Bayes-pFCL: Bayesian Personalized Federated Continual Learning

Qingyang Yu[1], Yang Hua[2], Hao Wang[3],
Yue Ning[1], Qizhen Zhang[4], Hao Wang[1]

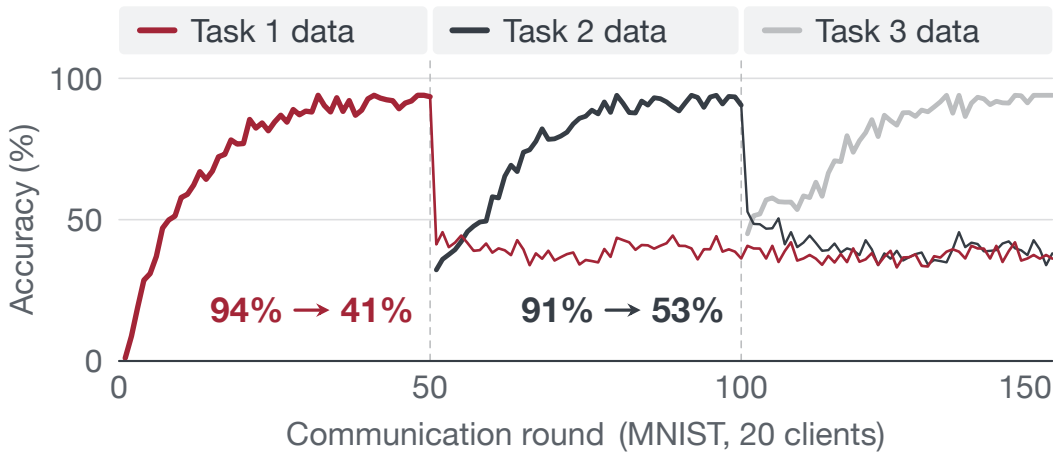
[1] Stevens Institute of Technology [2] Queen's University Belfast
[3] University of Illinois Urbana-Champaign [4] University of Toronto

IntelliSys Lab

Streaming, Distributed Data Degrade Models

91%

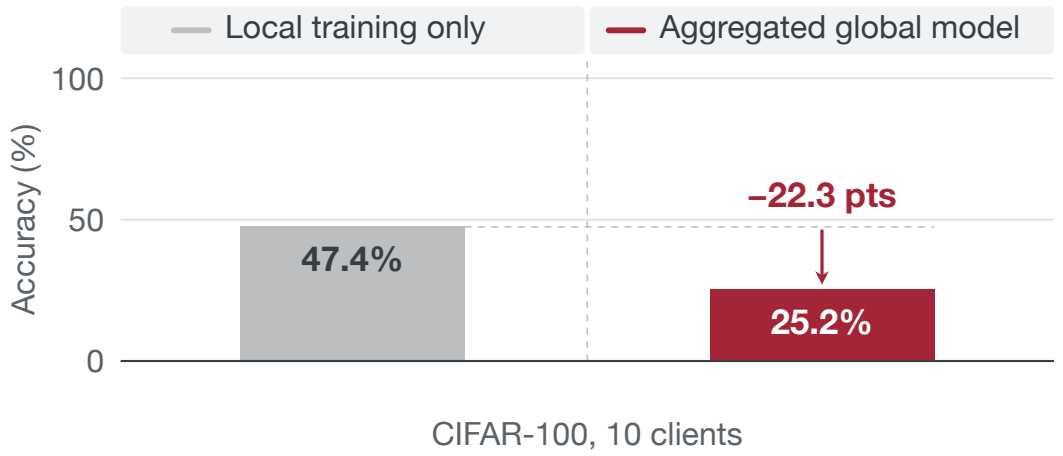
ML models degraded over time



Intra-model interference

67%

Sepsis patients missed at one hospital

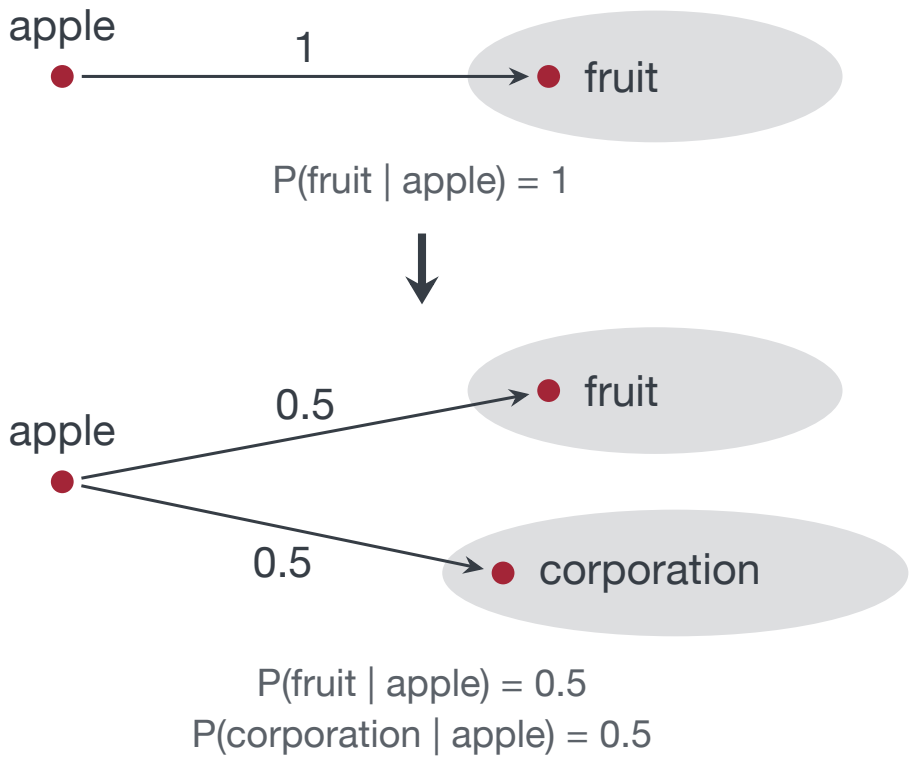


Inter-model interference

pFCL Must Handle Both

[1] Q. Yu et al., "Bayes-pFCL: Bayesian Personalized Federated Continual Learning," *NeurIPS*, 2026.
[2] D. Vela et al., "Temporal quality degradation in AI models," *Scientific Reports*, 2022.
[3] A. Wong et al., "External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients," *JAMA Internal Medicine*, 2021.

The Fan Effect: One Cue, Competing Memories



Knowledge

Probability of retrieving a memory from a cue

Interference

Uncertainty when memories share the same cue

We carry this view into pFCL.

[4] J. R. Anderson, "Retrieval of propositional information from long-term memory," *Cognitive Psychology*, 1974.
[5] J. R. Anderson et al., "An integrated theory of the mind," *Psychological Review*, 2004.

Defining Knowledge and Interference

1 Knowledge

Likelihood a model assigns to data

$$\mathcal{K}(\theta; \mathcal{D}) := p(\mathcal{D} \mid \theta)$$

2 Intra-model interference (CL)

New task data reshapes the model's belief

$$I^{\text{intra}} := I(\theta;\, \mathcal{D}_2 \mid \mathcal{D}_1)$$

3 Inter-model interference (FL)

Statistical dependence between models

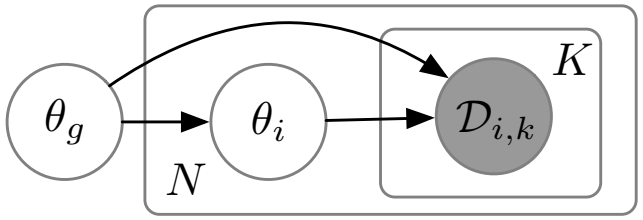
$$I^{\text{inter}} := I(\theta_1;\, \theta_2 \mid \mathcal{D})$$

Both interference terms are conditional mutual information.

[1] Q. Yu et al., "Bayes-pFCL: Bayesian Personalized Federated Continual Learning," *NeurIPS*, 2026.

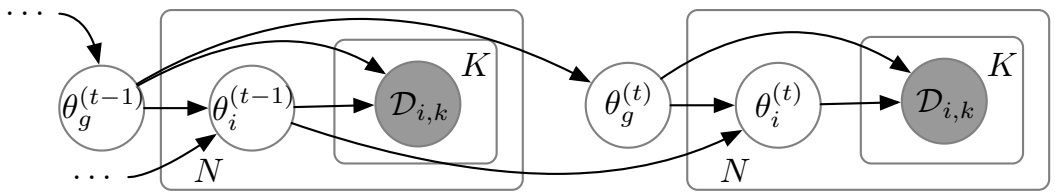
Static and Dynamic Bayesian Views of pFCL

Static view



One joint posterior over the global and personalized models, updated per task

Dynamic view



The posterior evolves at every communication round within a task

Interference for Client i at Task K

$$I_{i,K}^{\text{intra}} := I\left(\theta_i;\, \mathcal{D}_{1:N,K} \mid \mathcal{D}_{1:N,1:K-1}\right)$$

How task K shifts client i's posterior

$$I_{i,K}^{\text{inter}} := I\left(\theta_i;\, \theta_g \mid \theta_{-i},\, \mathcal{D}_{1:N,1:K}\right)$$

How strongly client i couples to the global model

With a Laplace approximation, both terms have computable Gaussian surrogates.

[1] Q. Yu et al., "Bayes-pFCL: Bayesian Personalized Federated Continual Learning," *NeurIPS*, 2026.

Interference Bounds Both Failure Modes

Intra-model interference

Bounds the expected forgetting loss

$$\mathbb{E}[F_{i,K}] \;\le\; \sqrt{\frac{2\,(L^{F}_{i,K})^{2}}{\tau^{F}_{i,K}}\; I^{\text{intra}}}$$

Inter-model interference

Bounds the excess personalized–global loss gap

$$\mathbb{E}[H_{i,K}] \;\le\; \mathbb{E}[B_{i,K}] + \sqrt{2\,\mathbb{E}\left[\sigma^{2}_{H,i,K}(\mathcal{D})\right]\; I^{\text{inter}}}$$

Lower Interference, Tighter Bounds

Assuming Lipschitz losses and posterior concentration

F: old-task loss increase. H: personalized–global loss gap. B: baseline with decoupled posteriors.

[1] Q. Yu et al., “Bayes-pFCL: Bayesian Personalized Federated Continual Learning,” *NeurIPS*, 2026.

Bayes-pFCL: Interference-Regularized Training

Static: refresh per task · Dynamic: refresh every communication round

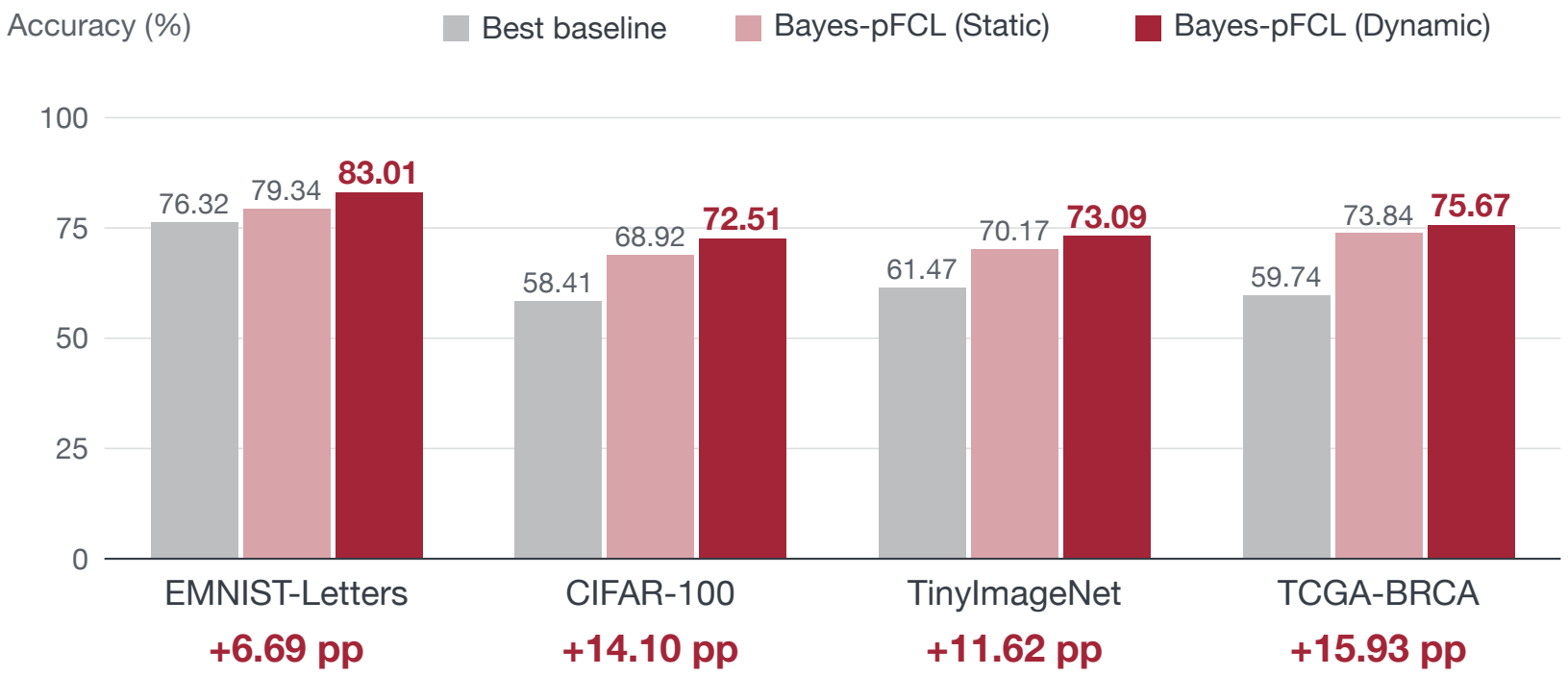
1. Retain the fixed pre-task posterior
2. Broadcast global model + context
3. Update personalized models locally
4. Aggregate models; refresh summaries

$$\mathcal{L}_{i,\text{new}} = \underbrace{\mathcal{L}_{i,\text{old}}(\theta_i; \mathcal{D}_{i,k})}_{\text{Fit the current task}} + \underbrace{\lambda_{\text{intra}}\, I^{\text{intra}}}_{\text{Limit posterior drift}} + \underbrace{\lambda_{\text{inter}}\, I^{\text{inter}}}_{\text{Regulate global–personalized coupling}}$$

Both penalties are computed from Laplace (Gaussian) posterior summaries kept in sync by the server and clients.

[1] Q. Yu et al., "Bayes-pFCL: Bayesian Personalized Federated Continual Learning," *NeurIPS*, 2026.

Higher Accuracy on All Four Benchmarks



75.67%

TCGA-BRCA accuracy

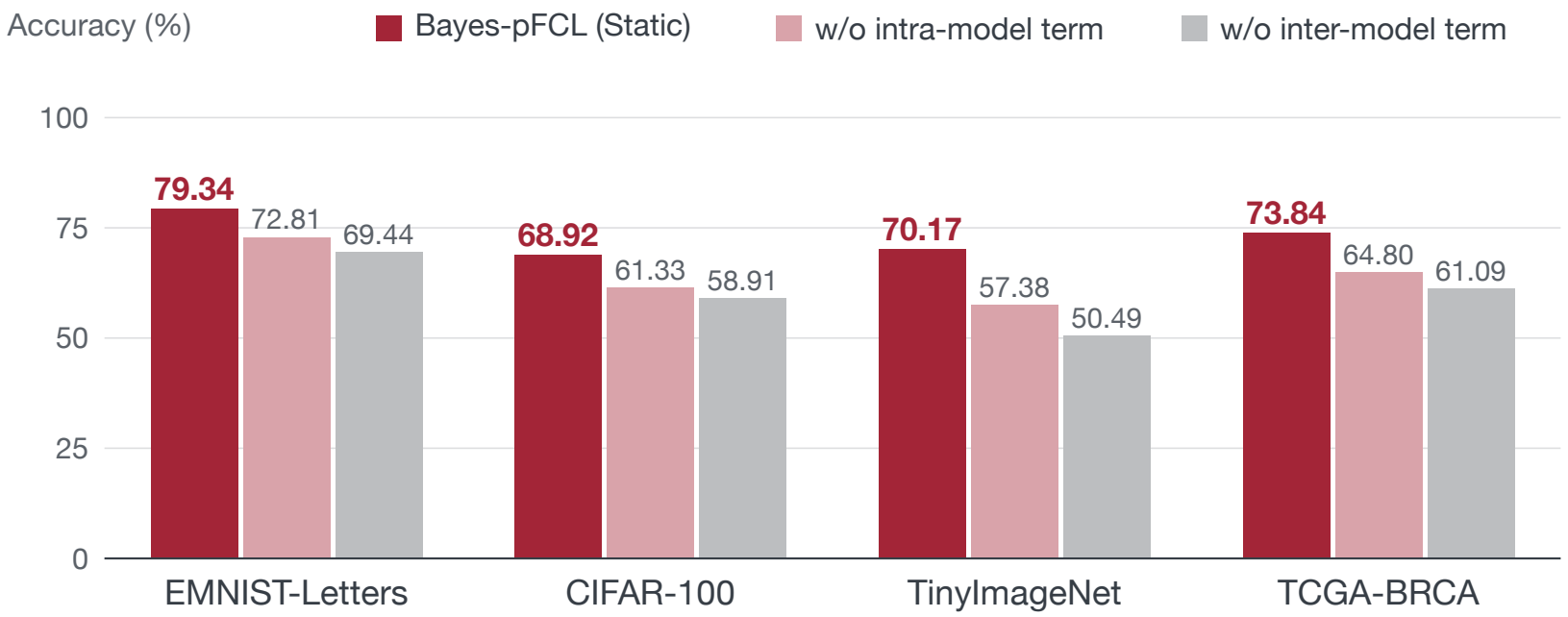
+15.93

percentage points vs. best baseline

Dynamic variant vs. the strongest baseline per dataset: pFedDIL (reproduced) on the image benchmarks, DCFCL on TCGA-BRCA.

[1] Q. Yu et al., "Bayes-pFCL: Bayesian Personalized Federated Continual Learning," *NeurIPS*, 2026.

Both Interference Terms Contribute



−19.68

pp without the inter-model term

−12.79

pp without the intra-model term

Largest drops, both on TinyImageNet. Ablation of Bayes-pFCL (Static): accuracy falls on all four datasets.

[1] Q. Yu et al., “Bayes-pFCL: Bayesian Personalized Federated Continual Learning,” *NeurIPS*, 2026.

Bayes-pFCL

Knowledge as posterior belief

Intra- and inter-model interference quantify forgetting and client heterogeneity in pFCL.

+6.7 to +15.9 pp accuracy

over the strongest baseline on four benchmarks

qyu13@stevens.edu

[1] Q. Yu et al., “Bayes-pFCL: Bayesian Personalized Federated Continual Learning,” *NeurIPS*, 2026.



STEVENS
INSTITUTE OF TECHNOLOGY
1870

Thank You

Homepage
ryougish1k1.github.io

IntelliSys Lab
intellisys.haow.us

Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030