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

Amiri, S.; Belloum, A.; Nalisnick, E.; Klous, S.; Gommans, L.

DOI
10.1109/DSN-W54100.2022.00018

Publication date
2022

Document Version
Final published version

Published in
Proceedings, 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshop volume

2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W) | 978-1-6654-0262-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/DSN-W54100.2022.00018

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

1st Saba Amiri
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
s.amiri@uva.nl

2nd Adam Belloum
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
a.s.z.belloum@uva.nl

3rd Eric Nalisnick
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
e.t.nalisnick@uva.nl

4th Sander Klous
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
s.klous@uva.nl

5th Leon Gommans
Air France - KLM
name of organization (of Aff.)
Amsterdam, The Netherlands
leon.gommans@klm.com

***Abstract*—*Federated Learning* enables distributed data holders to train a shared machine learning model on their collective data. It provides some measure of privacy by not requiring the data be pooled and centralized but still has been shown to be vulnerable to adversarial attacks. *Differential Privacy* provides rigorous guarantees and sufficient protection against adversarial attacks and has been widely employed in recent years to perform privacy preserving machine learning. One common trait in many of recent methods on federated learning and federated differentially private learning is the assumption of IID data, which in real world scenarios most certainly does not hold true. In this work, we empirically investigate the effect of non-IID data on node level on federated, differentially private, deep learning. We show the non-IID data to have a negative impact on both performance and fairness of the trained model and discuss the trade off between privacy, utility and fairness. Our results highlight the limits of common federated learning algorithms in a differentially private setting to provide robust, reliable results across underrepresented groups.**

***Index Terms*—differential privacy, federated learning, non-IID data**

I. Introduction

Federated Learning (FL) [1] is a decentralized scheme to train *Machine Learning (ML)* models on several nodes, each being able to contribute compute resources and/or their own private dataset. In most common FL scenarios, local models are trained at participant sites and model parameters are sent to a trusted orchestrator to be fused into one set of parameters and sent back to participants. Among other advantages, FL contributes to preserving privacy of participants by negating the need to share their private data with each other or the orchestrator. However, it has been shown that FL by itself does not guarantee privacy and has been known to be vulnerable to different adversarial attacks such as *Membership Inference Attacks* [2], [3], *Reconstruction Attacks* [4], [5] and *Inversion Attacks* [6], [7].

Among methods proposed to provide measurable guarantees of privacy, *Differential Privacy (DP)* is capable of providing algorithmic guarantees of privacy against linkage attacks [8], [9]. In the context of FL, one can employ Global DP - when the participants trust the orchestrator and privacy is provided on participant level - [10] or Local DP - when there is no trust between the orchestrator and participants and privacy guarantees are provided on a record level [11].

Although adopting FL and DP can lead to both flexibility in terms of distributed training of ML models and privacy guarantees, they also incur costs such as computational overhead [12], reduction of performance [10] and negative impact on fairness [13]. The latter two problem are especially prominent in a federated learning scenario where distribution among nodes is Non-IID, i.e. the distribution of the data on local FL nodes is different than the distribution of the global dataset.

In this work, we investigate the effect of non-IID data on differentially private federated learning of a discriminative model. We focus on utility - as defined by common metrics such as F1-score, accuracy, etc. - and fairness of the model. We will establish baselines for centralized DP and non-DP FL models using metrics for utility and fairness of the model output in presence of the skewness in distribution of target and non-target features. Then we extend our experiments to the case of FL and measure the impact of non-IID data among different nodes on utility and performance of the model.

A. Contributions

In this work, we set out to evaluate the impact of different aspects of non-IID data on federated, differentially private training of discriminative models. To the best of our knowledge, this is the first work to explore empirically the impact

2325-6664/22/$31.00 ©2022 IEEE
DOI 10.1109/DSN-W54100.2022.00018

of non-IID data in a DP federated machine learning setting. We present the following contributions:

- **Setting baseline and exploring the privacy-utility/fairness trade-offs in presence of non-IID data.** We show disparate impact of DP in centralized and fully-IID DP FL setups. We show the negative impact of DP on the fairness and utility of the centralized DP and fully-IID DP FL models.
- **Establishing the threeway trade-off of distribution-privacy-utility/fairness of differentially private FL.** We simulate different levels of non-IID data for DP FL and show how an increase in the non-IID aspect of the data also generally has more of a negative impact on utility and fairness for underrepresented classes.

II. Related Work

Impact of DP on model performance and fairness. Bagdasaryan et al. [14] empirically show that in both centralized and federated settings, adding DP to the deep model will exacerbate any potential "unfair" traits of the model towards underrepresented groups. They show this disparate impact on vision and language models. Gu et al. [15] research the impact of DP on model fairness in FL and conclude that although both the noise adding mechanism and the gradient clipping step have regularization impact on the outcome of the deep model, they also have a negative impact on the fairness of the model for underrepresented classes. Suriyakumar et al. [16] extensively research the impact of DP on machine learning in healthcare and empirically analyze the trade-offs between privacy, utility and fairness. They show that applying the DP-SGD mechanism [1] to machine learning models in healthcare has a negative impact on the robustness and fairness of the model in presence of classes and labels with long-tailed distributions. They also show that DP disparately impacts group fairness by looking at loss of influence for majority groups.

Impact of data distribution on utility and fairness. Zhao et al. [17] show steep degradation of machine learning utility in federated setting in presence of non-IID and highly skewed data. Farrand et al. [13] empirically show that adding DP to a centralized model will have disparate impact on both utility and fairness of the model outcome, even in presence of slight class imbalance. They further show that the disparate impact of adding DP is not limited to the high-privacy regime and also can be observed in low-privacy settings. Finally they conclude that increasing the privacy level will result in loss of utility across all classes, which makes the model less efficient but more fair. Ozdayi et al. [18] research the impact of data distribution on fairness and robustness of FL and report the negative impact of non-IID data on model performance and fairness. They also conclude that the impact of non-IID data on fairness is far greater than its impact on model utility.

Our work focuses on the impact of non-IID data on a federated, differentially private model. To the best of our knowledge this is the first work to address the interplay between federated learning, differential privacy and non-IID data.

III. Background

This section provides basic background information for the main ideas and algorithms used in this work.

A. Federated Deep Learning

The goal of training a discriminative ML model $\mathcal{M} : X \rightarrow Y$ with parameters θ is to fit function $\mathcal{M}$ to estimate the distribution $q(y \in Y|x \in X)$. In this work, we focus on the common case of one label per sample, in which case the output of q could be a one-hot encoding of the labels.

To perform the training process in a federated setup, *Federated Averaging* is one of the most prominent DL training algorithms. In FL setup, K participants work in tandem with an orchestrator by iteratively training local models on their private datasets and sending their local parameters sets $\theta_{k,i}$ back to the orchestrator in each communication round i. The orchestrator combines the received parameter sets into a single fused version and send it back to the participants. Participants update their local model using the global parameter set and repeat this process until convergence. In this work, we assume the orchestrator waits to receive all participant updates before performing the fusion and communication.

B. 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 $\mathcal{M}$, $S \subseteq \mathcal{R}$:

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

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 $1/samplesize$.

One of the most prominent methods bringing DP to the machine learning paradigm is DP-SGD, proposed by McMahan et al. in 2016 [1]. The DP-SGD method works by clipping the gradients to control the sensitivity of the mechanism and adding calibrated noise to the gradient values. An accountant has been proposed to keep track of the spent privacy budget. Other accountants have been proposed in the literature that provide tighter bounds on privacy costs, e.g. $Rènyi$-DP-based accountant [19].

C. Non-IID Data in FL

Consider the federated learning setup with n participants for a discriminative ML use-case. Training dataset X consists of samples $(x, y) \sim P(x, y)$ with x being the input feature vector and y being the target feature we are aiming to predict. Non-IID data for participant i includes target feature non-IID ($P_i(y) \neq P(y)$) and non-target feature non-IID for feature x_j ($P_i(x_j) \neq P(x_j)$).

Salary\Race	White	Black	Asian	Other
<=50K	193844	23364	6641	6021
>50K	16398	801	600	197

TABLE I
DISTRIBUTION OF PROTECTED FEATURE *Race* CATEGORIES

over the target class *Annual Salary*

IV. METHODOLOGY

In this work we measure the impact of non-IID data (as defined in Section III-C) on a DP-FL model.

A. Dataset

To understand the implications of having non-IID data in a DP FL setup, we chose the *Census-Income (KDD)* dataset from UCI Machine Learning Repository [20], which has been extensively used in research literature on fairness [21]–[23]. It contains demographic census data of U.S population. The target feature *Salary* indicates whether the person has an annual income of over or under 50k dollars. Protected feature is *Race*.

We perform extensive preprocessing on features such as education, marital status and employment status and group disparate values into high level categories to increase the baseline classification efficiency. The preprocessed dataset contains 30 predictive features - including one protected feature -, one target feature and 248466 samples. Table I shows the distribution of our protected feature *Race* against the target feature *Annual Salary*.

To measure the impact of DP and non-IID distribution on fairness, we define the group *White* in *Race* feature as the *Privileged Group* and *Annual Salary>50k* as the desired outcome. For the relevant fairness metrics, the rest of the categories in *Race* will be conditioned against the privileged group and the (un)fairness will be measured against the desired outcome.

B. Non-IID Data

To simulate non-IID data in an FL setting, we follow [18] and distribute the samples among different FL participants based on their target feature using Dirichlet distribution. We control the amount of deviation from IID using the concentration parameter of the Dirichlet distribution. Assuming prior binomial distribution q over target feature and a categorical distribution p over target feature from which samples are drawn independently on participant level, we sample p from $Dir(\alpha q)$. Noting that $\alpha \to \infty$ emulates fully-IID data while $\alpha \to 0$ results in fully-non IID data, we choose three parameters for α: 0.1, 1 and 100. These parameters based on our experiments will emulate extreme non-IID, medium non-IID and almost IID data respectively. Figure 1 illustrates the distribution of target class and the protected attribute (Race) among 15 participants of our FL scenario with different concentration parameters.

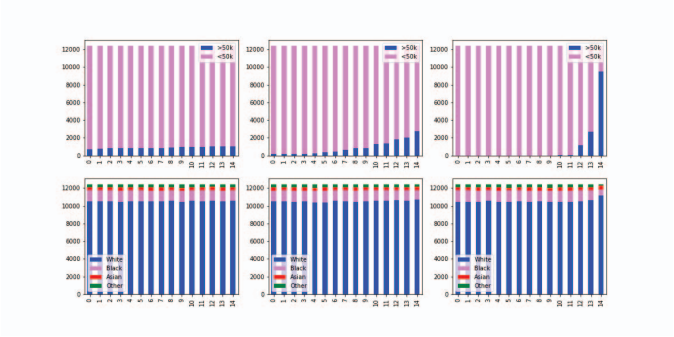

Fig. 1. Distribution of target feature and the protected class among 15 FL participants. Left column shows the IID data with $\alpha = 100$, middle column shows mild non-IID data with $\alpha = 100$ and right column shows extreme non-IID data with $\alpha = 100$

C. Metrics

To measure the utility of our experiments, we will report *F1-score*, *Precision* and *Recall*. In research area of fairness in ML, different metrics are defined and used for different usecases. For our binary discriminative model, we have chosen *Equalized Odds Rate* (EOR), *Generalize Entropy Index* (GEI) and *Differential Fairness Bias Amplification* (DFBA).

1) 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 $\mathcal{A} \subset x$ and $\mathbf{s}_i, \mathbf{s}_j \in A$ tuples of protected feature values. Randomized mechanism $\mathrm{M} : \mathcal{X} \to \mathcal{Y}$ is ϵ-Differentially Fair (DF) with respect to (A, Θ) if for all $(\mathbf{s}_i, \mathbf{s}_j) \in A \times A$ and $\mathbf{x} \sim \theta$:

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

for $\theta \in \Theta$ and $y \in \mathrm{Range}(M)$ where $P\left(\mathbf{s}_i \mid \theta\right) > 0, P\left(\mathbf{s}_j \mid \theta\right) > 0$ [24].

Intuitively ϵ-DF bounds the difference in log-likelihood of the probabilities of the outcomes of the randomized mechanism for any combination of protected feature values by ϵ.

To measure the impact of applying ϵ_1-DF mechanism M on ϵ_2-DF dataset $\mathcal{X}$ on the dataset's fairness, we can measure the Differential Fairness Bias Amplification (DFBA) which is the difference between ϵ values: $\epsilon_2 - \epsilon_1$ [24].

2) Generalized Entropy Index: Generalized Entropy Index (GEI) is a measure of inequality, originally defined in the context of economics to analyze the distribution of income and economic (in)equality [25].

For $\alpha \notin \{0,1\}$ and $b_i = \left(y_{predict_i} - y_{label_i} + 1\right)$, 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:

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

GEI is an individual and group level fairness metric and can be thought as the measure of redundancy in the data against the desired outcome from an information theoretic perspective.

3) Equal Odds Rate: Equal Odds Rate (EOR) [26] measures how close sensitivity of the outcome of the mechanism for privileged groups is to that of unprivileged groups in the protected feature with regard to the desired outcome. Equalized odds is achieved when there is absolute conditional independence between unprivileged groups in the protected feature and model outcome with regards to the desirable outcome.

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 $\mathbb{E}_{(\mathbf{x},y)\sim G}\left[\mathrm{M}(\mathbf{x}) \mid y = O\right] = \mathbb{E}_{(\mathbf{x},y)\sim G'}\left[\mathrm{M}(\mathbf{x}) \mid y = O\right]$

D. Implementation Details

We use a shallow feed-forward network with three fully connected layers. For federated setup, we simulate 15 clients and perform cross-silo FL. All our experiments are implemented using PyTorch[1] v1.10 and Opacus[2] v1.0.0 libraries. Experiments are run on a machine with a Tesla-P100 GPU and 32GB of RAM.

In this work we use *Rènyi*-DP [19] to achieve tighter privacy bounds. We assume no trust between participants and the orchestrator and opt for local DP by performing DP-SGD on each participant. We also assume a lack of availability of any public datasets to pre-train the models and/or to share between participants to mitigate the impact of non-IID data as has been suggested in some recent literature.

V. Experimental Results and Analysis

A. Impact on Utility

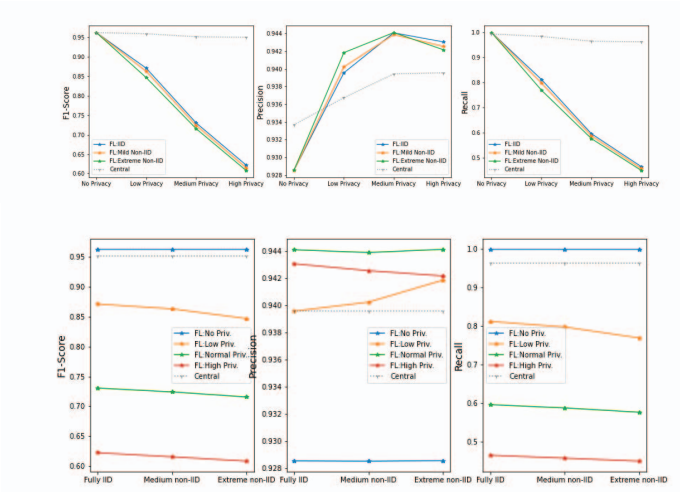

Fig. 2. Privacy and distribution trade-off against utility. The top row shows the privacy-utility trade-off and the bottom row shows the privacy-distribution trade off.

[1]https://pytorch.org/
[2]https://opacus.ai/

Figure 2 visualizes the privacy and distribution-utility trade-offs of our FL system. On privacy-utility front, considering the scale of the graphs makes it clear that recall is far more impacted than precision when the level of privacy increases. Looking at the F1-scores it can be deduced that the performance does decrease with increase in privacy level, with no significant difference between different distribution levels. However, looking at precision and recall graphs, we observe that while recall also drops with increase in privacy, precision shows an improvement which keeps its trend except for the high privacy regime. This could be contributed to the regularization effect of noise-adding mechanisms of DP. Although there is negligible drop of performance for the central model as we increase the privacy level, the positive impact of DP on precision can still be observed.

We have visualized the distribution-utility trade-off on the bottom row of Figure 2. We observe that going from fully-IID to extreme non-IID data, there is a drop in F1-Score and recall for privacy preserving regimes. However, while we see the same trend in precision for non-DP case and the high privacy regime, especially in case of low privacy regime we can see an increase in precision going from fully-IID to extreme non-IID. We can contribute this also to the regularization effect of DP which decreases the number of samples identified positively by the algorithm in this binary classification experiment, while increasing the number of samples incorrectly identified as negative.

B. Impact on Convergence Rate

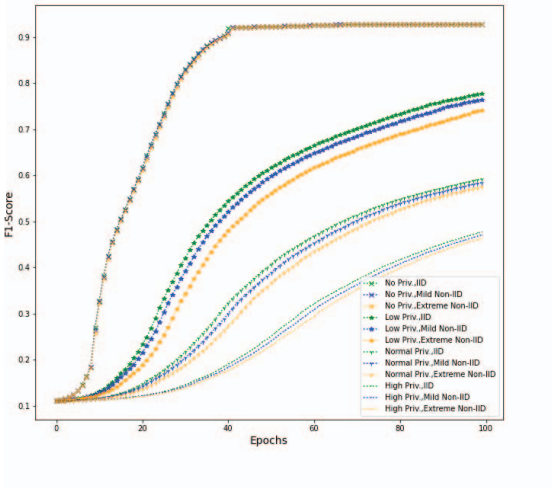

Fig. 3. F1-score of FL models with different privacy levels and distributions for 100 epochs.

Figure 3 depicts the impact of different privacy levels and distributions on the performance of our federated learning system. As can be observed, with non-DP setup the data distribution has no discernible impact on the performance and the convergence rate of the final model. However, while in

private setups the performance is negatively impacted, the data distribution has no significant impact on the convergence rate of the model.

C. Impact on Individual and Dataset-Level Fairness

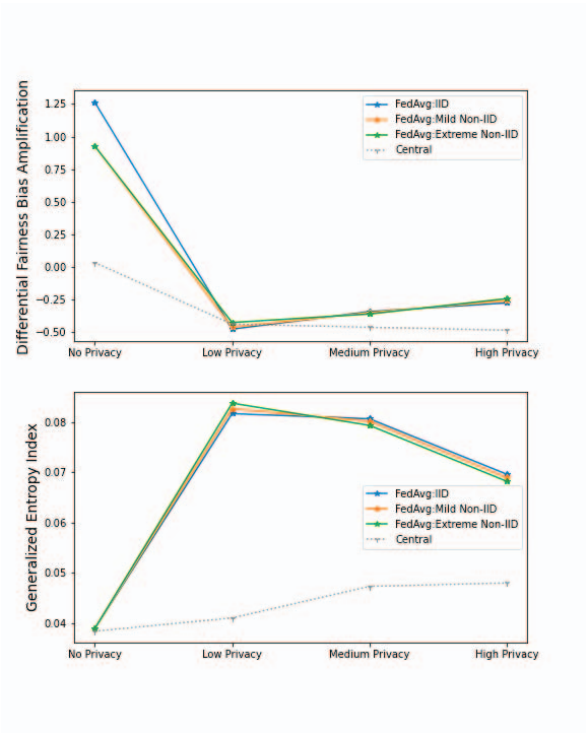

Fig. 4. Privacy trade-off against fairness

Figures 4 and 5 visualizes the privacy and distribution-fairness trade-off of our FL system for dataset level fairness metrics of differential fairness bias amplification and generalized entropy index. We observe that for centrally trained model, increasing the privacy level has a negative impact on both DFBA and GEI. However for the FL scenario the behavior of the system with regards to these two metrics is different. GEI is increased for all distribution scenarios when we go from no privacy to low privacy regime. However, increasing the privacy results in slight GEI drop, meaning that although the model is performing worse overall in terms of utility, increasing the DP reduces its negative impact on fairness by around 0.01. As for DFBA, we observe that in no privacy regime the classifier is causing bias to increase, while adding any measure of DP to the model will reduce the classifier bias to less than dataset bias. Increasing the DP level decreases this reduction in bias, but even in high privacy regime DP reduces the classifier bias to a level less than the dataset bias. It should be noted that this impact is not necessarily a positive one and should be evaluated independently for each specific use-case.

As for fairness-distribution trade-off, surprisingly DFBA decreases as we go from fully IID to extreme non-IID data. In

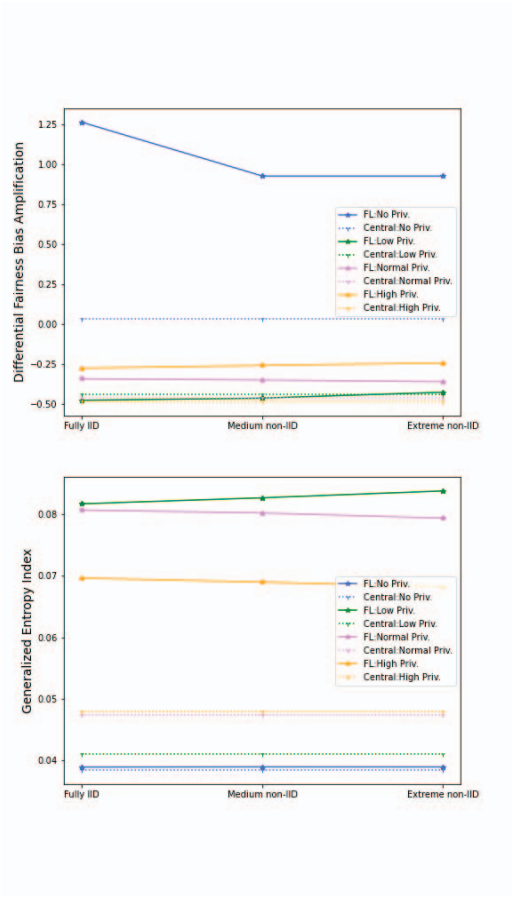

Fig. 5. Distribution trade-off against fairness

other privacy regimes, the DFBA stays the same for different data distributions. GEI stays the same for no privacy regime and increases slightly for low privacy regime, but shows a a moderate decrease in medium and high privacy regimes which could be attributed to the regularization effect of DP on model fairness.

D. Impact on Group-Level Fairness

Figures 6 and 7 show the impact of privacy and data distribution on EOR for underprivileged groups. We observe that adding privacy to the model has a small negative impact on the fairness of the model according to EOR. It is less prominent for the *Asian* group as their class imbalance is less severe than the other underprivileged groups. However in the federated setup, regardless of data distribution the negative impact of privacy is more prominent as we increase the privacy level. It is much more severe for the *Other* ethnicity group which is both under-represented in number of samples and in percentage of samples with desirable outcome.

In terms of distribution-fairness trade-off, we can observe that with no privacy the impact of data distribution on EOR is negligible. In private regimes the level of non-IID data has a negative impact on EOR fairness. The regularization effect of

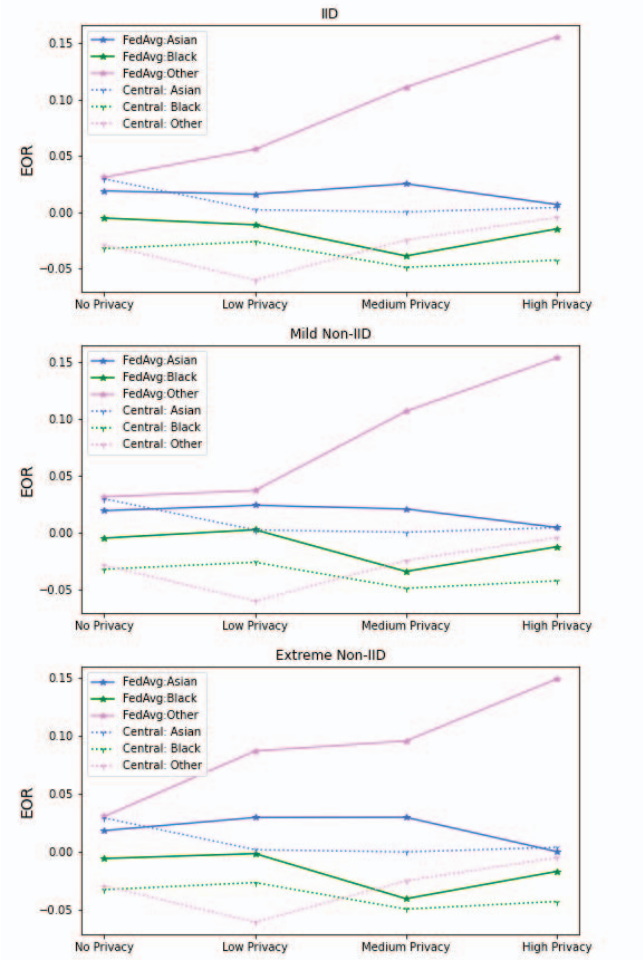

Fig. 6. Distribution trade-off against fairness. EOR is measured for each underprivileged group against the designated privileged group and for the desirable outcome of *Annual Salary* $> 50k$

high levels of DP can still be observed, as the negative impact on EOR is dampened going from low privacy to high privacy level.

VI. Conclusion and Future Works

In this work, we performed an extensive empirical analysis of the impact of non-IID data on the utility and fairness of differentially private federated deep learning models. Employing a real world dataset ubiquitous in ML fairness research with all three types of imbalance - class, label and node, the latter added during FL simulations - we explored the trade-

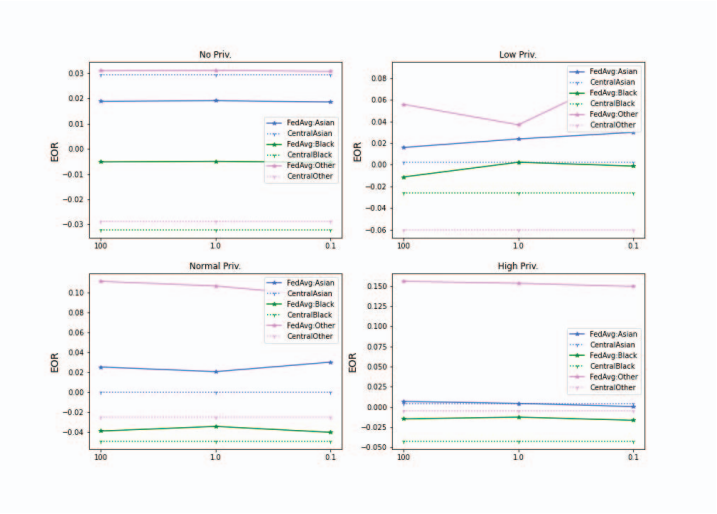

Fig. 7. Privacy trade-off against fairness. EOR is measured for each underprivileged group against the designated privileged group and for the desirable outcome of *Annual Salary* $> 50k$

offs between utility, privacy and data distribution. We showed that DP has a generally negative, although disparate, impact on both utility and performance of discriminative deep models for underprivileged groups. We also showed that non-IID data deepens the utility and fairness gap between minority and majority groups. We chose our utility and fairness metrics to enable us to explore the different aspects of non-IID data impact on our privacy preserving federated experimental setup to the fullest. Specifically differential fairness by definition has an organic connection to differential privacy, which encourages us to further explore how we can utilize this connection, e.g. providing lower bounds for differential fairness degradation by considering the privacy budget and non-IID aspects of the data.

Although consistent with our general intuitions about the impact of DP and non-IID data, this work is a limited exploration of the interplay between fairness and utility and their trade-off with privacy and data distribution. We are currently working on extending our experiments with a wide variety of datasets and more complicated models, e.g. large scale vision models. It would allow us to better understand the impact of non-IID data in privacy preserving FL.

We also plan on employing FL algorithms other than the vanilla *FedAvg* -e.g. *FedProx* [27], FairFed [28] - to see whether they can mitigate the utility and fairness degradation in non-IID FL setups. We have carefully eliminated the algorithms that have prerequisites which would make them unsuitable for DP-FL, e.g. sharing part of the dataset between participants.

Regardless, we have shown consistently that in presence of non-IID data, the utility and fairness of the privacy preserving discriminative models are negatively impacted which leads us to the conclusion that breaking the IID assumption hinders FL, and this must be carefully monitored when FL is deployed.

ACKNOWLEDGMENT

This research has been performed as part of the *Enabling Personalized Intervention* (EPI) project. The EPI project is funded by the Dutch Science Foundation in the Commit2Data program, grant number 628.011.028.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, pp. 1273–1282, PMLR, 2017.

[2] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Stand-alone and federated learning under passive and active white-box inference attacks," 2018.

[3] J. Zhang, J. Zhang, J. Chen, and S. Yu, "Gan enhanced membership inference: A passive local attack in federated learning," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2020.

[4] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond inferring class representatives: User-level privacy leakage from federated learning," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pp. 2512–2520, IEEE, 2019.

[5] M. Song, Z. Wang, Z. Zhang, Y. Song, Q. Wang, J. Ren, and H. Qi, "Analyzing user-level privacy attack against federated learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 10, pp. 2430–2444, 2020.

[6] Y. Huang, S. Gupta, Z. Song, K. Li, and S. Arora, "Evaluating gradient inversion attacks and defenses in federated learning," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[7] J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller, "Inverting gradients–how easy is it to break privacy in federated learning?," *arXiv preprint arXiv:2003.14053*, 2020.

[8] C. Dwork, A. Smith, T. Steinke, and J. Ullman, "Exposed! a survey of attacks on private data," *Annual Review of Statistics and Its Application*, vol. 4, pp. 61–84, 2017.

[9] C. Dwork, A. Roth, *et al.*, "The algorithmic foundations of differential privacy.," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, 2014.

[10] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. Quek, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.

[11] S. Truex, L. Liu, K.-H. Chow, M. E. Gursoy, and W. Wei, "Ldp-fed: Federated learning with local differential privacy," in *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*, pp. 61–66, 2020.

[12] T. T. Cai, Y. Wang, and L. Zhang, "The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy," *arXiv preprint arXiv:1902.04495*, 2019.

[13] T. Farrand, F. Mireshghallah, S. Singh, and A. Trask, "Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy," in *Proceedings of the 2020 Workshop on Privacy-Preserving Machine Learning in Practice*, pp. 15–19, 2020.

[14] E. Bagdasaryan, O. Poursaeed, and V. Shmatikov, "Differential privacy has disparate impact on model accuracy," *Advances in Neural Information Processing Systems*, vol. 32, pp. 15479–15488, 2019.

[15] X. Gu, T. Zhu, J. Li, T. Zhang, and W. Ren, "The impact of differential privacy on model fairness in federated learning," in *International Conference on Network and System Security*, pp. 419–430, Springer, 2020.

[16] V. M. Suriyakumar, N. Papernot, A. Goldenberg, and M. Ghassemi, "Chasing your long tails: Differentially private prediction in health care settings," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 723–734, 2021.

[17] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," *arXiv preprint arXiv:1806.00582*, 2018.

[18] M. S. Ozdayi and M. Kantarcioglu, "The impact of data distribution on fairness and robustness in federated learning," *arXiv preprint arXiv:2112.01274*, 2021.

[19] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pp. 263–275, IEEE, 2017.

[20] D. Dua and C. Graff, "UCI machine learning repository," 2017.

[21] V. Iosifidis and E. Ntoutsi, "Adafair: Cumulative fairness adaptive boosting," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 781–790, 2019.

[22] T. L. Quy, A. Roy, V. Iosifidis, and E. Ntoutsi, "A survey on datasets for fairness-aware machine learning," *arXiv preprint arXiv:2110.00530*, 2021.

[23] V. Iosifidis, A. Roy, and E. Ntoutsi, "Parity-based cumulative fairness-aware boosting," *arXiv preprint arXiv:2201.01148*, 2022.

[24] J. R. Foulds, R. Islam, K. N. Keya, and S. Pan, "An intersectional definition of fairness," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pp. 1918–1921, IEEE, 2020.

[25] A. F. Shorrocks, "The class of additively decomposable inequality measures," *Econometrica: Journal of the Econometric Society*, pp. 613–625, 1980.

[26] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," *Advances in neural information processing systems*, vol. 29, pp. 3315–3323, 2016.

[27] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.

[28] Y. H. Ezzeldin, S. Yan, C. He, E. Ferrara, and S. Avestimehr, "Fairfed: Enabling group fairness in federated learning," *arXiv preprint arXiv:2110.00857*, 2021.