Federated Learning Framework for Privacy-Preserving Global Weather Forecasting

M. Narender[1,2], Mosleh Mahmoud Khyyat[2]

[1] Post-Doctoral Researcher, Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia, pdf.narender@lincoln.edu.my

Associate Professor, Department of Computer Science and Engineering, ACE Engineering College, Hyderabad, India, machha.narender@gmail.com

[2] Information Systems and Technology College of Computer Science and Engineering, University of Jeddah, mkhayyat@uj.edu.sa

Abstract: Countries around the world rely on heterogeneous observations and reanalysis data disseminated by national and regional meteorological agencies at the time of weather forecasting. Cross-border sharing of data may, however, be restricted by sovereignty matters, security, commercial and regulatory matters. The authors of this paper propose a federated learning framework Fed-WeatherNet to allow participating countries to predict weather forecasts on global scale while retaining full privacy control over their atmospheric data. Instead of sharing atmospheric data, the participating countries just share privacy-preserving updates. The proposed framework used an STGN or Spherical Transformer backbone combined with Climate-Cluster Weighted Aggregation (CCWA) for climate-dependent non-IID data. Information leakage can be limited through the incorporation of client-side gradient clipping, differential privacy using DP-SGD and secure aggregation. Further protection is optionally provided by homomorphic encryption which offers a privacy mechanism comparable with secure aggregation. To lower the communication cost to bandwidth constraint participants, we introduce top-k sparsification, quantization, and asynchronous participation. A standardized evaluation protocol defines centralized as well as localized-only and federated baselines, with latitude-weighted RMSE, ACC, and CRPS as main forecast metrics. The specifics of experimentation are also covered in these Non-IID robustness, privacy-utility trade-offs, communication efficiency, participation fairness, adversarial robustness, and simulated cross-country ERA5 deployment.

Keywords: Federated learning; Global weather forecasting; Privacy preservation; Differential privacy; DP-SGD; Secure aggregation; Non-IID climate data; Climate-aware aggregation; ERA5; Spatiotemporal forecasting; Communication efficiency; WeatherBench.

1. Introduction

For global climate predictions, requires collaboration across observational-data-silos both national and regional space weather. In fact, many meteorological agencies suffer data-sovereignty, commercial, security, and cross border-sharing constraints. This leads to a lack of data and information sharing across borders, negatively impacting accuracy.

This paper presents this challenge as a federated learning problem, where countries keep their local atmospheric data and collaboratively train a model to predict future weather events using privacy-sanitized model updates. Federated learning enables collaborative model training while keeping source data decentralized [28], [27], [26].

The proposed framework, Fed-WeatherNet, is motivated by six knowledge gaps: the data-sovereignty versus forecast-accuracy trade-off at global scale; non-IID climate heterogeneity in federated aggregation; privacy-utility trade-offs for high-dimensional spatiotemporal atmospheric data; communication efficiency for high-resolution gridded data; the lack of standardized cross-country benchmarking protocols; and trust/incentive mechanisms for unequal data contributors.

1.1 Related Work

This document identifies existing numerical weather prediction and deep-learning systems, including GraphCast, FourCastNet, and Pangu-Weather, as examples of forecasting approaches trained on centrally pooled, homogenized reanalysis data such as ERA5. It states that a corresponding cross-country federated setting with privacy constraints has not been adequately quantified. Recent global AI weather systems include GraphCast, Pangu-Weather, and FuXi, while ERA5 is a major reanalysis resource for weather-model development [25]–[15].

1.2 Knowledge Gaps

Gap 1: Data sovereignty vs. forecast accuracy trade-off is underexplored at global scale.

Existing weather forecasting models are described as being trained on centrally pooled, homogenized reanalysis data. The source identifies the absence of a framework that quantifies forecast skill preserved or lost when collaboration shifts to privacy-constrained federated training across countries.

Gap 2: Non-IID climate heterogeneity across regions is not adequately modeled in FL aggregation.

The source notes that weather data differ across climate zones, seasonal cycles, sensor density, and atmospheric-variable ranges, while standard FedAvg assumes comparatively similar client distributions.

Gap 3: Privacy-utility trade-off for spatiotemporal atmospheric data is unresolved.

The source identifies uncertainty regarding how differential privacy and secure aggregation affect high-dimensional atmospheric grids and when privacy noise may compromise physical plausibility or forecast skill.

Gap 4: Communication efficiency for high-resolution gridded data in cross-national FL is unaddressed.

Weather-model updates can be substantially larger than those in typical vision or NLP federated-learning benchmarks. Bandwidth-constrained and asymmetric-connectivity participants are identified as underrepresented.

Gap 5: Lack of protocols for standard cross-country benchmarking.

According to the source, no standard protocols are currently used to compare federated with centralized and localized-only weather models across heterogeneous national data silos.

Gap 6: There are insufficient trust mechanisms for unequal data contributors.

The material discusses the inequality between the data-rich and data-poor countries and it also mentions allocation of incentives/rewards in a fair way of free-rider.

1.3 Key Contribution

Fed-WeatherNet is a tool to enhance the safety of global weather forecasting. The specified elements include climate-zone-aware weighted federated aggregation, secure aggregation, differential privacy

through gradient clipping and Gaussian noise, optional homomorphic encryption for comparison, communication-efficient quantization and top-k sparsification, adaptive/asynchronous participation, and a standardized evaluation framework for centralized, localized-only, and federated baselines. Common strategies for preserving privacy in Federated Learning are local training and the use of differential privacy among others.

2. Method, Experiments and Results

2.1 Problem Formulation

Consider there to be N participating countries (clients) C_1, C_2, …, CN, each of which holds a local dataset D_i of past and/or present atmospheric observations of the local domain, including temperature, pressure, humidity, wind vectors, and precipitation. The aim is to jointly learn the global forecasting model θ_G without exporting D_i outside of national borders.

2.2 Federated Architecture

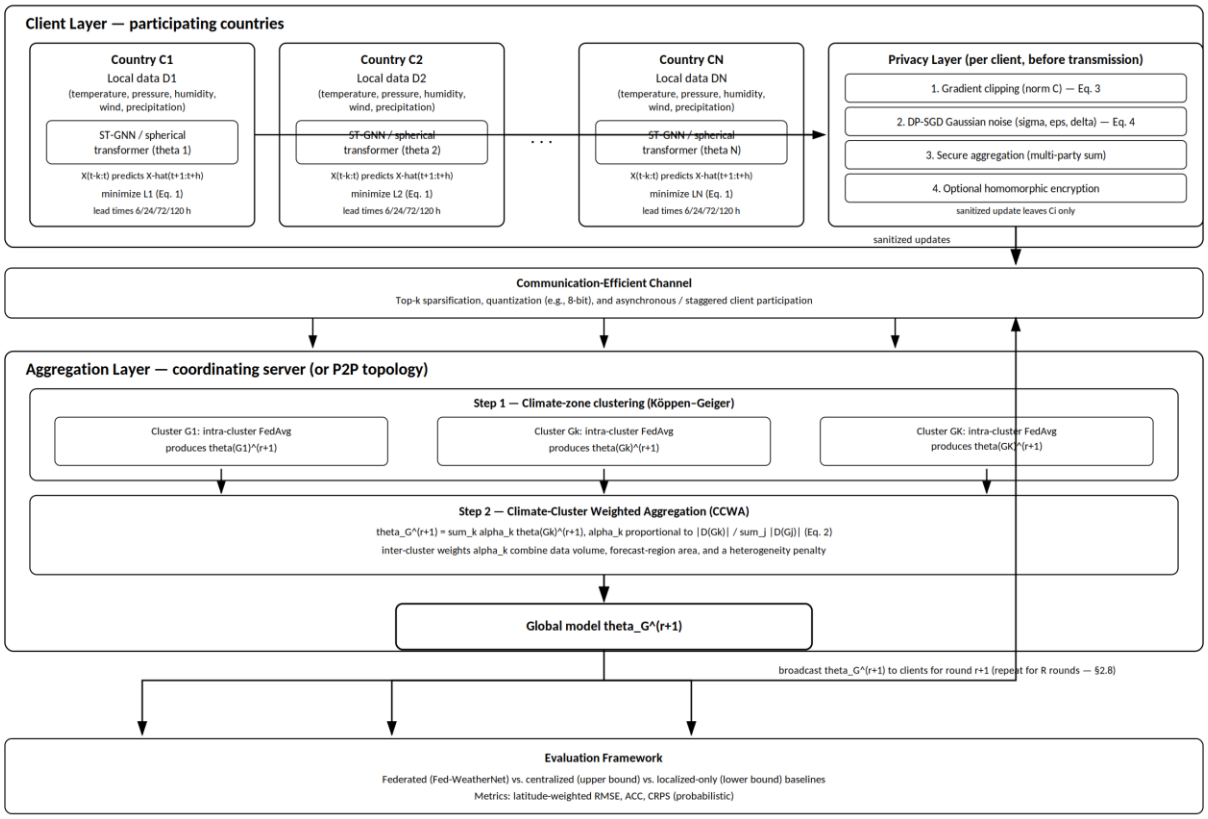


Fig. 1. Fed-WeatherNet federated architecture.

The framework consists of three different functional layers.

- The client layer: It indicates that each country is responsible for training a local spatiotemporal forecasting model using its own historical and near-real-time atmospheric data.
- Aggregation layer: Either a coordinating server, or if appropriate, a decentralized peer-to-peer topology (or network), receives sanitized local updates and produces a new global model.

- Privacy layer: gradient clipping, differential-privacy noise injection, and secure aggregation are applied before parameters leave client infrastructure.

For the local model, the source specifies a graph neural network (ST-GNN) or spherical transformer backbone operating on a latitude-longitude mesh. The model ingests a sequence of past atmospheric states $X_{(t-k:t)}$ and predicts future states $\hat{X}_{(t+1:t+h)}$ for lead times of 6 h, 24 h, 72 h, and 120 h.

Each country trains a local ST-GNN / spherical-transformer model on its own data; sanitized updates (after clipping, DP-SGD noise, and secure aggregation) pass through a communication-efficient channel to the server, which performs Climate-Cluster Weighted Aggregation (CCWA) in two steps (intra-cluster FedAvg, then inter-cluster weighting) to produce the global model, which is broadcast back for the next round.

2.3 Local Model Design

Each client-side model uses an ST-GNN or spherical transformer to represent spatial dependencies over the Earth's grid, including irregular station density and spherical geometry. Local training minimizes a latitude-weighted mean squared error, where $w(\varphi)$ corrects for grid-cell area distortion at latitude φ.

$$L_i = \frac{1}{|D_i|}\sum_{(x,y)\in D_i} w(\varphi)\left\|f_{\theta_i}(x) - y\right\|^2 \quad (1)$$

Where L_i is the local training loss for client (country) i; D_i is the local dataset held by client i, with $|D_i|$ its size; (x, y) is an input–target pair drawn from D_i; $w(\varphi)$ is the latitude-dependent weight that corrects for grid-cell area distortion at latitude φ; $f_{\theta_i}(x)$ is the prediction of the local model with parameters θ_i on input x; $\|\cdot\|$ denotes the Euclidean (L2) norm.

2.4 Federated Aggregation Strategy

Standard FedAvg is extended to Climate-Cluster Weighted Aggregation (CCWA) to address non-IID heterogeneity. Clients are grouped according to the Köppen–Geiger climate classification of their dominant geographic coverage. FedAvg is first performed within climate-similar groups, followed by inter-cluster aggregation using weights based on data volume and forecast-region area, with a heterogeneity penalty intended to reduce domination by large clusters.

$$\theta_G^{(r+1)} = \sum_{k=1}^{K} \alpha_k \theta_{G_k}^{(r+1)}, \quad \alpha_k \propto \frac{\left|D_{G_k}\right|}{\sum_j \left|D_{G_j}\right|} \quad (2)$$

Where θ_G is the global model parameter vector; superscript (r+1) denotes the value after communication round r+1; K is the number of Köppen–Geiger climate clusters G_k represents the k^{th} cluster, while $\theta_{G_k}^{r+1}$ denotes its intra-cluster FedAvg result. Furthermore, α_k refers to the weight assigned to the inter-cluster aggregation of the cluster G_k. This weight is deemed proportional to the respective share of the total data volume $\left|D_{G_k}\right|$ of that cluster. Notably, a heterogeneity penalty prevents domination by large clusters. Lastly, $\left|D_{G_k}\right|$ and $\left|D_{G_j}\right|$ represent the pooled local datasets of the clusters G_k and G_j, respectively.

2.5 Privacy-Preservation Mechanisms

- Gradient clipping restricts each local update to a maximum norm C prior to transmission.

- DP-SGD has incorporated calibrated Gaussian Noise $N(0, \sigma^2C^2)$ to the clipped gradients which contains a privacy budget ε to ensure forecasting quality and to limit the leak of information. DP design follows early proposals that focused on privacy-preserving learning.
- Secure aggregation employs a cryptographic multi-party summation protocol that ensures the server only observes the overall aggregated update rather than the contribution made by any individual country.
- Homomorphic encryption is designated as an optional comparison arm for specifications that require stronger cryptographic assertions, at increased computational cost.

2.6 Differential Privacy with Stochastic Gradient Descent

DP-SGD alters stochastic gradient descent to bound the influence of a single training example with calibrated noise. In any example, the source specifies the clipped gradient:

$$\bar{g}_i = \frac{g_i}{\max\left(1, \frac{\|g_i\|_2}{C}\right)} \tag{3}$$

Where g_i is the raw per-example gradient computed by client i; $\bar{g}_i$ is the clipped gradient (bar denotes clipping); $\|g_i\|_2$ is the L2 norm of the gradient; C is the clipping norm that bounds each per-example gradient's influence. Before the model is updated, Gaussian noise is added to the sum of clipped gradients for a batch of size L.

$$\tilde{g} = \frac{1}{L}\left[\sum_{i=1}^{L} \bar{g}_i + N(0, \sigma^2 C^2 I)\right] \tag{4}$$

In the formula, L refers to batch size, which is defined as "number of clipped per-example gradients averaged in one update." The variable $\hat{g}$ refers to the noisy, privacy-sanitized gradient used in the model update. Further, this gradient is defined as being in the form of $N(0, \sigma^2C^2I)$, which is isotropic Gaussian noise with zero mean and covariance σ^2C^2I. The symbol σ refers to the noise multiplier which controls the magnitude of the noise relative to the clipping norm C. Lastly, the two symbols ε and δ (which are referenced throughout the surrounding text) are known as the differential-privacy accounting parameters. In other words, ε refers to the privacy budget, while δ refers to the failure probability that the guarantee does not hold as stated. The privacy accounting parameters are denoted by ε and δ. It also highlight the privacy-utility trade-off: a smaller ε means stronger privacy protection, but greater noise in the output; a larger ε means weaker privacy protection, but more utility in the output. In the federated setting, clipping and noise are applied locally before the update leaves the client's infrastructure.

Table 1. Key Differential Privacy Parameters

Parameter	Role
Clipping norm C	Bounds per-example gradient influence.
Noise multiplier σ	Controls noise magnitude relative to C.
Privacy budget ε	Overall privacy guarantee; lower values provide stronger privacy.
Failure probability δ	Probability that the stated guarantee does not hold.
Number of rounds/epochs	Additional training rounds consume privacy budget.

2.7 Communication-Efficient Update Scheme

- Top-k sparsification transmits only the largest-magnitude gradient components.
- Quantization reduces update precision, with 8-bit updates given as an example.
- Asynchronous or staggered participation allows low-connectivity countries to contribute partial or delayed updates without blocking a global round.

2.8 Training Protocol

1. Initialize the global model θ_G^0 at the server and broadcast it to participating clients.
2. Each client trains locally for E epochs on D_i and produces its local model θ_i^r.
3. Clients apply clipping and differential-privacy noise, then submit updates through secure aggregation.
4. The server performs CCWA to produce $\theta_G^{(r+1)}$.
5. Repeat the local-training and aggregation cycle for R communication rounds or until validation convergence.

2.9 Evaluation Framework

The source specifies a standardized comparison among (1) fully centralized training, described as an upper-bound baseline; (2) fully localized or independent per-country models, described as a lower-bound baseline; and (3) the proposed federated privacy-preserving framework. A structured evaluation protocol is important for comparing utility, efficiency, and security dimensions in FL [8].

Forecast skill is evaluated using latitude-weighted RMSE, Anomaly Correlation Coefficient (ACC), and Continuous Ranked Probability Score (CRPS) for probabilistic variants, across held-out spatial regions and time periods. WeatherBench established a reproducible benchmark and common metrics for data-driven weather forecasting [24].

Experiments

Experiment 1: Baseline Comparison

Compare forecast skill (RMSE, ACC, CRPS) of the federated model against centralized and localized-only baselines across held-out test regions and time windows at 6 h, 24 h, 72 h, and 5-day lead times.

Experiment 2: Non-IID Robustness Study

Simulate varying degrees of climate heterogeneity using Köppen-zone grouping versus random grouping and measure convergence speed and final accuracy under FedAvg and climate-aware weighted aggregation.

Experiment 3: Privacy-Utility Trade-off Analysis

Sweep differential-privacy budgets ε = 0.1, 1, 5, 10, and ∞ and measure forecast-skill degradation and physical consistency, including mass/energy conservation checks.

Experiment 4: Communication Efficiency Study

Measure communication cost in MB per round and wall-clock convergence time under full-precision versus quantized/sparsified updates for low- and high-bandwidth client conditions.

Experiment 5: Fairness and Participation Study

Evaluate forecast accuracy for data-sparse participating countries under standard FedAvg and the proposed incentive-aware aggregation, quantifying accuracy gains for low-contributing clients.

Experiment 6: Adversarial/Robustness Test

Test resilience of secure aggregation and DP mechanisms against simulated malicious or colluding clients attempting model-inversion or membership-inference attacks on shared gradients.

Experiment 7: Real-World Case Study

Apply the framework to publicly available national meteorological datasets, such as simulated multi-country partitioning of ERA5 by country boundary, as a proxy for cross-border federated deployment.

Results

Not implanted methods, so results are not available.

Discussions

The source framework is designed to examine the interaction among forecast skill, climate heterogeneity, privacy protection, communication constraints, and participation fairness. In particular, the proposed experiments are intended to characterize the privacy-utility frontier, the robustness of climate-aware aggregation under non-IID conditions, and the communication and participation implications of federated weather forecasting.

Conclusions

1. Problem statement addressed / motivation: This paper identifies the need for cross-country weather-model collaboration while respecting national data-sovereignty constraints.
2. Method used: Fed-WeatherNet combines climate-cluster weighted federated aggregation, local DP-SGD, secure aggregation, communication-efficient updates, and a multi-baseline evaluation protocol.
3. Key findings: No empirical findings are reported in the source document; the listed experiments define the intended evaluation.
4. Limitations and future work: This paper identifies the absence of standardized cross-country benchmarks, unresolved privacy-utility behavior for atmospheric grids, communication constraints, unequal data contributions, and the difficulty of conducting genuine multi-agency pilots. These are proposed areas for further study.

References

1. J. Feng and W. Ji, “Global-regional AI weather forecasting on a stretched cubed sphere,” Geophys. Res. Lett., vol. 53, Art. no. e2026GL124199, 2026, doi: 10.1029/2026GL124199.
2. Y. H. He and P. W. Chan, “A study of the extended-range weather forecasts by artificial intelligence models Pangu-Weather and FengWu,” Weather, vol. 80, no. 4, pp. 134–138, 2025, doi: 10.1002/wea.7648.
3. L. Ma et al., “The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time,” Commun. Earth Environ., vol. 6, Art. no. 518, 2025, doi: 10.1038/s43247-025-02502-y.

4. Y. Li et al., “FuXi-En4DVar: An assimilation system based on machine learning weather forecasting model ensuring physical constraints,” Geophys. Res. Lett., vol. 51, Art. no. e2024GL111136, 2024, doi: 10.1029/2024GL111136.
5. X. Zhong et al., “FuXi-Extreme: Improving extreme rainfall and wind forecasts with diffusion model,” Sci. China Earth Sci., vol. 67, pp. 3696–3708, 2024, doi: 10.1007/s11430-023-1427-x.
6. M. Bonavita, “On some limitations of current machine learning weather prediction models,” Geophys. Res. Lett., vol. 51, no. 12, Art. no. e2023GL107377, 2024, doi: 10.1029/2023GL107377.
7. D. Kochkov et al., “Neural general circulation models for weather and climate,” Nature, vol. 632, pp. 1060–1066, 2024, doi: 10.1038/s41586-024-07744-y.
8. D. Chai, L. Wang, L. Yang, J. Zhang, K. Chen, and Q. Yang, “A survey for federated learning evaluations: Goals and measures,” IEEE Trans. Knowl. Data Eng., vol. 36, no. 10, pp. 5007–5024, 2024, doi: 10.1109/TKDE.2024.3382002.
9. A. Agiollo et al., “Anonymous federated learning via named-data networking,” Future Generation Computer Systems, vol. 152, pp. 288–303, 2024, doi: 10.1016/j.future.2023.11.009.
10. X. Zhou and G. Yang, “Communication-efficient and privacy-preserving large-scale federated learning counteracting heterogeneity,” Information Sciences, vol. 661, Art. no. 120167, 2024, doi: 10.1016/j.ins.2024.120167.
11. S. Saha et al., “A multifaceted survey on privacy preservation of federated learning: Progress, challenges, and opportunities,” Artif. Intell. Rev., vol. 57, Art. no. 184, 2024, doi: 10.1007/s10462-024-10766-7.
12. Z. Lu et al., “Federated learning with non-IID data: A survey,” IEEE Internet Things J., vol. 11, no. 11, pp. 19188–19209, 2024, doi: 10.1109/JIOT.2024.3376548.
13. W. Huang et al., “Federated learning for generalization, robustness, fairness: A survey and benchmark,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 46, no. 12, pp. 9387–9406, 2024, doi: 10.1109/TPAMI.2024.3418862.
14. D. Chiaro et al., “FL-Enhance: A federated learning framework for balancing non-IID data with augmented and shared compressed samples,” Information Fusion, vol. 98, Art. no. 101836, 2023, doi: 10.1016/j.inffus.2023.101836.
15. L. Chen et al., “FuXi: A cascade machine learning forecasting system for 15-day global weather forecast,” npj Climate and Atmospheric Science, vol. 6, Art. no. 190, 2023, doi: 10.1038/s41612-023-00512-1.
16. K. Bi et al., “Accurate medium-range global weather forecasting with 3D neural networks,” Nature, vol. 619, pp. 533–538, 2023, doi: 10.1038/s41586-023-06185-3.
17. R. Lam et al., “Learning skillful medium-range global weather forecasting,” Science, vol. 382, no. 6677, pp. 1416–1421, 2023, doi: 10.1126/science.adi2336.
18. D. Li et al., “Ubiquitous intelligent federated learning privacy-preserving scheme under edge computing,” Future Generation Computer Systems, vol. 144, pp. 205–218, 2023, doi: 10.1016/j.future.2023.03.010.
19. X. Yan et al., “Privacy-preserving asynchronous federated learning framework in distributed IoT,” IEEE Internet Things J., vol. 10, no. 15, pp. 13281–13291, 2023, doi: 10.1109/JIOT.2023.3262546.

20. T. Qi et al., “Differentially private knowledge transfer for federated learning,” Nature Commun., vol. 14, Art. no. 3785, 2023, doi: 10.1038/s41467-023-38794-x.
21. Q. Li et al., “A survey on federated learning systems: Vision, hype and reality for data privacy and protection,” IEEE Trans. Knowl. Data Eng., vol. 35, no. 4, pp. 3347–3366, 2023, doi: 10.1109/TKDE.2021.3124599.
22. A. El Ouadrhiri and A. Abdelhadi, “Differential privacy for deep and federated learning: A survey,” IEEE Access, vol. 10, pp. 22359–22380, 2022, doi: 10.1109/ACCESS.2022.3151670.
23. S. Shen et al., “From distributed machine learning to federated learning: In the view of data privacy and security,” Concurrency Computat.: Pract. Exper., 2022, doi: 10.1002/cpe.6002.
24. S. Rasp et al., “WeatherBench: A benchmark dataset for data-driven weather forecasting,” J. Adv. Model. Earth Syst., vol. 12, no. 11, Art. no. e2020MS002203, 2020, doi: 10.1029/2020MS002203.
25. H. Hersbach et al., “The ERA5 global reanalysis,” Quart. J. Roy. Meteorol. Soc., vol. 146, no. 730, pp. 1999–2049, 2020, doi: 10.1002/qj.3803.
26. M. Aledhari et al., “Federated learning: A survey on enabling technologies, protocols, and applications,” IEEE Access, vol. 8, pp. 140699–140725, 2020, doi: 10.1109/ACCESS.2020.3013541.
27. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” IEEE Signal Process. Mag., vol. 37, no. 3, pp. 50–60, 2020, doi: 10.1109/MSP.2020.2975749.
28. Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated machine learning: Concept and applications,” ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, Art. no. 12, 2019, doi: 10.1145/3298981.