Interaction-Aware Gaussian Weighting for Clustered Federated Learning

Alessandro Licciardi*, Davide Leo*, Eros Fanì, Barbara Caputo, Marco Ciccone

Presenting: **Alessandro Licciardi, Davide Leo**

Polytechnic University of Turin (PoliTO), Italy

Correspondence to alessandro.licciardi@polito.it

Forty-Second International Conference of Machine Learning
Vancouver, Canada, July 2025

Finanziato dall'Unione europea NextGenerationEU

Ministero delle Infrastrutture e dei Trasporti

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Introduction

Federated Learning framework and motivation

- Federated Learning (FL) is a distributed machine learning approach that trains models on user data guaranteeing **clients' data privacy** and **minimal communication overhead**

- Clients train a model locally and send updates to a central server, avoiding direct data sharing

Core Challenge: **Data Heterogeneity**

↓

Clients' data is not identically distributed (non-IID) leading to unstable training and poor model performance

Solutions

Regularization Methods | Model Personalization | Biased Client Selection | **Clustering**

Introduction

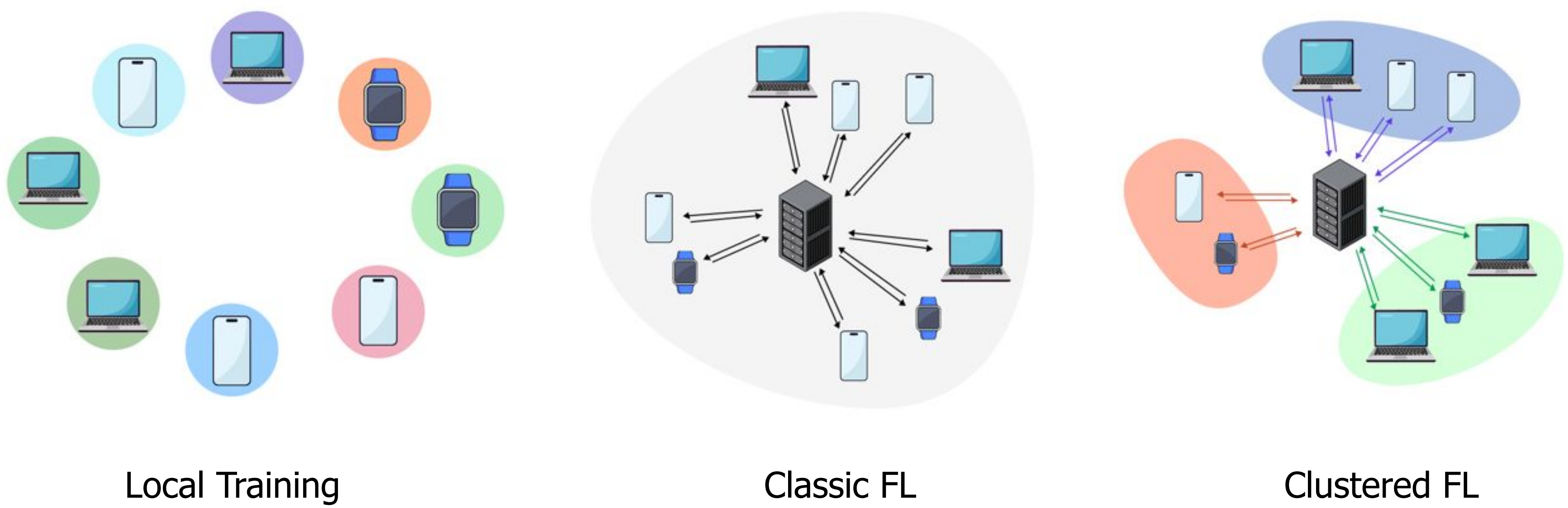
Clustered FL to tackle data heterogeneity

- **Clustered FL:** clients are partitioned into clusters based on the similarity of their data distributions.

  Instead of one global model, **each cluster trains its own model**

  Group-level personalization $\Longrightarrow$ less prone to overfitting than fully personalized FL

- **FedGWC** (Federated Gaussian Weighting Clustering) is a lightweight clustered FL algorithm that iteratively groups clients based on the learning pattern similarities through **local loss processes**



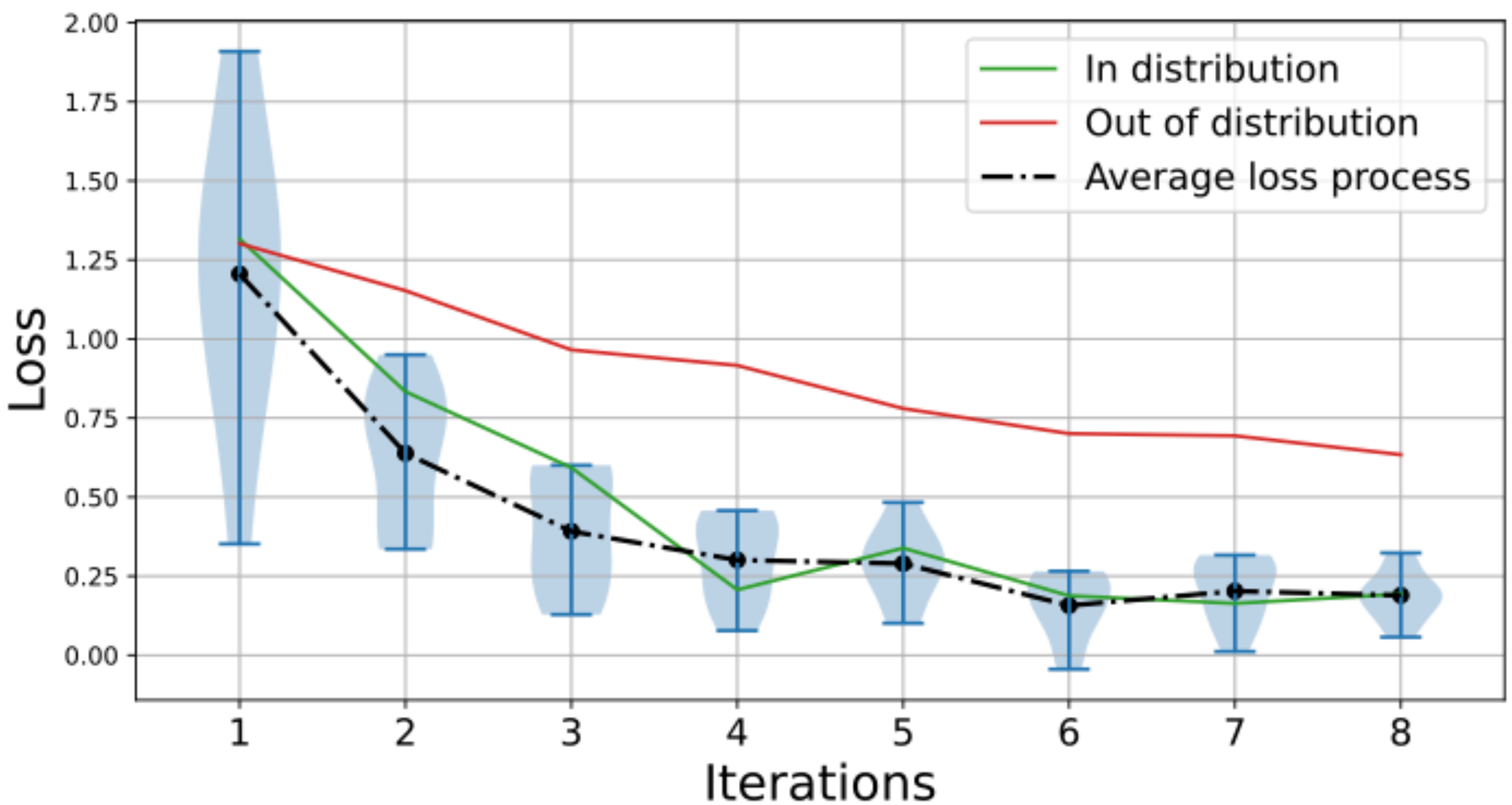
Method

FedGWC Weighting Algorithm

- At each round t the participating clients $\mathscr{P}_t$ **communicate the local loss** $l_k^{t,s} = \mathscr{L}_k(\theta_k^{t,s})$ for client k and updated local model $\theta_k^{t,s}$, where s denotes the local training iteration within the communication round
- The server computes the **Gaussian rewards** $\omega_k^t = 1/S \sum_{s=1}^{S} r_k^{t,s}$ with $r_k^{t,s} = \exp\left(-\frac{(l_k^{t,s} - \mu^{t,s})^2}{(\sigma^{t,s})^2}\right)$ w.r.t. the average loss process $\mu^{t,s}$

High rewards $(\omega_k^t \simeq 1)$ indicate client's loss is close to the average process $\Longrightarrow$ "in-distribution"

Low rewards $(\omega_k^t \simeq 0)$ indicate client's loss is far from the average process $\Longrightarrow$"out-of-distribution"



Method

FedGWC Clustering

Can we infer a complex communication structure from a single scalar value?

1. The server iteratively updates the **interaction matrix** P^t, where $P^t_{k,j}$ estimates how client k is perceived by client j

$$P^{t+1}_{k,j} = \begin{cases} (1-\alpha_t)P^t_{k,j} + \alpha_t \omega^t_k & (k,j) \in \mathcal{P}_t \times \mathcal{P}_t \\ P^t_{k,j} & (k,j) \notin \mathcal{P}_t \times \mathcal{P}_t \end{cases}$$

2. When MSE$(P^{t+1} - P^t) < \epsilon$ the server computes the symmetric affinity matrix W, applying a RBF kernel to the rows of P^{t+1}

3. Spectral clustering is performed on W with different number of clusters $n \in \{2, \ldots, n_{max}\}$ and for each cluster the **Davies-Bouldin score** DB_n is computed

   If $DB_n \geq 1 \, \forall n \in \{2, \ldots, n_{max}\}$ the server does not cluster clients $\Longrightarrow$ **no heterogeneity**

   Otherwise the server splits the clients into n_{cl} clusters, where $n_{cl} \in \arg\min_{n=2,\ldots,n_{max}} DB_n$

4. For each detected cluster, steps (1),(2), and (3) are recursively repeated

A new metric for Clustered FL

Wasserstein Adjusted Score

Problem: Standard clustering metrics are not effective in FL scenarios to evaluate cluster quality when client data is imbalanced

The Wasserstein Adjusted Score (WAS)

- We introduce WAS to quantify cluster cohesion in terms of class distribution. The distance between clients j and k is

$$d(j,k) = \left(\frac{1}{C} \sum_{i=1}^{C} (x_{(i)}^{k} - x_{(i)}^{j})^2 \right)^{1/2}$$

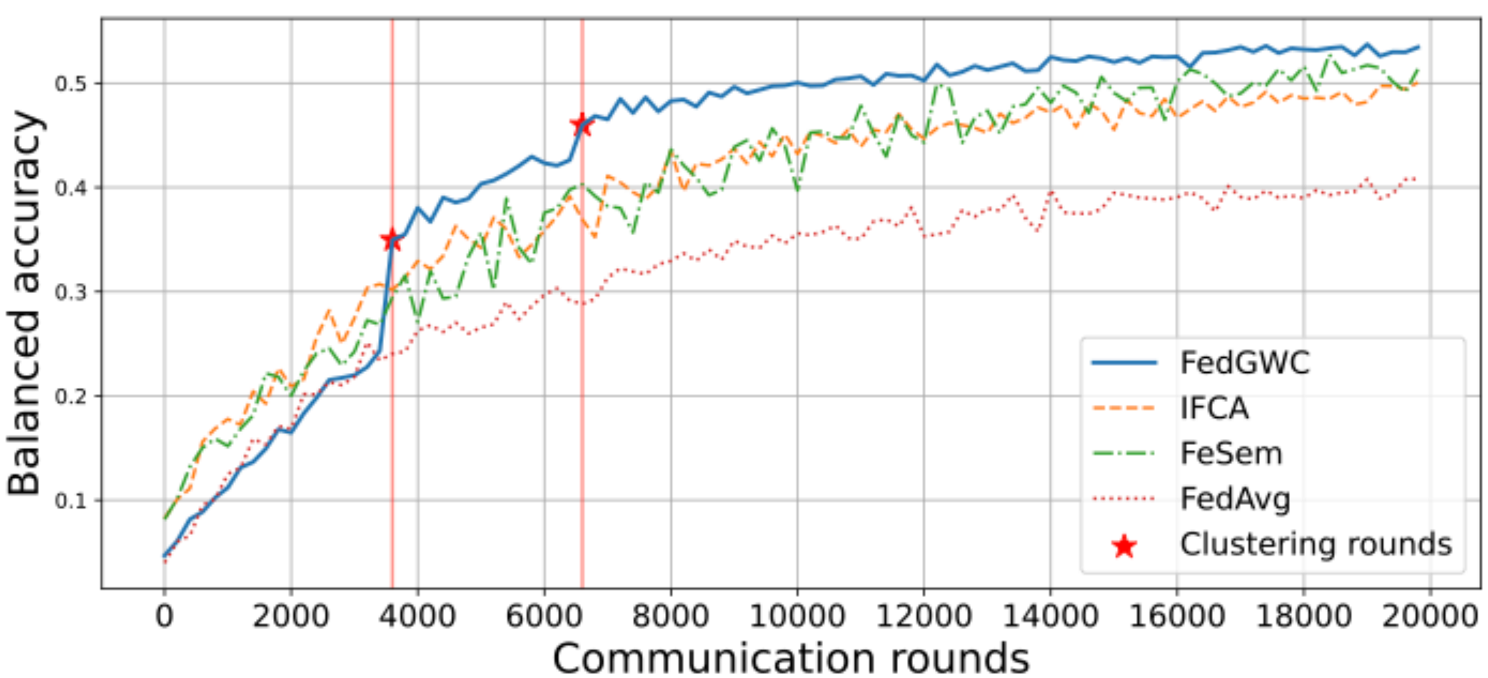
where C denotes the total number of classes, and $x_{(i)}^{k}$ and $x_{(i)}^{j}$ their **ranked class frequencies**

- This metric is theoretically equivalent to computing the **Wasserstein distance** between clients data empirical class distributions (Theorem B.3)

Experimental Results

FedGWC vs. Baselines

- On heterogeneous Cifar100 with 100 clients FedGWC outperforms clustering baselines; a significant increase in accuracy is observed when clusters are detected



		FL method	C	Automatic Cluster Selection	Acc	WAS ↑	WADB ↓
Cifar100	Clustered FL	IFCA	5	✗	47.5 ± 3.5	-0.8 ± 0.2	5.2 ± 5.1
		FeSem	5	✗	53.4 ± 1.8	-0.3 ± 0.1	38.4 ± 13.0
		CFL	1	✓	41.6 ± 1.3	/	/
		FedGWC	4	✓	**53.4** ± 0.4	**0.1** ± 0.0	**2.4** ± 0.4
	Classic FL	FedAvg	1	/	41.6± 1.3	/	/
		FedAvgM	1	/	41.5± 0.5	/	/
		FedProx	1	/	41.8± 1.0	/	/
Femnist	Clustered FL	IFCA	5	✗	76.7 ± 0.6	**0.3** ± 0.1	**0.5** ± 0.1
		FeSem	2	✗	75.6 ± 0.2	0.0 ± 0.0	25.6 ± 7.8
		CFL	1	✓	76.0 ± 0.1	/	/
		FedGWC	4	✓	76.1 ± 0.1	-0.2 ± 0.1	18.0 ± 6.2
	Classic FL	FedAvg	1	/	76.6± 0.1	/	/
		FedAvgM	1	/	**83.3**± 0.3	/	/
		FedProx	1	/	75.9± 0.2	/	/

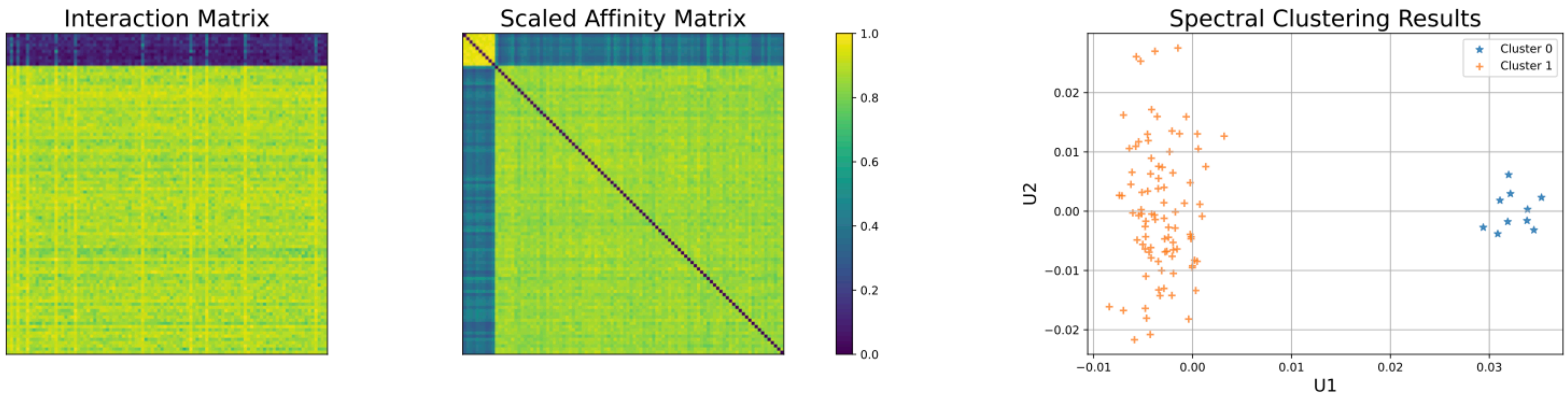
- FedGWC provides better performance also on large scale datasets (Google Landmarks Users-160K with around 800 clients and iNaturalist-Users-120k with round 2700 clients)

Dataset	FedGWC	CFL	IFCA	FedAvg	FedAvgM	FedProx	FairAvg
Google Landmarks	**57.4** ±0.3	40.5 ±0.2	49.4 ±0.3	40.5 ±0.2	36.4 ±1.3	40.2 ±0.6	39.0 ±0.3
iNaturalist	**47.8** ±0.2	45.3 ±0.1	45.8 ±0.6	45.3 ±0.1	37.7 ±1.4	44.9 ±0.2	45.1 ±0.2

Experimental Results

Cluster Analysis

- FedGWC successfully detects different heterogeneity levels, separating homogeneous clients (Cifar10 with $\alpha = 100$ labeled as Cluster 1) from heterogeneous clients (Cifar10 with $\alpha = 0.05$ labeled as Cluster 0)



- FedGWC successfully separates clients according to different visual domains (e.g. clients with blurred noisy data)

Dataset	(Clean, Noise, Blur)	Clustering method	C	Automatic Cluster Selection	Rand ↑ (max = 1.0)
Cifar100	(50, 50, 0)	IFCA	1	✗	0.5 ± 0.0
		FeSem	2	✗	0.49 ± 0.2
		FedGWC	2	✓	**1.0 ± 0.0**
	(50, 0, 50)	IFCA	1	✗	0.5 ± 0.0
		FeSem	2	✗	0.51 ± 0.1
		FedGWC	2	✓	**1.0 ± 0.0**
	(40, 30, 30)	IFCA	1	✗	0.33 ± 0.0
		FeSem	3	✗	0.55 ± 0.0
		FedGWC	4	✓	**0.6 ± 0.0**

Thank you for the kind attention!

Join us at the poster session on Wed. July 16th 11.00 AM- 1.30 PM (PDT)

Vancouver Convention Center, Vancouver, BC, Canada

Alessandro Licciardi **alessandro.licciardi@polito.it**

Davide Leo **davide.leo2000@gmail.com**

Polytechnic University of Turin (PoliTO), Italy

Interaction-Aware Gaussian Weighting for Clustered Federated Learning

A. Licciardi*, D. Leo*, E.Fanì, B. Caputo, M. Ciccone