Robust Collaborative Inference with Vertically Split Data Over Dynamic Device Environments

Surojit Ganguli, Zeyu Zhou, Christopher G. Brinton, David Inouye

PURDUE UNIVERSITY

Elmore Family School of Electrical and Computer Engineering

Outline

- Introduction
- Setup
- Method
- Results
- Conclusion

Outline

- Introduction
- Setup
- Method
- Results
- Conclusion

Introduction – Use Case

- ❑ Collaborative Learning (CL) over vertically split data is a scenario where features for the same entity are split across multiple devices

- ❑ Some real-world scenarios, where data is distributed and ability to handle dynamic changes in network is critical:
  - ➢ **Robustness:** Sensors observing different data modalities across an ocean or sea, may become unreliable given the harsh outdoor conditions. Thus, sensors may leave or join the network arbitrarily. System performance must be reliable even in such a scenario.
  - ➢ **Remote Monitoring:** Having a central server for surveilling a network of sensors in difficult to access harsh environments, such as the ocean, is challenging. Thus, there is a need for sensor network that operates without a central server.

Introduction – Research Question

Can we develop a cross-device decentralized collaborative learning (CL) method that maintains strong performance at test time even under near-catastrophic faults?

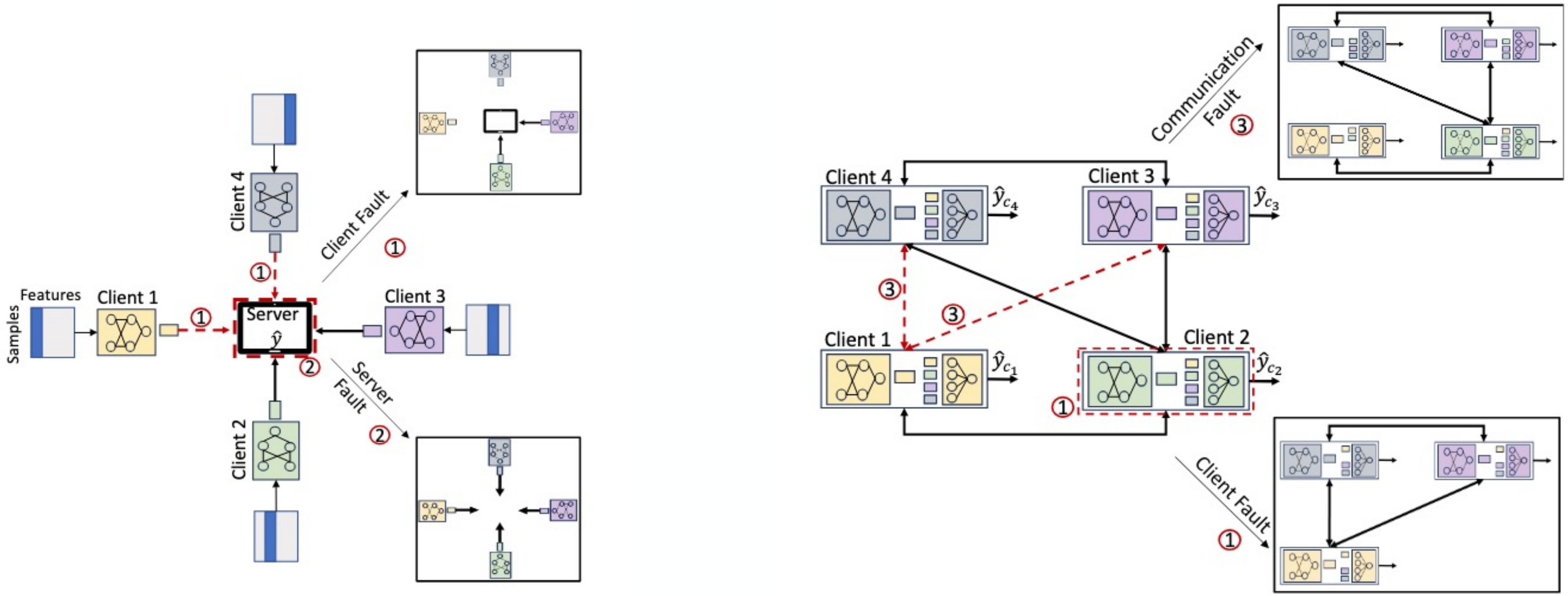
Introduction – Contributions

- Formalize the problem of decentralized collaborative learning under dynamic network conditions, called Dynamic Network Collaborative Learning (DN-CL)
- We develop MAGS, combining fault simulation, aggregator replication, and ensemble gossiping to provide fault tolerance during the inference stage of DN-CL
- Conduct extensive experiments and study the fundamental tradeoff between communication cost and robustness.

Outline

- Introduction
- Setup
- Method
- Results
- Conclusion

Setup – Dynamic Network Collaborative Learning (DN-CL)



(a) Collaborative Learning for Vertically Split Data (e.g., VFL)

(b) Dynamic Network Collaborative Learning (DN-CL)

Figure 1: Collaborative Learning (CL) with vertically split data (Figure 1a), like VFL, assumes samples are split across clients with a central server and the server is always available to communicate with at least one device. The data context in our study is the same as VFL where the features are split across clients. However, in our case, no centralized server node is assumed, and clients serve as data aggregators (Figure 1b). Our goal is to obtain robust test time performance even under highly dynamic networks such as client/device faults (1), server faults (2) and communication faults (3).

Setup – Client selection function (h)

- A criterion is required to select an aggregator
- We formulate a client selection function, h
- The function represents the communication to an external entity
- In the paper we present 4 functions but here focus on h_{active}

Active Client ($h_{active}(\hat{\boldsymbol{y}})$): Randomly selects active client

Setup – Dynamic Risk

Dynamic Risk

Client Selection Function

Distributed Inference Algorithm

$$R_h(\theta; \mathcal{G}(t)) \triangleq \mathbb{E}_{X,Y,\mathcal{G}(t),h}[\ell_h(\Psi(X; \theta, \mathcal{G}(t)), Y)]$$

$\Psi : \mathcal{X} \to \mathcal{Y}^C$ is a distributed inference algorithm parameterized by θ that outputs one prediction for each client and $\ell_h(\hat{\boldsymbol{y}}, y) \triangleq \ell(h(\hat{\boldsymbol{y}}; \mathcal{G}(T)), y)$ is a loss function where $h : \mathcal{Y}^C \to \mathcal{Y}$ (which could be stochastic) post-processes the client-specific outputs to create a single output based on the communication graph at the final inference time T, and ℓ could be any standard loss function.

Outline

- Introduction
- Setup
- Method
- Results
- Conclusion

Method – MAGS: Challenges

Challenges in cross-device decentralized collaborative learning (CL):

- At inference time, faults are encountered
- A single server is a point of catastrophic failure
- Variability among device predictions

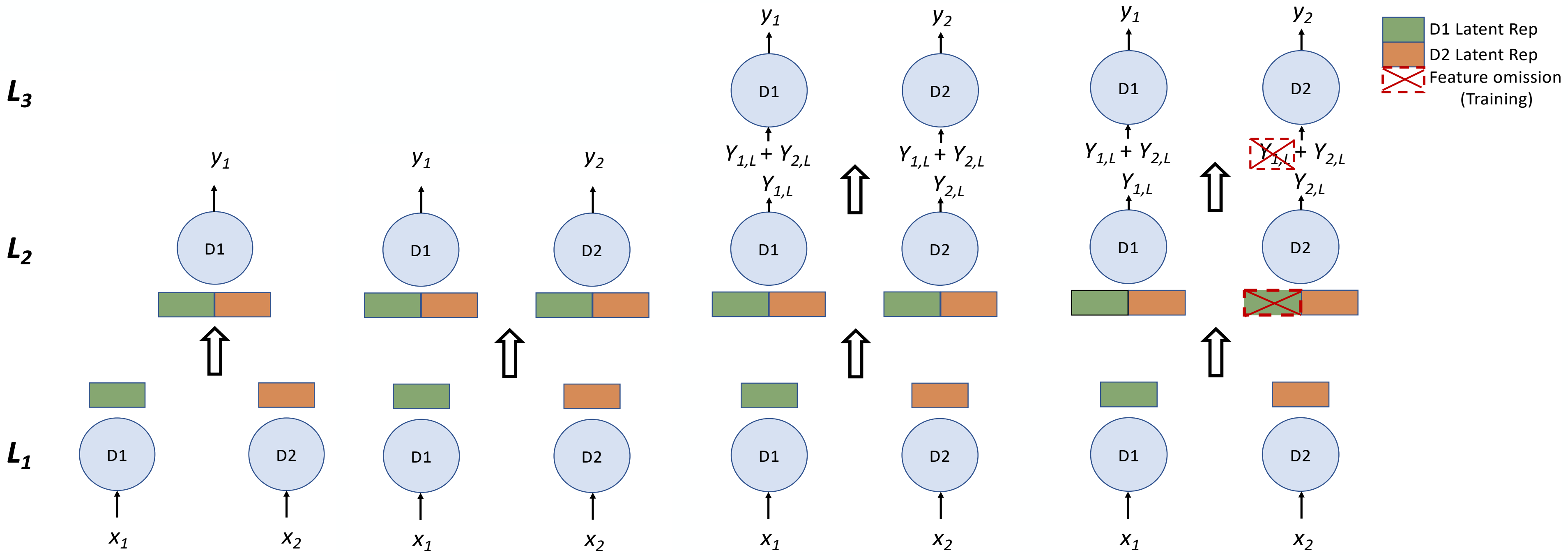
Method – MAGS: Solutions

Keys to fault tolerant inference over dynamic device network:

- ***Inference faults:*** Training with simulated faults
- ***Catastrophic failure:*** Replicating aggregators
- ***Prediction variability:*** Ensembling predictions via gossip

Multiple **A**ggregation with **G**ossip Rounds and **S**imulated Faults (**MAGS**)

Method – MAGS



(a) CL (e.g., VFL) (b) MACL (c) MACL-G (d) CD-MACL-G (MAGS)

Figure (a) CL setup with D1 acting as a client as well as the aggregating server. (b) MACL arrangement has both the devices acting as data aggregators aside from being clients. (c) MACL-G is an extension of MACL wherein the output log probabilities are averaged for final prediction. (d) CD-MACL-G is a variant of MACL-G where during the training phase, representation from D2 to D1 is not communicated by design.

Method – MAGS Dynamic Risk

Dynamic Risk Lower Bound

Risk with faults

Fault Rate

Number of Aggregators

Risk without faults

Risk of random predictor

$$R_h(\theta; \mathcal{G}_r(t)) \geq (1 - r^K) \cdot R_h(\theta; \mathcal{G}_{\text{base}}) + r^K \cdot R_{rnd}$$

Equation 1: Multiple aggregators, K, reduce the chance of catastrophic failure of the system, even for high failure rate, r

Method – MAGS Dynamic Risk

Dynamic Risk Upper Bound

Risk with faults

Failure Rate

Number of Aggregators

1

2

Non ensembled Risk of device k

Inference output of device k

Inference output of ensembled model

$$R_h(\theta;\mathcal{G}_r(t)) \leq \boxed{r^K \cdot R_{rnd}} + \boxed{(1-r^K)\left(\frac{1}{K}\sum_k R(k) - \frac{1}{K}\sum_k \mathbb{E}[\ell(\Psi_k,\Psi^{ens})]\right)}$$

$$\boxed{+(1-r^K)\sqrt{2K}\lambda^G\,\mathbb{E}\Big[\max_{j,j'\in\mathcal{K}}\|\Psi_j-\Psi_{j'}\|_2\Big]}$$

Risk of random predictor

3

Spectral Radius

Number of gossip rounds

1. Risk of catastrophic failure
2. Reduction in risk due to model ensembling
3. Approximation of ensembling using gossip averaging

Outline

- Introduction
- Setup
- Method
- Results
- Conclusion

Results – Experimental Setup

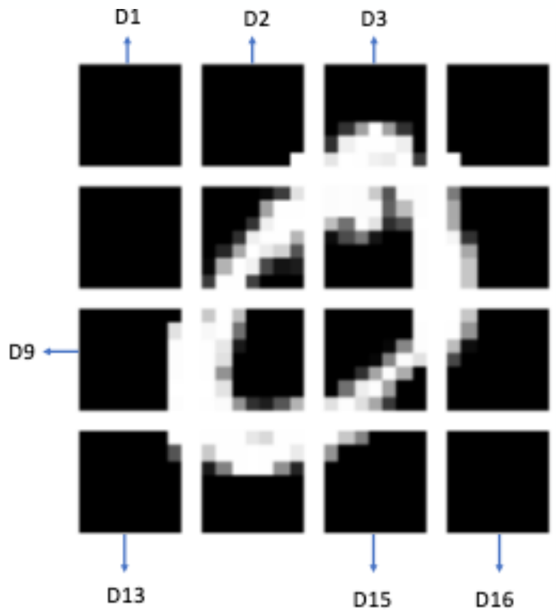


Figure: MNIST images split into 16 (49) sections for our evaluation. Each section is assigned to a particular device, denoted D_i

Datasets:
- MNIST
- CIFAR10/100
- StarCraftMNIST
- TinyImageNet

Baseline Communication Network:
- Complete Graph
- Grid Graph
- Ring Graph
- Random Geometric Graph

Baseline Methods:
- VFL
- PD-VFL
- Ablated versions of MAGS

Fault Models:
- Device
- Communication
- Temporal

Results – Multiple Fault Modes and Topologies

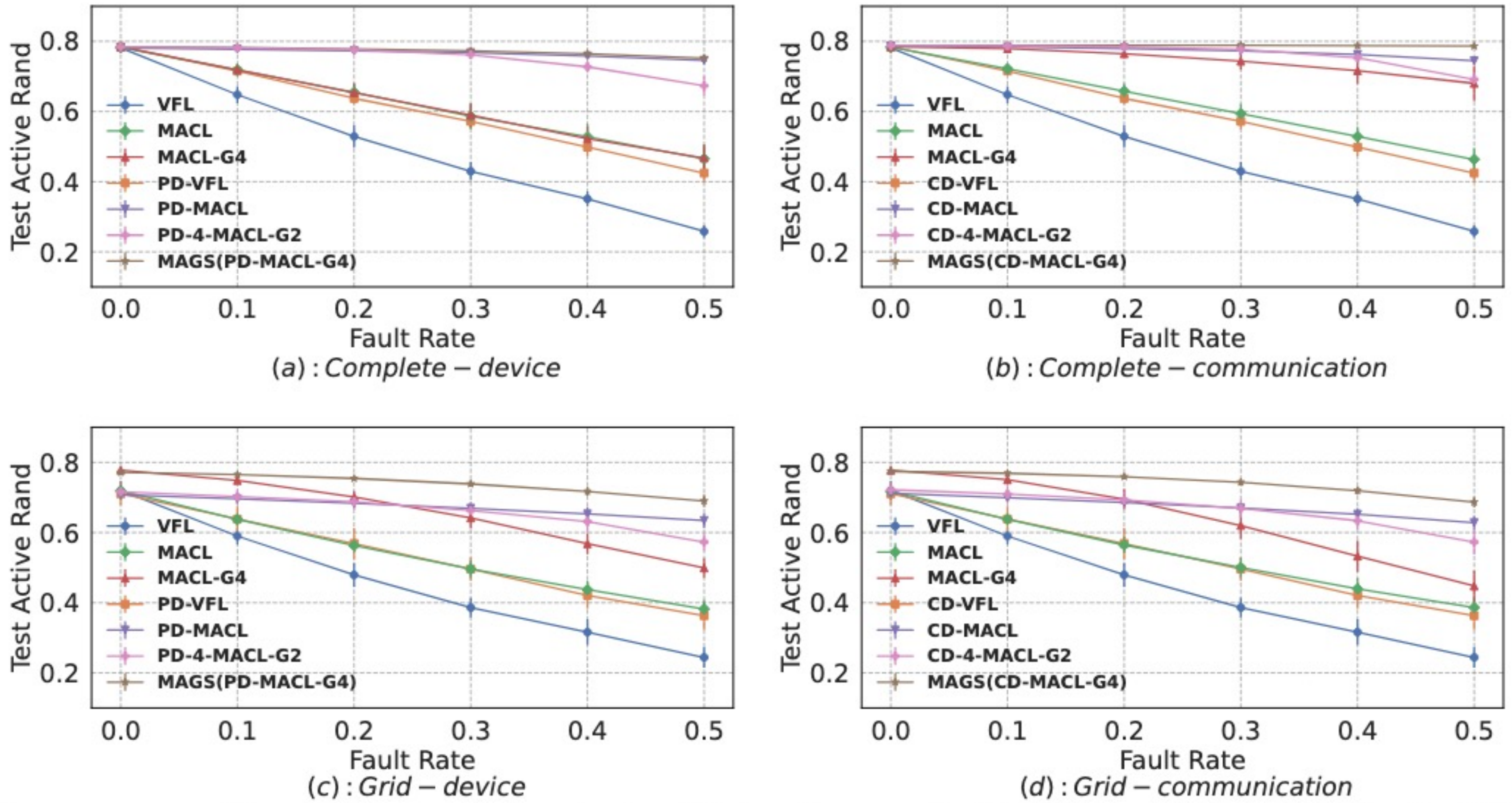


Figure: Test accuracy with and without communication (CD-) and party-wise (PD-) Dropout method for StarCraftMNIST with 16 devices. Here we include models trained under a dropout rate of 30% (marked by 'PD-' or 'CD-').

- Our proposed method, MAGS has the best performance for both device and communication faults
- Unlike VFL, performance degradation for MAGS at higher fault rates is graceful
- Low communication variants of MAGS also demonstrate robust performance to high fault rates

Results – K-MACL

	Complete		RGG $r = 2.5$		RGG $r = 2$		RGG $r = 1.5$		RGG $r = 1$		Ring	
	Avg	# Comm.	Avg	# Comm.	Avg	# Comm.	Avg	# Comm.	Avg	# Comm.	Avg	# Comm.
VFL	0.430	10.6	0.406	7.4	0.407	5.2	0.375	3.5	0.386	2.0	0.385	1.4
4-MACL	0.591	42	0.572	29	0.555	20.4	0.517	14.8	0.488	7.98	0.485	5.6
4-MACL-G2	0.687	126	0.661	87	0.623	61.2	0.566	44.8	0.491	23.94	0.484	16.8
MACL	0.594	168.5	0.581	114.9	0.558	80.8	0.528	58.7	0.503	33.5	0.507	22.7
MACL-G4	0.732	836.2	0.728	572.1	0.721	407.2	0.689	293.9	0.62	168.2	0.558	113.4

Table 1: Active Rand (Avg) performance at test time with 30% communication fault rate.

- Compared to VFL, MACL performs better but it comes at higher communication cost.
- Our proposed 4-MACL is shown to be a low communication cost alternative to MACL.
- Comparing performance and communication of 4-MACL and MACL, reveals an efficient trade-off between robustness and communication cost.
- 4-MACL with a poorly connected graph is still better than VFL with well connected graph

Results – Multiple Datasets

		MNIST				SCMNIST				CIFAR10			
		Active			Any	Active			Any	Active			Any
		Worst	Rand	Best	Rand	Worst	Rand	Best	Rand	Worst	Rand	Best	Rand
Test Fault Rate = 0.3	VFL	nan	0.507	nan	nan	nan	0.430	nan	nan	nan	0.267	nan	nan
	PD-VFL	nan	0.684	nan	nan	nan	0.572	nan	nan	nan	0.355	nan	nan
	4-MACL-G2	0.837	0.869	0.901	0.651	0.659	0.692	0.726	0.521	0.382	0.419	0.458	0.328
	MACL	0.098	0.69	**0.995**	0.512	0.075	0.592	**0.951**	0.444	0.013	0.342	**0.843**	0.269
	MACL-G4	0.945	0.945	0.946	0.691	0.742	0.743	0.743	0.548	0.478	0.479	0.480	0.362
	CD-4-MACL-G2	0.959	0.961	0.971	**0.717**	0.768	0.777	0.789	**0.592**	0.474	0.496	0.521	0.385
	MAGS(CD-MACL-G4)	**0.979**	**0.979**	0.979	0.714	**0.788**	**0.788**	0.788	0.582	**0.516**	**0.516**	0.516	**0.392**
Test Fault Rate = 0.5	VFL	nan	0.294	nan	nan	nan	0.258	nan	nan	nan	0.181	nan	nan
	PD-VFL	nan	0.488	nan	nan	nan	0.424	nan	nan	nan	0.263	nan	nan
	4-MACL-G2	0.564	0.612	0.721	0.423	0.466	0.519	0.620	0.368	0.251	0.303	0.387	0.228
	MACL	0.042	0.518	0.966	0.313	0.035	0.465	**0.925**	0.280	0.007	0.264	**0.762**	0.182
	MACL-G4	0.843	0.847	0.851	0.474	0.676	0.680	0.684	0.389	0.402	0.401	0.413	0.252
	CD-4-MACL-G2	0.863	0.852	0.923	**0.581**	0.693	0.691	0.763	**0.482**	0.392	0.422	0.499	0.305
	MAGS(CD-MACL-G4)	**0.974**	**0.975**	**0.976**	0.538	**0.785**	**0.786**	0.787	0.443	**0.501**	**0.504**	0.507	**0.301**

Table 2: Best models for 30% and 50% complete-communication test fault rates within 1 standard deviation are bolded.

- Across different datasets, MAGS has the best performance for the practical Active Rand client selection function
- Gossip's benefits for ML performance and reduced client variability is evidenced by the Best/Worst oracle metric gap

Outline

- Introduction
- Setup
- Method
- Results
- Conclusion

Conclusion

- We defined Dynamic Network Collaborative Learning (DN-CL) to address robust inference for vertically split data in decentralized networks that are prone to catastrophic faults.
- We proposed MAGS (Multiple Aggregation with Gossip Rounds and Simulated Faults), a novel method that ensures high fault tolerance by integrating three key features:
  - Fault Simulation
  - Aggregator Replication
  - Gossip Layers
- Experiments demonstrated that MAGS significantly outperforms standard VFL baselines, maintaining high performance even at extreme fault rates.
- Our analysis revealed a fundamental tradeoff between communication cost and robustness, which can be directly controlled by adjusting the number of aggregators.

Thank You