Challenges and Practical Use Cases for Federated Learning in Medical Imaging

Qi Dou

Assistant Professor

Department of Computer Science and Engineering
The Chinese University of Hong Kong

FL Tutorial @ MICCAI 2024

Outline

1. What is Federated Learning?

2. Challenges and Practical Use Cases

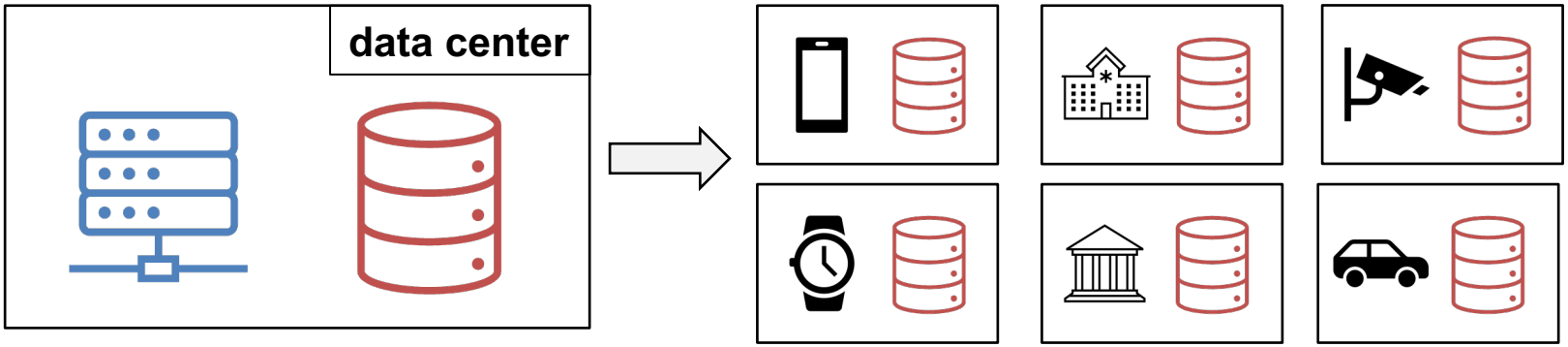
3. Summary

1. What is Federated Learning?

From Centralized to Decentralized Data

- The standard setting in Machine Learning / Depp Learning considers a centralized dataset processed in a tightly integrated system

- However, real-world data is often decentralized across many parties



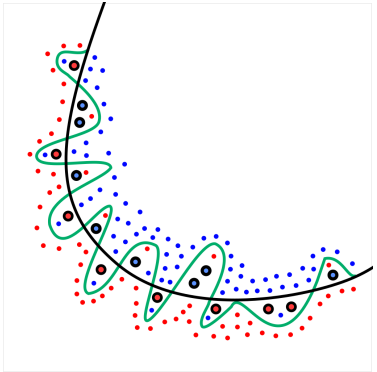
Challenges of Using Centralized Data

- Sending the data is too costly
  - Self-driving cars could generate TBs of data a day
  - Wireless devices have limited bandwidth/power

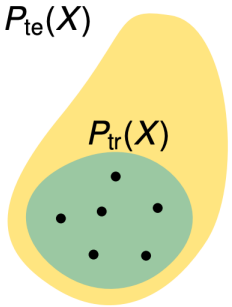
- Data itself is too sensitive
  - Growing public awareness on data privacy
  - Keeping control of data gives competitive advantages in business

Issues of Each Party Learning Locally

- Local dataset may be too small
  - Poor predictive performance (e.g., overfitting)
  - Non-statistically significant results



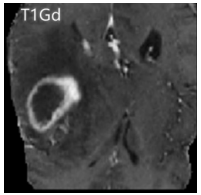
- Local dataset may be biased
  - Not representative of real data distribution



Medical AI Needs Large and Diverse Data

- **16,148** chest X-ray images and electronic medical record for COVID-19 clinical outcomes prediction
- **25,256** MRI scans for brain tumor segmentation
- **1.6 million** retinal images for disease diagnosis

[Nature Medicine, 2021]



[Nature Communications,2022]

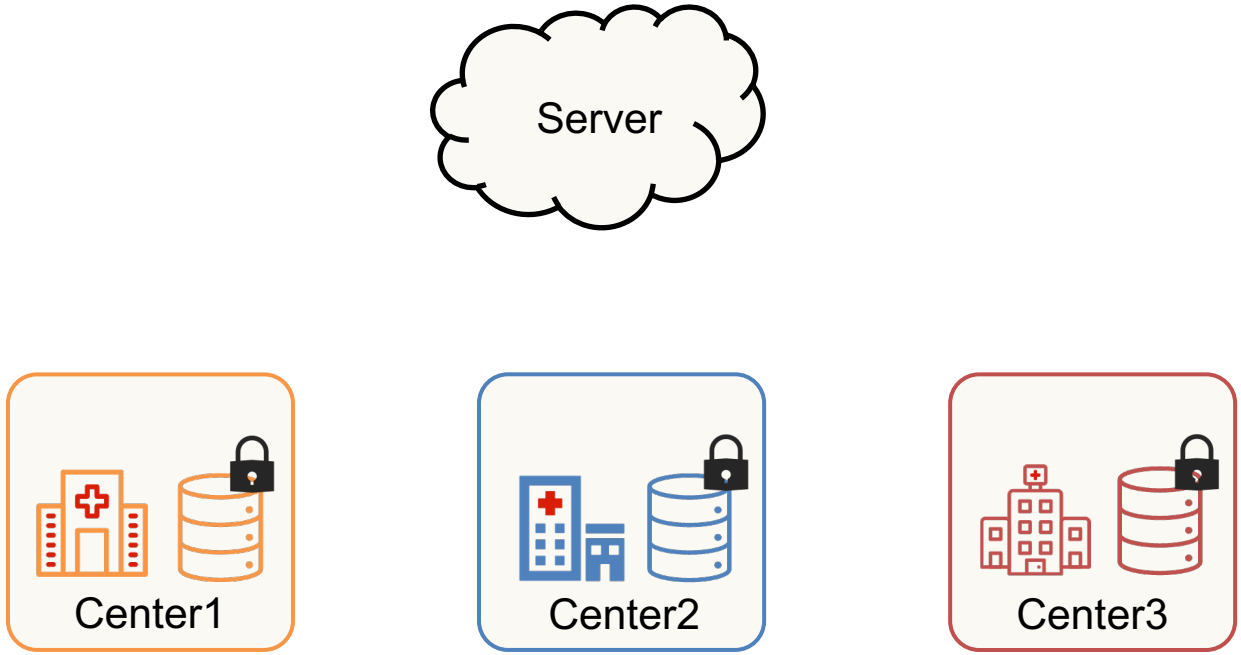
[Nature, 2023]

Collaborative development and validation is needed

- Joint effort to collect large-scale data
- Cover a wide variety of observations of diseases
- Present complexity as closer to real-world situations

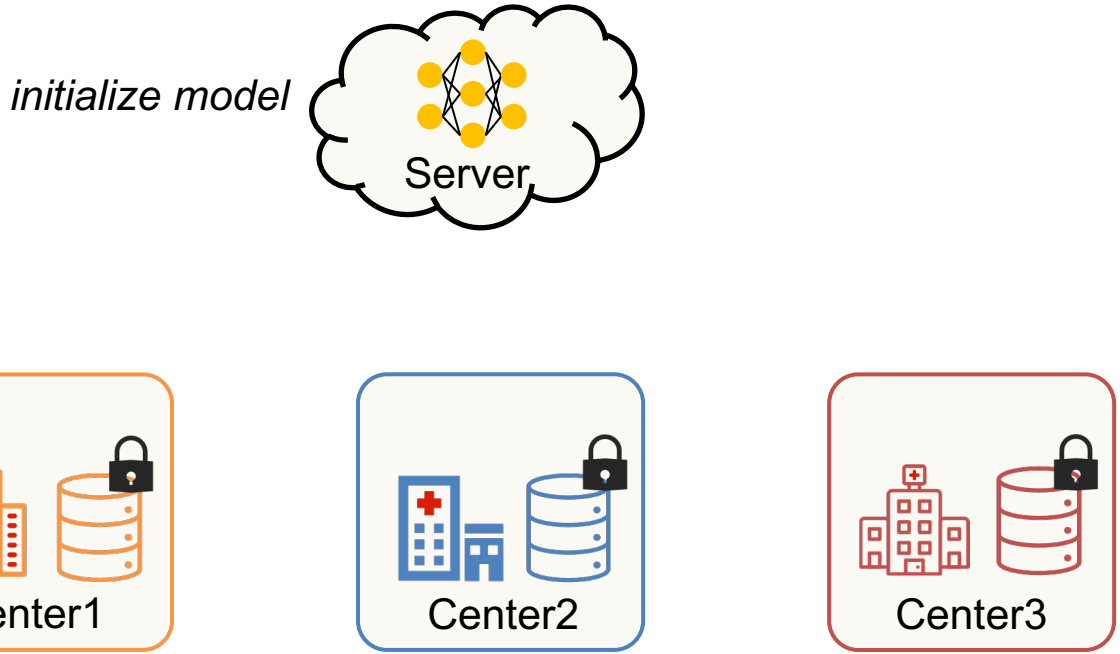
A Typical Paradigm of Federated Learning

- Federated Learning (FL) aims to collaboratively train a model while keeping the data decentralized



A Typical Paradigm of Federated Learning

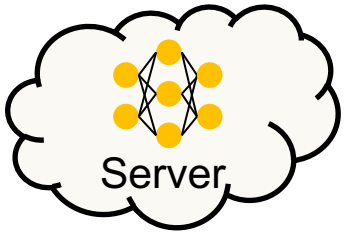
- Federated Learning (FL) aims to collaboratively train a model while keeping the data decentralized

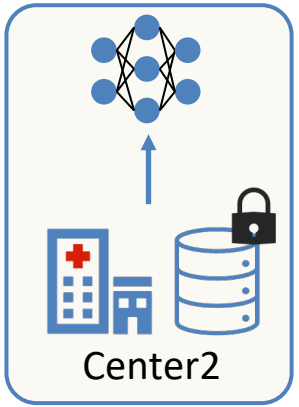


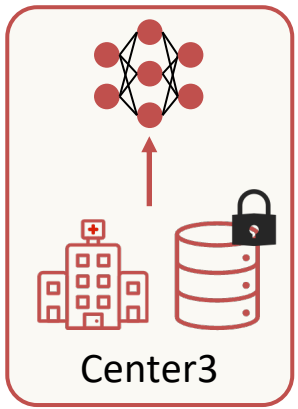
A Typical Paradigm of Federated Learning

- Federated Learning (FL) aims to collaboratively train a model while keeping the data decentralized

each center makes an update using its local dataset

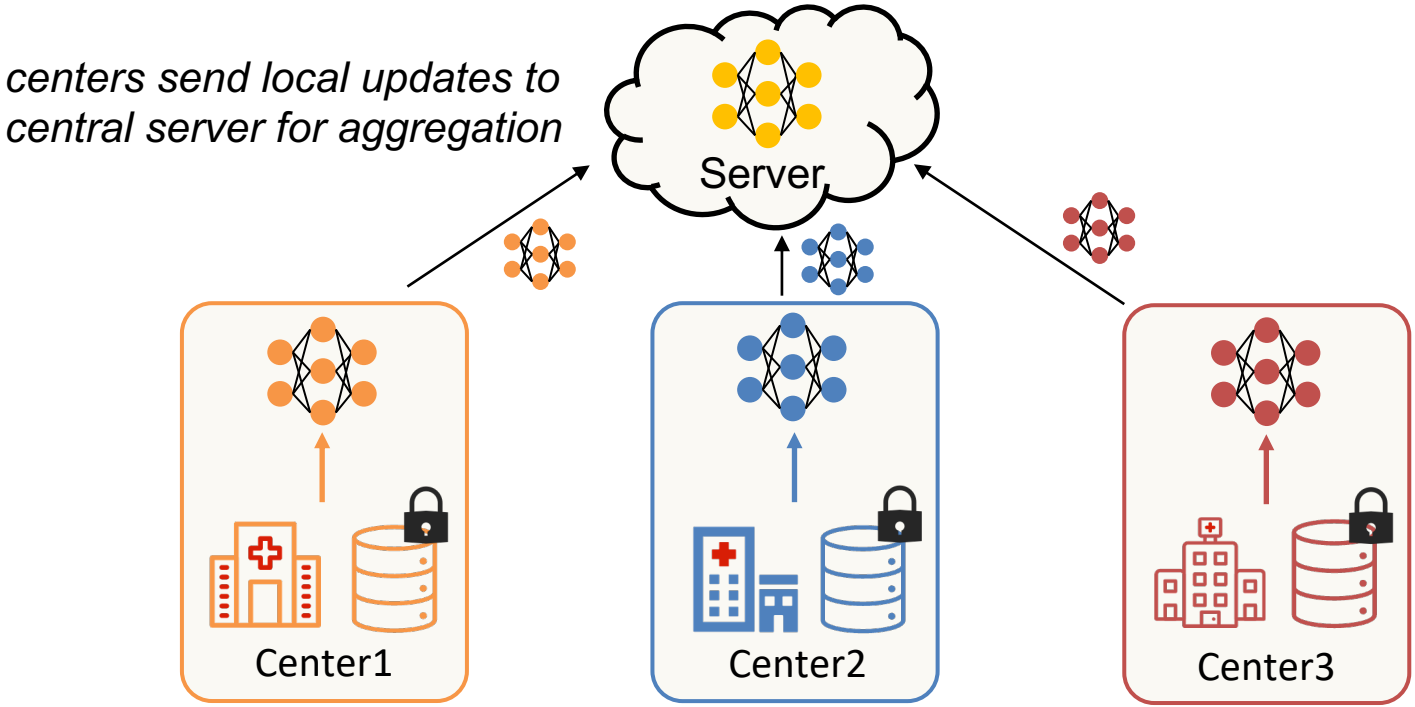






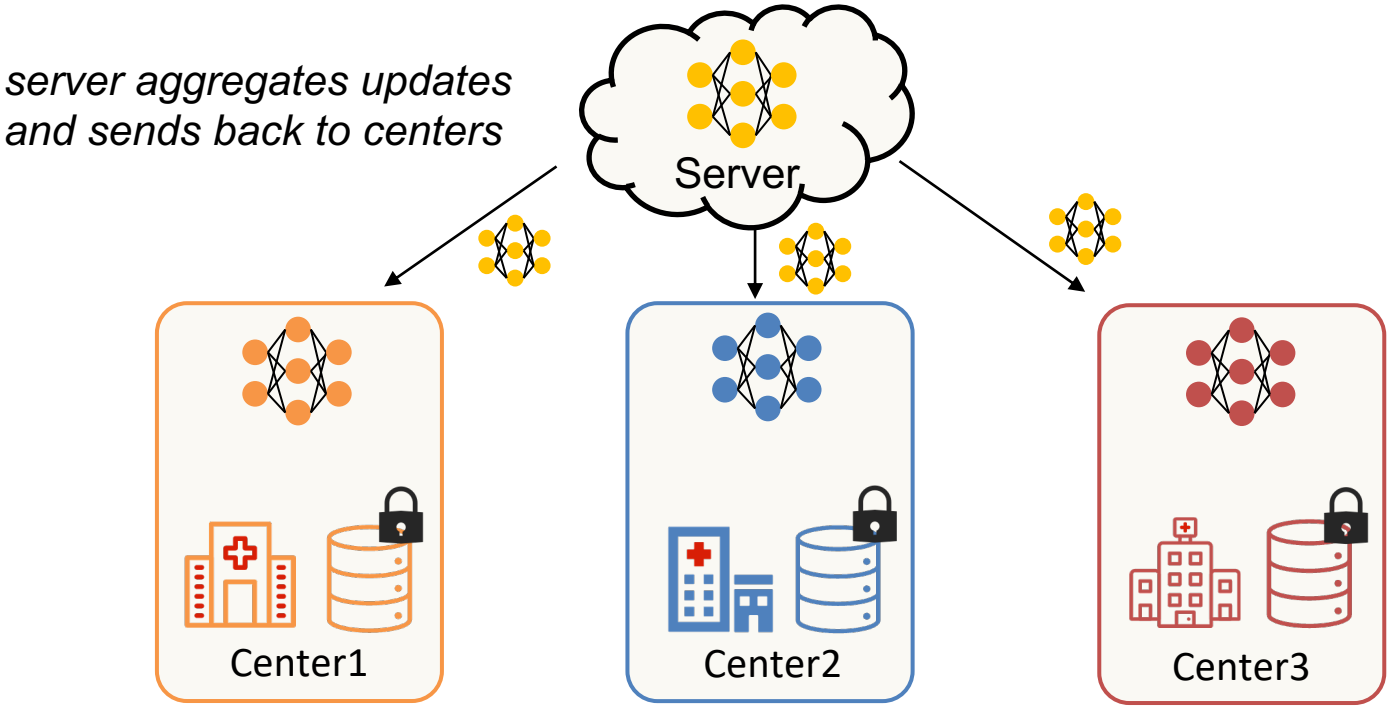
A Typical Paradigm of Federated Learning

- Federated Learning (FL) aims to collaboratively train a model while keeping the data decentralized



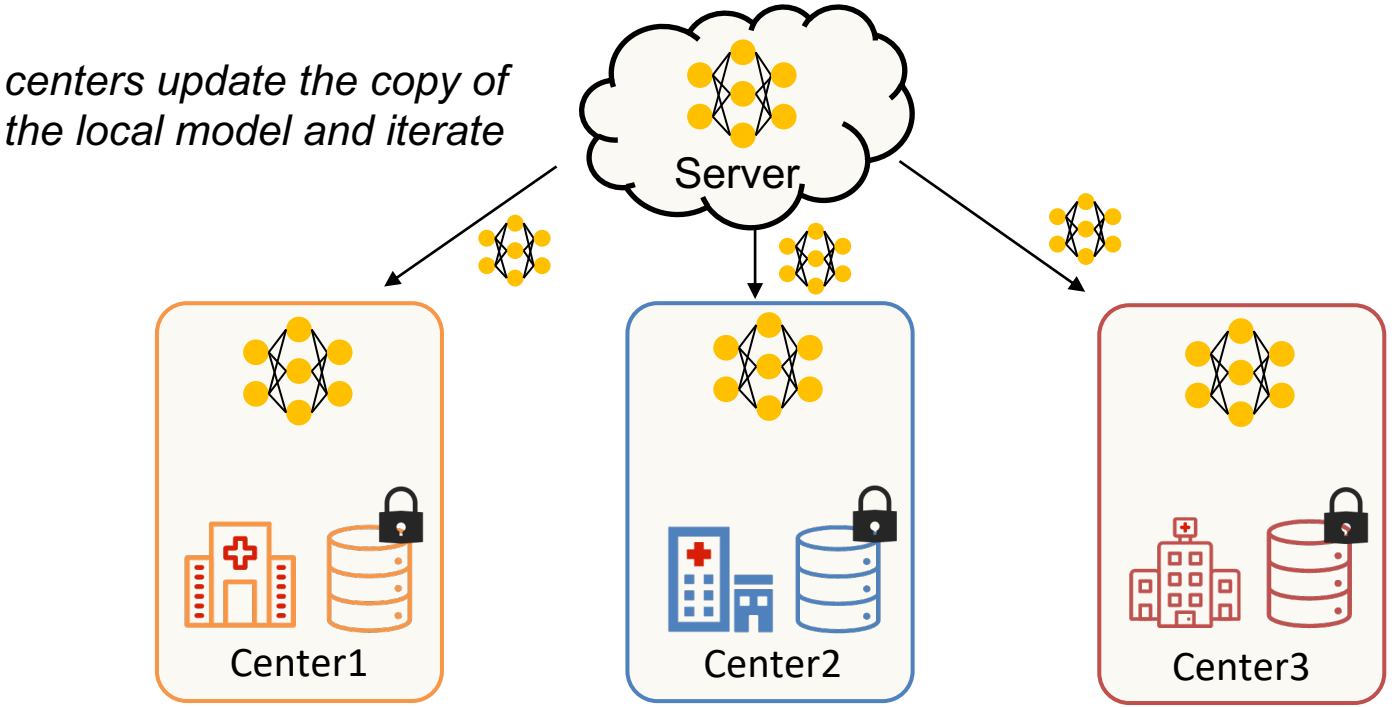
A Typical Paradigm of Federated Learning

- Federated Learning (FL) aims to collaboratively train a model while keeping the data decentralized



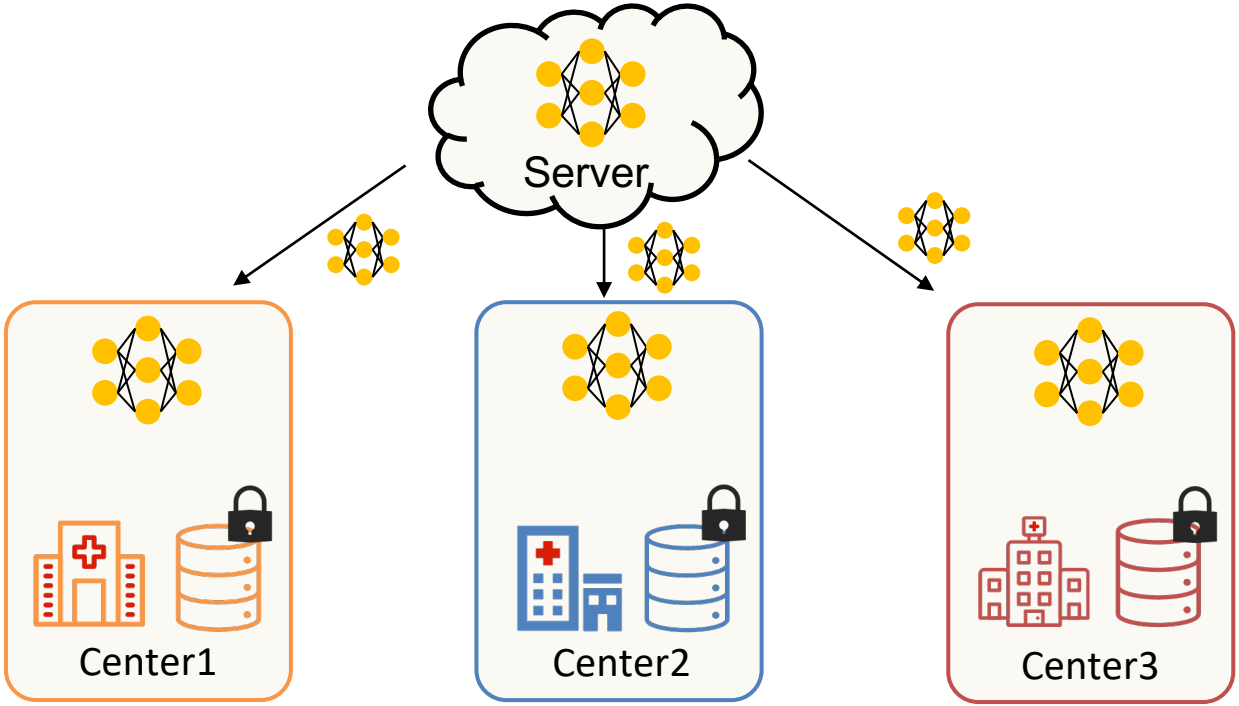
A Typical Paradigm of Federated Learning

- Federated Learning (FL) aims to collaboratively train a model while keeping the data decentralized

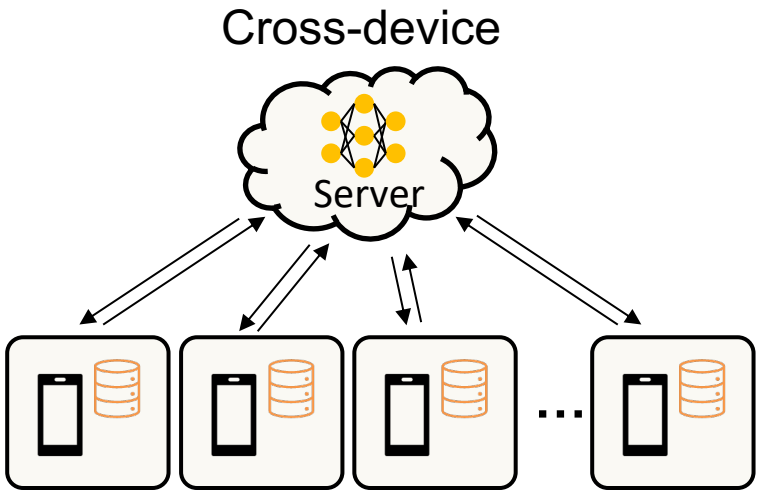


A Typical Paradigm of Federated Learning

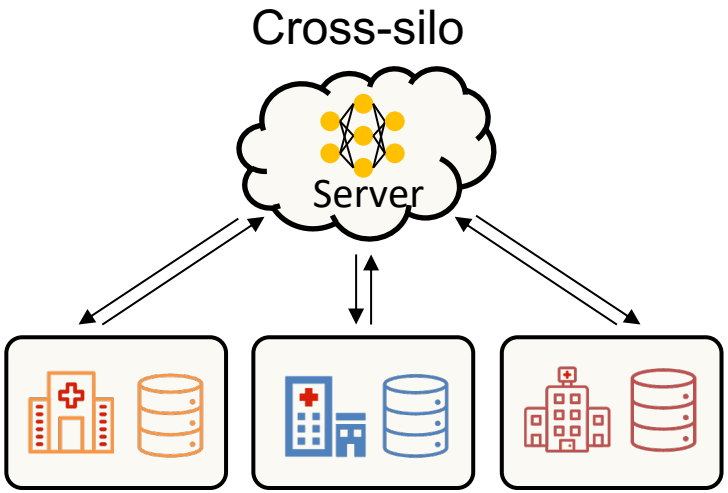
- The final model is expected to be **as good as the centralized solution** (ideally), or **better than what each center can learn on its own**



Cross-device FL v.s. Cross-silo FL



- Large number of clients (>>100)
- Small dataset per client
- Limited availability and reliability
- Some client may be malicious



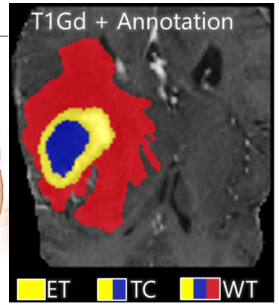
- 2-100 clients
- Medium to large dataset per client
- Reliable, almost always availability
- Clients are typically honest

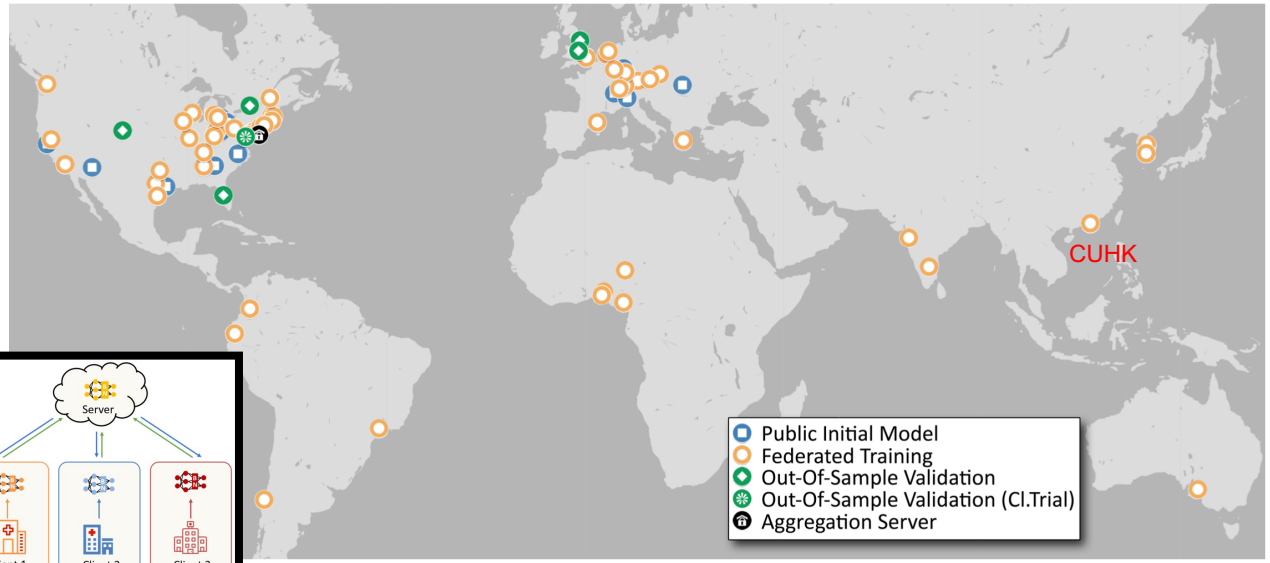
2. Challenges and Practical Use Cases

First Use Case

Federated learning – A Large-scale Global Study

- **Global study**: International collaboration of **71** institutions in **6** continents
- **Medical application:** Automated segmentation of glioblastoma (brain tumor) based on multi-sequence MR images





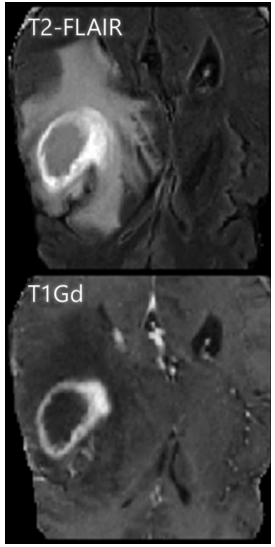
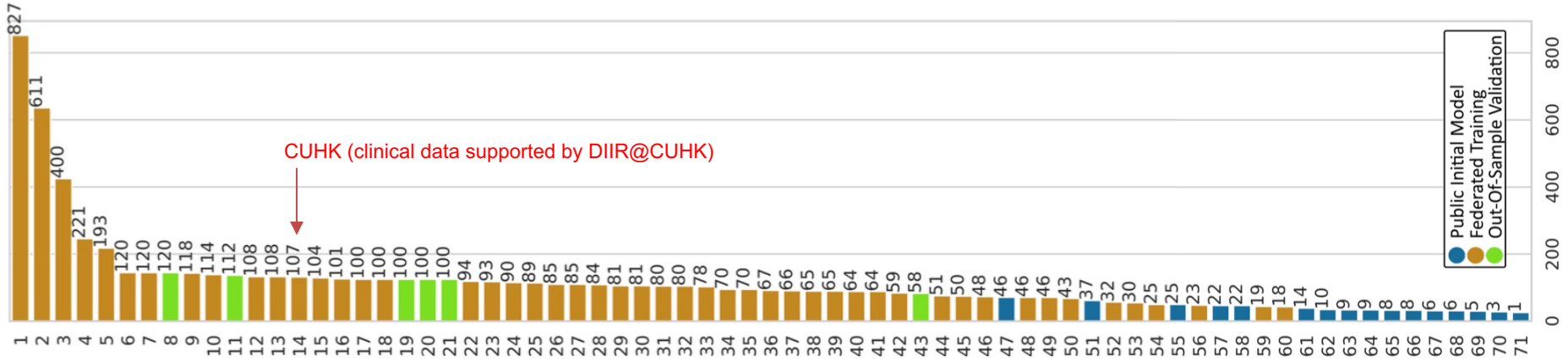


Image credit: https://www.mayoclinic.org/diseases-conditions/glioblastoma/cdc-20350148

First Use Case

Federated learning – A Large-scale Global Study

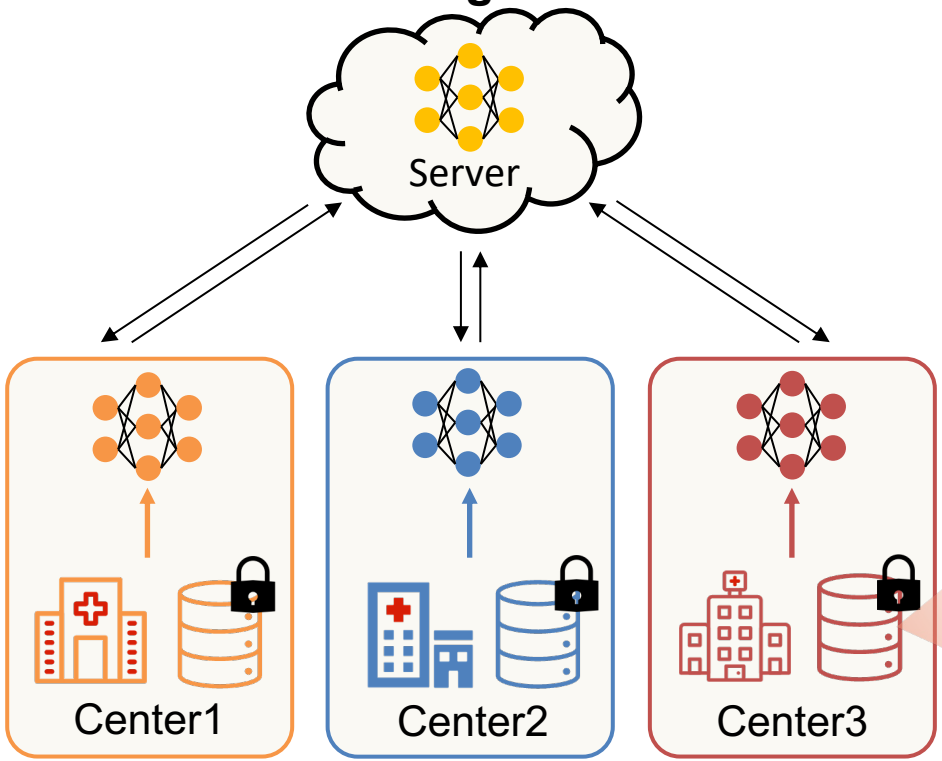
- **Largest-scale** real-world FL development in medical imaging field to-date
- A total of 25,256 MRI scans of **6,314** glioblastoma patients
- Scanners from various vendors (i.e., Siemens, GE, Philips, Hitachi, Toshiba)
- Male: Female ratio is 1.47: 1, age range is 7~94 years old



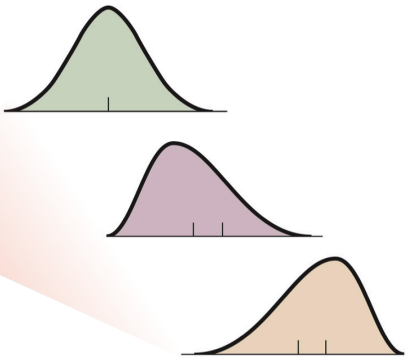
- Public Initial Model: 16 sites, 231 cases, for AI model (3D Residual-U-Net) initialization
- Federated Training: 49 sites, 5493 cases, 80% training, 20% local validation for client AI model
- Out-of-Sample Validation: 6 sites, 590 cases, did not join training, for model generalizability testing only

Challenges in Federated Learning : Heterogeneity

Federated Training

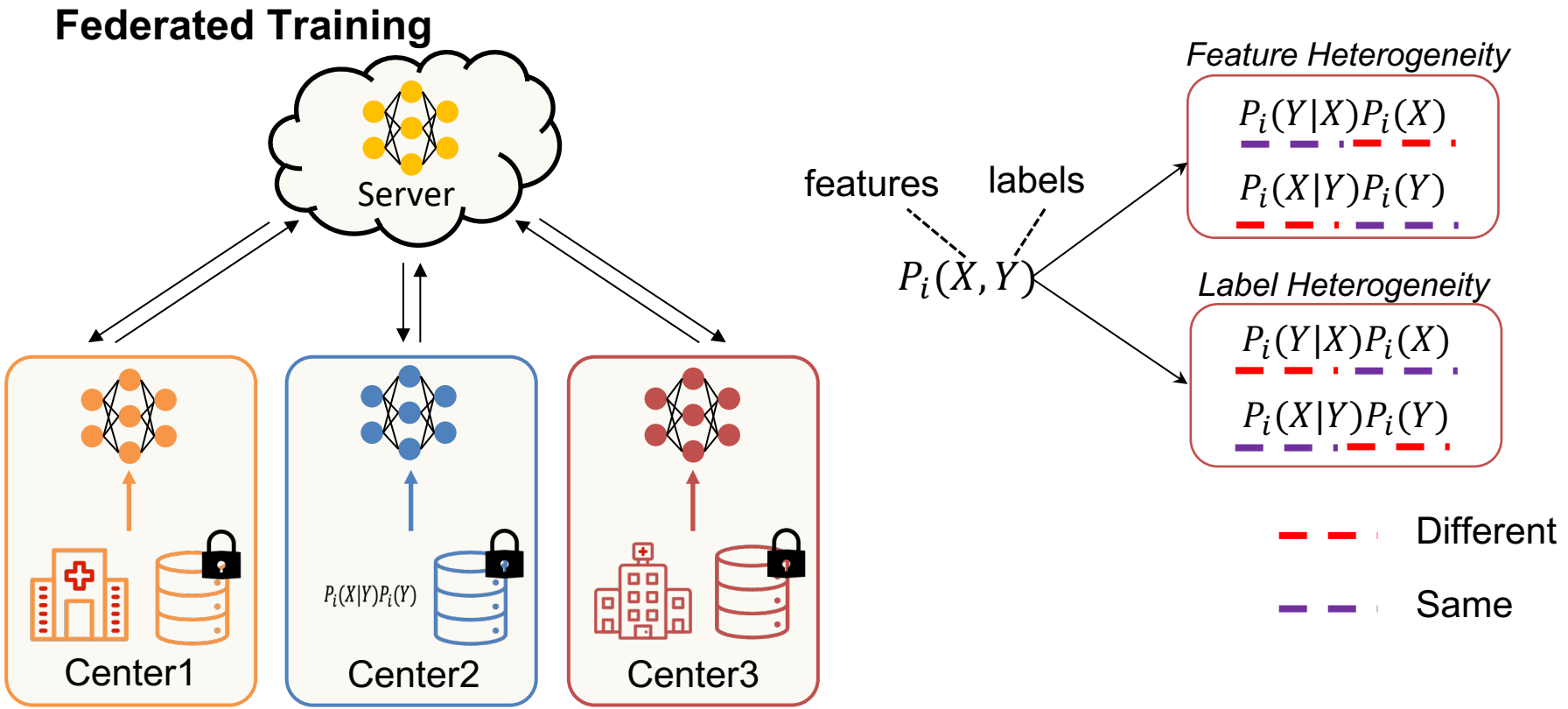


[1]

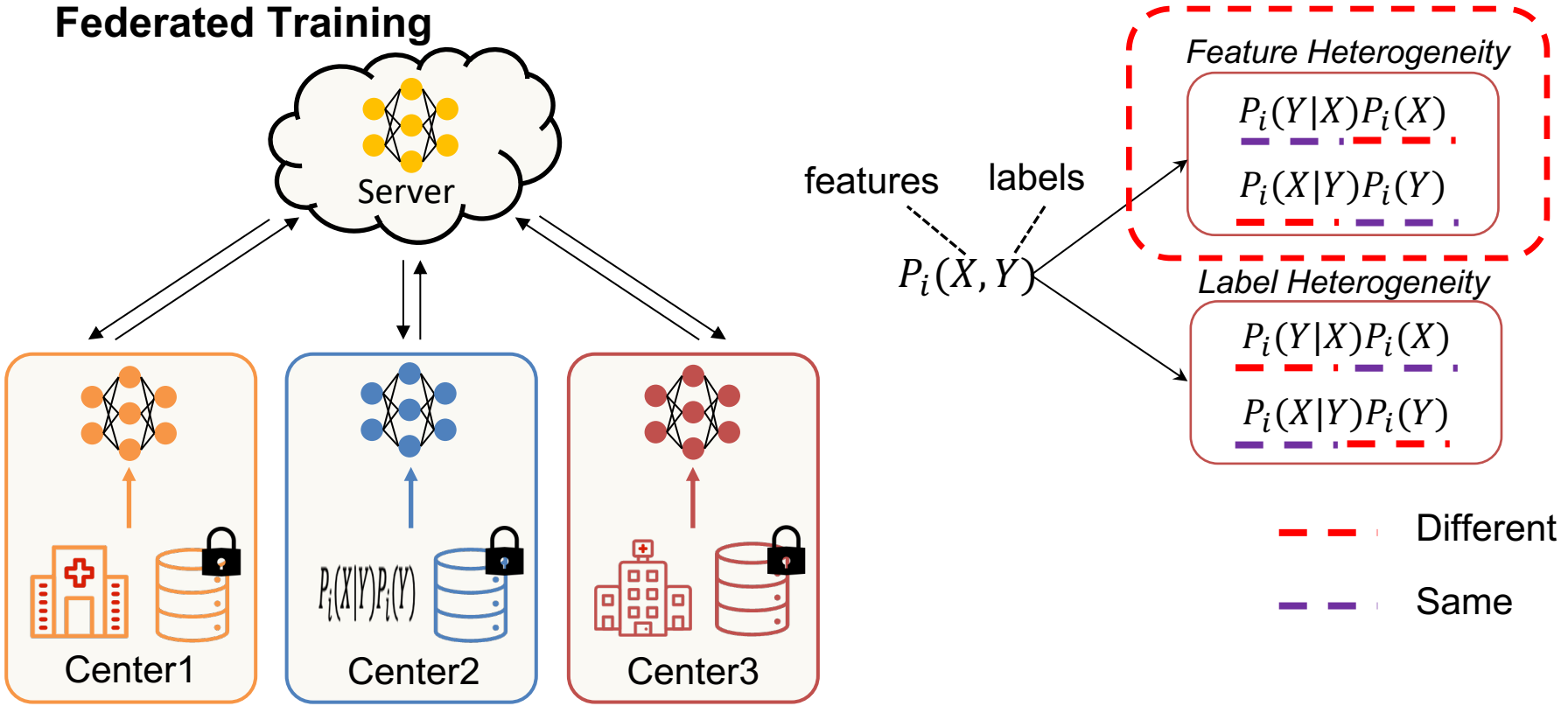


[1] Ye, M., Fang, X., Du, B., Yuen, P. C., & Tao, D. (2023). Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, *56*(3), 1-44.

Challenges in Federated Learning : Heterogeneity

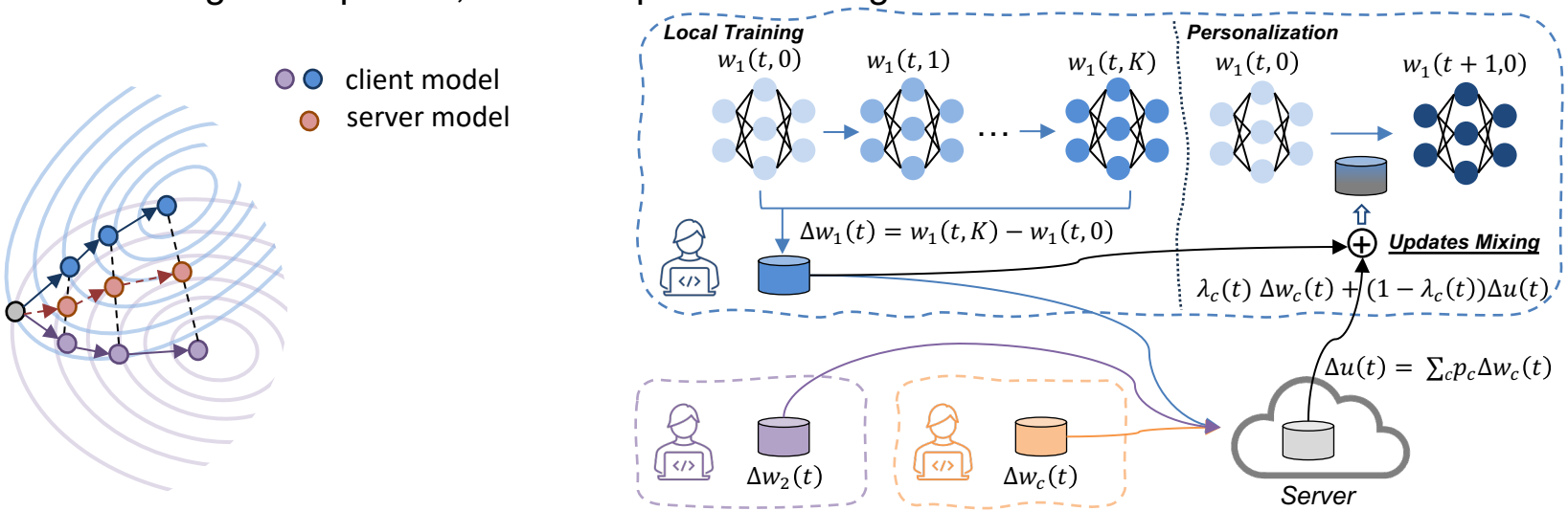


Challenges in Federated Learning : Heterogeneity



Overcoming Feature Heterogeneity by Personalization

- **Motivation:** A common global model may fail to fit various client data distributions. We aim to tackle feature heterogeneity by training personalized models.
- **Key idea:** Utilizing the local and global updates (naturally generated during FL training) to achieve personalization.
- **Method:** Mixing local and global updates via convergence ratio. By leveraging the convergence rate induced by Neural Tangent Kernel (NTK), we can assess the importance of local and global updates, and then perform mixing.



Jiang M, Le A, Li X, Dou Q. Heterogeneous Personalized Federated Learning by Local-Global Updates Mixing via Convergence Rate. *ICLR*, 2024.

Overcoming Feature Heterogeneity by Personalization

- **Mixing-ratio by NTK-Convergence**

  Evolution of prediction error for one step GD: $y - y(t+1) = (I - \eta H(t))(y - y(t))$

  error: $\xi(t)$

Proposition 1. *With the assumption that the error vector $\xi(t) := y - y(t)$ can be regarded as a random vector distributed uniformly in the space, by decomposing $H(t)$ and $\xi(t)$ into the eigenbasis of $H(t)$, we note that in gradient descent, the prediction error has an approximate convergence rate of $(1 - 2\eta\, tr(H(t))/n)$, where η is the learning rate and is small, and $tr(H(t))$ is the trace of $H(t)$.*

For client *c*, we can calculate the mixing ratio as:

$$\lambda_c(t) = \mathrm{tr}(H_c(t)) \;/\; (\mathrm{tr}(H_c(t)) + \mathrm{tr}(H_u(t)))$$

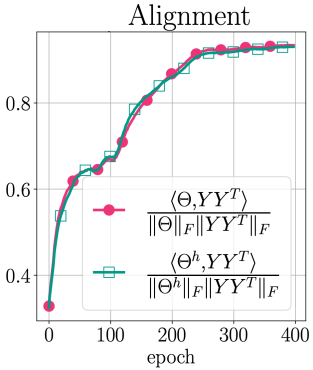
Jiang M, Le A, Li X, Dou Q. Heterogeneous Personalized Federated Learning by Local-Global Updates Mixing via Convergence Rate. *ICLR*, 2024.

Overcoming Feature Heterogeneity by Personalization

- **Approximate Trace Calculation**

  The H matrix will quickly grow with the dimension of the sample number, computing H matrix for practical networks is extremely challenging.

  The empirical NTK with entries of last-layer features highly aligns with the original NTK in terms of training dynamics. *[*Seleznova, Mariia, et al. "Neural (Tangent Kernel) Collapse." *arXiv:2305.16427* (2023).*]*



We approximate by considering the last layer dynamics only:

$$\mathrm{tr}(\hat{H(t)}) = \frac{1}{m} \sum_{i \in [n]} \sum_{r \in [m]} h_r(x_i) h_r(x_i) = \frac{1}{m} \times \mathrm{tr}(\Sigma(t))$$

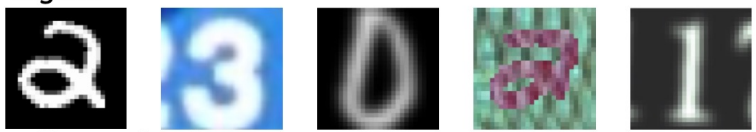
$$\text{where } \Sigma(t) = \underbrace{\sum_{i=1}^{n} h(x_i) h(x_i)^T}_{\text{feature matrix}}$$

Overcoming Feature Heterogeneity by Personalization

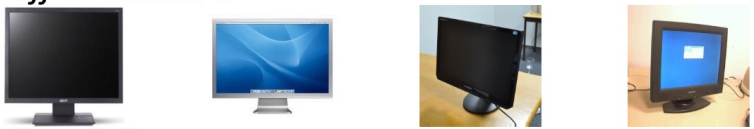
Experimental evaluations

- Classification on three natural image datasets

Digits5



Office-Caltech10

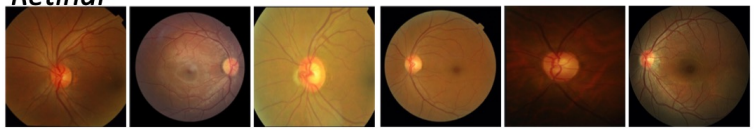


DomainNet

- Classification and segmentation on two medical image datasets

Camelyon

Retinal



Jiang M, Le A, Li X, Dou Q. Heterogeneous Personalized Federated Learning by Local-Global Updates Mixing via Convergence Rate. *ICLR*, 2024.

Overcoming Feature Heterogeneity by Personalization

Quantitative comparison with other FL methods - classification

Dataset	Digits5						Office-Caltech10					DomainNet							Camelyon17					
Client	A	B	C	D	E	Avg.	A	B	C	D	Avg.	A	B	C	D	E	F	Avg.	A	B	C	D	E	Avg.
FedAvg (PMLR 2017)	98.85 (0.03)	89.95 (0.09)	95.82 (0.48)	99.28 (0.0)	88.70 (0.37)	94.52 (0.14)	78.18 (2.36)	56.99 (1.69)	51.04 (6.51)	61.58 (3.91)	61.95 (1.13)	58.17 (0.0)	35.06 (1.65)	53.61 (2.52)	51.97 (0.99)	67.05 (1.95)	54.15 (0.72)	53.34 (0.53)	95.44 (1.05)	92.20 (0.62)	93.96 (0.88)	97.21 (0.27)	97.71 (0.33)	95.30 (0.28)
APFL (Arxiv)	96.62 (2.47)	90.07 (0.53)	96.94 (0.89)	99.12 (0.15)	91.17 (2.31)	94.78 (0.43)	78.05 (3.21)	53.48 (0.86)	64.76 (17.04)	57.91 (13.46)	63.55 (0.42)	58.11 (1.35)	36.87 (2.0)	53.29 (1.8)	46.60 (7.54)	69.02 (3.02)	58.00 (0.81)	53.65 (1.61)	96.27 (0.29)	92.94 (1.23)	96.70 (0.4)	97.95 (0.13)	98.25 (0.15)	96.42 (0.23)
L2SGD (NeurIPS 2020)	98.87 (0.05)	89.99 (0.07)	96.00 (0.2)	99.29 (0.01)	88.84 (0.37)	94.60 (0.1)	78.18 (2.36)	56.99 (1.69)	51.04 (6.51)	68.25 (1.86)	63.62 (1.18)	59.00 (0.72)	34.60 (2.13)	53.88 (1.28)	50.33 (2.73)	69.82 (2.33)	56.02 (1.67)	53.94 (0.26)	96.67 (0.33)	93.02 (0.32)	94.71 (0.18)	97.55 (0.15)	97.68 (0.26)	95.93 (0.13)
FedAlt (ICML 2022)	99.20 (0.05)	90.51 (0.27)	98.26 (0.38)	99.32 (0.03)	92.36 (0.08)	95.93 (0.08)	77.31 (2.69)	55.95 (2.01)	44.79 (1.8)	75.14 (4.27)	63.30 (0.37)	59.82 (0.48)	37.14 (0.97)	58.47 (1.47)	58.57 (3.65)	72.45 (2.0)	54.27 (0.73)	56.79 (0.69)	98.10 (0.18)	95.51 (0.13)	98.41 (0.01)	98.80 (0.01)	98.74 (0.11)	97.91 (0.06)
PerFedAvg (NeurIPS 2020)	99.05 (0.06)	89.55 (0.12)	96.11 (0.11)	99.23 (0.02)	89.53 (0.27)	94.69 (0.1)	71.73 (1.39)	56.55 (2.2)	61.46 (4.77)	74.01 (3.53)	65.94 (0.61)	59.57 (1.52)	35.42 (0.99)	55.99 (1.06)	48.47 (0.31)	67.60 (0.92)	56.08 (2.64)	53.85 (0.39)	96.71 (0.52)	93.05 (0.41)	95.06 (0.47)	97.68 (0.4)	97.92 (0.2)	96.08 (0.3)
FedBN (ICLR 2021)	99.22 (0.16)	91.48 (0.2)	96.20 (0.11)	99.32 (0.01)	91.14 (0.58)	95.47 (0.16)	80.10 (0.91)	**58.18** (2.46)	79.17 (3.61)	83.05 (4.48)	75.13 (1.75)	58.17 (0.87)	36.94 (1.14)	55.61 (1.75)	69.10 (1.91)	73.25 (3.61)	53.31 (4.68)	57.73 (0.69)	96.65 (0.49)	92.84 (0.45)	94.22 (0.4)	97.55 (0.13)	97.60 (0.31)	95.77 (0.21)
FedFomo (ICLR 2021)	98.83 (0.04)	90.49 (0.44)	95.95 (0.22)	99.33 (0.04)	89.19 (0.38)	94.76 (0.15)	74.76 (0.9)	54.69 (1.09)	54.58 (2.01)	67.34 (2.84)	62.84 (0.69)	59.28 (2.37)	36.38 (1.07)	56.43 (1.58)	49.40 (4.8)	69.33 (2.19)	57.34 (1.09)	54.69 (1.35)	96.67 (0.34)	92.25 (0.51)	95.18 (0.51)	97.60 (0.16)	96.42 (0.47)	95.62 (0.13)
FedRep (ICML 2021)	98.86 (0.12)	90.35 (0.09)	95.99 (0.47)	99.53 (0.01)	89.15 (0.19)	94.78 (0.16)	78.53 (1.05)	57.74 (0.93)	56.25 (5.41)	67.23 (5.95)	64.94 (2.76)	60.08 (3.67)	36.33 (0.98)	56.36 (0.76)	47.80 (4.51)	67.98 (2.05)	58.78 (0.91)	54.56 (0.95)	97.07 (0.15)	93.64 (0.37)	96.79 (0.05)	98.12 (0.15)	98.28 (0.1)	96.78 (0.08)
FedBABU (ICLR 2022)	98.85 (0.04)	90.15 (0.17)	95.75 (0.39)	**99.53** (0.01)	88.74 (0.49)	94.60 (0.17)	77.84 (0.6)	57.74 (1.29)	56.25 (6.25)	67.23 (7.06)	64.76 (2.74)	60.71 (3.17)	37.14 (0.72)	56.26 (1.65)	44.63 (1.03)	68.50 (3.06)	59.12 (1.31)	54.40 (0.75)	96.69 (0.06)	92.93 (0.53)	94.30 (0.66)	97.53 (0.19)	97.41 (0.2)	95.77 (0.2)
FedHKD (ICLR 2023)	98.11 (0.48)	90.38 (0.19)	95.41 (0.88)	99.47 (0.09)	90.06 (0.87)	94.69 (0.36)	77.14 (1.31)	56.52 (1.46)	53.98 (0.34)	65.48 (3.04)	63.28 (0.72)	59.04 (0.46)	36.78 (1.37)	54.01 (1.25)	48.81 (1.29)	67.82 (1.35)	56.84 (1.29)	53.88 (0.41)	96.32 (0.3)	93.91 (0.76)	94.75 (0.62)	96.91 (0.72)	97.56 (0.26)	95.89 (0.24)
LG-Mix (Ours)	**99.29** (0.03)	**92.35** (0.17)	**98.66** (0.05)	99.41 (0.02)	**95.77** (0.05)	**97.10** (0.04)	**80.45** (0.6)	56.55 (0.68)	**86.46** (1.8)	**93.79** (2.59)	**79.31** (1.06)	**60.84** (0.38)	**37.20** (0.95)	**61.49** (0.74)	**78.07** (1.04)	**78.62** (0.67)	**60.23** (1.15)	**62.74** (0.12)	**98.77** (0.02)	**97.98** (0.09)	**98.93** (0.07)	**99.13** (0.09)	**98.95** (0.04)	**98.75** (0.05)

Jiang M, Le A, Li X, Dou Q. Heterogeneous Personalized Federated Learning by Local-Global Updates Mixing via Convergence Rate. *ICLR*, 2024.

Overcoming Feature Heterogeneity by Personalization

Quantitative comparison with other FL methods - segmentation

Retinal Fundus Image Segmentation														
Client	A	B	C	D	E	F	Avg.	A	B	C	D	E	F	Avg.
	Dice Coefficient (Dice) ↑							Hausdorff Distance (HD) ↓						
FedAvg (PMLR 2017)	83.95 (1.73)	83.00 (1.05)	81.15 (2.1)	87.98 (0.51)	67.60 (0.66)	91.07 (0.32)	82.46 (0.12)	7.71 (0.42)	4.77 (0.02)	8.35 (2.59)	4.19 (0.9)	68.14 (31.41)	2.38 (2.38)	15.92 (5.04)
APFL (Arxiv)	74.03 (10.16)	72.76 (9.94)	70.03 (10.41)	81.30 (6.63)	66.27 (3.93)	90.56 (0.75)	75.83 (6.55)	27.91 (18.2)	52.47 (0.1)	43.07 (27.87)	18.00 (15.41)	76.69 (42.1)	5.73 (5.73)	37.31 (22.22)
L2SGD (NeurIPS 2020)	84.46 (0.42)	83.73 (0.29)	82.20 (1.57)	88.40 (0.19)	68.90 (0.66)	**91.12** (0.28)	83.14 (0.39)	7.55 (0.11)	4.81 (0.02)	7.95 (2.59)	4.18 (1.5)	57.76 (15.55)	2.48 (2.48)	14.12 (2.84)
FedAlt (ICML 2022)	85.01 (2.03)	85.19 (0.91)	84.06 (1.93)	88.98 (0.67)	64.48 (2.46)	91.05 (0.36)	83.13 (0.57)	6.20 (0.39)	5.07 (0.02)	5.25 (0.54)	3.59 (0.3)	75.60 (43.05)	2.58 (2.58)	16.38 (7.22)
PerFedAvg (NeurIPS 2020)	85.87 (0.4)	84.55 (0.75)	84.74 (1.53)	88.75 (0.38)	67.91 (1.77)	91.10 (0.12)	83.82 (0.29)	7.32 (0.23)	4.54 (0.02)	7.79 (4.21)	3.82 (0.61)	73.58 (39.54)	2.47 (2.47)	16.59 (6.04)
FedBN (ICLR 2021)	84.77 (0.29)	83.26 (0.51)	83.88 (0.61)	88.45 (0.5)	67.03 (1.55)	91.07 (0.09)	83.08 (0.4)	7.60 (0.25)	4.82 (0.01)	5.69 (0.4)	2.95 (0.15)	63.34 (18.31)	2.70 (2.7)	14.52 (2.99)
FedFomo (ICLR 2021)	71.48 (5.94)	80.47 (0.75)	76.62 (5.67)	86.19 (1.01)	55.10 (2.3)	89.87 (0.71)	76.62 (2.66)	12.29 (2.37)	4.71 (0.06)	12.08 (3.07)	5.10 (0.4)	154.01 (25.43)	2.72 (2.72)	31.82 (3.47)
LG-Mix (Ours)	**89.25** (0.54)	**86.76** (0.31)	**85.86** (0.67)	**89.79** (0.08)	**83.95** (1.11)	90.86 (0.07)	**87.75** (0.17)	**4.43** (0.25)	**3.63** (0.01)	**4.53** (0.15)	**3.57** (0.09)	**6.38** (1.2)	**2.34** (2.34)	**4.15** (0.22)

Jiang M, Le A, Li X, Dou Q. Heterogeneous Personalized Federated Learning by Local-Global Updates Mixing via Convergence Rate. *ICLR*, 2024.

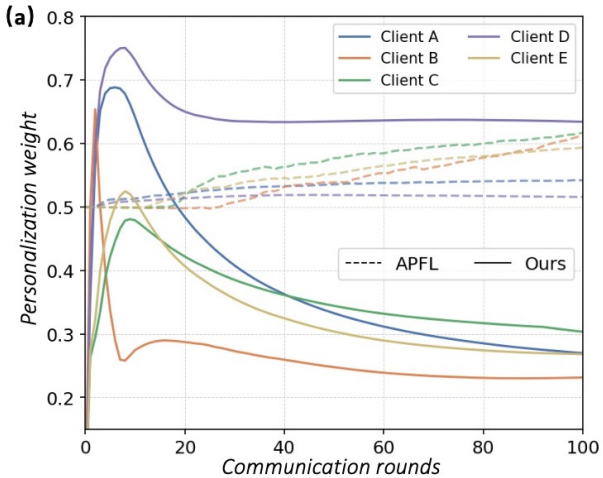
Overcoming Feature Heterogeneity by Personalization

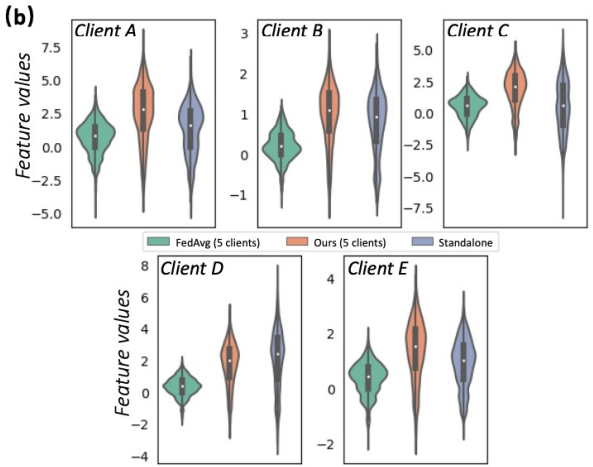
Analytical studies

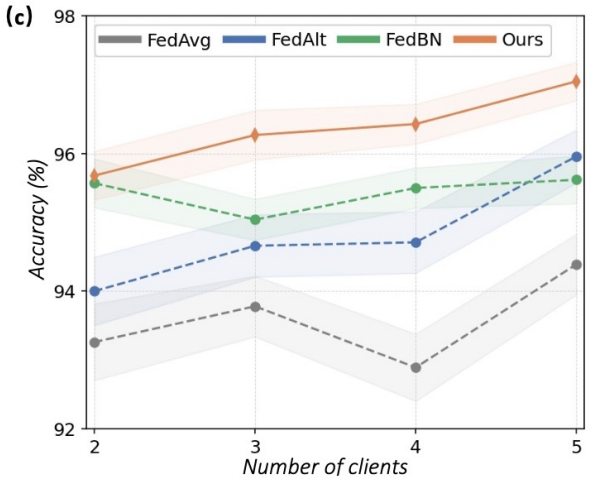
(a) Trend of the personalization weight.

(b) Feature value distribution by our personalized model.

(c) Client scalability study.

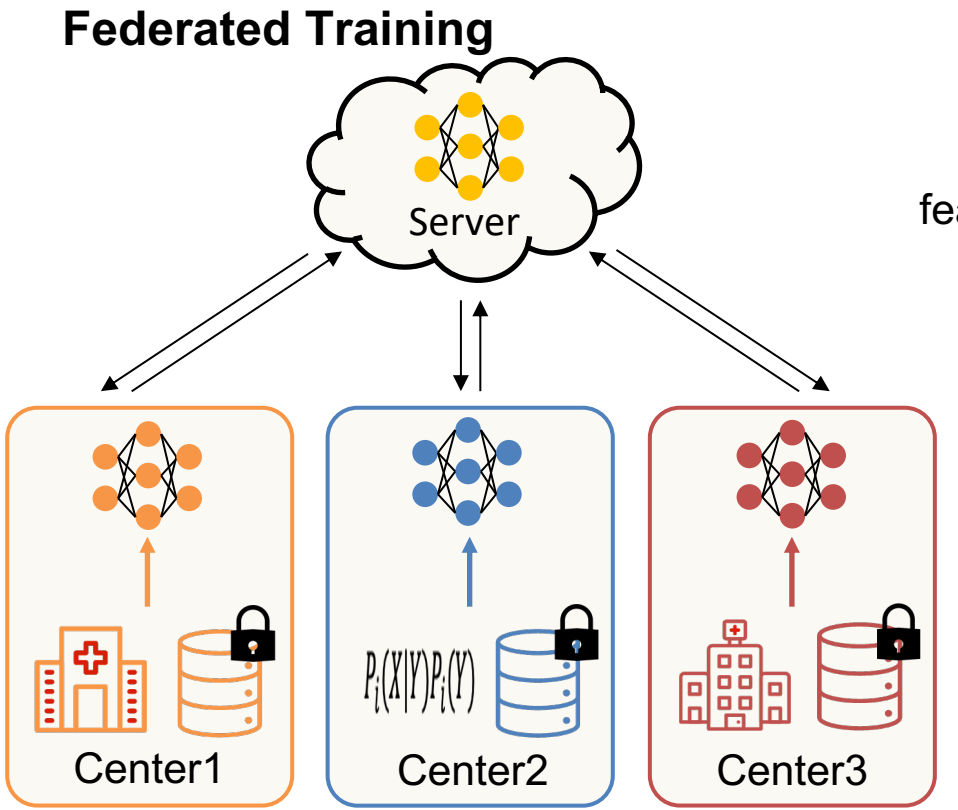


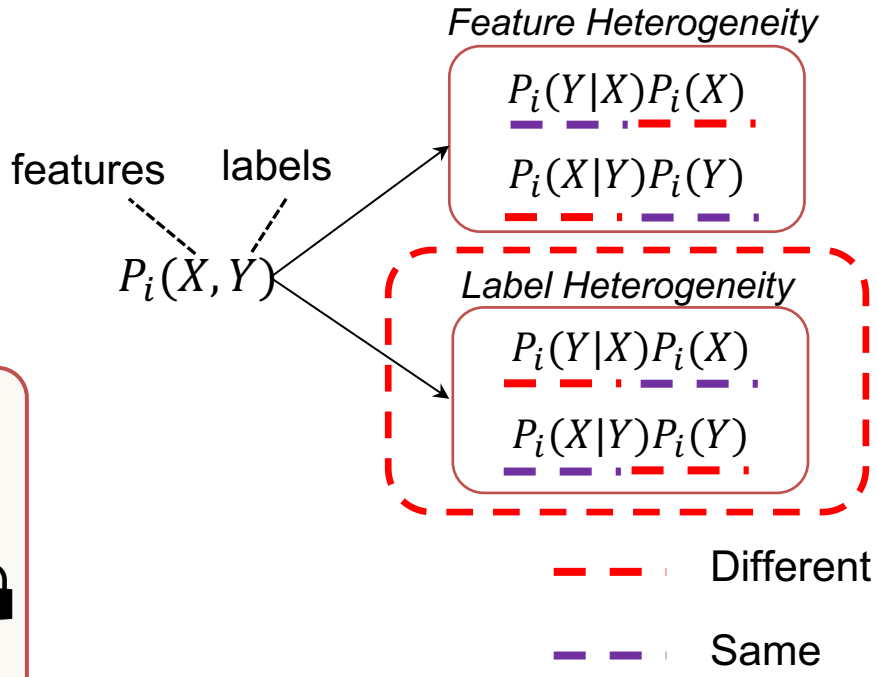




Jiang M, Le A, Li X, Dou Q. Heterogeneous Personalized Federated Learning by Local-Global Updates Mixing via Convergence Rate. *ICLR*, 2024.

Challenges in Federated Learning : Heterogeneity

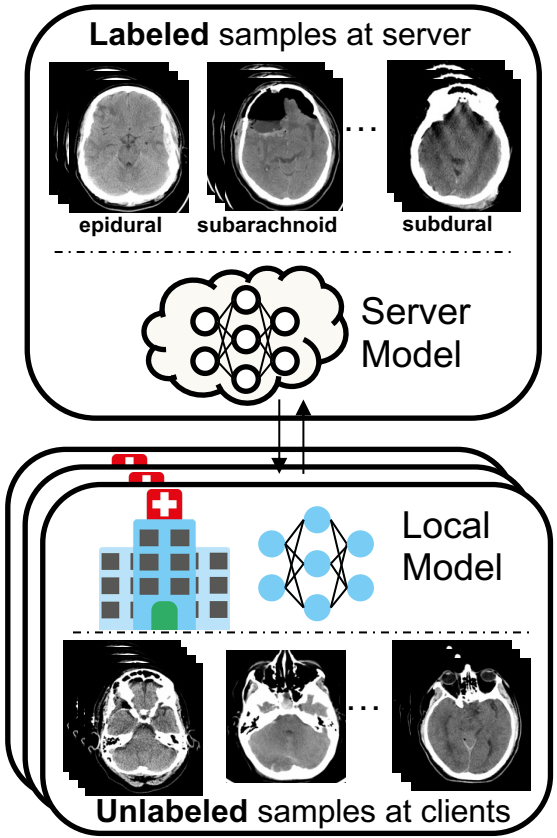




Overcoming Label Heterogeneity with imFed-Semi

Class imbalanced semi-supervised FL (*imFed-Semi*)

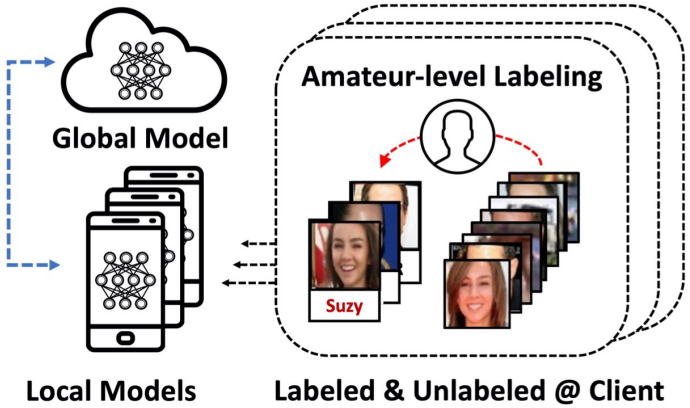
- Only the server holds a small amount of data
- All clients provide unlabeled data
- Label heterogeneity (class imbalance)



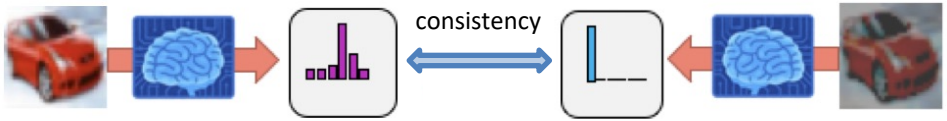
Overcoming Label Heterogeneity with imFed-Semi

Challenges of imFed-Semi

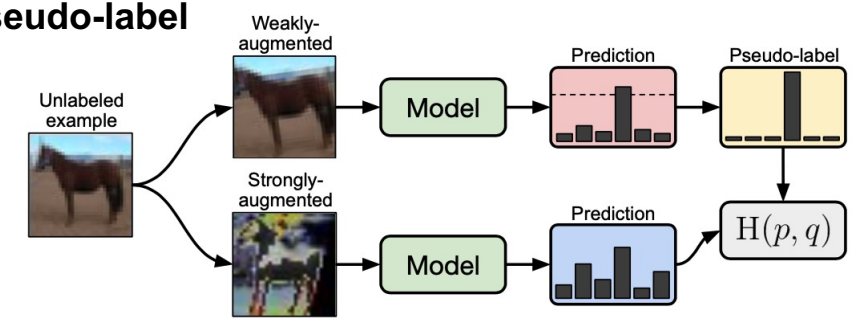
- Semi-supervised methods require data partially labeled.
- The widely used consistency or pseudo-label based method cannot distinguish class distribution differences.



Consistency



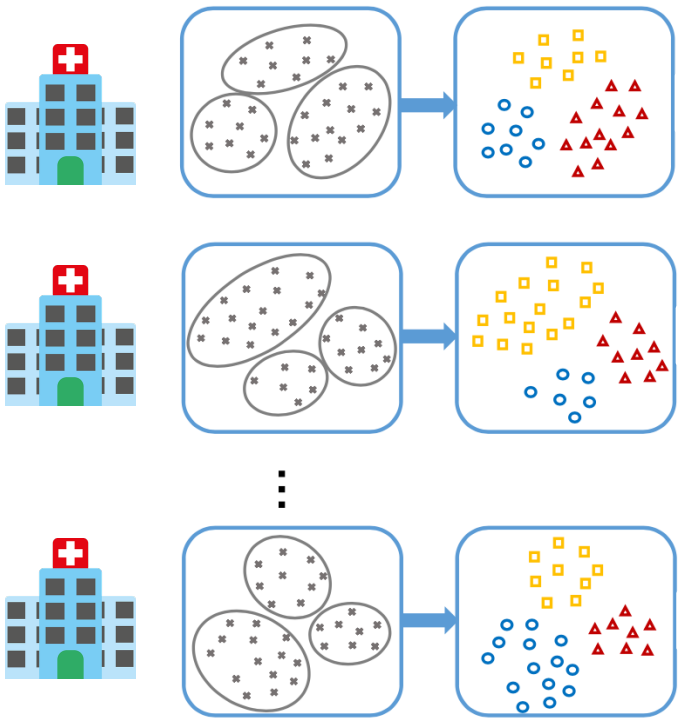
Pseudo-label



In imFed-Semi, all data at local clients are **unlabeled** with the presence of **class imbalance.**

Overcoming Label Heterogeneity with imFed-Semi

Learning Classifier from Unlabeled Data



Surrogate label

Step 1: Local surrogate task

- Use indexes of U sets as surrogate labels
- Formulate a surrogate supervised FL task

How to infer our desired classifier from the surrogate FL task?

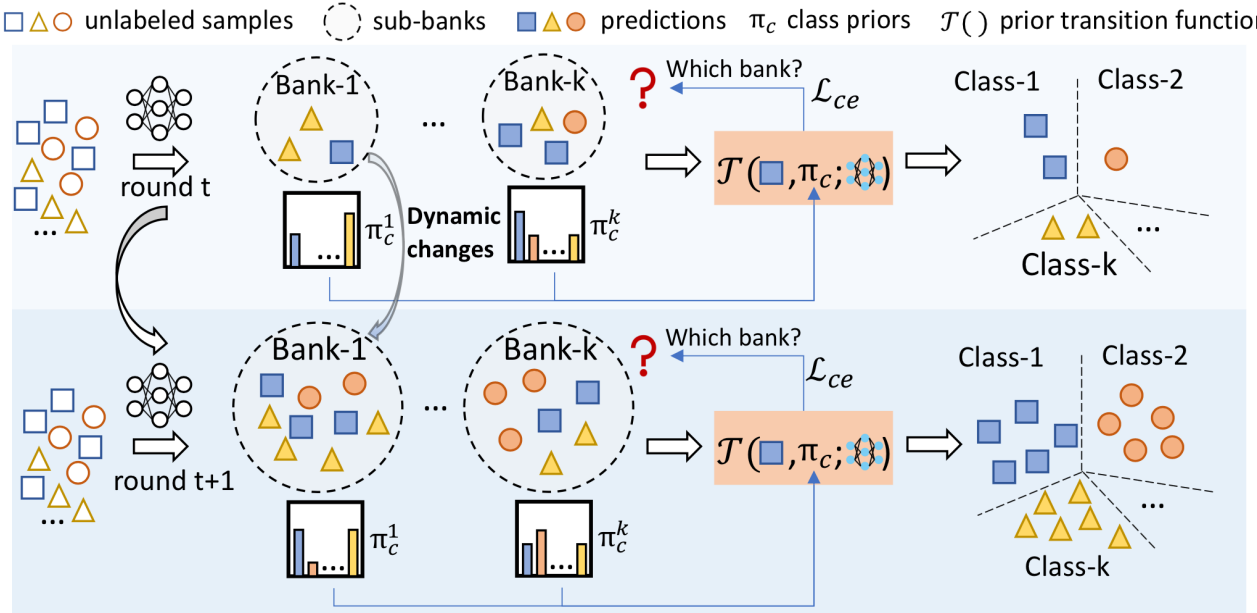
Step 2: Bridge class-posterior probabilities

- True class probability $\eta(x)$
- Surrogate class probability $\overline{\eta}_c(x)$
- Construct the relationship

Lu N, Wang Z, Li X, Niu G, Dou Q, Sugiyama M. "Federated Learning from Only Unlabeled Data with Class-Conditional-Sharing Clients." *ICLR* (2022).

Overcoming Label Heterogeneity with imFed-Semi

- **Method:** A novel bank learning algorithm to estimate the label proportions of each client in each round, the estimation serves as an approximation of true class probability, therefore utilizing unlabeled data and overcoming label heterogeneity.
- **Novelty:** Our method does not require the prior knowledge of true class probability and shows better performance against imbalanced label distributions.

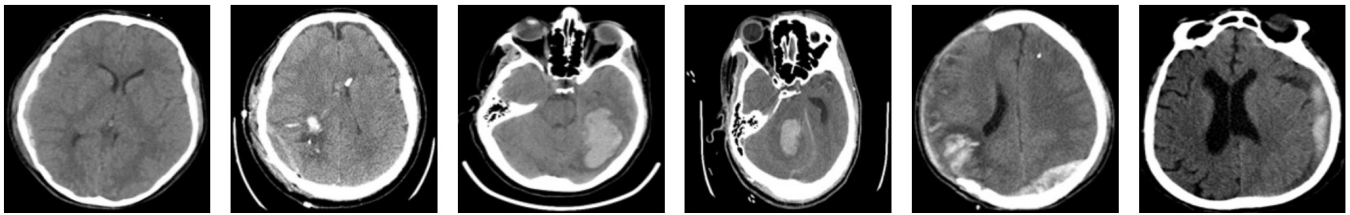


Jiang M, Yang H, Li X, Liu Q, Heng PA, Dou Q. Dynamic bank learning for semi-supervised federated image diagnosis with class imbalance. *MICCAI 2022.*

Overcoming Label Heterogeneity with imFed-Semi

Evaluation tasks

- Intracranial hemorrhage (ICH) diagnosis



(a) Healthy (b) Epidural (c) Intraparenchymal (d) Intraventricular (e) Subarachnoid (f) Subdural

- Skin lesion diagnosis

(a) Melanoma (b) Melanocytic Nevus (c) Basal Cell Carcinoma (d) Actinic Keratosis (e) Benign Keratosis (f) Dermatofibroma (g) Vascular Lesion

Evaluation metrics

- Accuracy, AUC, Specificity, Sensitivity, F1-score

Jiang M, Yang H, Li X, Liu Q, Heng PA, Dou Q. Dynamic bank learning for semi-supervised federated image diagnosis with class imbalance. *MICCAI 2022.*

Overcoming Label Heterogeneity with imFed-Semi

Quantitative comparison with other FL methods

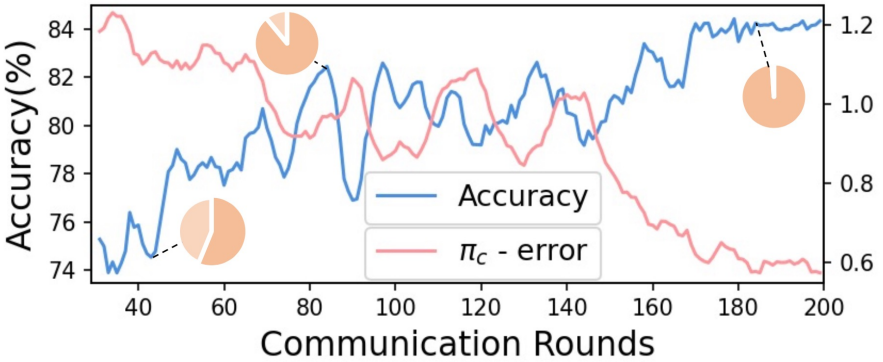
Intracranial Hemorrhage Diagnosis					
Methods	AUC	Accuracy	Specificity	Sensitivity	F1
FedAvg-SL	88.59±0.85	91.22±0.17	93.39±0.09	63.64±0.80	59.54±0.78
FedIRM [14]	60.41±0.93	72.27±0.40	82.93±0.29	30.13±0.55	22.93 ±0.31
FedMatch [10]	64.15±1.76	73.09±0.55	84.01±0.19	33.44±2.60	24.84 ±1.36
FSSL [29]	60.63±1.92	72.61±1.47	83.57±0.79	24.88±0.39	21.98±0.93
FedAvg-FM	61.54±1.63	74.50±1.31	86.12±1.79	25.01±0.55	22.55±0.07
FedProx-FM [12]	62.49±3.41	74.95±3.03	85.91±3.05	26.18±5.24	22.84±1.55
FedAdam-FM [18]	62.42±3.93	73.85±0.66	85.63±1.15	25.33±3.62	22.08±1.02
imFed-Semi (Ours)	**82.96±1.26**	**82.56±0.58**	**90.66±0.21**	**54.58±3.75**	**47.71±3.59**
Skin Lesion Diagnosis					
Methods	AUC	Accuracy	Specificity	Sensitivity	F1
FedAvg-SL	87.58±0.28	93.32±0.27	89.64±0.57	59.09±1.34	57.09 ±0.68
FedIRM [14]	65.29±2.26	79.05±1.33	90.39±1.08	28.72±2.22	23.52 ±1.21
FedMatch [10]	70.90±1.25	84.25±2.34	**93.33±1.74**	29.56±1.90	29.13 ±2.69
FSSL [29]	70.86±1.26	83.20±1.65	93.39±0.21	28.32±0.89	27.90±1.43
FedAvg-FM	70.61±1.79	82.67±0.91	91.92±1.67	30.65±0.91	29.09 ±0.94
FedProx-FM [12]	69.86±1.48	82.01±1.66	91.45±2.86	27.87±2.69	25.21±1.14
FedAdam-FM [18]	70.58±1.86	83.22±2.25	92.92±1.98	28.97±1.87	27.85±1.68
imFed-Semi (Ours)	**77.47±1.81**	**88.94±1.50**	89.81±2.64	**37.48±2.71**	**33.79±1.75**

Jiang M, Yang H, Li X, Liu Q, Heng PA, Dou Q. Dynamic bank learning for semi-supervised federated image diagnosis with class imbalance. *MICCAI 2022.*

Overcoming Label Heterogeneity with imFed-Semi

Analytical studies

- **The learning behavior regarding the dynamic bank construction and class prior estimation.**



- The effectiveness of dynamic selection comparing with a fixed threshold.

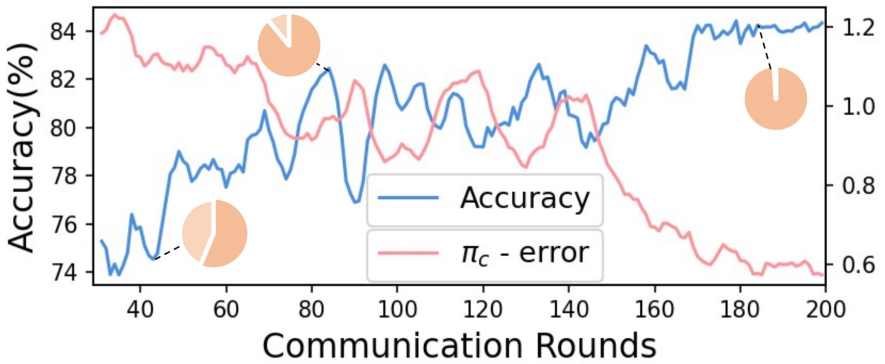
- How is the performance with more clients involved in federated training.

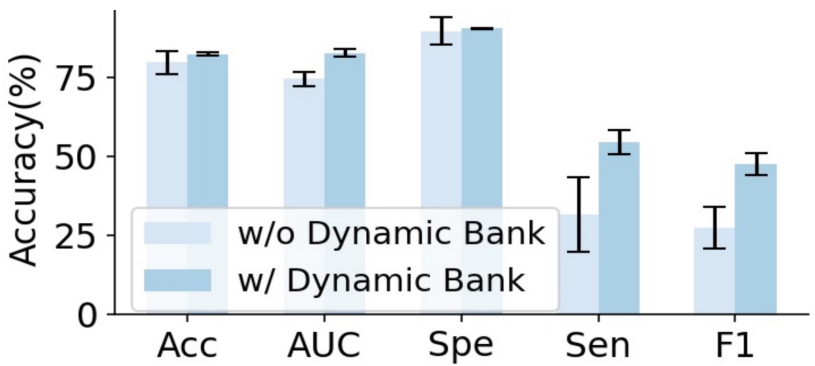
Jiang M, Yang H, Li X, Liu Q, Heng PA, Dou Q. Dynamic bank learning for semi-supervised federated image diagnosis with class imbalance. *MICCAI 2022.*

Overcoming Label Heterogeneity with imFed-Semi

Analytical studies

- The learning behavior regarding the dynamic bank construction and class prior estimation.
- **The effectiveness of dynamic selection comparing with a fixed threshold.**
- How is the performance with more clients involved in federated training.



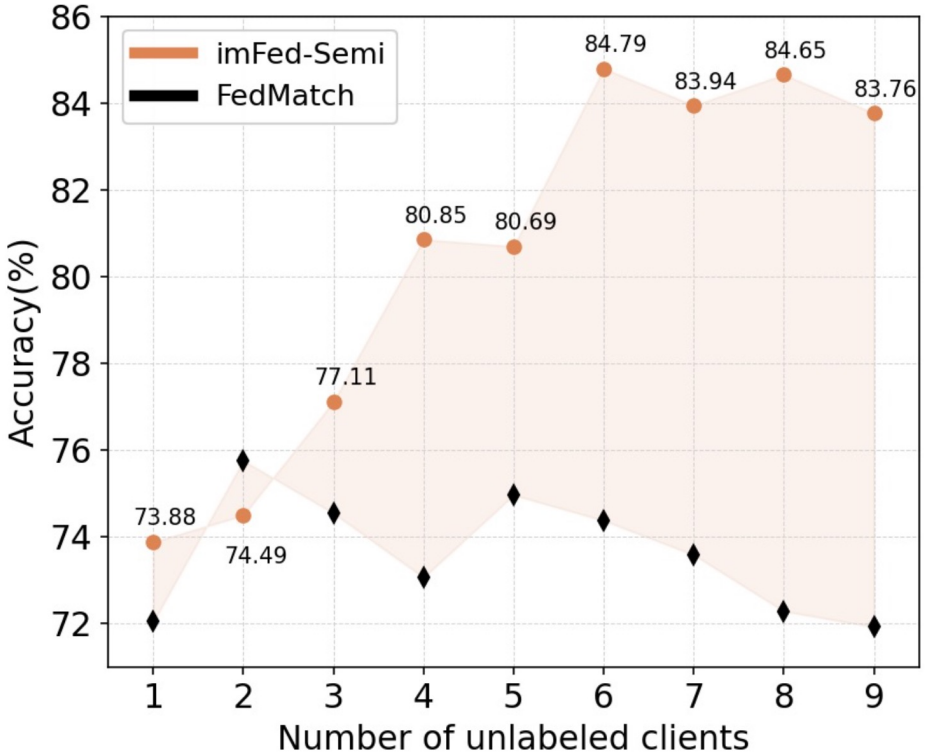


Jiang M, Yang H, Li X, Liu Q, Heng PA, Dou Q. Dynamic bank learning for semi-supervised federated image diagnosis with class imbalance. *MICCAI 2022.*

Overcoming Label Heterogeneity with imFed-Semi

Analytical studies

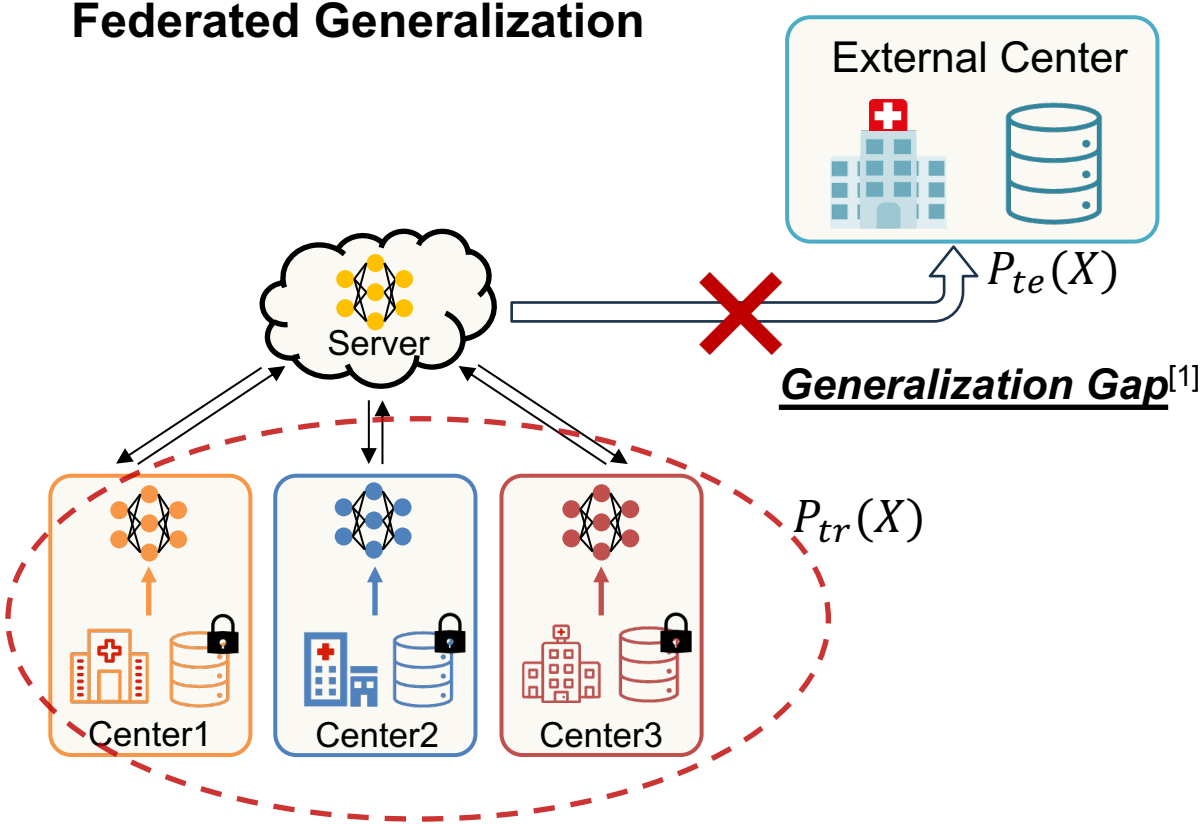
- The learning behavior regarding the dynamic bank construction and class prior estimation.
- The effectiveness of dynamic selection comparing with a fixed threshold.
- **How is the performance with more clients involved in federated training.**



Jiang M, Yang H, Li X, Liu Q, Heng PA, Dou Q. Dynamic bank learning for semi-supervised federated image diagnosis with class imbalance. *MICCAI 2022.*

Challenges in Federated Learning : Heterogeneity

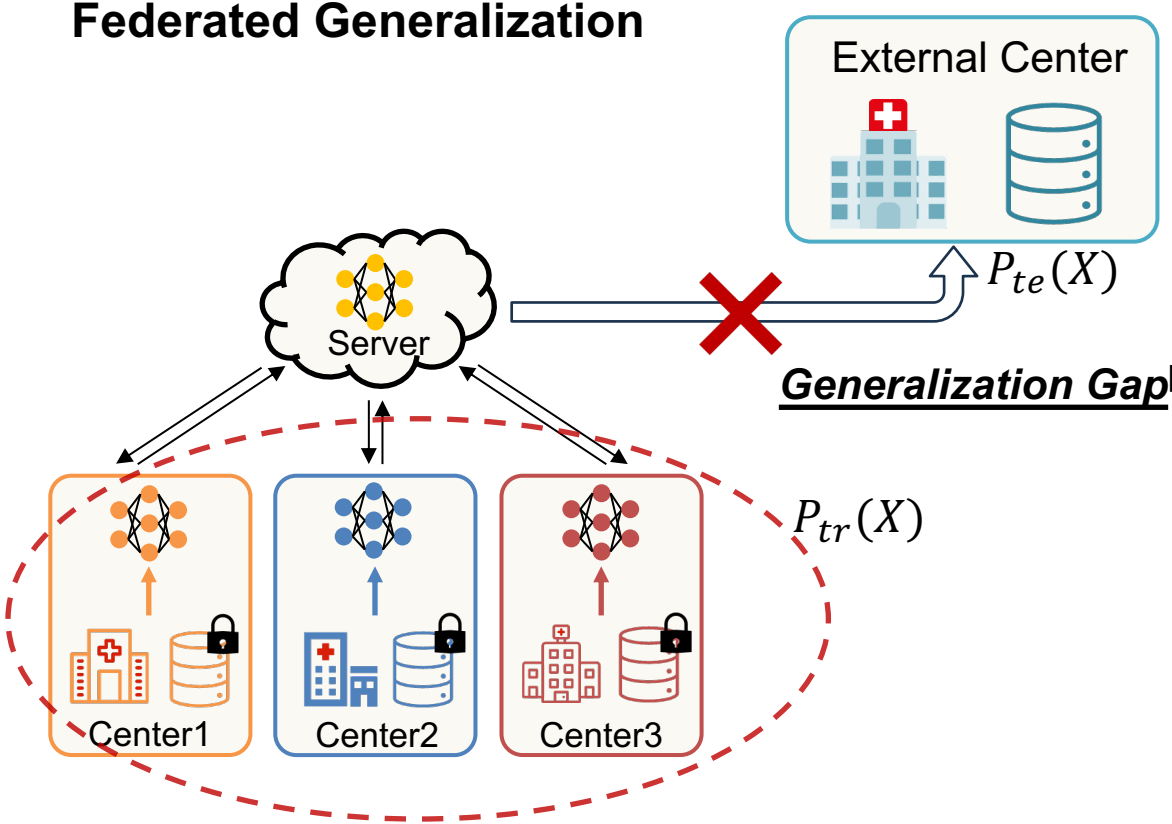
Federated Generalization

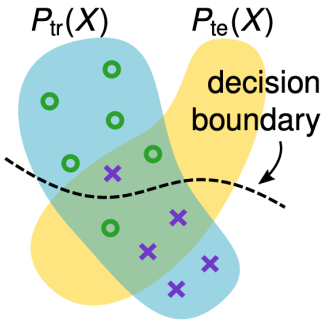


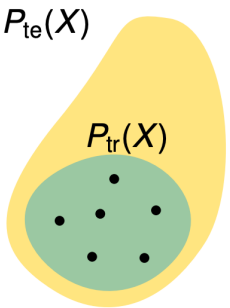
[1] Li, Y., Wang, X., Zeng, R., Donta, P. K., Murturi, I., Huang, M., & Dustdar, S. (2023). Federated domain generalization: A survey. *arXiv preprint arXiv:2306.01334*.

Challenges in Federated Learning : Heterogeneity

Federated Generalization





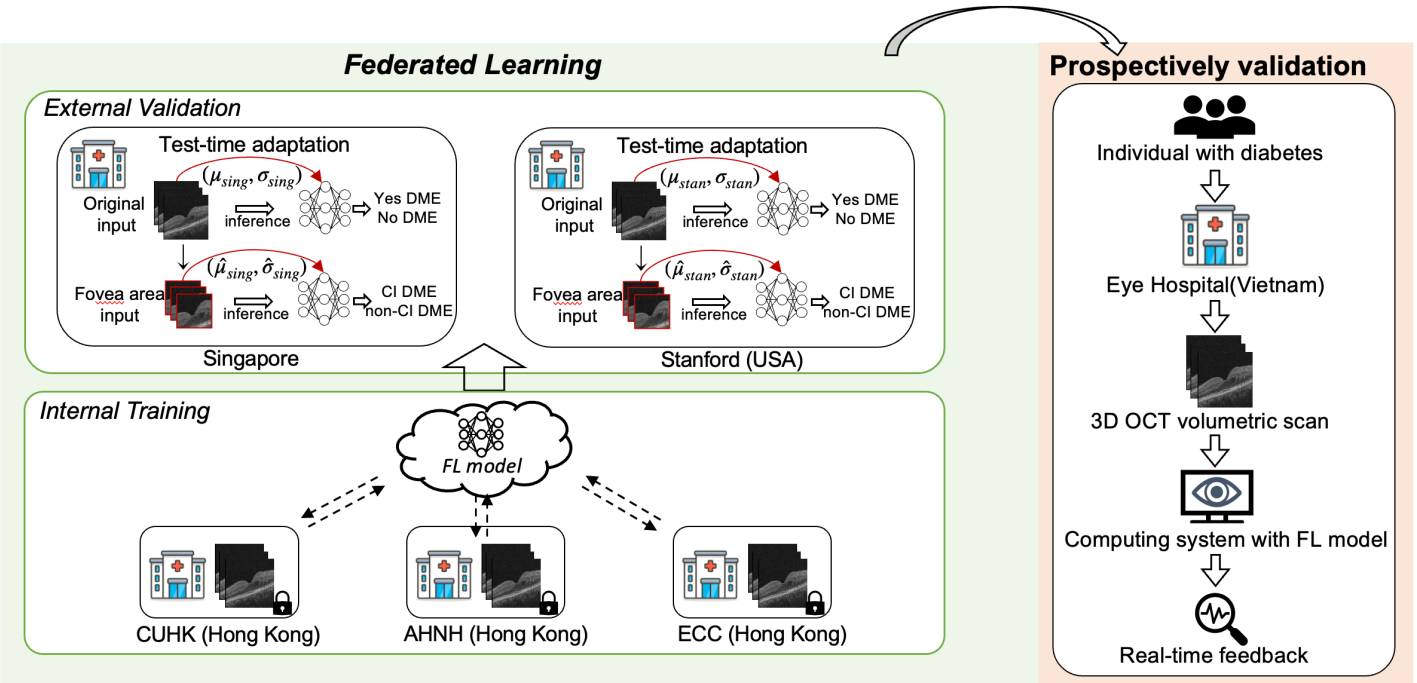


[D. Castro, I. Walker, and B. Glocker. 2019]

[1] Li, Y., Wang, X., Zeng, R., Donta, P. K., Murturi, I., Huang, M., & Dustdar, S. (2023). Federated domain generalization: A survey. *arXiv preprint arXiv:2306.01334*.

Improving Generalization by Test-time BatchNorm

- **Motivation:** Improving the generalizability of FL model for external unseen clients.
- **Framework:** Adjusting the statistics in the BatchNorm layer to improve the FL model performance on unseen clients. Further prospectively validate the FL model in real-world scenario with clients from the low- and middle-income countries.



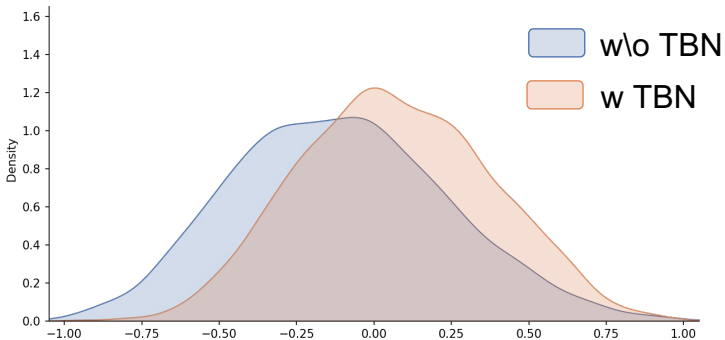
Improving Generalization by Test-time BatchNorm

- **Details:** Test-time BatchNorm (TBN) is designed during the inference time. It can reduce divergence between training and testing feature distributions, further improve the generalizability of FL model.

Federated testing:

Output: Prediction $\{\bar{y}_t\}_{t=1}^{n_t}$ for batch of test samples $\{\mathbf{x}_t\}_{t=1}^{n_t}$

1: **if** Testing on *internal clients* **then**
2: $\quad \{\bar{y}_t\}_{t=1}^{n_t} = f(\{\mathbf{x}_t\}_{t=1}^{n_t}; w_i^T)$
3: **end if**
4: **if** Testing on *external clients* **then**
5: $\quad$ During the forward of $\{\bar{y}_t\}_{t=1}^{n_t} = f(\{\mathbf{x}_t\}_{t=1}^{n_t}; w^T)$
6: $\quad$ **for** each layer l is BatchNorm **do**
7: $\quad\quad$ Re-estimate test-specific statistics for BatchNorm
8: $\quad\quad$ $\mu \leftarrow \tau\mu + (1-\tau)\frac{1}{n_t}\sum_{t=1}^{n_t}\mathbf{x}_t$
9: $\quad\quad$ $\sigma^2 \leftarrow \tau\sigma^2 + (1-\tau)\frac{1}{n_t}\sum_{t=1}^{n_t}(\mathbf{x}_t - \mu)^2$
10: $\quad$ **end for**
11: **end if**
12: **return** $\{\bar{y}_t\}_{t=1}^{n_t}$



Nguyen T, Jiang M, …, Dou Q, Chueng Y. Advancing Diabetic Macular Edema Detection from 3D Optical Coherence Tomography Scans: Integrating Privacy-Preserving AI and Generalizability Techniques - A Prospective Validation in Vietnam. *NEJM AI, 2024.*

Improving Generalization by Test-time BatchNorm

Experimental Evaluation

Internal validation:

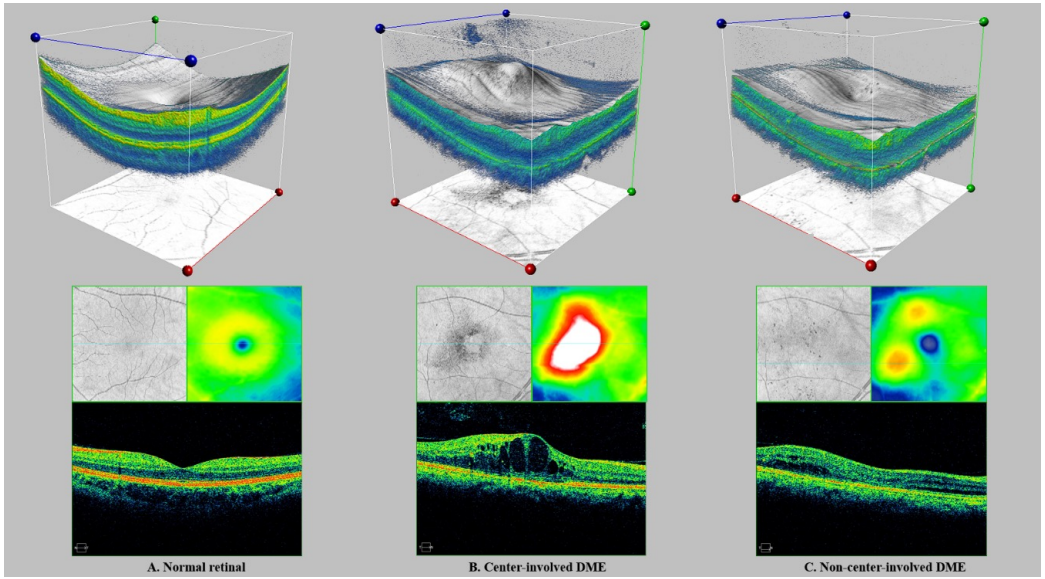
1) **FL Model**
2) **Centralized model** (Model trained on *centralized dataset*)

External validation:

1) **FL model**
2) **Centralized model**

Prospective validation: compare against human experts

1) **Retinal Specialist**
2) **3 Junior Ophthalmologists**



Nguyen T, Jiang M, …, Dou Q, Chueng Y. Advancing Diabetic Macular Edema Detection from 3D Optical Coherence Tomography Scans: Integrating Privacy-Preserving AI and Generalizability Techniques - A Prospective Validation in Vietnam. *NEJM AI, 2024.*

Improving Generalization by Test-time BatchNorm

Experimental Evaluation

Internal validation the performance of model in DME detection *(Yes/No DME)*

Model	Dataset	AUROC (95% CI)	Accuracy, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	p-value
Centralized	CUHK	0.937 (0.919-0.952)	88.97 (87.10-90.84)	85.14 (81.54-88.49)	91.23 (89.19-93.26)	0.073
FL	CUHK	0.950 (0.935-0.963)	89.81 (87.94-91.68)	85.89 (82.31-89.19)	92.12 (90.00-94.13)	-
Centralized	AHNH	0.912 (0.853-0.962)	88.54 (83.20-93.89)	80.33 (70.14-89.47)	95.71 (90.47-100)	0.096
FL	AHNH	0.949 (0.901-0.986)	93.89 (89.31-97.71)	93.44 (86.21-98.48)	94.29 (88.52-98.67)	-
Centralized	2010ECC	0.985 (0.948-1.000)	96.36 (90.09-100)	94.44 (85.71-100)	100.00 (100-100)	0.525
FL	2010ECC	0.993 (0.973-1.000)	96.36 (90.09-100)	94.44 (86.48-100)	100.00 (100-100)	-
Centralized	Average	0.937 (0.921-0.951)	88.37 (86.54-90.13)	85.43 (82.39-88.47)	90.29 (88.15-92.28)	**0.023**
FL	Average	**0.952 (0.938-0.963)**	90.60 (89.01-92.12)	84.21 (80.63-87.32)	94.75 (93.08-96.29)	-

Nguyen T, Jiang M, …, Dou Q, Chueng Y. Advancing Diabetic Macular Edema Detection from 3D Optical Coherence Tomography Scans: Integrating Privacy-Preserving AI and Generalizability Techniques - A Prospective Validation in Vietnam. *NEJM AI, 2024.*

Improving Generalization by Test-time BatchNorm

Experimental Evaluation

Internal validation the performance of model in DME classification *(CI/Non-CI-DME)*

Model	Dataset	AUROC (95% CI)	Accuracy, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	p-value
Centralized	CUHK	0.759 (0.709-0.804)	70.27 (65.49-74.81)	61.50 (54.58-67.80)	80.43 (74.56-86.02)	**0.003**
FL	CUHK	0.831 (0.792-0.868)	76.07 (71.78-80.10)	73.24 (66.83-79.16)	79.35 (73.18-85.21)	-
Centralized	AHNH	0.803 (0.681- 0.908)	77.05 (65.57-86.88)	81.08 (67.57-93.02)	70.83 (51.85-88.89)	0.159
FL	AHNH	0.877 (0.762-0.966)	90.16 (81.96-96.72)	100 (100-100)	75.00 (55.56-92.31)	-
Centralized	2010ECC	0.711 (0.453-0.968)	86.11 (71.34-93.91)	93.33 (78.67-98.15)	50.00 (18.76-81.23)	0.286
FL	2010ECC	0.827 (0.663-0.988)	66.67 (50.33-79.78)	60.00 (42.32-75.40)	100.00 (60.96-100)	-
Centralized	Average	0.777 (0.734-0.815)	72.06 (68.02-75.71)	67.87 (62.23-72.98)	77.57 (71.88-83.04)	**<0.001**
FL	Average	**0.849 (0.817-0.880)**	77.93 (74.09-81.38)	78.29 (73.23-83.10)	77.46 (71.88-82.55)	-

Nguyen T, Jiang M, …, Dou Q, Chueng Y. Advancing Diabetic Macular Edema Detection from 3D Optical Coherence Tomography Scans: Integrating Privacy-Preserving AI and Generalizability Techniques - A Prospective Validation in Vietnam. *NEJM AI, 2024.*

Improving Generalization by Test-time BatchNorm

Experimental Evaluation

External validation

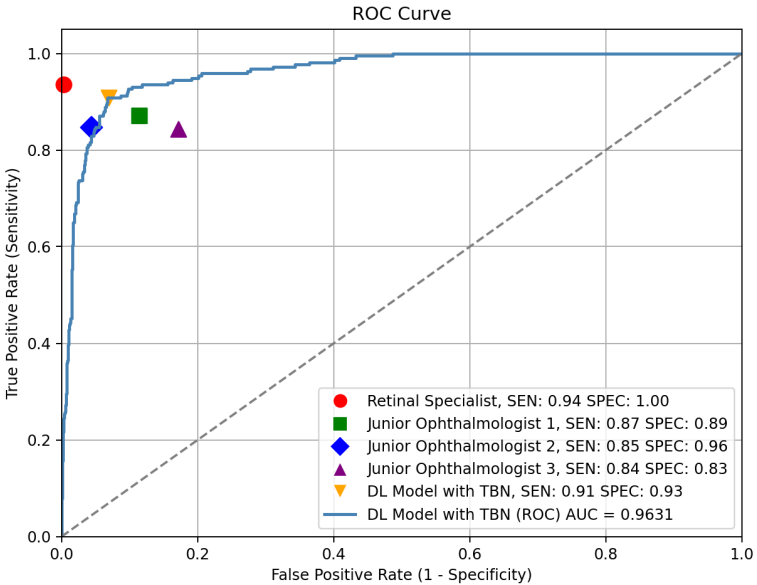
Model	AUROC (95% CI)	Accuracy, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	p-value
Testing of DME Detection on Retrospective Datasets					
Center-4					
FL	0.74 (0.67-0.81)	66.97 (64.52-69.43)	71.74 (57.77-84.44)	66.82 (64.18-69.29)	Ref
FL with TBN	**0.90 (0.85-0.95)**	88.29 (86.63-89.95)	82.61 (71.42-93.18)	88.49 (86.79-90.18)	**<0.001**
Center-5					
FL	0.87 (0.83-0.91)	84.39 (80.68-87.83)	88.85 (85.09-92.74)	74.58 (66.67-82.45)	Ref
FL with TBN	**0.96 (0.94-0.97)**	91.01 (87.83-93.91)	90.77 (87.12-94.14)	91.53 (86.21-96.33)	**<0.001**
Testing of Classification of ci-DME and non-ci-DME on Retrospective Datasets					
Center-4					
FL	0.84 (0.71-0.94)	78.26 (67.33-89.13)	90.90 (77.26-100)	66.67 (47.82-85.00)	Ref
FL with TBN	**0.89 (0.79-0.96)**	82.60 (71.68-93.47)	77.27 (58.82-94.11)	87.50 (72.00-100)	**0.039**
Center-5					
FL	0.86 (0.82-0.91)	77.69 (72.69-82.30)	75.12 (69.23-81.02)	87.27 (77.35-95.92)	Ref
FL with TBN	0.87 (0.82-0.92)	81.92 (77.31-86.54)	81.46 (76.09-86.53)	83.63 (73.91-92.85)	0.698

Nguyen T, Jiang M, …, Dou Q, Chueng Y. Advancing Diabetic Macular Edema Detection from 3D Optical Coherence Tomography Scans: Integrating Privacy-Preserving AI and Generalizability Techniques - A Prospective Validation in Vietnam. *NEJM AI, 2024.*

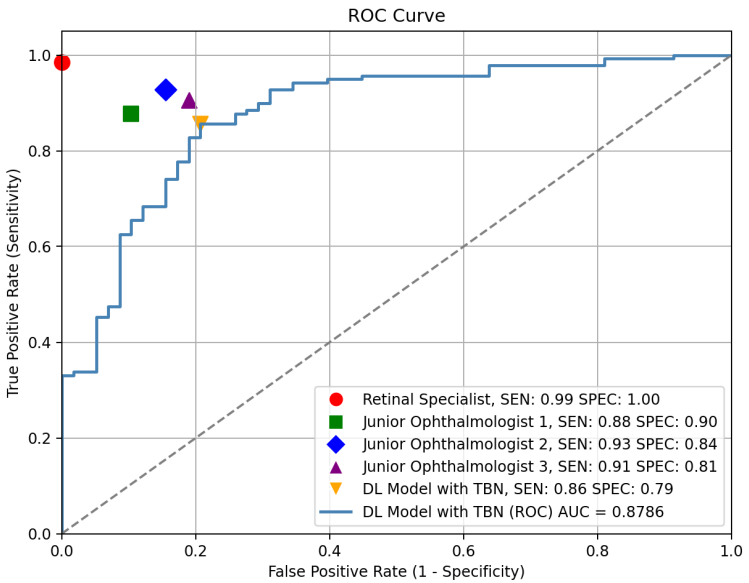
Improving Generalization by Test-time BatchNorm

Experimental Evaluation

Comparison of performances of FL models with TBN and human experts



(a) Detection of diabetic macular edema

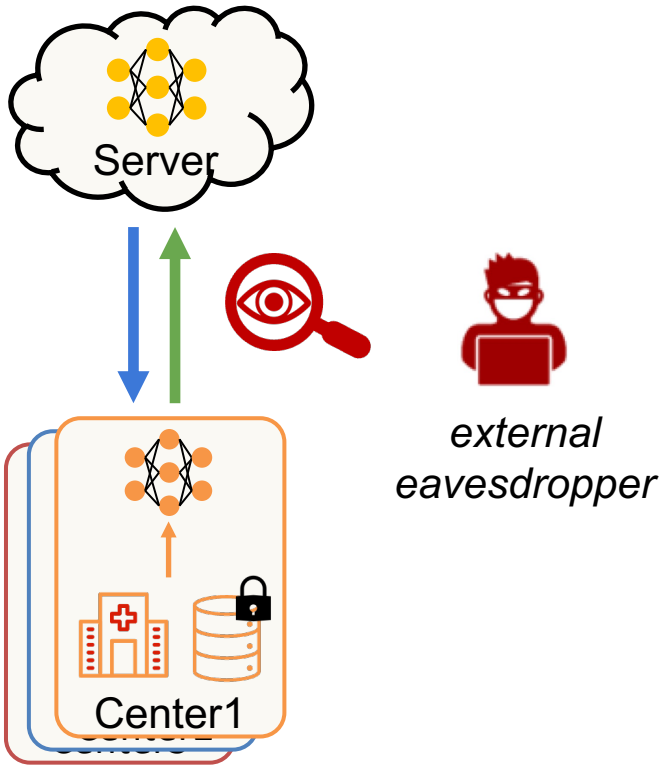


(b) Classification of CI-DME and Non-CI-DME

Nguyen T, Jiang M, …, Dou Q, Chueng Y. Advancing Diabetic Macular Edema Detection from 3D Optical Coherence Tomography Scans: Integrating Privacy-Preserving AI and Generalizability Techniques - A Prospective Validation in Vietnam. *NEJM AI, 2024.*

Challenges in Federated Learning : Privacy

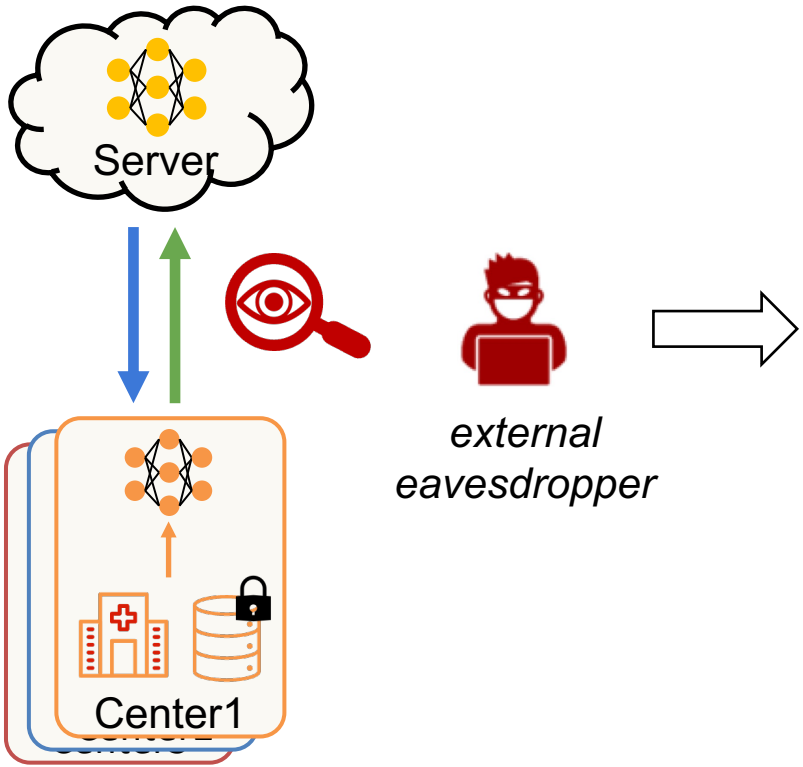
Federated Communication



[1] Chen, Y., Gui, Y., Lin, H., Gan, W., & Wu, Y. Federated learning attacks and defenses: A survey. In *2022 IEEE International Conference on Big Data (Big Data)*.

Challenges in Federated Learning : Privacy

Federated Communication

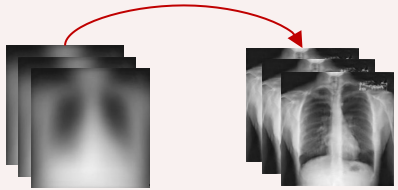


Threats[1]

1. *Model Extraction*

2. *Model Poisoning*

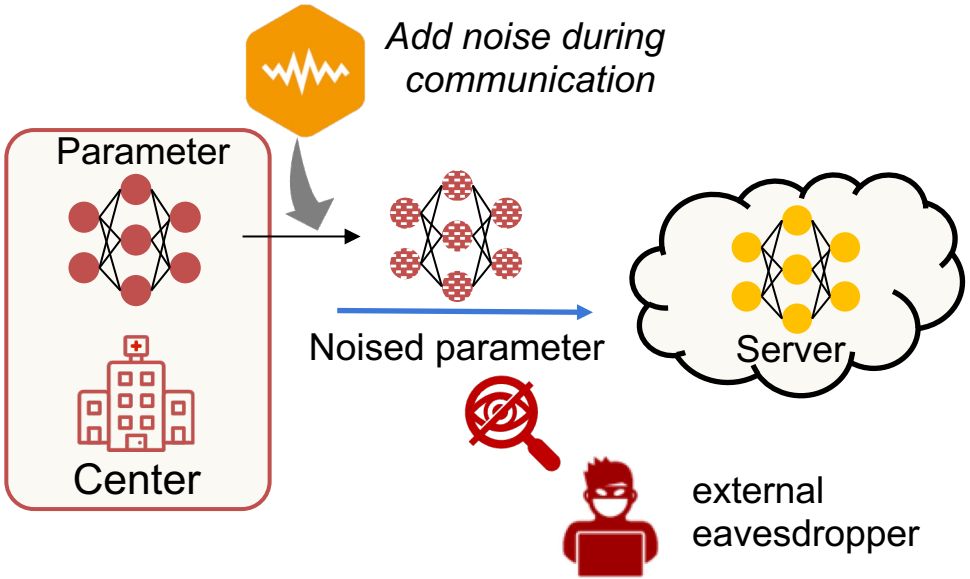
3. *Reconstruction Attack*



[1] Chen, Y., Gui, Y., Lin, H., Gan, W., & Wu, Y. Federated learning attacks and defenses: A survey. In *2022 IEEE International Conference on Big Data (Big Data)*.

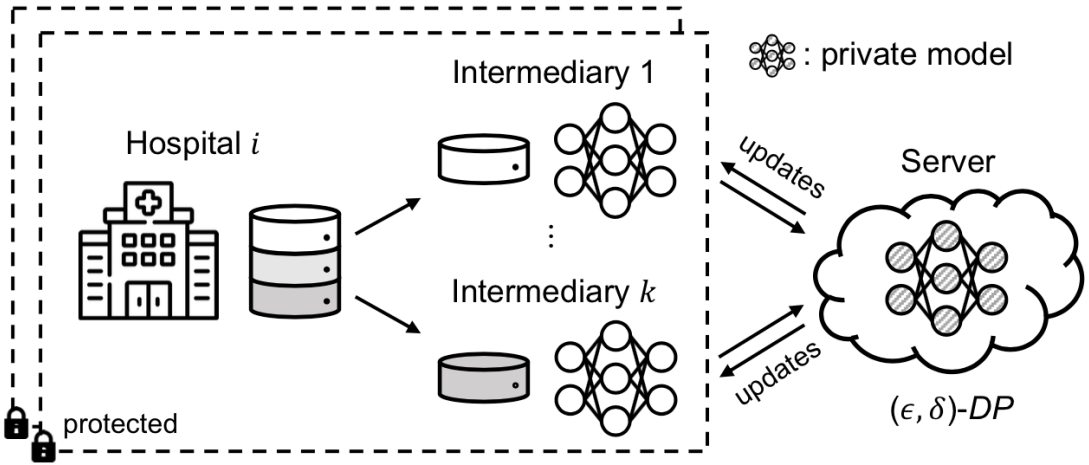
Differential Privacy for FL Clients with Client-DP

To further safeguard the parameter sharing, we use differential privacy (**DP**). We aim to protect each participating client via **client-level DP.**



Differential Privacy for FL Clients with Client-DP

- **Motivation:** Even without sharing data, model weights sharing in FL is also vulnerable to external eavesdropper.
- **Key Idea:** Introducing differential privacy in FL to further safeguard parameter sharing.
- **Method:** We propose to segment the client into several intermediaries, (i.e., sub-clients with non-overlapping datasets) to mitigate the noise introduced by DP, communication is then conducted among intermediaries.



Jiang M, Zhong Y, Le A, Li X, Dou Q. Client-Level Differential Privacy via Adaptive Intermediary in Federated Medical Imaging. *MICCAI* 2023.

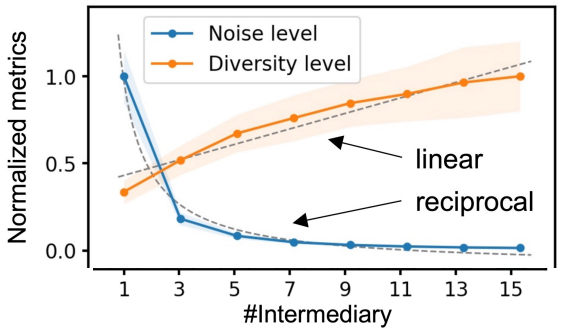
Differential Privacy for FL Clients with Client-DP

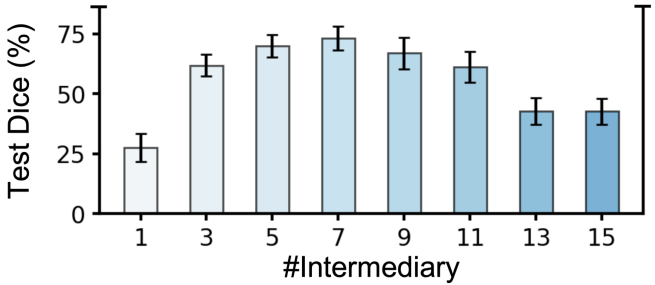
Analysis: The noise can be mitigated by segmenting client into several **intermediaries,** (i.e., sub-clients with non-overlapping datasets), where the communication is conducted among intermediaries.

Theorem 1. *If a randomized learning mechanism* $\mathcal{M} : \mathcal{X} \to \mathcal{R}$ *is* $(\epsilon, \delta) - DP$, *then its induced mechanism* $\tilde{\mathcal{M}} : \mathcal{Y} \to \mathcal{R}$ *is also* (ϵ, δ)*-DP.*

However, the gradient **diversity** can also be increased, bringing decreased performance.

We identify a **linear** and **reciprocal relationship** via experimental studies.





Jiang M, Zhong Y, Le A, Li X, Dou Q. Client-Level Differential Privacy via Adaptive Intermediary in Federated Medical Imaging. *MICCAI* 2023.

Differential Privacy for FL Clients with Client-DP

Quantitative comparison with other DP methods

Intracranial Hemorrhage Diagnosis ($N = 20$)								
Method	No Privacy		$z = 0.5$		$z = 1.0$		$z = 1.5$	
	AUC ↑	Acc ↑	AUC ↑	Acc ↑	AUC ↑	Acc ↑	AUC ↑	Acc ↑
DP-FedAvg [24]	$90.88_{\pm 0.15}$	$82.85_{\pm 0.26}$	$70.38_{\pm 0.61}$	$64.94_{\pm 0.28}$	$68.00_{\pm 1.43}$	$63.55_{\pm 1.12}$	$66.77_{\pm 0.12}$	$62.01_{\pm 0.58}$
+Ours	-	-	$82.42_{\pm 0.29}$	$74.87_{\pm 0.43}$	$80.84_{\pm 0.78}$	$73.37_{\pm 0.92}$	$80.77_{\pm 0.80}$	$72.95_{\pm 0.47}$
DP-FedAdam [28]	$91.85_{\pm 0.30}$	$84.16_{\pm 0.56}$	$75.91_{\pm 0.28}$	$68.75_{\pm 0.16}$	$70.75_{\pm 2.18}$	$65.34_{\pm 0.64}$	$70.89_{\pm 2.05}$	$63.73_{\pm 1.16}$
+Ours	-	-	$82.86_{\pm 0.47}$	$75.20_{\pm 0.27}$	$81.63_{\pm 0.38}$	$73.99_{\pm 0.79}$	$80.55_{\pm 0.60}$	$73.06_{\pm 0.68}$
DP-FedNova [31]	$90.89_{\pm 0.25}$	$83.00_{\pm 0.17}$	$71.84_{\pm 1.51}$	$66.25_{\pm 0.91}$	$69.26_{\pm 1.76}$	$63.75_{\pm 1.49}$	$68.45_{\pm 0.93}$	$63.21_{\pm 1.13}$
+Ours	-	-	$82.73_{\pm 0.38}$	$75.35_{\pm 0.27}$	$80.64_{\pm 0.57}$	$73.55_{\pm 0.78}$	$79.39_{\pm 0.41}$	$71.70_{\pm 0.35}$
DP^2-RMSProp [17]	$88.89_{\pm 0.02}$	$80.77_{\pm 0.25}$	$70.59_{\pm 1.16}$	$64.92_{\pm 1.19}$	$67.43_{\pm 0.60}$	$62.05_{\pm 0.23}$	$65.91_{\pm 1.40}$	$61.77_{\pm 1.13}$
+Ours	-	-	$81.60_{\pm 0.68}$	$74.47_{\pm 0.95}$	$80.23_{\pm 0.22}$	$73.15_{\pm 0.44}$	$81.32_{\pm 0.50}$	$74.21_{\pm 0.85}$

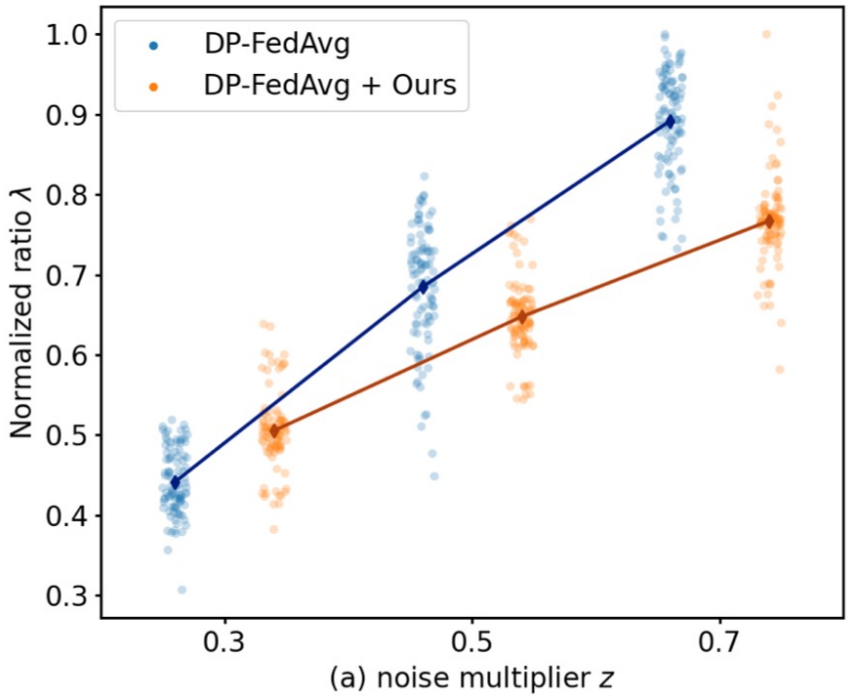
Prostate MRI Segmentation ($N = 6$)								
Method	No Privacy		$z = 0.3$		$z = 0.5$		$z = 0.7$	
	Dice ↑	IoU ↑	Dice ↑	IoU ↑	Dice ↑	IoU ↑	Dice ↑	IoU ↑
DP-FedAvg [24]	$87.69_{\pm 0.12}$	$79.62_{\pm 0.13}$	$41.43_{\pm 3.89}$	$29.28_{\pm 3.40}$	$22.45_{\pm 3.15}$	$13.50_{\pm 2.24}$	$13.59_{\pm 0.96}$	$7.41_{\pm 0.59}$
+Ours	-	-	$70.59_{\pm 1.55}$	$67.72_{\pm 0.47}$	$63.28_{\pm 4.69}$	$61.12_{\pm 0.50}$	$58.14_{\pm 4.71}$	$56.18_{\pm 7.21}$
DP-FedAdam [28]	$87.63_{\pm 0.16}$	$79.65_{\pm 0.20}$	$38.24_{\pm 2.86}$	$38.24_{\pm 1.38}$	$16.50_{\pm 1.82}$	$15.03_{\pm 2.28}$	$9.15_{\pm 2.19}$	$5.49_{\pm 2.22}$
+Ours	-	-	$69.68_{\pm 1.45}$	$61.31_{\pm 0.71}$	$57.11_{\pm 6.30}$	$57.23_{\pm 1.17}$	$43.99_{\pm 9.04}$	$47.49_{\pm 7.81}$
DP-FedNova [31]	$87.44_{\pm 0.35}$	$79.49_{\pm 0.29}$	$41.91_{\pm 6.34}$	$29.33_{\pm 5.78}$	$17.10_{\pm 7.45}$	$9.96_{\pm 4.81}$	$11.41_{\pm 0.64}$	$6.06_{\pm 0.38}$
+Ours	-	-	$70.80_{\pm 1.28}$	$66.64_{\pm 1.42}$	$68.63_{\pm 2.17}$	$63.99_{\pm 0.96}$	$58.99_{\pm 4.43}$	$59.14_{\pm 2.03}$
DP^2-RMSProp [17]	$87.46_{\pm 0.08}$	$80.00_{\pm 0.09}$	$38.33_{\pm 2.44}$	$24.73_{\pm 3.80}$	$16.74_{\pm 0.83}$	$10.75_{\pm 0.99}$	$7.77_{\pm 0.41}$	$4.00_{\pm 0.23}$
+Ours	-	-	$63.05_{\pm 2.60}$	$63.05_{\pm 2.60}$	$53.53_{\pm 4.97}$	$60.84_{\pm 5.81}$	$47.82_{\pm 2.01}$	$59.66_{\pm 2.76}$

Jiang M, Zhong Y, Le A, Li X, Dou Q. Client-Level Differential Privacy via Adaptive Intermediary in Federated Medical Imaging. *MICCAI* 2023.

Differential Privacy for FL Clients with Client-DP

Analytical studies

- **Our method effectively mitigates the effects from noise with increasing level of noise multiplier.**
- Our method shows more stable optimization directions with less variance among clients.
- Our method shows better scalability with varying number of clients.

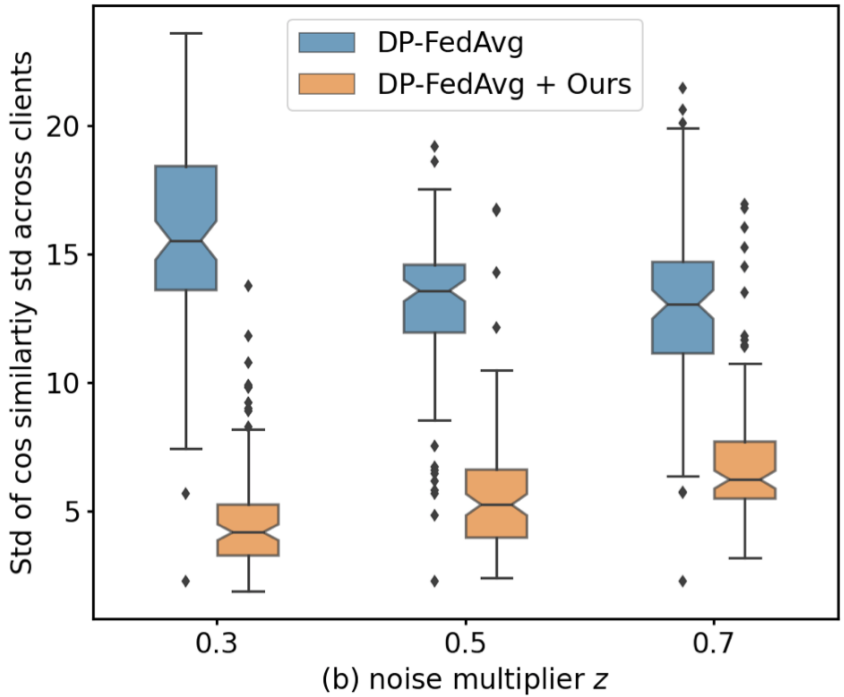


Jiang M, Zhong Y, Le A, Li X, Dou Q. Client-Level Differential Privacy via Adaptive Intermediary in Federated Medical Imaging. *MICCAI* 2023.

Differential Privacy for FL Clients with Client-DP

Analytical studies

- Our method effectively mitigates the effects from noise with increasing level of noise multiplier.
- **Our method shows more stable optimization directions with less variance among clients.**
- Our method shows better scalability with varying number of clients.

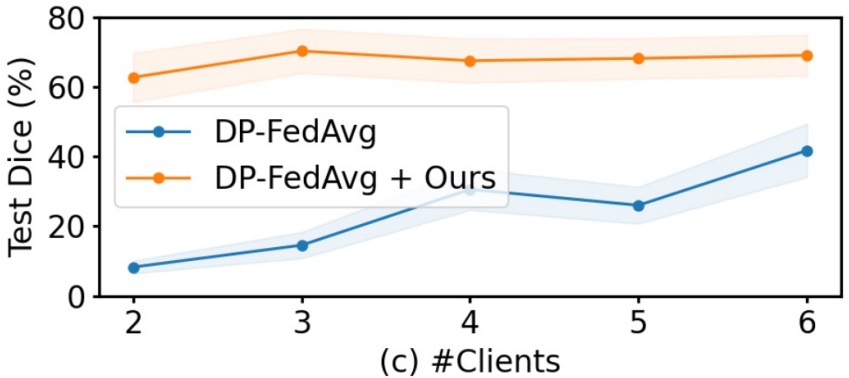


Jiang M, Zhong Y, Le A, Li X, Dou Q. Client-Level Differential Privacy via Adaptive Intermediary in Federated Medical Imaging. *MICCAI* 2023.

Differential Privacy for FL Clients with Client-DP

Analytical studies

- Our method effectively mitigates the effects from noise with increasing level of noise multiplier.
- Our method shows more stable optimization directions with less variance among clients.
- **Our method shows better scalability with varying number of clients.**



Jiang M, Zhong Y, Le A, Li X, Dou Q. Client-Level Differential Privacy via Adaptive Intermediary in Federated Medical Imaging. *MICCAI* 2023.

3. Summary

Summary

- FL is a promising approach to obtain powerful, accurate, safe, robust and generalizable models.

- By enabling multiple parties to train collaboratively without data sharing, FL neatly addresses issues related to egress of sensitive medical data.

- Not all technical questions have been answered yet, there are many open problems *(security; trustworthiness; fairness; traceability and accountability)* remained. FL will certainly be an active research area throughout the next decade.

FL Gallery

Platform
&
Library

FATE (fedai.org)

FedML (fedml.ai)

Fedleaner (Bytedance)

Flower (flower.dev)

LEAP (UBC)

NVFlare (NVIDIA)

Pysyft (DeepMind)

…

Open
Datasets
(multi-center)

Brain MRI (FeTS, Upenn)

Prostate MRI (SAML)

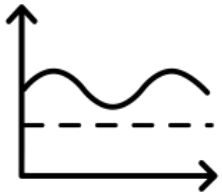
Dermoscopy skin lesion image

Retinal fundus image

Payhology Images (Camelyon17)

Autism Brain Imaging Data (ABIDE)

…



Baselines

FedAvg
FedProx
FedBN
FedNova
FedAdam
SCAFFOLD
Ditto
…

香港中文大學計算機科學與工程學系
Department of Computer Science and Engineering
The Chinese University of Hong Kong

Thanks for your attention!