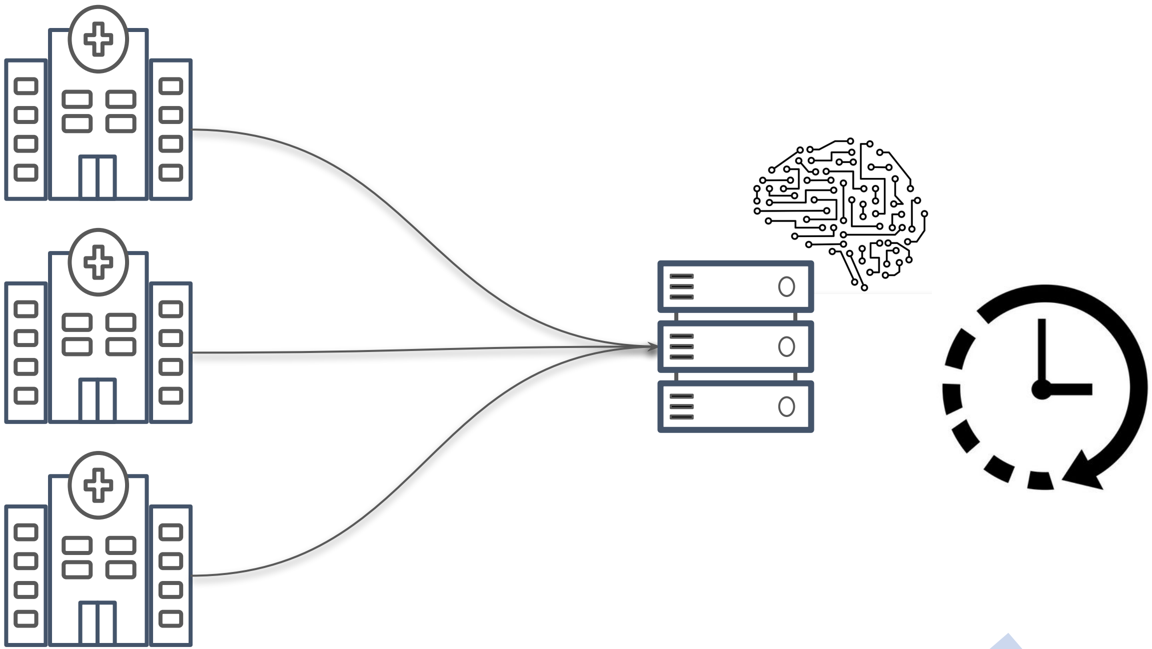
Federated Learning For Multi-Center Breast Cancer Classification in the Real World

Akis Linardos

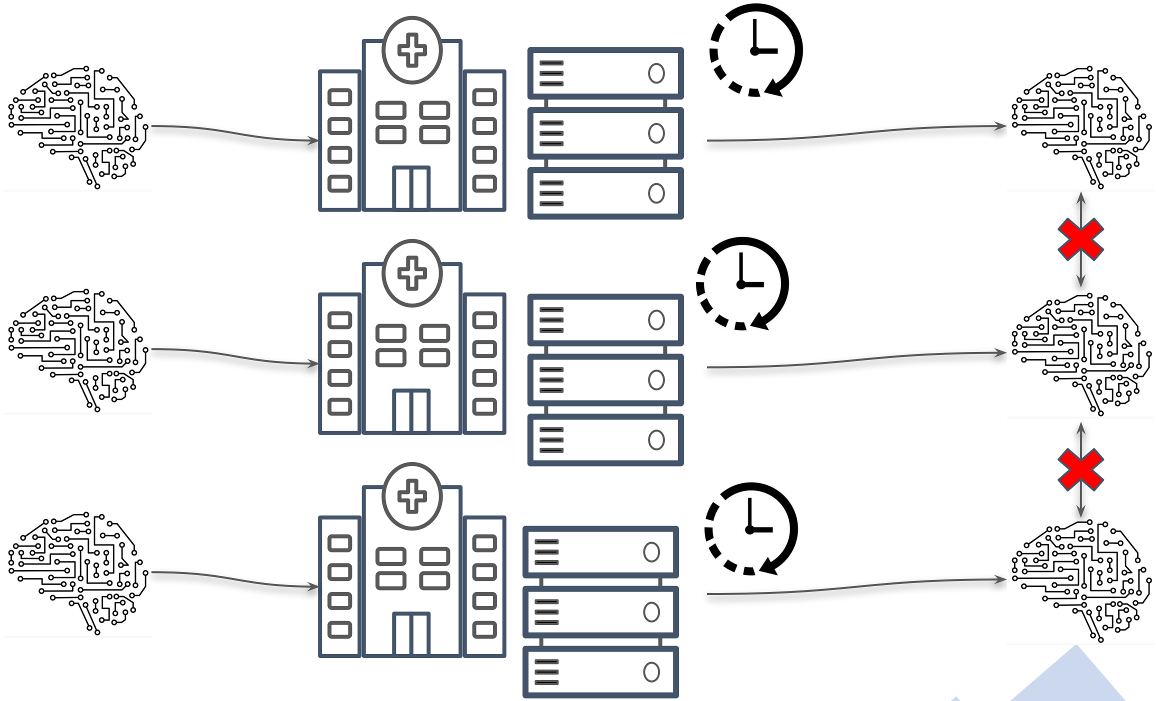
PhD student at BCN-AIM

Collaborative Data Sharing (CDS)



Data is pooled in one server where training occurs.

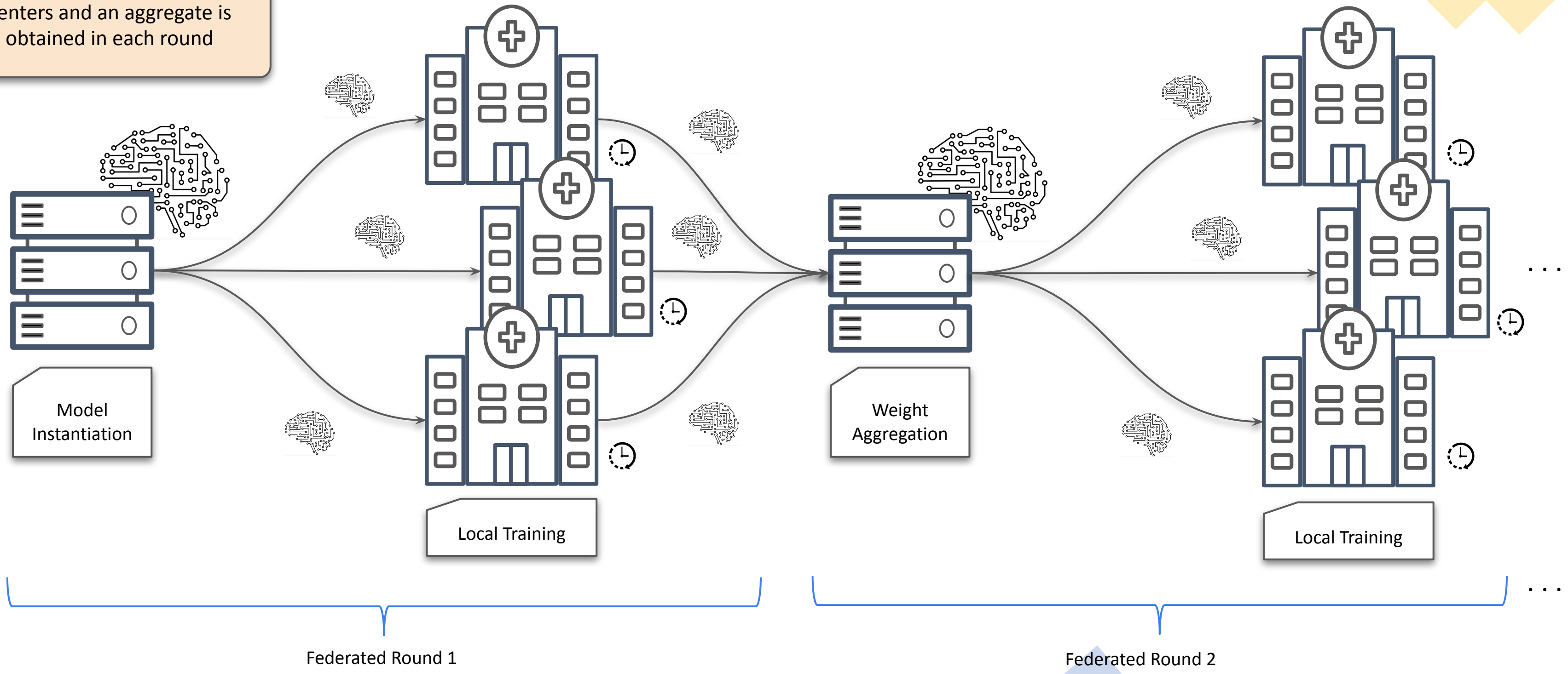
What really happens



Models are trained on single center data. Their generalization is questionable.

Federated Learning (FL)

A model is distributed across centers and an aggregate is obtained in each round



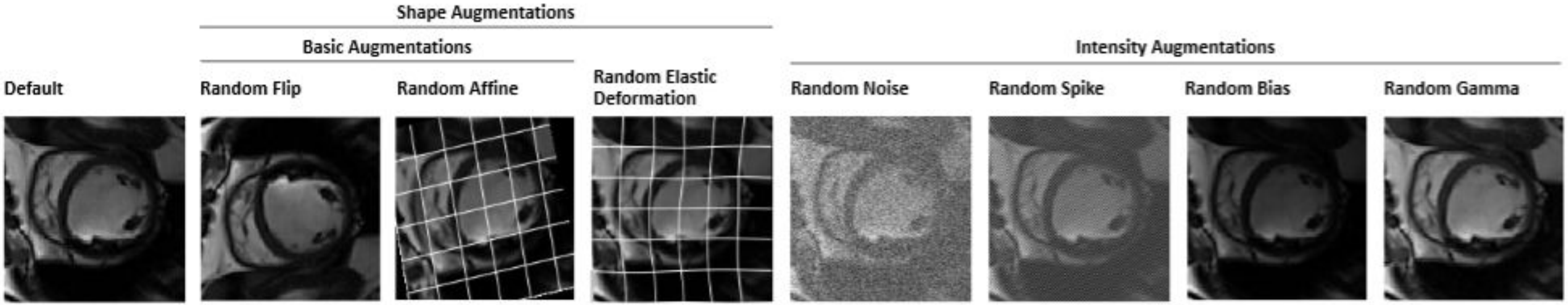
A **Simulation** Study in Cardiovascular Disease

https://arxiv.org/abs/2107.03901

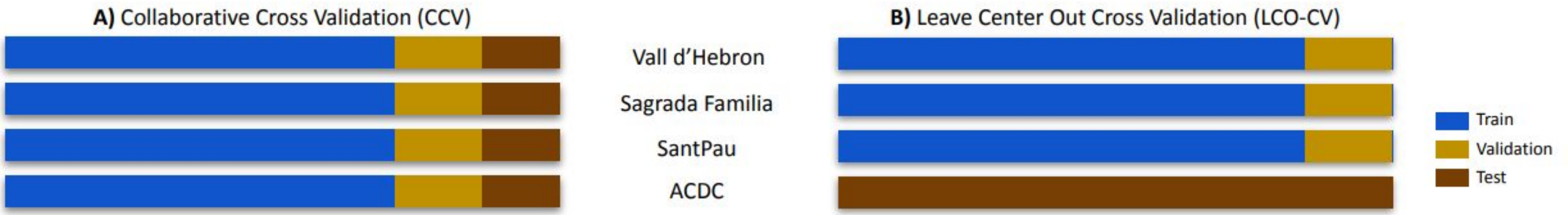
Linardos, Akis, et al. "Federated Learning for Multi-Center Imaging Diagnostics: A Study in Cardiovascular Disease." *arXiv preprint arXiv:2107.03901* (2021).

Cardiac MRI Multi-center Data

Center	Vendor	Spatial Resolution (mm)	Slice Thickness (mm^2)	NOR	HCM	Total
Vall d'Hebron	Philips	1.1516-1.2362	10.0	21	25	46
Sagrada Familia	Siemens	0.9765-1.6200	8.0-10.0	33	37	70
SantPau	Canon	0.7955-1.8228	10.0	14	10	24
ACDC	Siemens	1.3400-1.6800	5.0-10.0	20	20	40
Total				88	92	180

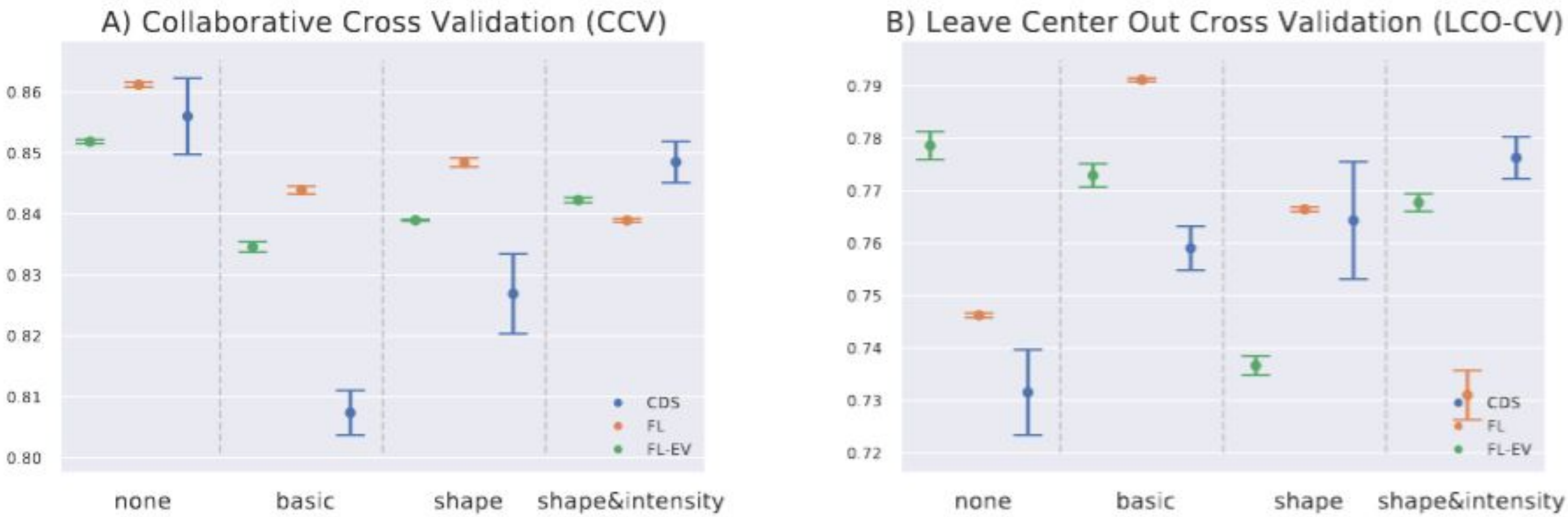


Principled Evaluation



LCO-CV gives us an estimate of *out-of-site* generalization performance, by testing iteratively on an unseen-center fold

Results: 5 seeds per configuration



FL-EV is an alternate aggregation technique we tried: contrast to the original **FL** algorithm, in this case each center gets an *Equal Vote*

FL outperforms **CDS** in many cases.

FL and **FL-EV** are more robust across different seeds, while **CDS** exhibits significant error bars.

Now onto the real challenge: Actual Deployment.

Federated Learning
For Multi-Center Breast Cancer Classification in the Real World

Our process: One Step at A Time

- Phase 1: Set up a federated network across collaborators. Tackle classification.
    - Test technical innovations in the real world (already found to work well in FeTS Challenge which was a simulation)
    - Labels: (**Normal / Benign Tumor / Malignant Tumor**)
- Phase 2: Improve classification, tackle other use cases. Required annotations on previously gathered data:
    - Bounding boxes around lesions

What technical innovations?

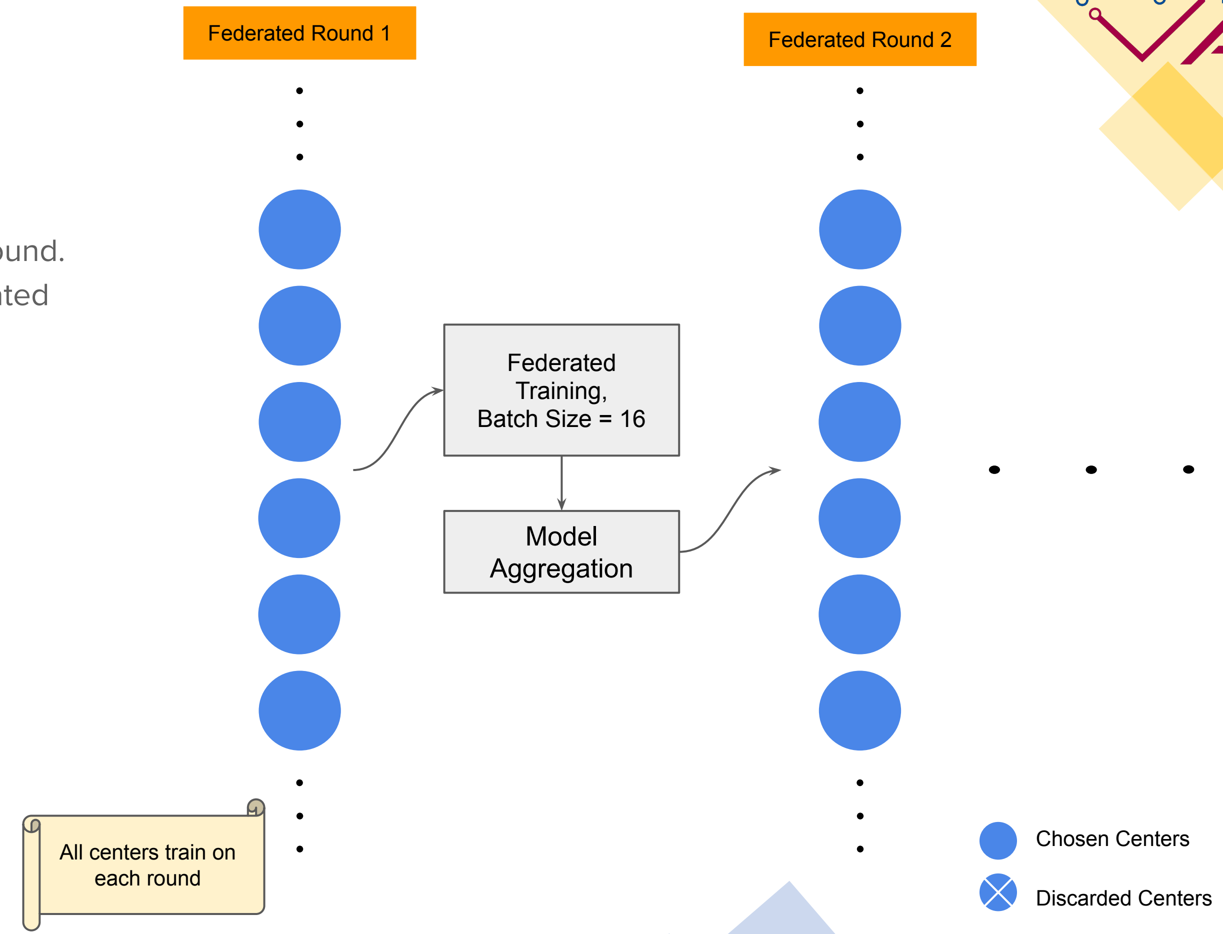
We proposed *Center Dropout* in the first Federated Challenge at MICCAI and scored ***1st*** on one of the leaderboards.

Since this was a simulation, we will now have the chance to test it in a real world setting.

Rank	Team Name	Institution	Lead Author
1	BCN-AIM	University of Barcelona	Akis Linardos
2	HT-TUAS	Turku University of Applied Sciences	Irfan Khan
3	Shoulders	Chinese University of Hong Kong	Quande Liu

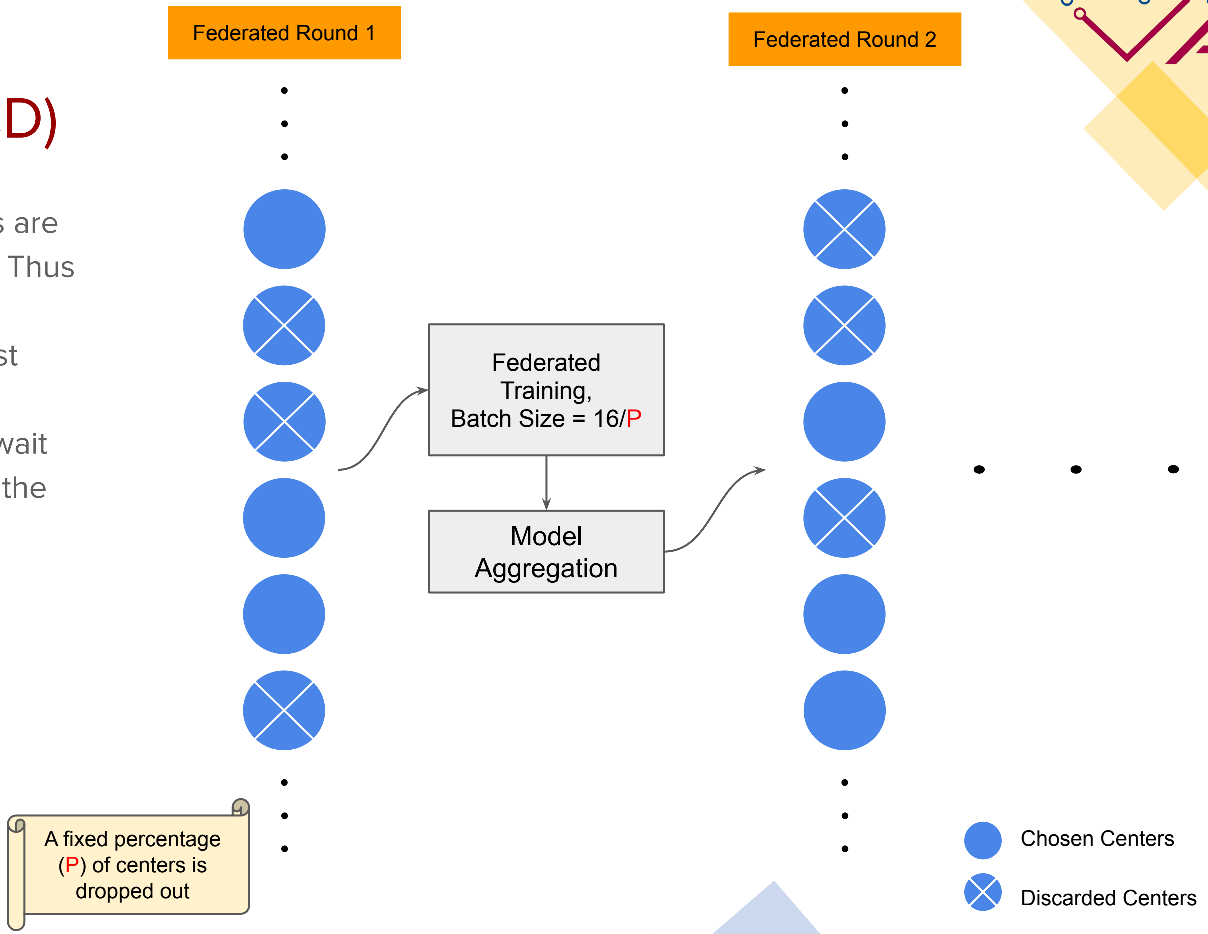
Vanilla FL

- All centers train in each round.
- Local models are aggregated



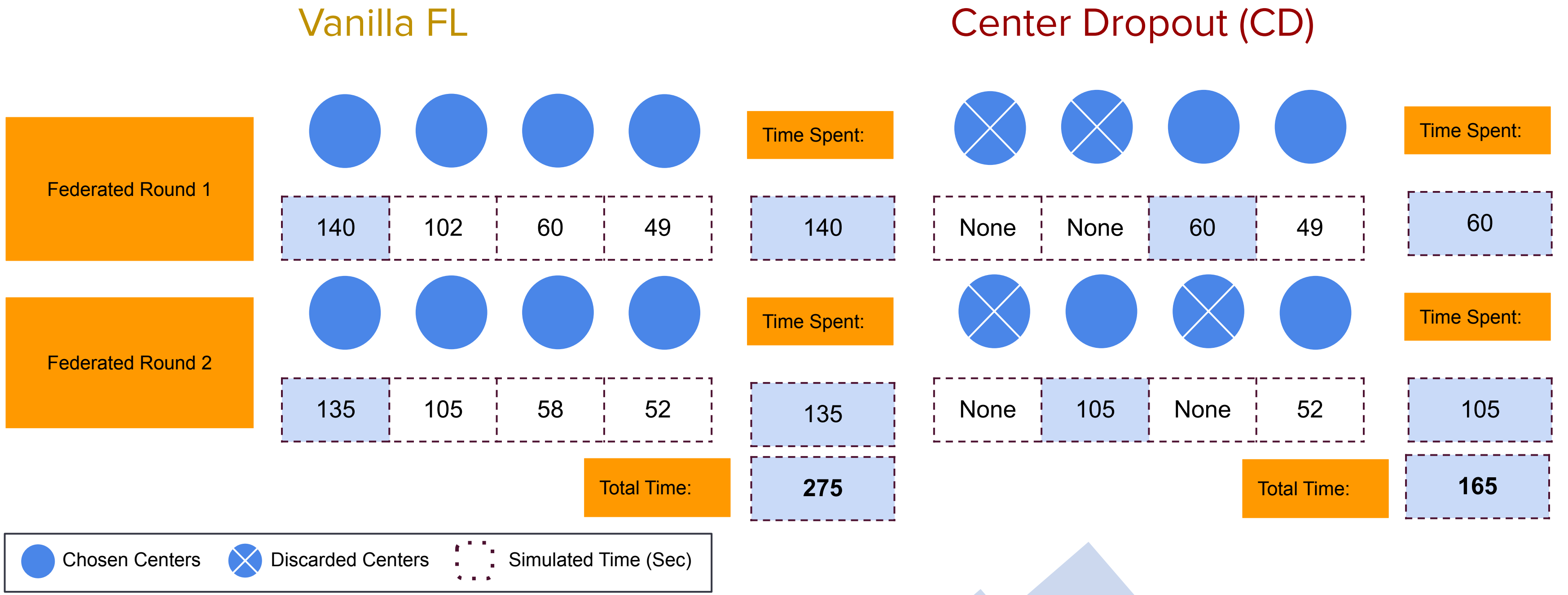
Center Dropout (CD)

- Local models of subgroups are aggregated in each round. Thus the vote is not always overwhelmed by the largest center.
- Training does not have to wait for the slowest member of the consortium.



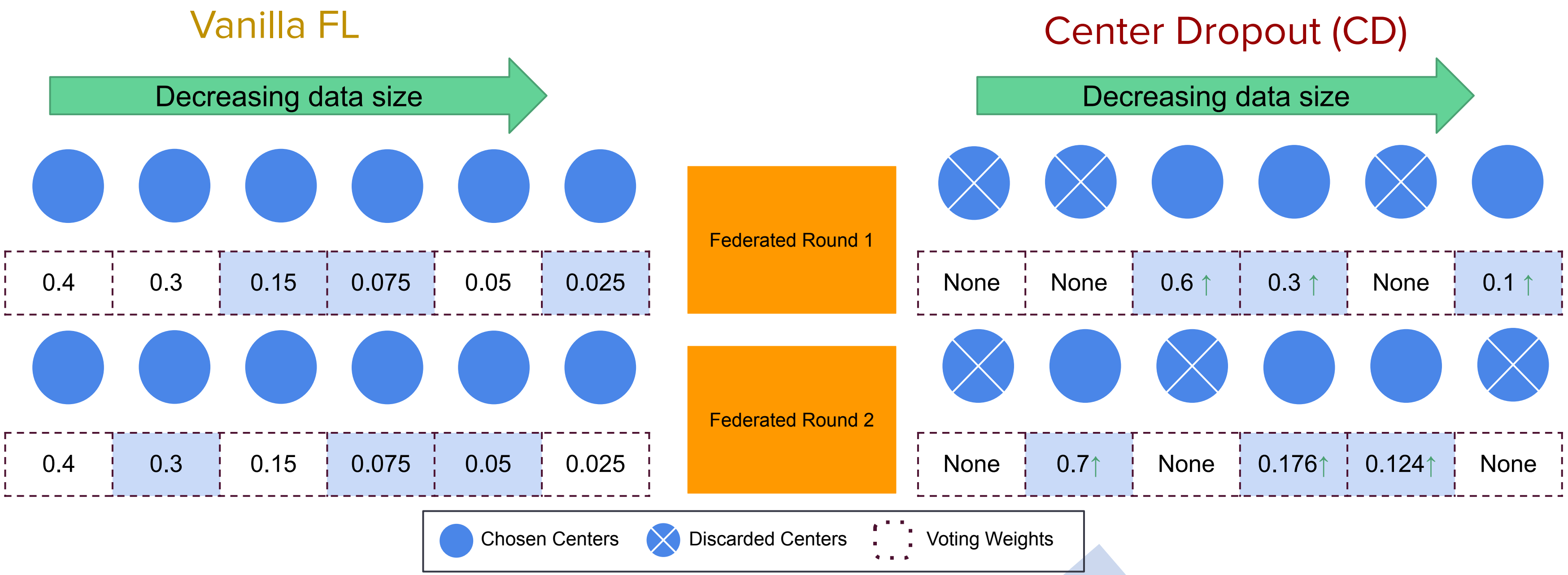
Intuitions: Speed

- In Vanilla FL, each round moves as slow as its slowest member. In CD the slowest member is different each round. We thus cut down on communication costs. By using a proportionally batch size, the same amount of training goes on with less communication.



Intuitions: Fairness

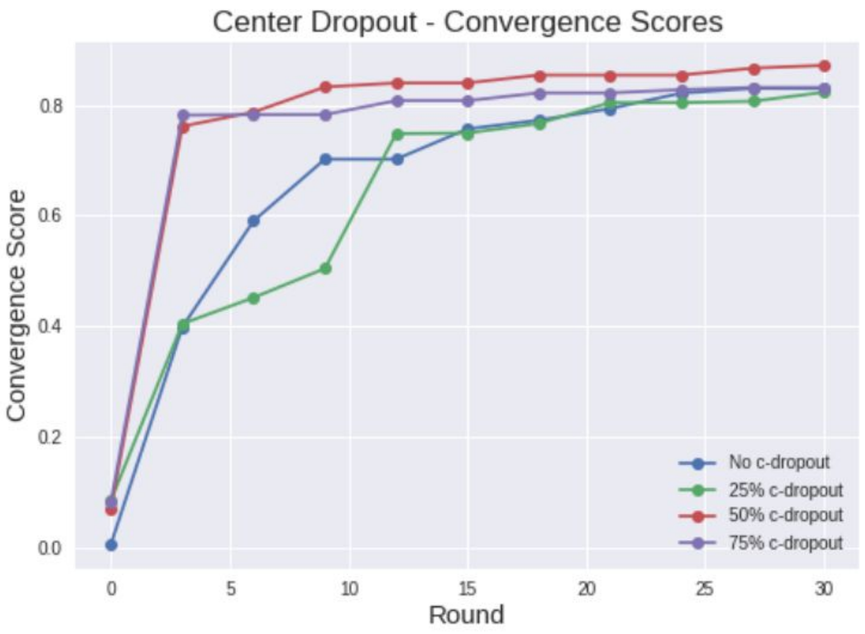
- In Vanilla FL, voting is the same in each round. In CD, smaller centers get the chance to have a higher vote, as they are not consistently overwhelmed.



Results: Better Performance, Faster Convergence

C. Dropout	Batch Size	Dice ET
None	16	0.6023
25%	21	0.58755
50%	32	**0.62857**
75%	64	0.62146

C. Dropout	Batch Size	Hausdorff95 ET
None	16	30.40639
25%	21	35.62406
50%	32	**25.439**
75%	64	33.87432



Plan Overview

Define Clinical Problem

Set up the Infrastructure

Test Connection and hardware with toy data

ETL Pipeline: Extract, Transform, Load

Data Harmonization

Deploy model, test with hospital data

Start Experiments

Analyze results

Bring the story together

Plan Overview

Define Clinical Problem

Set up the Infrastructure

Test Connection and hardware with toy data

ETL Pipeline: Extract, Transform, Load

Data Harmonization

Deploy model, test with hospital data

Start Experiments

Analyze results

Bring the story together

Plan Overview

Define Clinical Problem

Set up the Infrastructure

Test Connection and hardware with toy data

ETL Pipeline: Extract, Transform, Load

Data Harmonization

Deploy model, test with hospital data

Start Experiments

Analyze results

Bring the story together

Can we leverage Topological Data Analysis in Federated Learning?

- No work on this currently
- Federated Learning allows training without ever seeing the data. In this set-up, the only privacy-respecting data analysis of individual points would require noise-inserting techniques. The denoising properties of Topological Data Analysis could help us study this noisy data.
- Perhaps TDA can also help us tackle domain shift between multiple centers?

Thank you for your attention!

More about me and my past work at:

linardos.github.io

Prior experience:

- We have completed a CMR study (available on arxiv) and are in the process of publishing at Scientific Reports
- Key findings from that study:
    - The Federated Algorithm is more robust than Collaborative Data sharing, even when the exact same data is used.
    - Different Cross-Validation splits provide different results. Leave Center Out is proposed as best alternative.
    -

Results

- 4 configurations of Center Dropout (CD) were tested with different percentages (P)
- 50% CD outperforms all alternatives by a significant margin, including baseline.
- Possible room to wiggle in the range 40-60% for further fine tuning

C. Dropout	Batch Size	Dice ET	Dice WT	Dice TC
None	16	0.6023	0.75637	0.60354
25%	21	0.58755	0.70387	0.60216
50%	32	**0.62857**	**0.78498**	0.6417
75%	64	0.62146	0.74899	**0.64332**
C. Dropout	**Batch Size**	**Hausdorff95 ET**	**Hausdorff95 WT**	**Hausdorff95 TC**
None	16	30.40639	34.34818	30.59265
25%	21	35.62406	44.66591	38.50768
50%	32	**25.439**	**24.74043**	**25.31731**
75%	64	33.87432	38.05616	39.45167