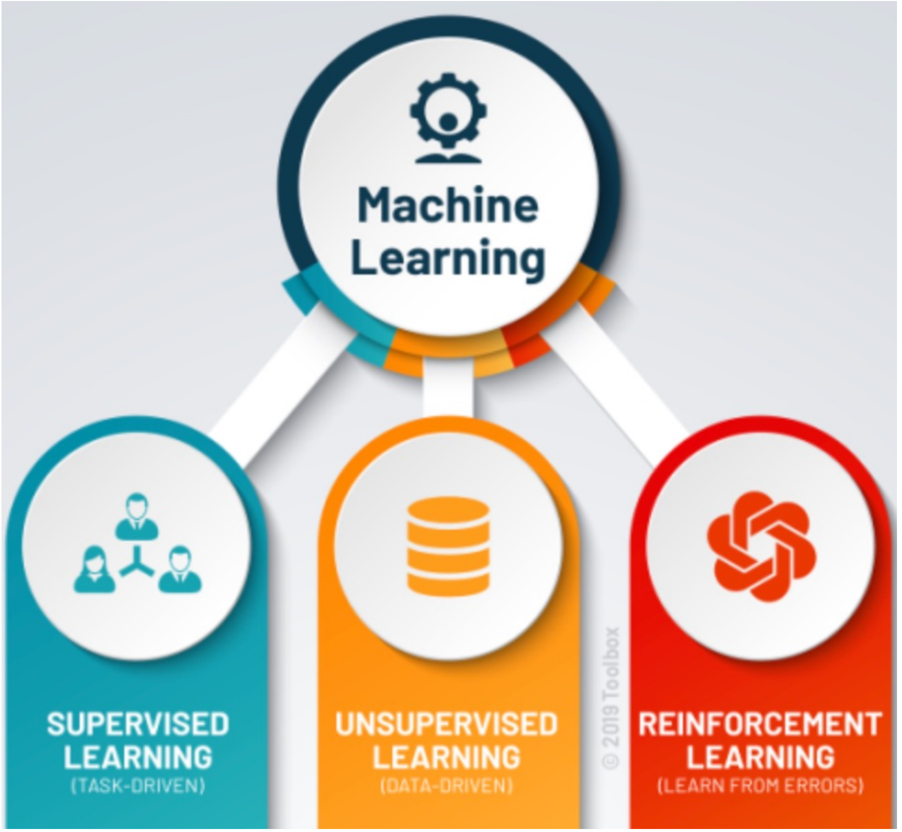
Federated Learning in Large Clinical Research Networks

Fei Wang
Associate Professor
Department of Population Health Sciences
Weill Cornell Medicine
feiwang.cornell@gmail.com
@feiwang03
https://wcm-wanglab.github.io/index.html

Machine Learning



https://www.potentiaco.com/what-is-machine-learning-definition-types-applications-and-examples/

Machine Learning

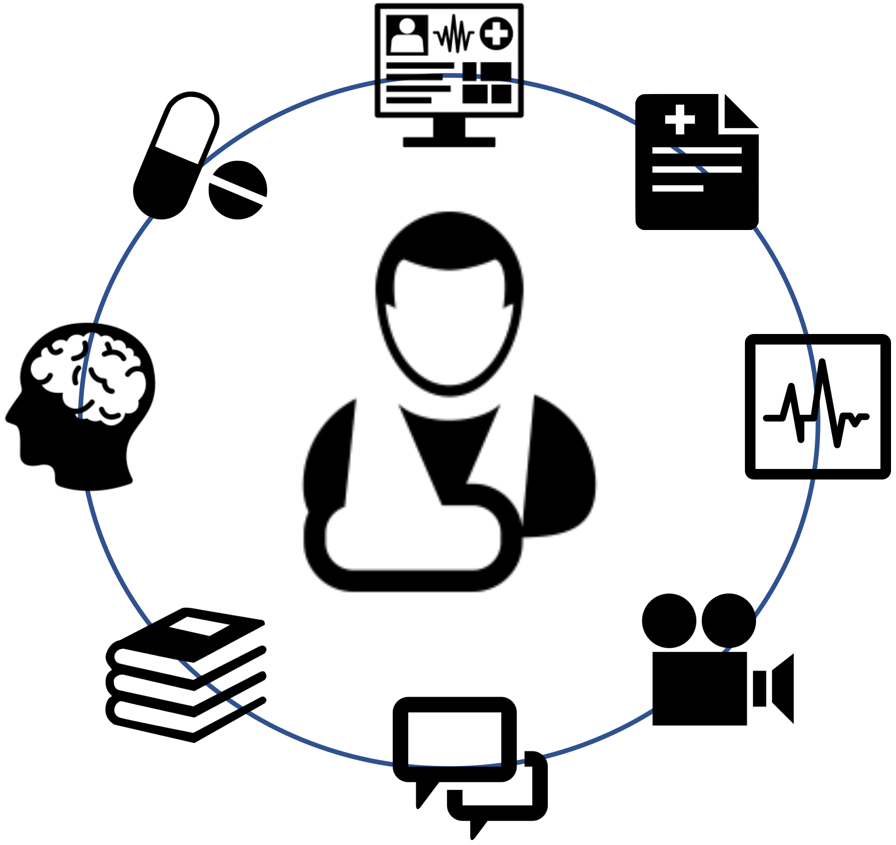
https://qbi.uq.edu.au/blog/2017/10/google-alphago-zero-masters-game-three-days

https://electrek.co/2017/04/29/elon-musk-tesla-plan-level-5-full-autonomous-driving/

https://www.shellypalmer.com/2017/01/5-awesome-illegal-uses-alexa/

https://siliconangle.com/2020/07/19/openais-latest-ai-text-generator-gpt-3-amazes-early-adopters/

Medicine



The NEW ENGLAND JOURNAL of MEDICINE

Perspective

A New Initiative on Precision Medicine

Francis S. Collins, M.D., Ph.D., and Harold Varmus, M.D.
N Engl J Med 2015; 372:793-795 | February 26, 2015 | DOI: 10.1056/NEJMp1500523

Comments open through March 4, 2015

Share:

Article | References | Citing Articles (784) | Comments (7) | Metrics

"Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier."

— President Barack Obama, State of the Union Address, January 20, 2015

President Obama has long expressed a strong conviction that science offers great potential for improving health. Now, the President has announced a research initiative that aims to accelerate progress toward a new era of precision medicine (www.whitehouse.gov/precisionmedicine). We believe that the time is right for this visionary initiative, and the National Institutes of Health (NIH) and other partners will work to achieve this vision.

AUDIO INTERVIEW

Interview with Dr. Francis Collins on what to expect from the recently announced Precision Medicine Initiative. (10:07)

Listen

Download

"The initiative will encourage and support the next generation of scientists to develop creative new approaches for detecting, measuring, and analyzing a wide range of biomedical information — including molecular, genomic, cellular, clinical, behavioral, physiological, and environmental parameters"

Machine Learning in Clinical Medicine

Original Investigation FREE

May 12, 2020

Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19

Wenhua Liang, MD[1,2]; Hengrui Liang, MD[1,2]; Limin Ou, MD[1]; et al

Author Affiliations | Article Information

JAMA Intern Med. 2020;180(8):1081-1089. doi:10.1001/jamainternmed.2020.2033

Table 3. Multivariable Logistic Regression Model for Predicting Development of Critical Illness in 1590 Patients Hospitalized With COVID-19 in Wuhan

Variables	Odds ratio (95% CI)	*P* value
X-ray abnormality (yes vs no)	3.39 (2.14-5.38)	<.001
Age, per y	1.03 (1.01-1.05)	.002
Hemoptysis (yes vs no)	4.53 (1.36-15.15)	.01
Dyspnea (yes vs no)	1.88 (1.18-3.01)	.01
Unconsciousness (yes vs no)	4.71 (1.39-15.98)	.01
No. of comorbidities	1.60 (1.27-2.00)	<.001
Cancer history (yes vs no)	4.07 (1.23-13.43)	.02
Neutrophil to lymphocyte ratio	1.06 (1.02-1.10)	.003
Lactate dehydrogenase, U/L	1.002 (1.001-1.004)	<.001
Direct bilirubin, μmol/L	1.15 (1.06-1.24)	.001
Constant	0.001	

Abbreviation: COVID-19, coronavirus disease 2019.

OXFORD ACADEMIC

Clinical Chemistry

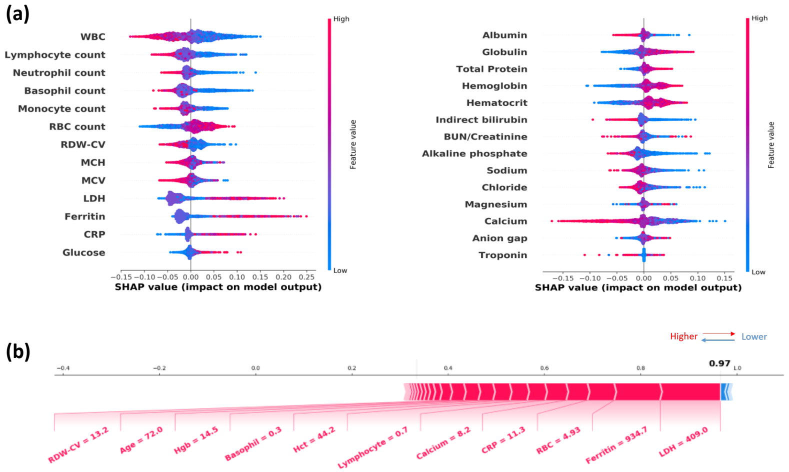
Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection Using Machine Learning FREE

He S Yang, Yu Hou, Ljiljana V Vasovic, Peter A D Steel, Amy Chadburn, Sabrina E Racine-Brzostek, Priya Velu, Melissa M Cushing, Massimo Loda, Rainu Kaushal ... Show more

Author Notes

Clinical Chemistry, Volume 66, Issue 11, November 2020, Pages 1396–1404, https://doi.org/10.1093/clinchem/hvaa200

Published: 30 October 2020 **Article history** ▾

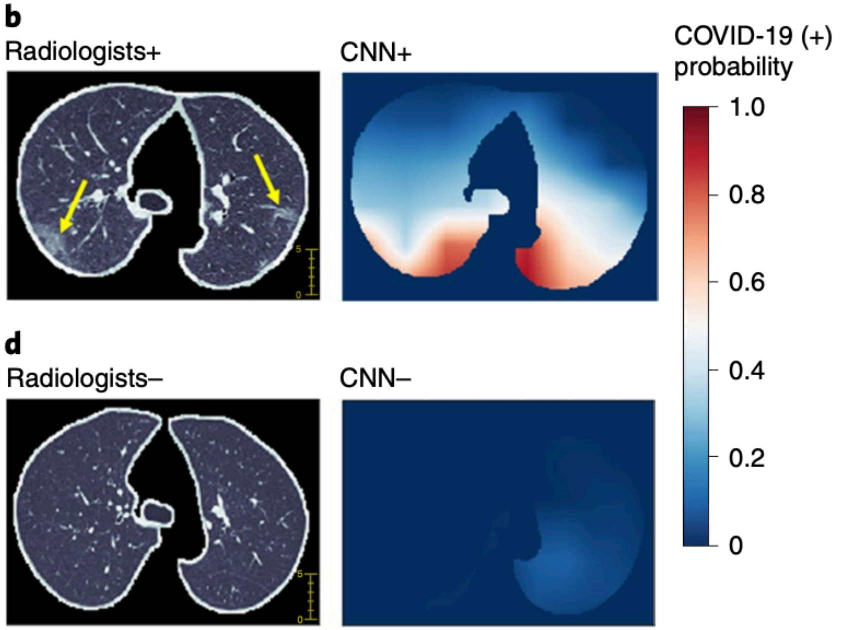


nature medicine

Letter | Published: 19 May 2020

Artificial intelligence–enabled rapid diagnosis of patients with COVID-19

Xueyan Mei, Hao-Chih Lee, [...] Yang Yang



Considerations


JAMA Network
JAMA Internal Medicine
Journals
Enter Search Term

New Online | Views 1,285 | Citations 0 | Altmetric 32

Viewpoint | Health Care Reform

ONLINE FIRST

December 17, 2018

Deep Learning in Medicine—Promise, Progress, and Challenges

Fei Wang, PhD[1]; Lawrence Peter Casalino, MD[1]; Dhruv Khullar, MD[1,2]

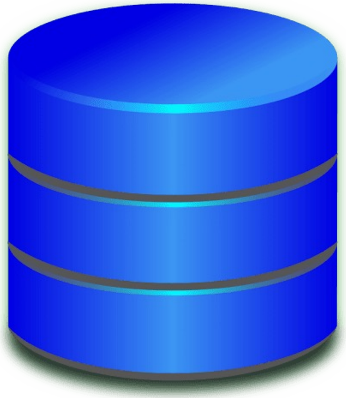
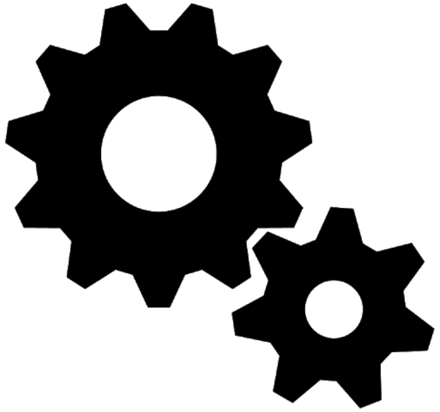
Author Affiliations

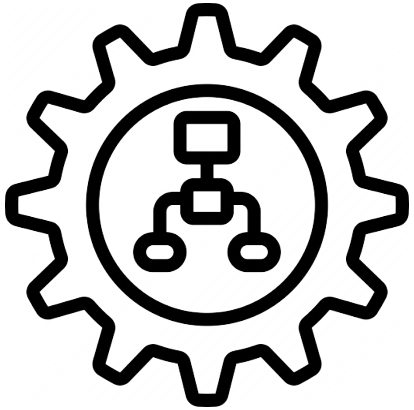
[1]Department of Healthcare Policy and Research, Weill Cornell Medicine, New York, New York

[2]Department of Medicine, Weill Cornell Medicine, New York, New York

JAMA Intern Med. Published online December 17, 2018. doi:10.1001/jamainternmed.2018.7117

More

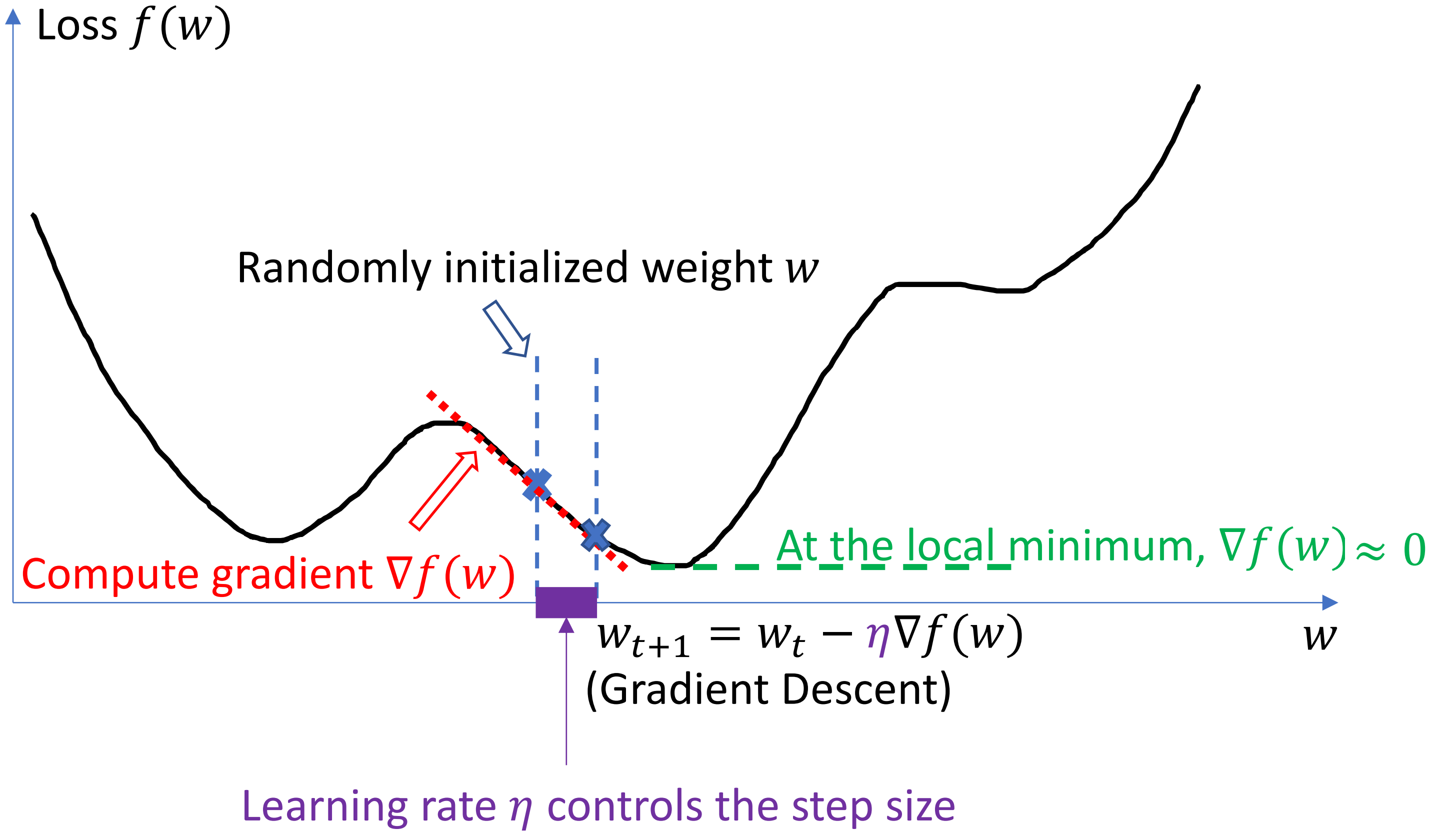



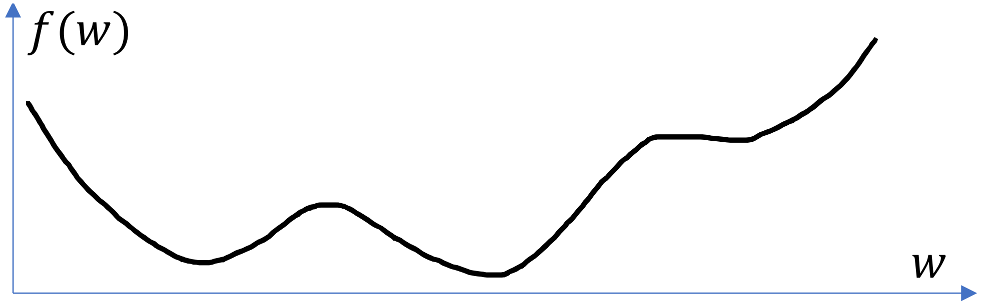



Model Training

$$\min_{w \in \mathbb{R}^d} f(w)$$

$$f(w) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} f_i(w)$$





How to stop? – when the update is small enough – converge.

$$\| w_{t+1} - w_t \| \leq \epsilon$$

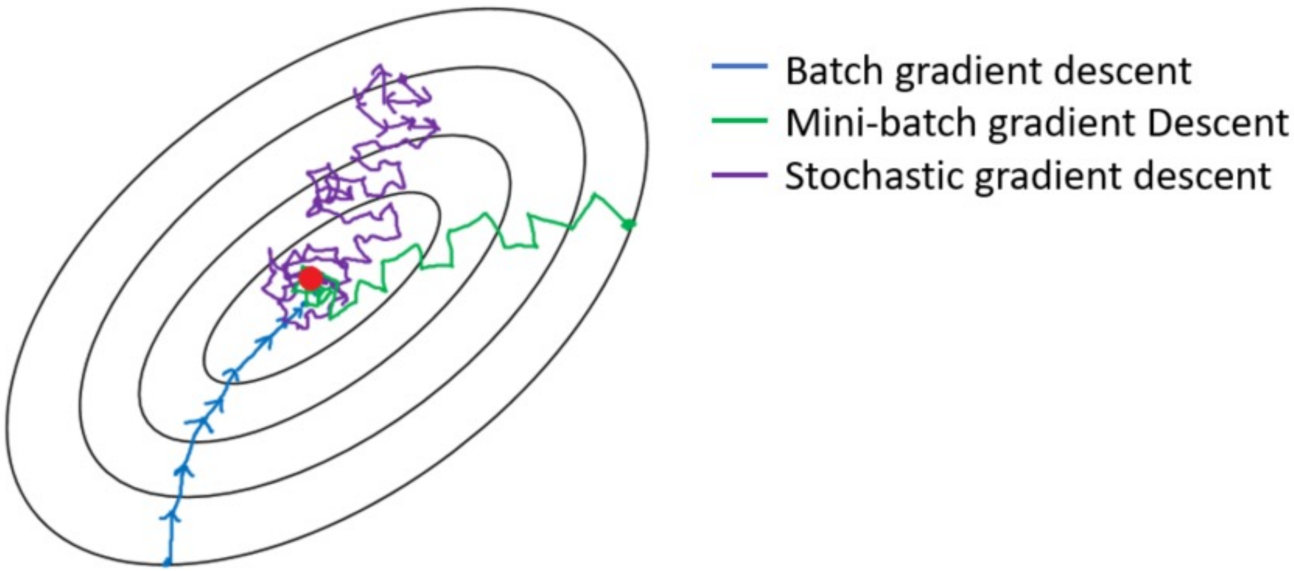
or $$\| \nabla f(w_t) \| \leq \epsilon$$

https://inst.eecs.berkeley.edu/~cs294-163/fa19/slides/federated-learning.pdf

Stochastic Gradient Descent

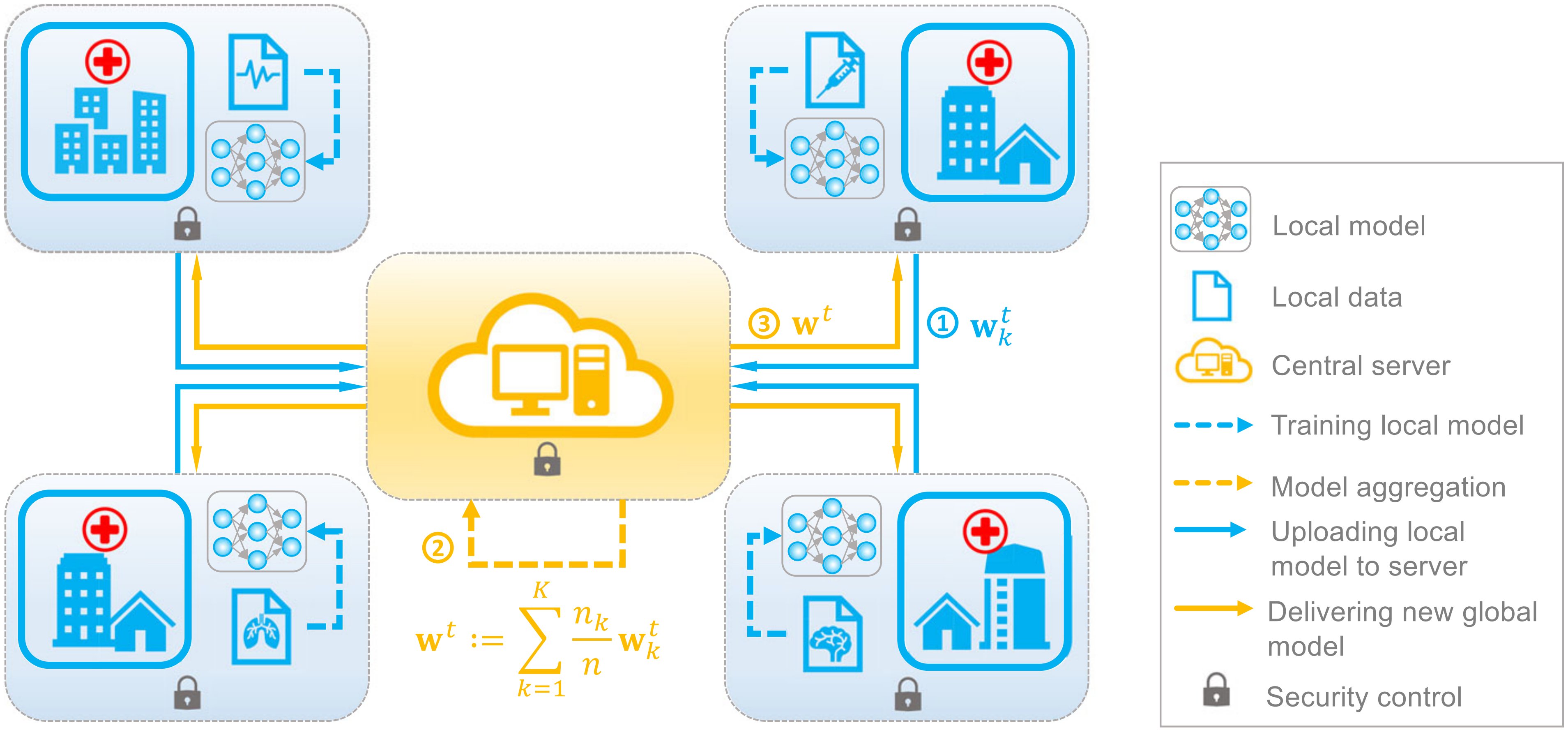
- At each step of gradient descent, instead of compute for all training samples, randomly pick a small subset (mini-batch) of training samples

$$w_{t+1} \leftarrow w_t - \eta \nabla f(w_t; x_k, y_k)$$



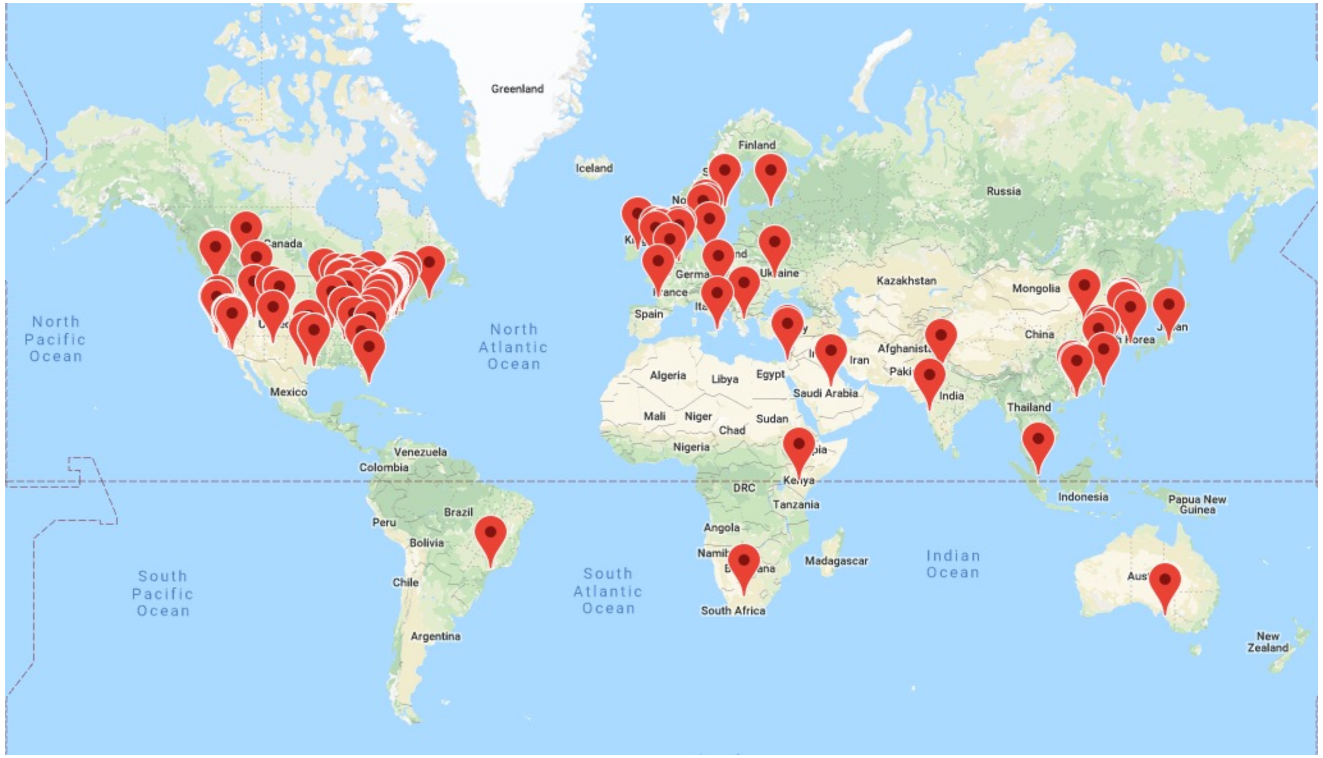
https://medium.com/analytics-vidhya/gradient-descent-vs-stochastic-gd-vs-mini-batch-sgd-fbd3a2cb4ba4

Federated Learning

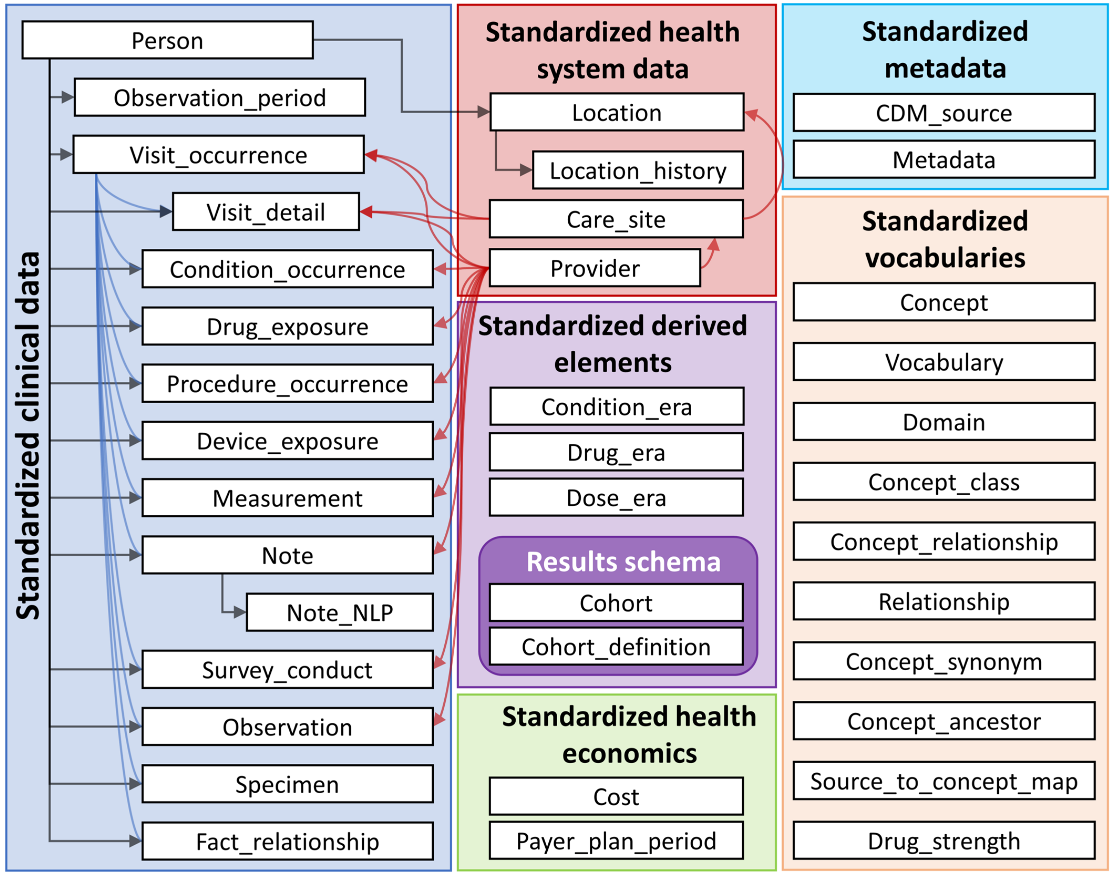


Xu, Jie, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and ***Fei Wang***. "Federated learning for healthcare informatics." *Journal of Healthcare Informatics Research* (2020): 1-19.

Clinical Research Networks



https://ohdsi.github.io/TheBookOfOhdsi/OhdsiCommunity.html



https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html

Federated SGD

- In a round t:
  - The central server broadcasts current model w_t to each client; each client k computes gradient: $g_k = \nabla F_k(w_t)$, on its local data.
    - Approach 1: Each client k submits g_k; the central server aggregates the gradients to generate a new model:
      - $w_{t+1} \leftarrow w_t - \eta \nabla f(w_t) = w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} g_k$.
    - Approach 2: Each client k computes: $w_{t+1}^k \leftarrow w_t - \eta g_k$; the central server performs aggregation:
      - $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

https://inst.eecs.berkeley.edu/~cs294-163/fa19/slides/federated-learning.pdf

Federated Averaging

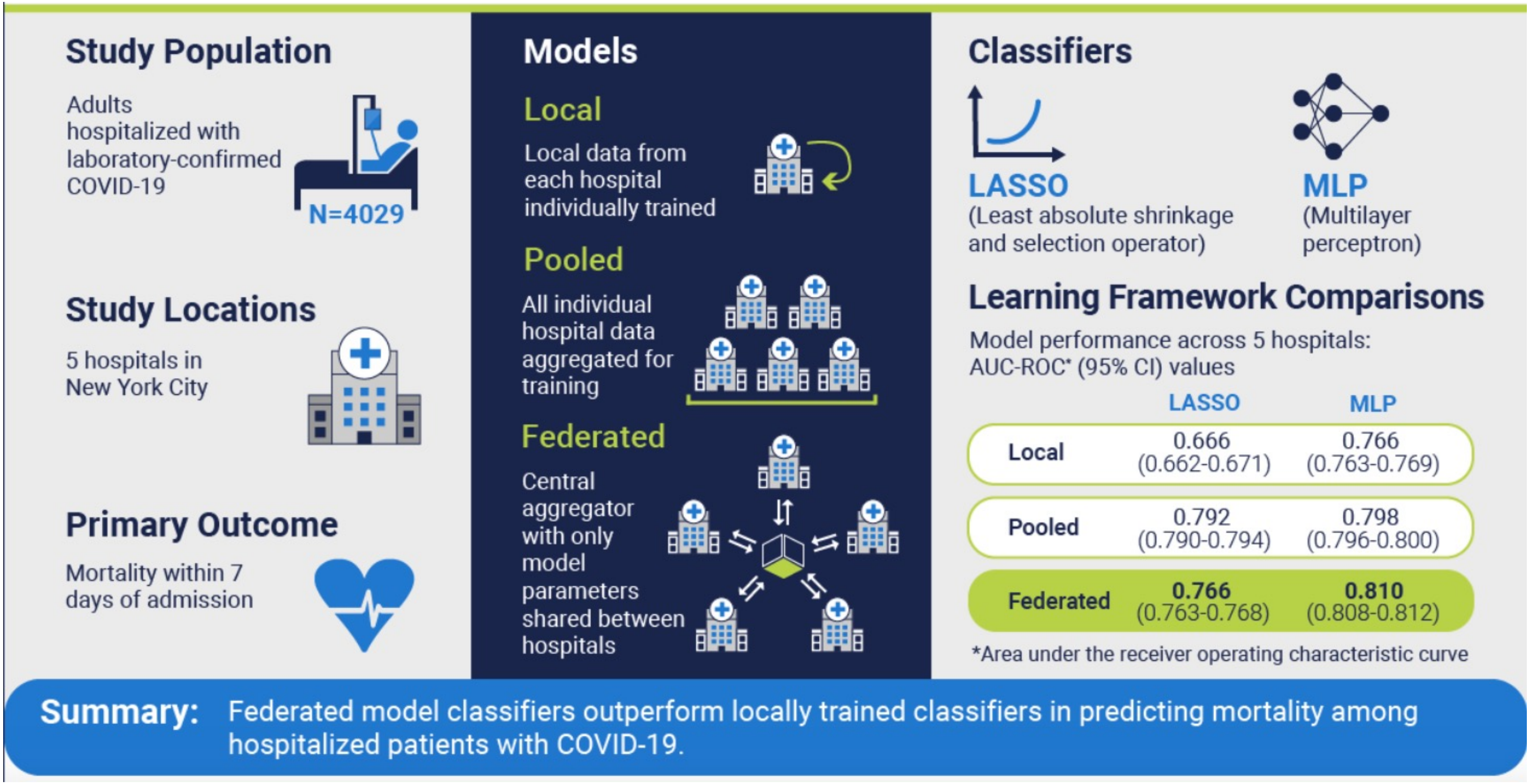
Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:
  initialize w_0
  for each round $t = 1, 2, \dots$ **do**
    $m \leftarrow \max(C \cdot K, 1)$
    $S_t \leftarrow$ (random set of m clients)
    for each client $k \in S_t$ **in parallel do**
      $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
    $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w)**:** // *Run on client k*
  $\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
  for each local epoch i from 1 to E **do**
    for batch $b \in \mathcal{B}$ **do**
      $w \leftarrow w - \eta \nabla \ell(w; b)$
  return w to server

1. At first, a model is randomly initialized on the central server.
2. For each round *t:*
   1. *A random set of clients are chosen;*
   2. *Each client performs local gradient descent steps;*
   3. *The server aggregates model parameters submitted by the clients.*

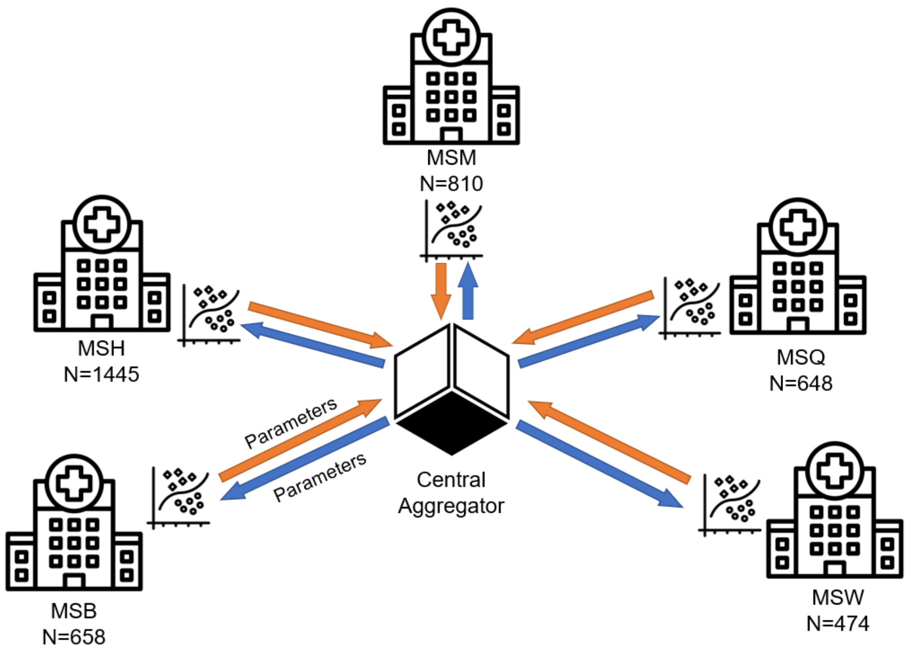
McMahan, Brendan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. "Communication-efficient learning of deep networks from decentralized data." In *Artificial intelligence and statistics*, pp. 1273-1282. PMLR, 2017.

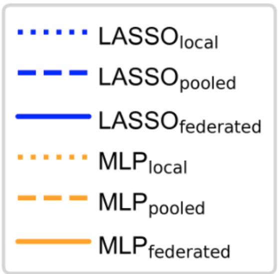


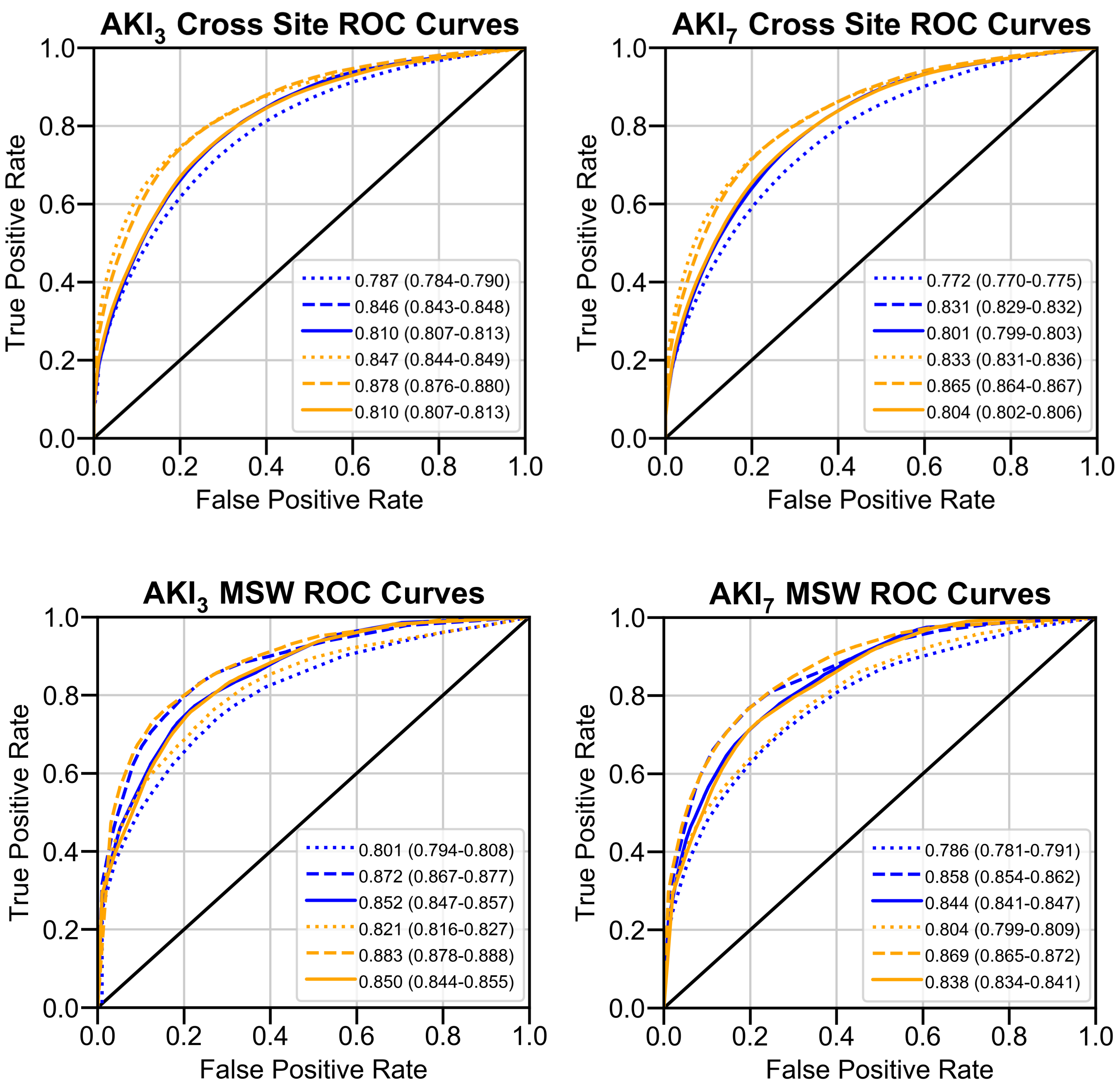
Akhil Vaid, Suraj K Jaladanki, Jie Xu, Shelly Teng, Arvind Kumar, Samuel Lee, Sulaiman Somani, Ishan Paranjpe, Jessica K De Freitas, Tingyi Wanyan, Kipp W Johnson, Mesude Bicak, Eyal Klang, Young Joon Kwon, Anthony Costa, Shan Zhao, Riccardo Miotto, Alexander W Charney, Erwin Böttinger, Zahi A Fayad, Girish N Nadkarni, **Fei Wang**, Benjamin S Glicksberg. "Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: Machine learning approach." *JMIR medical informatics* 9, no. 1 (2021): e24207.

Characteristic	Mount Sinai Brooklyn	Mount Sinai Hospital	Mount Sinai Morningside	Mount Sinai Queens	Mount Sinai West	*P* value
Number of patients, n	611	1644	749	540	485	—[b]
Gender, n (%)						
Male	338 (55.3)	951 (57.8)	411 (54.9)	344 (63.7)	257 (53.0)	.004
Female	273 (44.7)	693 (42.2)	338 (45.1)	196 (36.3)	228 (47.0)	.004
Age (years), median (IQR)	72.5 (63.6-82.7)	63.3 (51.3-73.2)	69.8 (57.4-80.3)	68.1 (57.1-78.8)	66.3 (52.5-77.6)	<.001
Ethnicity, n (%)						
Hispanic	21 (3.4)	460 (28.0)	259 (34.6)	198 (36.7)	111 (22.9)	<.001
Non-Hispanic	416 (68.1)	892 (54.3)	452 (60.3)	287 (53.1)	349 (72.0)	<.001
Unknown	174 (28.5)	292 (17.8)	38 (5.1)	55 (10.2)	25 (5.2)	<.001
Race, n (%)						
Asian	13 (2.1)	83 (5.0)	16 (2.1)	56 (10.4)	27 (5.6)	<.001
Black/African American	323 (52.9)	388 (23.6)	266 (35.5)	64 (11.9)	109 (22.5)	<.001
Other	54 (8.8)	705 (42.9)	343 (45.8)	288 (53.3)	164 (33.8)	<.001
Unknown	27 (4.4)	87 (5.3)	25 (3.3)	14 (2.6)	14 (2.9)	<.001
White	194 (31.8)	381 (23.2)	99 (13.2)	118 (21.9)	171 (35.3)	<.001

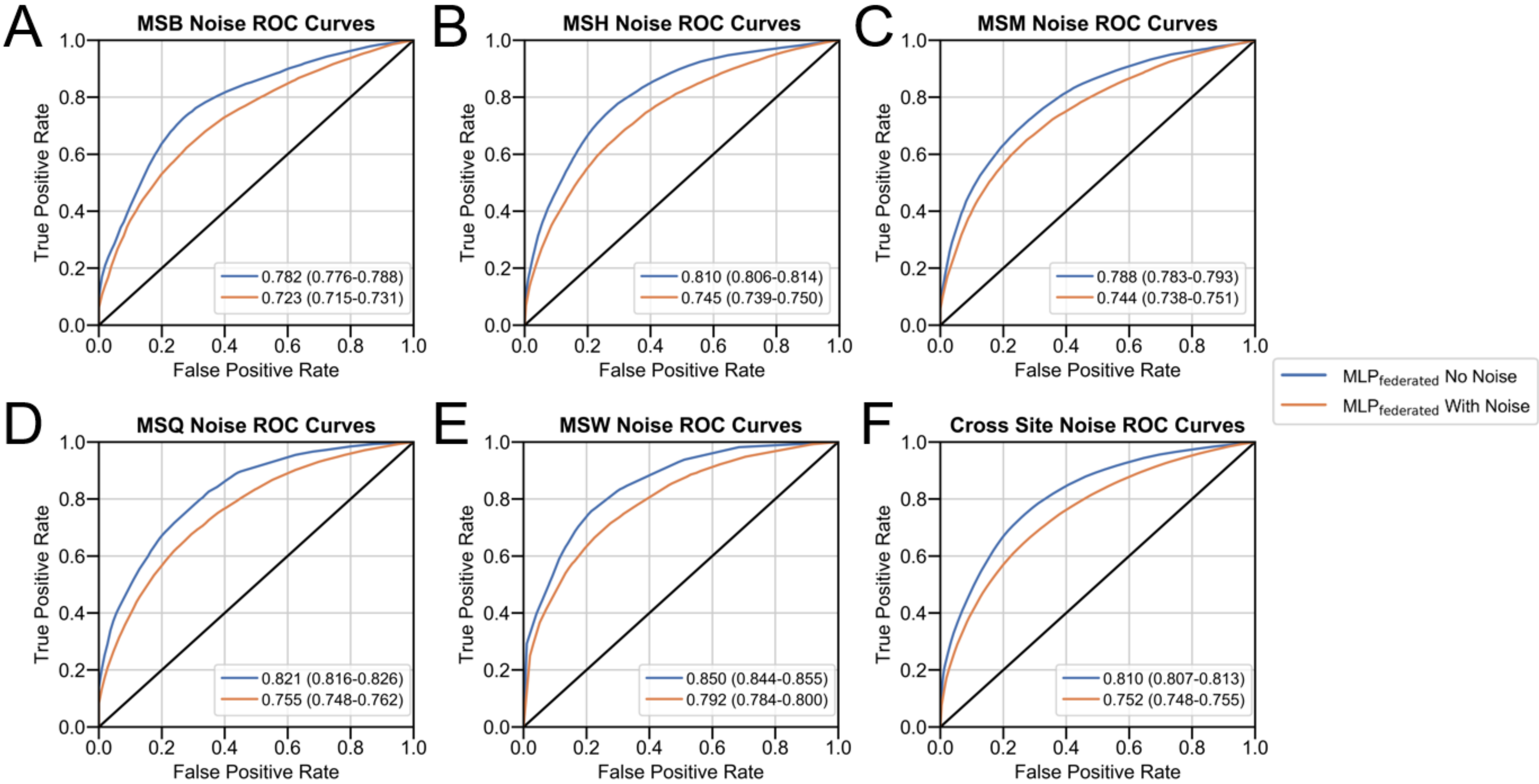
Model	Mount Sinai Brooklyn (n=611), AUROC (95% CI)	Mount Sinai Hospital (n=1644), AUROC (95% CI)	Mount Sinai Morningside (n=749), AUROC (95% CI)	Mount Sinai Queens (n=540), AUROC (95% CI)	Mount Sinai West (n=485), AUROC (95% CI)
LASSO model					
Local	0.791 (0.788-0.795)	0.693 (0.689-0.696)	0.66 (0.656-0.664)	0.706 (0.702-0.710)	0.482 (0.473-0.491)
Pooled	0.816 (0.814-0.819)	0.791 (0.788-0.794)	0.789 (0.785-0.792)	0.734 (0.730-0.737)	0.829 (0.824-0.834)
Federated	0.793 (0.790-0.796)	0.772 (0.769-0.774)	0.767 (0.764-0.771)	0.694 (0.690-0.698)	0.801 (0.796-0.807)
MLP model					
Local	0.822 (0.820-0.825)	0.750 (0.747-0.754)	0.747 (0.743-0.751)	0.791 (0.788 - 0.795)	0.719 (0.711-0.727)
Pooled	0.823 (0.820-0.826)	0.792 (0.789-0.795)	0.751 (0.747-0.755)	0.783 (0.779-0.786)	0.842 (0.837-0.847)
Federated (no noise	0.829 (0.826-0.832)	0.786 (0.782-0.789)	0.791 (0.788-0.795)	0.809 (0.806-0.812)	0.836 (0.83-0.841)







Jaladanki, Suraj K., Akhil Vaid, Ashwin S. Sawant, Jie Xu, Kush Shah, Sergio Dellepiane, Ishan Paranjpe, Lili Chan, Alexander W Charney, **Fei Wang**, Benjamin S Glicksberg, Karandeep Singh, Girish N Nadkarn "Development of a federated learning approach to predict acute kidney injury in adult hospitalized patients with COVID-19 in New York City." *medRxiv* (2021).



A
MSB Noise ROC Curves
0.782 (0.776-0.788)
0.723 (0.715-0.731)
B
MSH Noise ROC Curves
0.810 (0.806-0.814)
0.745 (0.739-0.750)
C
MSM Noise ROC Curves
0.788 (0.783-0.793)
0.744 (0.738-0.751)
D
MSQ Noise ROC Curves
0.821 (0.816-0.826)
0.755 (0.748-0.762)
E
MSW Noise ROC Curves
0.850 (0.844-0.855)
0.792 (0.784-0.800)
F
Cross Site Noise ROC Curves
0.810 (0.807-0.813)
0.752 (0.748-0.755)
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
$MLP_{federated}$ No Noise
$MLP_{federated}$ With Noise

nature medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄

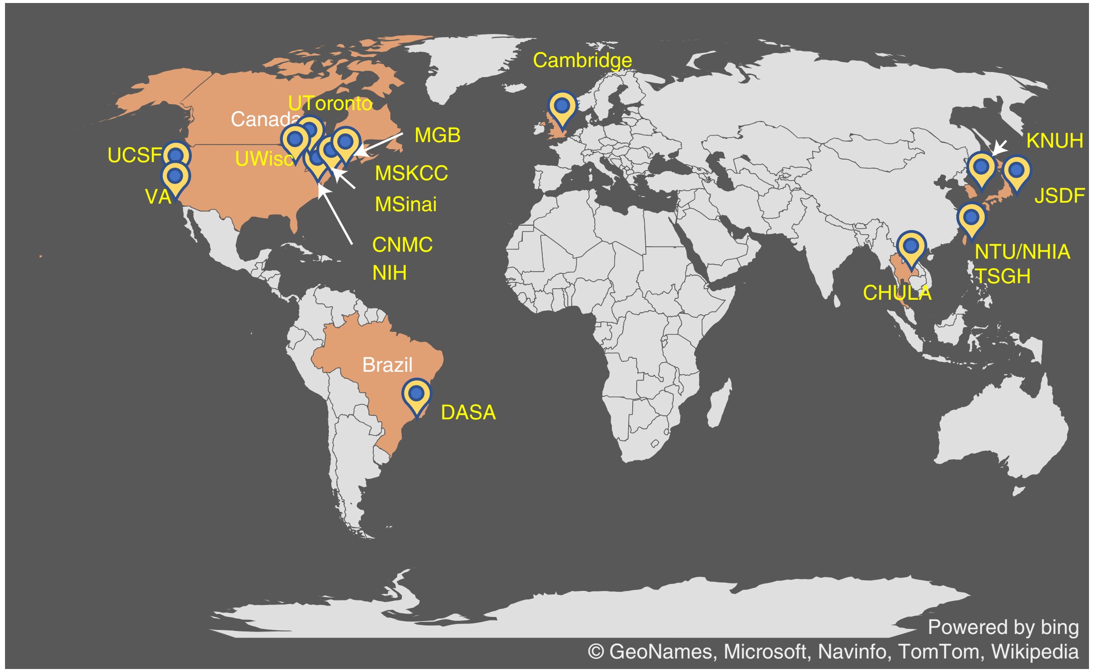
nature › nature medicine › articles › article

Article | Published: 15 September 2021

Federated learning for predicting clinical outcomes in patients with COVID-19

Ittai Dayan, Holger R. Roth, ... Quanzheng Li + Show authors

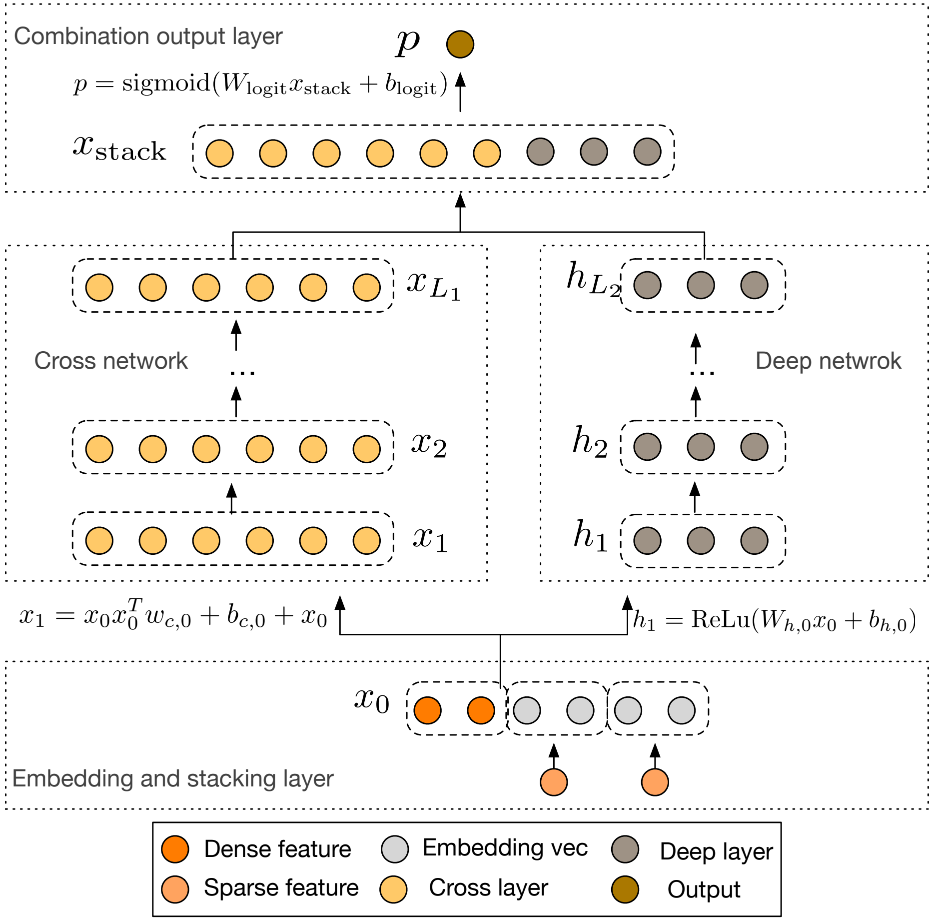
Nature Medicine **27**, 1735–1743 (2021) | Cite this article



Federated learning (FL) is a method used for training artificial intelligence models with data from multiple sources while maintaining data anonymity, thus removing many barriers to data sharing. Here we used data from 20 institutes across the globe to train a FL model, called EXAM (electronic medical record (EMR) chest X-ray AI model), that predicts the future oxygen requirements of symptomatic patients with COVID-19 using inputs of vital signs, laboratory data and chest X-rays. EXAM achieved an average area under the curve (AUC) >0.92 for predicting outcomes at 24 and 72 h from the time of initial presentation to the emergency room, and it provided 16% improvement in average AUC measured across all participating sites and an average increase in generalizability of 38% when compared with models trained at a single site using that site's data. For prediction of mechanical ventilation treatment or death at 24 h at the largest independent test site, EXAM achieved a sensitivity of 0.950 and specificity of 0.882. In this study, FL facilitated rapid data science collaboration without data exchange and generated a model that generalized across heterogeneous, unharmonized datasets for prediction of clinical outcomes in patients with COVID-19, setting the stage for the broader use of FL in healthcare.

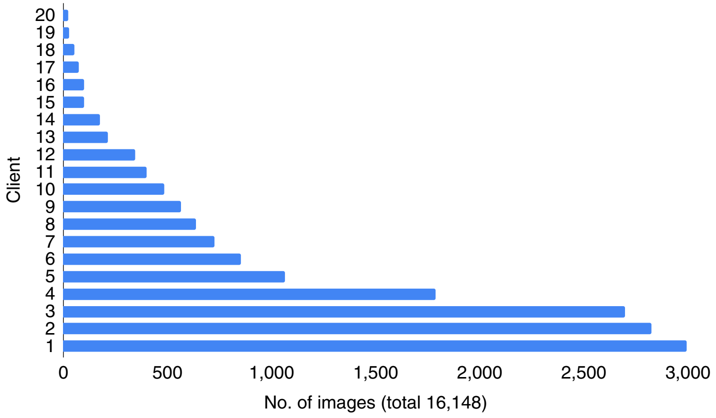
Data Types and Model Architecture

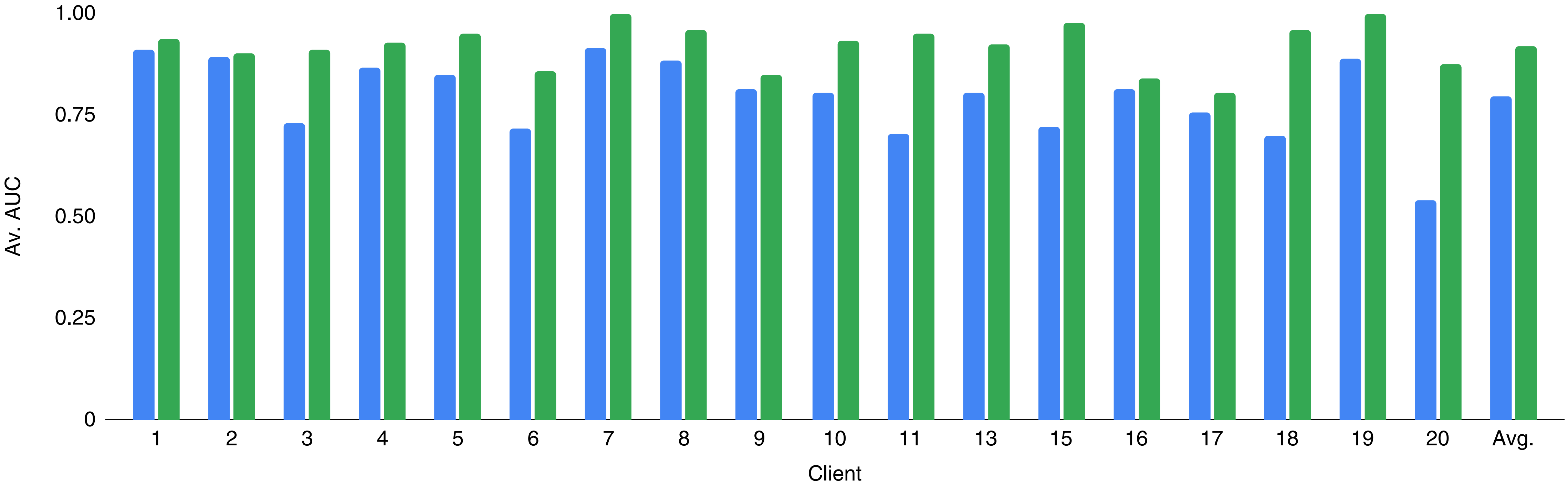
Category	Subcategory	Component name	Definition	Units	LOINC code
Demographic	–	Patient age	–	Years	30525-0
Imaging	Portable CXR	–	AP or PA portable CXR	–	36554-4
Lab value	C-reactive protein	C-reactive protein	Blood c-reactive protein concentration	mg l^{-1}	1988-5
Lab value	Complete blood count (CBC)	Neutrophils	Blood absolute neutrophils	10^9 l^{-1}	751-8
Lab value	CBC	White blood cells	Blood white blood cell count	10^9 l^{-1}	33256-9
Lab value	D-dimer	D-dimer	Blood D-dimer concentration	ng ml^{-1}	7799-0
Lab value	Lactate	Lactate	Blood lactate concentration	mmol l^{-1}	2524-7
Lab value	Lactate dehydrogenase	LDH	Blood LDH concentration	U l^{-1}	2532-0
Lab value	Metabolic panel	Creatinine	Blood creatinine concentration	mg dl^{-1}	2160-0
Lab value	Procalcitonin	Procalcitonin	Blood procalcitonin concentration	ng ml^{-1}	33959-8
Lab value	Metabolic panel	eGFR	Estimated glomerular filtration rate	ml min^{-1} 1.73 m^{-2}	69405-9
Lab value	Troponin	Troponin-T	Blood troponin concentration	ng ml^{-1}	67151-1
Lab value	Hepatic panel	AST	Blood aspartate aminotransferase concentration	IU l^{-1}	1920-8
Lab value	Metabolic panel	Glucose	Blood glucose concentration	mg dl^{-1}	2345-7
Vital sign	–	Oxygen saturation	Oxygen saturation	%	59408-5
Vital sign	–	Systolic blood pressure	Systolic BP	mmHg	8480-6
Vital sign	–	Diastolic blood pressure	Diastolic BP	mmHg	8462-4
Vital sign	–	Respiratory rate	Respiratory rate	Breaths min^{-1}	9279-1
Vital sign		COVID PCR test	PCR for RNA (not used as input to model)		95425-5
Vital sign	Oxygen device used in ED	Oxygen device	Ventilation, high-flow/NIV, low-flow, room air	–	41925-9
Outcome	24-h oxygen device	Oxygen device	Ventilation, high-flow/NIV, low-flow, room air	–	41925-9
Outcome	72-h oxygen device	Oxygen device	Ventilation, high-flow/NIV, low-flow, room air	–	41925-9
Outcome	Death	–	–	–	–
Outcome	Time of death	–	–	Hours	–



Wang, Ruoxi, Bin Fu, Gang Fu, and Mingliang Wang. "Deep & cross network for ad click predictions." In *Proceedings of the ADKDD'17*, pp. 1-7. 2017.

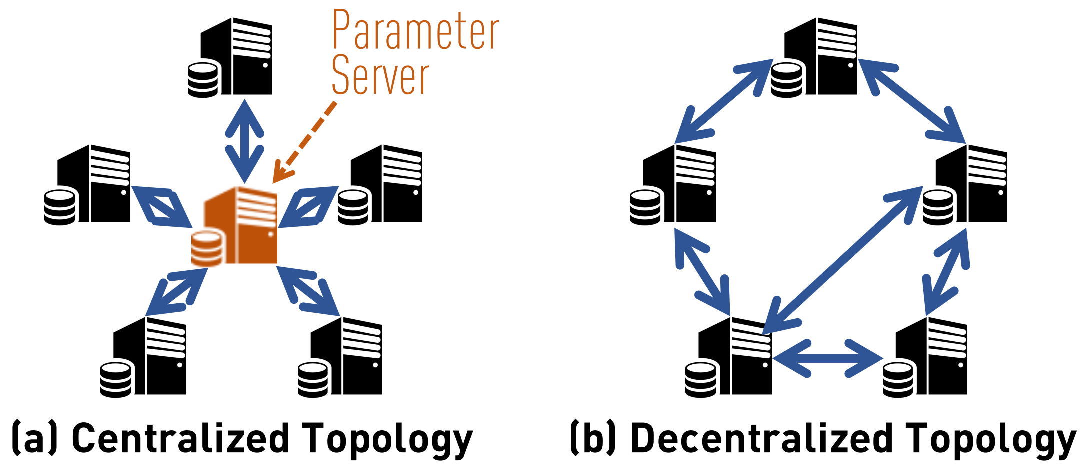
Model Performance

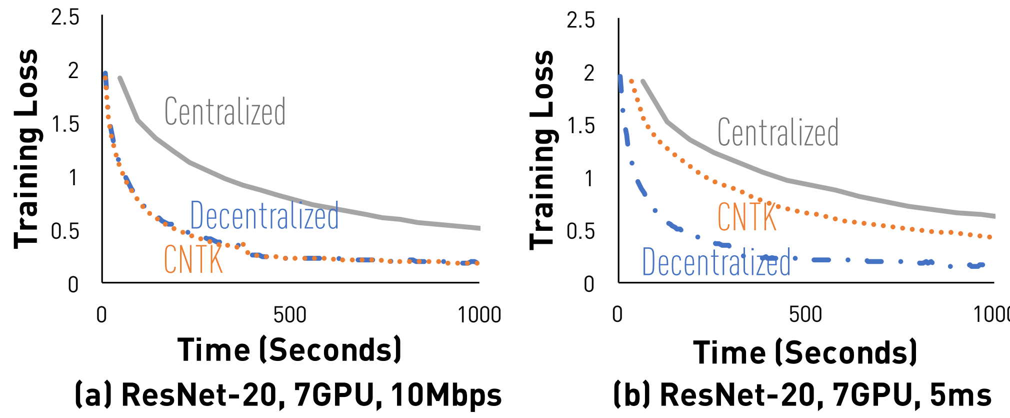




Client	1	2	3	4	5	6	7	8	9	10	11	13	15	16	17	18	19	20	Av.
Local	0.910	0.892	0.731	0.869	0.848	0.716	0.916	0.887	0.816	0.803	0.702	0.805	0.722	0.812	0.755	0.698	0.889	0.542	0.795
FL (gl. best)	0.938	0.902	0.912	0.929	0.950	0.857	1.000	0.961	0.849	0.935	0.950	0.925	0.979	0.839	0.806	0.958	1.000	0.875	0.920

Decentralized Optimization





Algorithm 1 Decentralized Parallel Stochastic Gradient Descent (D-PSGD) on the ith node

Require: initial point $x_{0,i} = x_0$, step length γ, weight matrix W, and number of iterations K

1: **for** $k = 0, 1, 2, \ldots, K-1$ **do**

2: Randomly sample $\xi_{k,i}$ from local data of the i-th node

3: Compute a local stochastic gradient based on $\xi_{k,i}$ and current optimization variable $x_{k,i}$: $\nabla F_i(x_{k,i}; \xi_{k,i})$ [a]

4: Compute the neighborhood weighted average by fetching optimization variables from neighbors: $x_{k+\frac{1}{2},i} = \sum_{j=1}^{n} W_{ij} x_{k,j}$ [b]

5: Update the local optimization variable $x_{k+1,i} \leftarrow x_{k+\frac{1}{2},i} - \gamma \nabla F_i(x_{k,i}; \xi_{k,i})$ [c]

6: **end for**

7: **Output:** $\frac{1}{n} \sum_{i=1}^{n} x_{K,i}$ [d]

Lian, Xiangru, Ce Zhang, Huan Zhang, Cho-Jui Hsieh, Wei Zhang, and Ji Liu. "Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent." *Advances in Neural Information Processing Systems* 30 (2017).

Algorithm 1 AD-PSGD (logical view)

Require: Initialize local models $\{x_0^i\}_{i=1}^n$ with the same initialization, learning rate γ, batch size M, and total number of iterations K.

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: Randomly sample a worker i_k of the graph G and randomly sample an averaging matrix W_k which can be dependent on i_k.

3: Randomly sample a batch
$$\xi_k^{i_k} := (\xi_{k,1}^{i_k}, \xi_{k,2}^{i_k}, \ldots, \xi_{k,M}^{i_k})$$
from local data of the i_k-th worker.

4: Compute the stochastic gradient locally
$$g_k(\hat{x}_k^{i_k}; \xi_k^{i_k}) := \sum_{j=1}^{M} \nabla F(\hat{x}_k^{i_k}; \xi_{k,j}^{i_k})$$
.

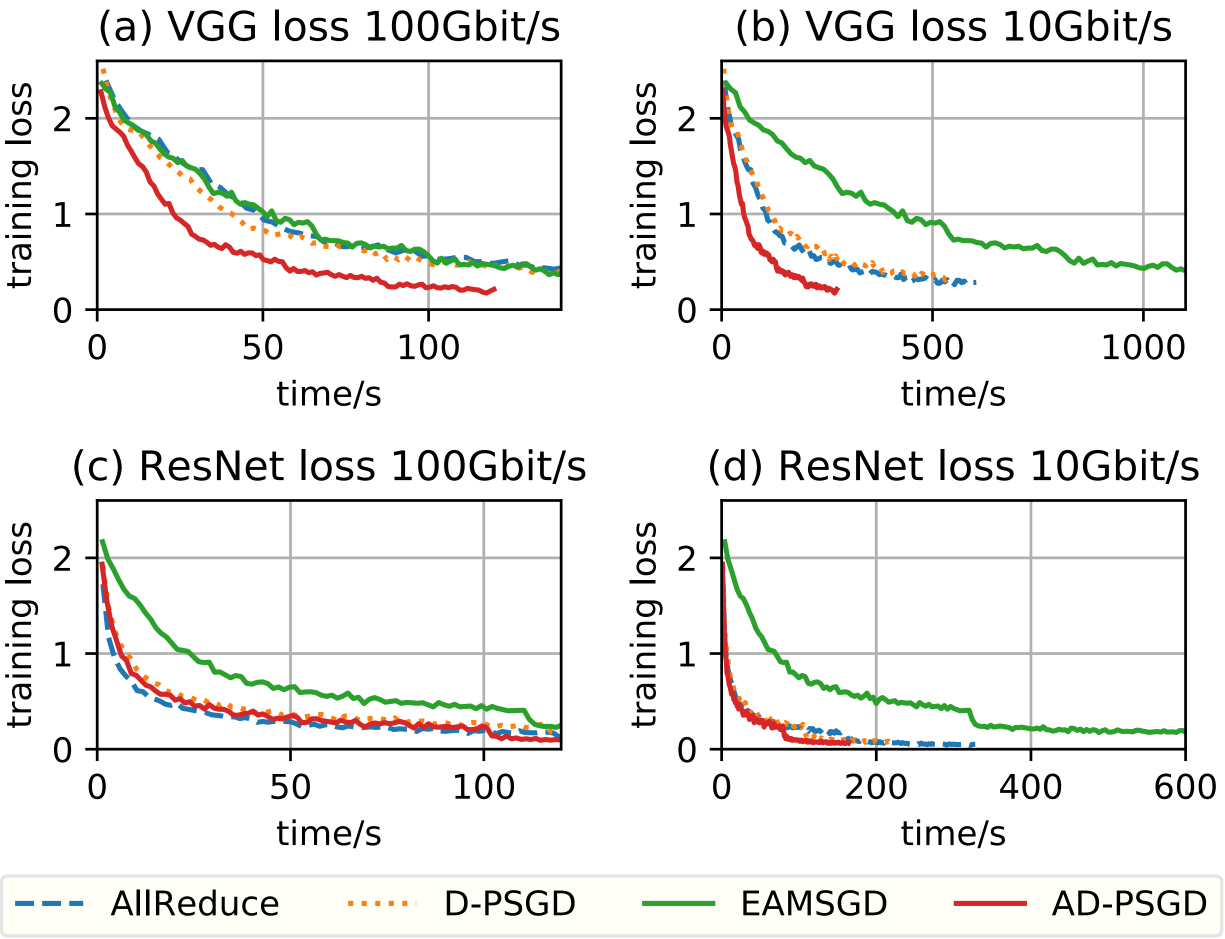
5: Average local models by [a]
$$[x_{k+1/2}^1, x_{k+1/2}^2, \ldots, x_{k+1/2}^n] \leftarrow [x_k^1, x_k^2, \ldots, x_k^n] W_k$$

6: Update the local model
$$x_{k+1}^{i_k} \leftarrow x_{k+1/2}^{i_k} - \gamma g_k(\hat{x}_k^{i_k}; \xi_k^{i_k}),$$
$$x_{k+1}^j \leftarrow x_{k+1/2}^j, \forall j \neq i_k.$$

7: **end for**

8: Output the average of the models on all workers for inference.

[a] Note that Line 4 and Line 5 can run in parallel.



Lian, Xiangru, Wei Zhang, Ce Zhang, and Ji Liu. "Asynchronous decentralized parallel stochastic gradient descent." In *International Conference on Machine Learning*, pp. 3043-3052. PMLR, 2018.

Algorithm 1 Differential Private AD-PSGD

Initialization: Initialize all local models $\{\mathbf{w}_k^0\}_{k=1}^K \in \mathbb{R}^d$ with $\mathbf{w}^0$, learning rate η, batch size B, privacy budget (ϵ, δ), and total number of iterations T.

Output: (ϵ, δ)-differentially private local models.

for $<t = 0, 1, ..., T-1>$ **do**

  Randomly sample a worker k^t of the graph G and randomly sample an doubly stochastic averaging matrix $\mathbf{A}_t \in \mathbb{R}^{K \times K}$ dependent on k^t;

  Randomly sample a batch $\xi_{k^t}^t := (\xi_{k^t}^{t,1}, \xi_{k^t}^{t,2}, ..., \xi_{k^t}^{t,B}) \in \mathbb{R}^{d \times B}$ from local data of the k^t-th worker with the sampling probability $\frac{B}{n_{k^t}}$;

  Compute stochastic gradient $g^t(\hat{\mathbf{w}}_{k^t}^t; \xi_{k^t}^t)$ locally

$$g^t(\hat{\mathbf{w}}_{k^t}^t; \xi_{k^t}^t) := \sum_{i=1}^{B} \nabla F_{k^t}(\hat{\mathbf{w}}_{k^t}^t; \xi_{k^t}^{t,i}) \tag{12}$$

  Add noise

$$\tilde{g}^t(\hat{\mathbf{w}}_{k^t}^t; \xi_{k^t}^t) = g^t(\hat{\mathbf{w}}_{k^t}^t; \xi_{k^t}^t) + \mathbf{n},$$

  where $\mathbf{n} \in \mathbb{R}^d \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ and σ is defined in Theorem 1.

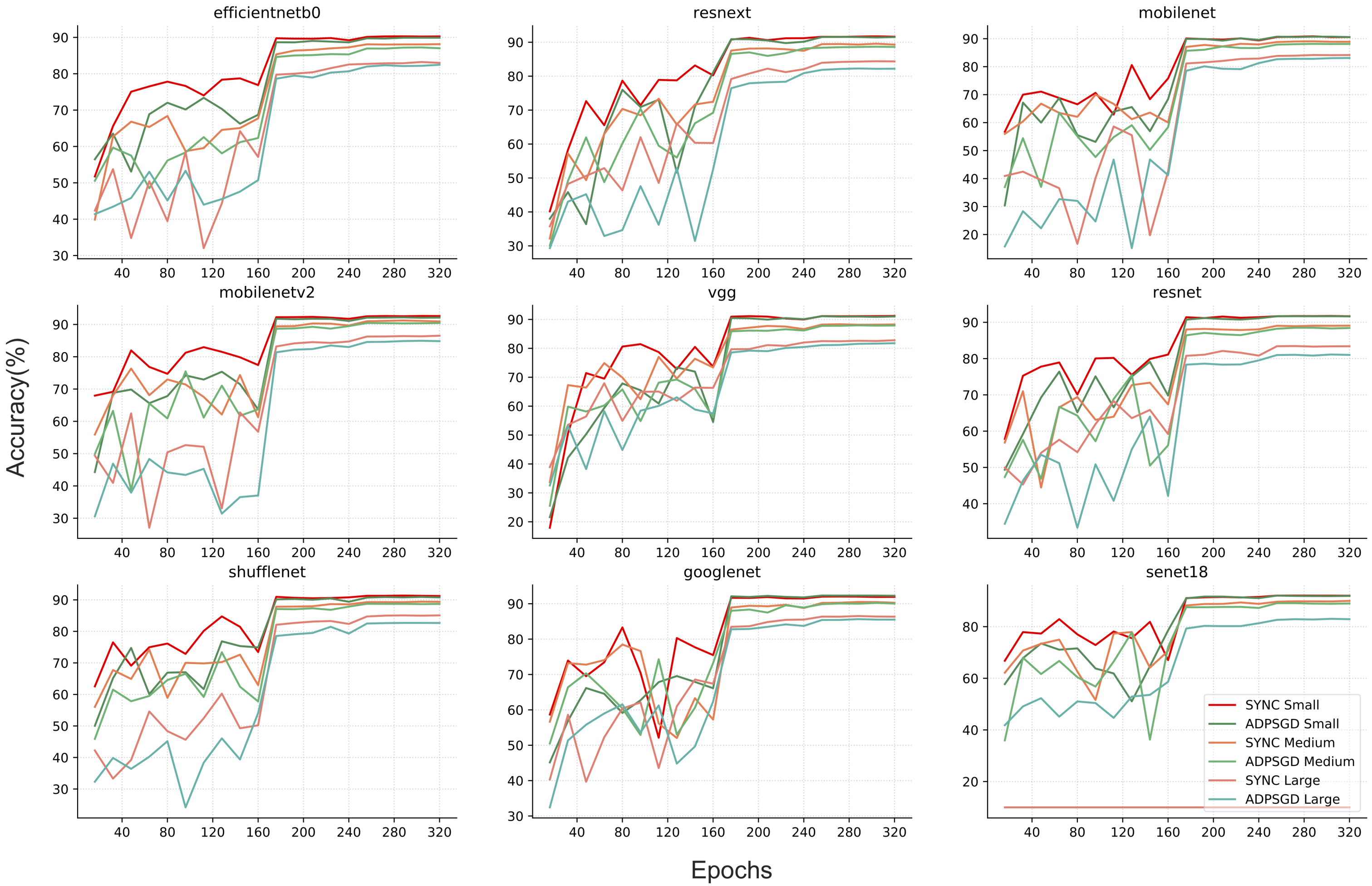
  Average local models by

$$[\mathbf{w}_1^{t+1/2}, \mathbf{w}_2^{t+1/2}, ..., \mathbf{w}_K^{t+1/2}] \leftarrow [\mathbf{w}_1^t, \mathbf{w}_2^t, ..., \mathbf{w}_K^t]\mathbf{A}_t; \tag{13}$$

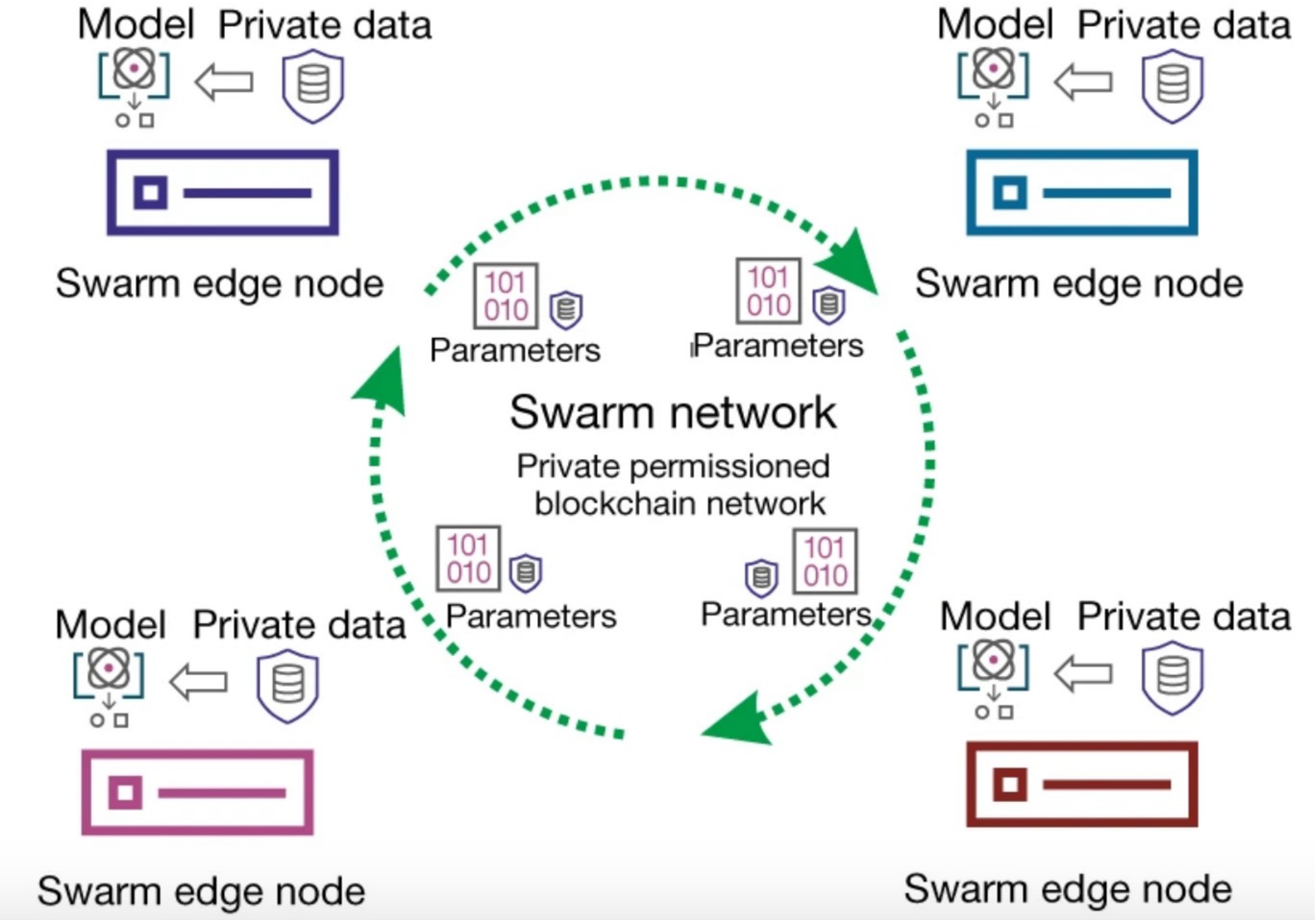
  Update the local model:

$$\mathbf{w}_{k^t}^{t+1} \leftarrow \mathbf{w}_{k^t}^{t+1/2} - \eta \tilde{g}^t(\hat{\mathbf{w}}_{k^t}^t; \xi_{k^t}^t),$$
$$\mathbf{w}_j^{t+1} \leftarrow \mathbf{w}_j^{t+1/2}, \forall j \neq k^t.$$

end for

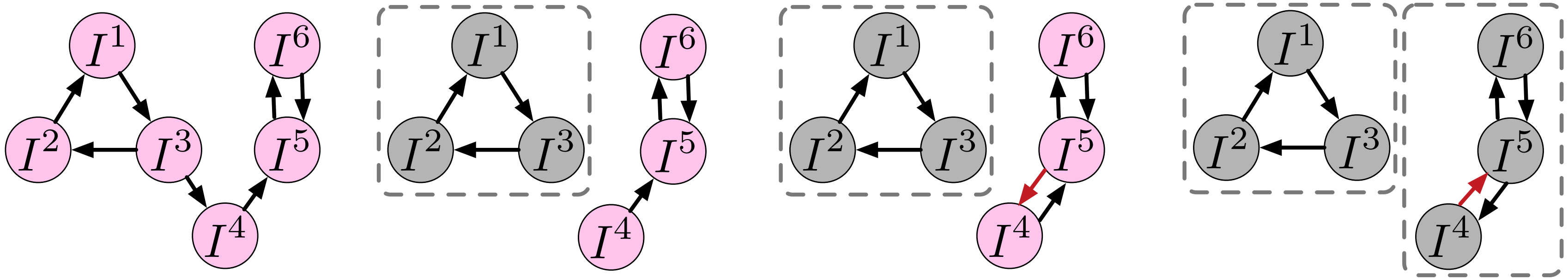


Xu, Jie, Wei Zhang, and **Fei Wang**. "A (DP) $^ 2$ SGD: Asynchronous Decentralized Parallel Stochastic Gradient Descent with Differential Privacy." *IEEE TPAMI To Appear* (2021).



Warnat-Herresthal, Stefanie, Hartmut Schultze, Krishnaprasad Lingadahalli Shastry, Sathyanarayanan Manamohan, Saikat Mukherjee, Vishesh Garg, Ravi Sarveswara et al. "Swarm Learning for decentralized and confidential clinical machine learning." *Nature* 594, no. 7862 (2021): 265-270.

Learning to Collaborate

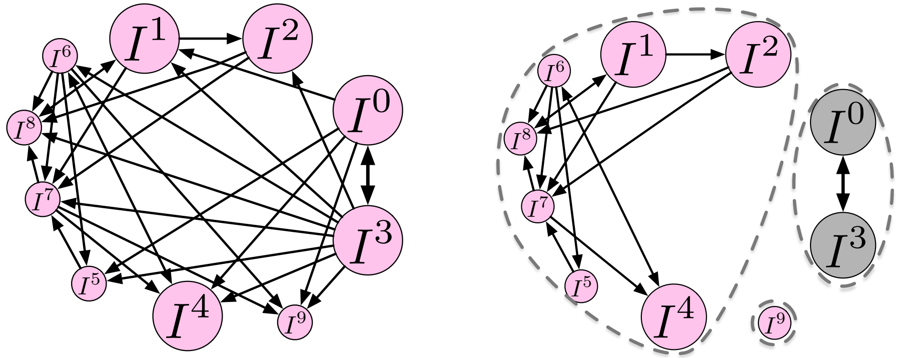


(a) Benefit Graph (b) Finding stable coalition (c) Re-build Benefit Graph (d) Collaboration Equilibrium

Algorithm 1: Achieving collaboration equilibrium

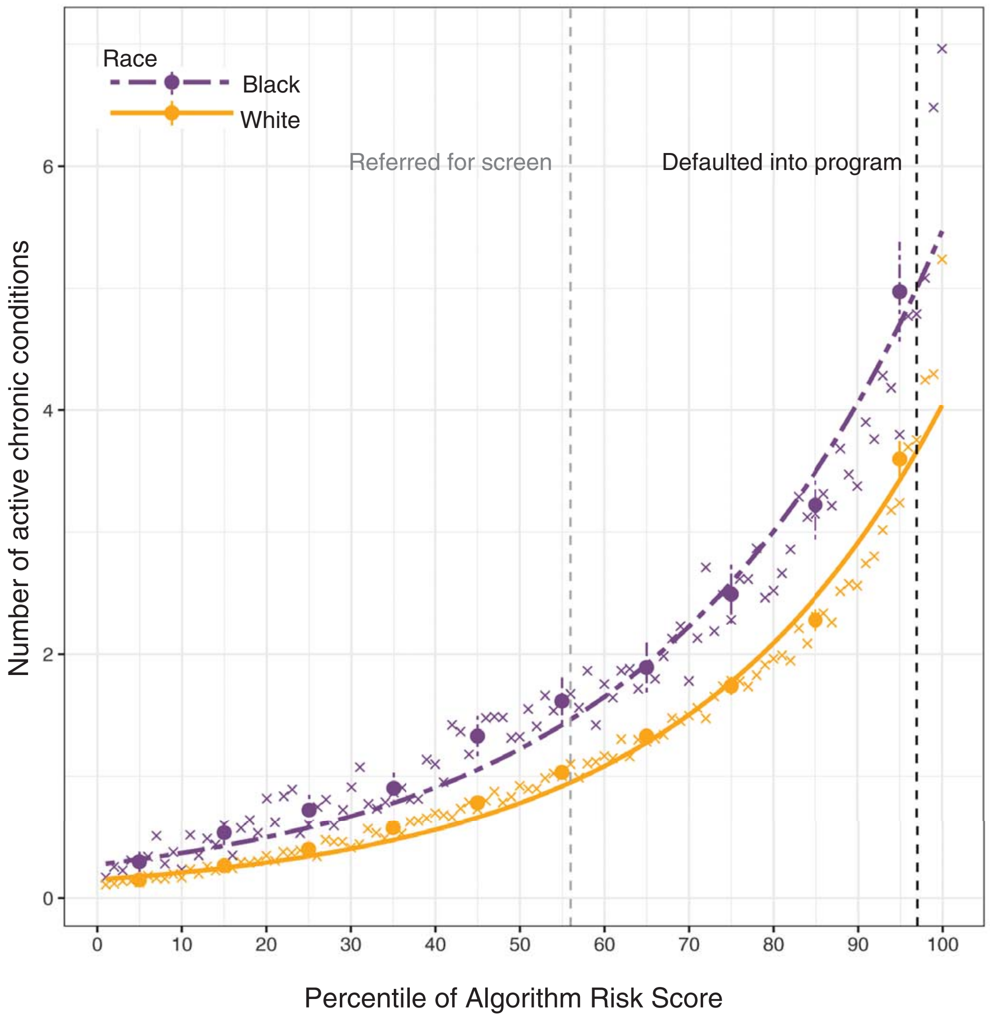
Input: N institutions $\boldsymbol{I} = \{I^i\}_{i=1}^N$ seeking collaborating with others
Set original client set $C \leftarrow \boldsymbol{I}$;
Set collaboration strategy $S \leftarrow \emptyset$;
while $C \neq \emptyset$ **do**
- **forall** *client* $I^i \in C$ **do**
  - Determine the OCS of I^i by SPO;
- Construct the benefit graph $BG(C)$;
- Search for all strongly connected components $\{C^1, C^2, ...C^k\}$ of $BG(C)$ using **Tarjan** algorithm;
- **forall** i = *1, 2, 3,... k* **do**
  - **if** C^i *is stable coalition* **then**
    - $C \leftarrow C \backslash C^i$;
    - $S \leftarrow S \cup \{C^i\}$;

Output: collaboration strategy S



Sen Cui, Jian Liang, Weishen Pan, Kun Chen, Changshui Zhang, **Fei Wang**. Learning to Collaborate. https://arxiv.org/abs/2108.07926 . 2021.

Model Bias



Science Contents News Careers Journals

SHARE

RESEARCH ARTICLE

Dissecting racial bias in an algorithm used to manage the health of populations

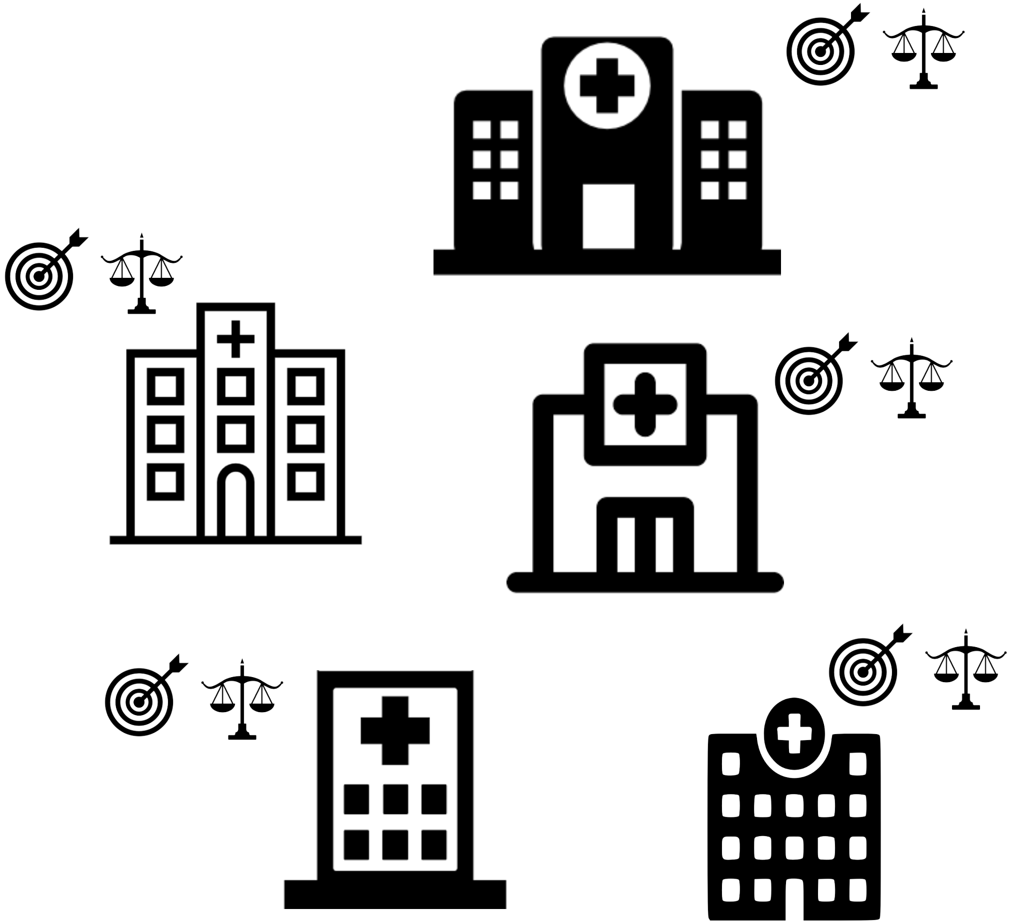
Ziad Obermeyer[1,2,*], Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5,*,†]

+ See all authors and affiliations

Science 25 Oct 2019:
Vol. 366, Issue 6464, pp. 447-453
DOI: 10.1126/science.aax2342

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

Federated Fairness



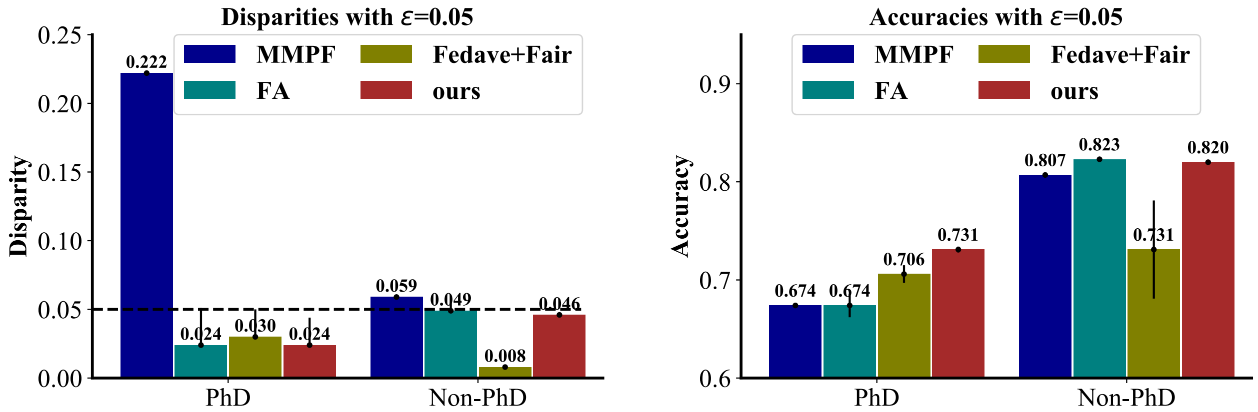
Definition 1 (Multi-client fairness (MCF)). *A learned model h achieves multi-client fairness if h meets the following condition:*

$$\Delta Dis_k(h) - \epsilon_k \leq 0 \quad \forall k \in \{1, ., N\} \tag{1}$$

where $\Delta Dis_k(h)$ denotes the disparity induced by the model h and ϵ_k is the given fairness budget of the k-th client. The disparity on the k-th client ΔDis_k can be measured by *demographic parity* (DP) [8] and *Equal Opportunity* (EO) [9] as follows:

$$\begin{aligned} \Delta DP_k &= |P(\hat{Y}^k = 1|A^k = 0) - P(\hat{Y}^k = 1|A^k = 1)| \\ \Delta EO_k &= |P(\hat{Y}^k = 1|A^k = 0, Y^k = 1) - P(\hat{Y}^k = 1|A^k = 1, Y^k = 1)| \end{aligned} \tag{2}$$

$$\min_{h \in \mathcal{H}} \quad [l_1(h), l_2(h), ...l_N(h)] \quad \text{s.t. } g_k(h) - \epsilon_k \leq 0 \quad \forall k \in \{1, ., N\}$$



Sen Cui, Weishen Pan, Jian Liang, Changshui Zhang, **Fei Wang**. Addressing Algorithmic Disparity and Performance Inconsistency in Federated Learning. https://arxiv.org/abs/2108.08435 . NeurIPS 2021.

Conclusions

- Clinical problems are typically complicated with limited sample size. Clinical data are sensitive. All these make federated learning important.
  - Data standardization/harmonization is important before federated learning can be applied.
  - To further protect the privacy, differential privacy/block chain techniques could be helpful.
  - Incentives/benefits are important to consider for participating in federated learning.
  - In addition to model accuracy, model fairness could be important as well.

Thank You!

feiwang.cornell@gmail.com

@feiwang03

https://wcm-wanglab.github.io/index.html