Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

FAIR Future Artificial Intelligence Research

Politecnico di Torino

spoke 7

# Accelerating Heterogeneous Federated Learning with Closed-form Classifiers


Raffaello Camoriano

*Politecnico di Torino*

*Joint work with:*

Eros Fanì

Barbara Caputo

Marco Ciccone


FAIR General Conference, Naples, Italy

September 24th, 2024

GENERAL CONFERENCE

FAIR Future Artificial Intelligence Research

23|24SET2024

NAPOLI

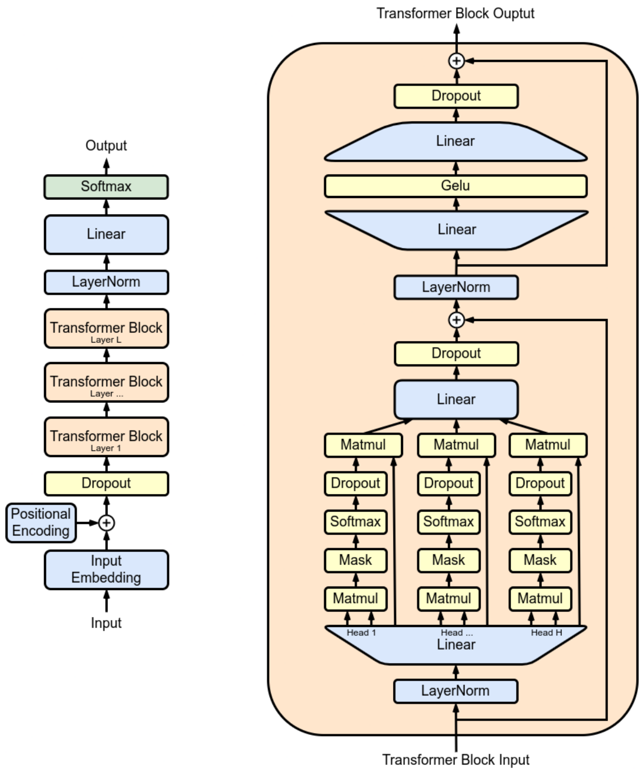
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

FAIR
Future Artificial Intelligence Research

# Leveraging large-scale data on edge devices could improve learning performance




Data usage restricted by:

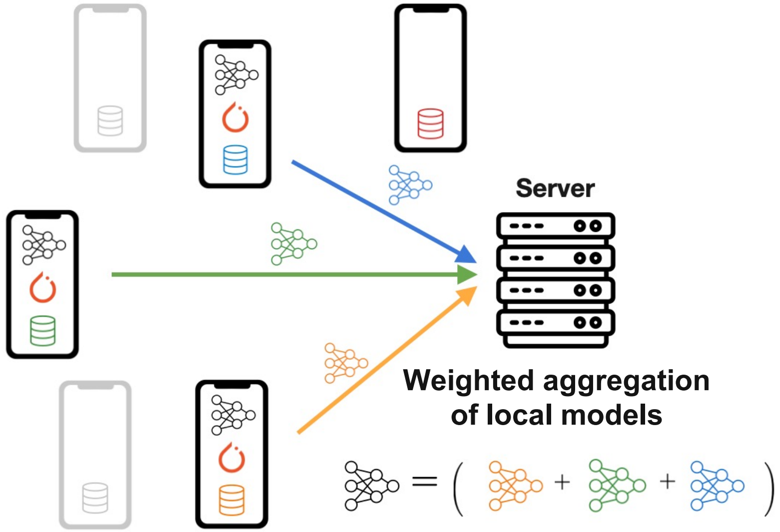
- Privacy requirements
- Communication costs
- Computational costs

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
Future Artificial Intelligence Research

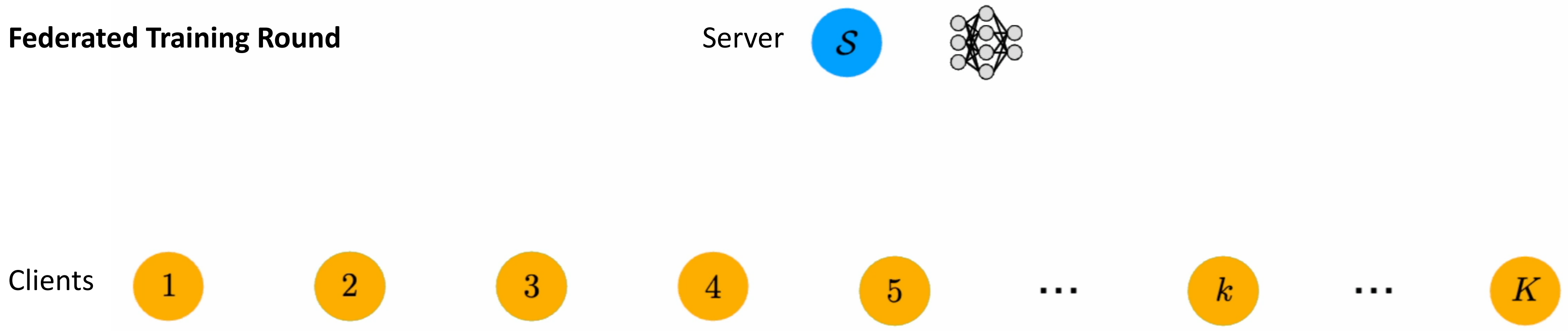
# Federated Learning

- **Population of users join a "federation" to train a model collaboratively**

  🔒 Data never leaves the device (privacy preserving)
  🏠 Training on data is local (distributed computation)
  📡 Communicate model updates
  💪 Powerful, but many challenges

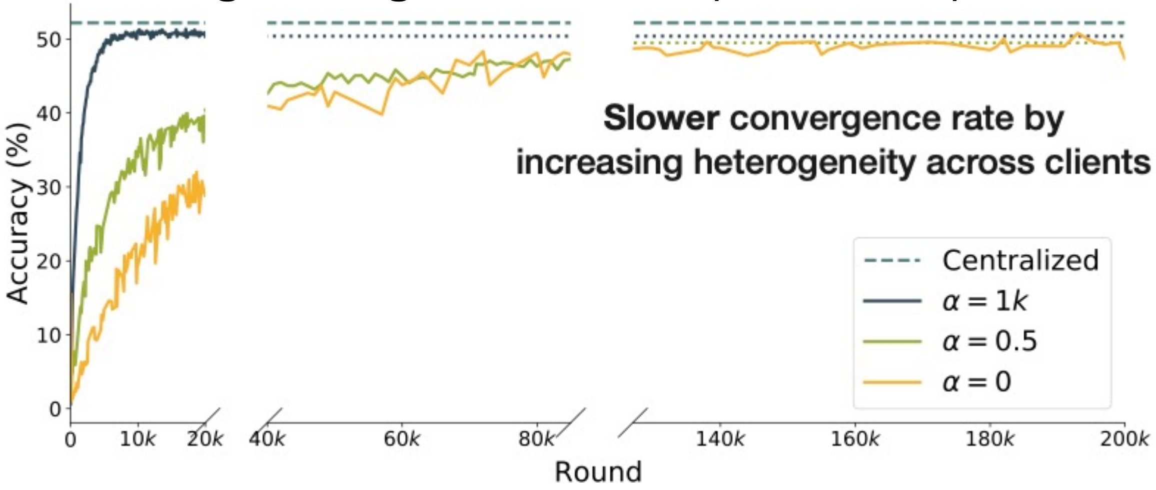



- **Allows training large-scale models while distributing computations**

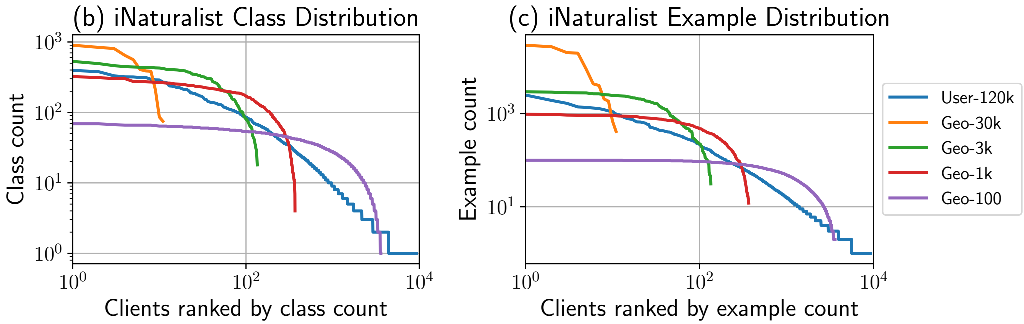


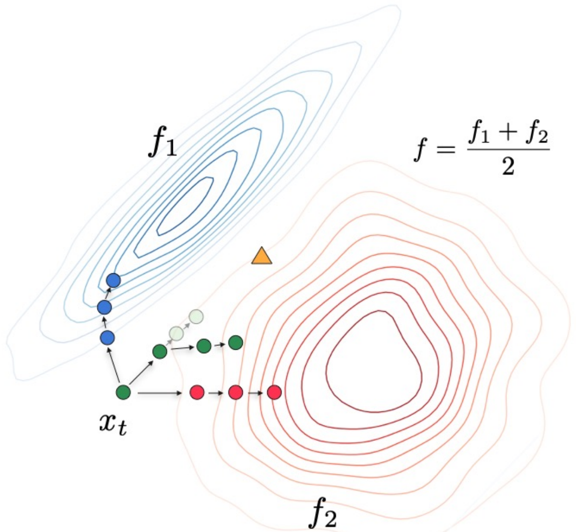
# Challenge of Statistical Heterogeneity

- Samples collected on the edge correlate to specific **user habits**, **preferences**, **location** (**Domain** Shift, **Label** Skewness, **Size** Imbalance)
- **Clients drift** from global solution by specializing on local data
- Training convergence is severely affected by client drift




Caldarola, Caputo, Ciccone (FedSAM, ECCV 2022)

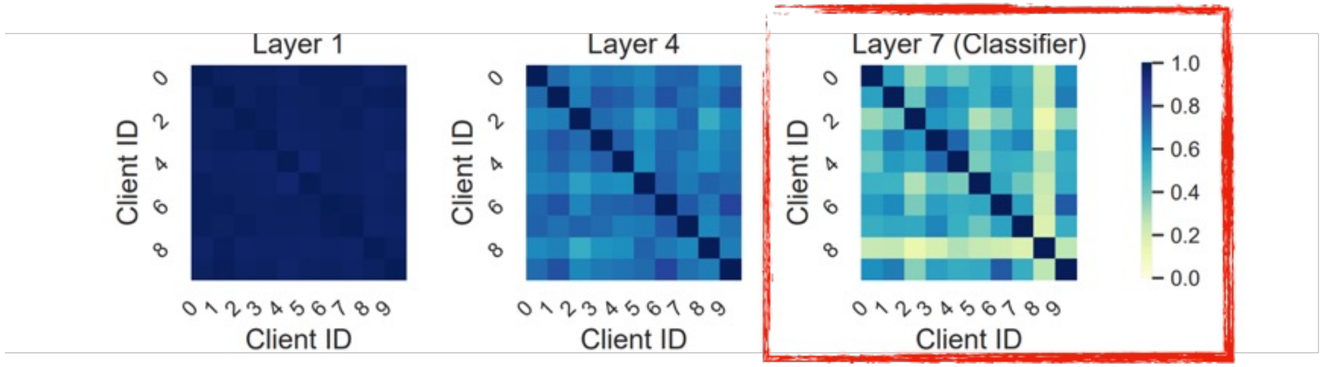



**Client Drift Problem**

# Which part of the model is most affected by heterogeneity?

- Recent work shows that **deeper layers are more subject to client drift** [1]
- Shown by CKA similarities of different layers in clients' local model pairs




- Heterogeneity and partial participation **mainly affect the prediction head** (classifier)

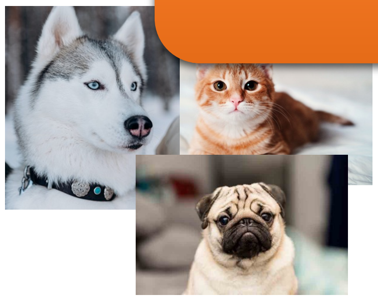
[1] Mi Luo, et al. "No Fear of Heterogeneity: Classifier Calibration for Federated Learning with Non-IID Data". 2021

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
Future Artificial Intelligence Research

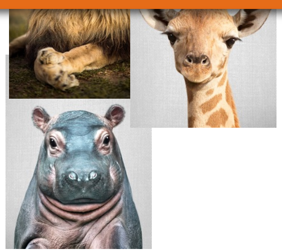
# Heterogeneity and issues with softmax classifier

- **The Softmax classifier** is sensitive to the direction of gradient updates
  - When **new classes** are added, **probabilities of previously learned classes may diminish**
  - This occurs even if logits of previous classes remain unchanged (because of softmax normalisation)
  - **Classifier is biased**
- Similar effect in Continual

Can we design an efficient FL method
<u>robust to client drift</u> in heterogeneous settings
and <u>unaffected by classifier bias</u>?



**Round 1**



**Round 2**

**...**

**Round T**

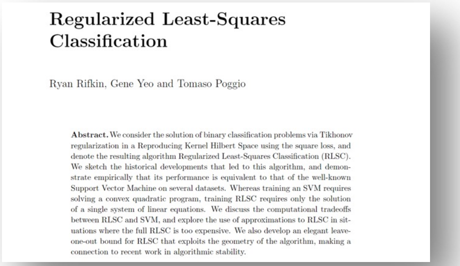
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

FAIR
Future Artificial Intelligence Research

# Background: Ridge Regression for Classification


Regularized Least-Squares Classification

Ryan Rifkin, Gene Yeo and Tomaso Poggio

**Abstract.** We consider the solution of binary classification problems via Tikhonov regularization in a Reproducing Kernel Hilbert Space using the square loss, and denote the resulting algorithm Regularized Least-Squares Classification (RLSC). We sketch the historical developments that led to this algorithm, and demonstrate empirically that its performance is equivalent to that of the well-known Support Vector Machine on several datasets. Whereas training an SVM requires solving a convex quadratic program, training RLSC requires only the solution of a single system of linear equations. We discuss the computational tradeoffs between RLSC and SVM, and explore the use of approximations to RLSC in situations where the full RLSC is too expensive. We also develop an elegant leave-one-out bound for RLSC that exploits the geometry of the algorithm, making a connection to recent work in algorithmic stability.

- We can use a 1-vs-all **Ridge Regression classifier** [2]: $f(x) = \arg\max_{y} \; w_y^\top \psi(x),$

$$W^* = [w_1 \dots w_K] = \arg\min_{W} \frac{1}{|D|} \sum_{(x,y)\in D} \left\| W^\top \psi(x) - \text{OneHot}(y) \right\|^2 + \lambda \left\| W \right\|^2$$

- Ridge Regression admits an **exact incremental formulation** [3,4]

### Closed-form solution

$$W^* = (\underbrace{X^\top X}_{A} + \lambda I)^{-1} \underbrace{X^\top Y}_{b}$$

$$X = [\psi(x_1) \dots \psi(x_N)]^\top$$
$$Y = [\text{OneHot}(y_1) \dots \text{OneHot}(y_N)]^\top$$

**Exact Model Updates**

### Online Update

$$w_z = (A^{\text{new}} + \lambda I)^{-1} b_z^{\text{new}} \quad \forall z \in \mathcal{Y}$$

$$A^{\text{new}} = A^{\text{old}} + \psi(x)\psi(x)^\top$$

$$b_z^{new} = \begin{cases} b_z^{\text{old}} + \psi(x) & \text{if } z = y \\ b_z^{\text{old}} & \text{otherwise.} \end{cases}$$

[2] Rifkin, R., Gene Yeo, and Tomaso Poggio. "Regularized least-squares classification." NATO Science Series Sub Series III Computer and Systems Sciences. 2003.
[3] Camoriano, R.*, Pasquale, G.*, et al. "Incremental robot learning of new objects with fixed update time." *IEEE ICRA*. 2017.
[4] Wang, R., Ciccone, M.*, et al. "Schedule-robust online continual learning." *arXiv preprint arXiv:2210.05561*. 2022.

# Fed3R: Federated Recursive Ridge Regression

**Client side**

Compute Local Ridge Statistics

$$\sum_{(x,y)\in\mathcal{D}_k} \varphi(x)\varphi(x)^\top = A_k$$

$$\sum_{(x,y)\in\mathcal{D}_k} \varphi(x)e_y^\top = b_k$$

$e_y \in \mathbb{R}^C$

**Server side**

Aggregate Statistics

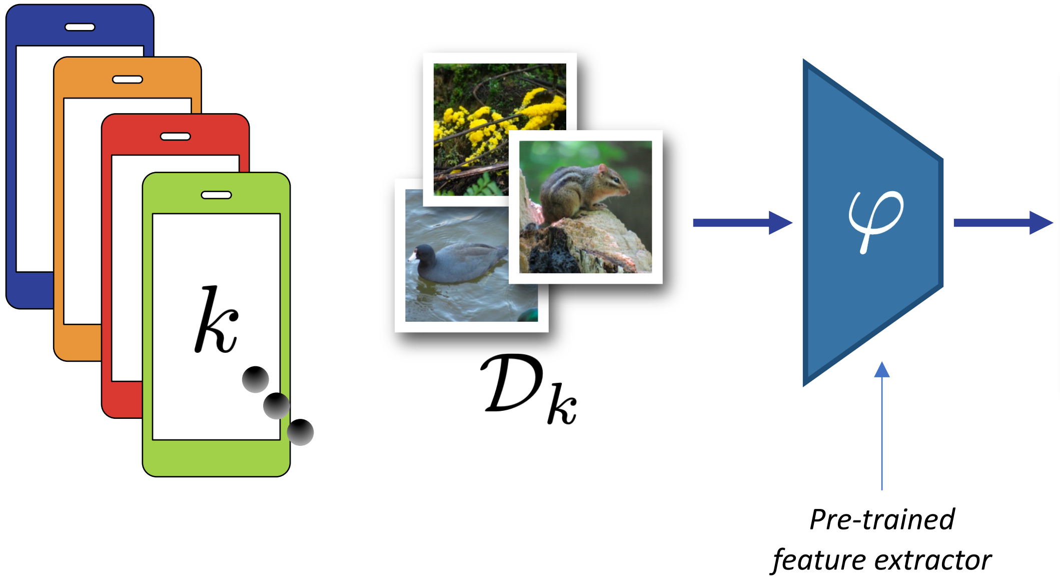
$$A = \sum_{k\in\mathcal{K}} A_k \qquad b = \sum_{k\in\mathcal{K}} b_k$$

$$W^* = (A + \lambda I_d)^{-1} b$$

- RR incremental formulation → **Allows for <u>exact aggregation in Fed3R</u>**
  - Immune to statistical heterogeneity
  - Faster convergence
  - Reduced computations and communication
- Fed3R-RF: non-linear variant based on random features for improved accuracy

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
Future Artificial Intelligence Research

# Fed3R: Step 1 (client side) - Local computations

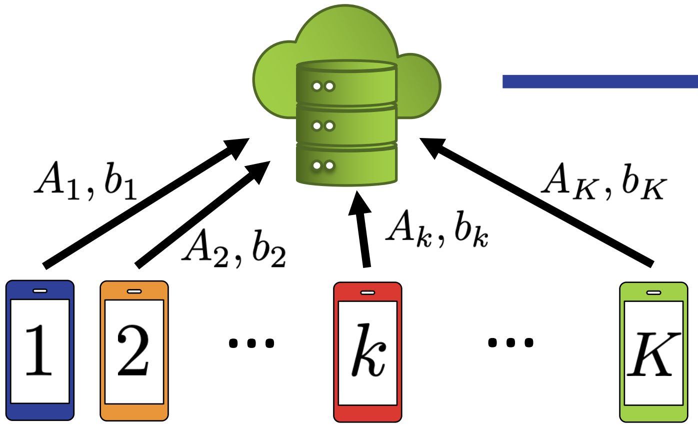



$$\sum_{(x,y)\in\mathcal{D}_k} \varphi(x)\varphi(x)^\top = A_k$$

$$\sum_{(x,y)\in\mathcal{D}_k} \varphi(x)e_y^\top = b_k$$

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Future Artificial Intelligence Research

# Fed3R: Step 2 (server side) - Exact aggregation




Compute the aggregate statistics

$$A = \sum_{(x,y)\in\mathcal{D}} \varphi(x)\varphi(x)^\top = \sum_{k\in\mathcal{K}} \sum_{(x,y)\in\mathcal{D}_k} \varphi(x)\varphi(x)^\top = \sum_{k\in\mathcal{K}} A_k$$

$$b = \sum_{(x,y)\in\mathcal{D}} \varphi(x)e_y^\top = \sum_{k\in\mathcal{K}} \sum_{(x,y)\in\mathcal{D}_k} \varphi(x)e_y^\top = \sum_{k\in\mathcal{K}} b_k$$

*Centralized dataset*

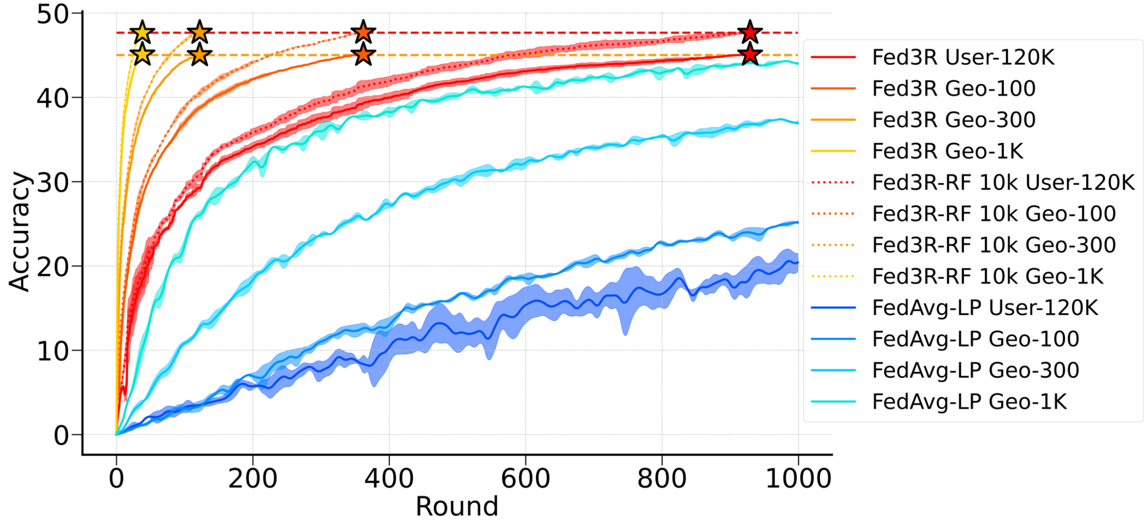
*Local datasets for each client k*
*(Valid for any federated split)*

Closed-form RR solution yields exact aggregate classifier: $W^* = (A + \lambda I_d)^{-1} b$

# Fed3R is immune to statistical heterogeneity

- Immune to heterogeneity: equivalent to exact centralized solution
- Convergence guaranteed in a single pass over clients
- Memory and computationally efficient

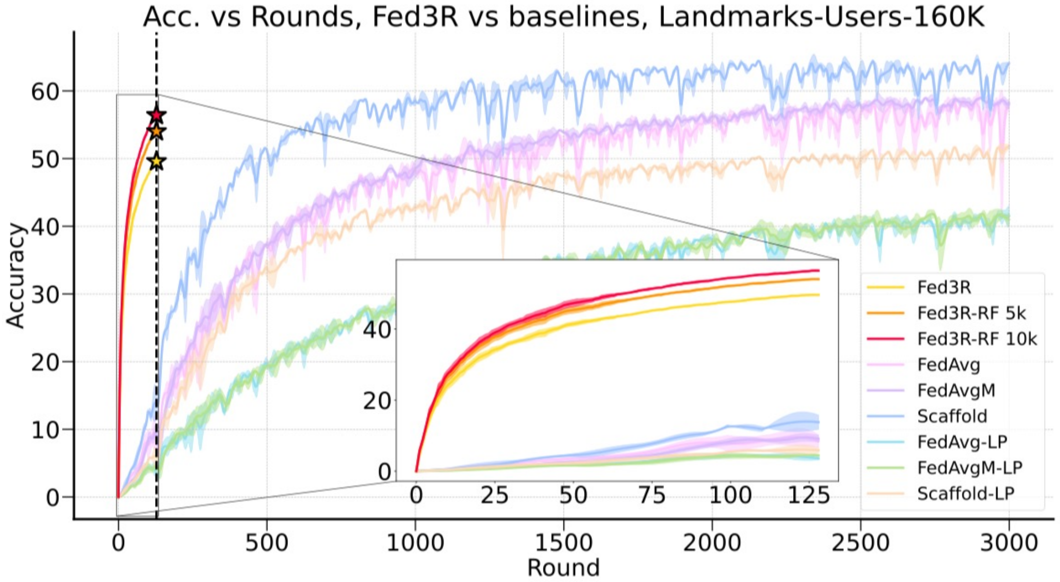
| | User-120K | Geo-100 | Geo-300 | Geo-1K |
|---|---|---|---|---|
| # clients | 9275 | 3606 | 1208 | 368 |
| # classes | 1203 | 1203 | 1203 | 1203 |

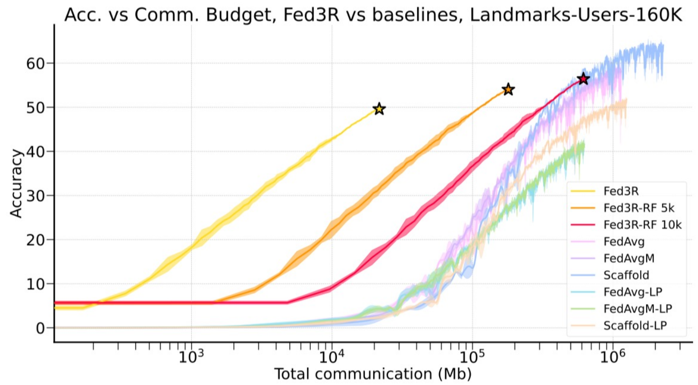


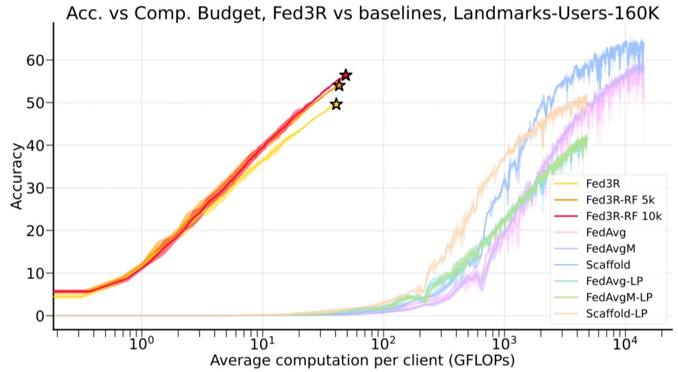

Fed3R performance is invariant to different federated splits of **iNaturalist** (Hsu, 2020). (MobileNetV2 pretrained on ImageNet-1k)

# Experiments with fixed feature extractor Fed3R & Fed3R-RF vs. baselines

Google Landmarks Dataset v2 - 2028 classes – 1262 clients



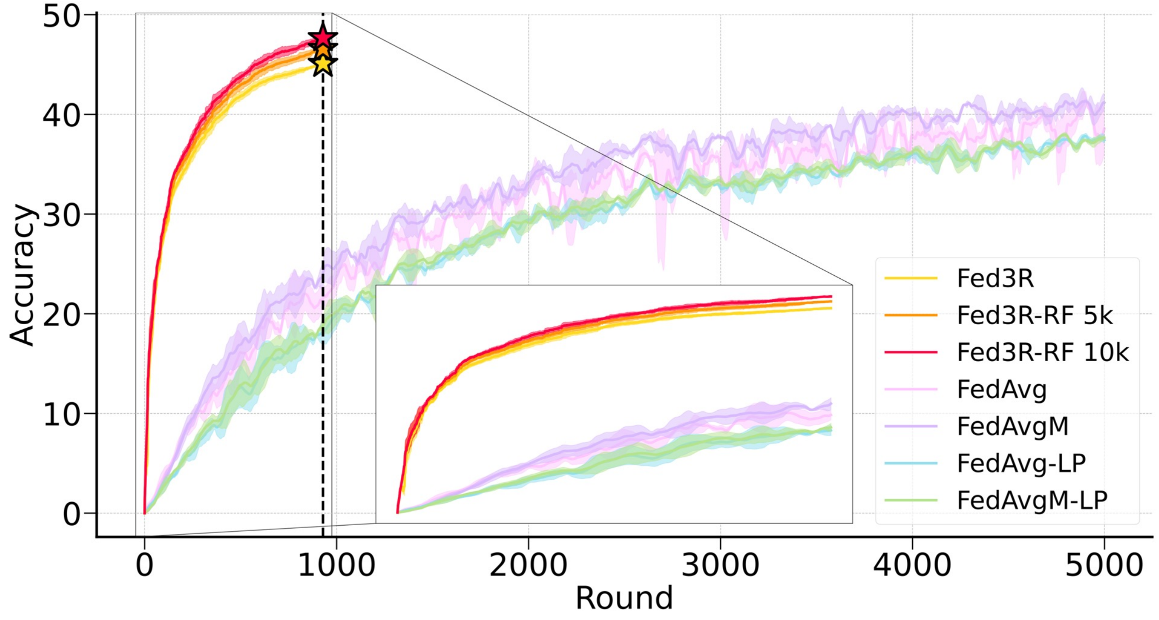



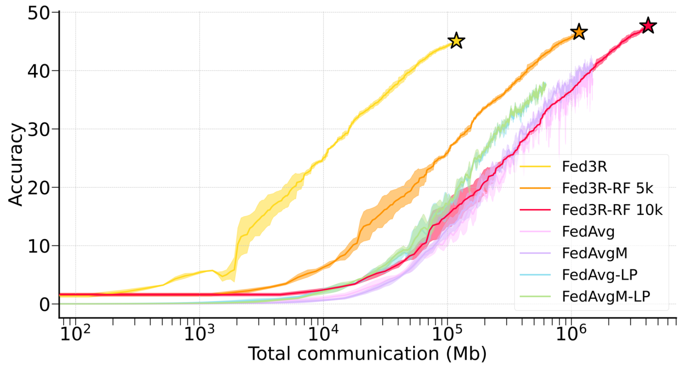



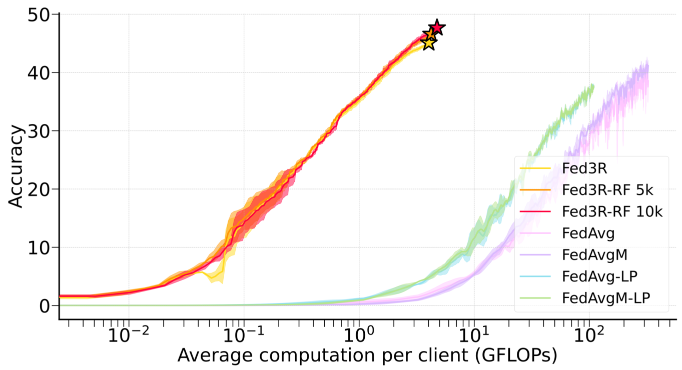
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Future Artificial Intelligence Research

# Experiments with fixed feature extractor
# Fed3R & Fed3R-RF vs. baselines

iNaturalist-Users120K - 1203 classes – 9275 clients



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

# Fed3R-FT: Fine-Tuning variant

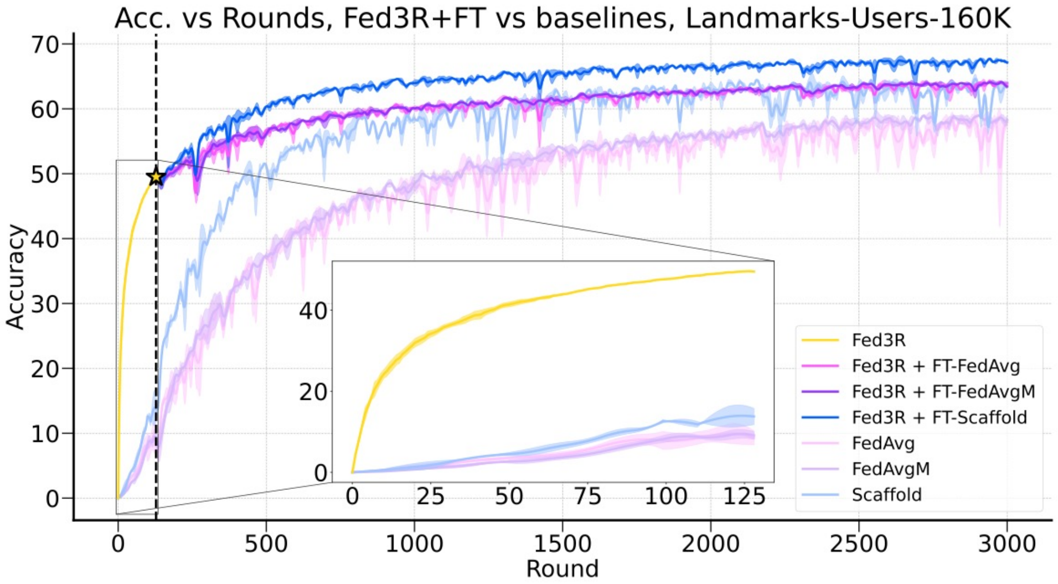
- Fed3R can also be used as robust initialization for *any* FL algorithm

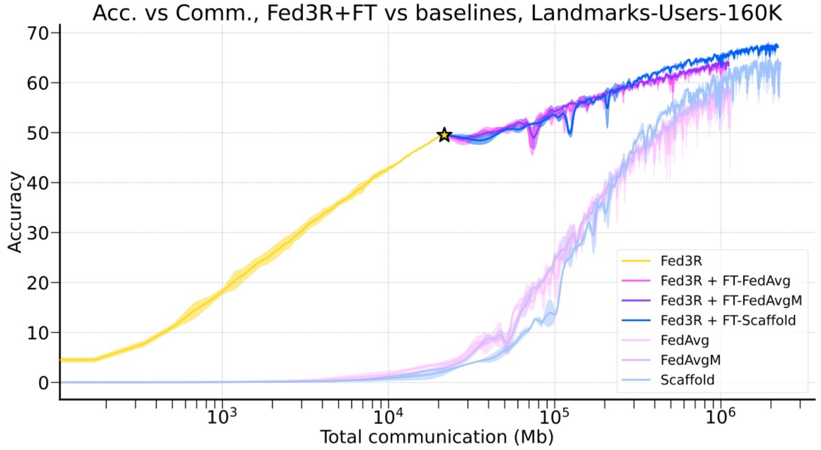
- **Fed3R+FT variant**: last layer initialization for faster fine-tuning
  - Fed3R+FT: fine-tune the whole model
  - Fed3R+FTlp: fine-tune only the classifier
  - Fed3R+FTfeat: fine-tune only the feature extractor

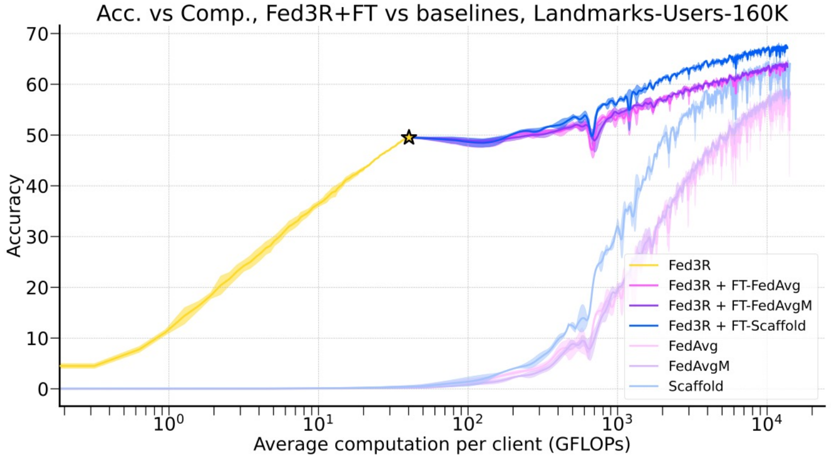
- Learning or fine-tuning in extreme cross-device settings is hard:
  - We show that **fine-tuning further improves final accuracy**
  - Fixing Fed3R and fine-tuning only the representation **stabilizes training**

# Experiments with Fine-Tuning

Gldv2 - 2028 classes – 1262 clients



Finanziato dall'Unione europea
NextGenerationEU

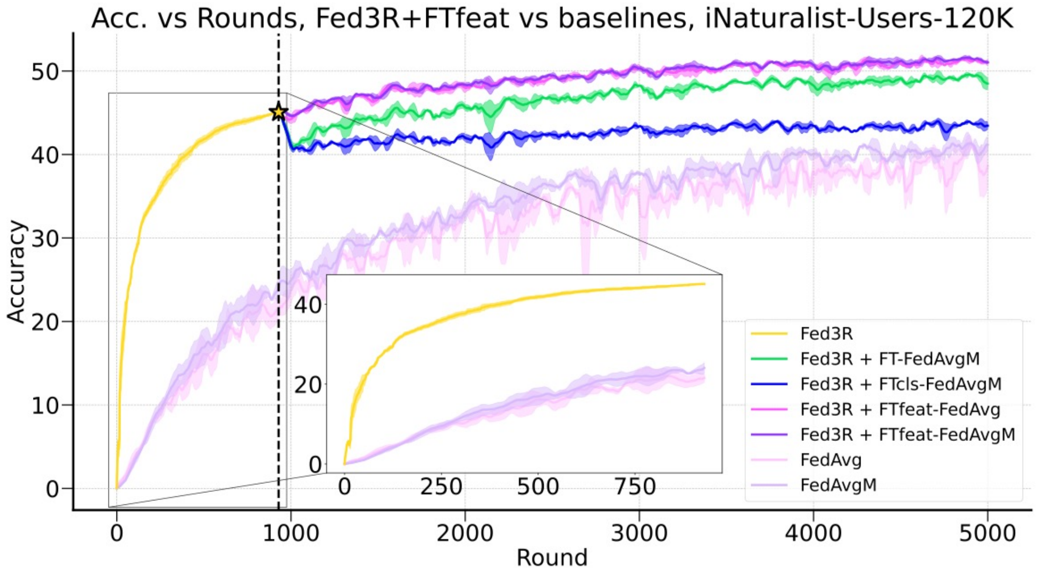
Ministero dell'Università e della Ricerca

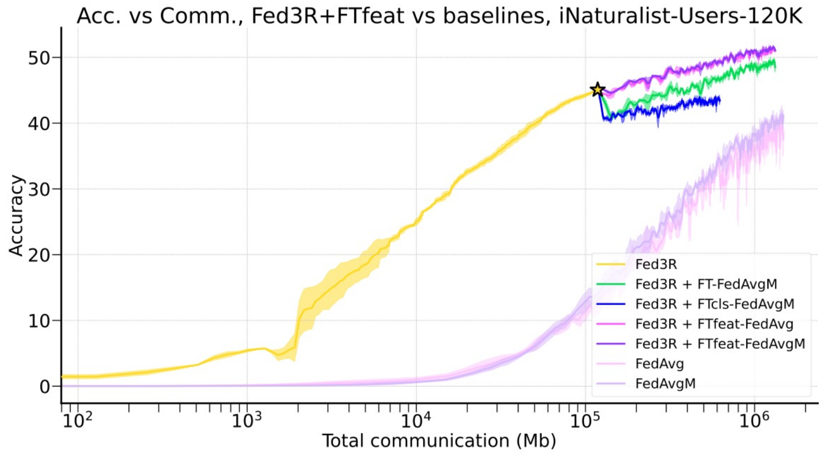
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

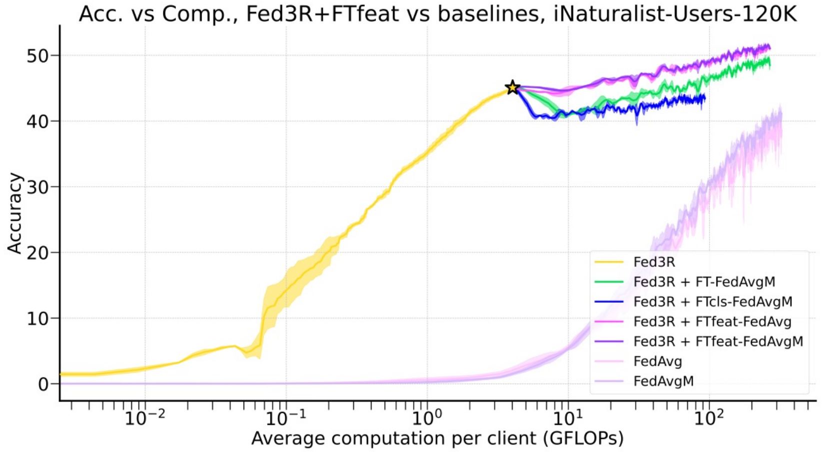
Future Artificial Intelligence Research

# Experiments with Fine-Tuning

iNaturalist - 1203 classes – 9275 clients



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

## Conclusions

- We introduce **Fed3R**, a robust FL algorithm with exact aggregation
  - Faster convergence than softmax-based classifiers
  - Communication and computationally efficient
- Fed3R-RF enables accuracy/communication cost trade-off
- Fed3R-FT yields improved accuracy and more stable training in cross-device settings

## Next steps

- Extension of Fed3R to the Personalized Federated Learning setting:
  - Fed3R model training
  - Additional fine-tuning on individual devices using client data for improved local performance (data stays private)

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

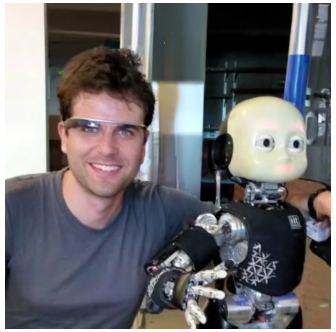
FAIR
Future Artificial Intelligence Research

# Accelerating Heterogeneous Federated Learning with Closed-form Classifiers

Eros Fanì, Raffaello Camoriano, Barbara Caputo, Marco Ciccone

*ICML 2024, Wien, Austria*

*https://fed-3r.github.io/*



SCAN ME

*Contact: eros.fani@polito.it*