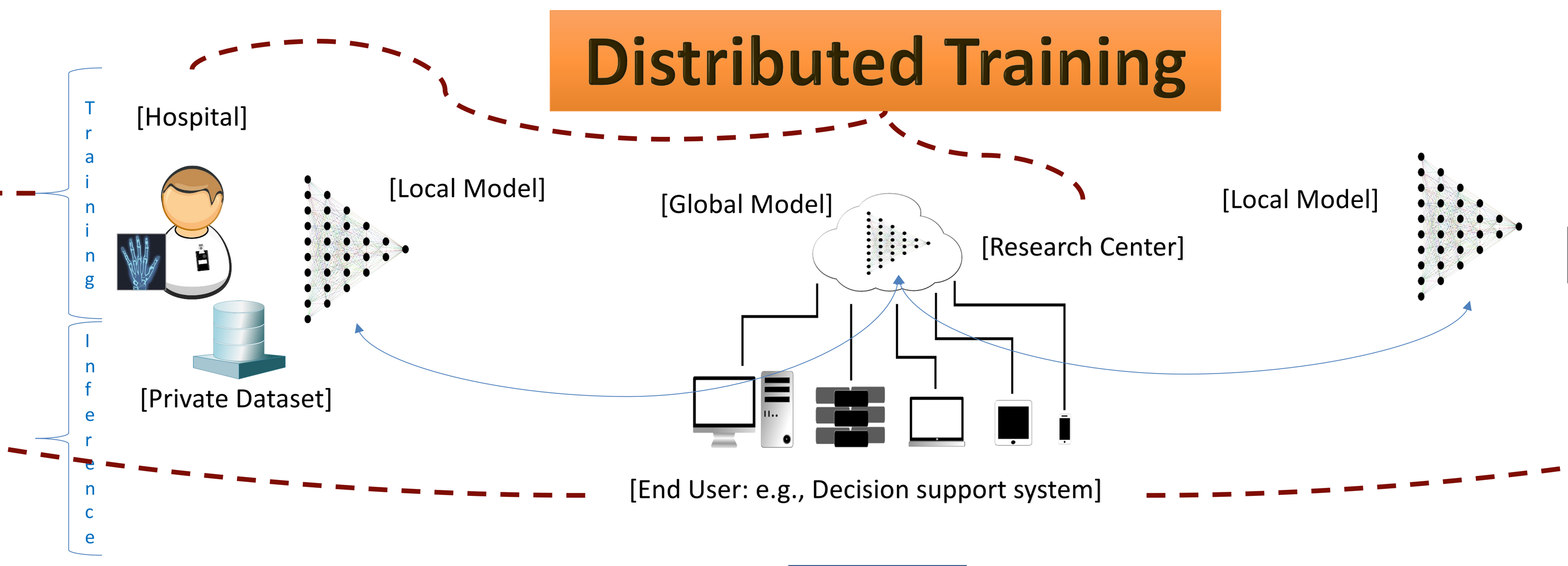
Impact of non-IID data on the performance and fairness of differentially private federated learning

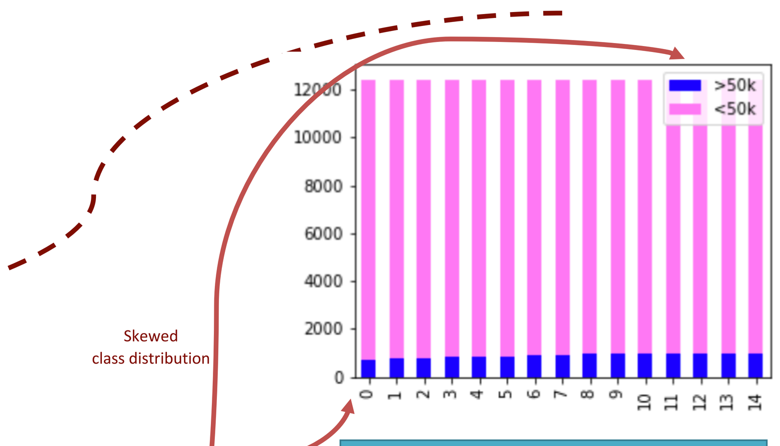
Saba Amiri, Adam Belloum, Eric Nalisnick, Sander Klous, Leon Gommans

University of Amsterdam
Informatics Institute

Issues with Distributed ML in Medical Domain

Privacy | **Distributed Training** | **Data Distribution**





Differential Privacy

Consider adjacent datasets $d, d' \in \mathcal{D}$ which only differ in one element. The randomized mechanism $\mathcal{M} : \mathcal{D} \rightarrow \mathcal{R}$ is (ϵ, δ)-differentially-private if for any subset of outputs of $\mathcal{M}$, $S \subseteq \mathcal{R}$

$$\Pr[\mathcal{M}(d) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(d') \in S] + \delta. \quad (1)$$

where ϵ is the privacy budget, setting the level of intended privacy. The lower the ϵ, the higher the privacy level. δ is a small probability of failure of the DP guarantee. As a rule of thumb, it is set as less than $1/samplesize$.

Differentially private SGD

1. Clip gradients
2. Add calibrated noise

Algorithm 2 Differentially-Private Stochastic Gradient Descent

Input: Training samples X, labels Y, training epochs E, batch size B, loss function $\mathcal{L}$, clipping threshold C, Gaussian noise scale σ, learning rate α , sampling probability p
Output: Model parameters θ

```
1: Initialize θ
2: for epoch e in [1 : E] do
3:    for batch b sampled from [X, Y] with prob(p) do
4:       for each sample b_i in b do
5:          g_i ← ∇_θ L(x_i, y_i, θ)
6:       end for
7:       g_b ← (1/B)(Σ g_i / max(1, ||g_i||_1 / C) + N(0, C²α²I))
8:    end for θ ← θ − α.g_b
9: end for
10: return θ, ε_spent
```

Federated Learning

Algorithm 1 Federated Averaging

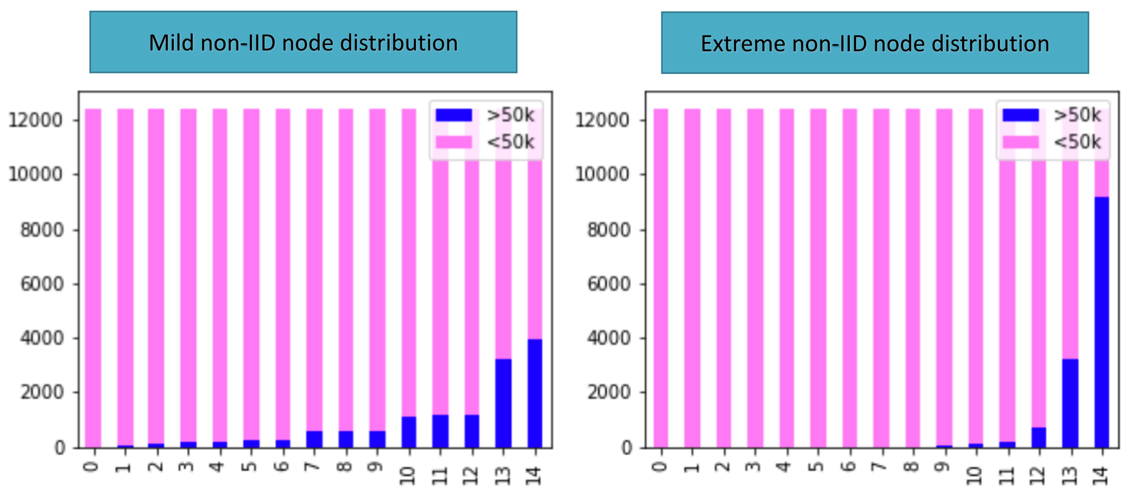
Server side operations for communication round i in $[1...C]$
Input: Updated parameter sets from K participant
Output: Model parameters θ

```
1: if i == 0 then
2:    initialize θ^i_global
3: else
4:    wait to receive k parameters sets {θ^i_1, ..., θ^i_k}
5:    θ^i_global = (1/K) Σ_{k=1}^{K} θ^i_k
6: end if
7: send θ^i_global to K participants
```

Participant side operations for communication round i in $[1...C]$
Input: Local Training samples X_p, labels Y_p, training epochs E, batch size B, loss function $\mathcal{L}$, learning rate α
Output: Model parameters θ

```
1: receive θ^i_global
2: θ^i_k ← θ^i_global
3: for epoch e in [1 : E] do
4:    for batch b in [X_p, Y_p] do
5:       θ^i_k ← θ^i_k − α∇L(b; θ^i_k)
6:    end for
7: end for
8: return θ^i_k
```

IID Assumption



- Real world data distribution is non-IID
  - Class imbalance: imbalance in target feature
  - Feature imbalance: imbalance in non-target feature
  - Node imbalance: imbalance in distribution of samples among nodes

Experimental Setup

Dataset

- Census Adult Income dataset
- ***Income*** as the target feature, ***“>50k”*** as desirable outcome
- ***Race*** as the protected feature, ***“White”*** as privileged group

Utility Metrics

- Precision
- Recall
- F1-Score

Fairness Metrics

- **Differential Fairness**

Let $P \subset \mathbb{R}^k \times \{0,1\}$ be the input space of a binary classifier model. Consider dataset $\mathcal{X}$ with feature set $x : \{x_1, x_2, ..., x_n\}$ and protected features set $A \subset x$ and $\mathbf{s}_i, \mathbf{s}_j \in A$ tuples of protected feature values. Randomized mechanism $\mathrm{M} : \mathcal{X} \rightarrow \mathcal{Y}$ is ϵ-Differentially Fair (DF) with respect to (A, Θ) if for all $(\mathbf{s}_i, \mathbf{s}_j) \in A \times A$ and $\mathbf{x} \sim \theta$:

$$e^{-\epsilon} \leq \frac{P_{M,\theta}(M(\mathbf{x}) = y \mid \mathbf{s}_i, \theta)}{P_{M,\theta}(M(\mathbf{x}) = y \mid \mathbf{s}_j, \theta)} \leq e^{\epsilon},$$

for $\theta \in \Theta$ and $y \in \mathrm{Range}(M)$ where $P(\mathbf{s}_i \mid \theta) > 0, P(\mathbf{s}_j \mid \theta) > 0$ [20].

- **Generalized Entropy Index**

For $\alpha \notin \{0,1\}$ and $b_i = (y_{predict_i} - y_{label_i} + 1)$, with N being the number of individual samples in dataset $\mathcal{X}$ the Generalized Entropy Index with mean $\mu = \frac{1}{N}\sum_N^1 b_i$ is defined as:

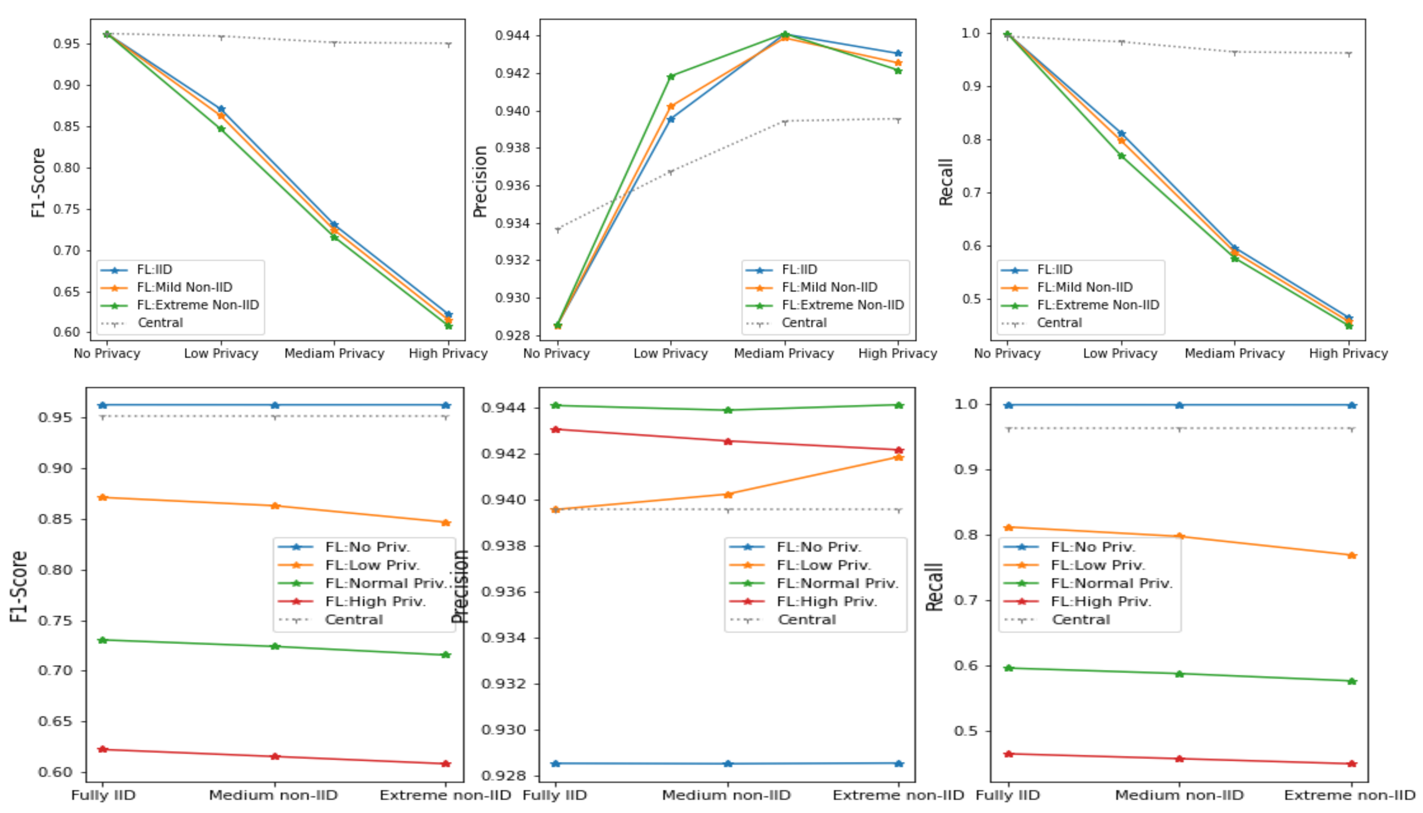
$$\frac{1}{N\alpha(\alpha-1)} \sum_{i=1}^{N} \left[\left(\frac{b_i}{\mu} \right)^{\alpha} - 1 \right]$$

- **Equal Odds Rate**

Formally, mechanism M exhibits absolute equal odds -i.e., is fair - for privileged group G and unprivileged group G' and desired outcome $O \in \{0,1\}$ if $\mathbb{E}_{(\mathbf{x},y)\sim G}[\mathrm{M}(\mathbf{x}) \mid y = O] = \mathbb{E}_{(\mathbf{x},y)\sim G'}[\mathrm{M}(\mathbf{x}) \mid y = O]$

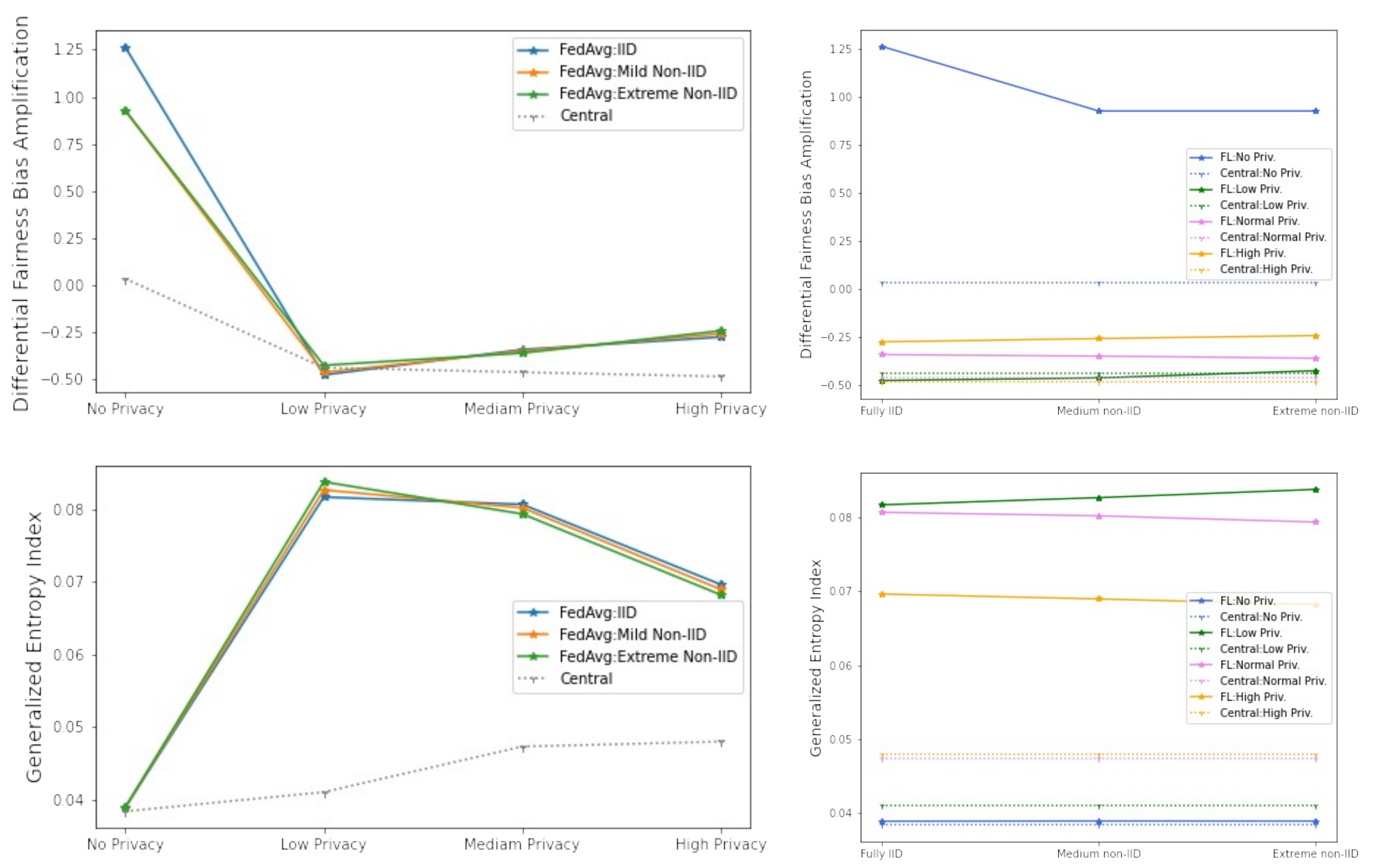
Impact of non-IID Data on Performance

- Performance drops with increase in privacy level
- Recall drops significantly while the difference in precision is prominent but negligible
- High privacy regimes act as a regularization method



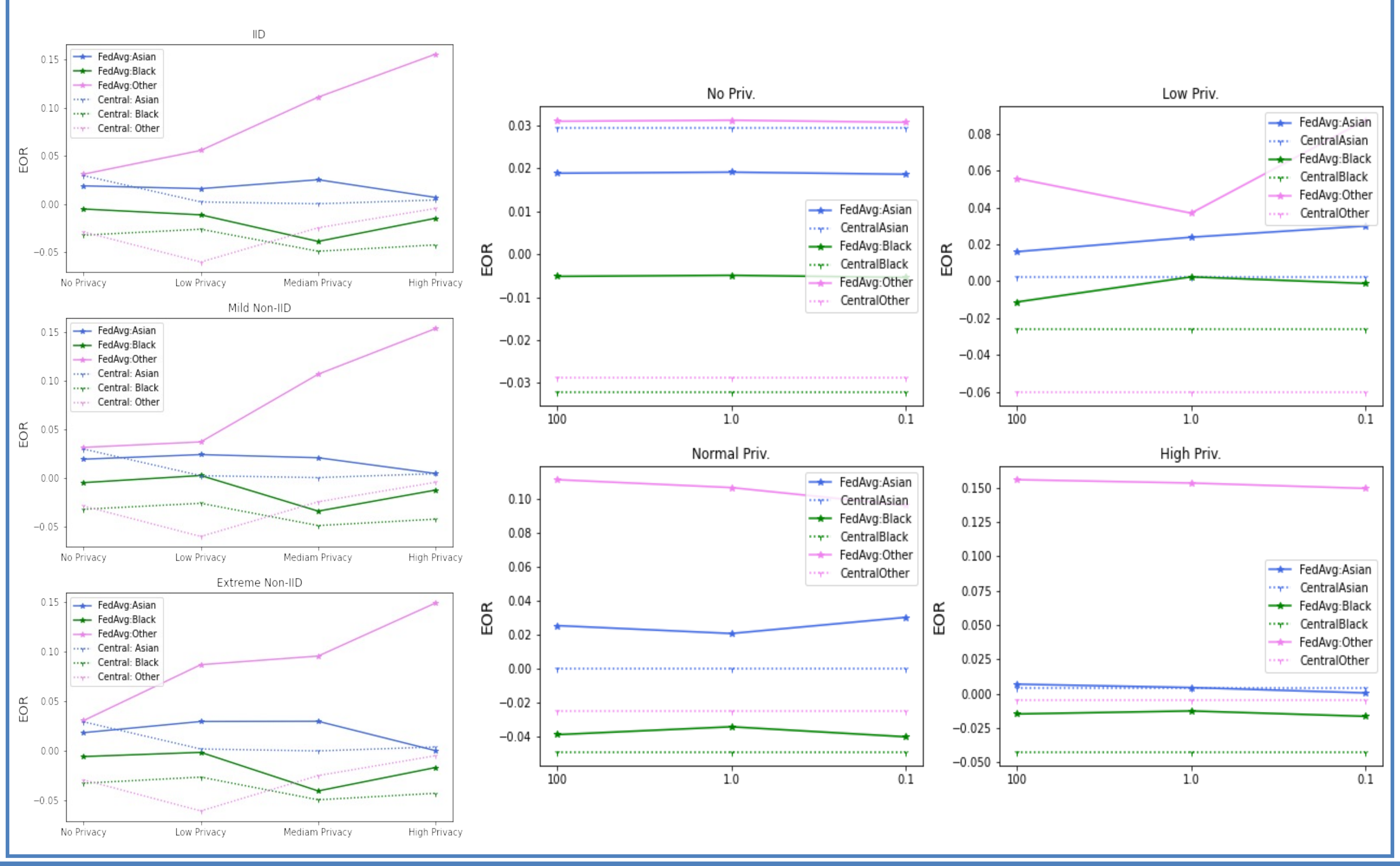
Impact of non-IID Data on Dataset-Level Fairness

- Fairness drops with increase in privacy level
- High privacy regimes act as a regularization method



Impact of non-IID Data on Group-Level Fairness

- Fairness drops with increase in privacy level, impact more prominent on more underprivileged groups
- Non-IID distribution has a negative impact in low privacy regimes, impact less prominent with increase in privacy level



More about EPI project: https://enablingpersonalizedinterventions.nl

Main references:

Verma, Sahil, and Julia Rubin. "Fairness definitions explained." *2018 ieee/acm international workshop on software fairness (fairware)*. IEEE, 2018.
Abadi, Martin, et al. "Deep learning with differential privacy." *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 2016.
Foulds, James R., et al. "An intersectional definition of fairness." *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020.