
CAP 5516

Medical Image Computing (Spring 2022)

Dr. Chen Chen

Center for Research in Computer Vision (CRCV)

University of Central Florida

Office: HEC 221

Address: 4328 Scorpius St., Orlando, FL 32816-2365

Email: chen.chen@crcv.ucf.edu

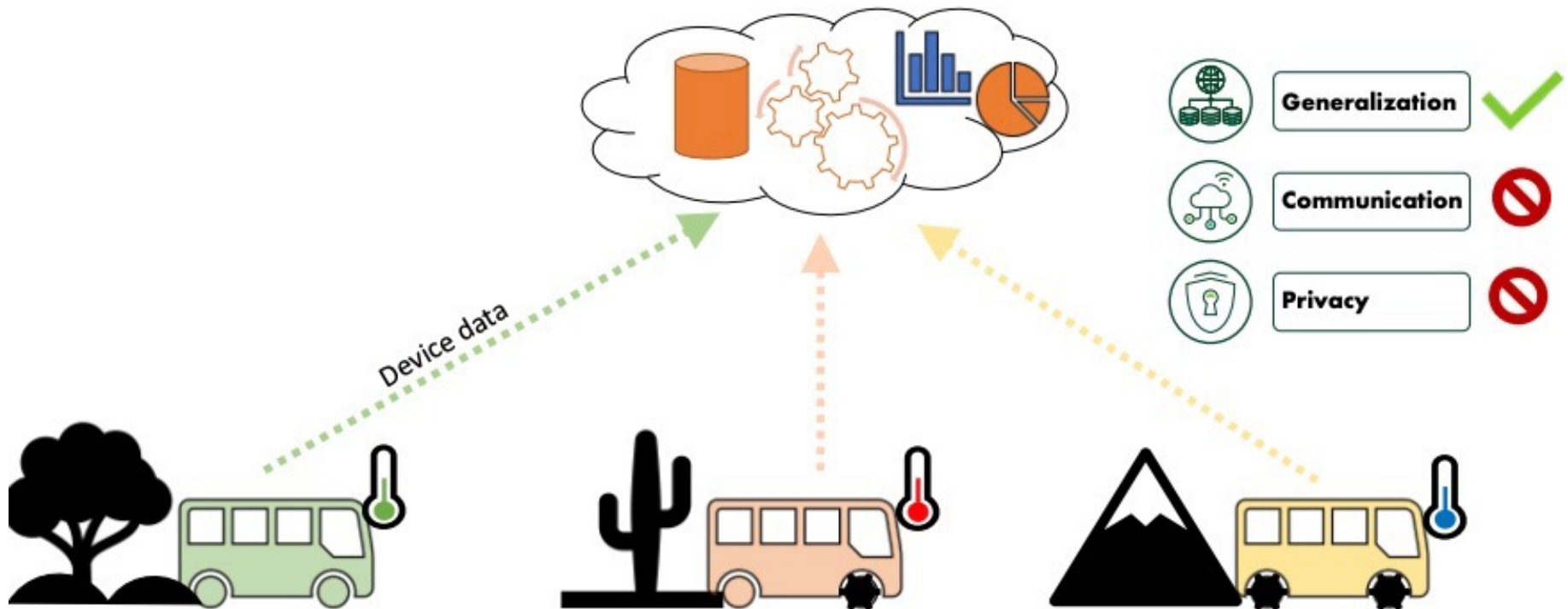
Web: <https://www.crcv.ucf.edu/chenchen/>

Lecture 13

Federated Learning and Its Application in Medical Image Computing

Centralized Machine Learning

- Traditional centralized learning
 - ML runs in the cloud, gathering info from all connected devices and sending back a model.



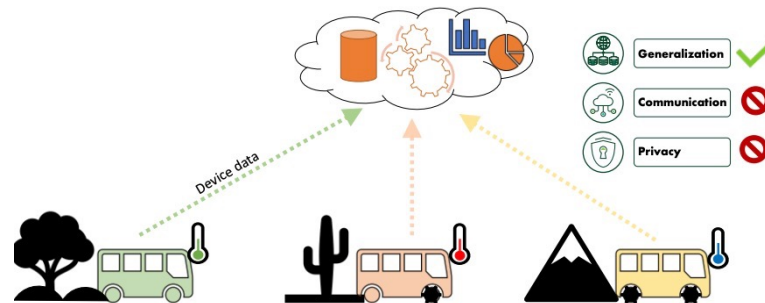
Ekkono Solutions, Short White Paper, May-20, SWP-openfika7-2005-01 “Federated Learning” https://www.ekkono.ai/wp-content/uploads/2020/12/SWP_Federated_Learning_Ekkono_Solutions_May_2020.pdf



Centralized Machine Learning

- Challenges

- **Connectivity** - data must be transmitted over a stable connection
- **Bandwidth** – e.g. a new ABB electrical substation could generate 5 GB/s. IoT devices are often mobile or distributed in very varying environments making just upholding a stable internet connection a challenge.
- **Latency** - real-time applications, e.g. automation, requires very low latency
- **Privacy** - sensitive operational data must remain on site
- **High demands on storage and computing capacities of the server**



A Shift of Paradigm: From Centralized to Decentralized Data

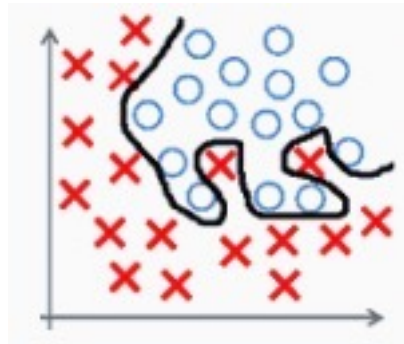
- How about each party learning on its own?



Figure credit: Aurélien Bellet (Inria)

A Shift of Paradigm: From Centralized to Decentralized Data

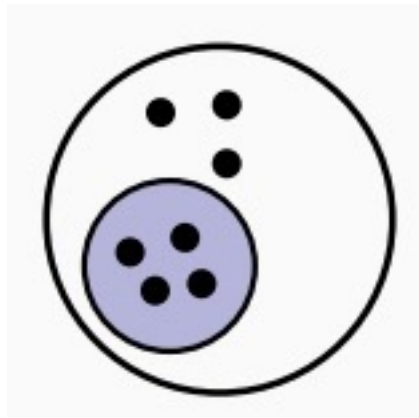
- How about each party learning on its own?
- The local dataset may be too small
 - Sub-par predictive performance (e.g., due to overfitting)
 - Non-statistically significant results (e.g., medical studies)



Credit: Aurélien Bellet (Inria)

A Shift of Paradigm: From Centralized to Decentralized Data

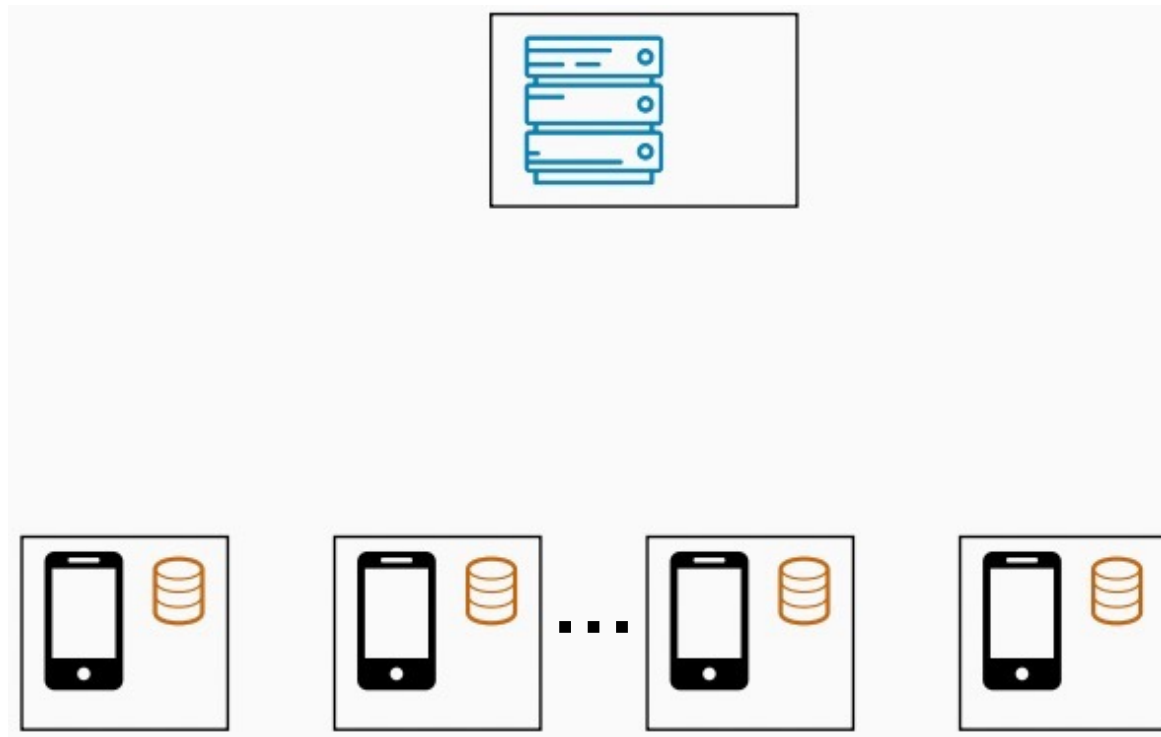
- How about each party learning on its own?
- The local dataset may be too small
 - Sub-par predictive performance (e.g., due to overfitting)
 - Non-statistically significant results (e.g., medical studies)
- The local dataset may be biased
 - Not representative of the target distribution



Credit: Aurélien Bellet (Inria)

Federated Learning

- A broad definition of federated learning
 - Federated Learning (FL) aims to collaboratively train an ML model while keeping the data decentralized

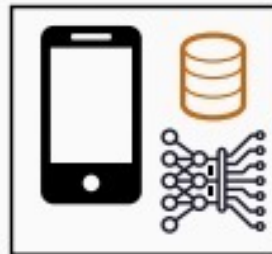
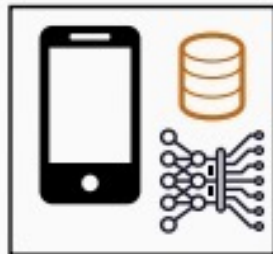
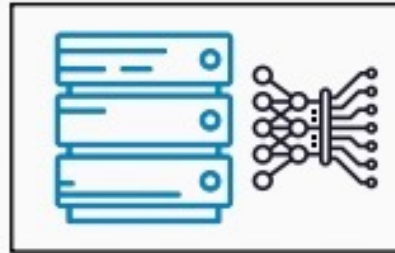


Credit: Aurélien Bellet (Inria)

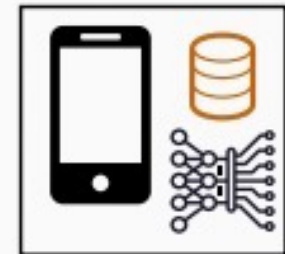
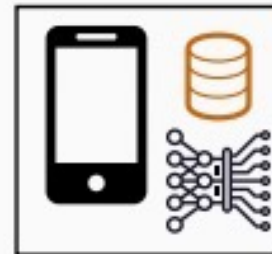
Federated Learning

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

initialize model



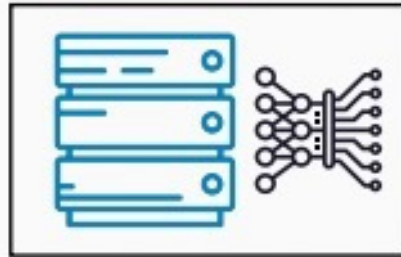
...



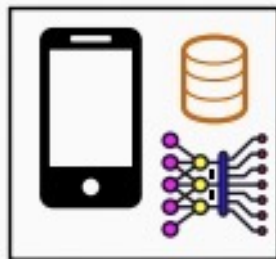
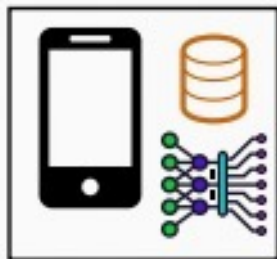
Federated Learning

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

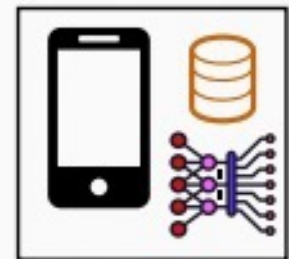
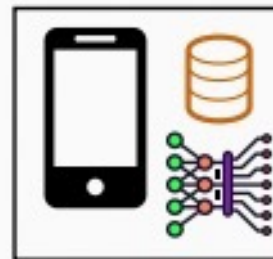
each party makes an update using its local dataset



Local model training

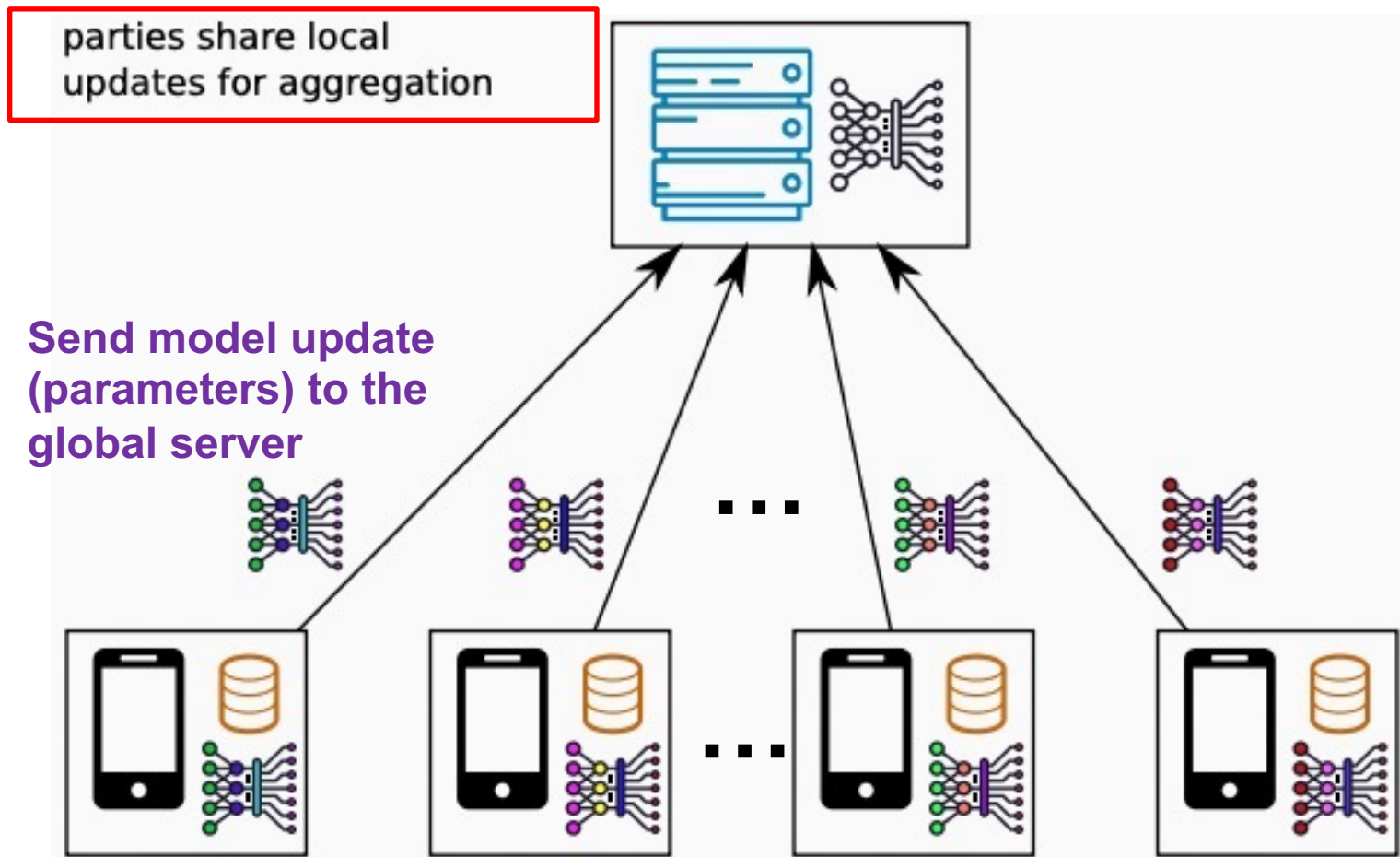


...



Federated Learning

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

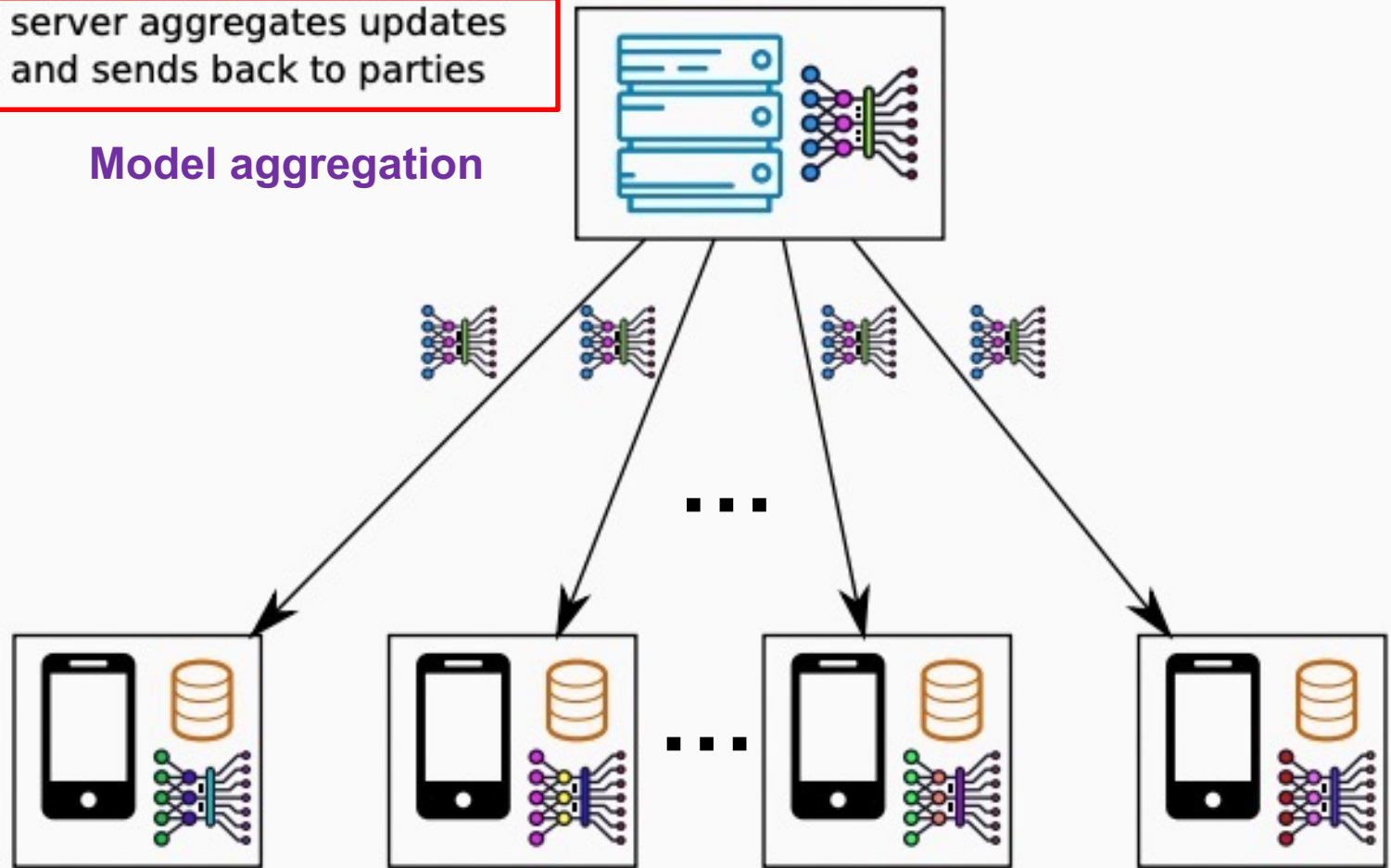


Federated Learning

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

server aggregates updates
and sends back to parties

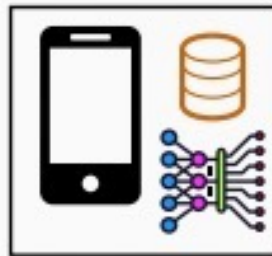
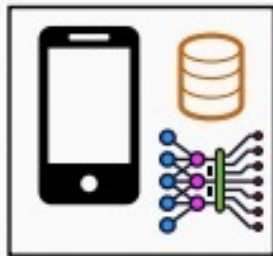
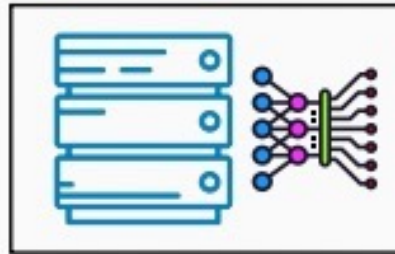
Model aggregation



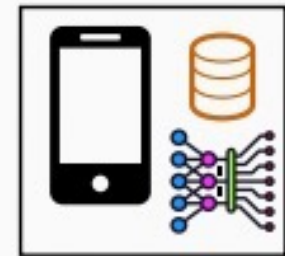
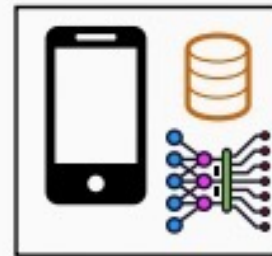
Federated Learning

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

parties update their copy of the model and iterate



...



Federated Learning

- We would like the final model to be as good as the centralized solution (ideally), or at least better than what each party can learn on its own

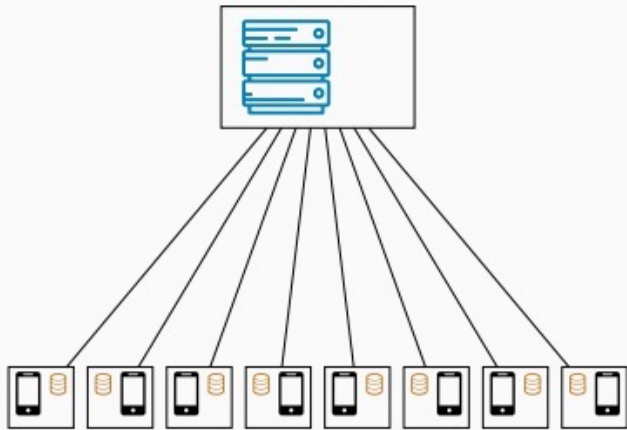
Key Differences with Distributed Learning

- Data distribution
 - In distributed learning, **data is centrally stored** (e.g., in a data center)
 - The main goal is just to **train faster**
 - We control how data is distributed across workers: usually, it is **distributed uniformly at random** across workers
 - In FL, data is naturally distributed and generated locally
 - Data is **not** independent and identically distributed (**non-i.i.d.**), and it is **imbalanced**

The discrepant data distributions will induce optimization inconsistency and feature divergence issues

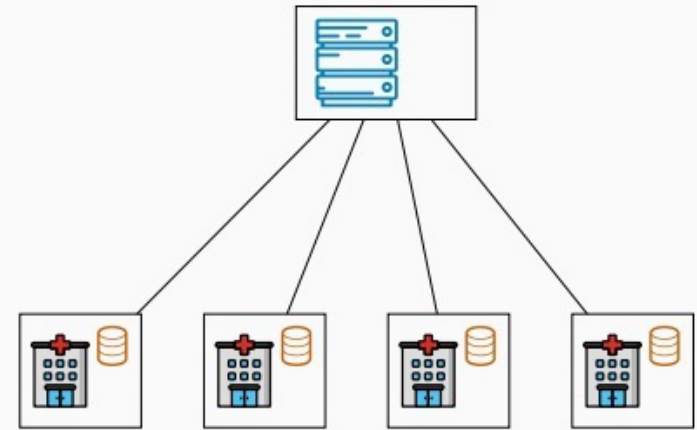
Cross-device vs. Cross-silo FL

Cross-device FL



- Massive number of parties (up to 10^{10})
- Small dataset per party (could be size 1)
- Limited availability and reliability
- Some parties may be malicious

Cross-silo FL



- 2-100 parties
- Medium to large dataset per party
- Reliable parties, almost always available
- Parties are typically honest

FL is a Booming Topic

- 2016: the term FL is first coined by Google researchers; 2020: more than 1,000 papers on FL in the first half of the year (compared to just 180 in 2018) [1]
- We have already seen some real-world deployments by companies and researchers
- Several open-source libraries are under development: PySyft, TensorFlow Federated, FATE, Flower, FedML
- FL is highly multidisciplinary: it involves machine learning, numerical optimization, privacy & security, networks, systems, hardware...

[1] <https://www.forbes.com/sites/robtoews/2020/10/12/the-next-generation-of-artificial-intelligence/>



A Baseline Algorithm: FedAvg

McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.

FedAvg (Basic Notations)

- We consider a set of K parties (clients)
- Each party k holds a dataset \mathcal{D}_k of n_k points
- Let $\mathcal{D} = \mathcal{D}_1 \cup \dots \cup \mathcal{D}_K$ be the joint dataset and $n = \sum_k n_k$ the total number of points
- We want to solve problems of the form $\min_{\theta \in \mathbb{R}^p} F(\theta; \mathcal{D})$ where:

$$F(\theta; \mathcal{D}) = \sum_{k=1}^K \frac{n_k}{n} F_k(\theta; \mathcal{D}_k) \quad \text{and} \quad F_k(\theta; \mathcal{D}_k) = \sum_{d \in \mathcal{D}_k} f(\theta; d)$$

- $\theta \in \mathbb{R}^p$ are model parameters (e.g., weights of a logistic regression or neural network)

FedAvg

Algorithm FedAvg (server-side)

Parameters: client sampling rate ρ

initialize θ

for each round $t = 0, 1, \dots$ **do**

$\mathcal{S}_t \leftarrow$ random set of $m = \lceil \rho K \rceil$ clients

for each client $k \in \mathcal{S}_t$ in parallel **do**

$\theta_k \leftarrow \text{ClientUpdate}(k, \theta)$

$\theta \leftarrow \sum_{k \in \mathcal{S}_t} \frac{n_k}{n} \theta_k$

Algorithm ClientUpdate(k, θ)

Parameters: batch size B , number of local steps L , learning rate η

for each local step $1, \dots, L$ **do**

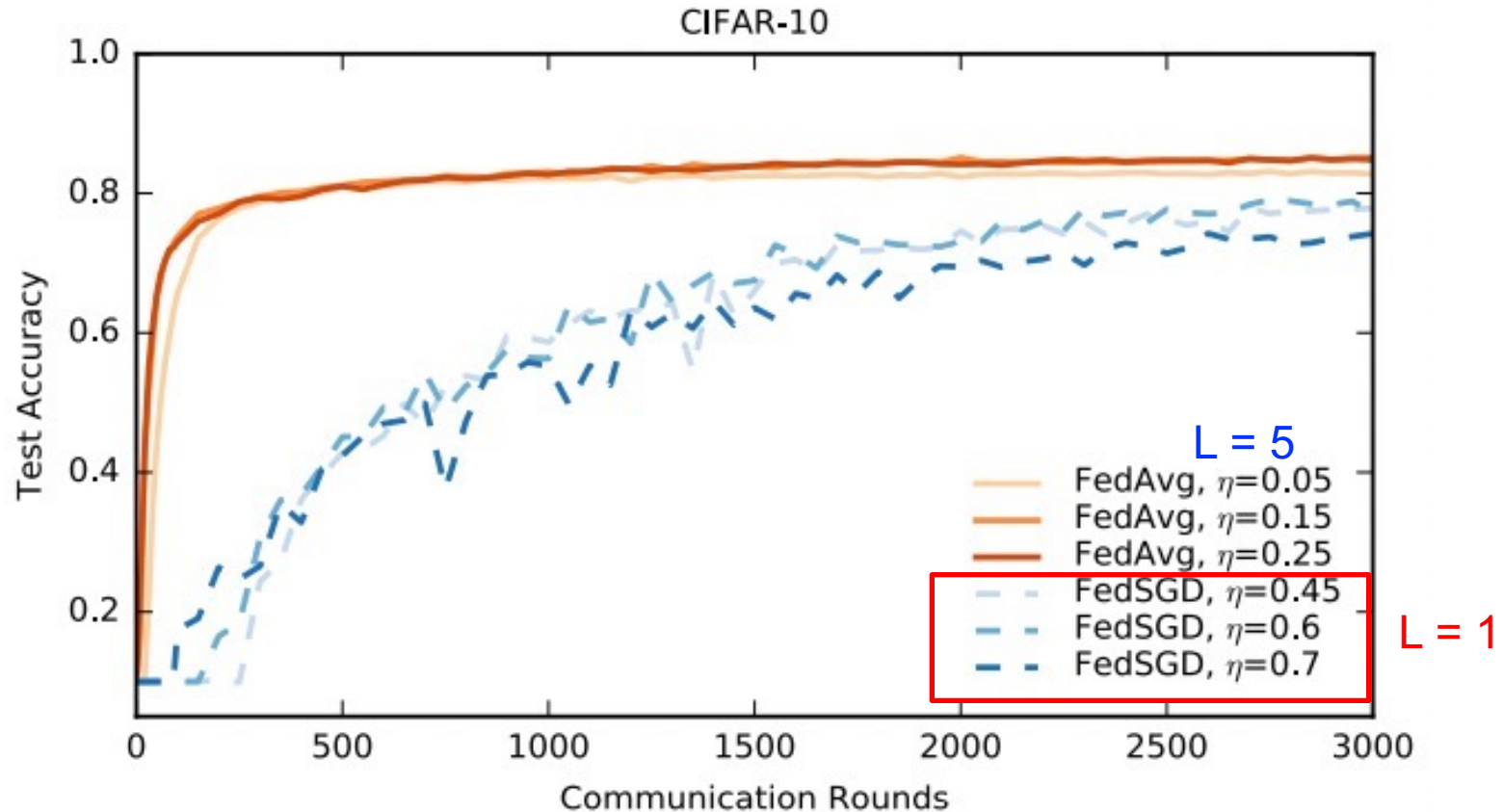
$\mathcal{B} \leftarrow$ mini-batch of B examples from \mathcal{D}_k

$\theta \leftarrow \theta - \frac{n_k}{B} \eta \sum_{d \in \mathcal{B}} \nabla f(\theta; d)$

 send θ to server

- For $L = 1$ and $\rho = 1$, it is equivalent to classic **parallel SGD**: updates are aggregated and the model synchronized at each step
- For $L > 1$: each client performs **multiple local SGD steps** before communicating

FedAvg

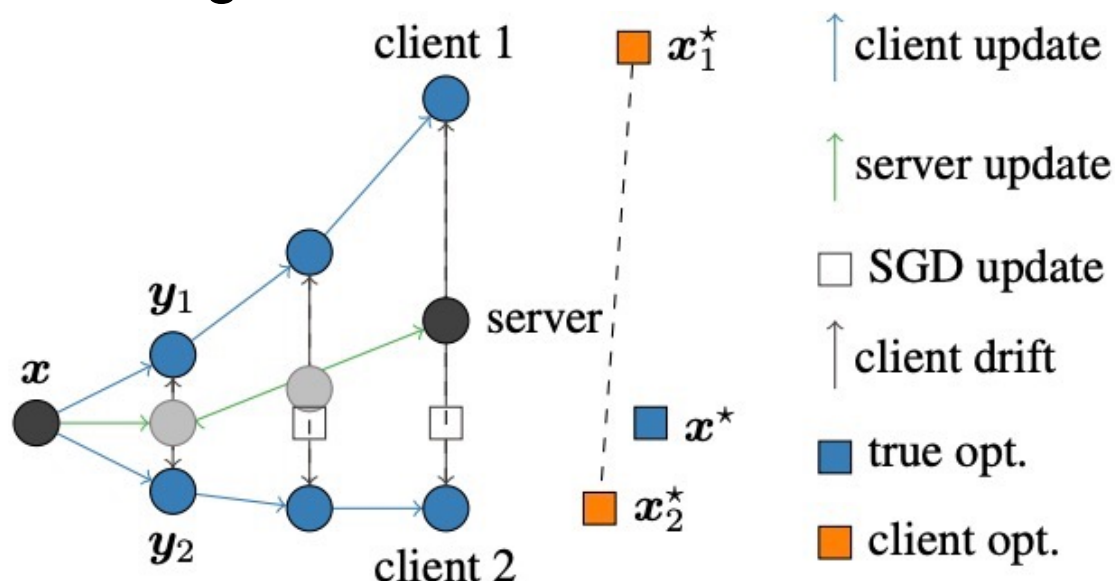


FedAvg with $L > 1$ allows to reduce the number of communication rounds, which is often the bottleneck in FL (especially in the cross-device setting)

McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.

Challenges in FL

- Dealing with non-i.i.d. data

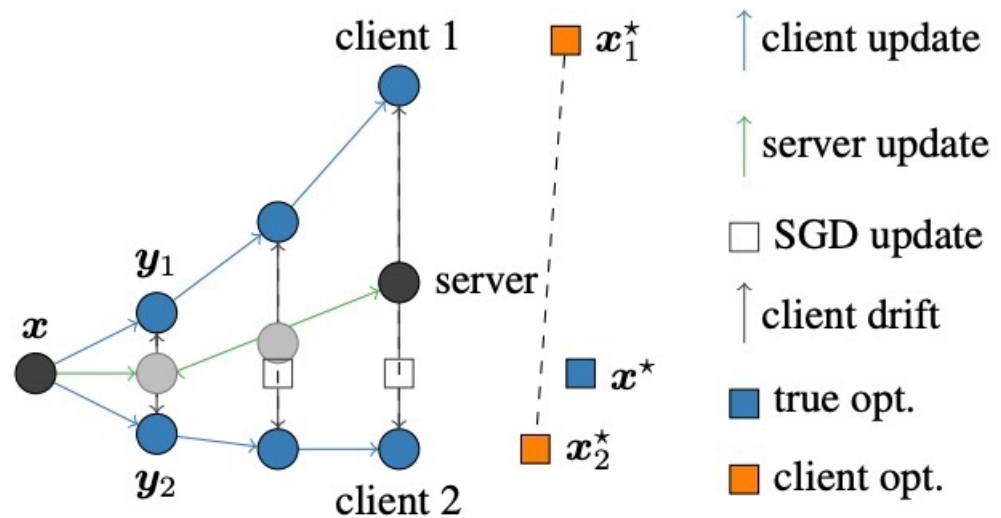


The client model often drifts away from the ideal global optimization point and overfits to its local objective.

Figure 1. Client-drift in FEDAVG is illustrated for 2 clients with 3 local steps ($N = 2$, $K = 3$). The local updates y_i (in blue) move towards the individual client optima x_i^* (orange square).

- When local datasets are non-i.i.d., FedAvg suffers from *client drift*
- To avoid this drift, one must use *fewer local updates and/or smaller learning rates*, which hurts convergence

FedProx



- **Solution**

- Proximal Term
- Limit the impact of local updates

$$\min_w h_k(w; w^t) = F_k(w) + \underbrace{\frac{\mu}{2} \|w - w^t\|^2}_{\text{proximal term}}$$

local function $F_k(\cdot)$

Global/server model from the previous round

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

FL of Personalized Models

- Learning from non-i.i.d. data is difficult/slow because each party wants the model to go in a particular direction
- If data distributions are very different, learning a single model which performs well for all parties may require a very large number of parameters
- Another direction to deal with non-i.i.d. data is thus to lift the requirement that the learned model should be the same for all parties (“one size fits all”)
- Instead, we can allow each party k to learn a (potentially simpler) personalized model θ_k but design the objective so as to enforce some kind of collaboration

Fallah, Alireza, Aryan Mokhtari, and Asuman Ozdaglar. "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach." *Advances in Neural Information Processing Systems* 33 (2020): 3557-3568.

Collins, Liam, et al. "Exploiting shared representations for personalized federated learning." *International Conference on Machine Learning*. PMLR, 2021.

Challenges in FL

- Preserving privacy (preventing data leakage)
 - Privacy concerns often motivate the need to keep raw data on each device local in federated settings.
 - However, sharing other information such as model updates as part of the training process can also potentially reveal sensitive information, either to a third party or to the central server.

Challenges in FL

- Preserving privacy (preventing data leakage)

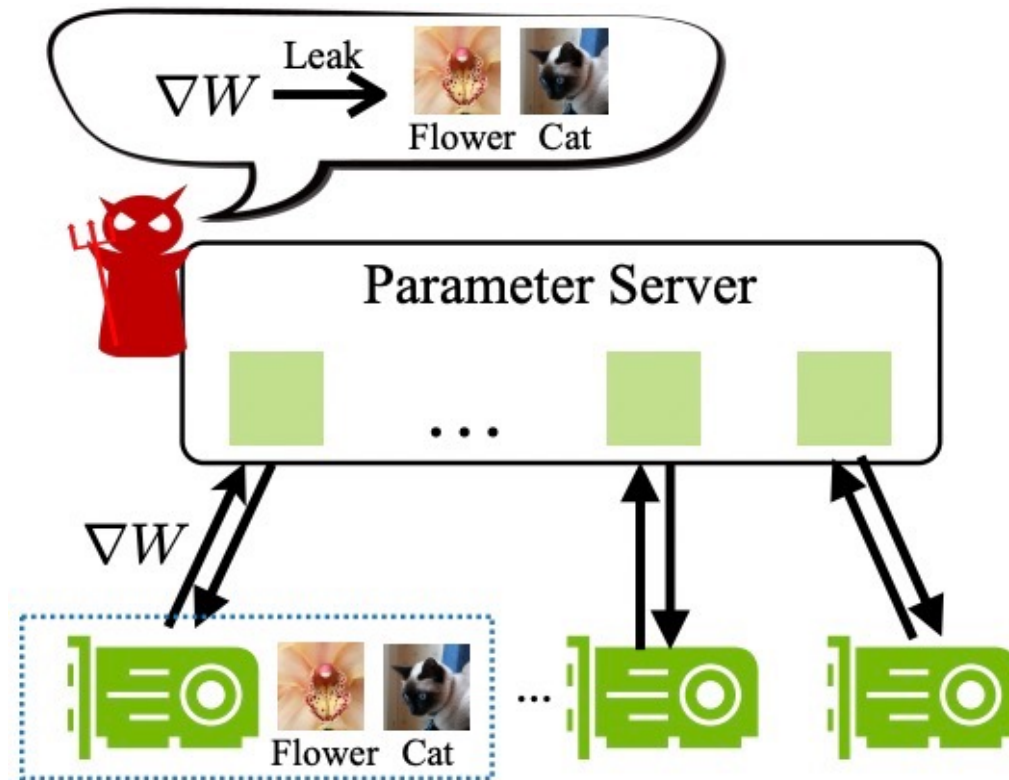


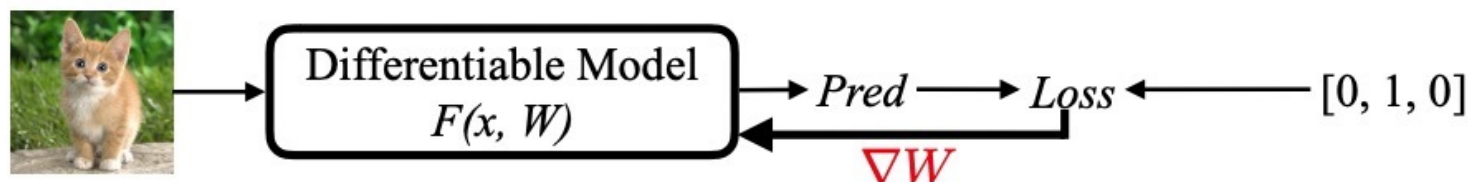
Figure from Zhu et al, 2019


Zhu, Ligeng, Zhijian Liu, and Song Han. "Deep leakage from gradients." *Advances in Neural Information Processing Systems* 32 (2019).

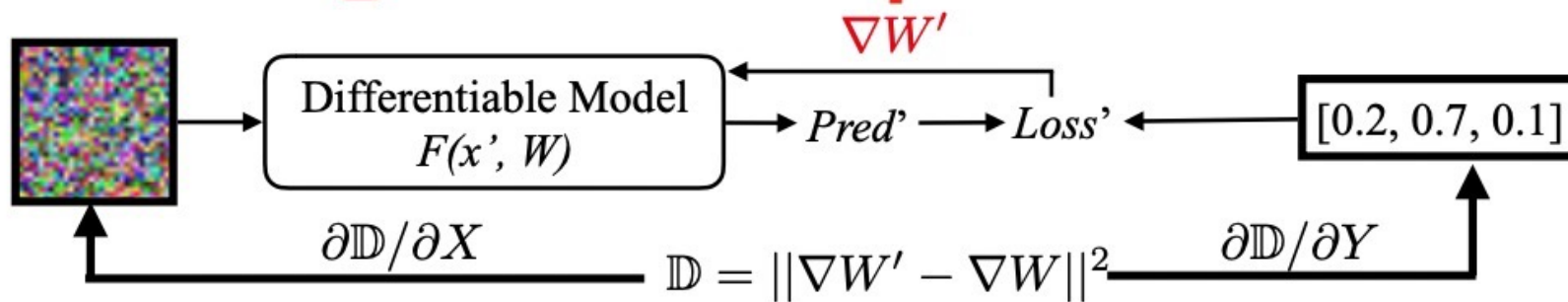
Challenges in FL

- Preserving privacy (preventing data leakage)

Normal Participant



Malicious Attacker 



Zhu, Ligeng, Zhijian Liu, and Song Han. "Deep leakage from gradients." *Advances in Neural Information Processing Systems* 32 (2019).

Challenges in FL

- Preserving privacy (preventing data leakage)

Algorithm 1 Deep Leakage from Gradients.

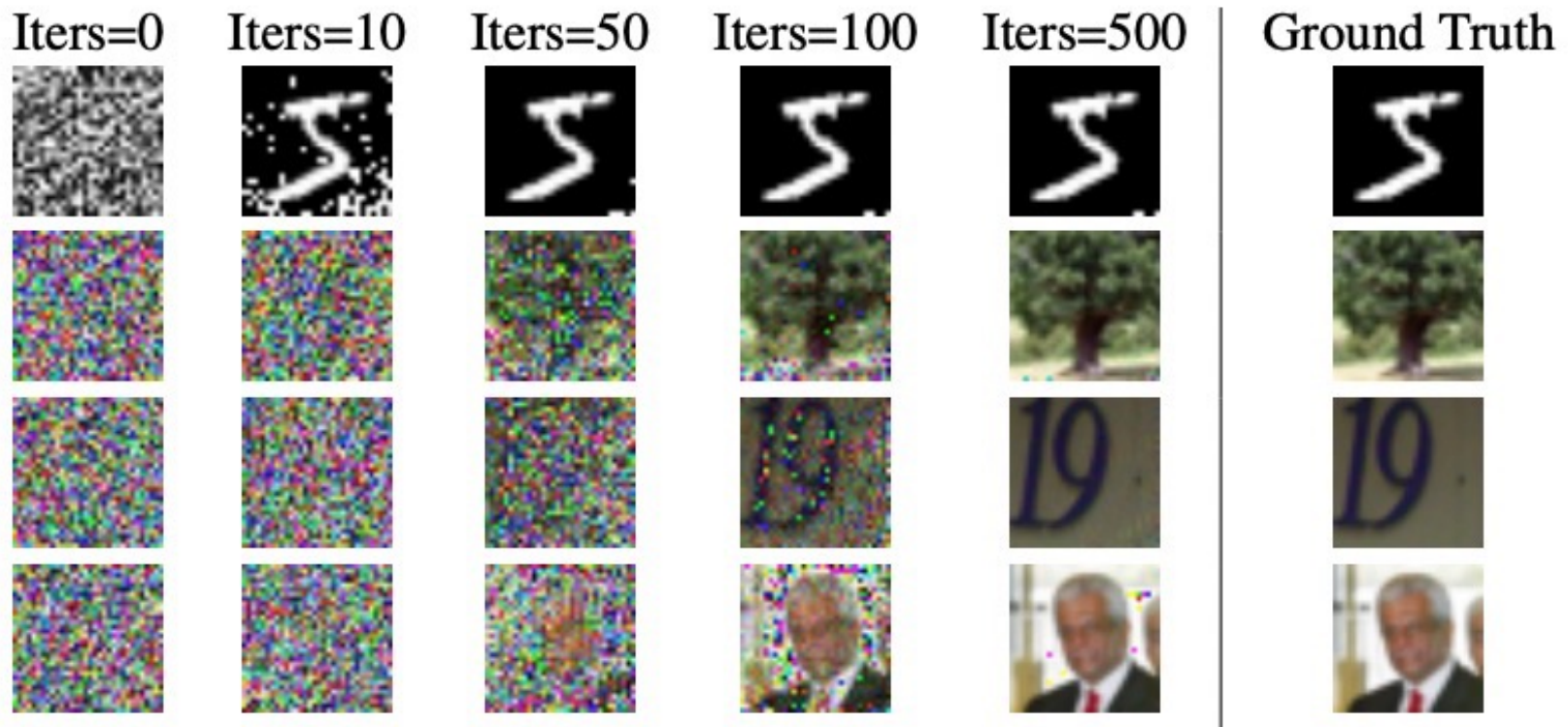
Input: $F(\mathbf{x}; W)$: Differentiable machine learning model; W : parameter weights; ∇W : gradients calculated by training data

Output: private training data \mathbf{x}, \mathbf{y}

```
1: procedure DLG( $F, W, \nabla W$ )  
2:    $\mathbf{x}'_1 \leftarrow \mathcal{N}(0, 1), \mathbf{y}'_1 \leftarrow \mathcal{N}(0, 1)$  ▷ Initialize dummy inputs and labels.  
3:   for  $i \leftarrow 1$  to  $n$  do  
4:      $\nabla W'_i \leftarrow \partial \ell(F(\mathbf{x}'_i, W_t), \mathbf{y}'_i) / \partial W_t$  ▷ Compute dummy gradients.  
5:      $\mathbb{D}_i \leftarrow \|\nabla W'_i - \nabla W\|^2$   
6:      $\mathbf{x}'_{i+1} \leftarrow \mathbf{x}'_i - \eta \nabla_{\mathbf{x}'_i} \mathbb{D}_i, \mathbf{y}'_{i+1} \leftarrow \mathbf{y}'_i - \eta \nabla_{\mathbf{y}'_i} \mathbb{D}_i$  ▷ Update data to match gradients.  
7:   end for  
8:   return  $\mathbf{x}'_{n+1}, \mathbf{y}'_{n+1}$   
9: end procedure
```

Challenges in FL

- Preserving privacy (preventing data leakage)



Zhu, Ligeng, Zhijian Liu, and Song Han. "Deep leakage from gradients." *Advances in Neural Information Processing Systems* 32 (2019).

Challenges in FL

- Preserving privacy (preventing data leakage)
 - Strategies for privacy-preserving
 - **Homomorphic Encryption** can be used to secure the learning process by computing on encrypted data.
 - **Secure multiparty computation.** [Cryptographic techniques](#) have been used to prevent [privacy disclosure](#) of client data in federated learning.
 - **Differential privacy.** Locally distorting client updates by adding noise with a distribution that offers ϵ -differential privacy is a very common approach to enhance privacy in FL

Blanco-Justicia, Alberto, et al. "Achieving security and privacy in federated learning systems: Survey, research challenges and future directions." Engineering Applications of Artificial Intelligence 106 (2021): 104468.

Federated Learning Applications

- There are multiple types of prominent federated learning applications:
 - **Smartphones.** Statistical models are used to power applications such as next-word prediction, face detection, and voice recognition by jointly learning user behavior across a large pool of mobile phones. However, users may not agree to share their data to protect their personal privacy or minimize the bandwidth or battery usage of their phones. Federated learning can be used to enable predictive features on smartphones without leaking private information or diminishing the user experience.

Source: <https://viso.ai/deep-learning/federated-learning/>

Federated Learning Applications

- There are multiple types of prominent federated learning applications:
 - **Organizations.** In the context of federated learning, entire organizations or institutions can also be viewed as “devices”. For example, hospitals are organizations that contain a large amount of patient data for predictive healthcare applications. However, hospitals operate under strict privacy practices and may face legal, administrative, or ethical constraints that require data to remain local. Federated learning is a solution for such applications because it can reduce strain on the network and enable private learning between various devices/organizations.

Source: <https://viso.ai/deep-learning/federated-learning/>

Federated Learning Applications

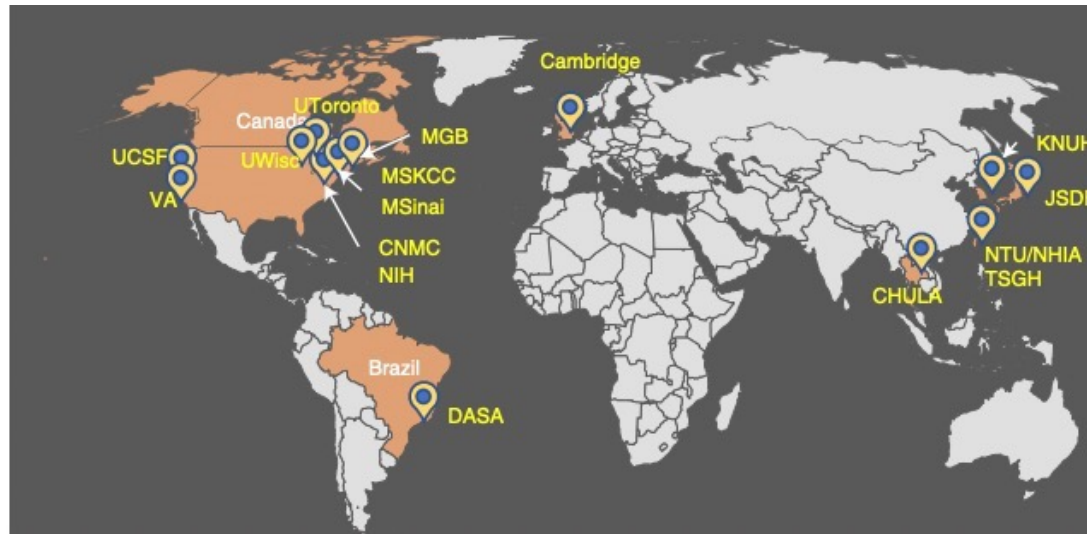
- There are multiple types of prominent federated learning applications:
 - **Internet of things.** Modern IoT networks, such as wearable devices, autonomous vehicles, or smart homes, use sensors to collect and react to incoming data in real-time. For example, a fleet of autonomous vehicles may require an up-to-date model of traffic, construction, or pedestrian behavior to operate safely. However, building aggregate models in these scenarios may be difficult due to privacy concerns and the limited connectivity of each device. Federated learning methods enable the training of models that efficiently adapt to changes in these systems while maintaining user privacy.

Source: <https://viso.ai/deep-learning/federated-learning/>

FL for Medical Imaging (Case Study)

Predicting oxygen requirements of patients with COVID-19

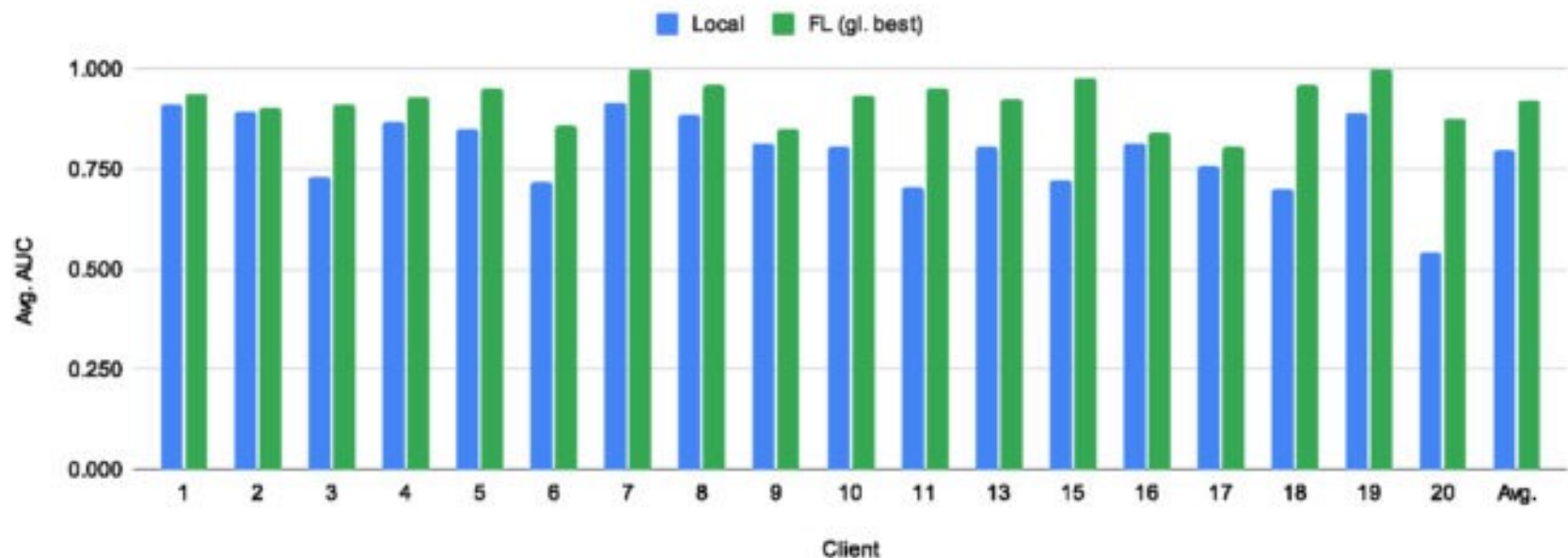
- Use data from 20 institutes across the globe to train a FL model, called EXAM (electronic medical record (EMR) chest X-ray AI model), that predicts the future oxygen requirements of symptomatic patients with COVID-19 using inputs of vital signs, laboratory data and chest X-rays.
- The study is led by Mass General Brigham and NVIDIA



20 different client sites contributing to the EXAM study

Predicting oxygen requirements of patients with COVID-19

- The global EXAM model, shared with all participating sites, resulted in a 16 percent improvement of the AI model's average performance. Researchers saw an average increase of 38 percent in generalizability when compared to models trained at any single site.



Source: <https://blogs.nvidia.com/blog/2021/09/15/federated-learning-nature-medicine/>



Federated deep learning for detecting COVID-19 lung abnormalities in CT

- Develop a federated learning method for detecting COVID-19 related CT abnormalities with external validation on patients from a multinational study.
- They recruited 132 patients from seven multinational different centers, with three internal hospitals from Hong Kong for training and testing, and four external, independent datasets from Mainland China and Germany, for validating model generalizability

Dou, Qi, et al. "Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study." NPJ digital medicine 4.1 (2021): 1-11.

Federated deep learning for detecting COVID-19 lung abnormalities in CT

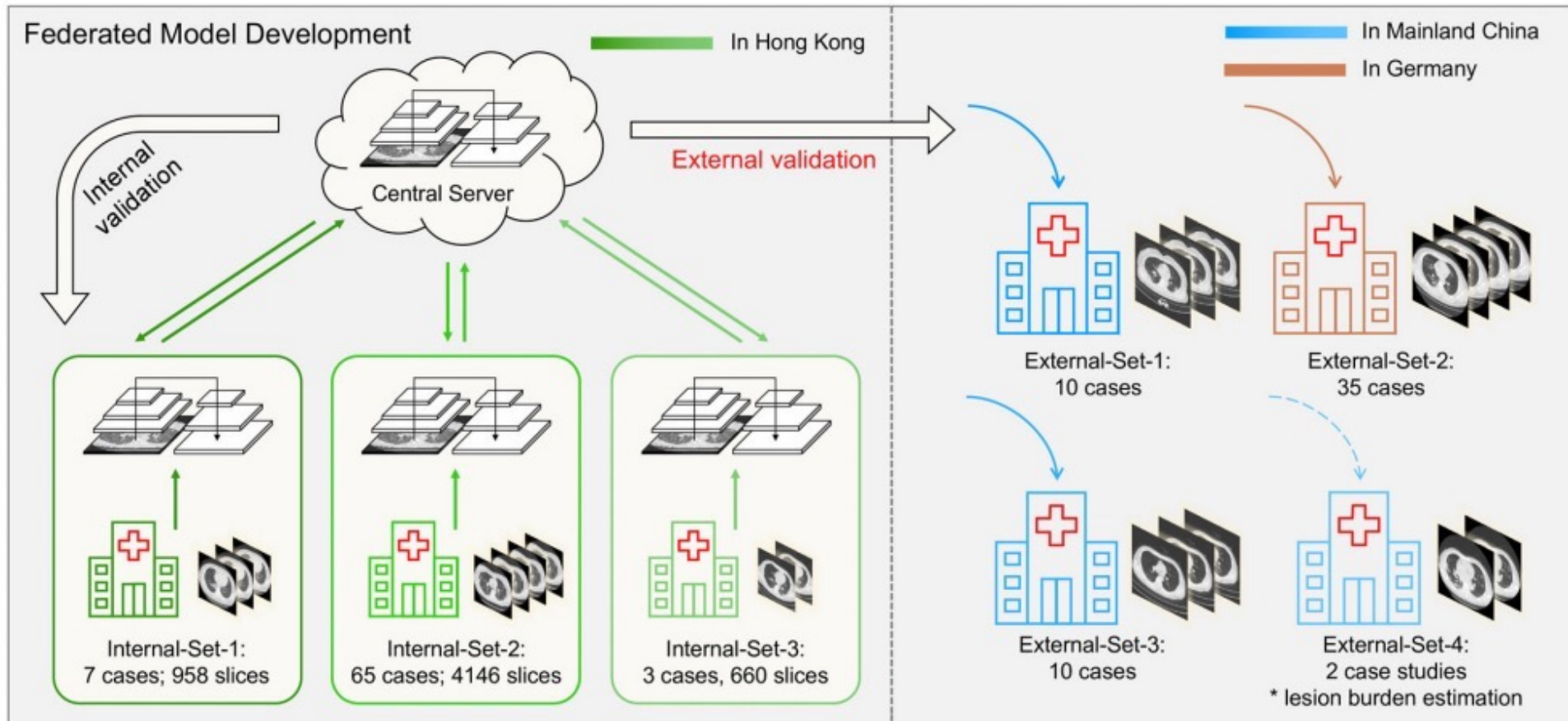


Fig. 1 Overview of our AI scheme to develop a privacy-preserving CNN-based model for detecting CT abnormalities in COVID-19 patients with a multinational validation study. A privacy-preserving AI system was developed with CT data from three hospitals in Hong Kong using federated learning, and then the generalizability was validated on external cohorts from Mainland China and Germany.

Dou, Qi, et al. "Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study." NPJ digital medicine 4.1 (2021): 1-11.

Healthcare Industry Adopting FL



MEDICAL IMAGING
Adopting NVIDIA Clara Federated Learning for Imaging



Healthcare Industry Adopting FL



This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement N° 831472. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA

PHARMA PARTNERS

- AMGEN
- astellas
- AstraZeneca
- Bayer
- Boehringer Ingelheim
- gsk
- janssen
- MERCK
- NOVARTIS
- SERVIER

PUBLIC PARTNERS

- Műgyetem 1782
- IKTOS
- KU LEUVEN
- loodse
- NVIDIA
- OWKIN
- SUBSTR FOUNDATION
- imi | innovative medicines initiative
- European Union flag
- efpia

PHARMA Machine Learning Ledger Orchestration for Drug Discovery

Impact of FL

- Increasing the value of AI for all healthcare stakeholders



Clinicians

Accurate assistance tools,
unbiased decisions



Patients

Accurate and unbiased AI,
cost reductions



Researchers

Safe collaboration, access
to large datasets, impact



Healthcare Providers

Access accurate AI while
monetizing data via FL



Manufacturers

continuous improvement of
ML-based systems



FL Workshops, Conferences, ...

- International Workshop on Federated Learning for User Privacy and Data Confidentiality, in conjunction with NeurIPS 2019
- International Workshop on Federated Learning for User Privacy and Data Confidentiality, in Conjunction with ICML 2020
- SpicyFL 2020 : NeurIPS Workshop on Scalability, Privacy, and Security in Federated Learning
- International Workshop on Federated Learning for User Privacy and Data Confidentiality, in Conjunction with ICML 2021
- FedVision (CVPR-2022 Workshop on Federated Learning for Computer Vision): <https://sites.google.com/view/fedvision>
- IEEE TMIM Special Issue on Federated Learning for Medical Imaging: <https://www.embs.org/wp-content/uploads/2021/02/IEEE-TMI-2022-FL-CFP.pdf>

References and resources

1. Kairouz, Peter, et al. "Advances and open problems in federated learning." *arXiv preprint arXiv:1912.04977* (2019).
2. Wang, Jianyu, et al. "A field guide to federated optimization." *arXiv preprint arXiv:2107.06917* (2021).
3. Lim, Wei Yang Bryan, et al. "Federated learning in mobile edge networks: A comprehensive survey." *IEEE Communications Surveys & Tutorials* 22.3 (2020): 2031-2063.
4. Li, Qinbin, et al. "A survey on federated learning systems: vision, hype and reality for data privacy and protection." *IEEE Transactions on Knowledge and Data Engineering* (2021).
5. Federated learning framework - FedML: <https://fedml.ai/>
6. <https://github.com/FedML-AI/FedML>



Thank you!

Question?