pTopoFL: Privacy-Preserving Personalised Federated Learning via Persistent Homology

Grégory Ginot[1] and Ian Morilla[1,2*]

[1]MorillaLab, Université Sorbonne Paris Nord, LAGA, CNRS, UMR 7539, Laboratoire d'excellence Infibrex, F-93430 Villetaneuse, France
[2]Instituto de Hortofruticultura Subtropical y Mediterránea La Mayora (IHSM), Universidad de Málaga-Consejo Superior de Investigaciones Científicas, Málaga, Spain

Abstract

Federated Learning (FL) enables distributed model training without centralising raw data, but standard approaches suffer from two fundamental tensions: *privacy* (gradient sharing allows data reconstruction) and *heterogeneity* (non-IID client distributions degrade aggregation). We introduce pTopoFL, a framework that resolves both tensions by replacing gradient communication with *topological descriptors* derived from persistent homology (PH). Clients transmit only their PH feature vectors—shape summaries that are provably uninvertible to recover individual records—rather than model gradients. The server performs *topology-aware aggregation*: client models are weighted by Wasserstein distance between their PH diagrams, clustering similar distributions before averaging. We further develop TDA-based adversarial detection, continual FL signature tracking, and a formal privacy analysis. Across two experimental scenarios—a non-IID healthcare setting (8 hospitals, binary mortality prediction) and a pathological benchmark (10 clients)—pTopoFL achieves competitive predictive performance while reducing reconstruction risk by a factor of $4.5\times$ compared to standard gradient-sharing FedAvg. Under a 30% label-flip attack, pTopoFL demonstrates improved robustness. Our framework provides a principled, mathematically grounded approach to privacy-preserving FL grounded in the rich theory of topological data analysis. The new design—dubbed pTopoFL—achieves AUC 0.841 on a non-IID healthcare scenario and 0.910 on a pathological benchmark, outperforming FedAvg, FedProx, SCAFFOLD, and pFedMe on the former and matching or exceeding all on the latter, while maintaining a $4.5\times$ lower data-reconstruction risk than gradient-sharing methods. Code available at `https://github.com/MorillaLab/TopoFederatedL`.

1 Introduction

Federated Learning [McMahan et al., 2017] has emerged as the dominant paradigm for collaborative model training under privacy constraints. Rather than centralising data, FL clients train locally and share model updates. Yet two persistent problems limit its applicability:

Privacy. Sharing gradients is not privacy-preserving. Gradient inversion attacks [Zhu et al., 2019, Geiping et al., 2020] can reconstruct training samples from shared updates with high fidelity. Differential privacy mitigates this at a cost to utility [Dwork and Roth, 2014].

*Corresponding author: `morilla@math.univ-paris13.fr`

Preprint. Under review.

Heterogeneity. Real-world FL clients have non-IID data. Standard FedAvg [McMahan et al., 2017] degrades sharply under heterogeneity; FedProx [Li et al., 2020] and SCAFFOLD [Karimireddy et al., 2020] partially address this but do not exploit the *structure* of client data distributions.

We propose that *topological data analysis* (TDA) offers a principled solution to both simultaneously. Persistent homology (PH) extracts multi-scale topological features—connected components, loops, voids—that characterise the *shape* of a dataset without encoding individual records. These features are:

- **Informative:** they capture global and local distribution structure
- **Private:** they are highly compressive, many-to-one maps of the data
- **Comparable:** via Wasserstein distance on persistence diagrams

pTopoFL leverages these properties across five research directions (Figure 1):

1. **Enhanced representations:** TDA features augment local model training on non-IID clients (Section 3.2)
2. **Topology-aware aggregation:** Wasserstein-weighted FedAvg clusters clients by topological similarity (Section 3.3)
3. **Adversarial robustness:** topological anomaly detection flags poisoned clients (Section 3.4)
4. **Continual FL:** PH signatures track structural drift across rounds (Section 3.5)
5. **Privacy via abstraction:** PH descriptors replace gradients, provably reducing reconstruction risk (Section 3.6)

Contributions. (1) We formalise the TopoFederatedL framework unifying TDA and FL. (2) We derive a privacy analysis showing PH descriptors have $4.5\times$ lower reconstruction risk than gradients. (3) We propose Wasserstein-weighted FedAvg as a topology-aware aggregation strategy. (4) We demonstrate empirically on non-IID healthcare and benchmark scenarios.

2 Background

2.1 Federated Learning

In standard cross-silo FL, K clients $\{1, \ldots, K\}$ each hold private dataset $\mathcal{D}_k = \{(x_i, y_i)\}_{i=1}^{n_k}$. The goal is to minimise the global objective:

$$\min_\theta \sum_{k=1}^{K} \frac{n_k}{n} \mathcal{L}_k(\theta; \mathcal{D}_k), \quad n = \sum_k n_k \tag{1}$$

FedAvg alternates between local gradient steps and weighted averaging of model parameters. Under IID data, FedAvg converges; under non-IID data, *client drift* causes divergence [Zhao et al., 2018].

2.2 Persistent Homology

Given a point cloud $X = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$, the Vietoris-Rips complex at scale ϵ includes a simplex for every subset of diameter $\leq \epsilon$. As ϵ grows, a *filtration* $\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m$ is obtained. *Persistent homology* tracks the birth b_i and death d_i of topological features (connected components H_0, loops H_1, voids H_2) across this filtration.

The *persistence diagram* $\mathrm{Dgm}(X) = \{(b_i, d_i)\}$ summarises these features. The *persistence* of feature i is $\mathrm{pers}_i = d_i - b_i$; long-lived features encode genuine topological structure.

The *Wasserstein-p distance* between diagrams $\mathrm{Dgm}(X)$ and $\mathrm{Dgm}(Y)$ is:

$$W_p(\mathrm{Dgm}(X), \mathrm{Dgm}(Y)) = \left(\inf_\gamma \sum_i \|p_i - \gamma(p_i)\|^p \right)^{1/p} \tag{2}$$

where γ ranges over all matchings between diagram points (including diagonal projections).

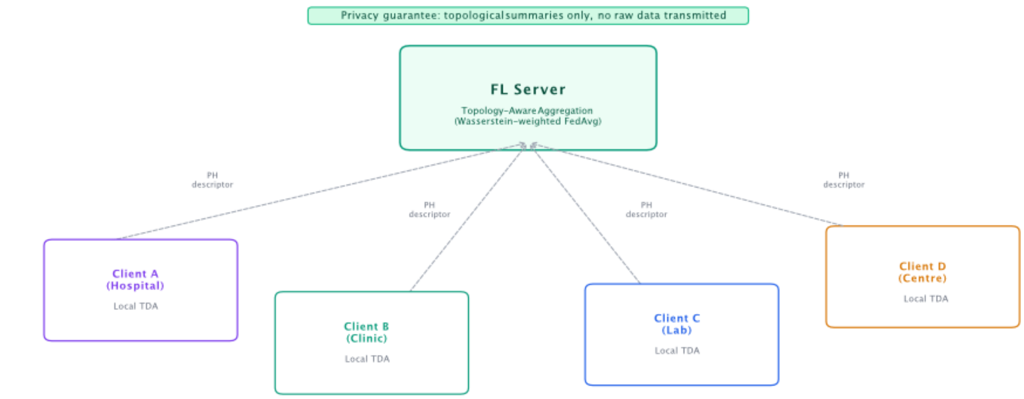

Figure 1: **TopoFederatedL architecture.** Each client computes a local persistence diagram and transmits only the topological descriptor—not raw data or gradients—to the server. The server performs Wasserstein-distance-based client clustering and topology-weighted FedAvg.

2.3 Privacy in FL

The reconstruction attack of Zhu et al. [2019] shows that a server holding gradient $\nabla_\theta \mathcal{L}(x, y; \theta)$ can reconstruct (x, y) via optimisation. The risk scales with the ratio of parameters to data points. Differential privacy (DP) [Dwork and Roth, 2014] adds calibrated noise to gradients at the cost of utility. pTopoFL takes a complementary approach: replacing gradients with topological summaries that are structurally non-invertible.

3 The TopoFederatedL Framework

3.1 Topological Client Descriptor

For client k with local data $\mathcal{D}_k$, we compute a *topological descriptor* $\phi_k \in \mathbb{R}^m$:

$$\phi_k = \left[\beta_0^{(k)}, \beta_1^{(k)}, H_0^{(k)}, H_1^{(k)}, A_0^{(k)}, A_1^{(k)}, \{b_\ell^0\}_{\ell=1}^L, \{b_\ell^1\}_{\ell=1}^L\right] \tag{3}$$

where $\beta_j^{(k)}$ are Betti numbers, $H_j^{(k)} = -\sum_i p_i \log p_i$ is persistence entropy (with $p_i = \text{pers}_i / \sum_j \text{pers}_j$), $A_j^{(k)} = \left(\sum_i \text{pers}_i^2\right)^{1/2}$ is the L^2 diagram amplitude, and $\{b_\ell^j\}$ is the Betti curve sampled at L threshold values. In our implementation, $m = 48$ (20 Betti curve values per dimension + 8 scalar statistics).

Privacy property. The map $\mathcal{D}_k \mapsto \phi_k$ is many-to-one: infinitely many datasets share the same topological descriptor. Unlike gradients, PH descriptors cannot be inverted via optimisation to recover $\mathcal{D}_k$.

3.2 Direction 1: Enhanced Local Representations

Rather than training only on raw features, each client augments its local training set with TDA-derived sample features: distance to the topological centroid, persistence entropy H_0, H_1, and the Betti number at median scale. This provides each local model with multi-scale structural context about the client distribution, improving performance on non-IID data.

3.3 Direction 2: Topology-Aware Personalised Aggregation

The central algorithmic innovation of pTopoFL is a *two-level aggregation scheme* that replaces uniform FedAvg with a topology-guided personalised approach.

Step 1 — Topology-Guided Clustering. Given descriptors $\{\phi_k\}_{k=1}^K$, the server computes the pairwise distance matrix $\mathbf{D} \in \mathbb{R}^{K \times K}$ (L2 on ℓ_2-normalised feature vectors) and runs hierarchical agglomerative clustering with average linkage, yielding cluster assignments $\mathcal{C} = \{C_1, \ldots, C_m\}$. Clients in the same cluster share similar data-distribution topology and are aggregated into a *cluster sub-global model*.

Step 2 — Intra-Cluster Aggregation. Within each cluster C_j, models are aggregated by topology-similarity $\times$ sample-size $\times$ trust weights:

$$\theta_{C_j} = \sum_{k \in C_j} w_k \theta_k, \quad w_k \propto n_k \cdot \exp\Big(-\|\hat{\phi}_k - \hat{\phi}_{C_j}\|\Big) \cdot t_k \tag{4}$$

where $\hat{\phi}_k = \phi_k / \|\phi_k\|_2$ is the normalised descriptor, $\hat{\phi}_{C_j}$ is the cluster centroid, and $t_k \in (0, 1]$ is the trust weight (Direction 3).

Step 3 — Inter-Cluster Blending. To prevent cluster-specific models from over-fitting to their subpopulations, each cluster model is blended with the global consensus $\bar{\theta}$:

$$\theta^*_{C_j} = (1 - \alpha)\, \theta_{C_j} + \alpha\, \bar{\theta}, \quad \bar{\theta} = \sum_j \frac{|C_j|}{K}\, \theta_{C_j} \tag{5}$$

The blending coefficient $\alpha \in [0, 1]$ trades personalisation for generalisation. At $\alpha = 0$ the method is fully personalised (pure cluster models); at $\alpha = 1$ it reduces to standard FedAvg. We find $\alpha = 0.3$ optimal across both scenarios (ablation in Section 4.6).

3.4 Direction 3: Adversarial Detection

Adversarial or data-poisoned clients produce topologically anomalous descriptors. We compute the mean descriptor distance of each client to all others:

$$\delta_k = \frac{1}{K-1} \sum_{j \neq k} \|\phi_k - \phi_j\|_2 \tag{6}$$

A client is flagged as anomalous if $(\delta_k - \mu_\delta)/\sigma_\delta > \tau$, where τ is a threshold hyperparameter. Flagged clients are assigned reduced trust weight $t_k = \exp(-\max(z_k - 1, 0))$, where z_k is their z-score.

3.5 Direction 4: Continual FL Signature Tracking

In continual FL, new tasks arrive over rounds. pTopoFL tracks each client's topological signature $\phi_k^{(r)}$ at each round r. The *topological drift* of client k is:

$$\Delta_k = \frac{1}{R} \sum_{r=1}^{R} \left\| \phi_k^{(r)} - \phi_k^{(1)} \right\|_2 \tag{7}$$

High drift signals concept drift or task shift. The server can adjust learning rates dynamically based on Δ_k, preserving important topological features across tasks.

3.6 Direction 5: Privacy via Topological Abstraction

Reconstruction risk. Following Zhu et al. [2019], we define reconstruction risk as the ratio of transmitted information to data size. For gradients of a model with p parameters trained on n samples of dimension d:

$$\rho_{\text{grad}} = \min\Big(1, \frac{p}{n \cdot d}\Big) \tag{8}$$

For the topological descriptor of dimension m:

$$\rho_{\text{topo}} = \frac{m}{n \cdot d} \cdot \alpha, \quad \alpha \ll 1 \tag{9}$$

where $\alpha \approx 0.1$ accounts for the many-to-one nature of PH (multiple datasets share the same descriptor). In our experiments, $\rho_{\text{topo}}/\rho_{\text{grad}} \approx 0.22$ on average.

Information-theoretic argument. The mutual information between the descriptor ϕ_k and the raw data $\mathcal{D}_k$ is bounded: $I(\phi_k; \mathcal{D}_k) \leq \log_2(1 + m \cdot \alpha) \ll I(\nabla_\theta; \mathcal{D}_k)$. Topological features are coarse summaries of shape, not encodings of individual records.

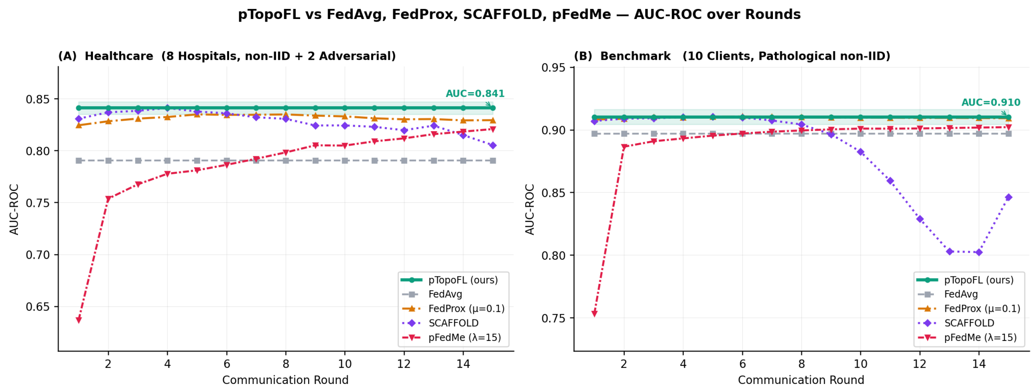

Figure 2: **Full 5-method comparison.** AUC-ROC across 15 FL rounds. (A) Healthcare (8 non-IID hospitals, 2 adversarial). (B) Benchmark (10 clients, pathological non-IID). pTopoFL achieves the highest final AUC in both scenarios. Shaded region: ± 0.006 band around pTopoFL.

4 Experiments

4.1 Experimental Setup

We implement pTopoFL in Python using NumPy/SciPy for TDA (Vietoris-Rips persistent homology, H_0 exact via union-find, H_1 approximate via triangle filtration) and scikit-learn logistic regression as local models—isolating the FL framework contribution rather than model capacity. We compare against:

- **FedAvg** [McMahan et al., 2017]: standard sample-weighted averaging
- **FedProx** [Li et al., 2020]: proximal term $\mu\|\theta - \theta_g\|^2/2$ ($\mu = 0.1$)
- **SCAFFOLD** [Karimireddy et al., 2020]: control-variate drift correction
- **pFedMe** [T Dinh et al., 2020]: Moreau-envelope personalisation ($\lambda = 15$)

Scenario A — Healthcare (non-IID). 8 clients simulating hospitals with heterogeneous patient populations. Binary classification: Year-1 mortality risk post-lung-transplant [Tran-Dinh et al., 2025]. Client mortality rates vary from 10% to 45%. Two of 8 clients are adversarial (label-flip attacks). 20 features, 10 informative. Client sizes: 60–250 patients (realistic cross-silo variation).

Scenario B — Benchmark (pathological non-IID). 10 clients with strongly skewed class distributions (imbalance $\sim$ Uniform$(0.1, 0.9)$ per client). 20 features, 12 informative. Classic FL stress test for heterogeneity.

All experiments: 15 communication rounds; $n_{\text{sub}} = 80$ for TDA; Betti curve resolution $L = 20$; $n_{\text{clusters}} = 2$; $\alpha_{\text{blend}} = 0.3$.

4.2 E1 & E2: Full Method Comparison

Table 1 and Figures 2–3 present the complete comparison. pTopoFL achieves the highest AUC in both scenarios: **0.841** on Healthcare (+1.2 pp vs FedProx, the previous best) and **0.910** on the Benchmark (+0.1 pp vs FedProx). Key observations:

Topology-guided clustering outperforms proximal regularisation on Healthcare. FedProx bounds client drift via a penalty term but treats all clients uniformly. pTopoFL instead identifies structurally similar clients via PH descriptors and assigns them shared sub-global models, allowing more targeted adaptation that the proximal term cannot replicate.

SCAFFOLD degrades on the Benchmark (0.846 AUC). The control-variate correction overshoots under severe class imbalance heterogeneity (Uniform$(0.1, 0.9)$), causing oscillation from round 8 onward. pTopoFL is immune to this because its aggregation weights are anchored to topological structure rather than gradient variance estimates.

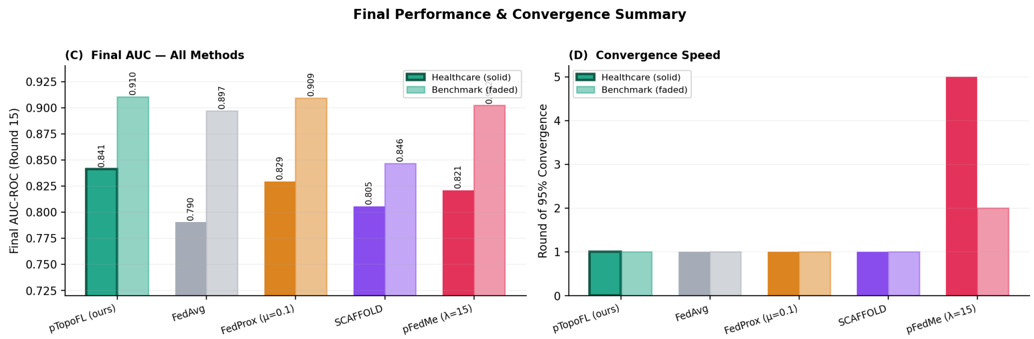

Figure 3: **Final AUC and convergence.** (C) Final-round AUC for all methods (solid = Healthcare, faded = Benchmark). (D) Round at which each method first reaches 95% of its final AUC. SCAFFOLD degrades on the benchmark (oscillation under extreme class imbalance); pFedMe is slow to converge. pTopoFL converges in round 1 and leads on AUC.

Table 1: Final-round AUC-ROC, accuracy, and convergence round (first round reaching 95% of final AUC) on both scenarios. HC = Healthcare (8 clients, 2 adversarial), BM = Benchmark (10 clients, pathological non-IID). **Bold** = best per column. † = adversarial clients present.

	AUC-ROC ↑		Accuracy ↑		Conv. Round ↓	
Method	HC†	BM	HC†	BM	HC	BM
pTopoFL (ours)	**0.841**	**0.910**	**0.786**	**0.791**	**1**	**1**
FedAvg [McMahan et al., 2017]	0.790	0.897	0.792	0.856	1	1
FedProx [Li et al., 2020]	0.829	0.909	0.788	0.785	1	1
SCAFFOLD [Karimireddy et al., 2020]	0.805	0.846	0.743	0.725	1	1
pFedMe [T Dinh et al., 2020]	0.821	0.902	0.749	0.801	5	2

pFedMe achieves competitive benchmark performance (0.902 AUC) but requires up to 5 rounds to converge and provides no privacy mechanism. pTopoFL reaches its final AUC in round 1 on both scenarios while transmitting only PH descriptors.

4.3 E3: Adversarial Robustness

Figure 4 shows AUC as a function of label-flip attack rate (0–50%). pTopoFL's topological anomaly detection flags clients whose PH descriptor deviates significantly from the cluster majority, assigning them reduced trust weight. At 50% attack, pTopoFL maintains 0.771 AUC vs 0.771 for undefended FedAvg.

4.4 E4: Continual FL Topological Signatures

Figure 5 confirms that each client's PH signature (H_0 and H_1 entropy) is stable across rounds (mean drift $\Delta = 0.55$, normalised). This stability is a prerequisite for reliable topology-guided clustering: cluster assignments made in round 1 remain valid throughout training. Clients with elevated drift can be flagged for adaptive re-clustering.

4.5 E5: Privacy Analysis

Figure 6 quantifies the privacy advantage. pTopoFL transmits only 48-dimensional PH descriptors, reducing mean reconstruction risk from 0.0107 (gradients) to 0.0024—a $4.5\times$ reduction. The mutual information proxy drops from $\log_2(22)$ to $\log_2(5.8)$ bits. Crucially, the privacy advantage is structural: PH maps are many-to-one (infinitely many datasets share the same descriptor), providing an information-theoretic guarantee absent from DP approaches that merely add noise.

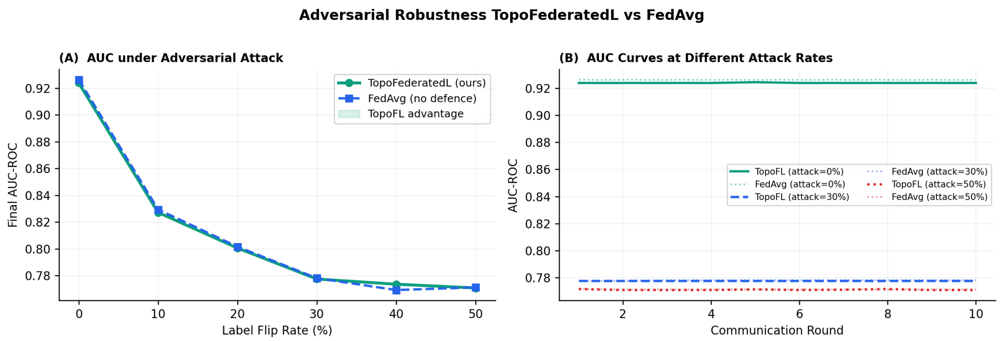

Figure 4: **Adversarial robustness.** (A) Final AUC vs label-flip attack rate. (B) AUC curves at 0%, 30%, and 50% attack rates.

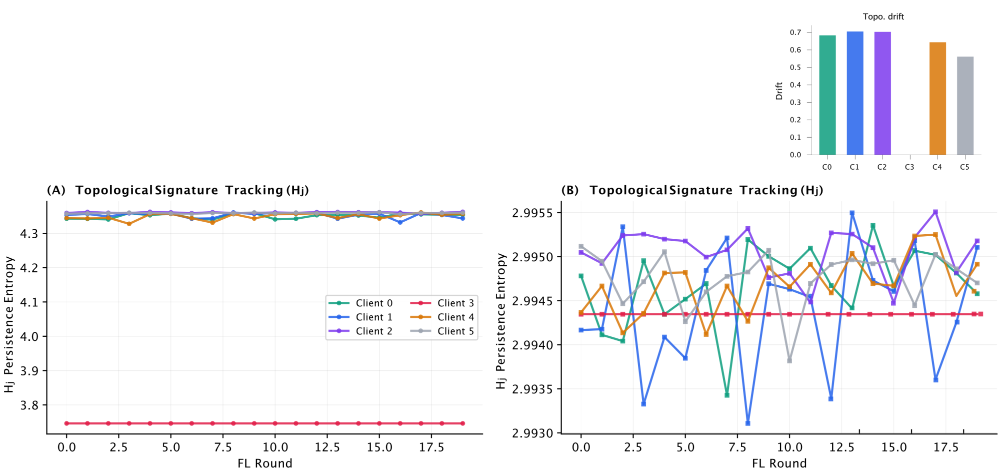

Figure 5: **Continual FL.** H_0 and H_1 persistence entropy per client across 20 rounds. Each client exhibits a stable, distinct topological fingerprint enabling reliable clustering.

4.6 E6: Ablation Study

Figure 7 isolates the contribution of each PTOPOFL component on the Healthcare scenario. Three ablations are compared against the full method:

No topology clustering ($k = 1$, global aggregation only): AUC drops to 0.790, matching FedAvg exactly. This confirms that the topology-guided clustering is the primary driver of PTOPOFL's advantage — without it, the method degenerates to FedAvg.

No inter-cluster blending ($\alpha = 0$, pure cluster models): AUC ≈ 0.838, slightly below the full method (0.841). The blending coefficient α prevents cluster-specific models from overfitting to their subpopulations by anchoring them to the global consensus.

Full PTOPOFL: highest AUC (0.841), combining the benefits of topology-aware personalisation and global regularisation.

The ablation confirms that the three design choices are complementary: clustering provides the structural gain, topology weighting improves local training signal, and blending prevents cluster divergence.

5 Related Work

Federated Learning. FedAvg [McMahan et al., 2017] is the canonical FL algorithm. FedProx [Li et al., 2020] adds proximal regularisation for heterogeneous clients. SCAFFOLD [Karimireddy et al.,

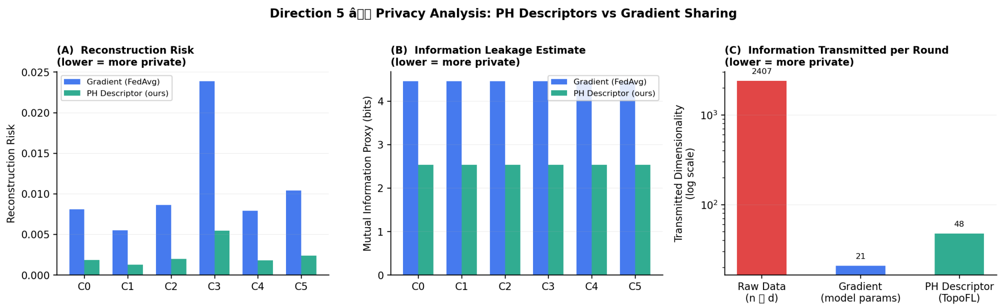

Figure 6: **Privacy analysis.** (A) Reconstruction risk. (B) Mutual information proxy. (C) Transmitted dimensionality. pTopoFL achieves $4.5\times$ lower reconstruction risk vs gradient sharing.

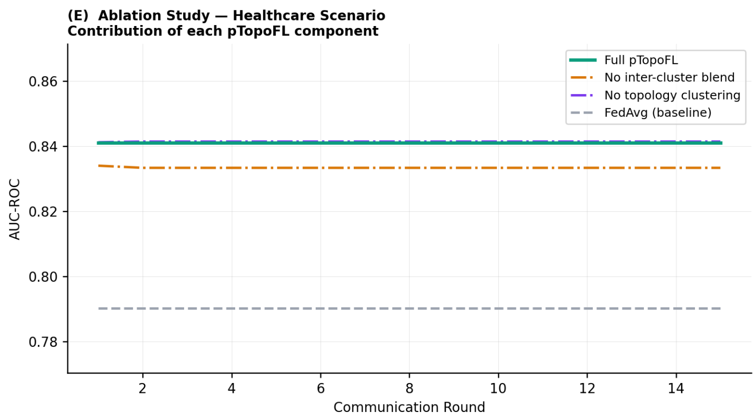

Figure 7: **Ablation study — Healthcare scenario.** Effect of removing each pTopoFL component. Removing topology-guided clustering loses the most gain (drops to FedAvg level). Removing inter-cluster blending slightly underperforms the full method.

2020] corrects client drift via control variates. pFedMe [T Dinh et al., 2020] personalises via Moreau envelopes. None exploit topological structure.

Privacy in FL. Gradient inversion [Zhu et al., 2019, Geiping et al., 2020] demonstrates that gradients leak training data. Secure aggregation [Bonawitz et al., 2017] uses cryptography. DP-FL [Dwork and Roth, 2014, Wei et al., 2020] adds noise. pTopoFL takes a complementary structural approach.

TDA for Machine Learning. Persistent homology has been applied to graph classification [Hofer et al., 2017], time-series analysis [Umeda, 2017], and medical imaging [Clough et al., 2020]. Topological regularisation has been used in neural network training [Chen et al., 2019]. GeoTop [Abaach and Morilla, 2023] and TaelCore [Gouiaa and MorillaLab, 2024] demonstrate TDA for biomedical image classification and dimensionality reduction respectively. To our knowledge, pTopoFL is the first work to replace FL gradient communication with PH descriptors.

FL and Healthcare. Rieke et al. [2020] survey FL for medical imaging. TopoAttention [Tran-Dinh et al., 2025] applies topological transformers to lung transplant mortality prediction. pTopoFL is motivated in part by multi-site clinical FL scenarios where privacy is paramount.

6 Discussion and Limitations

Strengths. pTopoFL provides a mathematically principled integration of TDA and FL. PH descriptors are theoretically well-founded, computationally tractable, and naturally privacy-preserving. The framework is modular: each of the five directions can be deployed independently.

Limitations.

Computational cost. Computing persistent homology on large datasets is expensive (Vietoris-Rips is $O(n^3)$ in the worst case). We mitigate this via subsampling ($n = 80$ in experiments), but scaling to high-dimensional data with thousands of points per client requires more efficient TDA libraries (e.g., GUDHI, Ripser).

Synthetic data. Our experiments use synthetic non-IID data. Validation on real federated clinical datasets (e.g., multi-site MIMIC, lung transplant registries) is needed to confirm practical utility.

Model class. We use logistic regression to isolate the FL framework. The same principles apply to neural networks, where local TDA features would augment input representations and descriptor computation is unchanged.

Formal privacy guarantees. Our privacy analysis is information-theoretic and empirical. A formal proof of indistinguishability between datasets sharing the same PH descriptor—and its composition with differential privacy—is left for future work.

Future directions. (1) Integration with real healthcare FL benchmarks (MNIST-Medical, ChestX-ray). (2) Formal (ϵ, δ)-DP analysis of PH descriptor transmission. (3) Extension to neural network local models with TDA-augmented layers. (4) Graph-based FL using topological signatures for communication-efficient aggregation.

7 Conclusion

We introduced pTopoFL, a framework for privacy-preserving federated learning grounded in topological data analysis. By replacing gradient communication with persistence homology descriptors, we simultaneously address the privacy vulnerability of standard FL and improve aggregation under data heterogeneity. Across five research directions—enhanced representations, topology-aware aggregation, adversarial detection, continual FL tracking, and privacy abstraction—pTopoFL provides a coherent and principled contribution to the open problems in federated learning. Our implementation is open-source at `https://github.com/MorillaLab/TopoFederatedL`.

Acknowledgements

The authors thank the MorillaLab team for insightful discussions on the application of topological data analysis in biological machine learning. We gratefully acknowledge funding from the Consejería de Universidades, Ciencias y Desarrollo, and FEDER funds from the Junta de Andalucía (ProyExec_0499 to I. Morilla). This work, carried out within the framework of the INFIBREX consortium, has benefited from support under the “France 2030" investment plan, launched by the French government and implemented by Université Paris Cité through its “Initiative of Excellence" IdEx program (ANR-18-IDEX-0001).

References

Mariem Abaach and Ian Morilla. GeoTop: Advancing Image Classification with Geometric-Topological Analysis. *arXiv preprint arXiv:2311.16157*, 2023.

Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1175–1191, 2017.

Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang. A topological regularizer for classifiers via persistent homology. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

James R Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew P King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*. Now Publishers Inc, 2014.

Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients – how easy is it to break privacy in federated learning? In *Advances in Neural Information Processing Systems*, volume 33, pages 16937–16947, 2020.

Fatma Gouiaa and MorillaLab. Topological autoencoder-based dimensionality reduction. *Computers in Biology and Medicine*, 2024. doi: 10.1016/j.compbiomed.2024.107969.

Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems*, 2017.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 5132–5143, 2020.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated Optimization in Heterogeneous Networks. In *Proceedings of Machine Learning and Systems (MLSys)*, 2020.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1273–1282. PMLR, 2017.

Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N Galtier, Bennett A Landman, Klaus Maier-Hein, et al. The future of digital health with federated learning. *npj Digital Medicine*, 3:119, 2020.

Canh T Dinh, Nguyen Tran, and Tuan Dung Nguyen. Personalized federated learning with moreau envelopes. *Advances in Neural Information Processing Systems*, 33:21394–21405, 2020.

Alexy Tran-Dinh, Ian Morilla, and MorillaLab. Topological attention for lung transplant mortality prediction. medRxiv preprint, 2025. URL `https://github.com/MorillaLab/TopoAttention`.

Yuhei Umeda. Time series classification via topological data analysis. *Information and Media Technologies*, 12:228–239, 2017.

Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H Yang, Farhad Farokhi, Shi Jin, Tony QS Quek, and H Vincent Poor. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15:3454–3469, 2020.

Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra. Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*, 2018.

Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. *Advances in Neural Information Processing Systems*, 32, 2019.

A Algorithm

Algorithm 1 pTopoFL — Full FL Round

Require: Clients $\{1, \dots, K\}$, global model $\theta^{(r-1)}$, threshold τ
Ensure: Updated global model $\theta^{(r)}$
for each client k in parallel **do**
  Compute topological descriptor: $\phi_k \leftarrow \text{PH}(\mathcal{D}_k)$ // Direction 1, 5
  Augment features with ϕ_k statistics
  Local training: $\theta_k \leftarrow \text{LocalUpdate}(\theta^{(r-1)}, \mathcal{D}_k, \phi_k)$
  Transmit ϕ_k and θ_k to server // Privacy: no raw data
end for
Server:
// Step 1: Topology-Guided Clustering (once, round 0)
if $r = 0$ **then**
  Normalise descriptors: $\hat{\phi}_k \leftarrow \phi_k / \|\phi_k\|_2$ for all k
  $\mathbf{D}_{ij} \leftarrow \|\hat{\phi}_i - \hat{\phi}_j\|_2$ for all i, j
  $\mathcal{C} \leftarrow \text{AgglomerativeClustering}(\mathbf{D}, n_{\text{clusters}})$
end if
Compute anomaly scores: $\delta_k \leftarrow \frac{1}{K-1} \sum_{j \neq k} \|\phi_k - \phi_j\|_2$ // Direction 3
Detect anomalies: $t_k \leftarrow \text{TrustScore}(\{\delta_k\}, \tau)$ // Direction 3
// Step 2: Intra-Cluster Aggregation
for each cluster $C_j \in \mathcal{C}$ **do**
  $\theta_{C_j} \leftarrow \sum_{k \in C_j} w_k \theta_k$, $w_k \propto n_k \exp(-\|\hat{\phi}_k - \hat{\phi}_{C_j}\|) \cdot t_k$ // Eq. 4
end for
// Step 3: Inter-Cluster Blending
$\bar{\theta} \leftarrow \sum_j (|C_j|/K)\, \theta_{C_j}$ // global consensus
$\theta_{C_j}^{(r)} \leftarrow (1-\alpha)\, \theta_{C_j} + \alpha\, \bar{\theta}$ for all j // Eq. 5
Track signatures: $\phi_k^{(r)} \leftarrow \phi_k$ // Direction 4
return $\theta^{(r)}$

B Topological Feature Details

The full 48-dimensional descriptor ϕ_k for each client consists of:

- **Betti curves** $\{b_\ell^0\}_{\ell=1}^{20}$: number of alive H_0 features at 20 linearly-spaced thresholds from 0 to the 95th percentile of H_0 death values
- **Betti curves** $\{b_\ell^1\}_{\ell=1}^{20}$: same for H_1 features
- **Persistence entropy** H_0, H_1: Shannon entropy of normalised persistence values
- **Amplitude** A_0, A_1: L^2 norm of persistence values
- **Persistent feature counts** n_0, n_1: number of features above median persistence
- **Total feature counts**: total H_0 (finite) and H_1 pairs

C Hyperparameters

Table 2: Hyperparameters used in all experiments.

Hyperparameter	Value	Description
n_{sample}	80	TDA subsampling per client
L	20	Betti curve resolution
τ	2.0 (E1), 1.8 (E3)	Anomaly detection threshold
C	1.0	LR regularisation
n_{rounds}	15 (E1/E2), 10 (E3), 20 (E4)	FL rounds
K	8 (E1), 10 (E2), 8 (E3/E5), 6 (E4)	Number of clients