Beyond Aggregation: Guiding Clients in Heterogeneous Federated Learning

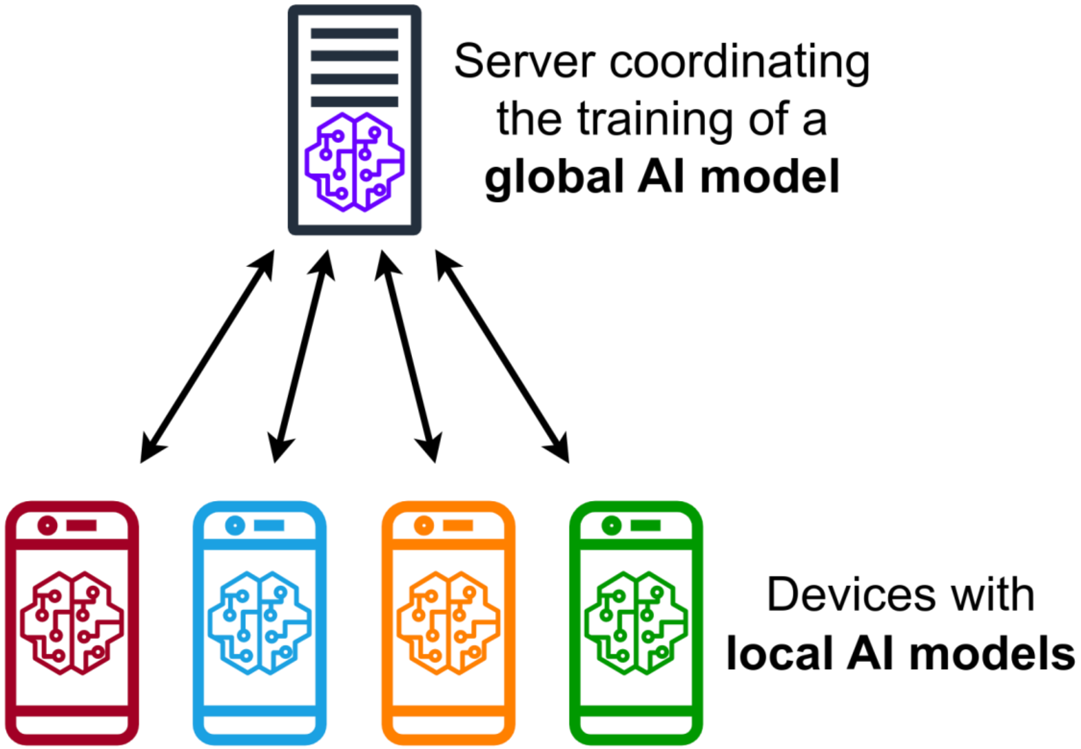
Zijian Wang[1], Xiaofei Zhang[2], Xin Zhang[3], Yukun Liu[4], Qiong Zhang[1,*]

[1]Renmin University of China, [2]Zhongnan University of Economics and Law

[3]Meta, [4]East China Normal University, [*]Corresponding

Background

Federated learning (FL) trains models across distributed clients without sharing raw data.



Challenges

In traditional FL, statistical heterogeneity is often treated as an obstacle.

Paradigm	**Existing Work**	**Local Adaptation**	**System Coordination**
Globalization	FedAvg, FedProx, ...	👎 Weak	👍 Strong
Personalization	FedRep, Ditto, ...	👍 Strong	👎 Weak

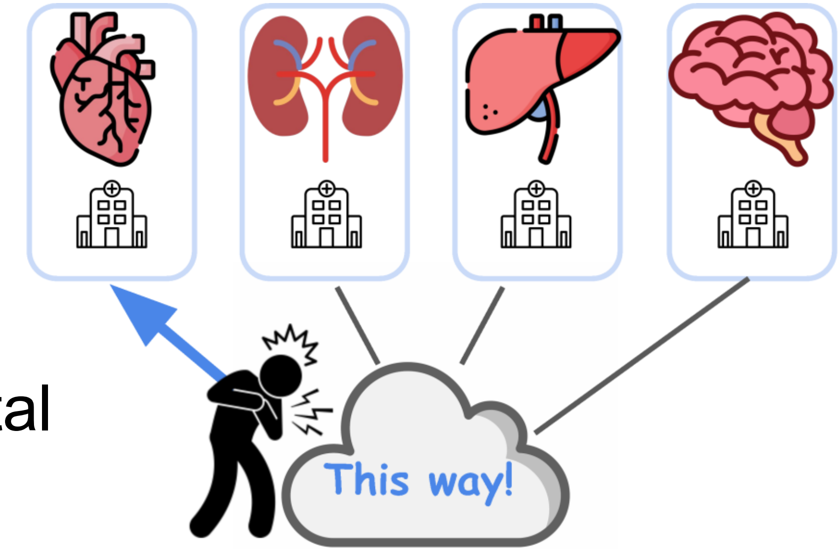
🤔 Can we achieve the best of both worlds?

Motivations

When a new patient arrives:

❌ Aggregate opinions from all hospitals

✅ Route the patient to the most suitable hospital



💡 Client routing enables system-level intelligence.

💡 Statistical heterogeneity can be leveraged as a feature, not just a bug.

Problem Formulation

- Client Routing → Feature Distribution Identification

$$c(x) := \arg\max_i P_X^{(i)}(x)$$

- Sample Classification → Predictive Distribution Estimation

$$\hat{y}_i(x) := \arg\max_k \mathbb{P}^{(i)}(Y = k | X = x)$$

Density Ratio Modeling

- Model client predictive distributions via personalized heads atop a shared backbone.

$$\mathbb{P}^{(i)}(Y = k \mid X = x) = \frac{\exp(\boxed{\alpha_{ik} + \beta_{ik}^\top} g_{\boxed{\theta}}(x))}{\sum_{k'} \exp(\alpha_{ik'} + \beta_{ik'}^\top g_\theta(x))}$$

- Model client feature distributions as exponential tilts of a shared base.

$$\frac{\mathrm{d}P_X^{(i)}}{\boxed{\mathrm{d}P_X^{(0)}}}(x) = \exp\left(\gamma_i + \xi_i^\top h_\tau(g_\theta(x))\right)$$

Empirical Likelihood Inference

❗ Parametric base assumptions (e.g., Gaussian/Gamma) may be misspecified.

✅ Approximate the shared base with an empirical distribution.

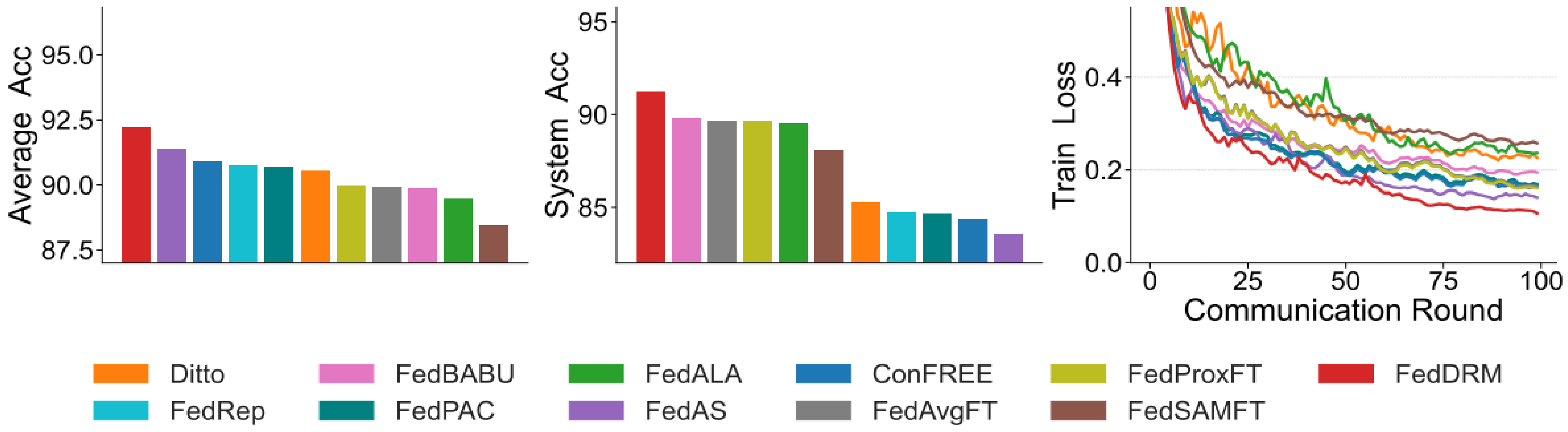
$$\frac{\mathrm{d}P_X^{(i)}}{\boxed{\mathrm{d}P_X^{(0)}}}(x) = \exp\left(\gamma_i + \xi_i^\top h_\tau(g_\theta(x))\right)$$

Maximize empirical likelihood to yield a unified objective of two cross-entropy losses.

$$\ell(\zeta) = -p\ell_N(\zeta) = \sum_{i,j} \ell_{\mathrm{CE}}(i, h_\tau(g_\theta(x_{ij})); \gamma, \xi) + \sum_{i,j} \ell_{\mathrm{CE}}(y_{ij}, g_\theta(x_{ij}); \alpha, \beta)$$

Experiments

New Evaluation Metric: System Accuracy (Server Routing → Local Prediction)



- FedDRM achieves high system accuracy while maintaining superior standard average accuracy.
- FedDRM excels on the real-world medical dataset RETINA, highlighting its potential for clinical deployment.

Conclusion

- Introduce a new FL paradigm for system-level intelligence via client routing.
- Develop a statistically grounded FedDRM framework based on density ratio models and empirical likelihood.
- Provide a practical path toward adaptive, expertise-aware, and resource-efficient real-world FL systems.