Robust Privacy-Preserving Federated Learning for 6G Network Resource Allocation

Nizamuddin Maitlo[1*], Samina Rajper[2], Manahil Shaikh[3], Hina Kareem[4]

Abstract: **Sixth-generation (6G) radio access networks must allocate resources across heterogeneous service slices—enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communication (URLLC), massive Machine-Type Communication (mMTC), and Extended Reality (XR)—while meeting strict service-level targets and preserving privacy. We present a robust, privacy-compatible federated learning (FL) pipeline that learns slice-aware admission and serving decisions from gNB-side telemetry without centralizing raw logs. Using eight non-IID clients with 150 telemetry windows per client sampled every 90 s, we train a compact client-side classifier for 20 global rounds under partial participation (approximately 85%). Labels are defined by next-window SLA satisfaction (downlink throughput ≥ 40 Mbps and p95 latency ≤ 50 ms). The global model improves from 0.63 to 0.91 accuracy and from 0.62 to 0.85 macro-F1, while the training loss drops from 0.92 to 0.18. Fairness, measured as the gap between the best and worst mean per-client accuracy, is 0.08, with most per-round gaps falling in the 0.05–0.14 range. A heterogeneity snapshot confirms strong non-IID conditions across PRB utilization, CQI, p95 latency, UE density, and slice-mix entropy. The pipeline outputs reproducible CSV logs and figures and is designed to remain compatible with client-level differential privacy (norm clipping with Gaussian noise), secure aggregation, and over-the-air (AirComp) update fusion to reduce uplink time. These results indicate that FL can achieve accurate and fairness-aware 6G resource allocation under realistic participation and stringent SLAs while supporting privacy and communication optimizations.**

Keywords: **6G, federated learning, radio access networks, network slicing, resource allocation, non-IID data, fairness, differential privacy, secure aggregation, over-the-air computation.**

I. INTRODUCTION

To meet strict service requirements, such as Ultra-Reliable Low-Latency Communications (URLLC), massive Machine-Type Communications (mMTC), and Extended Reality (XR), 6G radio access networks (RANs) must schedule spectrum, time-frequency resources, and slice budgets across diverse traffic classes under time-varying channel conditions. At the same time, operators must protect user and operational data while keeping control-plane overhead low. Although centralized learning can support resource-management decisions, it requires collecting raw telemetry, such as per-UE KPIs and slice mixtures, which increases the attack surface and consumes backhaul resources.

[1,2,3,4] Institute of Computer Science, Shah Abdul Latif University, Khairpur
Country: Pakistan
Email: *nizamuddin.cs@gmail.com

Federated learning (FL) keeps data at devices or edge locations and communicates model updates rather than raw data [1]. This architectural shift fits RAN constraints, but practical deployment faces three major realities:

1) *Heterogeneity:* Client data are often non-IID and size-imbalanced across cells and slices, with temporal drift and distinct radio conditions. These factors can slow convergence and create fairness gaps [2], [3].
2) *Privacy:* Many RAN settings require client-level privacy guarantees. Differential Privacy (DP), based on norm clipping and Gaussian noise, can be combined with secure aggregation to prevent inspection of individual client updates [4]–[6].
3) *Efficiency:* When channel inversion and power control are available, uplink latency can be reduced through analog-domain summation. This enables efficient aggregation through payload compression and Over-the-Air Computation (AirComp) [7], [8].

The problem setting of this work is learning slice-aware resource decisions from windowed gNB-side telemetry. For each client, represented by a sector, gNB, or cluster, the logged features include PRB utilization, CQI, SINR, downlink/uplink BLER, HARQ retransmissions, UE counts, handover counts, throughput and latency summaries, and slice-mix fractions. The next window is labeled with a Boolean variable indicating whether SLA thresholds are satisfied, while a bounded reward summarizes the allocation quality of the window. The learning task is to train a privacy-preserving and round-efficient binary classifier that remains applicable across clients with different IID and non-IID distributions.

The main challenges for FL in 6G RANs are as follows:

- *Statistical skew:* Client slice composition and load regimes vary across the network. Label imbalance and size skew make FedAvg challenging and may favor dominant clients [2], [3].
- *System skew:* Partial participation and backhaul variation affect client sampling. Therefore, stale or missing updates should not disrupt training [9].
- *Privacy at scale:* Client-level DP necessitates careful clipping and scheduling of updates as well as secure aggregation, which are essential to preserve privacy on a large scale; the details of these algorithms and methods are left out of scope here [4]–[6].
- *Round efficiency:* The wider the federations and the larger the models, the higher the uplink cost. This can be lowered by compression and AirComp, which incur engineering trade-offs, as demonstrated in [7], [8], [10].

In this work, we define a repeatable FL pipeline that leverages windowed gNB KPIs and slice-level features and test the pipeline

under strict SLA constraints and under realistic partial participation. In contrast to traditional demos, we report on convergence behavior, fairness according to the best-worst per-client accuracy gap, and heterogeneity statistics based on PRB utilization, CQI, latency, number of UEs and slice entropy. These results are generated from reproducible CSV logs and figures. We use tighter SLA targets, including downlink throughput of at least 40 Mbps and p95 latency of at most 50 ms. We also inject partial participation of approximately 85% and client impairments that increase PRB pressure and reduce CQI for selected clients. The overall framework is shown in Figure 3, while the data and file flows are presented in Figure 1, and the classifier architecture is shown in Figure 2.

Study Contributions

- We develop a privacy-friendly 6G FL pipeline that learns slice-aware decisions from gNB-side telemetry while avoiding the centralization of raw data. The pipeline also includes hooks for client-level DP and secure aggregation.
- We provide a comprehensive assessment under non-IID data and partial participation. The evaluation reports global accuracy, F1-score, error rate, loss decay, per-client means, per-round fairness gaps, and telemetry-based heterogeneity snapshots.
- We release reproducibility artifacts, including schema-checked CSV files such as fl_rounds.csv and client_*_eval.csv, telemetry files, and figures for global accuracy, F1-score, per-client scatter plots, and fairness gaps. These artifacts support independent verification and ablation studies.

II. BACKGROUND AND RELATED WORK

Federated learning (FL) decouples raw-data movement from statistical learning by periodically aggregating models trained locally at distributed clients. In the standard FedAvg protocol [1], the global model is updated in each communication round using a sample-size-weighted average of the locally updated client models. Each selected client performs local training steps on its own data, and the resulting local model is then aggregated at the server. Production deployments extend this cycle with availability-aware client sampling, straggler tolerance, robust telemetry, and isolation mechanisms that allow large and unreliable fleets to train reliably and economically [9]. In radio access networks (RANs), FL is particularly attractive because per-UE and per-slice telemetry, such as PRB utilization, CQI, and latency statistics, can remain at the edge. This minimizes the potential of exposing data and the backhaul footprint of centralized learning [11].

Client heterogeneity is one of the major practical challenges in FL, including non-IID data, sample imbalance, and temporal drift. User geography, radio conditions, user mobility patterns and slice composition can create user distributions that are skewed to one side in a cellular environment. The above factors can lead to convergence delay and performance difference between clients [2], [3]. This problem is tackled by several algorithmic families. To ensure stability for local optimization, FedProx introduces a proximal term [12]. Control-variate methods account for any client drift and enhance the stability of the system with local updates that are biased [13]. However, methods of personalization can learn shared representations without losing any client-specific components [14], [15]. Furthermore, fairness-aware client selection and optimization methods seek to minimize the degradation to the worst client and minimize performance gaps between the clients [16].

Another crucial requirement in operational FL is privacy. Typically, client-level Differential Privacy (DP) truncates the client updates to a norm bound and introduces calibrated Gaussian noise to the updates based on their sensitivity. This gives privacy guarantees that can be quantified, and typically with a tolerable loss in utility [4], [5]. Secure aggregation is used in addition to DP so that the server will only see the aggregate client update instead of any individual updates [6]. RAN telemetry is particularly critical because raw features can be used to uncover sensitive user behavior or operational information. The integration of DP and secure aggregation into the FedAvg message flow, without altering the interface between the learning task, has been proposed in the literature.

For wireless and edge FL, the key to communication efficiency is often the round time. Common approaches reduce communication overhead through update compression, quantization, and payload reduction, although these methods may introduce additional noise or bias [10]. Over-the-Air Computation (AirComp) can further reduce uplink latency by exploiting the additivity of the wireless multiple-access channel [7], [8].

Beyond convergence, calibrated probabilities are important for threshold-based operational decisions. Post-hoc calibration methods, such as temperature scaling, can improve the reliability of predicted probabilities [17]. Since marginal gains often diminish after later communication rounds, early stopping can reduce unnecessary training cost [18]. Overall, existing surveys emphasize the joint design of learning, scheduling, and communication to satisfy latency and energy constraints in mobile networks [11], [19].

III. METHODOLOGY

This section describes the end-to-end pipeline: telemetry windowing and labeling, the client-side classifier and local objective, the federated protocol under partial participation, and the (optional) privacy and communication enhancements.

A. Data Schema, Windowing, and Labeling

Each client $k \in \{1, \ldots, K\}$ (e.g., a gNB sector/site) logs gNB-side telemetry and derives fixed-length windows of duration $\Delta = 90$ s, producing n_k windows (in our experiments, $n_k = 150$ per client and $K = 8$). Each window aggregates resource, channel, link, load/mobility, and service measurements, including (examples) PRB utilization, CQI/SINR, BLER and HARQ summaries, UE counts and handovers, throughput and latency percentiles (p50/p95), and slice-mix fractions $p = (p_{\text{eMBB}}, p_{\text{URLLC}}, p_{\text{mMTC}}, p_{\text{XR}}, p_{\text{BE}})$ with $\Sigma_s\, p_s = 1$.

We define a binary label indicating whether the next window satisfies strict service-level thresholds:

$$y_i^{(k)} = \mathbb{I}\left[DLthr_{mean} \geq T_{thr} \wedge lat_{p95} \leq T_{lat}\right] \quad (1)$$

where $T_{\text{thr}} = 40$ Mbps and $T_{\text{lat}} = 50$ ms. For reporting (not training), we also compute a bounded reward score:

$$r_i^{(k)} = \frac{1}{2} min\left(1, \frac{DLthr_{mean}}{T_{thr}}\right) + \frac{1}{2} min\left(1, \frac{T_{lat}}{lat_{p95} + \varepsilon}\right) \quad (2)$$

with a small $\varepsilon > 0$ for numerical stability. Numeric features are scaled locally (min-max) to [0, 1]. Each client performs an 80/20 stratified split by label for evaluation.

B. Client-side Classifier and Local Objective

We use a compact MLP $f_\theta : \mathbb{R}^d \to [0, 1]$ with hidden sizes (64, 32, 16), ReLU activations, dropout $p = 0.3$ after the first two hidden layers, and a sigmoid output $\hat{y}$. The local loss is binary cross-entropy (BCE). For a mini-batch $B^{(k)}$:

$$\ell(\theta; B^{(k)}) = -\frac{1}{|B^{(k)}|}\sum_{(x,y)\in B^{(k)}} [y \log \hat{y} + (1 - y) \log(1 - \hat{y})] \quad (3)$$

and the client objective is

$$F_k(\theta) = \frac{1}{n_k}\sum_{i=1}^{n_k} \ell(\theta; x_i^{(k)}, y_i^{(k)}) \quad (4)$$

Local optimization uses Adam (batch size 64).

C. Federated Protocol with Partial Participation

In round t, the server selects a subset of available clients S_t (partial participation). Each selected client downloads the current global model $\theta^{(t)}$, performs local training to obtain $\theta_k^{(t)}$, and uploads the update (or model parameters). The server aggregates using FedAvg:

$$\theta^{(t+1)} = \frac{\sum_{k\in S_t} n_k \theta_k^{(t)}}{\sum_{k\in S_t} n_k} \quad (5)$$

We log per-round global accuracy, macro-F1, loss, $|S_t|$, and wall-clock time.

D. Privacy and Secure Aggregation Compatibility

The pipeline is designed to remain compatible with client-level Differential Privacy (DP) and secure aggregation. A standard DP mechanism clips client updates to a norm bound C and injects Gaussian noise:

$$\tilde{g}_k = clip(\nabla F_k(\theta^{(t)}), C), \quad \hat{g}_k = \tilde{g}_k + \mathcal{N}(0, \sigma^2 C^2 I) \quad (6)$$

where σ controls the privacy-utility trade-off. Secure aggregation ensures the server observes only the sum (or average) of client updates, preventing inspection of any single client update.

E. Over-the-Air (AirComp) Uplink Compatibility

To shorten uplink time, the protocol can exploit AirComp, which leverages the multiple-access channel to compute sums in the analog domain. With channel inversion and power control (weights $w_k \approx h_k^{-1}$), the receiver observes an approximate sum:

$$y \approx \sum_{k\in S_t} s_k + n \quad (7)$$

where s_k denotes a transmitted (possibly compressed) update representation and n is noise. AirComp reduces latency but introduces synchronization and physical-layer constraints; the learning API remains unchanged.

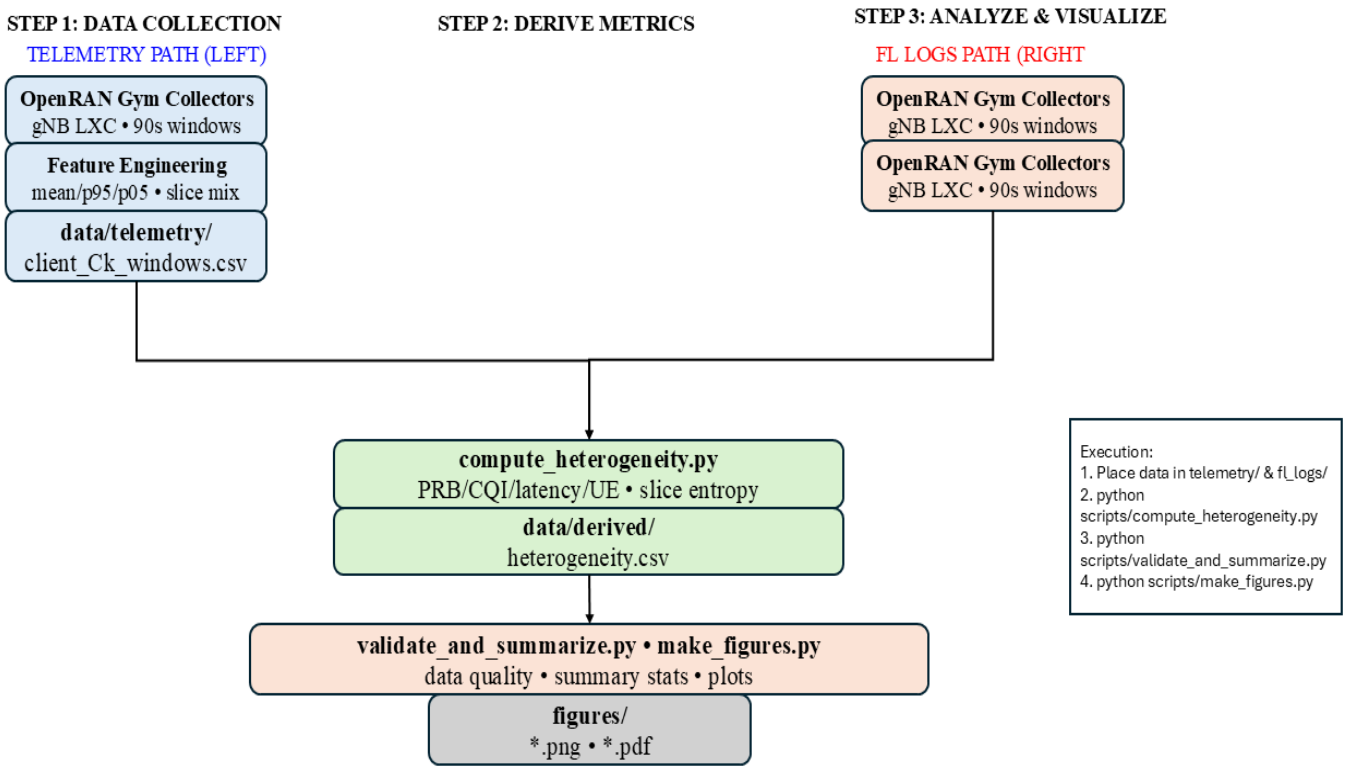


Figure 1: Data pipeline: telemetry collection, windowing and metric derivation, heterogeneity computation, and generation of logs/figures

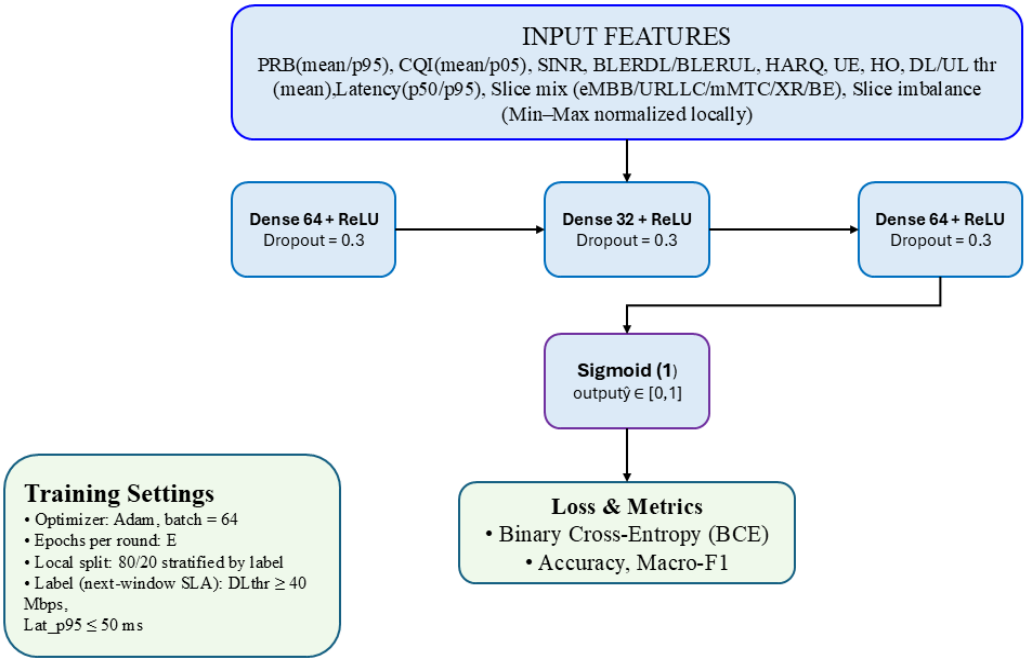


Figure 2: Client-side binary classifier architecture and training configuration

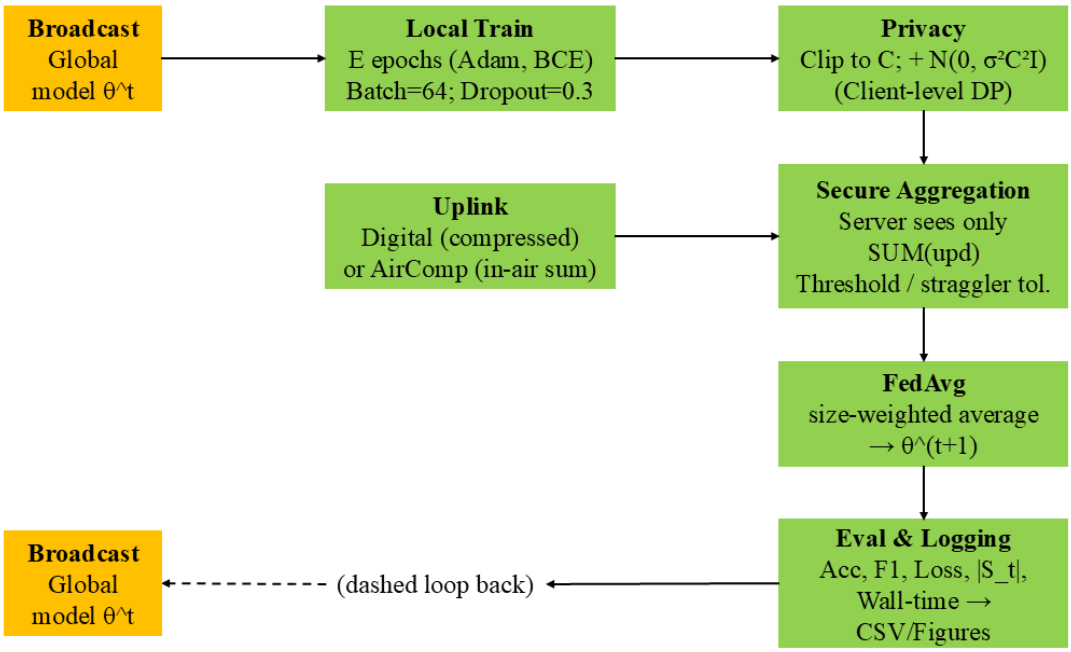


Figure 3: Federated learning round workflow including local training, optional client-level DP, optional secure aggregation, and FedAvg update

F. Metrics, Fairness, and Heterogeneity

We track global accuracy, macro-F1, and loss across rounds, along with client-level behavior. We quantify per-round fairness dispersion using the accuracy gap among participating clients:

$$\Delta_t = \max_{k \in S_t} a_{k,t} - \min_{k \in S_t} a_{k,t} \qquad (8)$$

and also report the gap between the best and worst mean per-client accuracies over training. To summarize slice diversity (non-IID evidence), we compute slice-mix entropy:

$$H_k = -\sum_s \bar{p}_{k,s} \log_2 \left(\bar{p}_{k,s} + \varepsilon\right) \qquad (9)$$

where $\bar{p}_{k,s}$ is the mean fraction of slice s for client k.

G. Reproducibility Artifacts

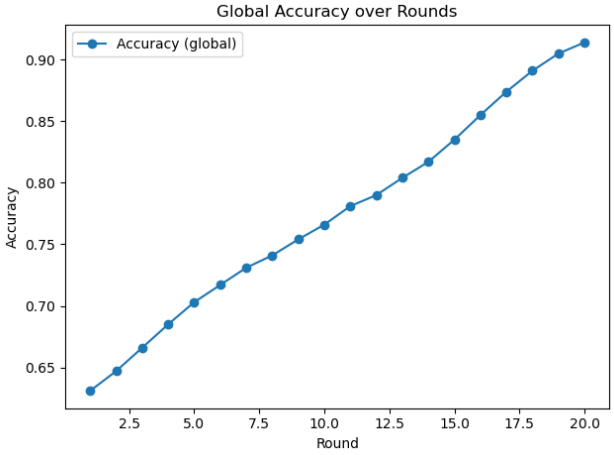
The implementation writes schema-checked CSVs for telemetry windows, per-round server logs, per-client evaluations, and derived heterogeneity metrics, and generates figures for convergence and fairness. This enables replication and ablation studies without manual post-processing.

IV. RESULTS

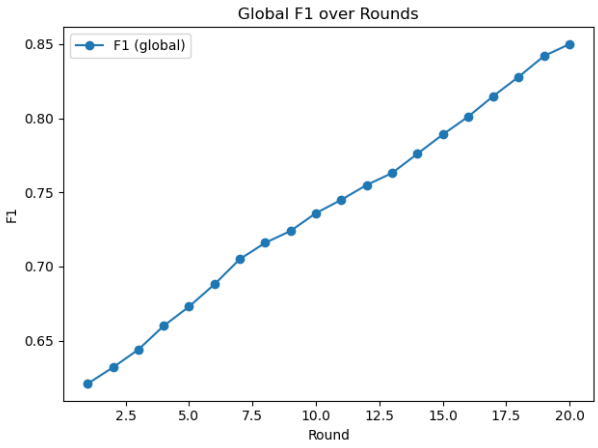
We report convergence, client-level dispersion, fairness, and non-IID evidence using the logged per-round metrics and derived heterogeneity statistics.

A. Convergence

Across 20 global rounds under partial participation (6–8 clients per round), the global model improves steadily. Global accuracy increases from 0.631 to 0.914 and macro-F1 from 0.621 to 0.850, while the training loss decreases from 0.921 to 0.180. Performance gains become marginal after approximately rounds 16–18, suggesting that early stopping near this region can reduce training time with limited impact on final performance.



(a) Global accuracy



(b) Global macro-F1

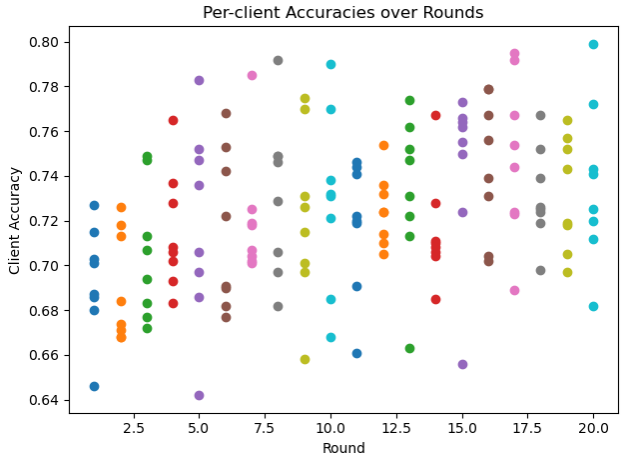
Figure 4: Global performance over federated rounds

TABLE I: KEY ROUNDS FROM SERVER LOGS

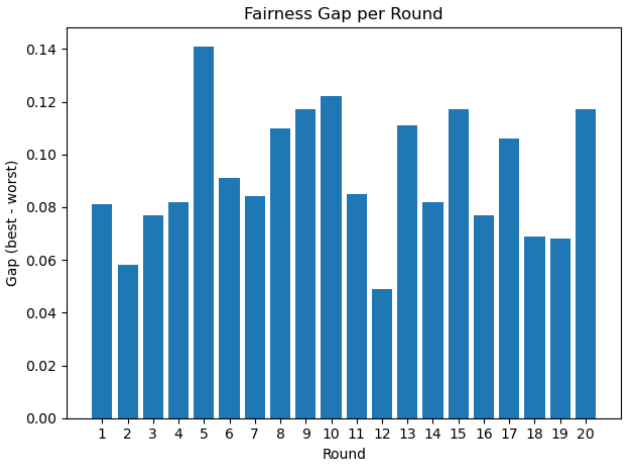
Round	Acc.	Macro-F1	Loss	Participants	Wall (s)
1	0.631	0.621	0.921	7	36.4
10	0.766	0.736	0.528	7	40.8
20	0.914	0.850	0.180	6	38.3

B. Fairness and Client-level Dispersion

We evaluate fairness using (i) the per-round accuracy gap among participating clients and (ii) the gap between the best and worst mean per-client accuracies over training. Per-round gaps typically fall within 0.05–0.14, with a maximum around round 5 (0.141) and a minimum around round 12 (0.049). Over the full run, the best mean per-client accuracy is 0.7592 (C7) and the worst is 0.6796 (C8), yielding a mean gap of 0.0796. Overall, no client collapses; all clients remain within roughly 8–10 percentage points of the best client, indicating bounded dispersion under the evaluated non-IID setting.



(a) Per-client accuracies across rounds



(b) Fairness gap per round (best-worst)

Figure 5: Client-level dispersion and fairness across federated rounds.

C. Heterogeneity (Non-IID Evidence)

To contextualize the learning dynamics, we summarize non-IID conditions using client-level averages of key telemetry-derived statistics. The snapshot indicates substantial cross-client variation in PRB utilization, CQI, p95 latency, UE density, and slice-mix entropy. In particular, heavier-load clients show elevated PRB usage and UE counts and higher slice entropy, while several clients exhibit lower CQI and higher p95 latency, consistent with challenging channel or mobility conditions. These discrepancies help explain both the convergence behavior and the observed fairness gaps.

TABLE II: HETEROGENEITY SNAPSHOT (AVERAGES OVER WINDOWS)

Client	PRB Mean	CQI Mean	Lat p95	UE Mean	Slice Entropy
C1	47.485	8.191	71.367	9.493	1.826
C2	54.121	8.777	75.224	11.153	1.906
C3	66.992	8.293	86.345	14.460	2.017
C4	74.899	8.735	92.101	15.907	2.033
C5	72.853	10.168	83.823	17.620	2.106
C6	80.522	10.603	87.555	17.300	2.134
C7	87.858	10.537	94.214	19.133	2.168
C8	89.489	11.461	92.358	21.587	2.244

D. Operator-facing Summary

The pipeline offers three benefits on the operations side: (i) high overall accuracy of 0.914 and macro-F1 of 0.850 at round 20; (ii) low dispersion at the client level (mean gap 0.0796, with a per-round mean fairness gap of 0.0895); and (iii) measurable evidence of non-IID behavior, including heterogeneity indicators such as load, channel quality, latency and slice diversity. Based on these observations, early stopping around rounds 16–18 appears to be a viable option to shorten training duration. Moreover, fairness-aware sampling or light personalization can be used to further narrow the performance gap between clients.

V. DISCUSSION

The results indicate that a small federated classifier trained on windowed gNB telemetry can realize high accuracy of the whole system with low dispersion of the clients under non-IID data. Yet, there are issues of privacy, communication efficiency, fairness, and system-level issues to consider when applying it in practice. These aspects are discussed below.

A. DP and Secure Aggregation

The proposed training run is privacy-compatible, but by enabling client-level Differential Privacy (DP) and secure aggregation, a trade-off between the utility and privacy of the training can be measured. It is important to appropriately choose the clipping norm C and noise multiplier σ in DP-based FL. In practice, this also means that it is necessary to have a privacy accountant tracking the total privacy budget. The larger the σ, the more privacy is protected, but it might decrease the final accuracy and slow down convergence.

Secure aggregation ensures that the server sees just the aggregated update, not any individual client's update, for sensitive operational telemetry. If the deployment involves setting a conservative clipping strategy with low noise levels, then the noise can be monitored and adjusted for convergence, and the value of σ can be tuned accordingly. This enables the system to comply with privacy constraints while reducing the loss of utility.

B. Communication Efficiency, Compression, AirComp and Early Stopping

Communication can be the predominant aspect of round time, particularly as federation size and/or model width expands. Standard update-compression techniques (such as quantization or sparsification) can be used to reduce uplink payloads, but these can add further stochasticity. Another option is Over-the-Air Computation (AirComp), which takes advantage of the additivity of the wireless multiple-access channel to perform computations directly over the channel. This can help minimize the uplink delay if the synchronization and power control can be achieved.

The fact that performance is observed to saturate around rounds 16–18 suggests that early stopping is an optimization with low risk involved. It can decrease the overall training expense without sacrificing the majority of the achievable accuracy and macro-F1.

C. Fairness and Non-IID Effects

The fairness analysis reveals that the differences in accuracy at the client level are still significant for the case of strong heterogeneity regarding PRB load, CQI, latency, UE density and slice entropy. In the real world, this means that a global-only model can still serve sub-optimal clients or high-load clients, despite positive aggregate figures.

There are two mitigation strategies of particular relevance:

1) *Selection and weighting:* Weaker clients can be reweighted, oversampled, or explicitly considered using worst-client loss minimization.
2) *Light personalization:* A small client-specific component can enhance tail-client performance while retaining global-model performance.

Calibration and decision thresholding. Probability calibration is relevant for consistent thresholding in the context of the model output being a decision signal (e.g., SLA satisfaction likelihood). Post-hoc calibration (e.g., temperature scaling) can enhance the reliability of confidence estimates, especially when there is distribution shift among clients. In reality, operators can perform calibration on held-out telemetry windows and then choose the desired thresholds based on their risk tolerance, such as strict admission to guarantee URLLC latency.

D. System Robustness and Limitations

Partial participation, stragglers, and occasional dropouts of clients occur in realistic RAN deployment. Typically, robust scheduling mechanisms should enforce a minimum quorum, prevent stale updates, for example by damping or using round-based staleness control, and make sure that integrity checks are carried out on updates before they are aggregated. Some limitations of this study are that the size of the federation is relatively small and the classifier is purposefully kept small. Scaling to larger federations, additional slices, or richer model families may change absolute performance and the fairness profile. Moreover, while the pipeline is designed to support DP, secure aggregation, and AirComp, a full privacy accounting and hardware-in-the-loop validation of communication savings remain important future steps.

VI. CONCLUSION AND FUTURE WORK

A comprehensive federated learning pipeline has been proposed for 6G network resource allocation without centralization of raw operational logs that is compatible with privacy. Strong non-IID client behaviour, partial participation, and strict SLA-based labelling resulted in convergence of the global model in 20 rounds with accuracy of 0.914, macro-F1 score of 0.850 and a loss of only 0.180. Fairness dispersion was also kept within bounds: the difference between the best and worst mean per-client accuracy was 0.0796 and the difference between rounds was in the range of 0.05–0.14 on average. A heterogeneity snapshot was taken, and confirmed there was significant diversity in load, channel quality, latency, UE density and slice-mix entropy across clients, giving meaningful context to the progress of convergence and client-level gaps. The pipeline produces reproducible logs and figures, and is designed to be compatible with client-level differential privacy

(clipping with Gaussian noise), secure aggregation, and communication accelerators like update compression and over-the-air computation.

Future work will include laboratory and field validation of AirComp and compression in hardware-in-the-loop or field trials to demonstrate end-to-end AirComp uplink savings, expanding to larger federations and greater skew while maintaining AirComp robustness under system heterogeneity, quantifying the AirComp privacy-utility frontier with full DP accounting under secure aggregation, and reducing remaining gaps at the client level through fairness-aware objectives and light personalization based on slice composition and radio conditions.

REFERENCES

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), PMLR, vol. 54, 2017, pp. 1273-1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[2] T.-M. H. Hsu, H. Qi, and M. Brown, "Measuring the effects of non-identical data distribution for federated visual classification," arXiv preprint, 2019. [Online]. Available: https://arxiv.org/abs/1909.06335

[3] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," arXiv preprint, 2018. [Online]. Available: https://arxiv.org/abs/1806.00582

[4] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in Theory of Cryptography Conference (TCC), LNCS, vol. 3876. Springer, 2006, pp. 265-284.

[5] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in Proc. 2016 ACM SIGSAC Conf. Computer and Communications Security (CCS). New York, NY, USA: ACM, 2016, pp. 308-318.

[6] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in Proc. 2017 ACM SIGSAC Conf. Computer and Communications Security (CCS). New York, NY, USA: ACM, 2017, pp. 1175-1191.

[7] J. Zhang, S. C. Liew, and L. Dai, "Over-the-air computation for wireless federated learning," IEEE Communications Magazine, vol. 58, no. 12, pp. 57-63, 2020.

[8] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 2022-2035, 2020.

[9] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konecny, S. Mazzocchi, H. B. McMahan, T. V. Overveldt, D. Petrou, D. Ramage, and J. Roselander, "Towards federated learning at scale: System design," arXiv preprint, 2019. [Online]. Available: https://arxiv.org/abs/1902.01046

[10] D. Alistarh, D. Grubic, J. Li, R. Tomioka, and M. Vojnovic, "QSGD: Communication-efficient SGD via gradient quantization and encoding," in Advances in Neural Information Processing Systems, 2017, pp. 1709-1720.

[11] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 2031-2063, 2020.

[12] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," arXiv preprint, 2018. [Online]. Available: https://arxiv.org/abs/1812.06127

[13] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in Proc. 37th Int. Conf. Machine Learning (ICML), PMLR, vol. 119, 2020, pp. 5132-5143. [Online]. Available: https://proceedings.mlr.press/v119/karimireddy20a.html

[14] V. Smith, C.-K. Chiang, M. Sanjabi, and A. Talwalkar, "Federated multi-task learning," in Advances in Neural Information Processing Systems, 2017, pp. 4424-4434.

[15] T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: Fair and robust federated learning through personalization," in Proc. 38th Int. Conf. Machine Learning (ICML), PMLR, vol. 139, 2021, pp. 6357-6368. [Online]. Available: https://proceedings.mlr.press/v139/li21h.html

[16] T. Li, M. Sanjabi, A. Beirami, and V. Smith, "Fair resource allocation in federated learning," in Int. Conf. Learning Representations (ICLR) Workshop, 2020. [Online]. Available: https://openreview.net/forum?id=ByexElSYDr

[17] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. 34th Int. Conf. Machine Learning (ICML), PMLR, vol. 70, 2017, pp. 1321-1330. [Online]. Available: https://proceedings.mlr.press/v70/guo17a.html

[18] L. Prechelt, "Early stopping - but when?" in Neural Networks: Tricks of the Trade. Springer, 1998, pp. 55-69.

[19] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," IEEE Transactions on Wireless Communications, vol. 20, no. 1, pp. 269-283, 2021.

[20] N. Noonari, R. Dembani, and N. Maitlo, "HGR: Hand-gesture-recognition based text input method for AR/VR wearable devices," in 2020 IEEE Int. Conf. Systems, Man, and Cybernetics (SMC). IEEE, 2020, pp. 744-751.

[21] N. Maitlo, N. Noonari, S. A. Ghanghro, S. Duraisamy, and F. Ahmed, "Color recognition in challenging lighting environments: CNN approach," in 2024 IEEE 9th Int. Conf. for Convergence in Technology (I2CT). IEEE, 2024, pp. 1-7.

[22] N. Maitlo, N. Noonari, K. Arshid, N. Ahmed, and S. Duraisamy, "AINS: Affordable indoor navigation solution via line color identification using mono-camera for autonomous vehicles," in 2024 IEEE 9th Int. Conf. for Convergence in Technology (I2CT). IEEE, 2024, pp. 1-7.

[23] N. Maitlo, S. K. Bhutto, M. Mahdi, and S. A. Mangi, "GDTII: Gesture driven text input for immersive interfaces," ILMA Journal of Technology & Software Management, vol. 5, no. 2, 2024.

[24] J. Konecny, H. B. McMahan, D. Ramage, and P. Richtarik, "Federated optimization: Distributed machine learning for

on-device intelligence," arXiv preprint, 2016. [Online]. Available: https://arxiv.org/abs/1610.02527