Fair Resource Allocation in Federated Learning

***Tian Li (CMU)**, Maziar Sanjabi (Facebook AI), Ahmad Beirami (Facebook AI), Virginia Smith (CMU)*

tianli@cmu.edu

Federated Learning

Federated Learning

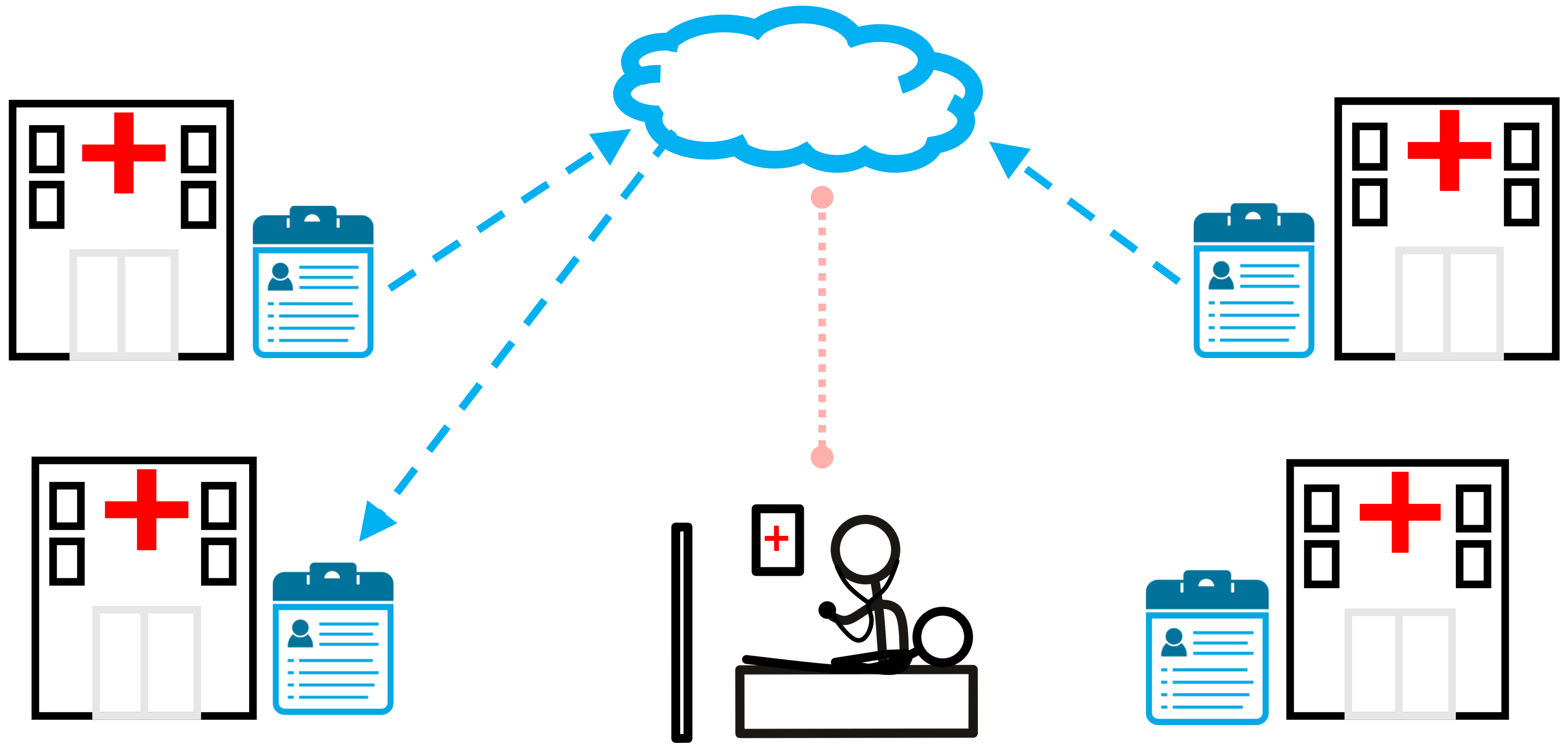
Privacy-preserving training in heterogeneous, (potentially) massive networks

Federated Learning

Privacy-preserving training in heterogeneous, (potentially) massive networks

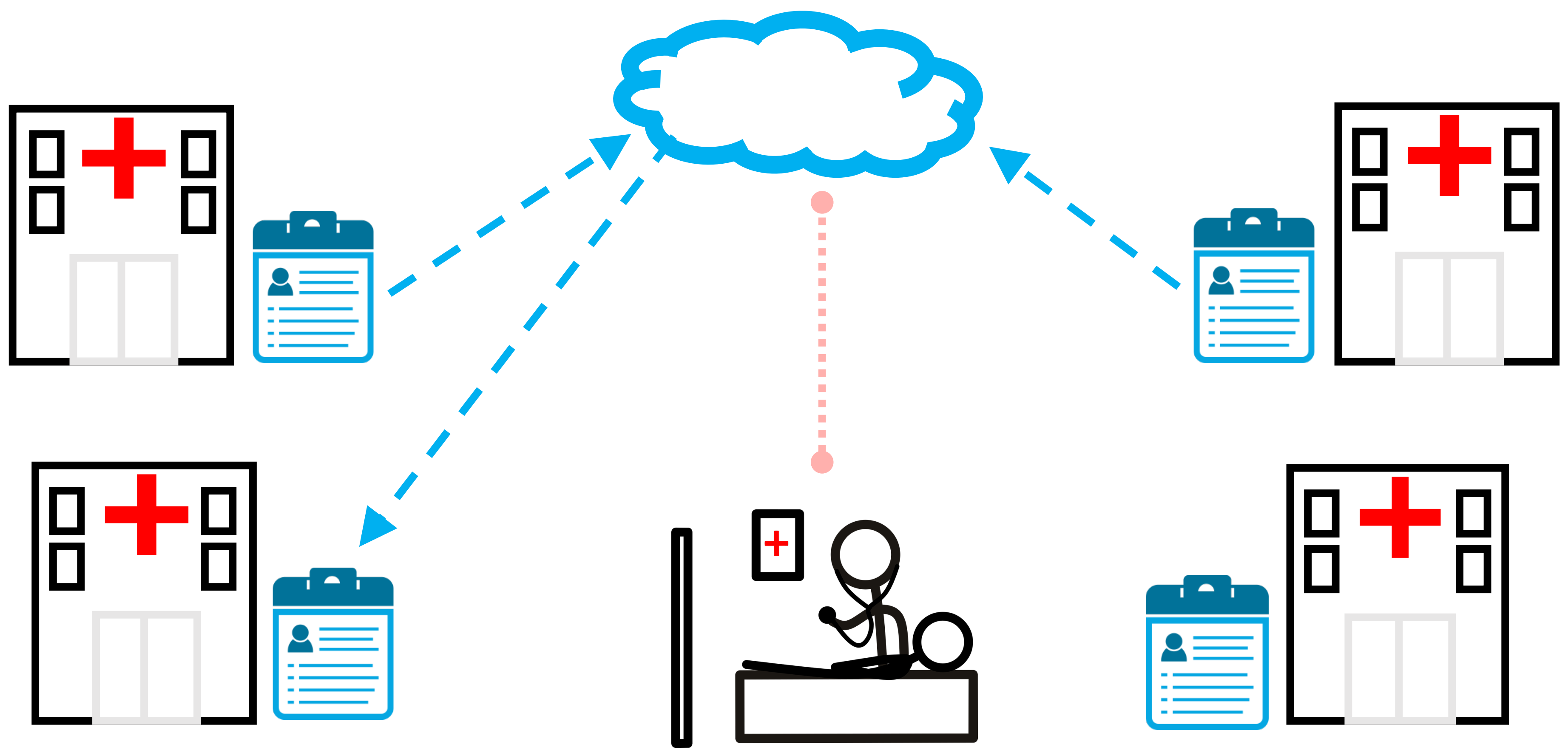
Federated Learning

Privacy-preserving training in heterogeneous, (potentially) massive networks



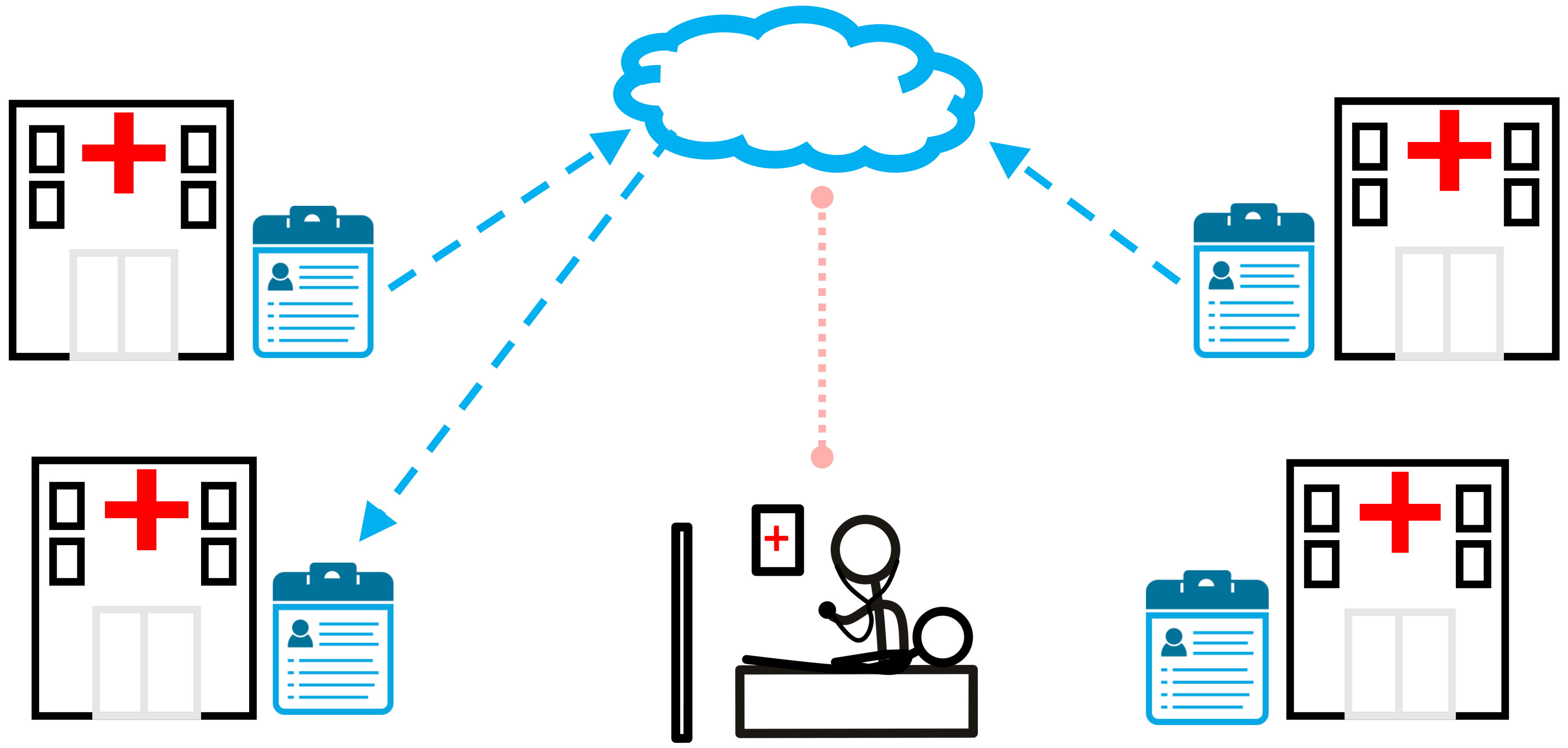
Federated Learning

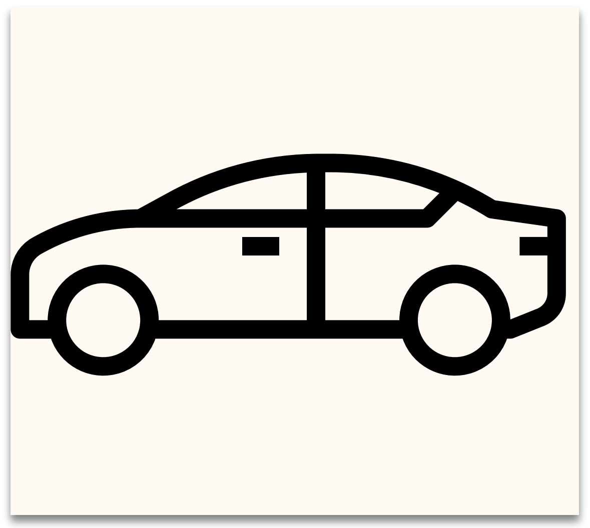
Privacy-preserving training in heterogeneous, (potentially) massive networks



Federated Learning

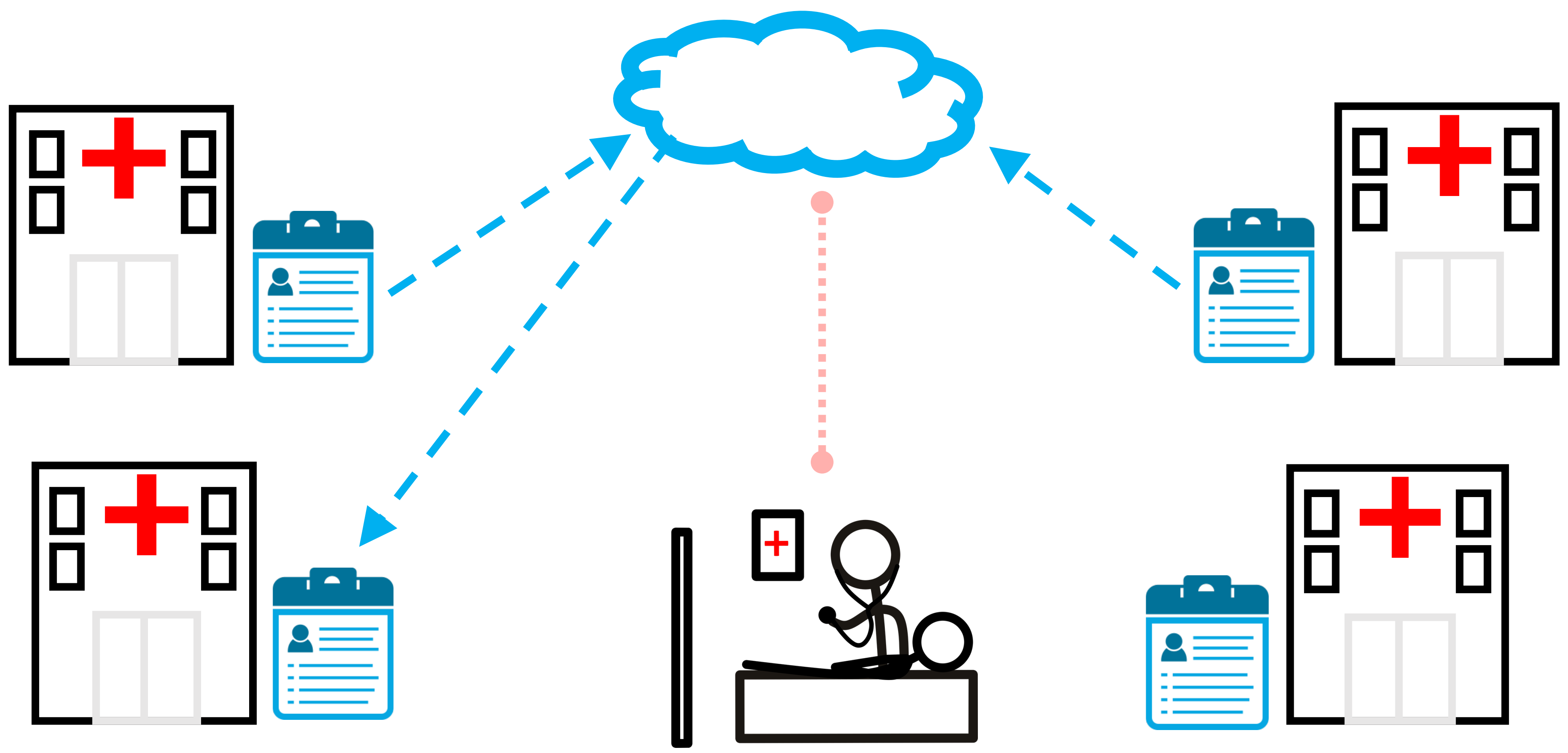
Privacy-preserving <u>training</u> in heterogeneous, (potentially) massive networks

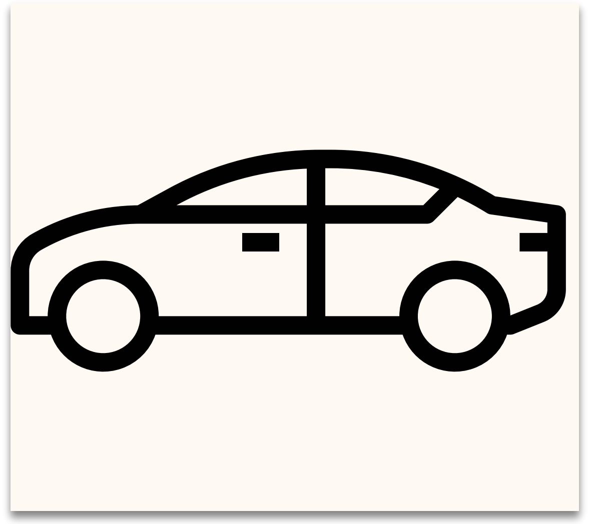


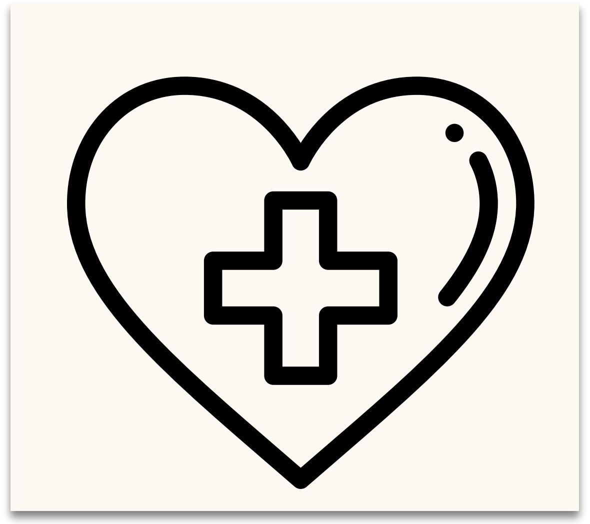


Federated Learning

Privacy-preserving training in heterogeneous, (potentially) massive networks

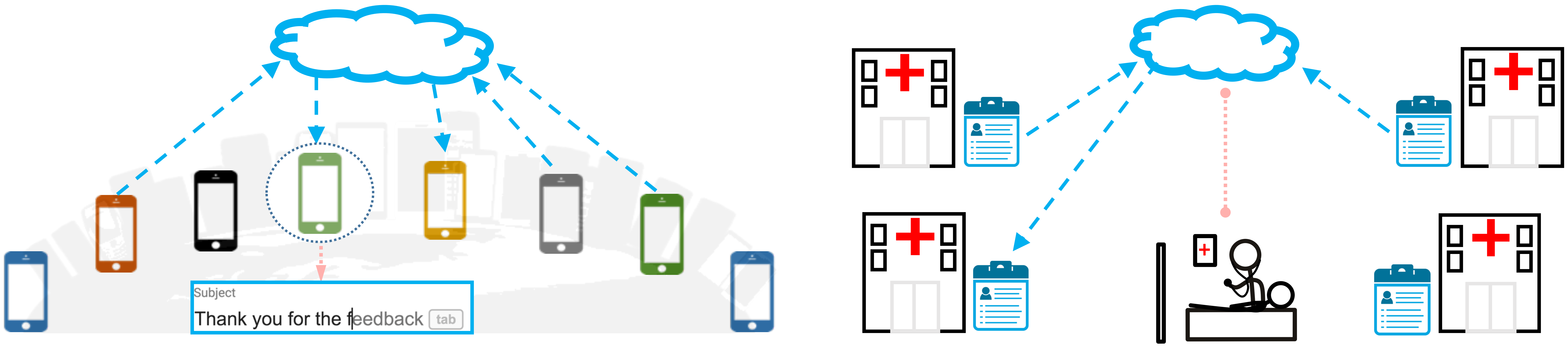


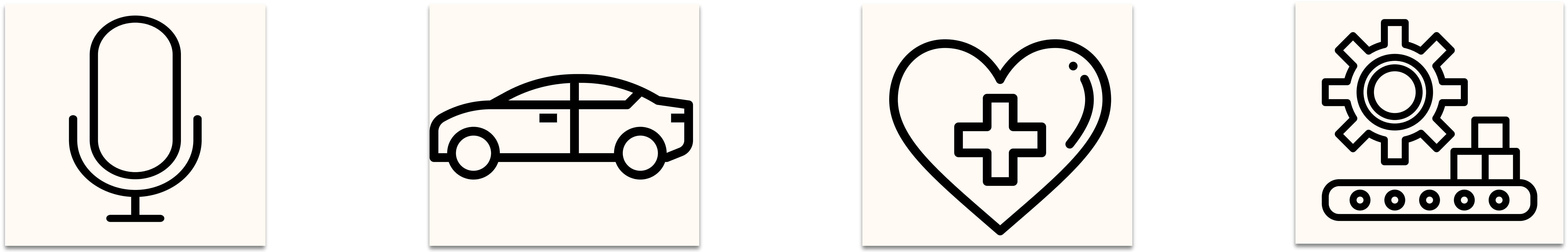




Federated Learning

Privacy-preserving training in heterogeneous, (potentially) massive networks





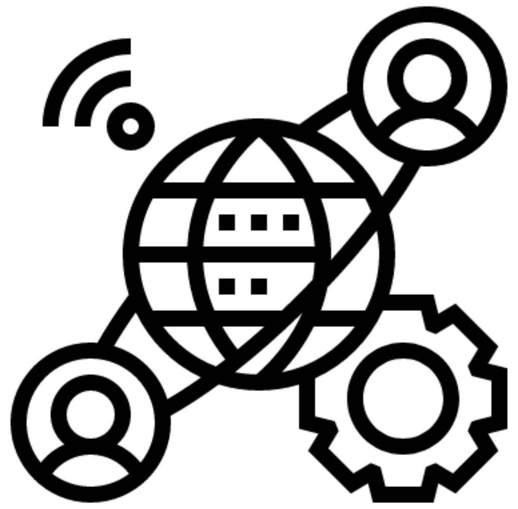
Challenges

Challenges

objective: $$\min_{w} \left(p_1 F_1 + p_2 F_2 + \cdots + p_N F_N\right)$$

Challenges

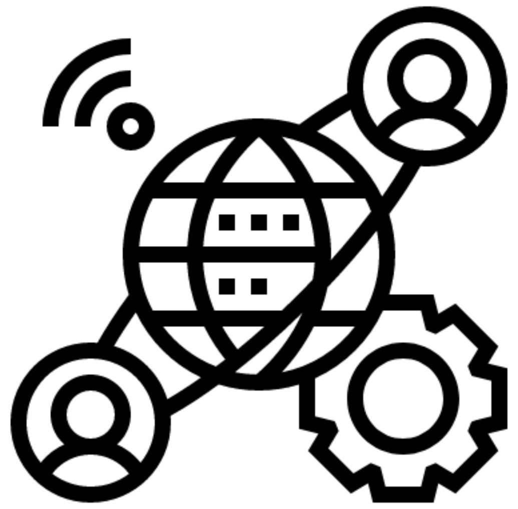
objective: $\min_{w} \left(p_1 F_1 + p_2 F_2 + \cdots + p_N F_N \right)$



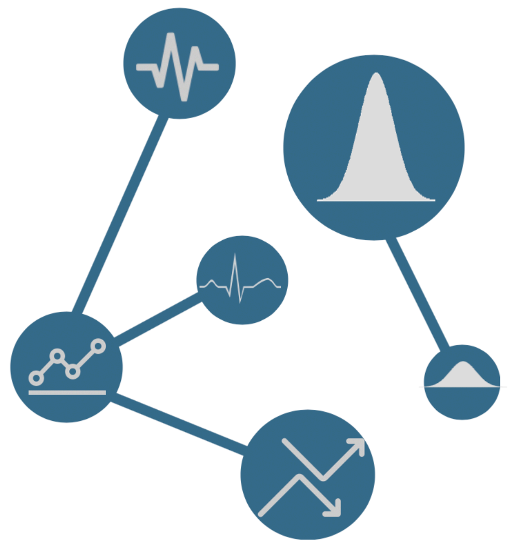
no accuracy guarantee for individual devices

Challenges

objective: $\min_{w} \left(p_1 F_1 + p_2 F_2 + \cdots + p_N F_N \right)$



no accuracy guarantee for individual devices

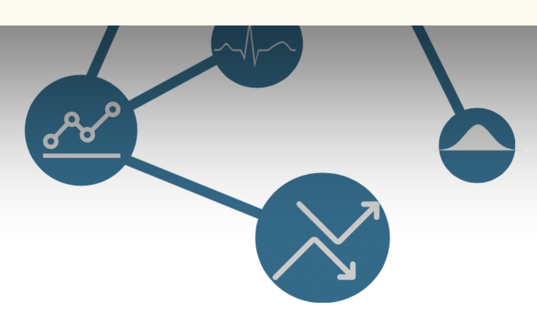


model performance can vary widely

Challenges

objective: $$\min_{w} \left(p_1 F_1 + p_2 F_2 + \cdots + p_N F_N \right)$$

Can we devise an **efficient** federated optimization method to encourage a more **fair** (i.e., more **uniform**) distribution of the model performance across devices?



model performance can vary widely

Fair Resource Allocation Objective

Fair Resource Allocation Objective

$$\min_{w} \left(p_1 F_1 + p_2 F_2 + \cdots + p_N F_N \right)$$

Fair Resource Allocation Objective

q-FFL: $$\min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

Fair Resource Allocation Objective

q-FFL: $$\min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

- Inspired by α-fairness for fair resource allocation in wireless networks

Fair Resource Allocation Objective

q-FFL: $$\min_w \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

- Inspired by α-fairness for fair resource allocation in wireless networks
- A tunable framework ($q = 0$: previous objective; $q = \infty$: minimax fairness)

Fair Resource Allocation Objective

q-FFL: $$\min_w \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

- Inspired by α-fairness for fair resource allocation in wireless networks
- A tunable framework ($q = 0$: previous objective; $q = \infty$: minimax fairness)
- Theory

Fair Resource Allocation Objective

q-FFL: $$\min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

- Inspired by α-fairness for fair resource allocation in wireless networks
- A tunable framework ($q = 0$: previous objective; $q = \infty$: minimax fairness)
- Theory

Fair Resource Allocation Objective

$$\textbf{q-FFL:} \quad \min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

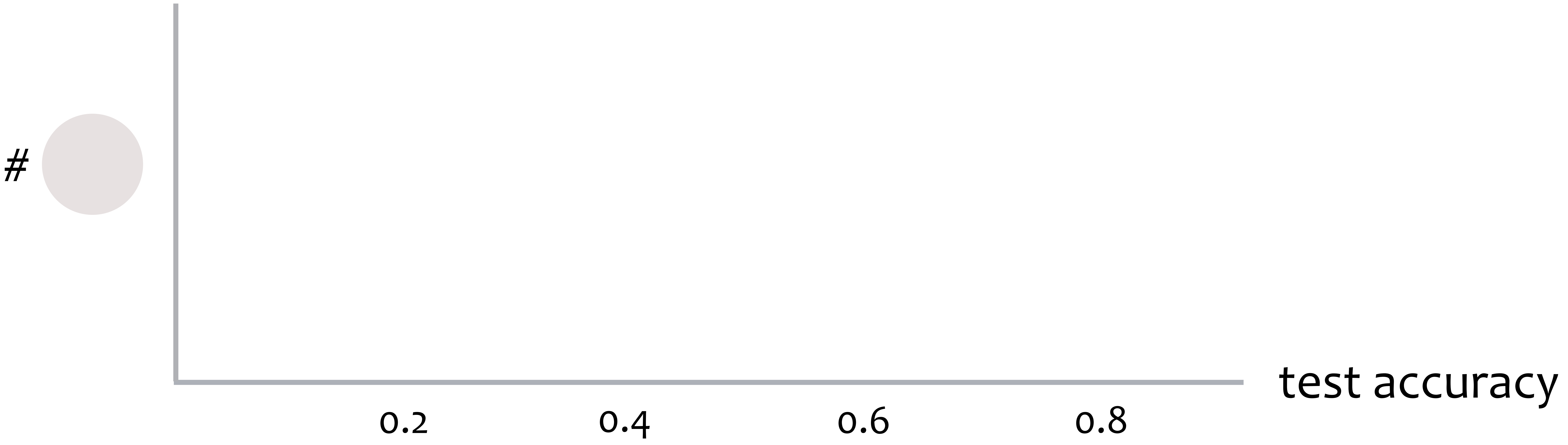
- Inspired by α-fairness for fair resource allocation in wireless networks
- A tunable framework ($q = 0$: previous objective; $q = \infty$: minimax fairness)
- Theory
  - ✓ Generalization guarantees
  - ✓ Increasing q results in more ‘uniform’ accuracy distributions (in terms of various uniformity measures such as variance)

Fair Resource Allocation Objective

q-FFL: $$\min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

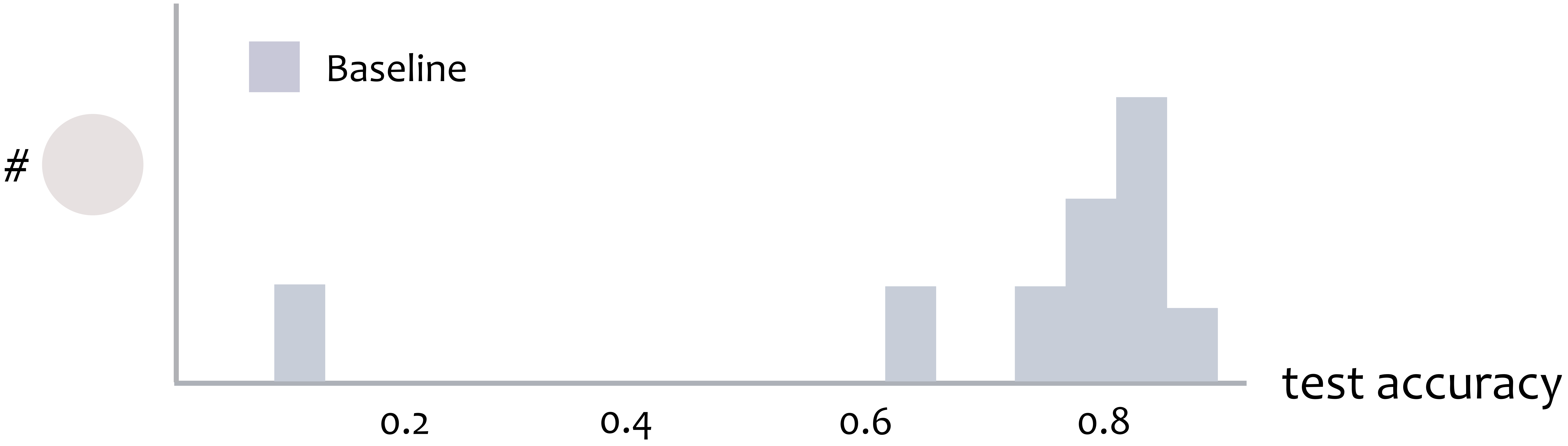
Fair Resource Allocation Objective

q-FFL: $$\min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$



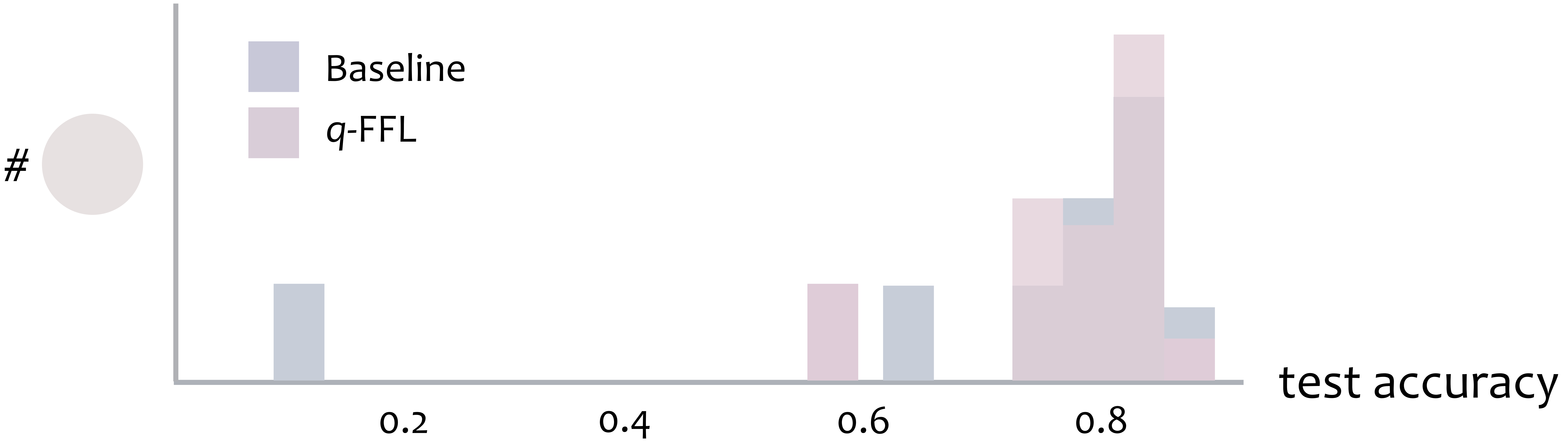
Fair Resource Allocation Objective

q-FFL: $$\min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$



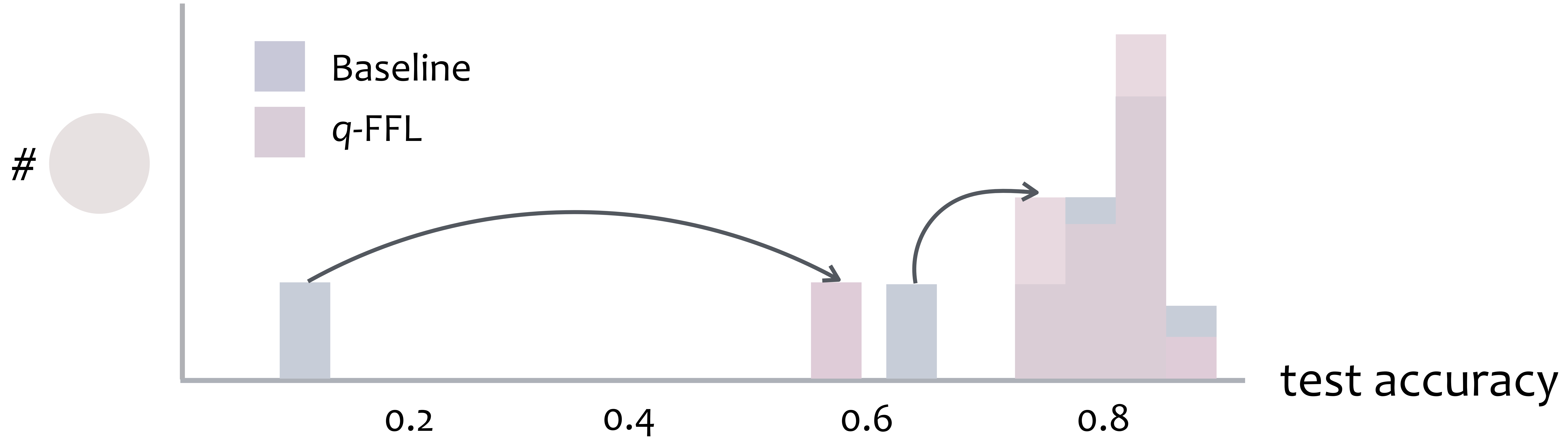
Fair Resource Allocation Objective

q-FFL: $$\min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$



Fair Resource Allocation Objective

$$\textbf{q-FFL:} \quad \min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$



Efficient Solver

Efficient Solver

Challenges

Efficient Solver

Challenges

- Different fairness/accuracy tradeoffs: different q's

Efficient Solver

Challenges

- Different fairness/accuracy tradeoffs: different q's

Efficient Solver

Challenges

- Different fairness/accuracy tradeoffs: different q’s
- Heterogeneous networks, expensive communication

Efficient Solver

Challenges

- Different fairness/accuracy tradeoffs: different q's →
- Heterogeneous networks, expensive communication →

Efficient Solver

Challenges

- Different fairness/accuracy tradeoffs: different q's →
- Heterogeneous networks, expensive communication →

Efficient Solver

Challenges

High level ideas

- Different fairness/accuracy tradeoffs: different q's →
- Heterogeneous networks, expensive communication →

Efficient Solver

Challenges

- Different fairness/accuracy tradeoffs: different q's
- Heterogeneous networks, expensive communication

High level ideas

- Dynamically estimate the step sizes associated with different q's

Efficient Solver

Challenges

- Different fairness/accuracy tradeoffs: different q's
- Heterogeneous networks, expensive communication

High level ideas

- Dynamically estimate the step sizes associated with different q's

Efficient Solver

Challenges

- Different fairness/accuracy tradeoffs: different q's
- Heterogeneous networks, expensive communication

High level ideas

- Dynamically estimate the step sizes associated with different q's
- Allow for low device participation, local updating

Empirical Results

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespeare	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
LEAF *(leaf.cmu.edu)*

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespeare	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
LEAF (leaf.cmu.edu)

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespe	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
LEAF (leaf.cmu.edu)

similar average accuracy

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespe	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
LEAF *(leaf.cmu.edu)*

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespe	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
***LEAF** (leaf.cmu.edu)*

decrease variance significantly

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespe	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
LEAF (leaf.cmu.edu)

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespe	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
LEAF (leaf.cmu.edu)

increase the accuracy of the worst 10% devices

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespe	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
LEAF (leaf.cmu.edu)

slightly decrease
the accuracy of the
best devices

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Synthetic	$q = 0$	80.8	18.8	100.0	724
	$q = 1$	79.0	**31.1**	100.0	**472**
Vehicle	$q = 0$	87.3	43.0	95.7	291
	$q = 5$	87.7	**69.9**	94.0	**48**
Sent140	$q = 0$	65.1	15.9	100.0	697
	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespe	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

Empirical Results

Benchmark:
LEAF (leaf.cmu.edu)

slightly decrease
the accuracy of the

- on average, reduce the variance of accuracy across all devices by 45%
- solving the objective orders-of-magnitude more quickly than other baselines

	$q = 1$	66.5	**23.0**	100.0	**509**
Shakespe	$q = 0$	51.1	39.7	72.9	82
	$q = .001$	52.1	**42.1**	69.0	**54**

q-FFL Extended: meta-learning

q-FFL Extended: meta-learning

Fair meta-learning (q-FFL+MAML): fair initializations across tasks

q-FFL Extended: meta-learning

Fair meta-learning (q-FFL+MAML): fair initializations across tasks

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Omniglot	$q = 0$	79.1	61.2	94.0	93
	$q = .1$	79.3	**62.5**	93.8	**86**

q-FFL Extended: meta-learning

Fair meta-learning (q-FFL+MAML): fair initializations across tasks

Dataset	Objective	Average	Worst 10%	Best 10%	Variance
Omniglot	$q = 0$	79.1	61.2	94.0	93
	$q = .1$	79.3	**62.5**	93.8	**86**

More broadly, an alternative/generalization of minimax optimization

Many other scenarios

code & paper: OpenReview / *cs.cmu.edu/~litian*

Thanks!

code & paper: OpenReview / *cs.cmu.edu/~litian*