Towards Fair and Privacy Preserving Federated Learning for the Healthcare Domain

***Abstract*—Federated learning (FL) enables data sharing in healthcare contexts where it might otherwise be difficult due to data-use-ordinances or security and communication constraints. Distributed and shared data models allow models to become generalizable and learn from heterogeneous clients while addressing data security, privacy, and vulnerability considerations as data itself is not shared across a given learning network's nodes. On the other hand, FL models often struggle with variable client data distributions and operate on an assumption of independent and identically distributed (IID) data. As the field has grown, the notion of fairness aware federated learning (FAFL) mechanisms have also been introduced and is of distinct significance to the healthcare domain where many sensitive groups and protected classes exist. In this paper we create a benchmark methodology for FAFL mechanisms under various heterogeneous conditions on datasets in the healthcare domain typically outside the scope of current federated learning benchmarks, such as medical imaging and waveform data formats. Our results indicate considerable variation in how various FAFL schemes respond to high levels of data heterogeneity and additionally, that doing so under privacy preserving conditions can create significant increases in network communication cost and latency compared to the typical federated learning scheme.**

***Index Terms*—Federated Learning, Healthcare, Medical Imaging, Privacy, Fairness**

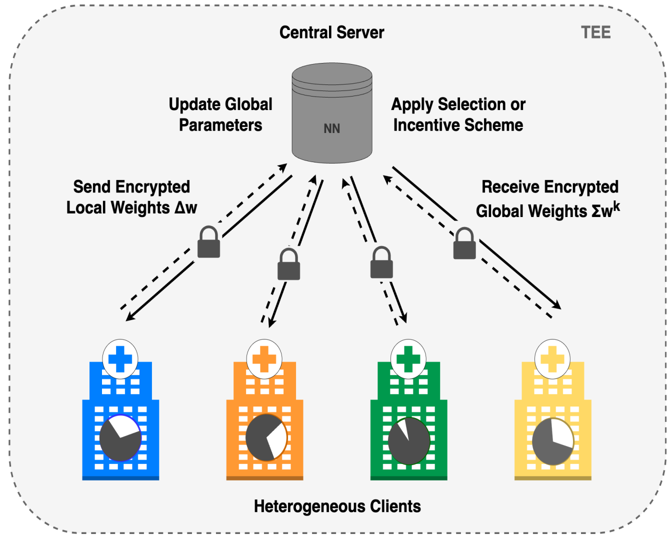

Fig. 1. Discrete generalized federated learning lifecycle with client node level and server level processes and interactions. After model training and inference, key client server interactions include sending and downloading weights during aggregation steps.

I. Introduction

Privacy and security have been a growing concern with Machine Learning systems. Having huge amounts of data in the same place is a security risk. Therefore, many domains such as healthcare regulate the centralization of data. Addressing this concern of centrally located data and privacy, McMahan et al. contributed to the concept of federated learning (FL) and distributed architecture [1]. In a federated learning system, decentralized clients load and train data on their own devices and communicate weights to a global server and is therefore ideal for the sensitive nature of clinical environments.

Fairness within FL systems can refer to both individual and group parity, performance, and participation. FAFL can involve every stage of the FL lifecycle as seen in Fig. 1, including selection, collaboration, and incentives [2]. Like with federated learning in healthcare, there is no standard usage of FAFL mechanisms and analysis is typically constrained to one method at a time. There remains no existing benchmark on the relationship and performance of fair and privacy preserving FL implementations in the healthcare context.

There exist many health data ordinances that ensure patient privacy and confidentiality, commonly through compliance fixtures [3]. Machine Learning in healthcare typically involve multiple networks with differing communication and capacities and trusted execution environments (TEEs) in healthcare are constrained in communication and specialized [4]. Differential privacy is also relevant as it enables utilization of features while minimizing available information on individuals in a group. Moreover, healthcare incorporates a wide variety of data, including those from minimally encrypted emitting instruments with heterogeneous capabilities or whose infrastructure is split between client and server. This can leave them vulnerable to hijacking or inversion [5].

The lack of standard FL in healthcare usage remains a challenge as implementations remain context specific. The contributions of this paper will include jointly implementing FAFL mechanisms under an AES-GCM encryption scheme and proposing additional benchmarks on common healthcare datasets as described in Table II.

II. Related Work

A. Federated Learning Benchmarks

Many traditional benchmarks of federated learning implementations on health datasets, such as signal datasets and clinical databases like MIMIC III, do not incorporate fairness mechanisms [6]. Several applications for differential privacy and encrypted computation have been separately proposed for health contexts, notably PySyft and Microsoft SEAL [7] [8].

Unified benchmark systems like FedScale have an apparent lack of built in health datasets and encryption methods.

B. Fairness Approaches

1) Client Selection: Client selection and filtering are stages of nodes selection based on on potential to improve the global model [2]. Clients' resource constraints can determine its capacity to contribute to the global model, leading to clients being filtered out before the first round. Clients are rewarded or removed from the global model depending on performance.

2) Contribution Fairness: Contribution evaluation typically consider fairness in node filtering and rewards. This approach is key for differential privacy, as individual features can be removed for deidentification [9]. Participants are rewarded based on novel global contributions often by rewarding participants based on test accuracy of the client's model [10].

3) Group Fairness: Shi et al. defines group fairness as in-group performance parity, such as basic assured accuracy. Hu et al. [2] satisfy global model fairness requirements through a conditional group loss, enforcing minimum fairness metrics but not requiring them to be equal [12]. Huang et al. [15] presented long-term client selection fairness constraints.

4) Agnostic Learning: Agnostic loss based federated learning (AFL) models of fairness prevent overfitting by optimizing for variable client mixtures instead of assuming distribution parity between target and client. They have been noted struggle to generalize with large clients [13].

5) Expectation and Regret Fairness: Under a regret fairness scheme, nodes follow an objective function that balances latency with incentive payouts [2]. This mimics the free-rider problem by preventing high quality contributors from being represented [14]. Expectation fairness minimizes inequity through incentive payouts that occur at different periods, which is useful for schemes for variable incentive budgets [16].

III. Experimental Design

A. Heterogeneous Client Distribution

Typical FL schemes assume client data is independent and identically distributed (IID) while real world environments typically possess high levels of heterogeneity [17]. We simulate non-IID clients by creating 10 data partitions and implementing a parameter α to simulate the percentage of client data that is non-IID for low, medium, and high values as depicted in Figure 2. To further simulate heterogeneity, we also enforce a maximum unique label limit of 2 per client. For reference, we also include results from an IID sampled model.

B. AES Encryption

Advanced Encryption Standard-Galois Counter Mode (AES-GCM), a classic Authenticated Encryption mode with highly secure symmetric encryption, is incorporated in every federated round we train to protect privacy and client-server communication [18]. Parallel computing is a key component in our framework and the widely utilized AES-GCM is suitable due to the size and sensitive nature of healthcare-focused datasets [19]. We further plan to extend this and launch our models in a Trusted Execution Environment (TEE) to monitor model behavior in realistic privacy-preserving environments.

IV. Benchmark and Analysis

We demonstrate our methodology on the NIH Chest X-Ray Dataset by formulating a triclass problem of classifying presence of Pneumonia, a non-Pneumonia specified lung conditions, or no findings. Fairness approaches (see Table II) we account for are (1) the *long-term client selection fairness constraint* and (2) two model schemes utilizing the *reputation threshold contribution fairness*, one featuring an adversarial agent and one scheme without any agents. All models trained with 5 base clients for 50 epochs and 100 rounds of federation.

Our results demonstrate clear decreases in validation accuracy with increasing client heterogeneity, with different fairness mechanisms responding to variability differently. The reputation scheme with an additional adversarial agent outperforms the scheme without any agents at all levels of heterogeneity. Notably, the long term fairness constraint mechanism does not perform much differently than the standard federated scheme at low levels of heterogeneity but eventually converges to the optimal scheme as heterogeneity increases.

V. Conclusions and Futue Work

Federated learning in healthcare is a natural extension of the standard FL architecture and is aligned with the healthcare domain focuses of security, privacy, and group fairness. In this paper we create a novel benchmark for FAFL mechanisms in healthcare contexts and propose augmenting it. Applying FAFL in privacy preserving settings also brings challenges of communication overhead due to additional feedback loops and enforced security. These challenges will continue to be relevant

Fairness Approach	**Implementation**	**Local Protocol**	**Global Protocol**
(1) Client Selection	Long Term Fairness [9]	$C^2MAB(\eta, c_k^t, \Delta w_{i*}^t)$	$\frac{1}{K}\Sigma_{i=1}^K w_i^t$ if $I_k^t = 1$
(2) Contribution Fairness	Reputation Threshold [10] [11]	$w_k^t + \Delta w_k^t + \Delta w_{i*}^t, r_k^t$	$\Sigma_{i=1}^R r_i^{t-1} w_i^t \times \gamma / \|\Delta w_i^t\|$
(3) Group Fairness	Bounded Group Loss [12]	$w_k^t - (\lambda^T r + \Delta w_{i*}^t)$	$w_i^t + \frac{1}{K}\Sigma_{i=1}^K g_i^{t+1}$
(4) Agnostic Fairness	Agnostic Loss [13]	$\min\max L(\lambda_i^t, \Delta w_{i*}^t)$	$\frac{1}{K}\Sigma_{i=1}^K w_i^t, \frac{1}{K}\Sigma_{i=1}^K \lambda_i^t$
(5) Expectation Fairness	Incentive Sharing [14]	$\max V(s_k^t, \Delta w_{i*}^t)^k$	$\frac{1}{K}\Sigma_{i=1}^K w_i^t, \pi_i^t(w_i^t, s_i^t)$

TABLE I: Generalized formulation of fairness mechanisms. The client and global parameter set for the next time step are represented by w_k^{t+1} and w_i^{t+1} respectively. Protocols may feature unique terms such as the gradient regularization parameter γ. We let Δw_i^t be implicitly representative of a designated loss function $f_k(w_i^t)$ applied to weights at round i for K clients.

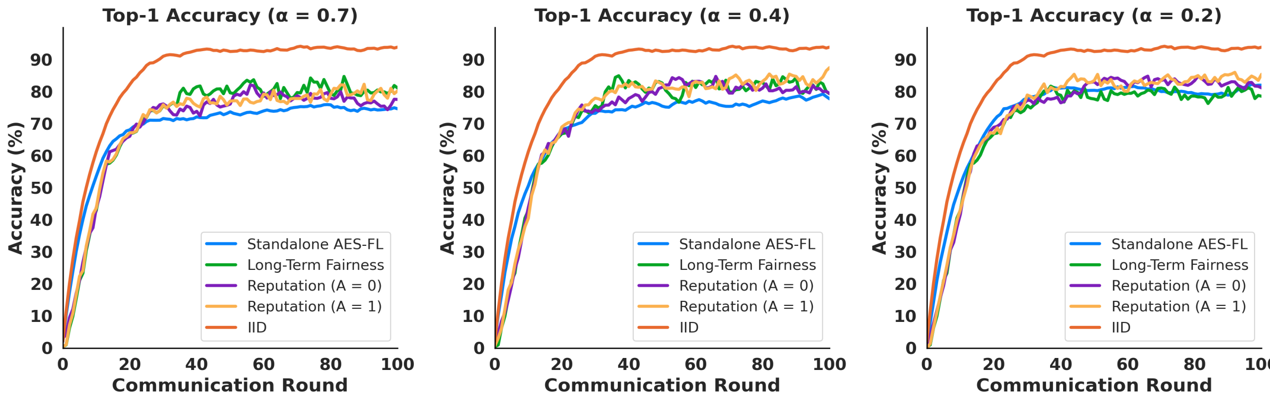

Fig. 2. Benchmarks for inference on NIH Chest X-Ray Dataset with baseline and distributed learning classification models and varying ratios of non-IID data of clients. Top-1 accuracy is reported for for standalone standard FL, long term client selection fairness constraint, reputation mechanism without adversarial agent, and reputation mechanism with adversarial agent schemes. Optimal scheme is long term fairness for alpha = 0.7 (**Left**) and reputation with adversarial agent for alpha = 0.4 (**Center**) and alpha = 0.2 (**Right**).

Dataset	Data Type	Format	Total Classes	Size	Base Model
NIH Chest X-Ray	Adult X-Ray Images	PNG	13	25k	ResNet-34
Fast MRI	Brain Tissue Scans	3D Arrays	2	8k	U-Net
PTB Dataset	ECG Waveforms	Signal	10	$<$1k	ResNet-18
MIMIC II	Clinical and Vital Signs	Tabular, Signal	59	30k	XGBoost

TABLE II: Proposed benchmark health dataset and characteristic datatype, format, classes, and size. A base transfer learning model is given for each dataset and classification problem to be benchmarked.

as the field of FL, fairness, and their use in healthcare evolve to include more complex and multi-notion implementations.

We propose future work to extend our analysis to the datasets listed in Table II and a wider range of mechanisms and schemes, including ones built to address heterogeneity such as FedProx. To further augment our analysis, we propose benchmarking estimators such as communication and execution times to further simulate real world settings.

References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," 2016.

[2] Y. Shi, H. Yu, and C. Leung, "Towards fairness-aware federated learning," 2021.

[3] M. A. Ahmad, A. Patel, C. Eckert, V. Kumar, and A. Teredesai, "Fairness in machine learning for healthcare," in *Proceedings of the 26th acm sigkdd international conference on knowledge discovery & data mining*, pp. 3529–3530, 2020.

[4] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," *International Journal of Intelligent Networks*, vol. 3, pp. 58–73, 2022.

[5] A. R. Javed, I. Razzak, and G. Xu, "Federated learning for privacy preservation of healthcare data from smartphone-based side-channel attacks," *IEEE journal of biomedical and health informatics*, vol. PP, 05 2022.

[6] G. H. Lee and S.-Y. Shin, "Federated learning on clinical benchmark data: Performance assessment," *J Med Internet Res*, vol. 22, p. e20891, Oct 2020.

[7] A. Budrionis, M. Miara, P. Miara, S. Wilk, and J. G. Bellika, "Benchmarking pysyft federated learning framework on mimic-iii dataset," *IEEE Access*, vol. 9, pp. 116869–116878, 2021.

[8] F. Wibawa, F. O. Catak, S. Sarp, M. Kuzlu, and U. Cali, "Homomorphic encryption and federated learning based privacy-preserving cnn training: Covid-19 detection use-case," 2022.

[9] S. K. Shyn, D. Kim, and K. Kim, "Fedccea : A practical approach of client contribution evaluation for federated learning," 2021.

[10] X. Xu and L. Lyu, "A reputation mechanism is all you need: Collaborative fairness and adversarial robustness in federated learning," 2020.

[11] T. Huang, W. Lin, W. Wu, L. He, K. Li, and A. Y. Zomaya, "An efficiency-boosting client selection scheme for federated learning with fairness guarantee," *CoRR*, vol. abs/2011.01783, 2020.

[12] S. Hu, Z. S. Wu, and V. Smith, "Fair federated learning via bounded group loss," 2022.

[13] M. Mohri, G. Sivek, and A. T. Suresh, "Agnostic federated learning," *CoRR*, vol. abs/1902.00146, 2019.

[14] H. Yu, Z. Liu, Y. Liu, T. Chen, M. Cong, X. Weng, D. Niyato, and Q. Yang, "A fairness-aware incentive scheme for federated learning," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, AIES '20, (New York, NY, USA), p. 393–399, Association for Computing Machinery, 2020.

[15] W. Huang, T. Li, D. Wang, S. Du, and J. Zhang, "Fairness and accuracy in federated learning," *CoRR*, vol. abs/2012.10069, 2020.

[16] J. Fitzsimons, M. Osborne, and S. Roberts, "Intersectionality: Multiple group fairness in expectation constraints," 2018.

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

[18] D. A. McGrew and J. Viega, "The security and performance of the galois/counter mode (gcm) of operation.," in *Indocrypt*, vol. 3348, pp. 343–355, Springer, 2004.

[19] M. Dworkin, "Recommendation for block cipher modes of operation: Galois/counter mode (gcm) and gmac," tech. rep., National Institute of Standards and Technology, 2007.