# FedAdaPriv-CPU: Adaptive Differential Privacy with Frozen Backbone for Resource-Constrained Federated Medical Imaging

**Anusha Randhawa , Muhammad Hamza Zaman and Faisal Iradat**
*Department of Computer Science, School of Mathematics and Computer Science*

*Institute of Business Administration, Karachi, Pakistan*

*Correspondence to: Muhammad Hamza Zaman, m.zaman.27162@khi.iba.edu.pk*

## 1. Introduction

Privacy-preserving federated learning (FL) allows hospitals to collaboratively train diagnostic models without sharing patient data, with differential privacy (DP) providing formal guarantees against gradient-based reconstruction attacks [1]. Yet every existing deployment applies DP-SGD (DP-Stochastic Gradient Descent) uniformly across all model parameters, all clients, and all rounds, ignoring the fundamental heterogeneity of real hospital networks. This design choice is both statistically wasteful and computationally prohibitive. Training a full ResNet or MobileNetV2 end-to-end with DP-SGD requires GPU infrastructure that the majority of hospitals in low- and middle-income countries(LMIC) simply do not possess.

**FedAdaPriv-CPU** is a novel federated framework that (i) freezes a pretrained MobileNetV2 backbone so that all client-side computation is CPU-feasible, (ii) applies a three-signal adaptive privacy budget allocation assigning a per-client, per-round $\varepsilon_{k,r}$ instead of a static global ε, and (iii) couples ε allocation directly to aggregation weighting.

## 3. Substantial section

### *3.1 Related work*

This work extends Tanveer et al. [2], who explicitly identify adaptive privacy budgets and resource-constrained deployment as future work, from tabular stroke data to four-class chest X-ray (CXR) classification. It also directly addresses the open problem stated by Li et al. [3] where the authors presented a simple model on devices within multi-level FL frameworks.

Moreover, McMahan et al. [4] introduced DP-FedAvg, applying DP-SGD uniformly to all parameters across all rounds with a fixed noise scale, the static baseline all subsequent work inherits. [6] propose a model that allocates a budget based on similarity. ALDP-FL [7] uses a multi-signal scoring function (accuracy, loss, rounds, dataset size). However, in these studies, DP-SGD is applied to the *full model*, requiring GPU-capable clients, and none couple their ε allocation to aggregation weighting. In medical imaging, Ho et al. [8] demonstrate FL with local DP for COVID-19 CXR, but use full-model training and static ε. The systematic review in [9] covering 300+ DP-FL papers confirms that partial-layer DP confined to a frozen-backbone classifier head has not been studied.

### *3.2 Main Contribution*

The key research question we intend to address is, "What happens to the privacy-utility tradeoff when DP-SGD is applied exclusively to a small-parameter MLP classifier head (approx 50K–200K parameters) atop a frozen pretrained backbone, and how does adaptive per-client per-round epsilon allocation interact with this constrained noise target?". FedAdaPriv-CPU addresses three simultaneous gaps in privacy-preserving federated medical imaging: (C1) the first characterisation of the DP-SGD noise regime for frozen-backbone FL, deriving an empirical εmin threshold, (C2) a three-signal adaptive per-client per-round ε formula coupled bidirectionally to aggregation weighting, (C3) the first CPU-feasibility benchmark for DP-FL on multi-class CXR under non-IID conditions, directly extending [2] to imaging and answering the open resource-constrained edge deployment question raised in [3].

### *3.3 Framework Overview*

Fig. 1 illustrates the four-phase pipeline executed each federation round. A lightweight convolutional backbone pre-trained on ImageNet is frozen on all clients, it is never trained or updated. Only a three-layer MLP classifier head (720K parameters) is trained, and only its gradients are MobileNetV2 as the backbone, selected for its efficient inference profile suitable for resource-constrained edge deployment.

However, the framework is architecture-agnostic and compatible with any fixed feature extractor, including EfficientNet-B0 and ResNet-18.

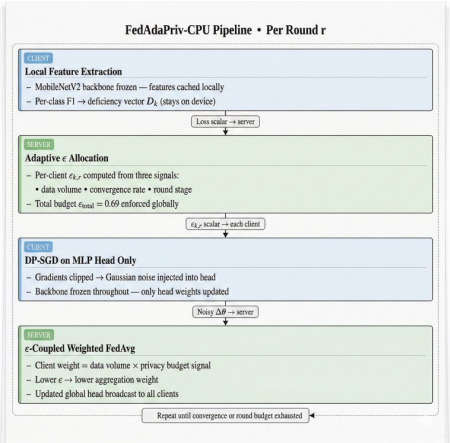


Fig. 1: FedAdaPriv-CPU four-phase pipeline. Backbone never updates. Only head weights (KB) traverse the network.

### *3.4 Methodology*

#### *3.4.1 Local Feature Extraction*

In the first phase, each client independently performs local feature extraction using a frozen MobileNetV2 backbone, projecting raw input images into a 1280-dimensional feature cache $\{f_i\} \in \mathbb{R}^{1280}$. Each client computes a per-class F1 score to derive a local deficiency vector $D_k$, which quantifies class-level convergence gaps.

#### *3.4.2 Adaptive ε Allocation: Three Signal Formula*

The server assigns $\varepsilon_{k,r}$ using three multiplicative signals:

**Signal 1 — data volume:** More data → higher ε → less noise.

$$\alpha_{vol}(k) = \log(n_k+1) / \log(n_{max}+1).$$

**Signal 2 — convergence:** Still learning fast → more noise; near convergence → less noise. Here, Δloss(k,r) is derived from the client's local deficiency vector $D_k$ as the weighted mean of per-class F1 shortfalls.

$$\alpha_{conv}(k,r) = 1 - \exp(-\lambda\Delta loss(k,r)).$$

**Signal 3 — round stage:** Curriculum decay from privacy-first to utility-first.

$$\alpha_{round}(r) = r/R$$

The global budget constraint $\sum_r \varepsilon_{k,r} \leq \varepsilon_{total}$ is tracked via the moments accountant, calibrated so $\varepsilon_{total} = 0.69$ matching [2].

$$\varepsilon_{k,r} = \varepsilon_{base} \times \alpha_{vol}(k) \times \alpha_{conv}(k,r) \times \alpha_{round}(r)$$

$$\Sigma\, \varepsilon_{k,r} \leq \varepsilon_{total} = 0.69$$

The sum of all per-client, per-round privacy budgets must not exceed a total cumulative epsilon of 0.69 [2].

#### *3.4.3 The Frozen-Head Noise Regime*

A key theoretical finding motivates proposed design. Standard DP-SGD distributes noise across $P_{full}$ parameters (approx 3.4M for MobileNetV2). When confined to a frozen MLP head with $P_{head} \approx 720K$ parameters, the per-parameter signal-to-noise ratio changes by $\sqrt{(P_{full}/P_{head})} \approx 2.2x$ in favour of the head. However, the head lacks the redundancy of a large model: below a minimum viable ε ($\varepsilon_{min}$), accuracy collapses catastrophically rather than degrading gracefully.

This study empirically characterise this $\varepsilon_{min}$ threshold predicted at approx 0.3–0.4 — and show that above $\varepsilon_{min}$ the frozen-head approach outperforms full-model DP-SGD on CPU at equal ε, producing a crossover accuracy curve absent from all prior work.

#### *3.4.4 ε-Coupled Weighted FedAvg*

The server performs aggregation using a privacy-aware weighting scheme termed ε-Coupled Weighted FedAvg. Rather than aggregating client updates equally, each client is assigned a weight $W_k = (n_k / n_total) \times \alpha vol(k)$, which jointly accounts for the client's data volume and its allocated privacy budget. Clients operating under tighter constraints contribute proportionally less to the global update, reflecting the higher noise corruption in their gradients.

The server then computes the new global head as Θ_global as ($\Theta_global \leftarrow \Sigma_k\, W_k\, \Delta\Theta_k$) and distributes the updated MLP head weights back to all participating clients. Critically, only the lightweight head parameters are transmitted rather than the full model, keeping communication overhead minimal.

#### *3.4.5 Experimental Setup & Expected Results*

COVID-19 Radiography Dataset (Kaggle, 21,165 CXR images; 4 classes, partitioned into 3 hospital clients: Hospital A (70% COVID-19), Hospital B (70% Lung Opacity), Hospital C (70% Viral Pneumonia, fewest samples ~3,200). All client-side computation runs CPU-only; R = 25 federated rounds, E = 5 local epochs, b = 64, $\delta = 10^{-5}$. The framework is expected to achieve competitive accuracy at equal privacy cost with an 8–10× reduction in client-side training time, making privacy-preserving federated CXR diagnosis genuinely deployable in low-resource hospital settings.

**Acknowledgments**

The authors gratefully acknowledge Department of Computer Science for their support.

**References**

[1] Abadi, M. et al. Deep learning with differential privacy. ACM CCS, pp. 308–318, 2016.
[2] Tanveer, F., Iradat, F., Iqbal, W. et al. Balancing privacy and performance in healthcare: A federated learning framework for sensitive data. Digital Health, SAGE, 2025. DOI: 10.1177/20552076251381769
[3] Li, X. et al. MLA-FedFSL: Multi-level aggregation for federated few-shot learning. Neurocomputing, 2025. DOI: 10.1016/j.neucom.2025.012469
[4] McMahan, H.B., Ramage, D., Talwar, K., Zhang, L. Learning differentially private recurrent language models. ICLR, 2018.
[5] Fu, J., Chen, Z., Han, X. Adap DP-FL: Differentially private federated learning with adaptive noise. IEEE TrustCom, 2022.
[6] Chen, L. et al. Differentially private deep learning with dynamic privacy budget allocation. IEEE T-IFS, 18, 4422–4435, 2023.
[7] ALDP-FL: Adaptive local differential privacy in federated learning. Scientific Reports, Nature, 2025.
[8] Ho, T.T., Tran, K.D., Huang, Y. FedSGDCOVID: Federated SGD COVID-19 detection under local DP using chest X-ray images. Sensors, 22, 3728, 2022.
[9] Systematic review: Differentially private federated learning (300+ papers). arXiv:2405.08299, 2025.