CIS520 Project Report

A validation of decentralized federated learning for clinical computer vision

Jordan Matelsky, Felipe Parodi
Team Name: Fed Heads

Abstract

Federated learning is a promising new approach to cooperative training of machine learning models. Because of its additional data-privacy characteristics, federated learning has seen overwhelming adoption in edge computing, but only modest adoption in other domains, due to the logistics requirements of provisioning a large, central, trusted server for parameter fusion. In this work, we explore the emerging technology of decentralized federated learning, wherein each node in the federated community performs its own peer-to-peer parameter fusion locally. We apply this generalized form of federated learning to the domain of clinical imagery analysis, which has historically suffered from the logistical hurdles of centralizing patient data for training. We demonstrate that decentralized federated learning enables disparate sites to train classifier models without sharing data and without the need for a high-burden, always-on, central server. We share several use-case studies of the decentralized federated learning technology, ranging from simple toy examples to complex, state-of-the-art radiology classifiers, with training-data splits ranging from the simple (uniform distribution of classes between nodes) to the pathological (one class label per compute node), and we find that decentralized federated learning enables otherwise intractable machine learning deployments.

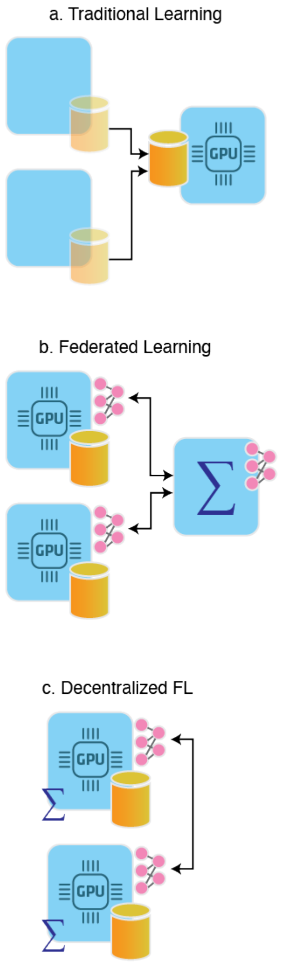

Figure 1: Three types of learning examined in this work. **a. Traditional learning.** Edge nodes provide only data. A central server performs all computation. **b. Federated learning.** Edge nodes perform local computation and training. A central server performs parameter fusion. **c. Decentralized federated learning.** Edge nodes perform local computation and training. Each node performs fusion with its networked neighbors.

1 Motivation

Among the most inspiring of recent advances in computer vision (CV) is the extraordinary performance of deep neural networks on clinical radiology imagery: neural networks can now perform radiologist-level disease classification from imagery alone [1]. One major setback in the field of clinical CV is the need for diverse and extensive training data. Clinical datasets are protected under local and federal privacy law, which dramatically increases the difficulty of amassing a large, multi-site training dataset for deep learning researchers.

Federated learning (FL) is a technique in which multiple compute nodes independently train the same neural network architecture, and then “agree” upon network parameters through a centralized fusion step [2]. In recent work, this process has been adapted to a *decentralized* (DFL) approach in which each compute node performs its own local fusion step (rather than relying upon a central authority). [3] In this work, we share an application of this novel DFL approach to the domain of clinical CV. We demonstrate that DFL learning on distributed datasets — *i.e.*, cooperatively learning a shared model across hospital boundaries — is a viable replacement for the much more logistically and legally nuanced aggregation of data in a centralized repository. We compare DFL approaches with the traditional centralized approach, and illustrate considerations for future work in the novel training architecture.

2 Related Work

Centralized federated learning (FL) has been in active production use for many years. Perhaps the most commonly encountered implementation is *GBoard*, a virtual keyboard that trains a local predictive network locally on each phone, the parameters of which are then uploaded to a central Google server to be fused and re-distributed. [2] Decentralized federated learning (DFL) is a novel approach that removes the central server and instead enforces a peer-to-peer network topology on the federated compute nodes. Though this application has been explored experimentally in the past, it has not to our knowledge been applied in practice, nor in a production environment. [4, 5]. The first known generalized framework for DFL — named *Scatterbrained* — was released in 2021, which we use here [3].

3 Data

In order to compare decentralized federated learning and traditional ML approaches, we used an existing, well-studied clinical CV dataset. Ideal options included the **NIH Chest X-ray** dataset [6] or **CheXpert** [1]. In the case that these large datasets and their reference network complexities overextend our compute resources, we have identified several smaller binary classification datasets, such as the **Pneumonia Chest X-Ray Images** dataset on Kaggle [7], as viable alternatives, in order to de-risk our project. A summary of these datasets is available in **Table 1**.

4 Problem Formulation

We examined two types of machine learning in this work: Traditional learning, and decentralized federated learning (**a** and **c** in **Fig. 1**). We compared the performance of a reference network (the same between each type of learning) by modifying its access to data in order to simulate real-world data imbalances.

We first separated the data into multiple hospital “sites,” where each hospital has a different balance of each class label. (For example, the Hospital of the University of Pennsylvania sees more *pneumonia* patients and no *edema* patients; the Johns Hopkins hospital sees *edema* patients but no *pneumonia* patients; etc.) We then provided the network access to data thus:

Traditional Learning. We provide the network serial access to all of the data, randomly shuffled (the control case, to illustrate the conventional approach to learning on large datasets). Then we provide a new, untrained network with *serial* access to each of the split hospital datasets, one after the other, to illustrate the closest analog of federation on a single node.

Decentralized Federated Learning. We provide each node its own dataset from the split detailed above. The nodes communicate in an all-to-all network topology for peer-to-peer parameter fusion.

Following training, we will compare the performance of each network in order to determine which of the methods best captured the variance of the training data, and we will then make suggestions for clinical application of federated learning based upon the strengths and weaknesses we discover.

In our analysis, we also consider the implications of data transfer as well as differential privacy [8], two aspects in which federated learning has considerable strength.

5 Methods

We first identify a target dataset of interest (see **Data**). We then retrieve and validate the performance of a reference neural network model. Reusing a community-contributed model enables us to better contextualize against existing baseline performance, as well as de-risk the technical components of this work. We then produce two federated learning

Dataset	Dataset Size	Task Size	Reference
NIH Chest X-ray	108,948 Images	8 Multilabels	[6]
CheXpert	224,316 Images	14 Multilabels	[1]
Chest X-Ray Images (Pneumonia) – Kaggle	5,863 Images	Binary Classification	[7]

Table 1: Options for dataset use. Two larger datasets are viable candidates for multilabel assignment; a binary classification example dataset is also provided as a simpler alternative. All options have several high-performing reference network implementations available online.

topologies: A centralized, "conventional" FL network (hub-and-spoke), and a novel, all-to-all decentralized FL network, using the *Scatterbrained* library [3].

We train each network architecture entirely independently, though we reuse the same train/test data splits for each (see **Problem Formulation**). This helps to ensure that we are fairly comparing each network topology and training architecture.

Finally, we compare the performance of each architecture, based upon the metrics detailed in **Evaluation**. We report these, alongside a "go/no-go" analysis of the different networks.

5.1 Evaluation

We evaluated the training architectures along several axes. In particular, we measured mean accuracy, and time-to-best-performance. In addition to these conventional metrics, we also planned to measure the volume of data transfer per model (i.e., the cost of bandwidth), the wall clock time to train, and the maximum achieved performance.

In addition to these quantitative metrics, we will also report on qualitative discoveries during the implementation of this novel approach. If FL tools are difficult to adapt on our intended timeline, this is a relevant and noteworthy weakness of the methods proposed, and we will note it accordingly in the evaluation.

6 Experiments & Results

6.1 Experiment 1: MNIST Classification with SGD

We classified MNIST digits [9] using a simple linear SVM with stochastic gradient-descent (SGD) learning, leveraging the `sklearn.linear_model.SGDClassifier` implementation from *sklearn.* We produced three benchmark results: a single-site training benchmark (using no federated learning), a centralized federated learning approach, and a decentralized federated learning approach. We evaluated mean-accuracy overall — i.e., where n is the number of samples,

$$\texttt{accuracy}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}(\hat{y}_i = y_i)$$

Compared to the standard model, both federated learning models were trained on less data per individual node, and therefore (as expected) performed worse overall. Of the two federated models, the decentralized (fully-connected) federation outperformed the centralized federated model both in terms of absolute score after each epoch as well as in rate of learning for the untrained nodes (**Fig. 2**).

6.2 Experiment 2: Class-Isolated Learning on MNIST

In this experiment, we trained a decentralized, fully-connected federation of `sklearn.linear_model.SGDClassifier` learners in which each node was given access to only one class in the training dataset. In other words, *Node 4* saw only MNIST training samples with the class label of "4." We were surprised to discover that the entire community trained in lock-step, and after each parameter-fusion step, all nodes were equally performant when evaluated on *all* digits, not just the digit upon which the node had been trained (**Fig. 3**).

7 Conclusions & Discussion

From these early results, we have illustrated that decentralized federated learning is a powerful technique that can be used to improve the performance of a community of learners. In our upcoming work, we will explore the potential of this technique in a healthcare environment, and we will examine potential privacy and performance considerations.

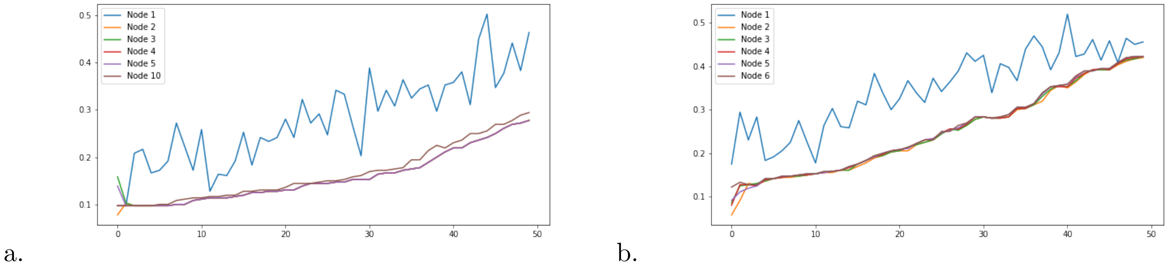

Figure 2: Early results comparing a simple linear SGD classifier in (a) a traditional, centralized federated learning regime, and (b) a decentralized federated learning regime. Notice that the performance of the untrained nodes in the decentralized approach learn a useful model more rapidly than those in the centralized approach. x-axis is training epoch step; y axis is mean-accuracy.

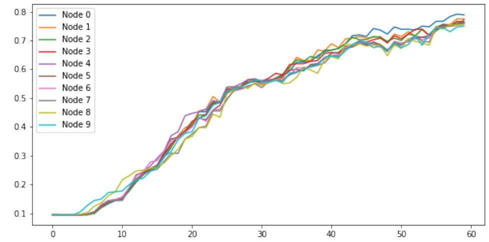

Figure 3: Results of class-isolated training of a DFL network. Each node saw in its training set only one class label each. Despite this, all nodes perform equally well on all digits during validation. x-axis is training epoch step; y axis is mean-accuracy.

References

[1] P. Rajpurkar, J. Irvin *et al.*, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” 2017.

[2] B. McMahan and D. Ramage, “Federated learning: Collaborative machine learning without centralized training data,” https://ai.googleblog.com/2017/04/federated-learning-collaborative.html, 2017, accessed: 2020-09-24.

[3] M. Wilt, J. K. Matelsky, and A. S. Gearhart, “Scatterbrained: A flexible and expandable pattern for decentralized machine learning,” *In submission*, 2021.

[4] A. Lalitha, O. C. Kilinc, T. Javidi, and F. Koushanfar, “Peer-to-peer federated learning on graphs,” *arXiv preprint arXiv:1901.11173*, 2019.

[5] A. G. Roy, S. Siddiqui *et al.*, “Braintorrent: A peer-to-peer environment for decentralized federated learning,” *arXiv preprint arXiv:1905.06731*, 2019.

[6] X. Wang, Y. Peng *et al.*, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017.

[7] D. S. Kermany, M. Goldbaum *et al.*, “Identifying medical diagnoses and treatable diseases by image-based deep learning,” *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[8] M. Naseri, J. Hayes, and E. D. Cristofaro, “Local and central differential privacy for robustness and privacy in federated learning,” 2021.

[9] Y. LeCun, C. Cortes, and C. Burges, “MNIST handwritten digit database,” *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, vol. 2, 2010.