FedSLD: Federated Learning with Shared Label Distribution for Medical Image Classification

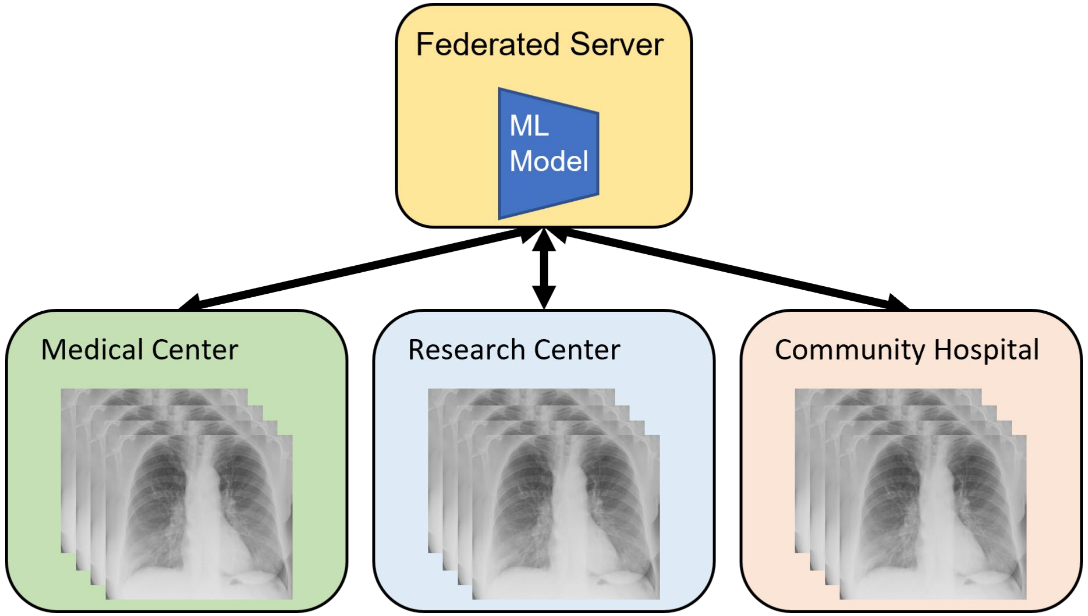
Jun Luo[a] and Shandong Wu[a,b,c,d]

[a]Intelligent Systems Program / [b]Dept. of Radiology / [c]Dept. of Biomedical Informatics / [d]Dept. of Bioengineering,
University of Pittsburgh, Pittsburgh, PA USA

Background

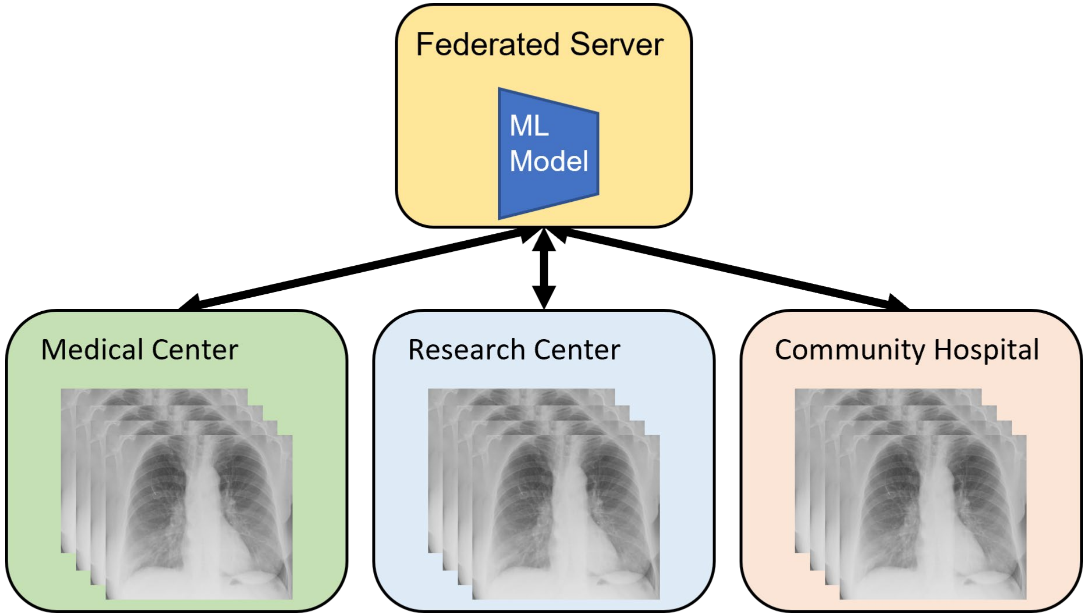
- Deep learning requires a large amount of data
  - Large medical datasets are difficult to collect
  - Medical data is privacy-sensitive
  - Laws and regulations (e.g. HIPAA, GDPR) make it hard to share data

- Federated learning (FL) – privacy preserving machine learning
  - Push model to the clients
  - Only model weights are shared while keeping the data decentralized



Background

- Federated learning poses data heterogeneity challenge
  - Data heterogeneity – non-IID
  - Medical datasets are often non-IID
    - Different data acquisition protocols
    - Different local demographics
    - Etc.
  - Potential influence
    - slower convergence
    - inferior performance
    - Loss of clients’ incentives to participate in the federation

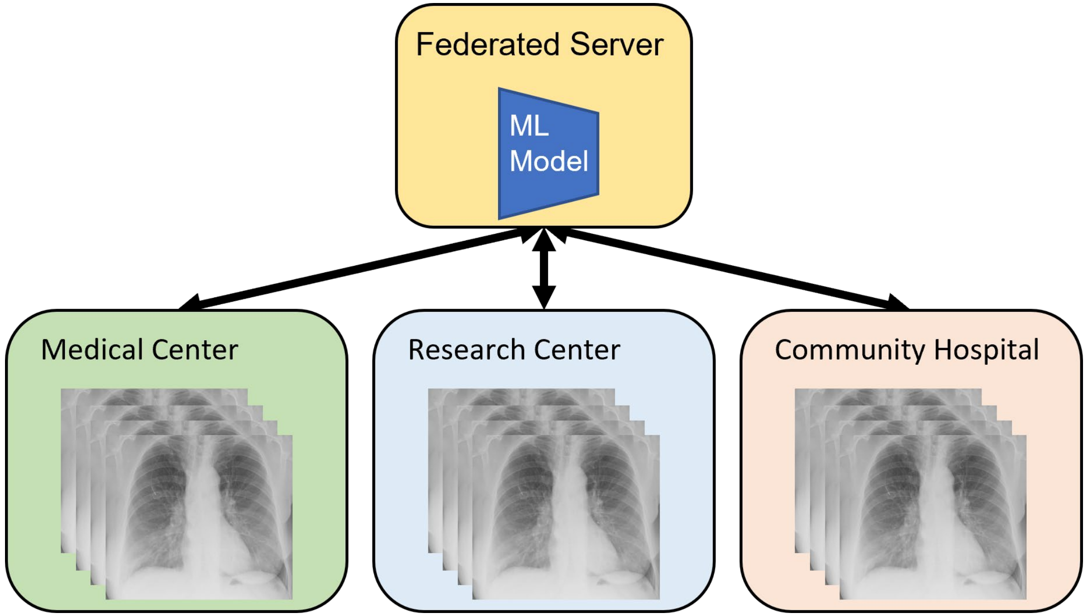


Purpose

- Investigate a federated learning algorithm, ***Fed****erated Learning with* ***S****hared* ***L****abel* ***D****istribution* (**FedSLD**), for classification task, under a cross-silo (medical institutions) setting

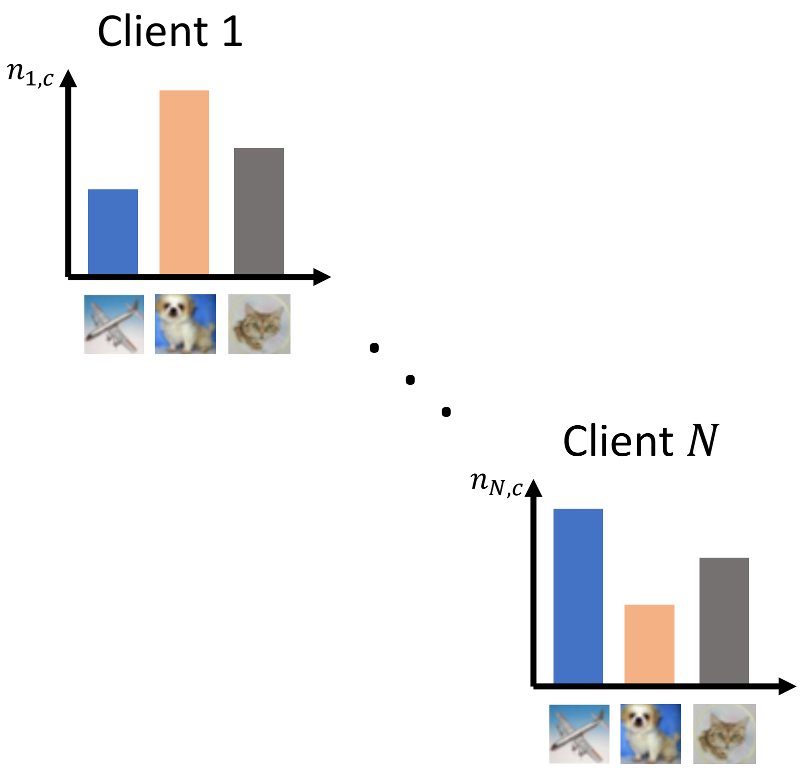
- Focus on the data heterogeneity challenge of federated learning, assuming legitimate for the clients to share the number of samples in each class

- Evaluate the proposed algorithm on four datasets under two kinds of non-IID data distributions



Method

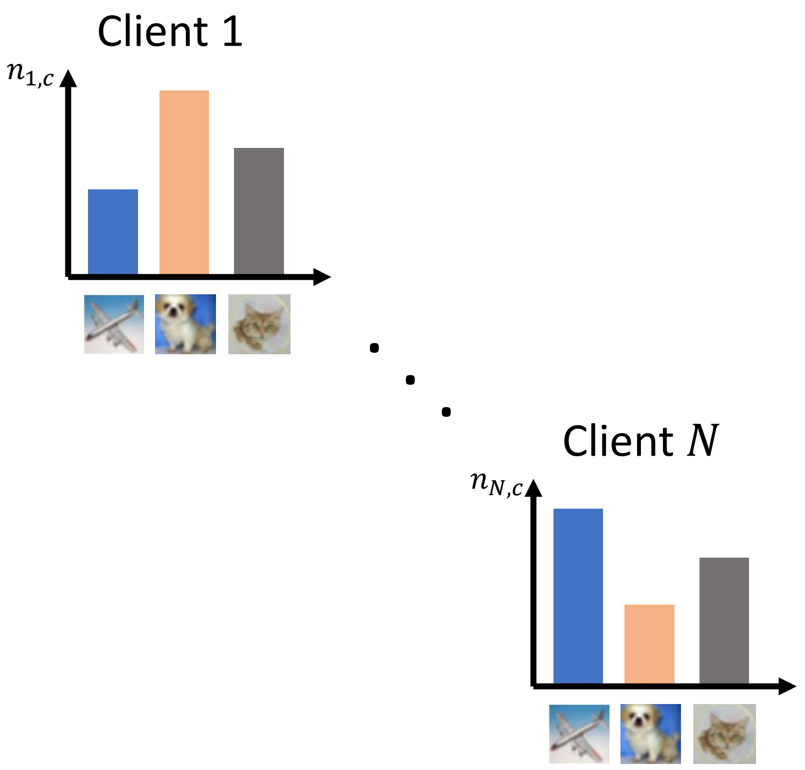
- Assumption
  - FedAvg [1] assumption
    - Weighted sum of local empirical risks
    - Weights are often $n_i / \sum_j n_j$
    - Assumes knowledge of number of samples
  - FedSLD
    - Assumes knowing number of samples in each class
      - This assumption usually holds true for cross-silo FL, including medical setting
      - Estimate of label distribution



Method

- Estimation of label distribution
  - Non-IID: $\mathcal{P}_i(x, y) \neq \mathcal{P}_j(x, y)$
  - By Bayes' theorem, $\mathcal{P}_i(x|y)\mathcal{P}_i(y) \neq \mathcal{P}_j(x|y)\mathcal{P}_j(y)$
  - Aggregate knowledge of #samples in each class, estimate $\mathcal{P}(y)$ by

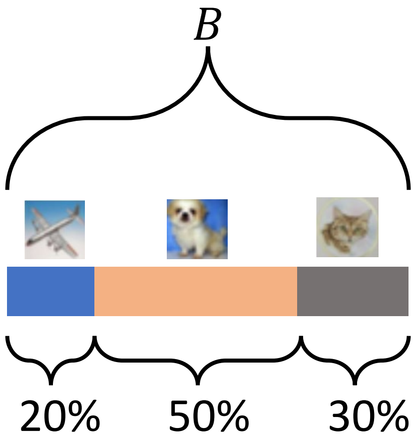
$$\tilde{\mathcal{P}}(y = c) = \frac{\sum_{i=1}^{N} n_{i,c}}{\sum_{i=1}^{N} n_i}$$



Method

- Compute the percentage of each class in each mini-batch
  - During local update, given a batch of data $\{(x_k, y_k)\}_{k=1}^{B}$ with B data samples, compute

$$p_b(y = c) = \frac{\sum_{k=1}^{B} [\![y_k = c]\!]}{B}$$



Method

- Weigh each data samples' contribution to the loss based on
  - The estimation of the prior of each class
  - The percentage of each class in each mini-batch

- Final loss of the mini-batch

$$\mathcal{L}_b\left(\{(x_k, y_k)\}_{k=1}^{B}\right) = -\sum_{k=1}^{B}\left(\frac{p_b(y = y_k)}{\tilde{\mathcal{P}}(y = y_k)} \cdot \sum_{c=1}^{C} y_{k,c} \log\left(f_i(x_k)\right)_c\right)$$

- Aggregate the model at the end of each training round as in FedAvg

Algorithm 1 FedSLD.

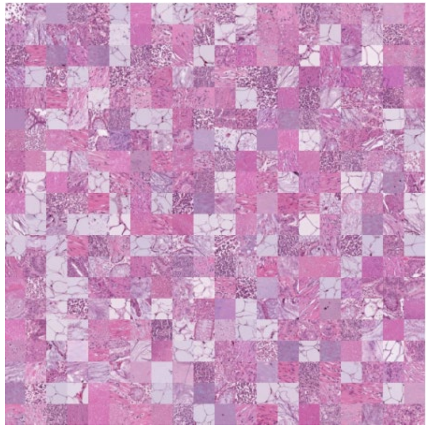
Input: Initialized model parameter weights w^0, number of clients N, number of local epochs E, batch size B, is the batch size, learning rate η, number of rounds R.

```
∀i ∈ [N], c ∈ [C], acquire n_{i,c}, client i's numbers of samples of each class c.
∀c ∈ [C], P̃(y = c) = Σ_{i=1}^N n_{i,c} / Σ_{i=1}^N n_i  // compute estimated prior label distribution.
for r ← 1, 2, ..., R do
    ∀i ∈ [N] w_i^r = w^{r−1} // broadcast model parameters.
    for i ← 1, 2, ..., N in parallel do
        for {x_k, y_k}_{k=1}^B in all minibatches do
            ∀c, p_b(y = c) ← Σ_{k=1}^B [[y_k = c]]/B
            Compute loss L_b by Equation 3.
            w_i^r ← w_i^r − η∇_w L_b
        end for
    end for
    w^r = Σ_{i=1}^N (n_i/n) w_I^r // aggregate model updates
end for
return w^R
```

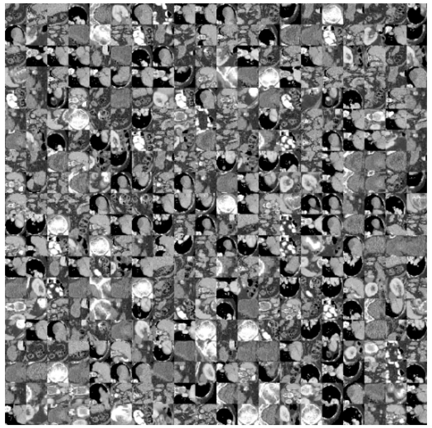
Experiments

- Datasets
  - Two benchmark datasets
    - MNIST
    - CIFAR10
  - Two medical imaging datasets from MedMNIST [2] collection
    - OrganMNIST (axial) (11-class liver tumor images)
    - PathMNIST (9-class colorectal cancer images)

PathMNIST



OraganMNIST (axial)



[2] Yang, Jiancheng, Rui Shi, and Bingbing Ni. "Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis." 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021.

Experiments

- Two non-IID settings
  - Pathological non-IID
    - Randomly select 2 classes for each client
    - In each class, assign a random number of images
  - Practical non-IID
    - Randomly partition each class of the dataset into 12 shards (10 x 1%, 1 x 10%, 1 x 80%)
    - Randomly assign one shard from each class to each client
    - Allows each client to have images from all classes, with more images from some classes while less from others
    - A simulation that is closer to real-world medical applications

- Compared baselines
  - FedAvg
  - FedProx [3]

[3] Li, Tian, et al. "Federated optimization in heterogeneous networks." Proceedings of Machine Learning and Systems 2 (2020): 429-450.

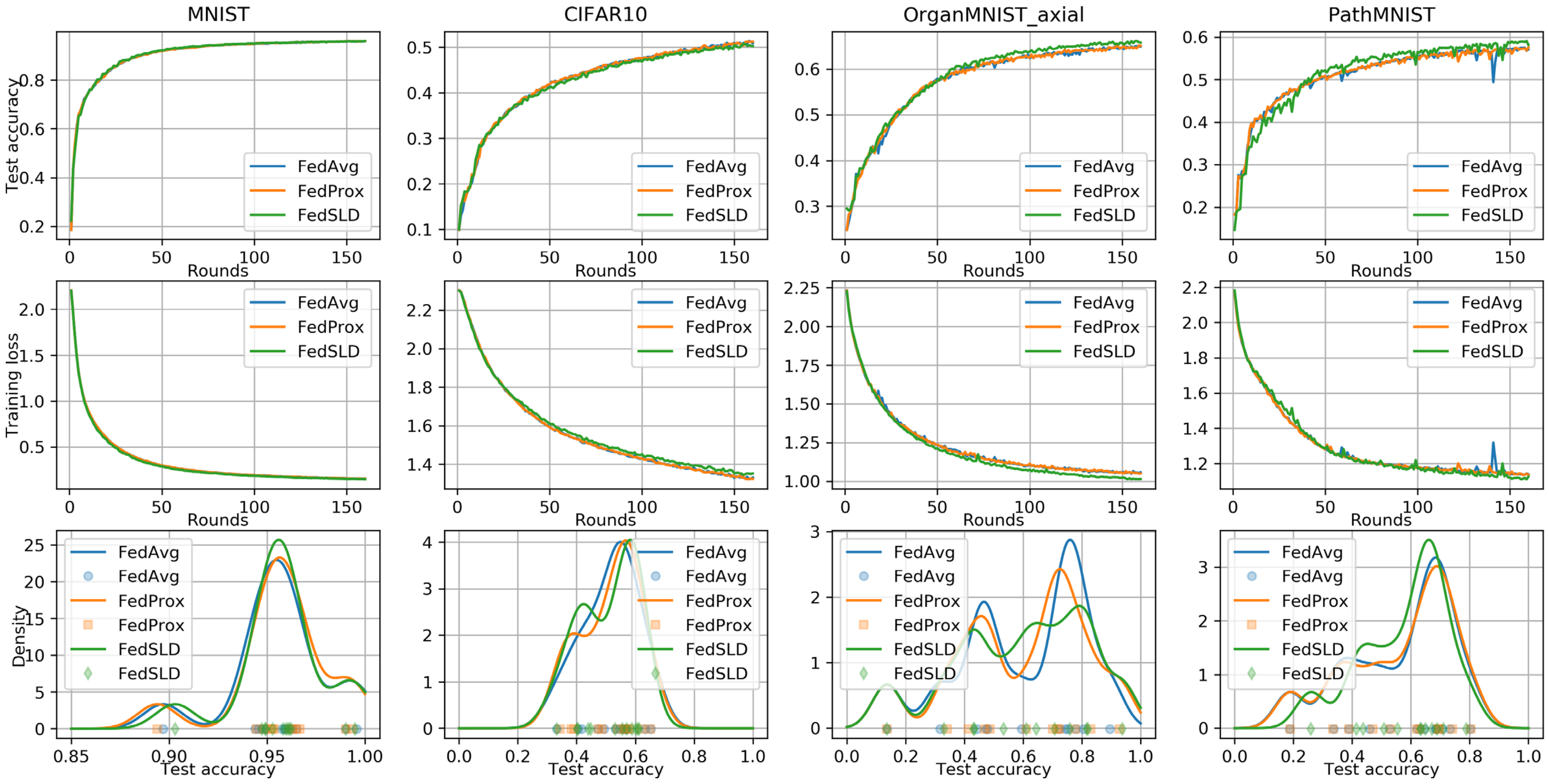
Experiments

- Evaluation metrics
  - Numerical metrics: two types of test accuracies
    - ***B**est **M**ean **C**lient **T**est **A**ccuracy (**BMCTA**)*
      - Mean over all clients
      - Best over all rounds
    - ***B**est **T**est **A**ccuracy (**BTA**)*
      - Computed the highest test accuracy for the combined test set from each client
  - Plots
    - Training loss curve
    - Test accuracy curve
    - For fairness, density estimation on the clients' test accuracies
      - Higher density at higher accuracy reflects better result

Results

- Pathological non-IID results

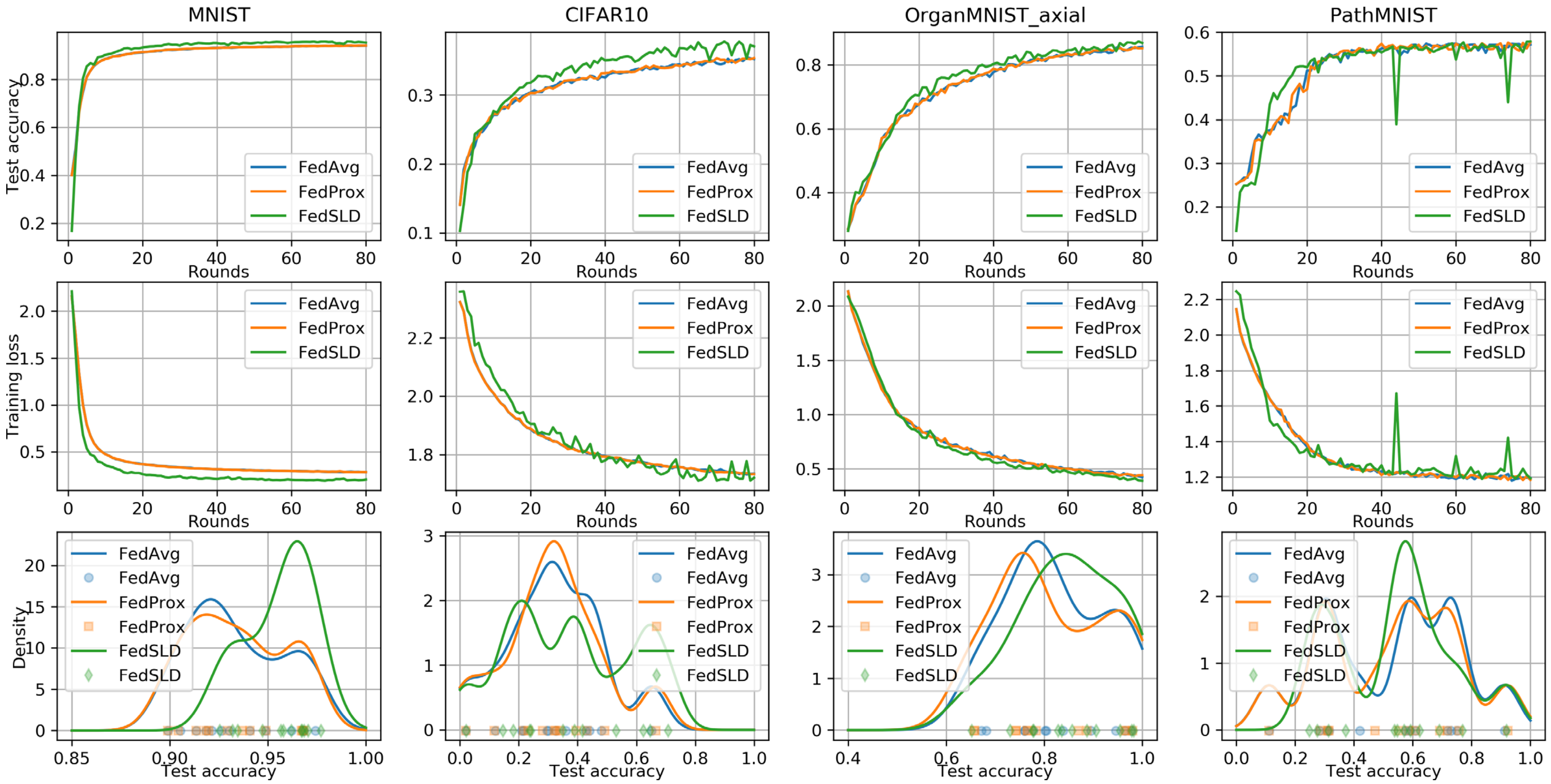
BMCTA/BTA	MNIST	CIFAR10	Organ-MNIST	Path-MNIST
FedAvg	95.60/95.92	**51.50/51.39**	59.52/64.99	95.60/95.92
FedProx	95.71/95.98	51.39/51.24	59.44/65.10	95.71/95.98
FedSLD (Ours)	**95.74/96.03**	50.81/50.71	**59.70/66.13**	**95.74/96.03**



Results

- Practical non-IID results

BMCTA/BTA	MNIST	CIFAR10	Organ-MNIST	Path-MNIST
FedAvg	93.41/94.15	32.07/35.46	82.32/85.69	52.70/57.38
FedProx	93.45/94.20	31.98/35.38	81.53/85.54	52.77/57.72
FedSLD (Ours)	**95.56/95.85**	**37.48/37.79**	**84.75/84.75**	**53.87/57.90**



Discussions

- We designed a novel federated learning algorithm for medical image classification task, simulating a real-world cross-silo (medical institutions) setting.
  - Leverage the information of number of samples in each class as knowledge of clients’ label distribution
  - Weigh each sample’s contribution to the local empirical risk
  - Introduce a practical non-IID setting that aims to mimic real-world medical setting

- Results show that our FL algorithm outperforms the baselines in most cases on four datasets under two non-IID settings
  - Faster convergence and better performance
  - Reduced variance of clients’ test accuracy implies a more fair training

Conclusion

- Our work proposed a novel FL algorithm for classification tasks that aims to mitigate the negative influence of data heterogeneity in cross-silo medical applications.

- Our method demonstrates that leveraging the information in terms of the shared label distribution will produce a faster and better convergence, and encourage a fair training across all clients.

- As information regarding the dataset at medical silos is used, the proposed FedSLD can perform better on heterogenous data for federated learning in medical domains.

Acknowledgements:

Intelligent **C**omputing for **C**linical **I**maging (**ICCI**) Lab, University of Pittsburgh

- NIH/NCI #1R01CA218405, a NSF grant (CICI:SIVD:2115082), the grant 1R01EB032896 as part of the NSF/NIH Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science Program, and a Pilot Research Project from the Scaling Grant of the Pitt Momentum Funds for the Pittsburgh Center for AI Innovation in Medical Imaging.
- Extreme Science and Engineering Discovery Environment (XSEDE), supported by NSF grant number ACI-1548562, NSF award number ACI-1928147, at the Pittsburgh Supercomputing Center (PSC).
- Pilot Research Project of the Pittsburgh Center for AI Innovation in Medical Imaging and the associated Pitt Momentum Funds of a Scaling grant from the University of Pittsburgh (2020)

Thank you!

Jun Luo
jul117@pitt.edu