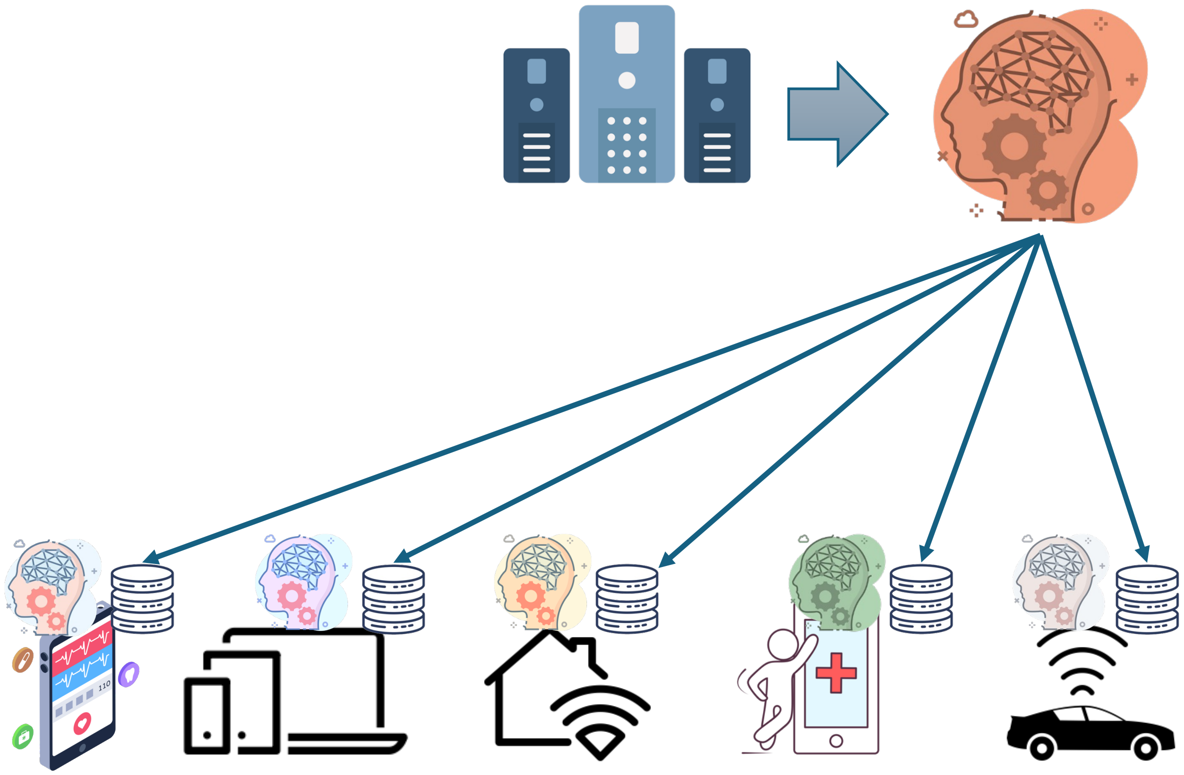
FedCert: Federated Accuracy Certification

Minh Hieu Nguyen[1*], **Huu Tien Nguyen[1*]**, Trung Thanh Nguyen[2],
Manh Duong Nguyen[1], Trong Nghia Hoang[3], Truong Thao Nguyen[4], Phi Le Nguyen[1]

[*]Equal Contribution

[1]Hanoi University of Science and Technology, Vietnam
[2]Nagoya University, Japan
[3]Washington State University, United States
[4]National Institute of Advanced Industrial Science and Technology (AIST), Japan

Federated Learning

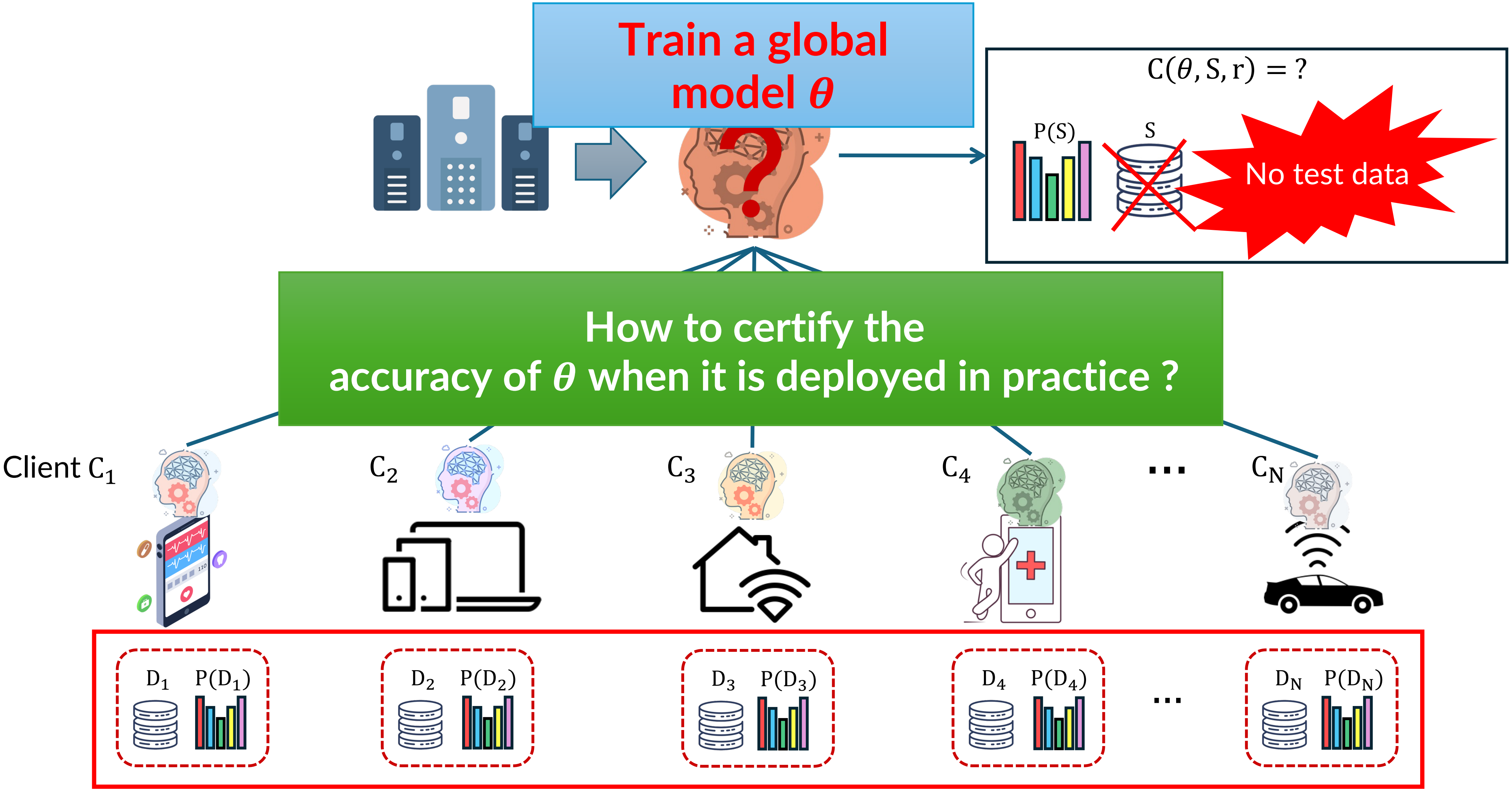


Federated Learning (FL)

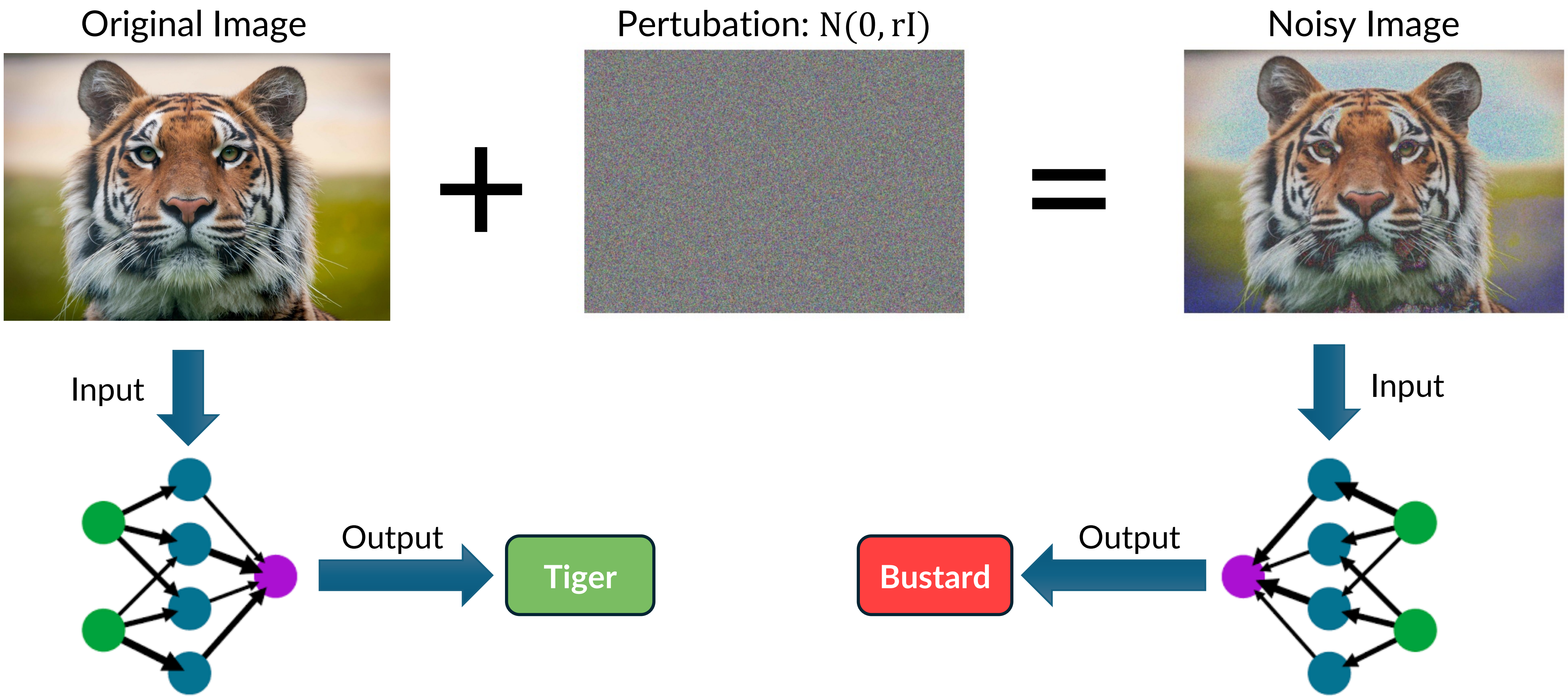
- Each client trains a local model using its data
- All the local models are aggregated to generate a global model

- Preserve data privacy
  - Healthcare
  - Finance
- Leverage computing resources from multiple clients

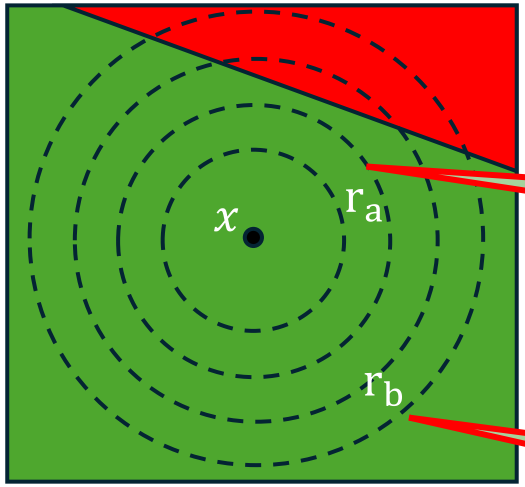
Problem Definition



Certified Accuracy



Certified Accuracy



$f(x+\epsilon)$ remain **correct,** $\forall\ \epsilon \sim N(0, r_a I)$

Classifier f is robust at sample x within radius r_a

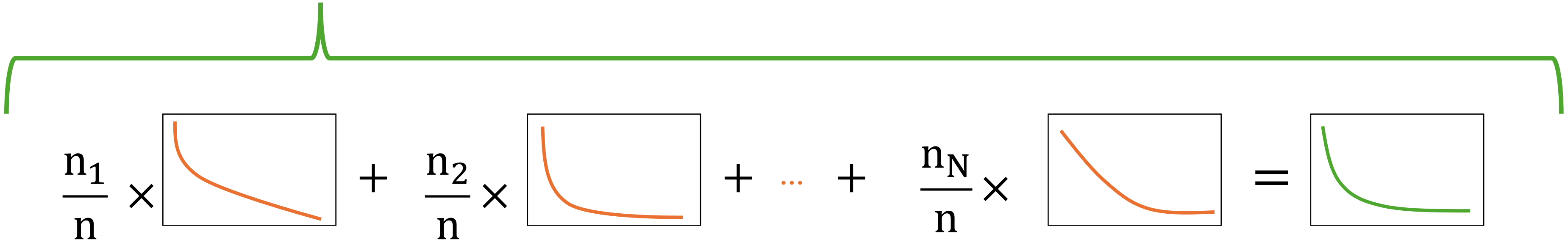
$f(x+\epsilon)$ do not remain **correct,** $\forall\ \epsilon \sim N(0, r_b I)$

Classifier f is not robust at sample x within radius r_b

Certified Accuracy: $C(f,\ S,\ r) = \frac{n_S^{robust}}{n_S}$ $\begin{cases} \text{Dataset } S \text{ with } n_S \text{ samples} \\ \text{Classifier } f \text{ is robust at } n_S^{robust} \text{ samples within radius } r \end{cases}$

Related Work

- VW: Volume-based Weighted-sum Method
  - $n = \sum_{i=1}^{N} n_i$ with n_i is the cardinality of the local dataset D_i
    - $c(\theta, S, r) \approx \sum_{i=1}^{N} \frac{n_i}{n} c(\theta, D_i, r)$

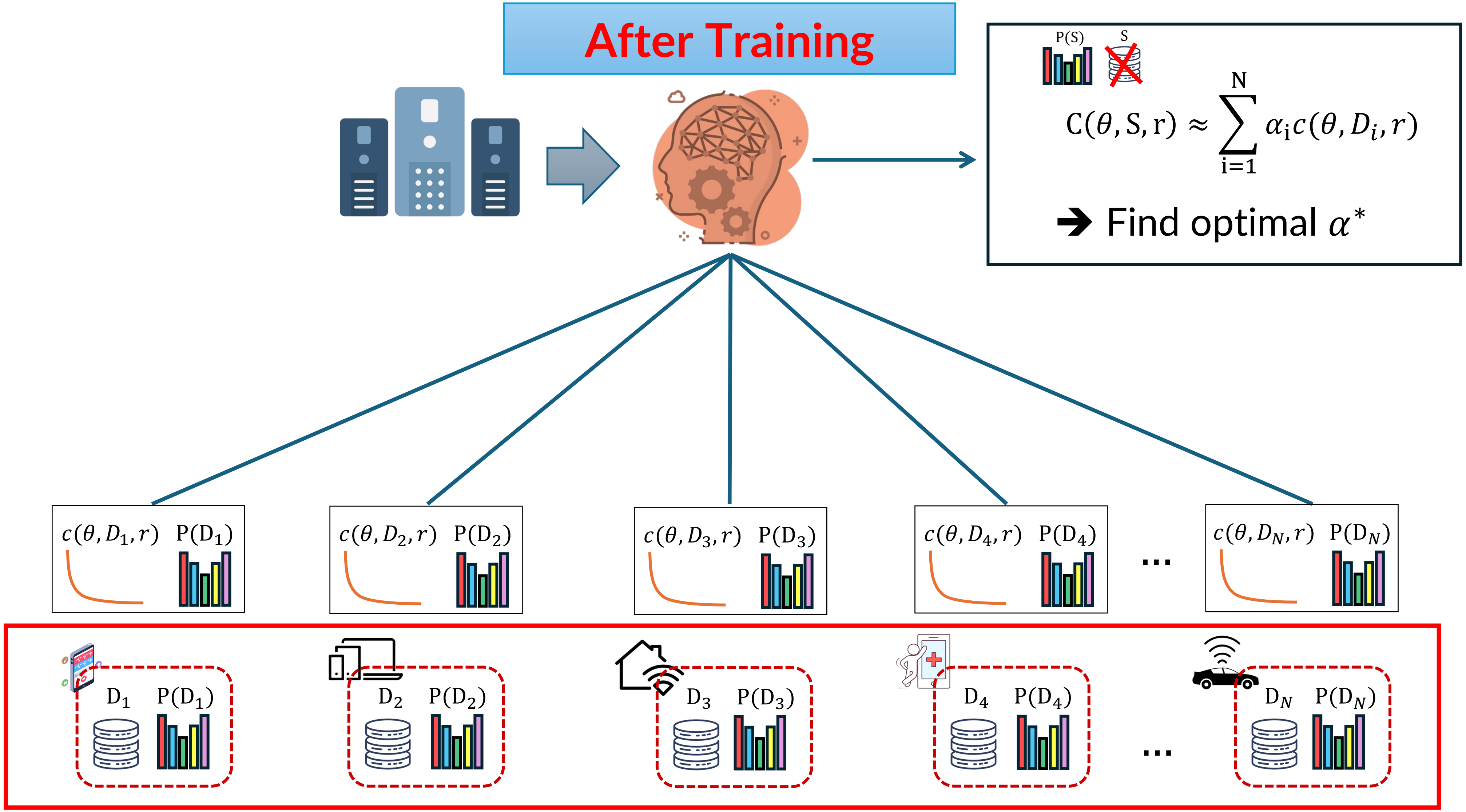


Approximation
certified accuracy

Drawback: VW leads to less reliable evaluations of the global model's performance when client data is highly heterogeneous

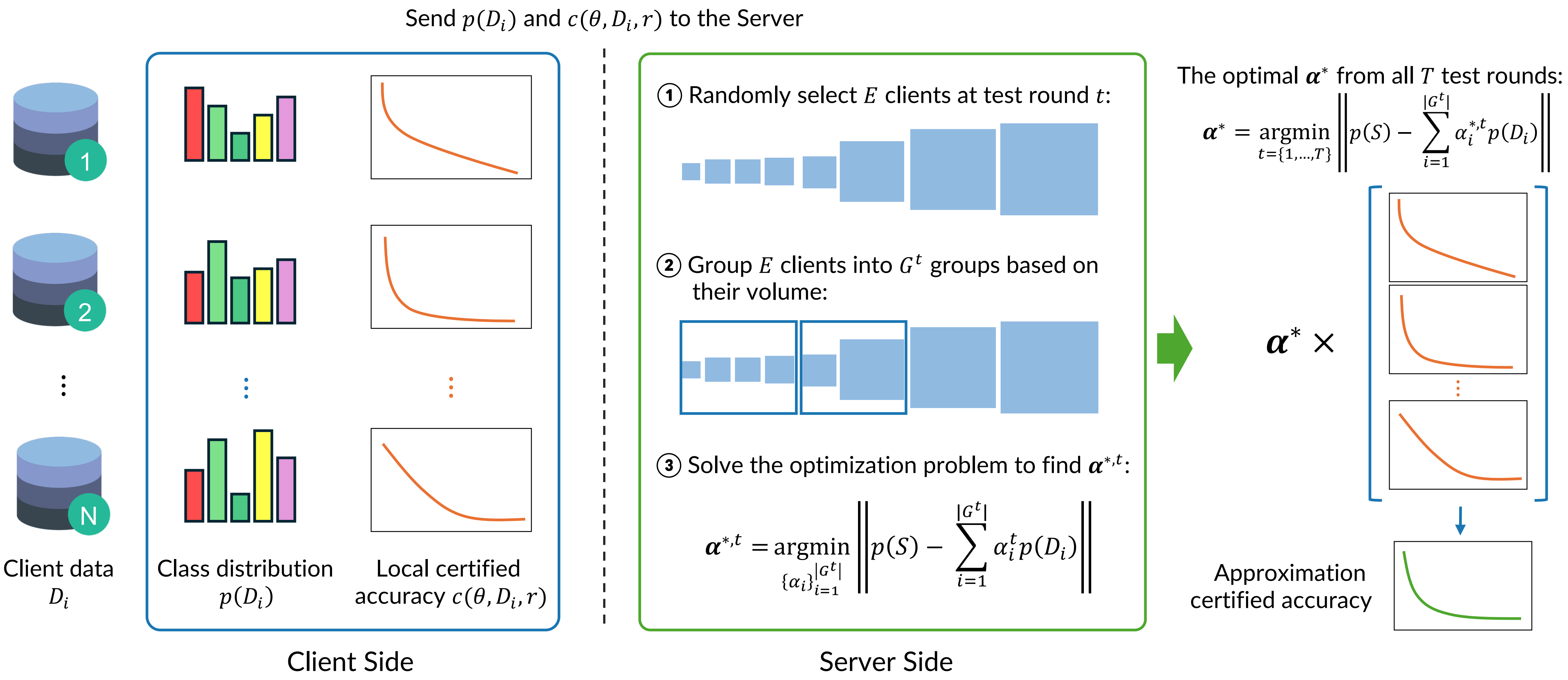
[1] H. R. Roth *et al.*, "NVIDIA FLARE: Federated learning from simulation to real-world," *Computing Research Repository arXiv Preprints*, arXiv:2210.13291, 2022

Motivation

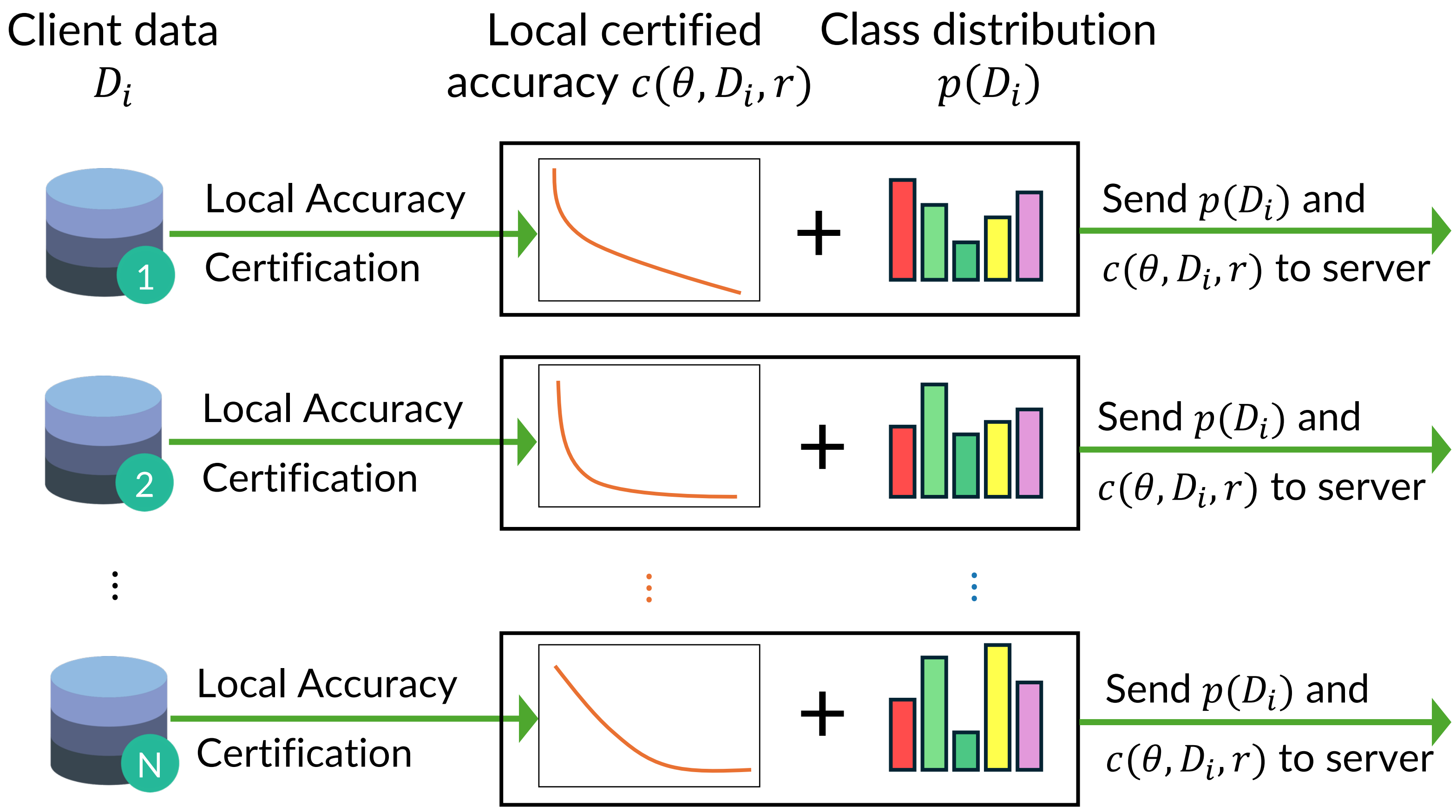


After Training
P(S)
S
$C(\theta, S, r) \approx \sum_{i=1}^{N} \alpha_i c(\theta, D_i, r)$
➔ Find optimal α^*
$c(\theta, D_1, r)$ $P(D_1)$
$c(\theta, D_2, r)$ $P(D_2)$
$c(\theta, D_3, r)$ $P(D_3)$
$c(\theta, D_4, r)$ $P(D_4)$
…
$c(\theta, D_N, r)$ $P(D_N)$
D_1 $P(D_1)$
D_2 $P(D_2)$
D_3 $P(D_3)$
D_4 $P(D_4)$
…
D_N $P(D_N)$

Methodology – Overview



Methodology – Client Side



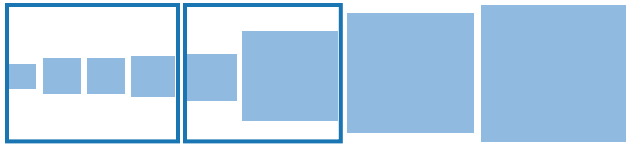
[2] J.Cohen *et al.*,"Certified adversarial robustness via randomized smoothing," in *Proceedings of the 36th International Conference on Machine Learning*, 2019, pp. 1310–1320.

Methodology – Server Side

for t = 1 to T do

① Random select E clients at test round t:

② Group the clients into V groups based on their volume:



Algorithm 1 Grouping Algorithm

```
Input: Small clients 𝒮𝒞, large clients ℒ𝒞, threshold τ
Sort 𝒮𝒞 in ascending order of data size n_i
Initialize 𝒱 ← ∅; Q ← Queue(𝒮𝒞)
while Q ≠ ∅ do
    Initialize virtual client V ← ∅
    while n_V < τ and Q is not empty do
        C ← Q.dequeue(); V ← V ∪ C
    end while
    Add V to 𝒱
end while
G ← ℒ𝒞 ∪ 𝒱
Return G
```

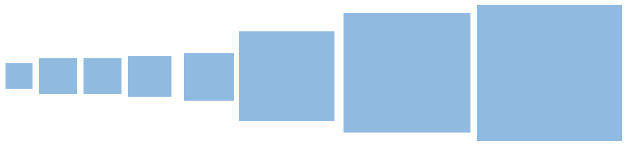
Virtual client V:

- $n_V = \sum_{j=1}^{m} n_j$
- $p(D_V) = \sum_{j=1}^{m} \frac{n_j}{n_V} p(D_j)$
- $c(\theta, D_V, r) = \sum_{j=1}^{m} \frac{n_j}{n_V} c(\theta, D_j, r)$

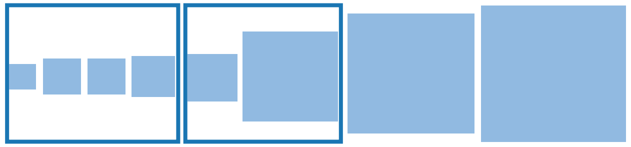
Methodology – Server Side

for t = 1 to T do

① Random select E clients at test round t:



② Group the clients into V groups based on their volume:



③ Solve the optimization problem to find $\boldsymbol{\alpha}^{*,t}$

$$\boldsymbol{\alpha}^{*,t} = \underset{\{\alpha_i\}_{i=1}^{|G^t|}}{\operatorname{argmin}} \left\| p(S) - \sum_{i=1}^{|G^t|} \alpha_i^t p(D_i) \right\|$$

Using CVXPY to solve:

$$\boldsymbol{\alpha}^{*,t} = \underset{\{\alpha_i\}_{i=1}^{|G^t|}}{\operatorname{argmin}} \left\| p(S) - \sum_{i=1}^{|G^t|} \alpha_i^t p(D_i) \right\|,$$

subject to:

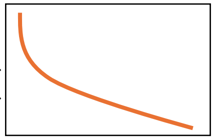
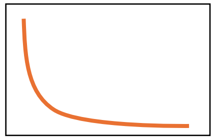
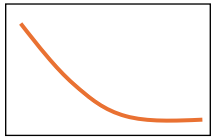
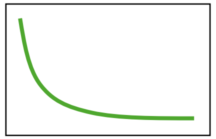
$$\sum_{i=1}^{|G^t|} \alpha_i^t = 1, 0 \leq \alpha_i^t \leq 1, \forall \text{ i} \in [1, |\text{G}^t|].$$

Methodology – Server Side

Find the optimal $\boldsymbol{\alpha}^*$ from all T test rounds:

$$\boldsymbol{\alpha}^*, \mathrm{G}^* = \underset{t=\{1,\dots,T\}}{\mathrm{arg}min} \left\| p(S) - \sum_{i=1}^{\mathrm{V}} \alpha_i^t p(D_i) \right\|$$

$$\mathrm{c}(\theta, \mathrm{S}, \mathrm{r}) \approx \sum_{\mathrm{i}=1}^{|\mathrm{G}^*|} \boldsymbol{\alpha}_{\mathrm{i}}^* c(\theta, D_i, r)$$

$\boldsymbol{\alpha}_1^* \times$  $+ \boldsymbol{\alpha}_2^* \times$  $+ \cdots + \boldsymbol{\alpha}_{|\mathrm{G}*|}^* \times$  $=$ 

Approximation
certified accuracy

Experiment Settings

- **Methods**
  - VW: Volume-based Weighted-sum
  - AP: ***FedCert*** without client grouping
  - GA: ***FedCert*** with client grouping
- **Backbone of $\boldsymbol{\theta}$**
  - ResNet-18
  - MobileNetV2
- **FL training algorithm**
  - FedAvg
  - FedProx
  - Scaffold
- **Datasets**
  - CIFAR-10
  - CIFAR-100
- **Split**
  - 50000 images for the local datasets
  - 10000 images for the target test dataset
- **The local datasets are distributed to clients using different types of non-IID distributions**
  - Pareto
  - Dirichlet

Performance of Approximation Method

TABLE I: Performance of three approximation methods for estimating certified accuracy with different FL settings

	Dataset	Client Partition	RMSE			MAPE		
			AP	GA	VW	AP	GA	VW
Resnet-18	CIFAR-10	Dirichlet	0.021	**0.014**	0.061	0.059	**0.055**	0.192
Resnet-18	CIFAR-10	Pareto	0.014	**0.008**	0.032	0.044	**0.016**	0.102
Resnet-18	CIFAR-100	Dirichlet	0.061	**0.036**	0.056	0.464	**0.273**	0.445
Resnet-18	CIFAR-100	Pareto	0.019	**0.007**	0.052	0.370	**0.187**	1.036
Mobilenetv2	CIFAR-10	Dirichlet	0.103	**0.050**	0.109	0,285	**0.145**	0.337
Mobilenetv2	CIFAR-10	Pareto	0.034	**0.009**	0.062	0.249	**0.048**	0.556
Mobilenetv2	CIFAR-100	Dirichlet	0.003	**0.001**	0.006	0.187	**0.039**	0.060
Mobilenetv2	CIFAR-100	Pareto	0.008	**0.005**	0.060	0.227	**0.084**	1.579

GA consistently outperforms both AP and VW methods

Client grouping improving the performance of FL systems

Impact of the non-IID degree

TABLE II: Impact of the data distribution on the performance of proposed methods (ResNet-18, CIFAR-10 dataset, FedAvg)

Client Partition	β	RMSE			MAPE		
		AP	GA	VW	AP	GA	VW
Dirichlet	0.1	0.021	**0.014**	0.061	0.059	**0.055**	0.192
	0.3	0.046	**0.025**	0.122	0.179	**0.073**	0.464
	0.5	0.037	**0.014**	0.088	0.106	**0.032**	0.252
	1	**0.053**	0.065	0.142	**0.124**	0.181	0.447
	2	**0.030**	0.079	0.134	**0.126**	0.330	0.576
	3	**0.033**	0.053	0.152	**0.077**	0.153	0.475
Pareto	2	0.026	**0.011**	0.125	0.113	**0.049**	0.552
	3	0.014	**0.008**	0.032	0.044	**0.016**	0.102
	4	0.021	**0.017**	0.024	0.146	**0.110**	0.155
	5	0.017	**0.005**	0.122	0.054	**0.011**	0.364
	6	0.019	**0.005**	0.052	0.038	**0.011**	0.112

For the Pareto partition, GA consistently shows superior performance with the lowest RMSE and MAPE values in most cases.

Impact of the non-IID degree

TABLE II: Impact of the data distribution on the performance of proposed methods (ResNet-18, CIFAR-10 dataset, FedAvg)

Client Partition	β	RMSE			MAPE		
		AP	GA	VW	AP	GA	VW
Dirichlet	0.1	0.021	**0.014**	0.061	0.059	**0.055**	0.192
	0.3	0.046	**0.025**	0.122	0.179	**0.073**	0.464
	0.5	0.037	**0.014**	0.088	0.106	**0.032**	0.252
	1	**0.053**	0.065	0.142	**0.124**	0.181	0.447
	2	**0.030**	0.079	0.134	**0.126**	0.330	0.576
	3	**0.033**	0.053	0.152	**0.077**	0.153	0.475
Pareto	2	0.026	**0.011**	0.125	0.113	**0.049**	0.552
	3	0.014	**0.008**	0.032	0.044	**0.016**	0.102
	4	0.021	**0.017**	0.024	0.146	**0.110**	0.155
	5	0.017	**0.005**	0.122	0.054	**0.011**	0.364
	6	0.019	**0.005**	0.052	0.038	**0.011**	0.112

AP shows competitive performance and outperforms GA

GA outperforms both AP and VW methods at β = [0.1, 0.3, 0.5]

Impact of the non-IID degree

TABLE II: Impact of the data distribution on the performance of proposed methods (ResNet-18, CIFAR-10 dataset, FedAvg)

Client Partition	β	RMSE			MAPE		
		AP	GA	VW	AP	GA	VW
Dirichlet	0.1	0.021	**0.014**	0.061	0.059	**0.055**	0.192
	0.3	0.046	**0.025**	0.122	0.179	**0.073**	0.464
	0.5	0.037	**0.014**	0.088	0.106	**0.032**	0.252
	1	**0.053**	0.065	0.142	**0.124**	0.181	0.447
	2	**0.030**	0.079	0.134	**0.126**	0.330	0.576
	3	**0.033**	0.053	0.152	**0.077**	0.153	0.475
Pareto	2	0.026	**0.011**	0.125	0.113	**0.049**	0.552
	3	0.014	**0.008**	0.032	0.044	**0.016**	0.102
	4	0.021	**0.017**	0.024	0.146	**0.110**	0.155
	5	0.017	**0.005**	0.122	0.054	**0.011**	0.364
	6	0.019	**0.005**	0.052	0.038	**0.011**	0.112

When the data distribution becomes less skewed, grouping the data from two or more clients may result in group imbalance

Robustness to the FL algorithm

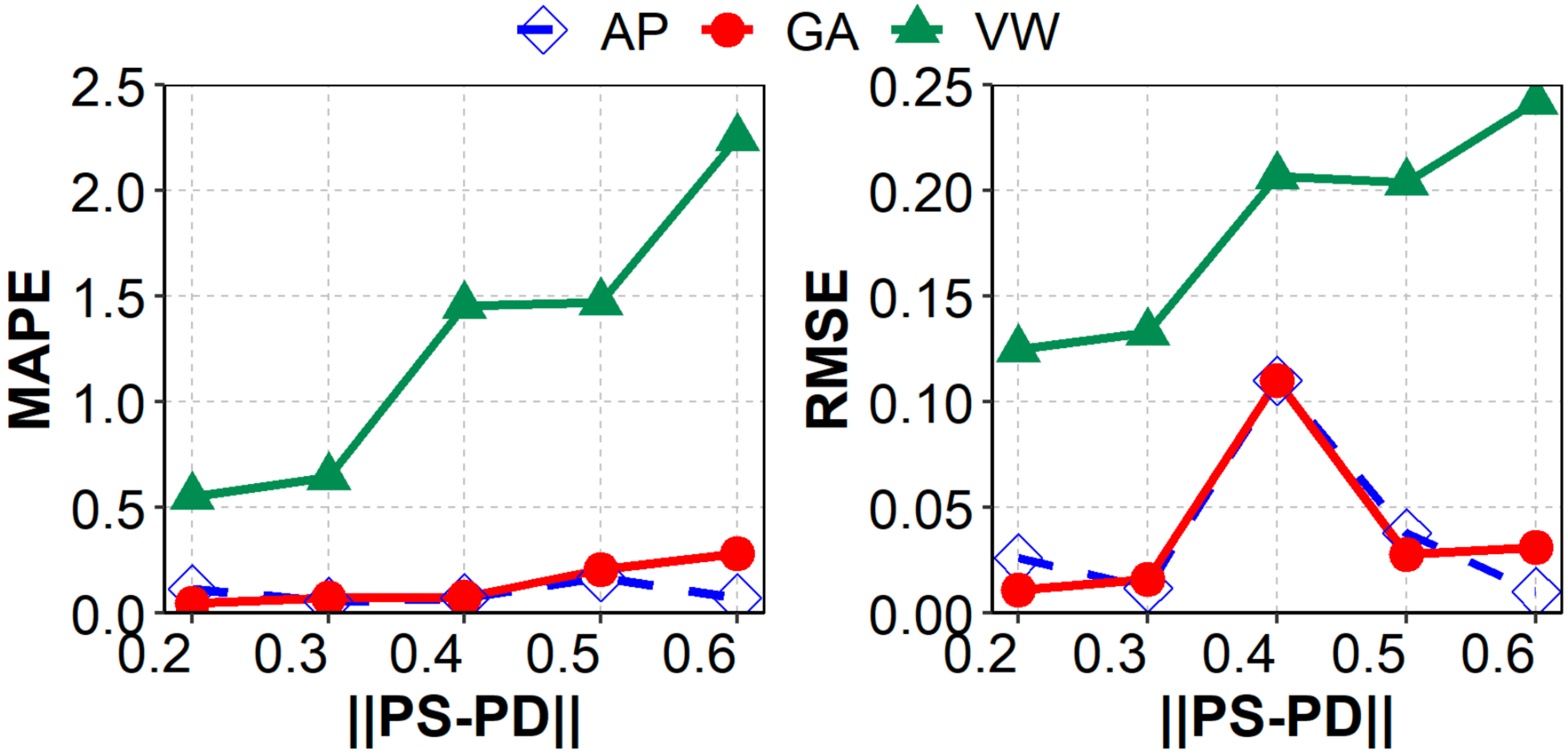
TABLE III: Robustness of the proposed methods to the FL algorithm (ResNet-18, CIFAR-10 dataset, Dirichlet, β = 0.1)

FL Algorithm	RMSE			MAPE		
	AP	GA	VW	AP	GA	VW
FedAvg [1]	0.021	**0.014**	0.061	0.059	**0.055**	0.192
FedProx [19]	0.121	**0.096**	0.128	0.500	**0.386**	0.528
Scaffold [20]	0.006	**0.005**	0.010	0.014	**0.013**	0.034

GA method consistently outperforms both the AP and VW methods across all metrics for all algorithms

Different desired data distributions

Figure 1. Performance under different desired data distributions (PS) and the test sample distributions of all clients (PD). (ResNet-18, CIFAR-10 dataset, Pareto, β = 2, FedAvg)



Conclusion

- **Propose a novel algorithm – FedCert**
  - Incorporating the client grouping algorithm
  - Leveraging certified accuracy principle
  - Offers a structured approach to enhance the robustness of FL models against adversarial perturbations

- **Results**
  - Significant improvements in accurately evaluating the robustness of the FL system on the CIFAR-10 and CIFAR-100 datasets

- **Future Work**
  - Further optimizing the algorithm and exploring its applicability to diverse datasets and FL scenarios.