning, and lightweight security, and clarified their correspondence to the core challenges.

However, this paper is primarily based on literature and secondary data, with limited empirical experimental validation on real satellite platforms or in high-fidelity simulation environments. Future work should investigate software-defined radio (SDR) and co-simulations with open-source satellite simulation platforms such as STK and NS3 to construct a SAGIN-FL testbed that more closely approximates practical conditions, enabling quantitative evaluation and comparison of the technologies reviewed herein.

References

[1] Recommendation, I. T. U. R. (2023). Framework and overall objectives of the future development of IMT for 2030 and beyond. International Telecommunication Union (ITU) Recommendation (ITU-R).
[2] Jin, Y., Fan, J., Xi, E., Dong, H., & Yu, H. (2025). Survey on federated learning-based communication technologies for space-air-ground integrated networks. Application Research of Computers, 42(11), 3215–3224.
[3] Li, L., Zhu, L., & Li, W. (2025). Privacy-preserving federated learning for space–air-ground integrated networks: a bi-level reinforcement learning and adaptive transfer learning optimization framework. Sensors, 25(9), 2828.
[4] Oh, J., Lee, D., Truong, T. P., Hur, D., Hong, S., & Cho, S. (2025, February). A survey on satellite networks with federated learning to analyze data or manage resource. In 2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 0022-0027). IEEE.
[5] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282). Pmlr.
[6] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems, 2, 429-450.
[7] Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020, November). Scaffold: Stochastic controlled averaging for federated learning. In International conference on machine learning (pp. 5132-5143). PMLR.
[8] Li, L., Zhu, L., & Li, W. (2025). HISATFL: A hierarchical federated learning framework for satellite networks with cross-domain privacy adaptation. Electronics, 14(16), 3237.
[9] Alistarh, D., Grubic, D., Li, J., Tomioka, R., & Vojnovic, M. (2017). QSGD: Communication-efficient SGD via gradient quantization and encoding. Advances in neural information processing systems, 30.
[10] Bernstein, J., Wang, Y. X., Azizzadenesheli, K., & Anandkumar, A. (2018, July). signSGD: Compressed optimisation for non-convex problems. In International conference on machine learning (pp. 560-569). PMLR.
[11] Wen, W., Xu, C., Yan, F., Wu, C., Wang, Y., Chen, Y., & Li, H. (2017). Terngrad: Ternary gradients to reduce communication in distributed deep learning. Advances in neural information processing systems, 30.
[12] Lin, Y., Han, S., Mao, H., Wang, Y., & Dally, W. J. (2017). Deep gradient compression: Reducing the communication bandwidth for distributed training. arXiv preprint arXiv: 1712.01887.
[13] Xu, X., Li, Y., & Lu, L. (2025). RAFL: Reputation-aware federated learning with hierarchical aggregation in LEO satellite networks. Journal of Systems Architecture, 103565.
[14] Xu, A., Zhang, Z., Liu, Z., & Cui, S. (2026). EMS-FL: Federated tuning of mixture-of-experts in satellite-terrestrial networks via expert-driven model splitting. arXiv preprint arXiv: 2602.19485.
[15] Nguyen, L. X., Hassan, S. S., Park, Y. M., Tun, Y. K., Han, Z., & Hong, C. S. (2025). SemSpaceFL: A collaborative hierarchical federated learning framework for semantic communication in 6G LEO satellites. IEEE Transactions on Communications, 74, 1269-1286