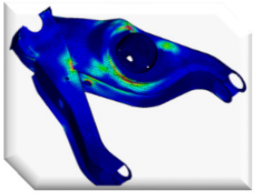
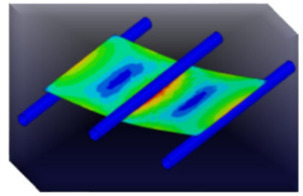
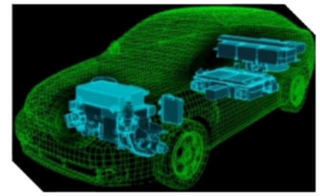



Uncertainty-Aware Federated Learning for Machine Health Monitoring

Chao Hu

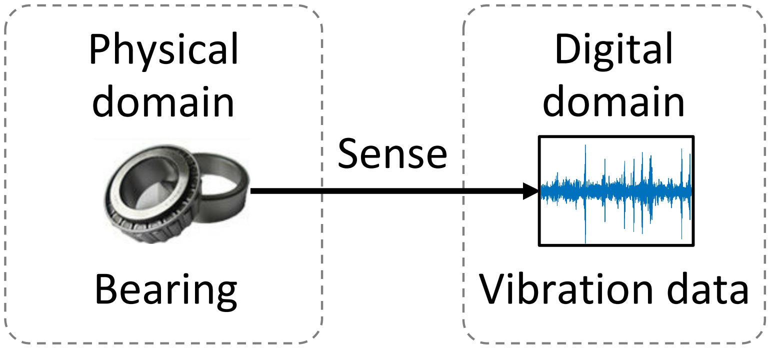
Associate Professor
Department of Mechanical Engineering
University of Connecticut

Collaborators: Olga Fink (EPFL), Hao Lu (UPC), Adam Thelen (ISU), Simon Laflamme (ISU)

UCONN REIL Reliability Engineering & Informatics Laboratory

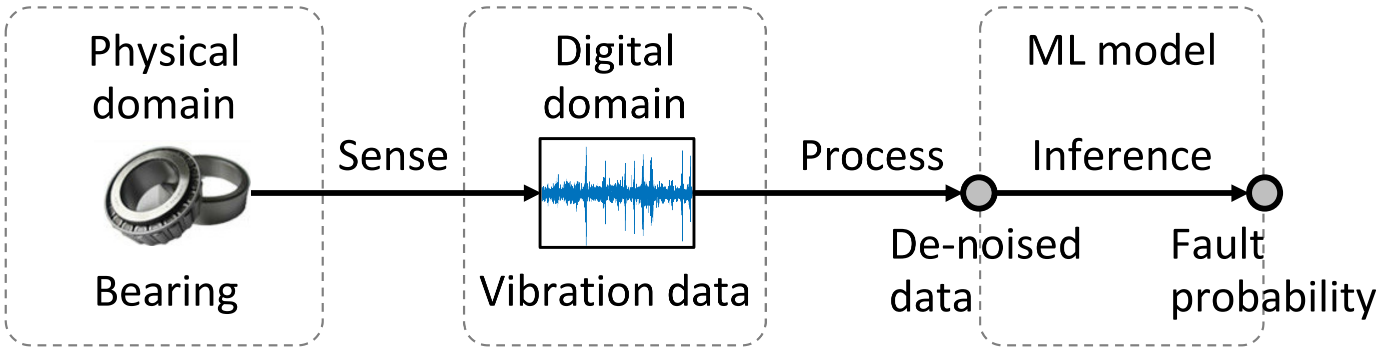
Problem Definition

ML Pipeline for Predictive Maintenance



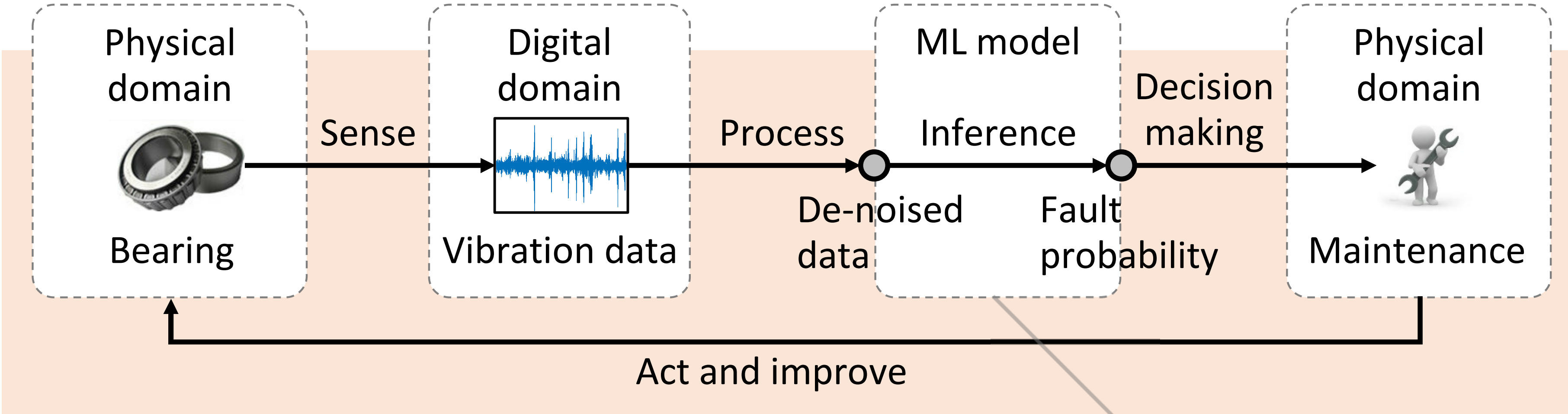
Case Study: Bearing Fault Classification

ML Pipeline for Predictive Maintenance



Case Study: Bearing Fault Classification

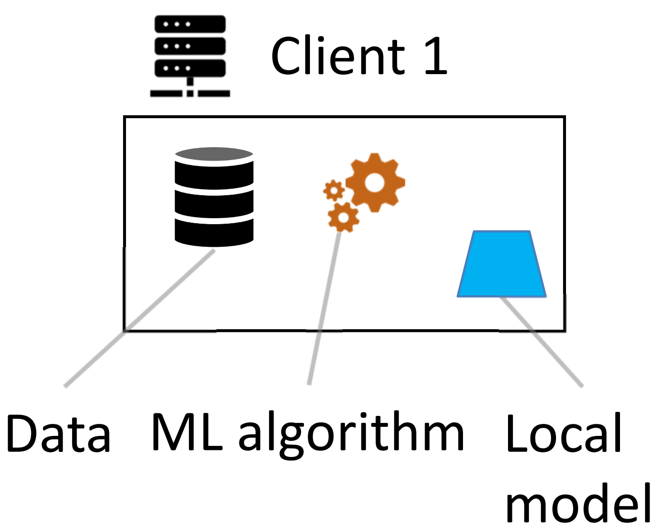
ML Pipeline for Predictive Maintenance



We focus on bearing fault classification.

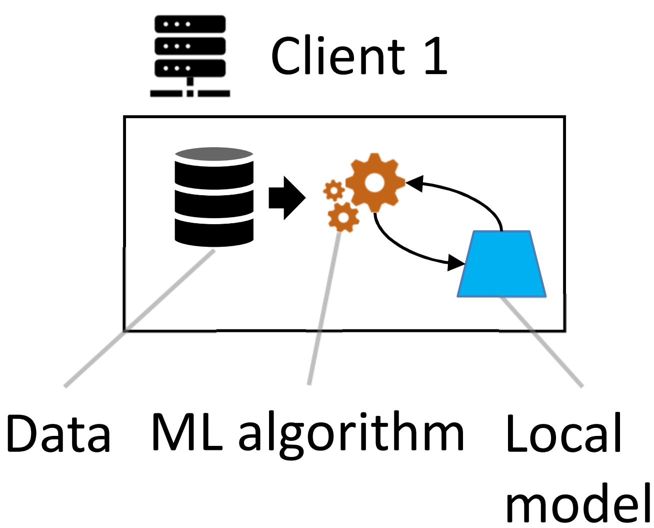
Distributed Learning

Looking at one local client



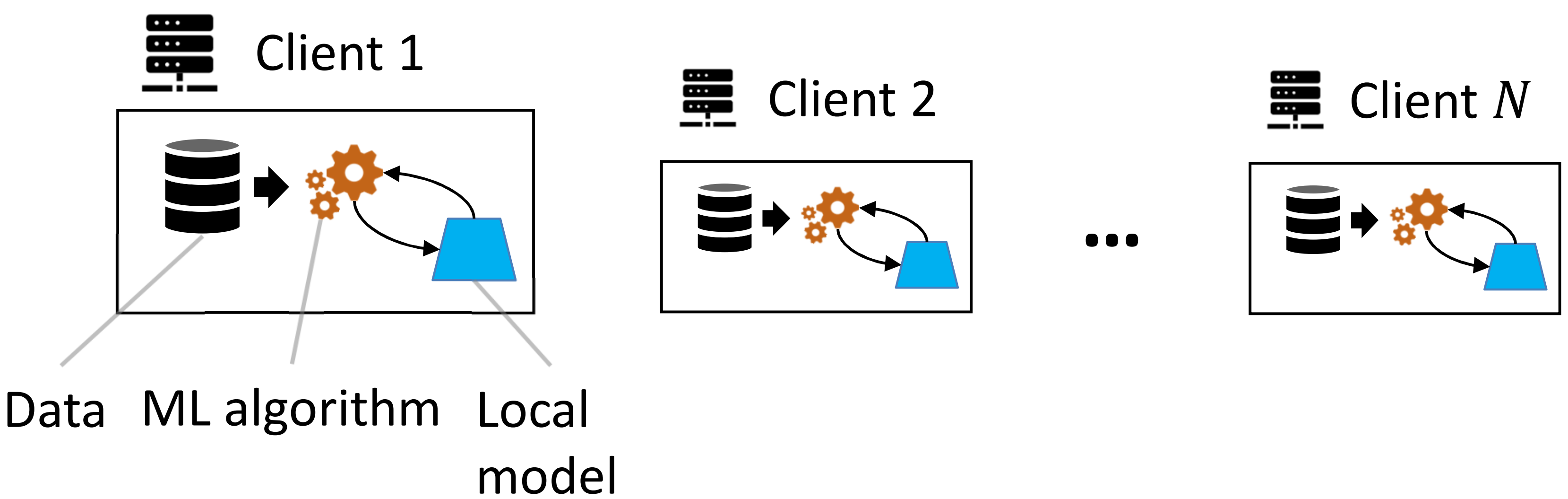
Distributed Learning

Model training locally at one client



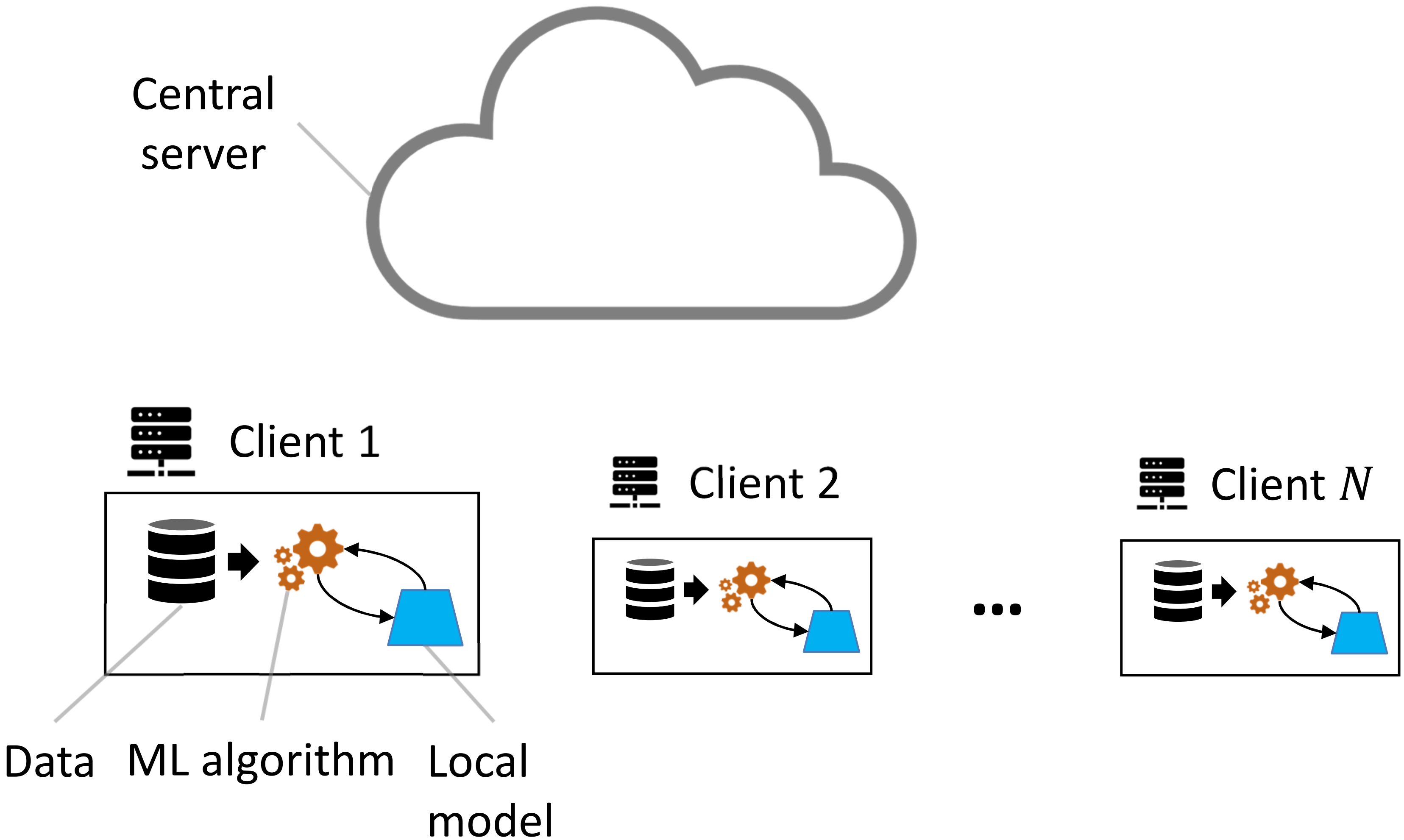
Distributed Learning

Looking at multiple local clients



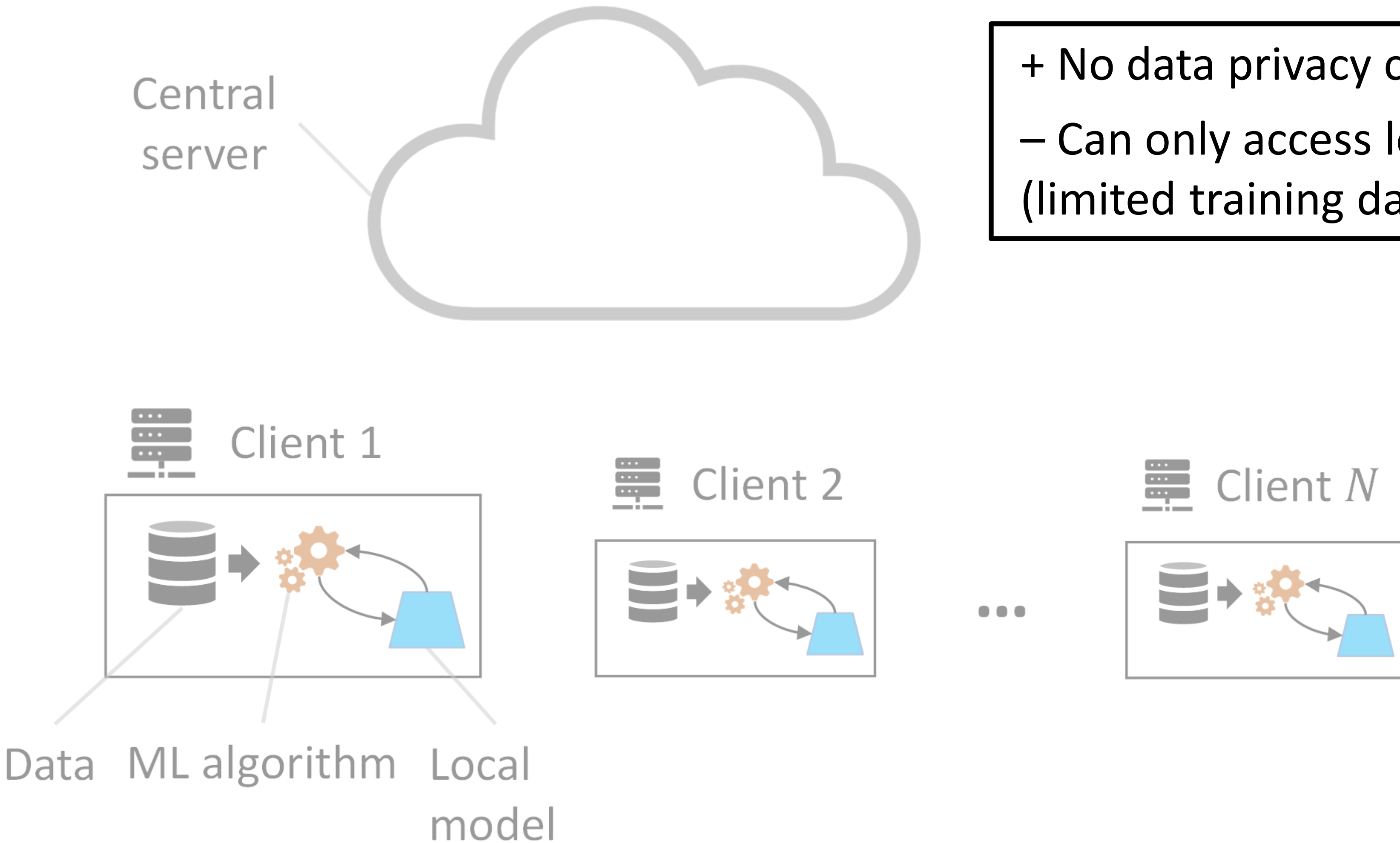
Distributed Learning

Learning with no data sharing



Distributed Learning

Learning with no data sharing

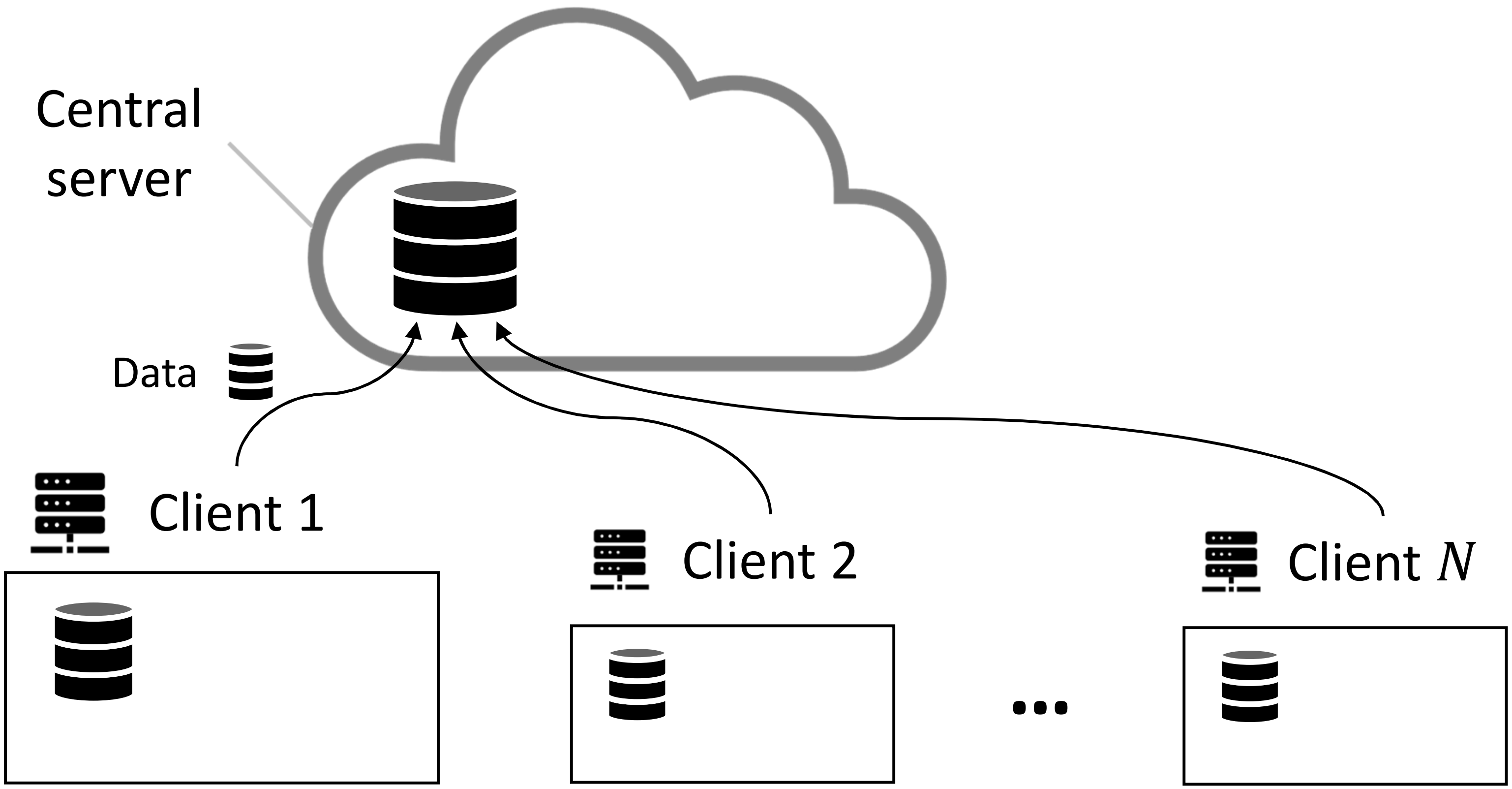


+ No data privacy concerns

– Can only access local data (limited training data)

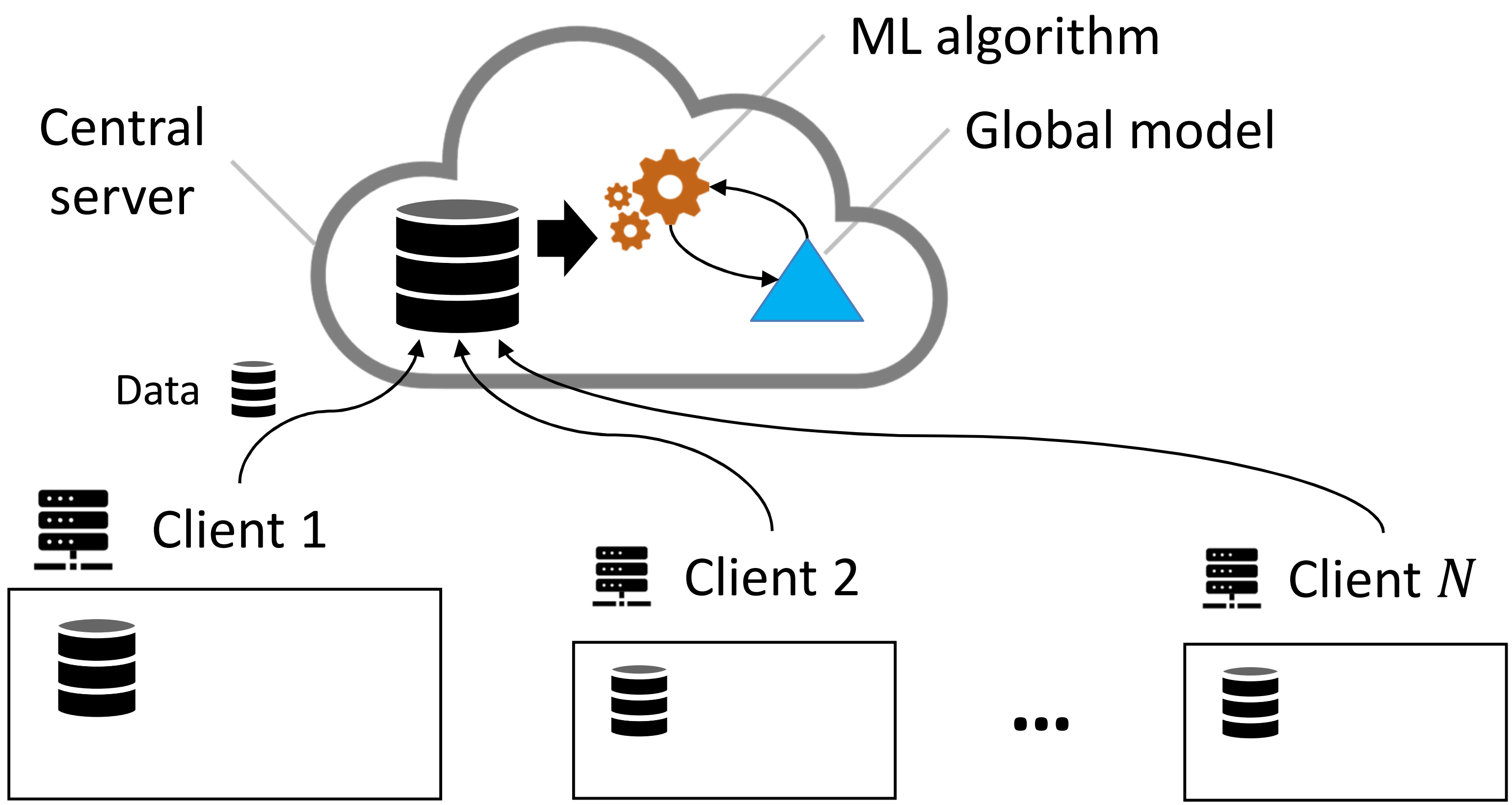
Centralized Learning

Sharing data with central server



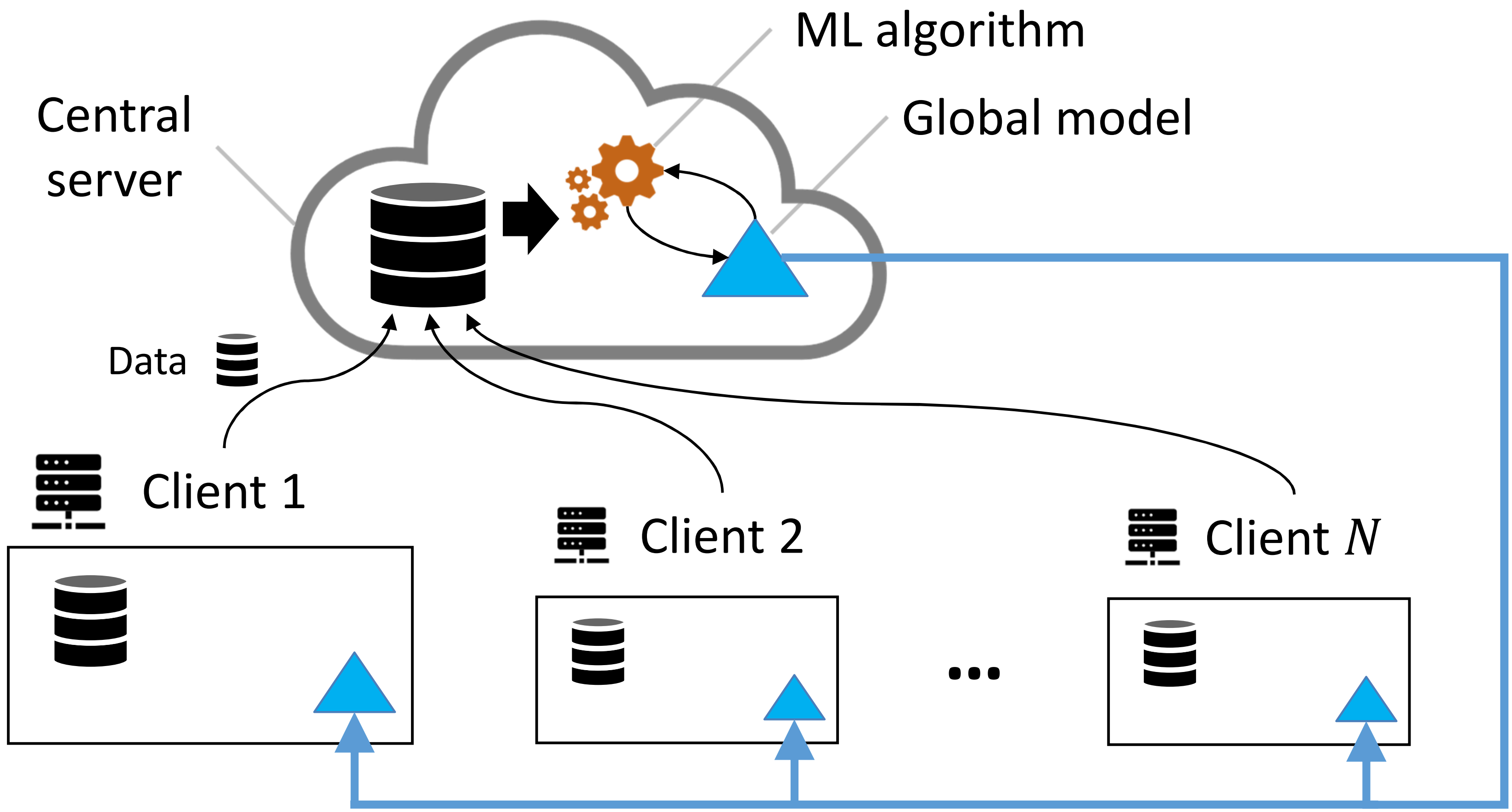
Centralized Learning

Sharing data with central server



Centralized Learning

Sharing data with central server

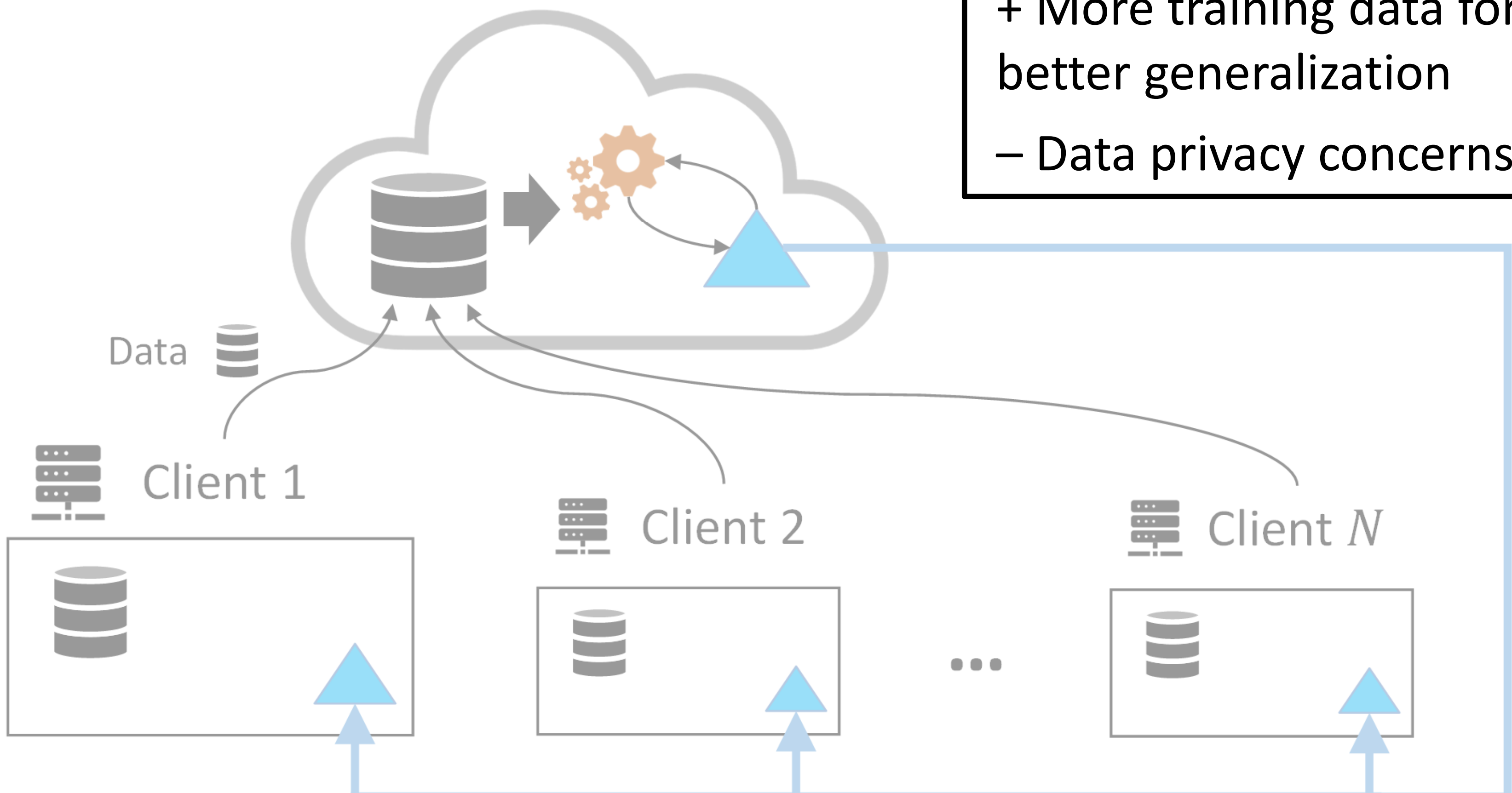


Centralized Learning

Sharing data with central server

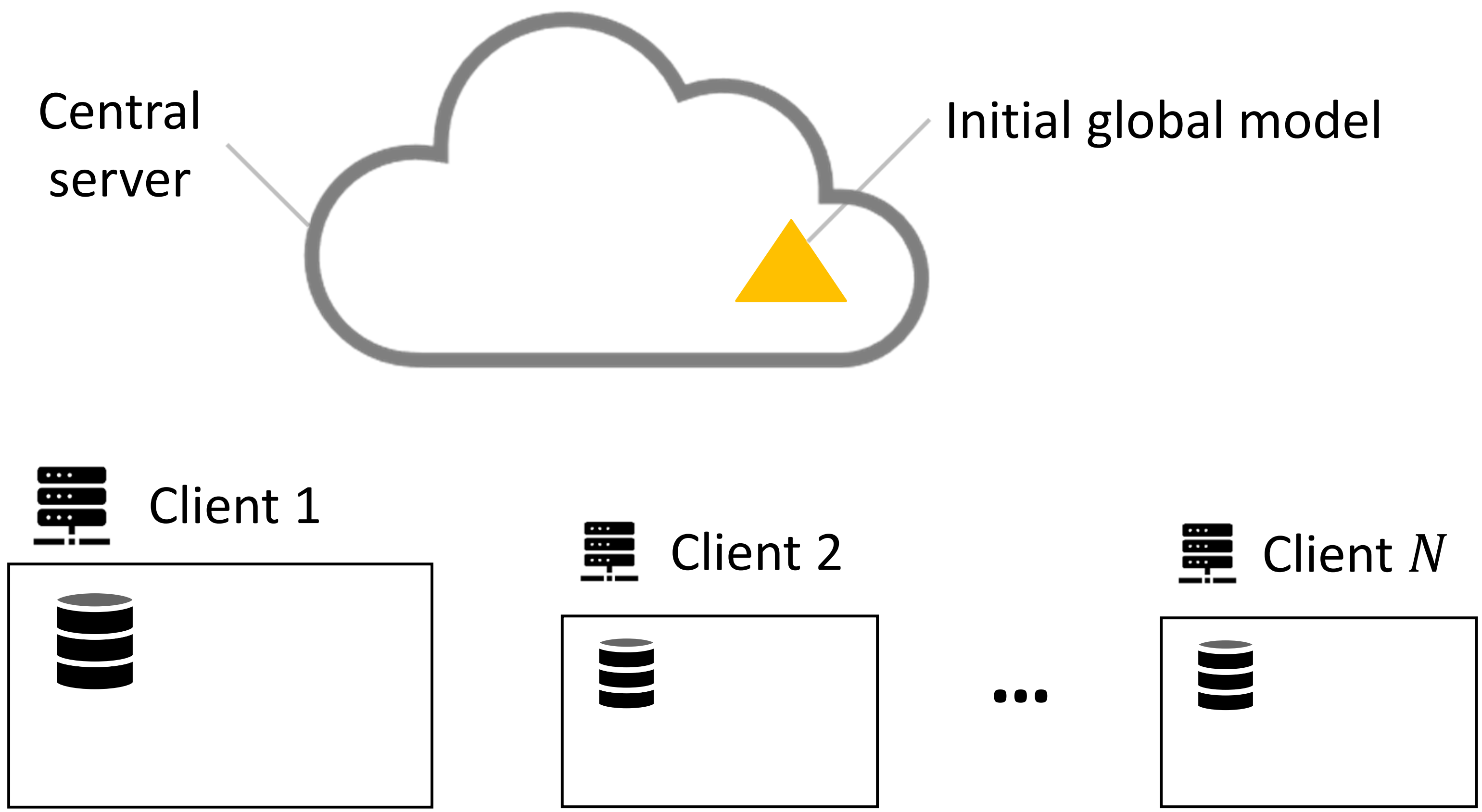
+ More training data for better generalization

– Data privacy concerns



Federated Learning

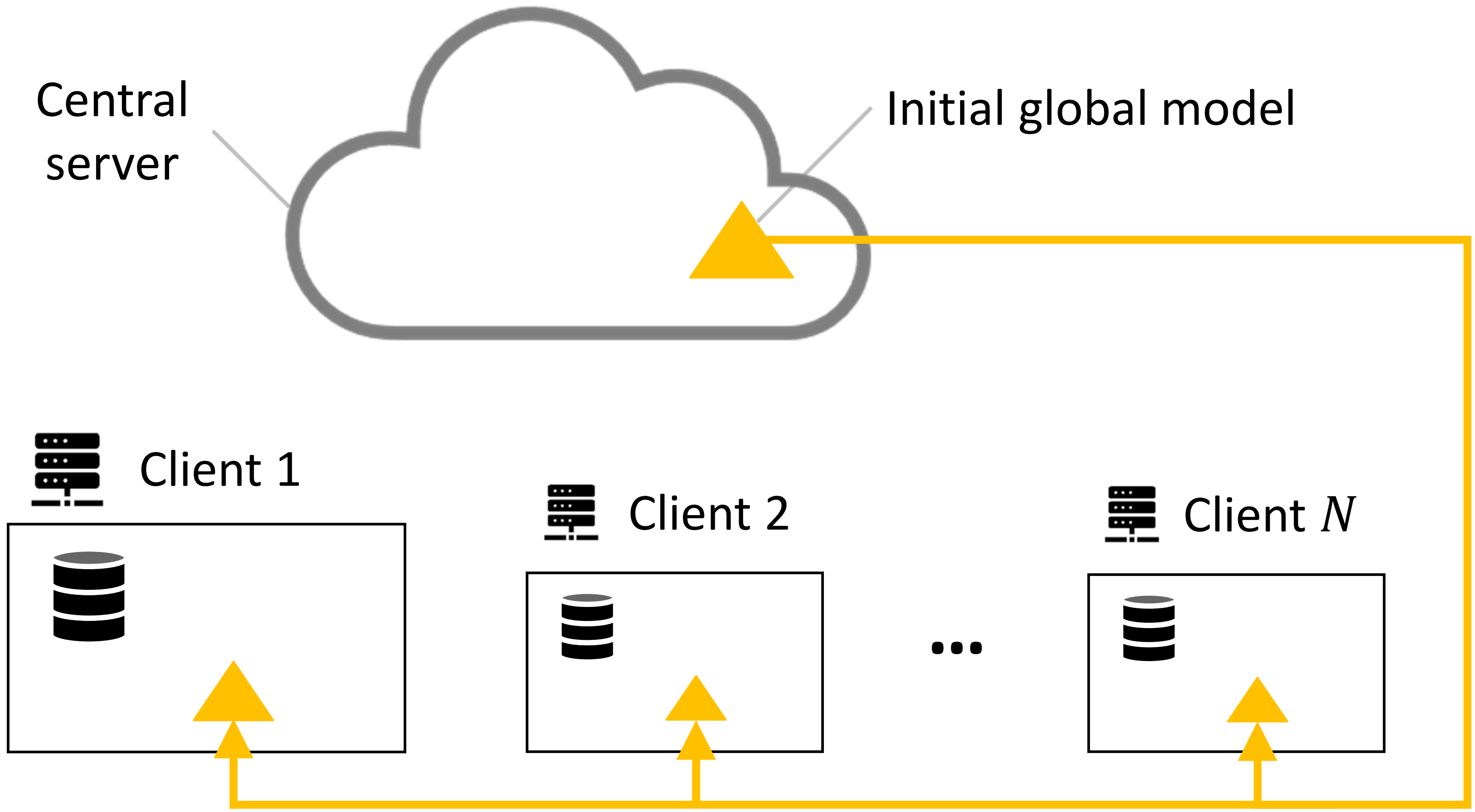
Step 1: Initialize a global model



McMahan, B., Moore, E., Ramage, D., Hampson, S. and y Arcas, B.A., 2017, April. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR.

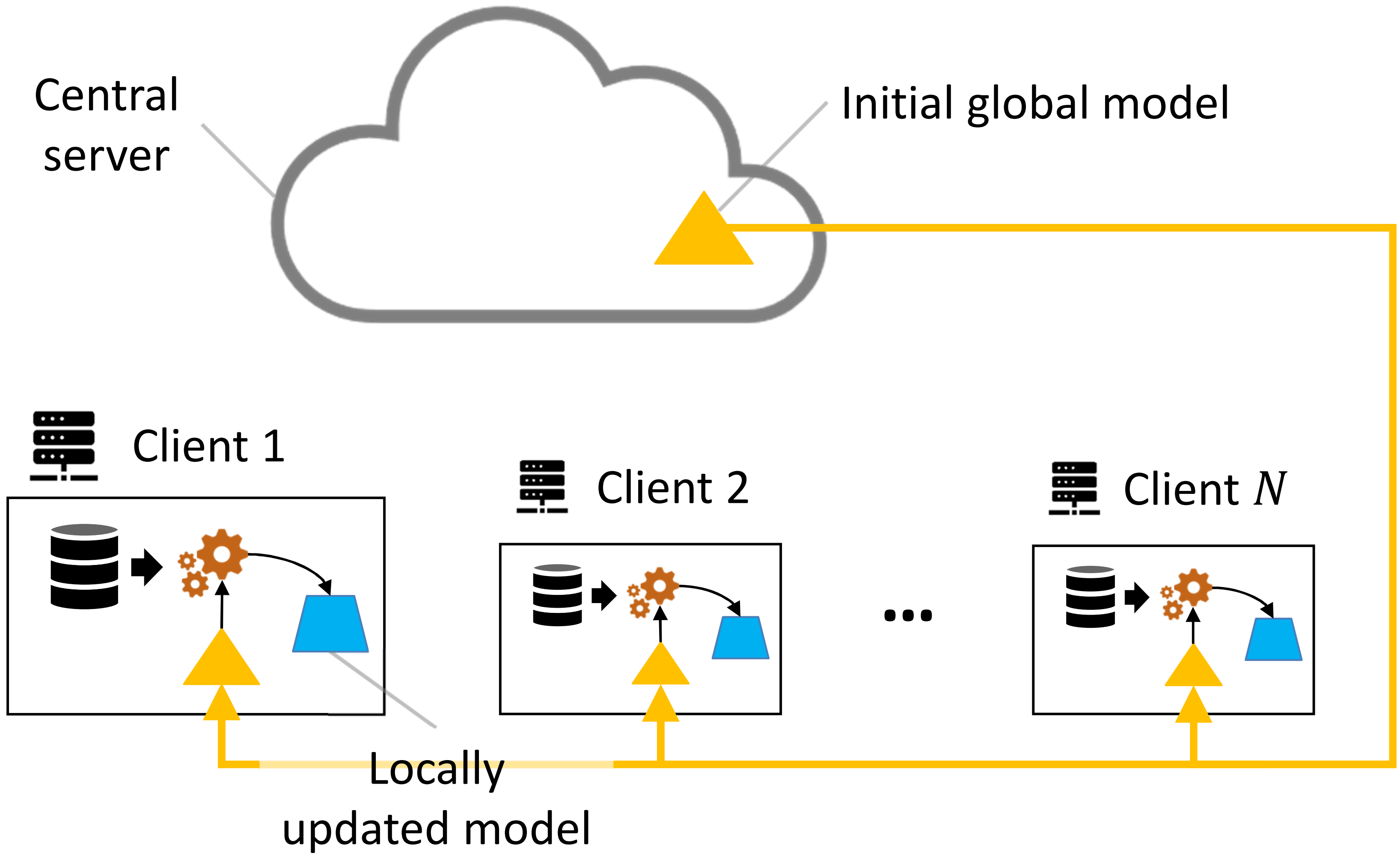
Federated Learning

Step 2: Update the global model locally



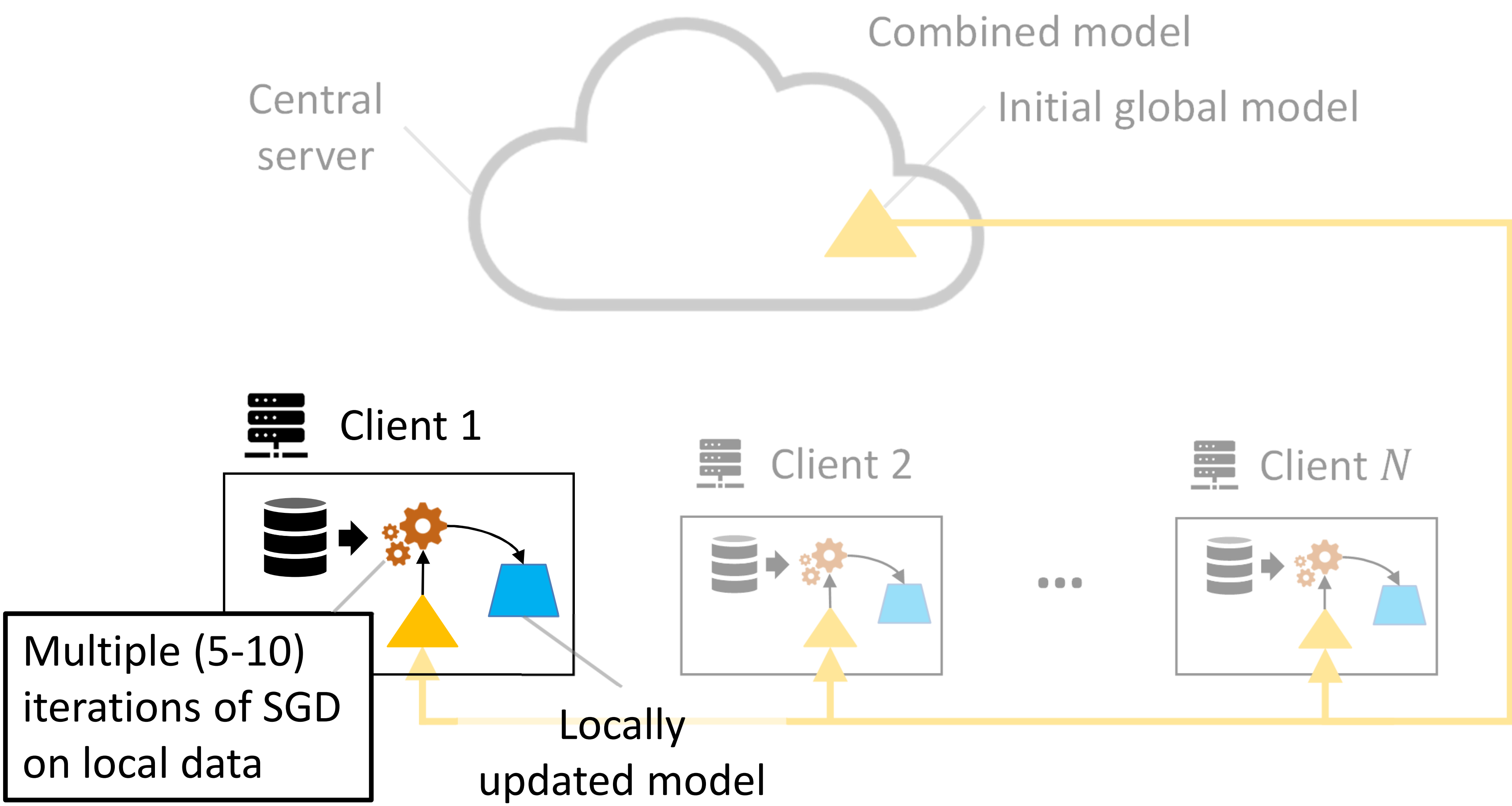
Federated Learning

Step 2: Update the global model locally



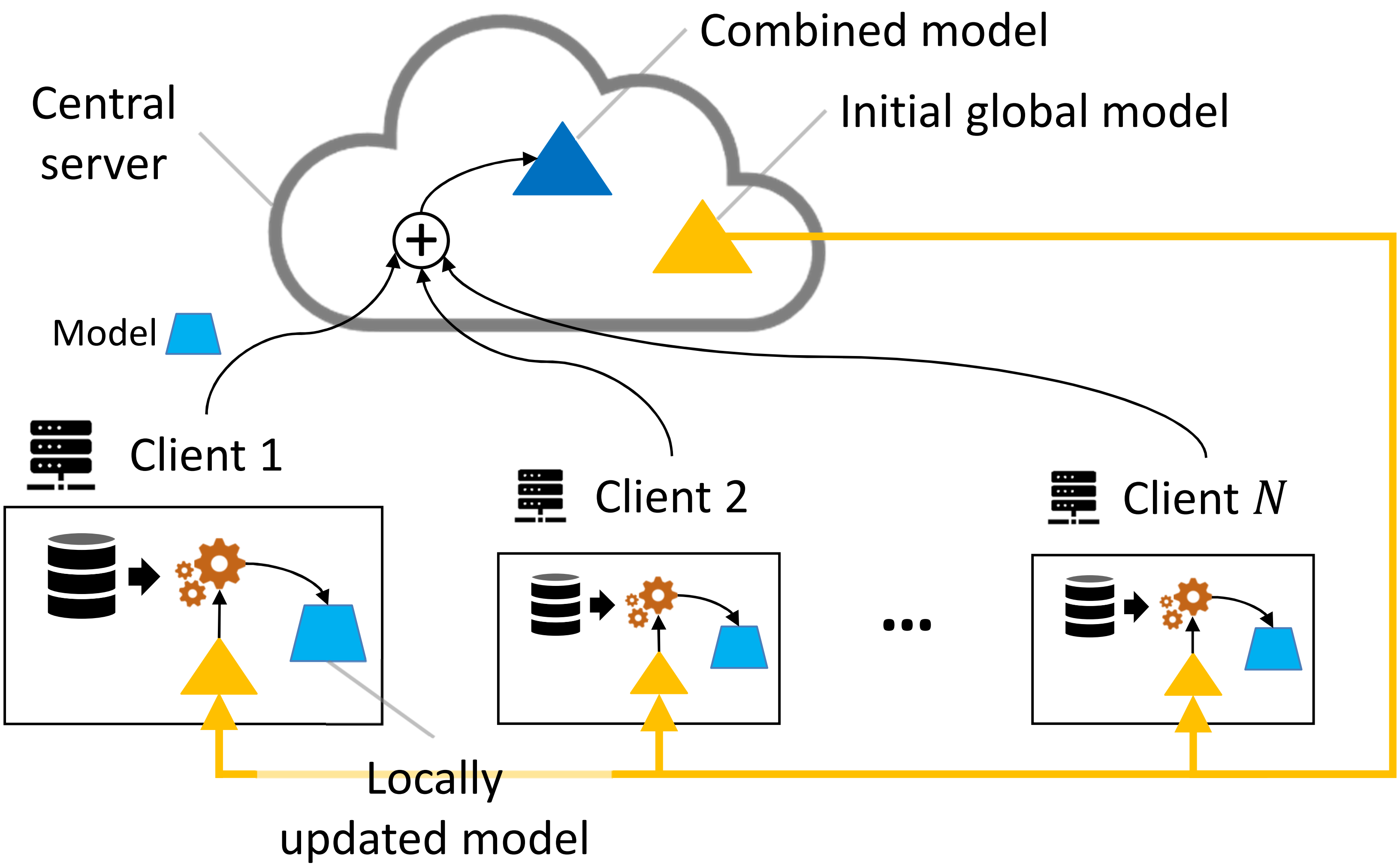
Federated Learning

Step 3: Aggregate local models $\{\theta_t^n\}_{n=1,\ldots,N}$



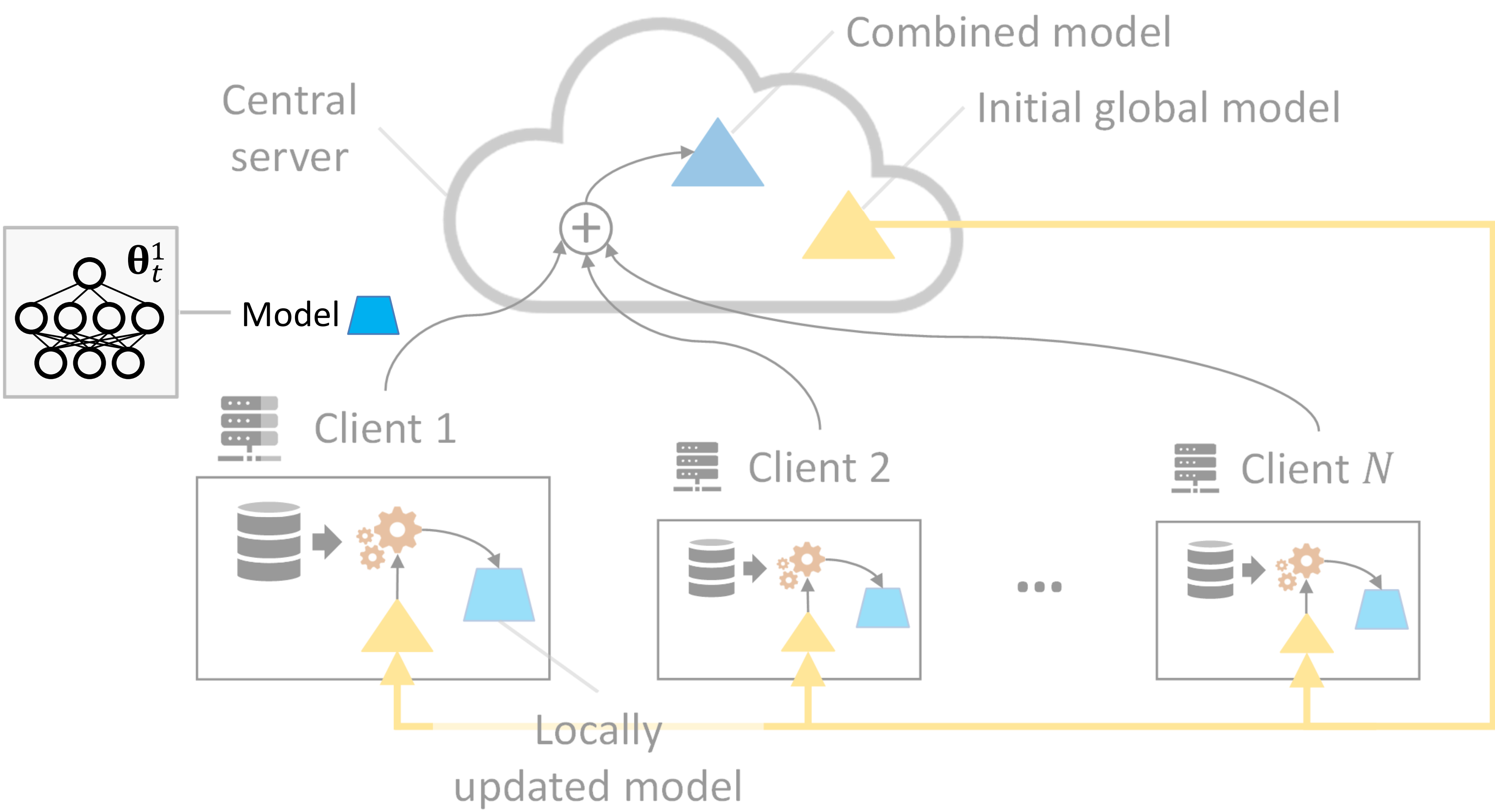
Federated Learning

Step 3: Aggregate local models $\{\boldsymbol{\theta}_t^n\}_{n=1,\dots,N}$



Federated Learning

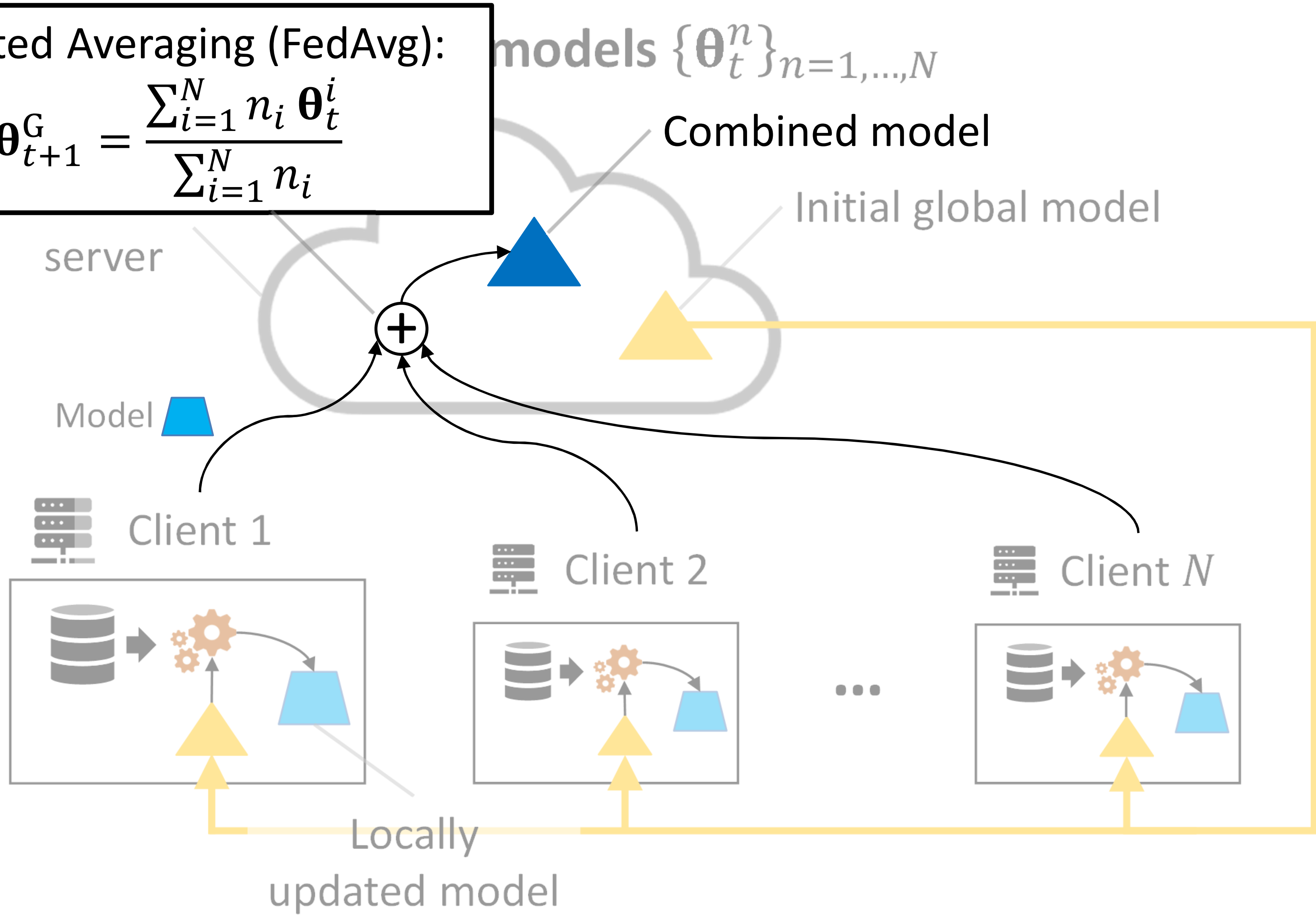
Step 3: Aggregate local models $\{\theta_t^n\}_{n=1,\ldots,N}$



Federated Learning

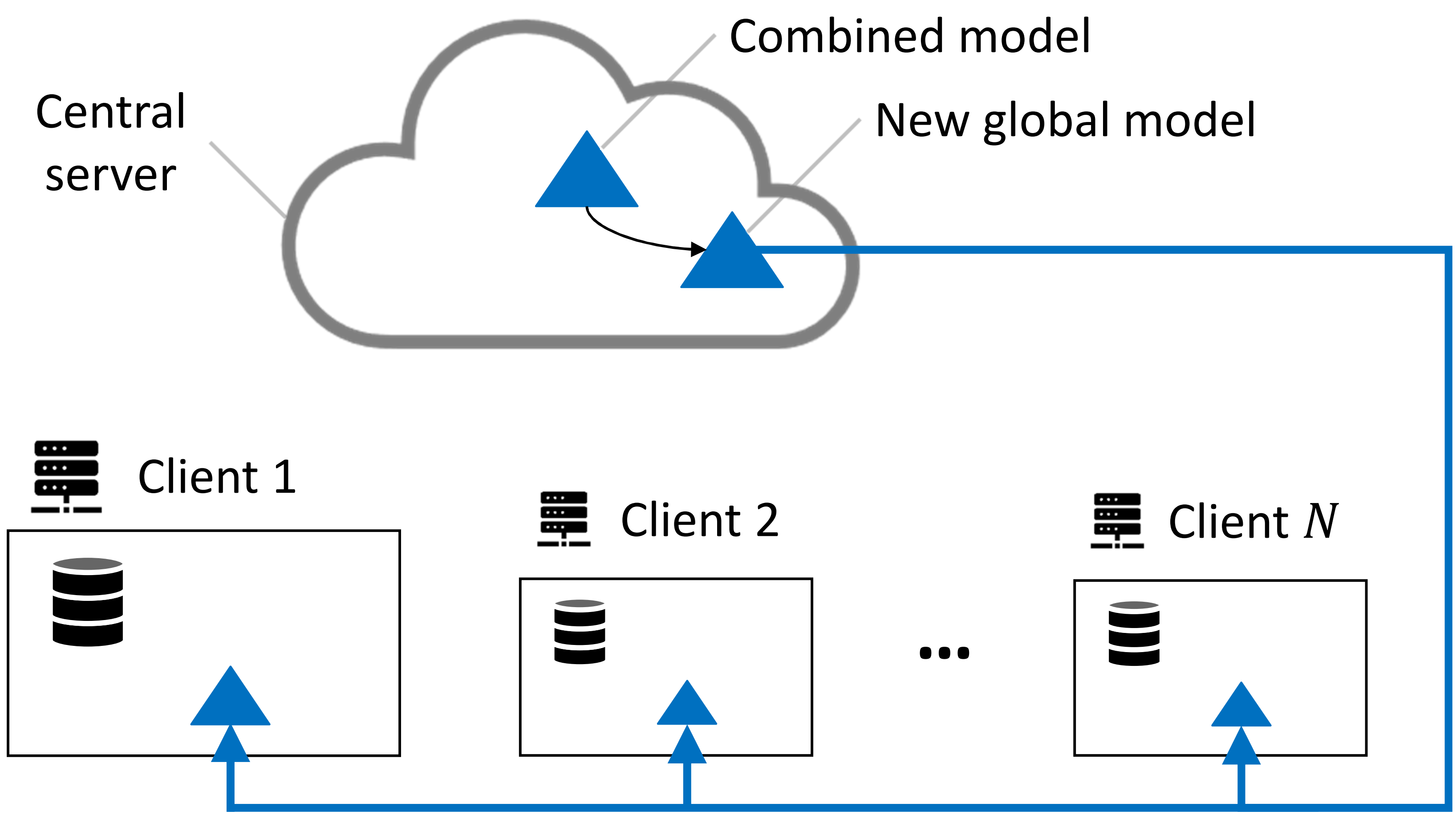
Federated Averaging (FedAvg):

$$\boldsymbol{\theta}_{t+1}^{\mathrm{G}} = \frac{\sum_{i=1}^{N} n_i \, \boldsymbol{\theta}_t^i}{\sum_{i=1}^{N} n_i}$$



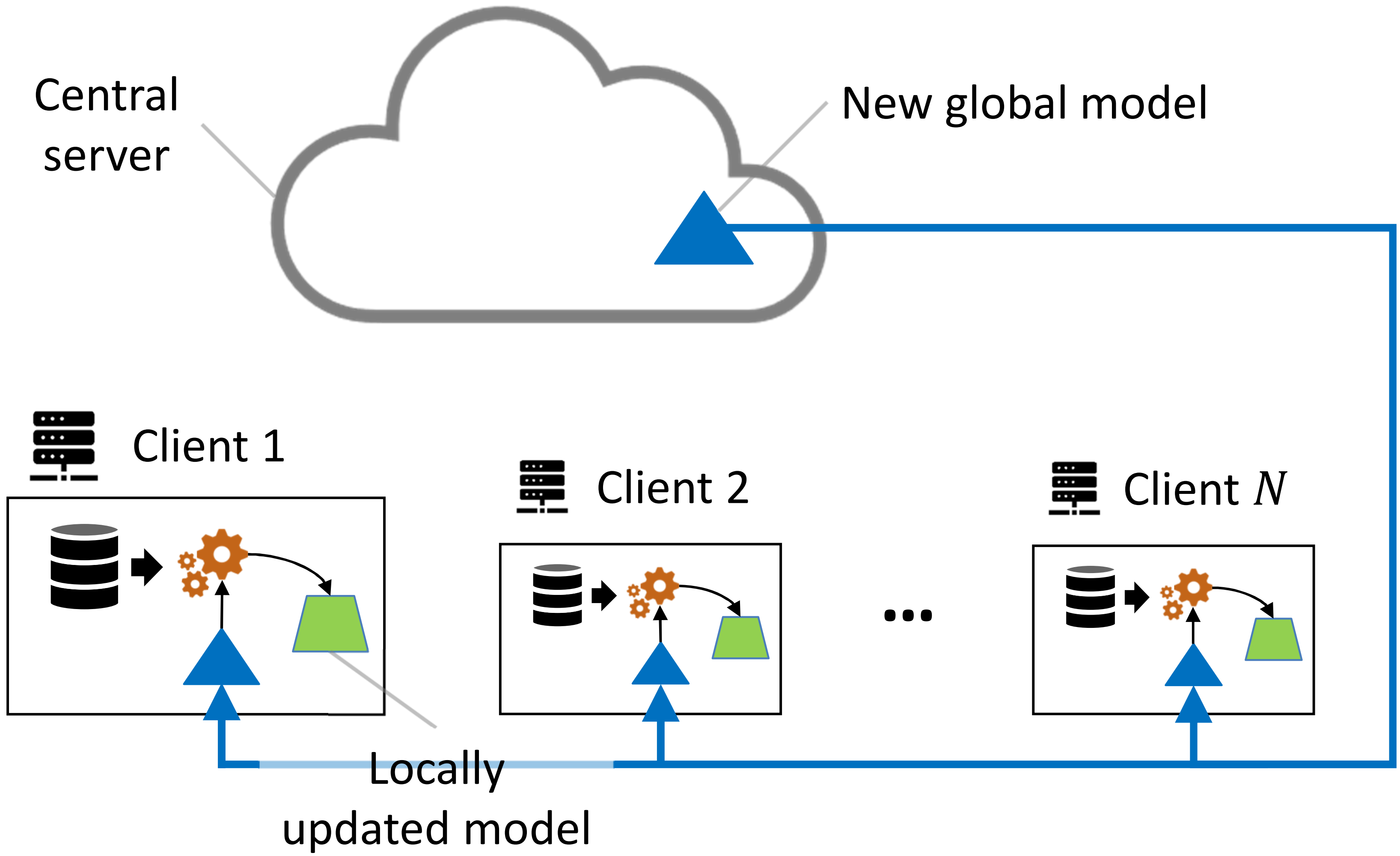
Federated Learning

Start next round (repeat Steps 1–3)



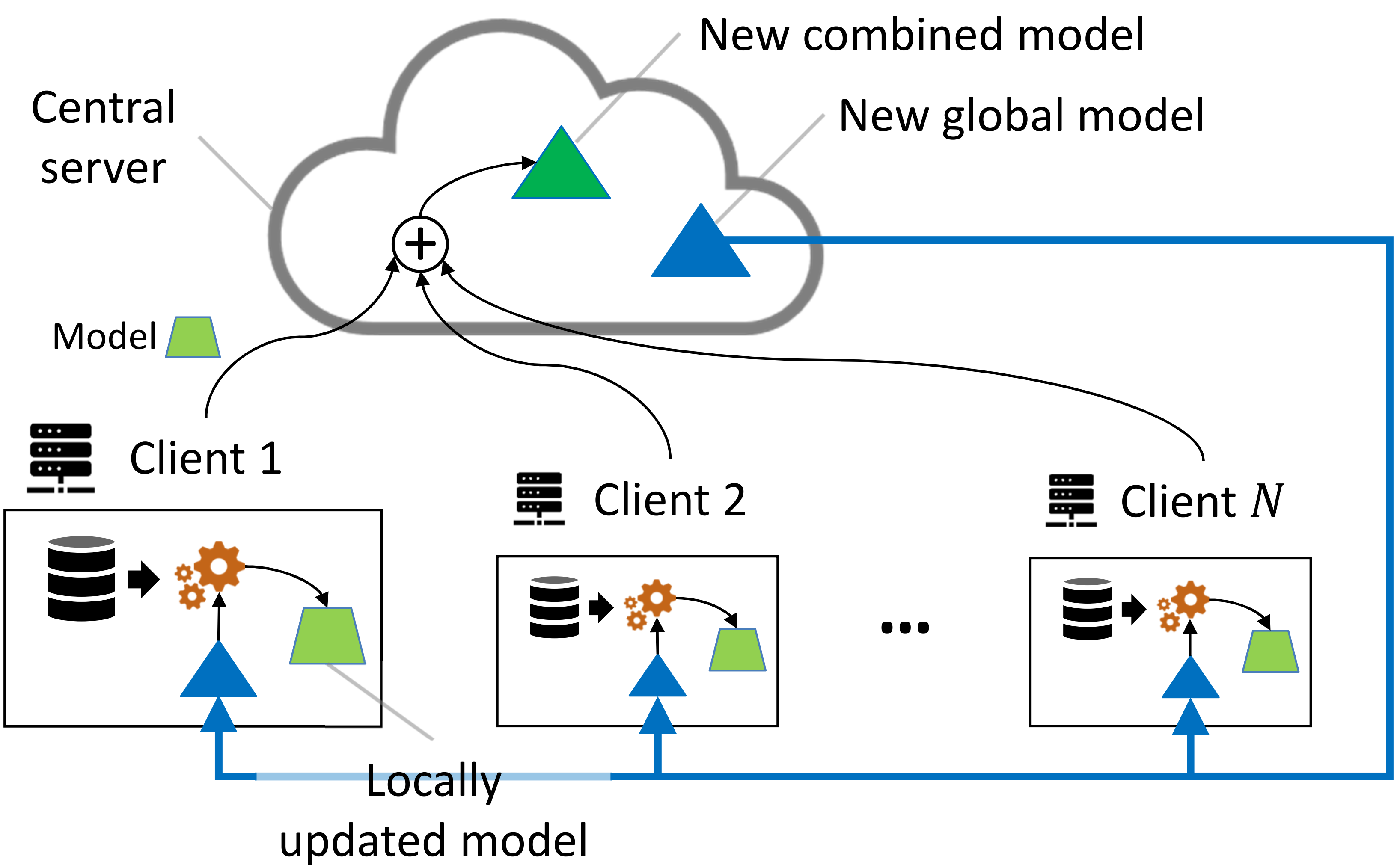
Federated Learning

Start next round (repeat Steps 1–3)



Federated Learning

Start next round (repeat Steps 1–3)



What Can Go Wrong?

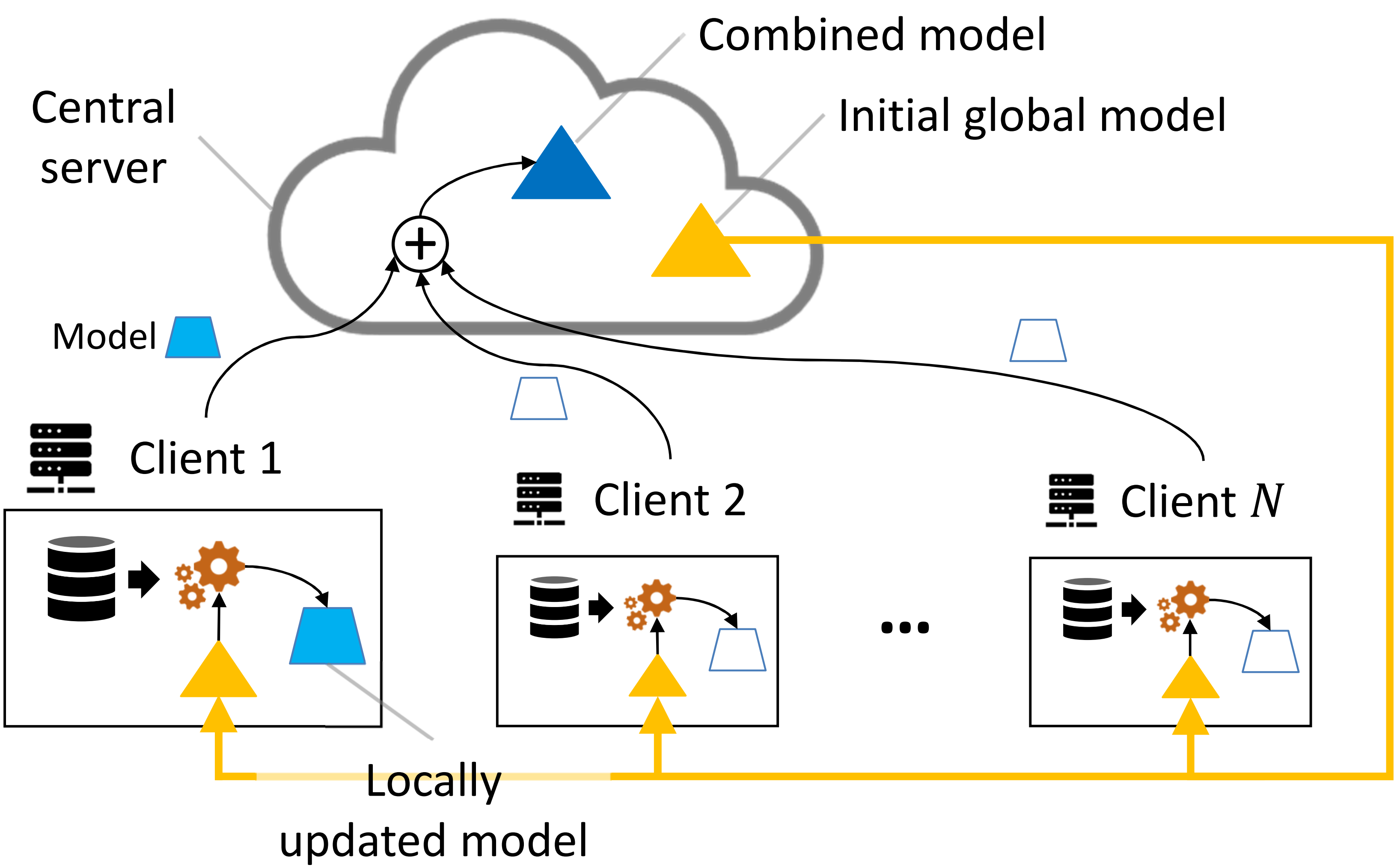
What Can Go Wrong?

Data heterogeneity in industrial settings:

1. *Domain shift:* Clients do not share a similar data distribution.
2. *Label heterogeneity:* Clients' datasets have different quantities and types of faults.

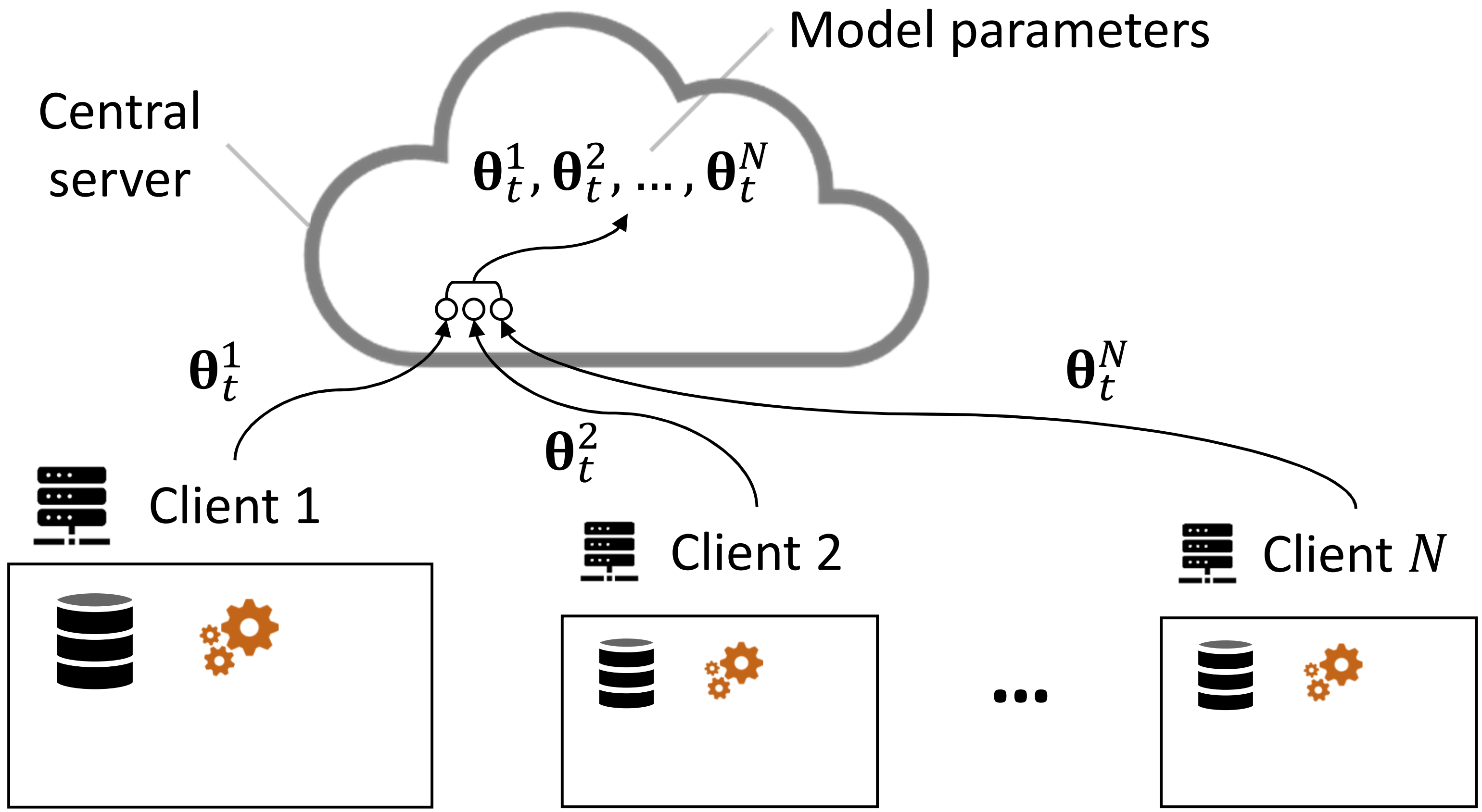
Can We Aggregate Models Selectively?

Step 3 in FedAvg: Aggregate local models $\{\boldsymbol{\theta}_t^n\}_{n=1,\dots,N}$



Can We Aggregate Models Selectively?

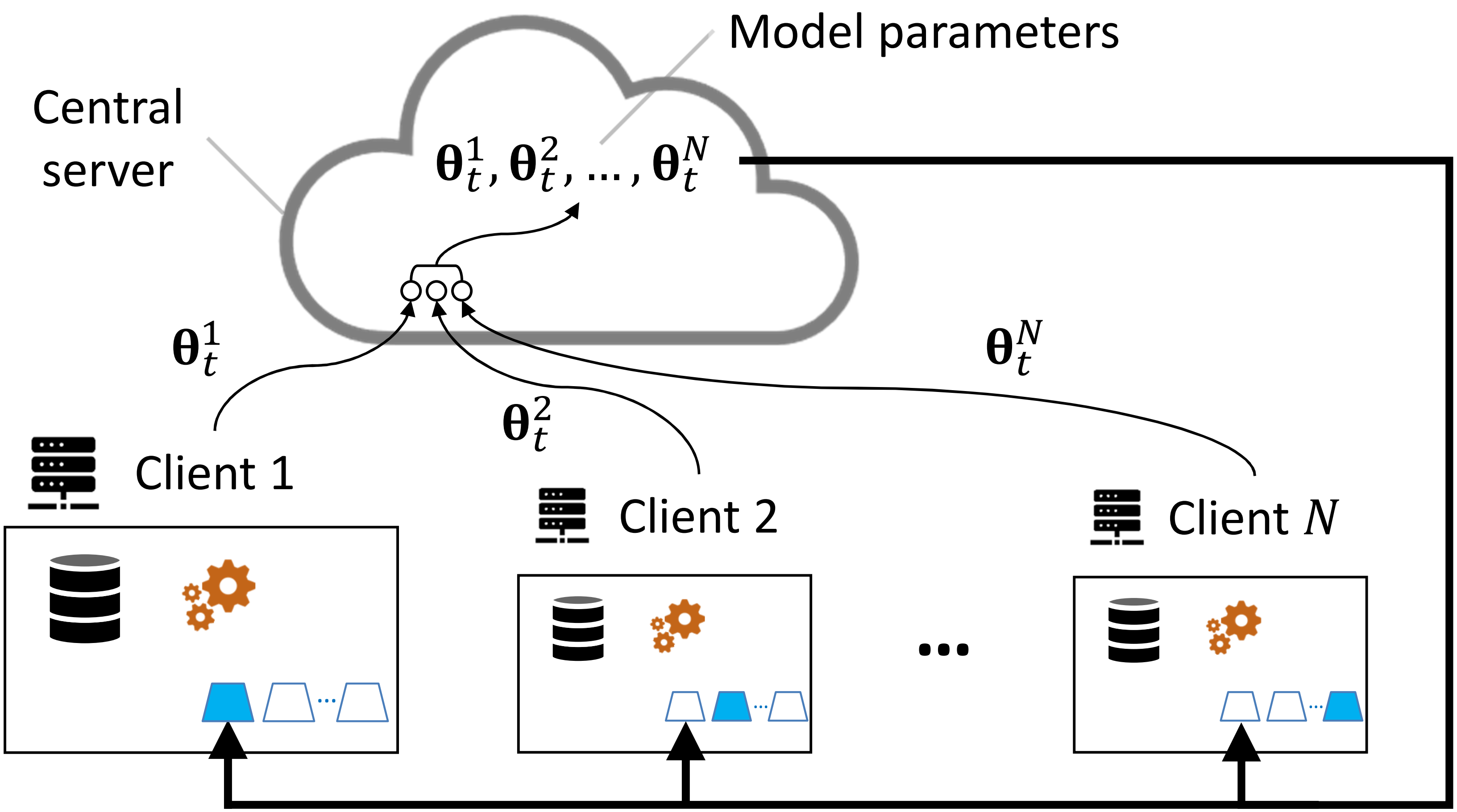
Step 3.1: Collect parameters of all models



Lu, H., Thelen, A., Fink, O., Hu, C. and Laflamme, S., 2023. Federated Learning with Uncertainty-Based Client Clustering for Fleet-Wide Fault Diagnosis. *arXiv preprint arXiv:2304.13275*.

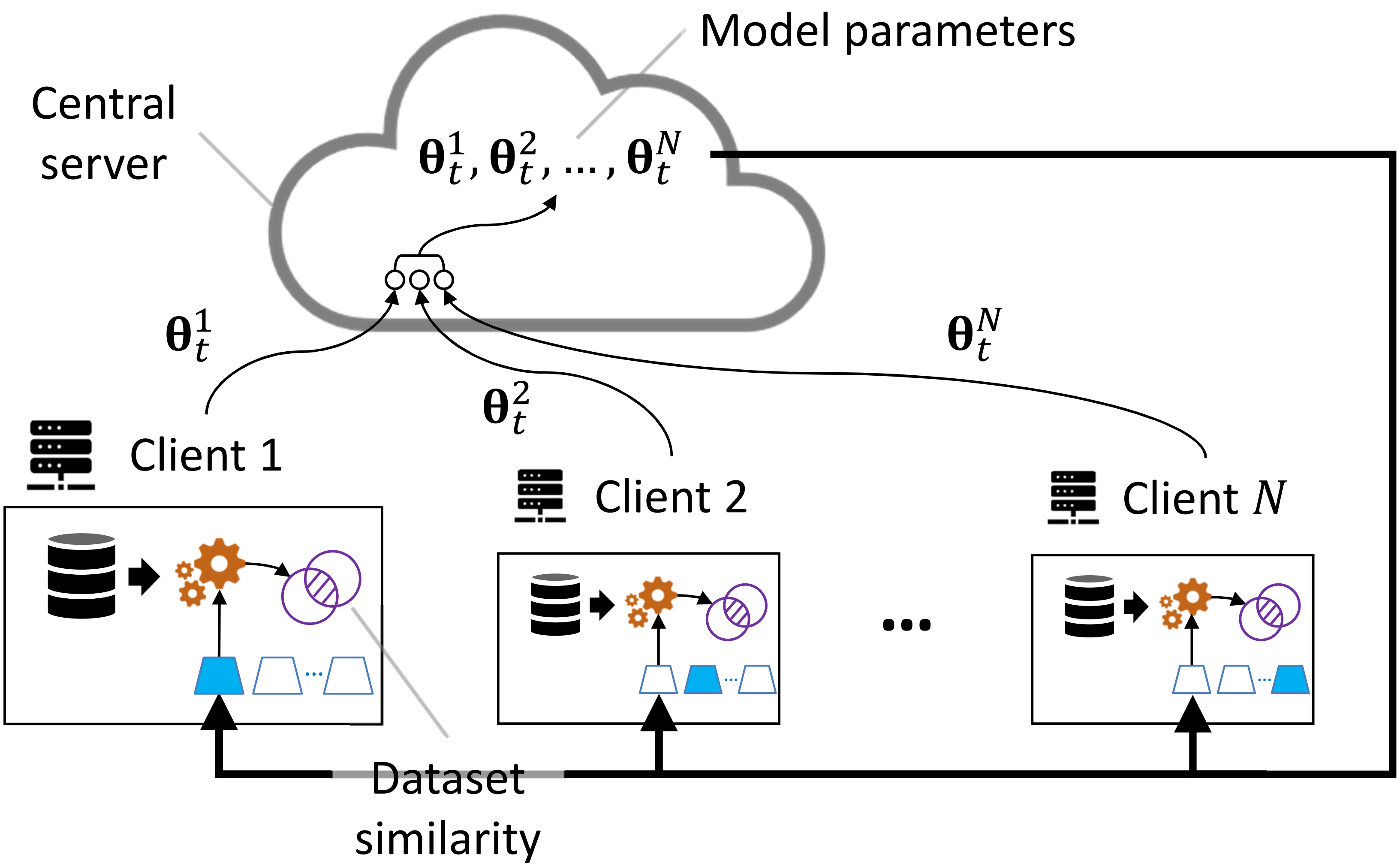
Can We Aggregate Models Selectively?

Step 3.2: Share models with clients



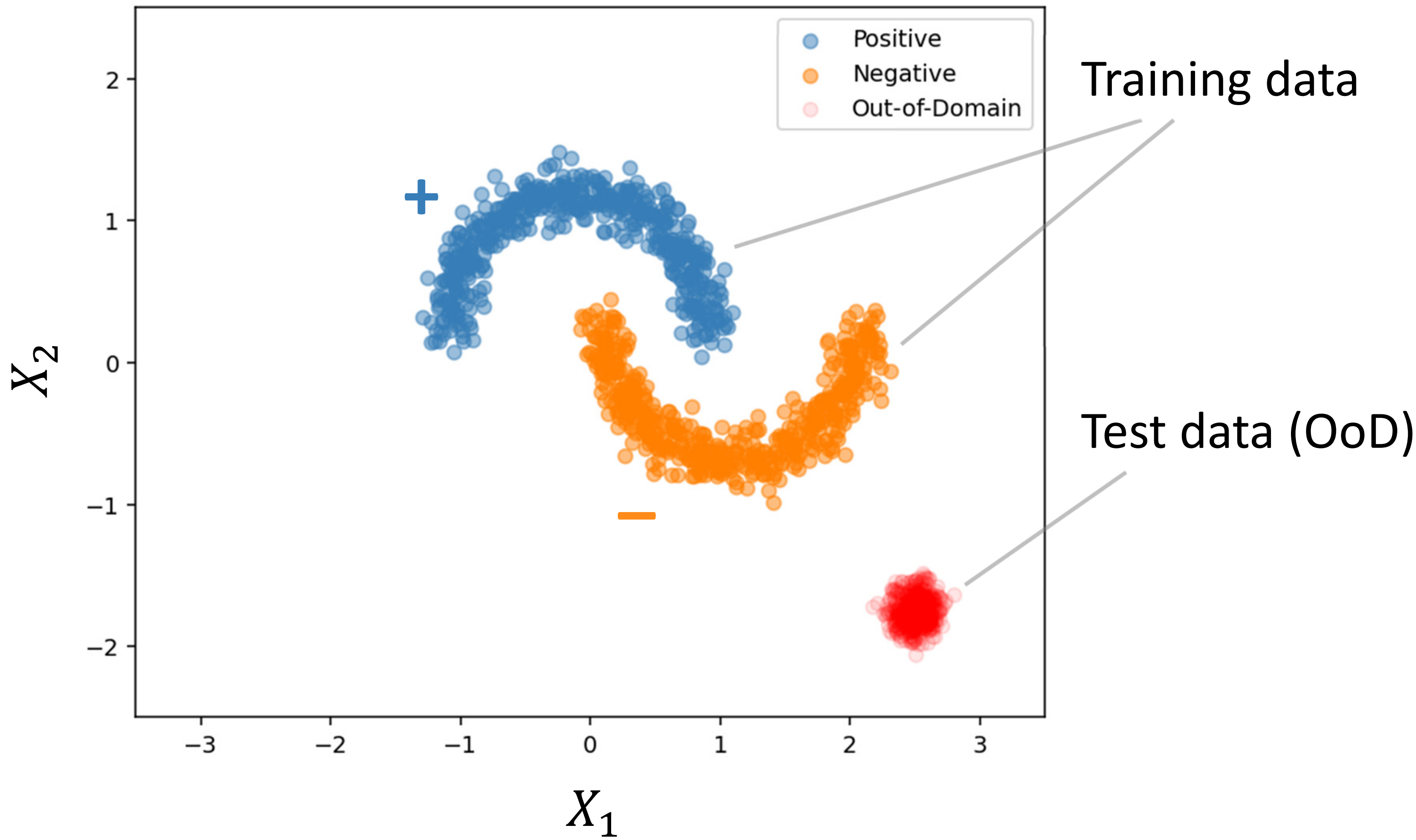
Can We Aggregate Models Selectively?

Step 3.3: Compute dataset similarity



How to Estimate Dataset Similarity?

Looking at a 2D binary classification problem



How to Estimate Dataset Similarity?

Quantify predictive uncertainty

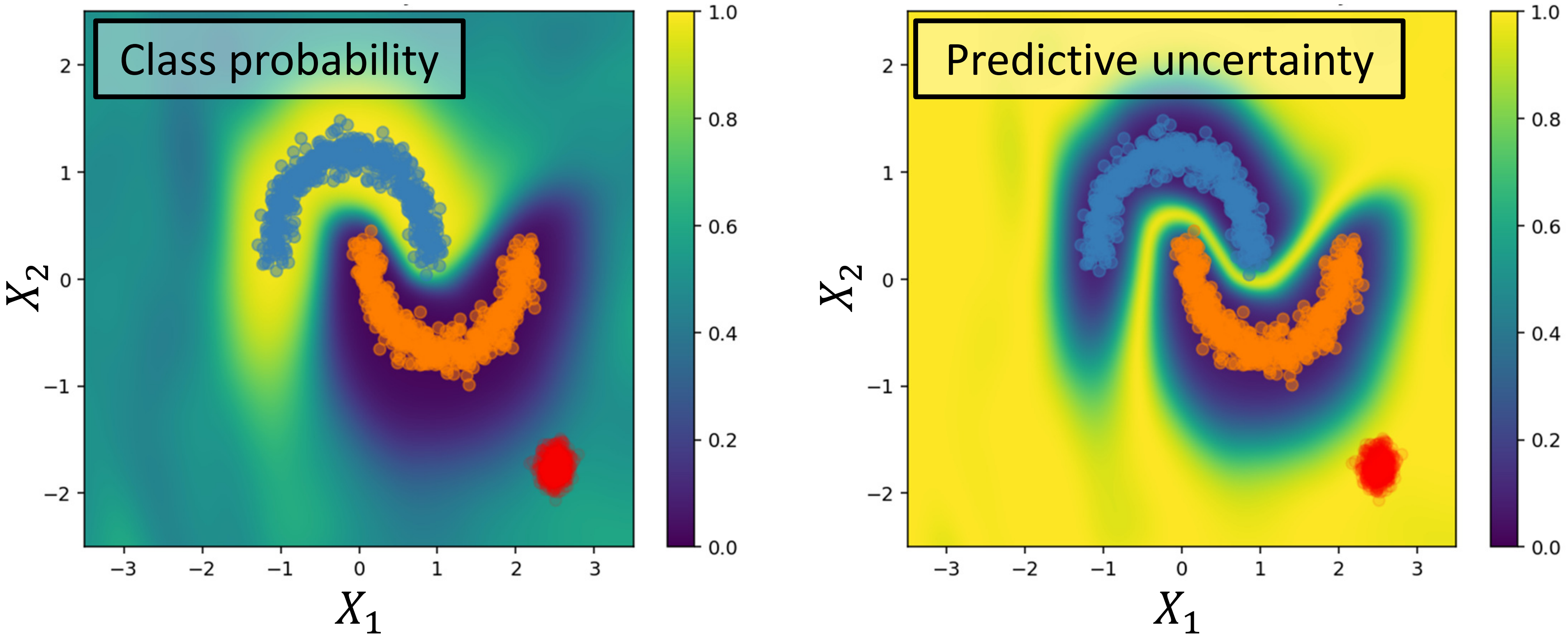
Spectral-normalized neural Gaussian process (SNGP) for distance-aware UQ

Liu, J., Lin, Z., Padhy, S., Tran, D., Bedrax Weiss, T. and Lakshminarayanan, B., 2020. Simple and principled uncertainty estimation with deterministic deep learning via distance awareness. *NIPS*, *33*, pp.7498-7512.

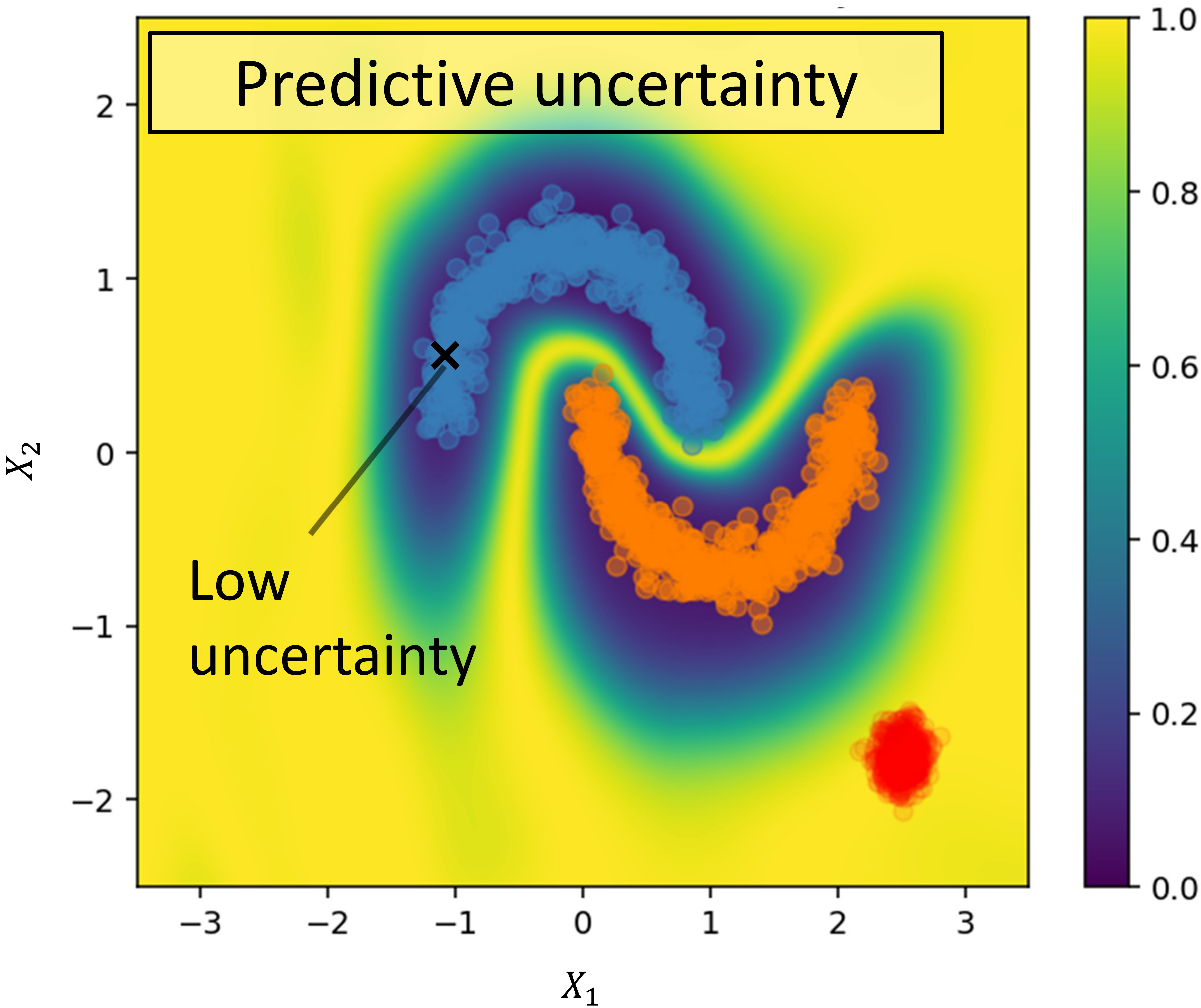
How to Estimate Dataset Similarity?

Quantify predictive uncertainty

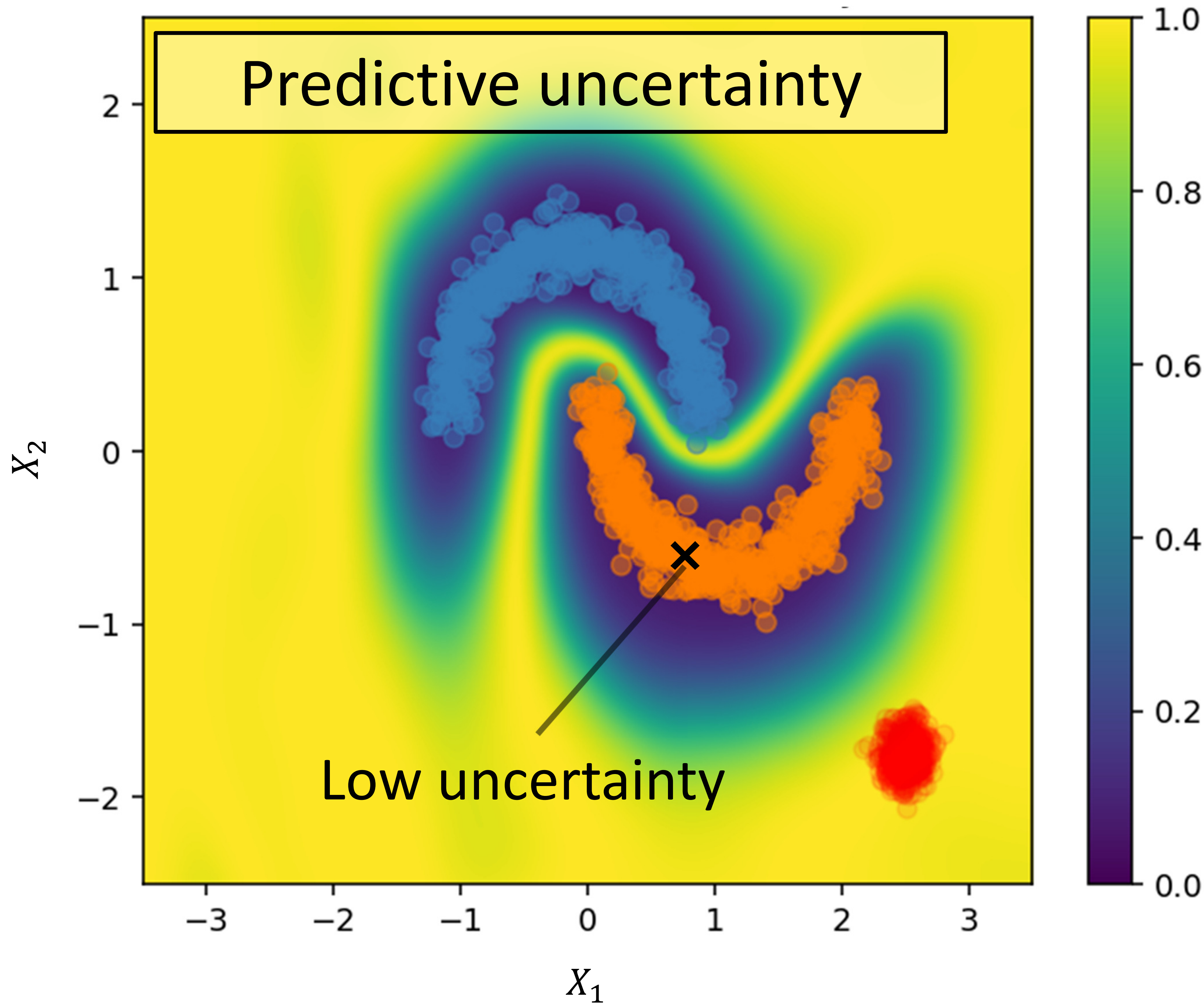
Spectral-normalized neural Gaussian process (SNGP) for distance-aware UQ

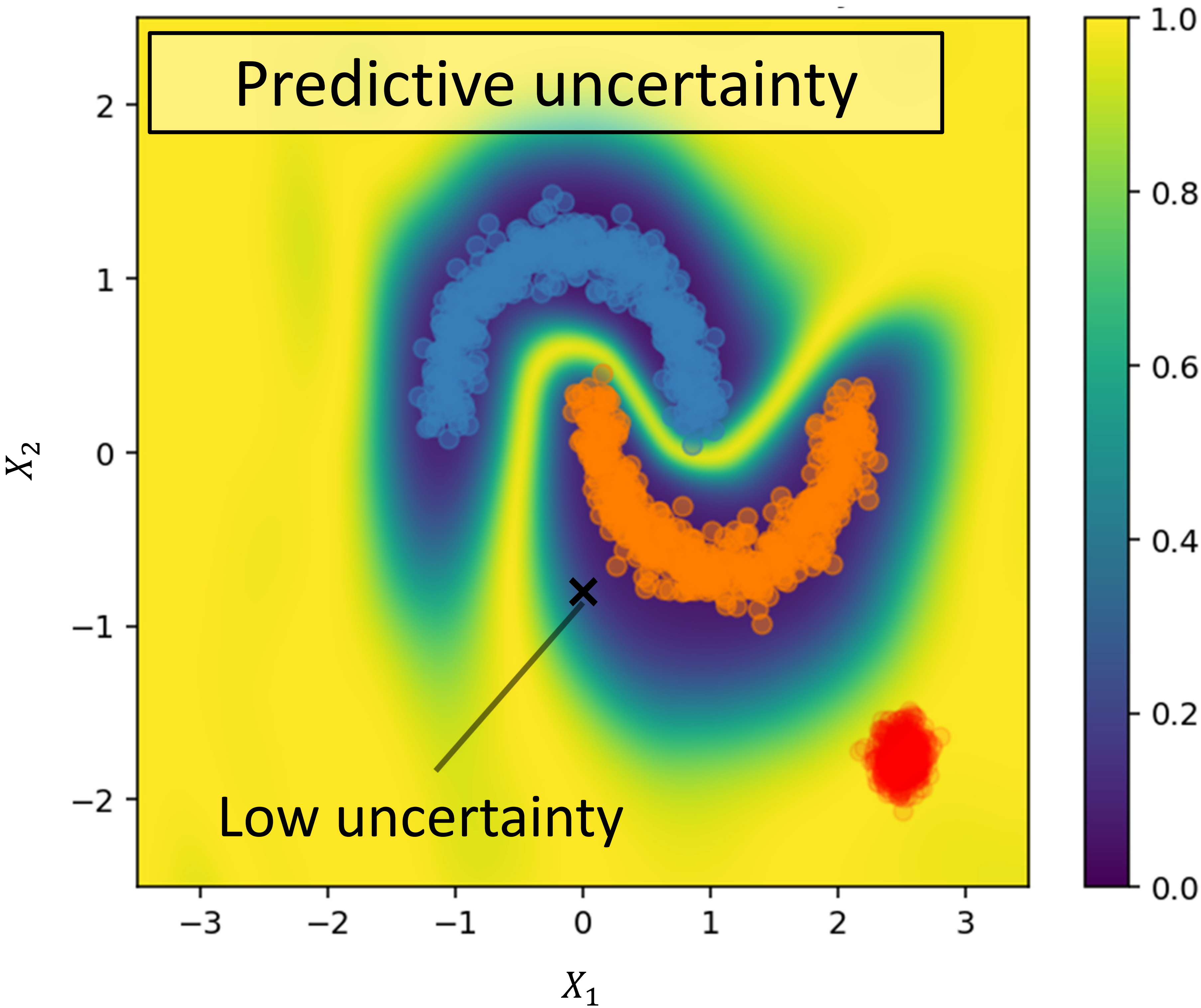


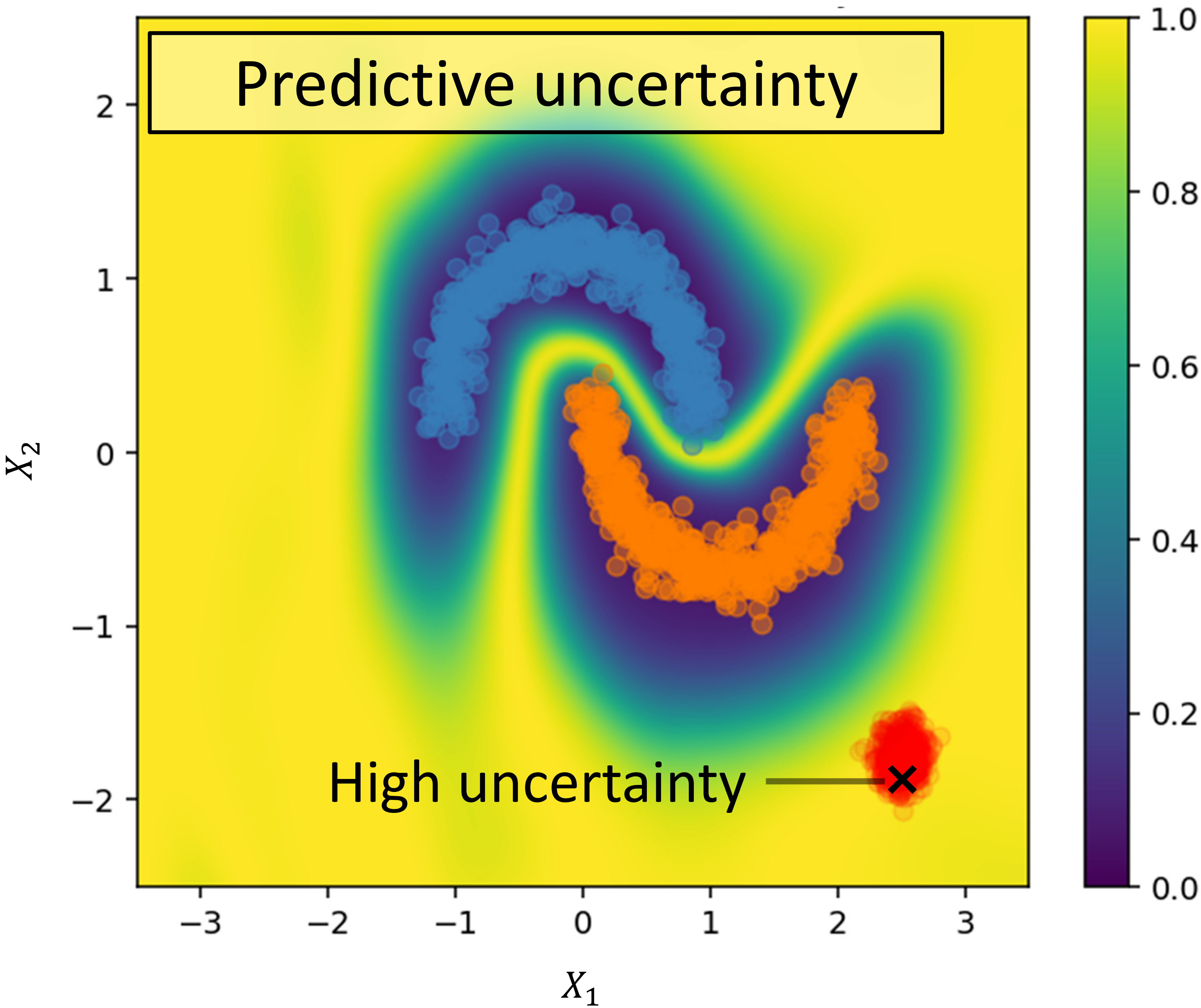
Liu, J., Lin, Z., Padhy, S., Tran, D., Bedrax Weiss, T. and Lakshminarayanan, B., 2020. Simple and principled uncertainty estimation with deterministic deep learning via distance awareness. *NIPS*, *33*, pp.7498-7512.



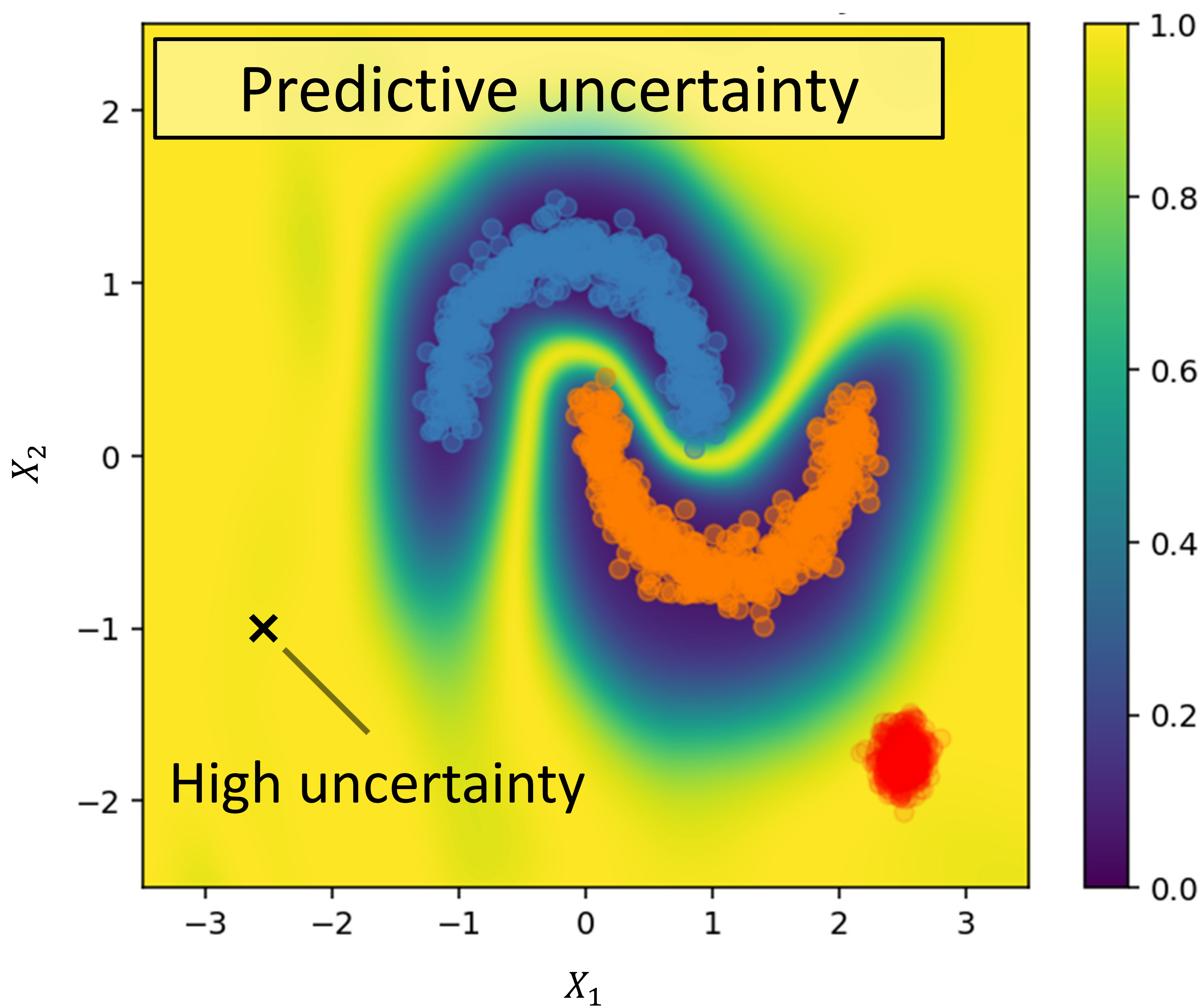
Predictive uncertainty
Low uncertainty
X_1
X_2
−3 −2 −1 0 1 2 3
−2 −1 0 1 2
0.0 0.2 0.4 0.6 0.8 1.0


Predictive uncertainty
Low uncertainty
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
X_1
X_2
1.0
0.8
0.6
0.4
0.2
0.0

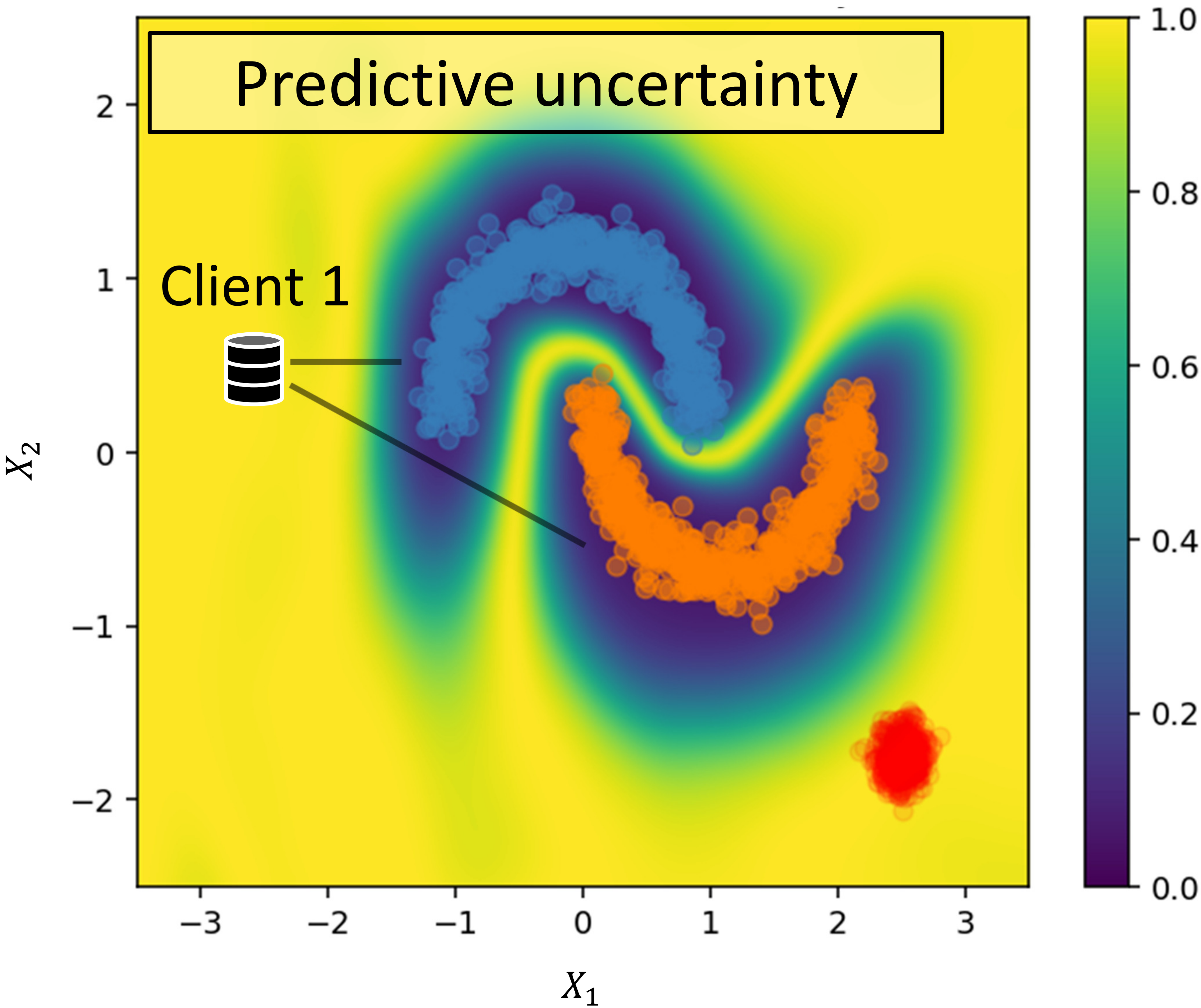

Predictive uncertainty
Low uncertainty
X_1
X_2
−3
−2
−1
0
1
2
3
−2
−1
0
1
2
0.0
0.2
0.4
0.6
0.8
1.0



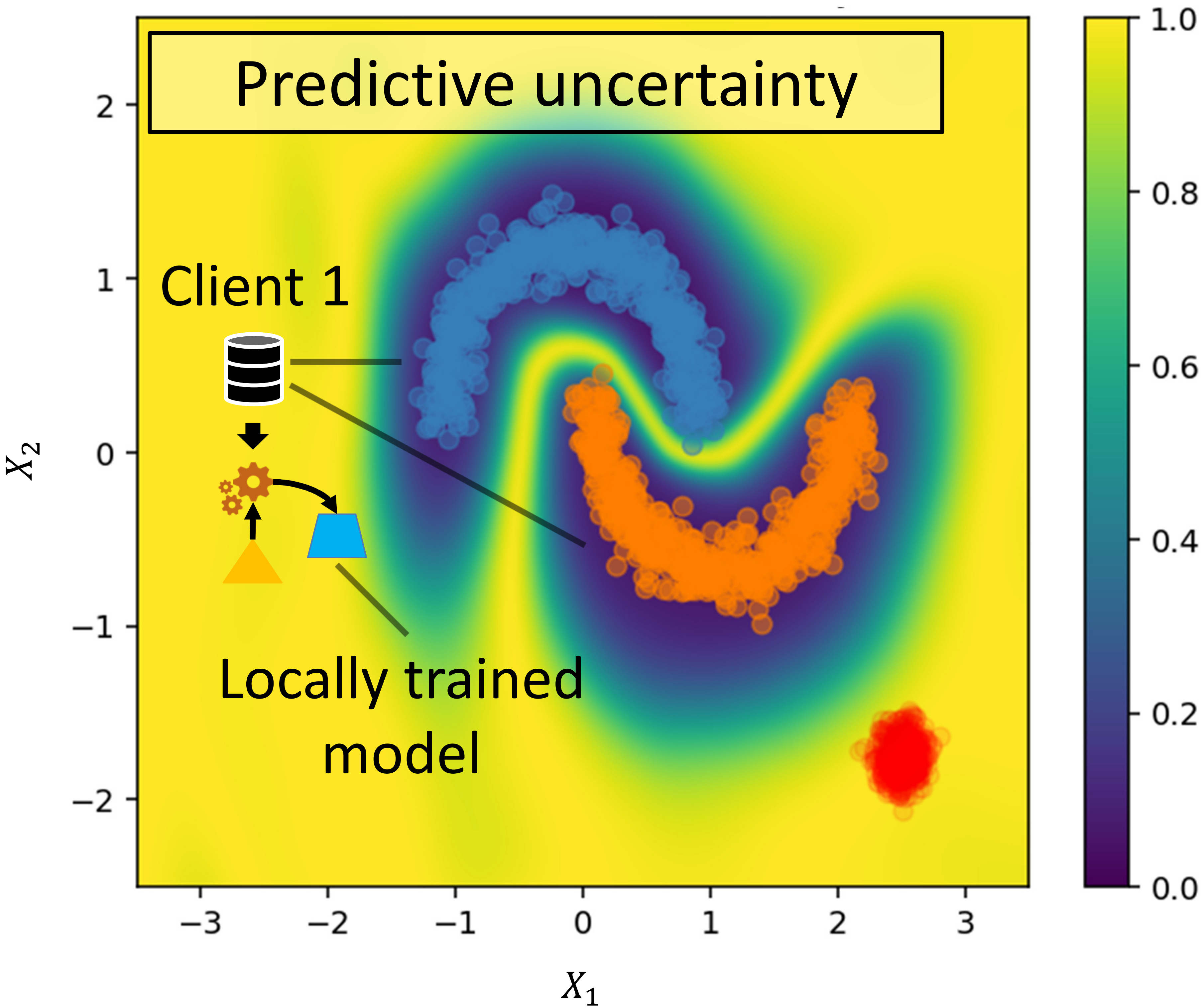
Predictive uncertainty
High uncertainty
1.0
0.8
0.6
0.4
0.2
0.0
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
X_1
X_2

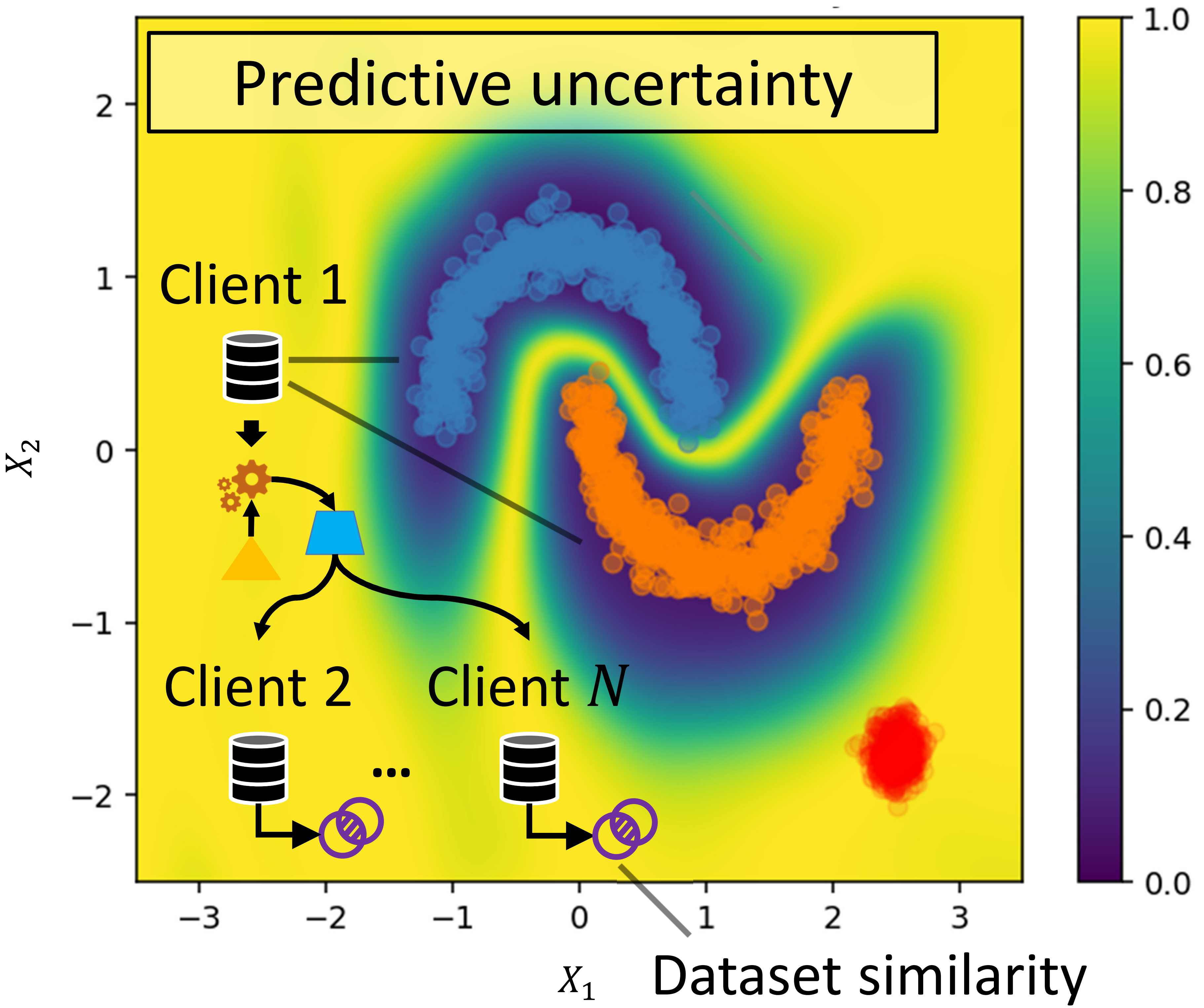


Predictive uncertainty
High uncertainty
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
X_1
X_2
1.0
0.8
0.6
0.4
0.2
0.0

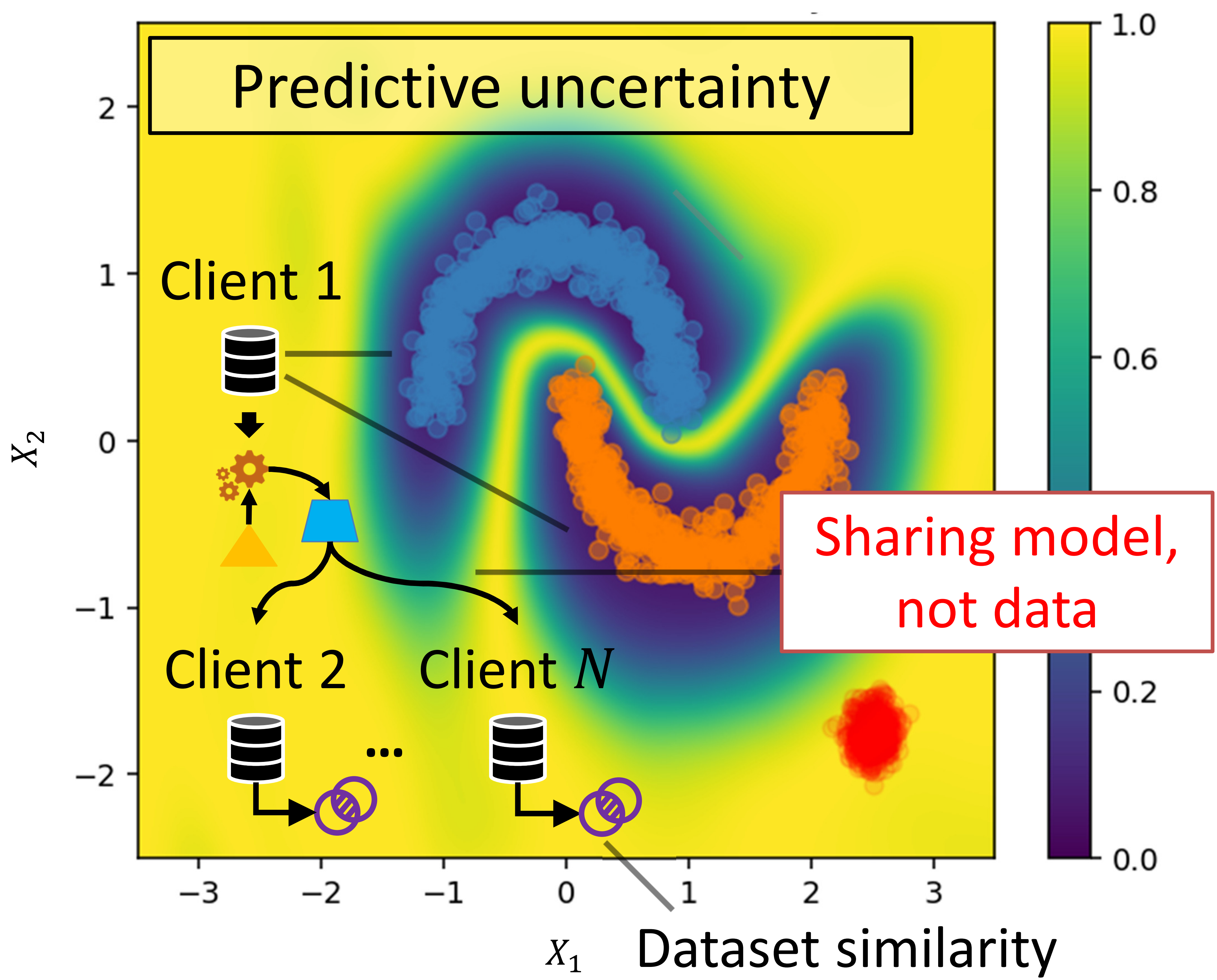


Predictive uncertainty
Client 1
X_1
X_2
−3
−2
−1
0
1
2
3
2
1
0
−1
−2
1.0
0.8
0.6
0.4
0.2
0.0


Predictive uncertainty
Client 1
Locally trained model
X_1
X_2
−3
−2
−1
0
1
2
3
1.0
0.8
0.6
0.4
0.2
0.0



Predictive uncertainty
Client 1
Client 2
Client N
...
Dataset similarity
X_1
X_2
−3 −2 −1 0 1 2 3
2 1 0 −1 −2
1.0 0.8 0.6 0.4 0.2 0.0



Predictive uncertainty
Client 1
Client 2
Client N
...
Sharing model, not data
Dataset similarity
X_1
X_2
−3
−2
−1
0
1
2
3
2
1
0
−1
−2
1.0
0.8
0.6
0.2
0.0

Matrix of Predictive Variances (Normalized)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Dataset 1	0.0	0.02	0.01	0.72	0.76	0.72	0.01	0.02	0.01	0.27	0.24	0.24
Dataset 2	0.0	0.0	0.01	0.58	0.57	0.57	0.0	0.0	0.01	0.3	0.21	0.27
Dataset 3	0.0	0.02	0.0	0.87	0.84	0.89	0.01	0.02	0.0	0.17	0.17	0.13
Dataset 4	1.0	1.0	1.0	0.0	0.04	0.02	1.0	1.0	1.0	0.93	0.99	0.92
Dataset 5	0.76	0.78	0.77	0.01	0.0	0.04	0.78	0.81	0.78	1.0	1.0	1.0
Dataset 6	0.71	0.68	0.7	0.01	0.08	0.0	0.68	0.67	0.68	0.51	0.59	0.49
Dataset 7	0.0	0.01	0.01	0.72	0.79	0.72	0.0	0.01	0.0	0.25	0.2	0.2
Dataset 8	0.0	0.0	0.01	0.58	0.62	0.57	0.0	0.0	0.01	0.32	0.22	0.27
Dataset 9	0.0	0.01	0.0	1.0	1.0	1.0	0.0	0.01	0.0	0.17	0.15	0.14
Dataset 10	0.65	0.69	0.65	0.49	0.56	0.49	0.65	0.72	0.69	0.0	0.02	0.01
Dataset 11	0.28	0.29	0.36	0.54	0.68	0.53	0.33	0.33	0.38	0.0	0.0	0.02
Dataset 12	0.57	0.63	0.51	0.4	0.32	0.46	0.54	0.62	0.55	0.0	0.03	0.0

Color scale: 0, 0.2, 0.4, 0.6, 0.8, 1.0

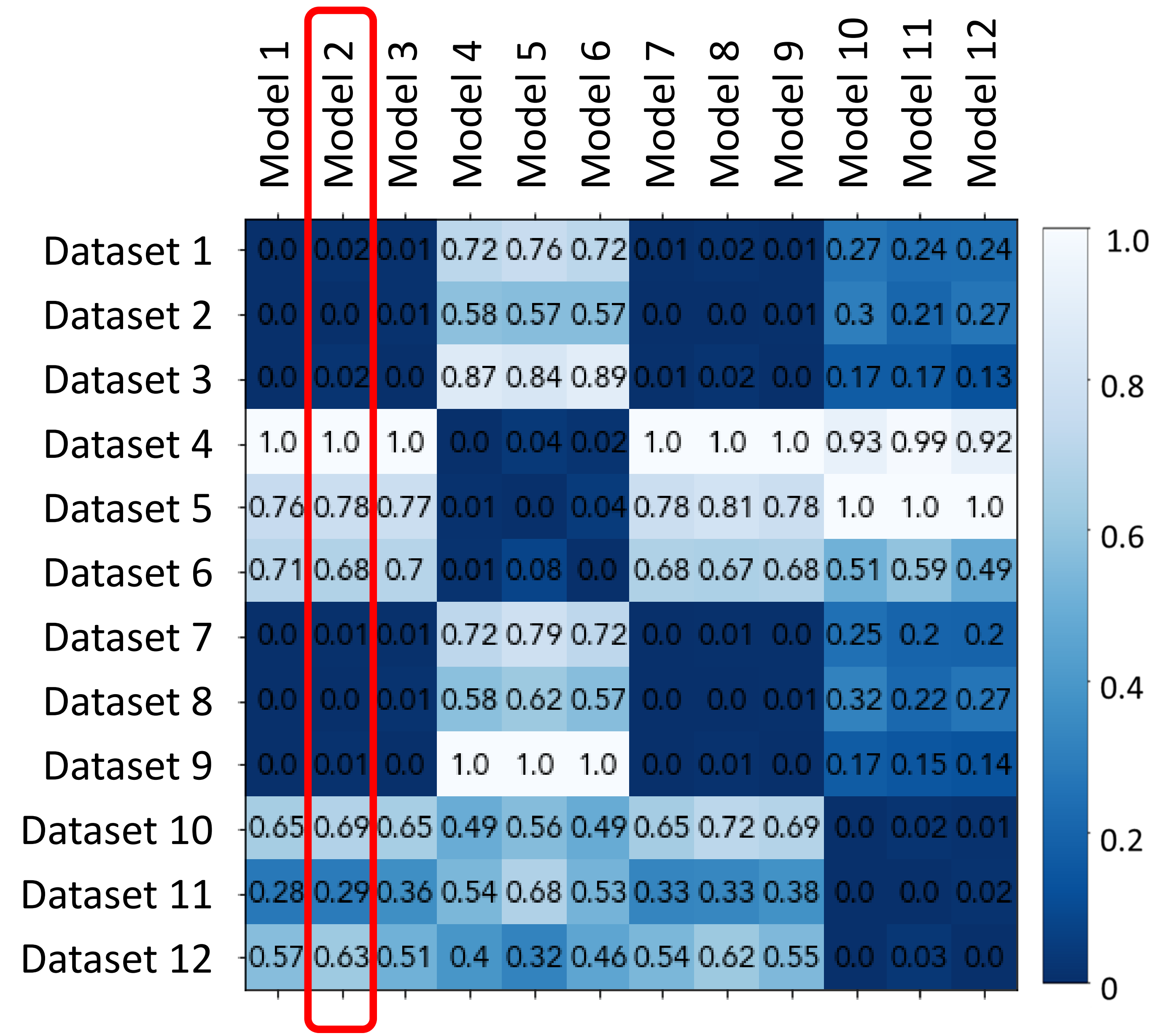
Predictive uncertainty: proxy of dataset similarity

Matrix of Predictive Variances (Normalized)

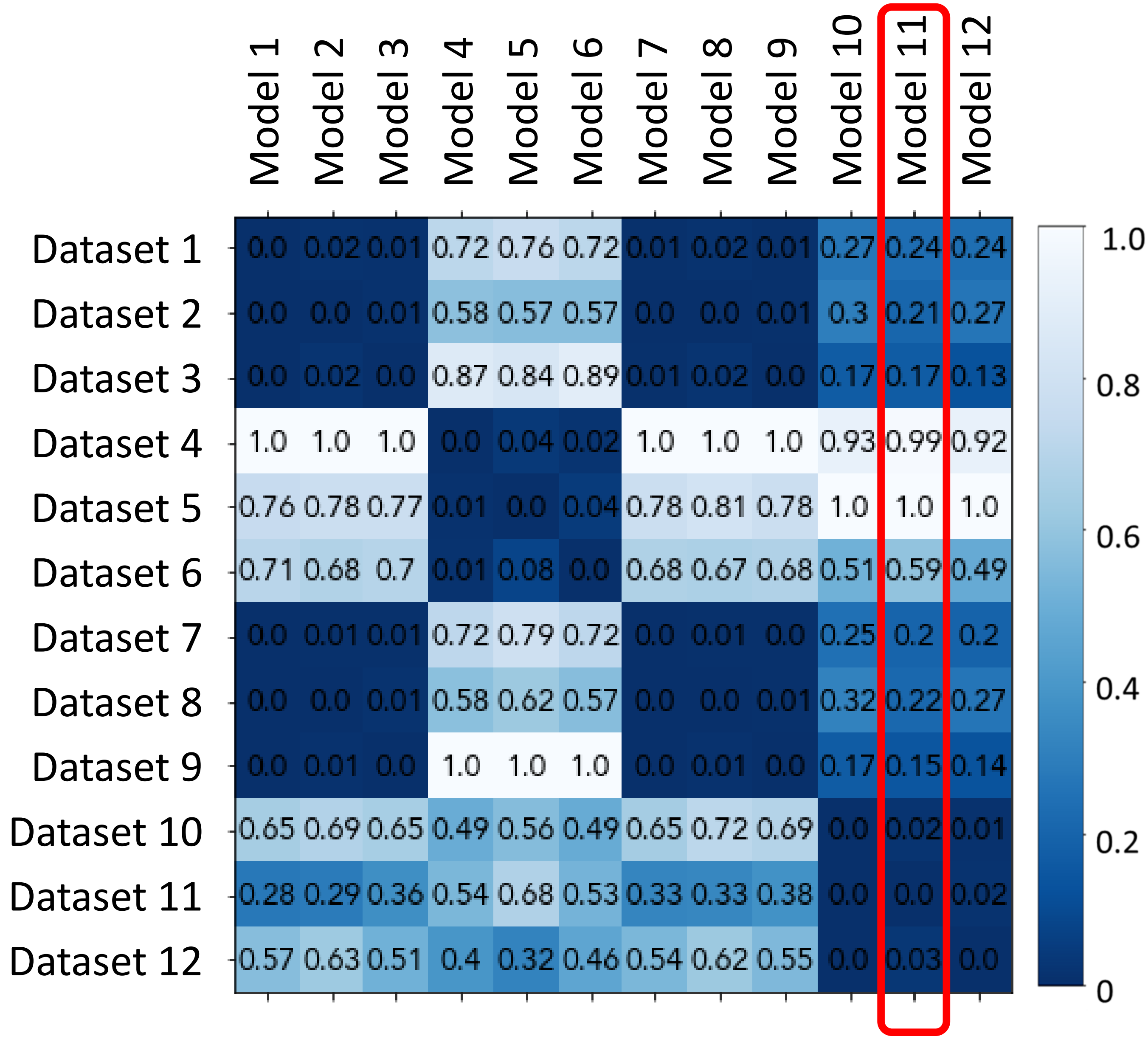
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Dataset 1	0.0	0.02	0.01	0.72	0.76	0.72	0.01	0.02	0.01	0.27	0.24	0.24
Dataset 2	0.0	0.0	0.01	0.58	0.57	0.57	0.0	0.0	0.01	0.3	0.21	0.27
Dataset 3	0.0	0.02	0.0	0.87	0.84	0.89	0.01	0.02	0.0	0.17	0.17	0.13
Dataset 4	1.0	1.0	1.0	0.0	0.04	0.02	1.0	1.0	1.0	0.93	0.99	0.92
Dataset 5	0.76	0.78	0.77	0.01	0.0	0.04	0.78	0.81	0.78	1.0	1.0	1.0
Dataset 6	0.71	0.68	0.7	0.01	0.08	0.0	0.68	0.67	0.68	0.51	0.59	0.49
Dataset 7	0.0	0.01	0.01	0.72	0.79	0.72	0.0	0.01	0.0	0.25	0.2	0.2
Dataset 8	0.0	0.0	0.01	0.58	0.62	0.57	0.0	0.0	0.01	0.32	0.22	0.27
Dataset 9	0.0	0.01	0.0	1.0	1.0	1.0	0.0	0.01	0.0	0.17	0.15	0.14
Dataset 10	0.65	0.69	0.65	0.49	0.56	0.49	0.65	0.72	0.69	0.0	0.02	0.01
Dataset 11	0.28	0.29	0.36	0.54	0.68	0.53	0.33	0.33	0.38	0.0	0.0	0.02
Dataset 12	0.57	0.63	0.51	0.4	0.32	0.46	0.54	0.62	0.55	0.0	0.03	0.0

1.0
0.8
0.6
0.4
0.2
0

Matrix of Predictive Variances (Normalized)

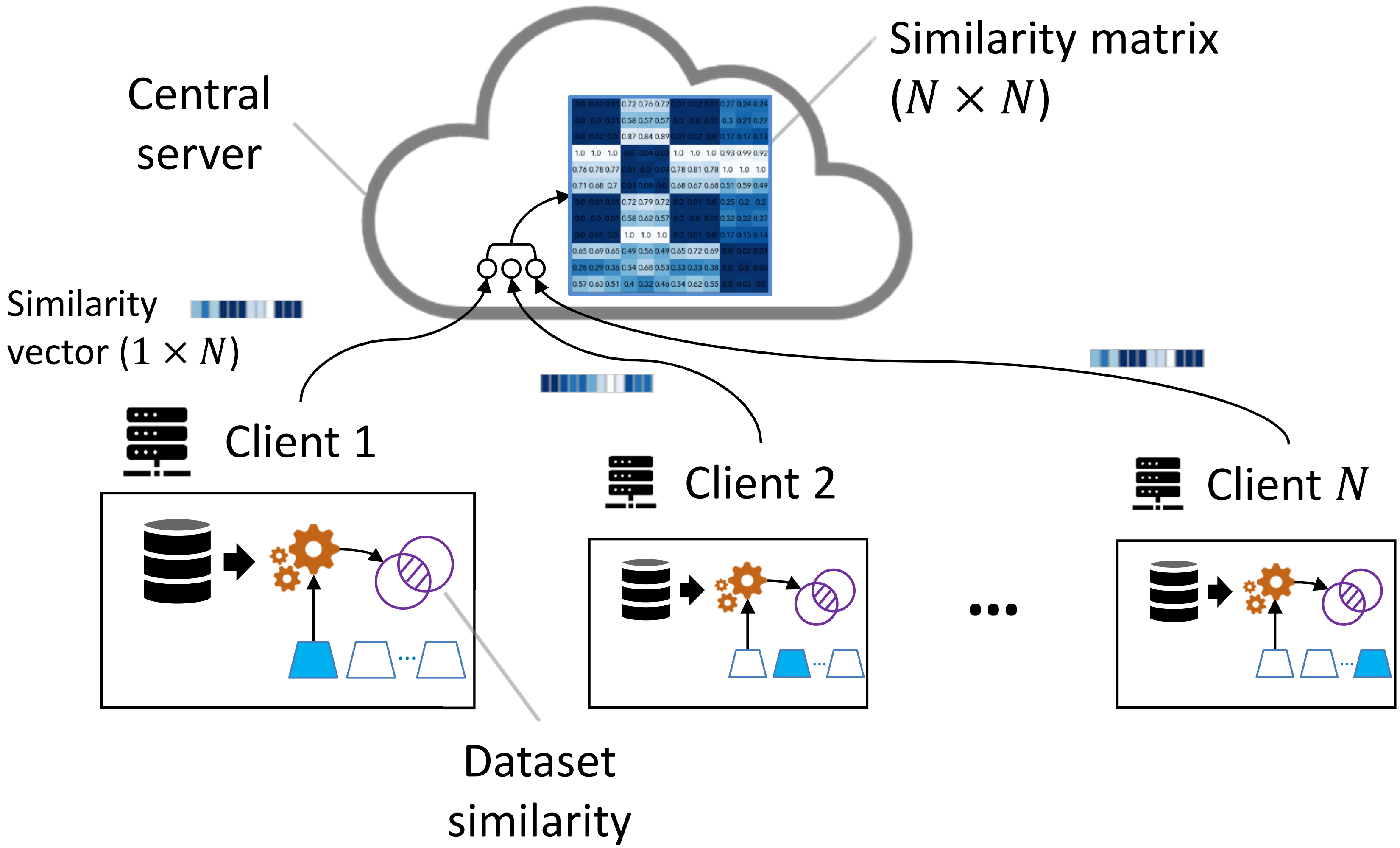


Matrix of Predictive Variances (Normalized)



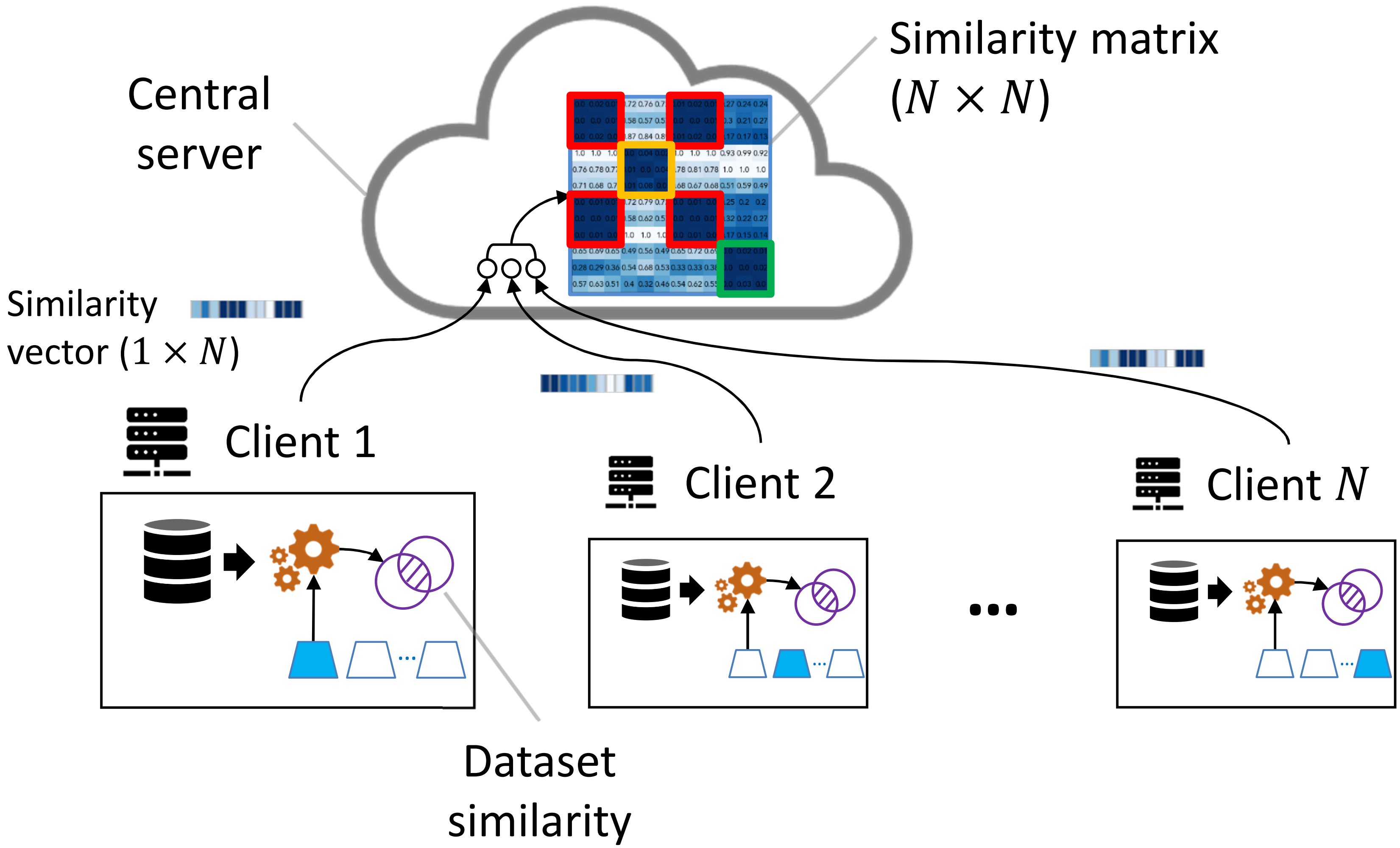
Can We Aggregate Models Selectively?

Step 3.4: Collect dataset similarity from clients

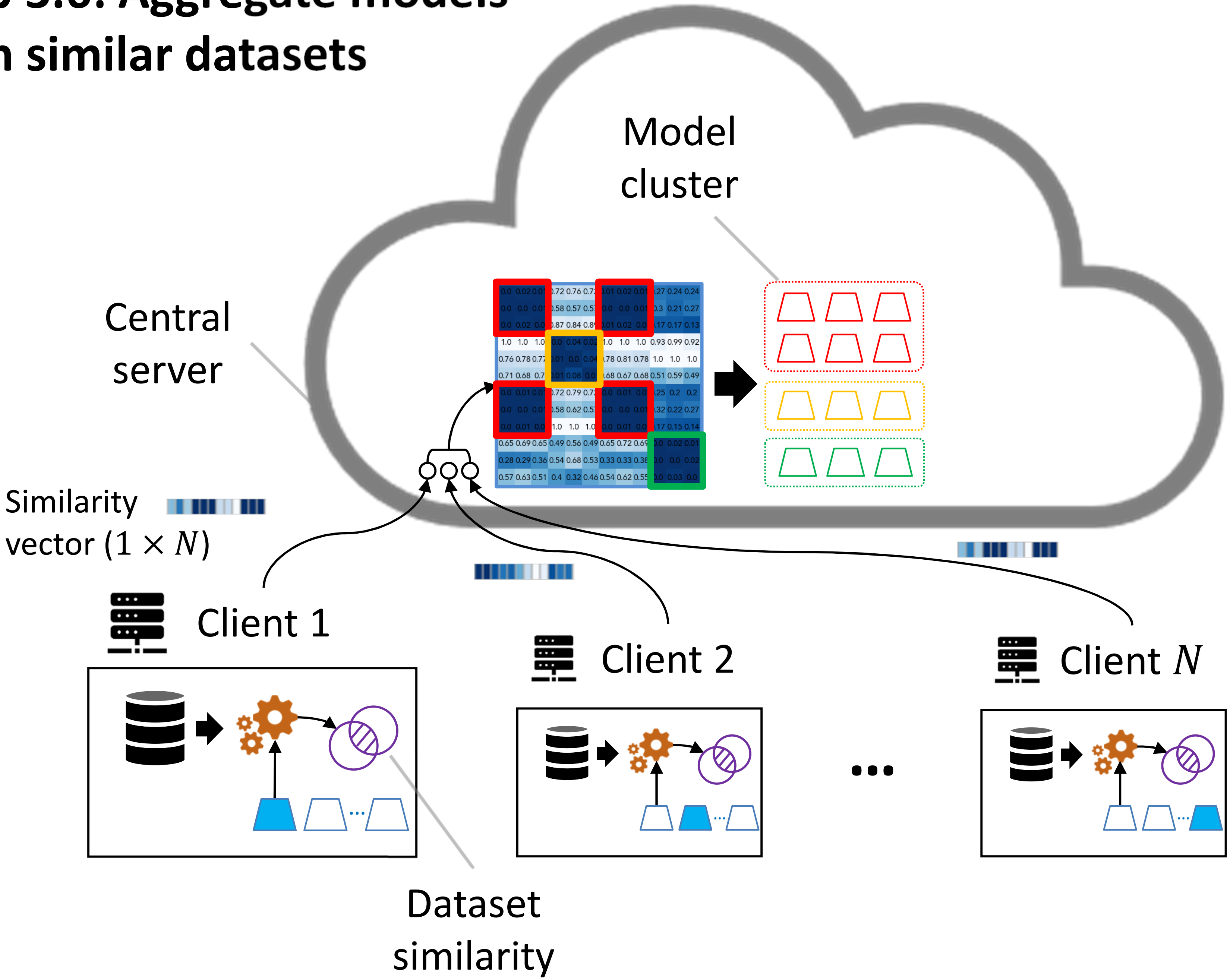


Can We Aggregate Models Selectively?

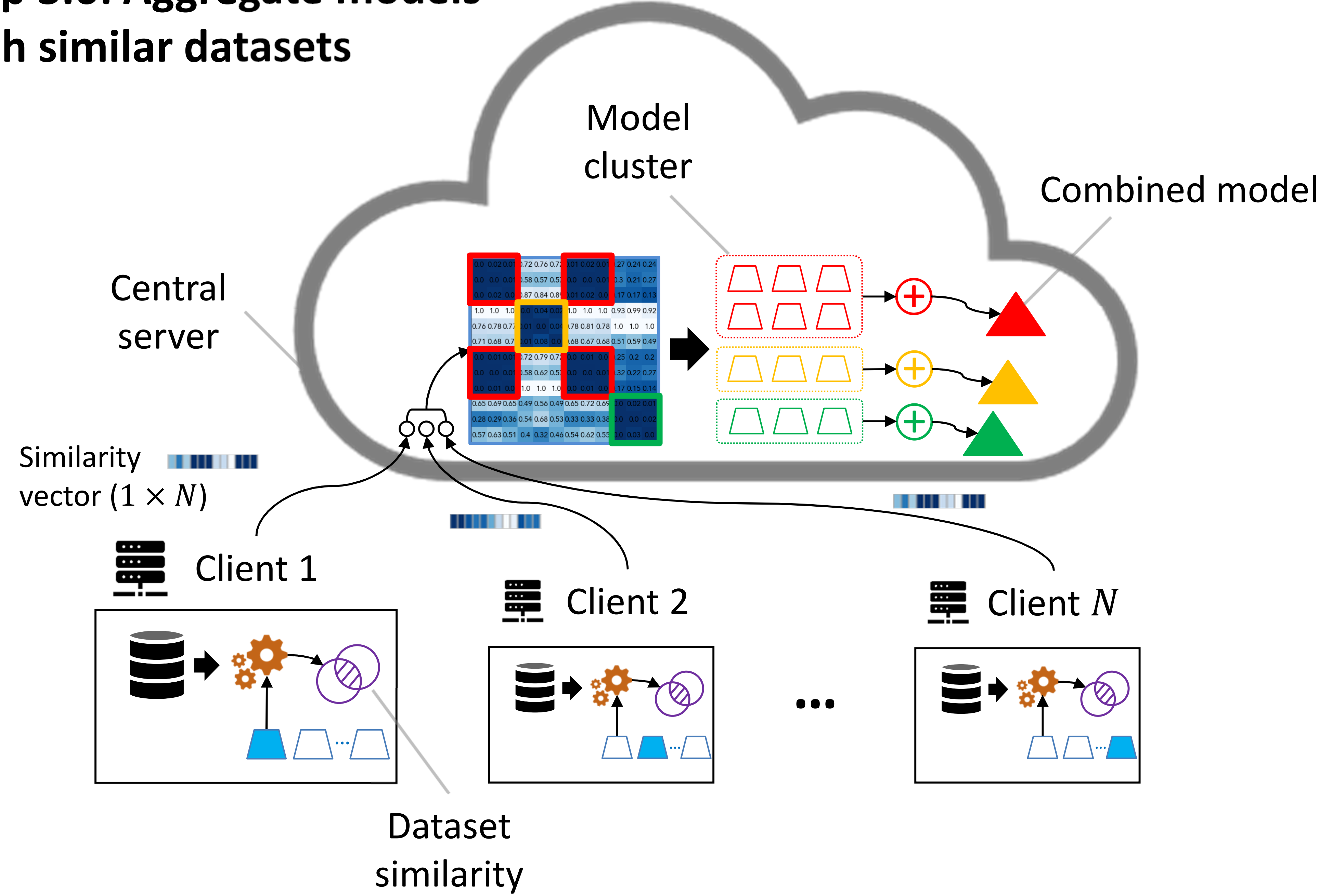
Step 3.5: Cluster clients based on dataset similarity



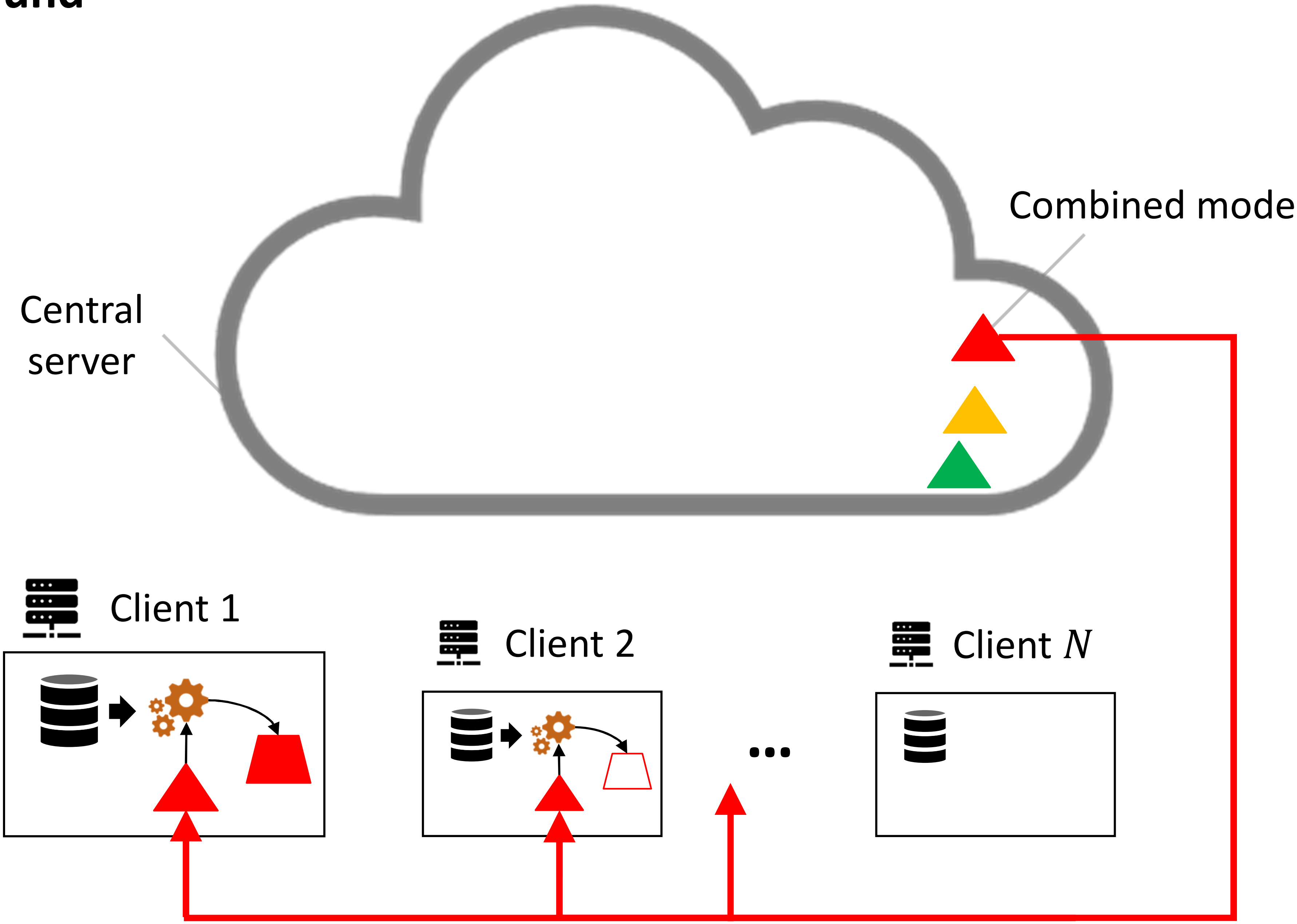
Step 3.6: Aggregate models with similar datasets



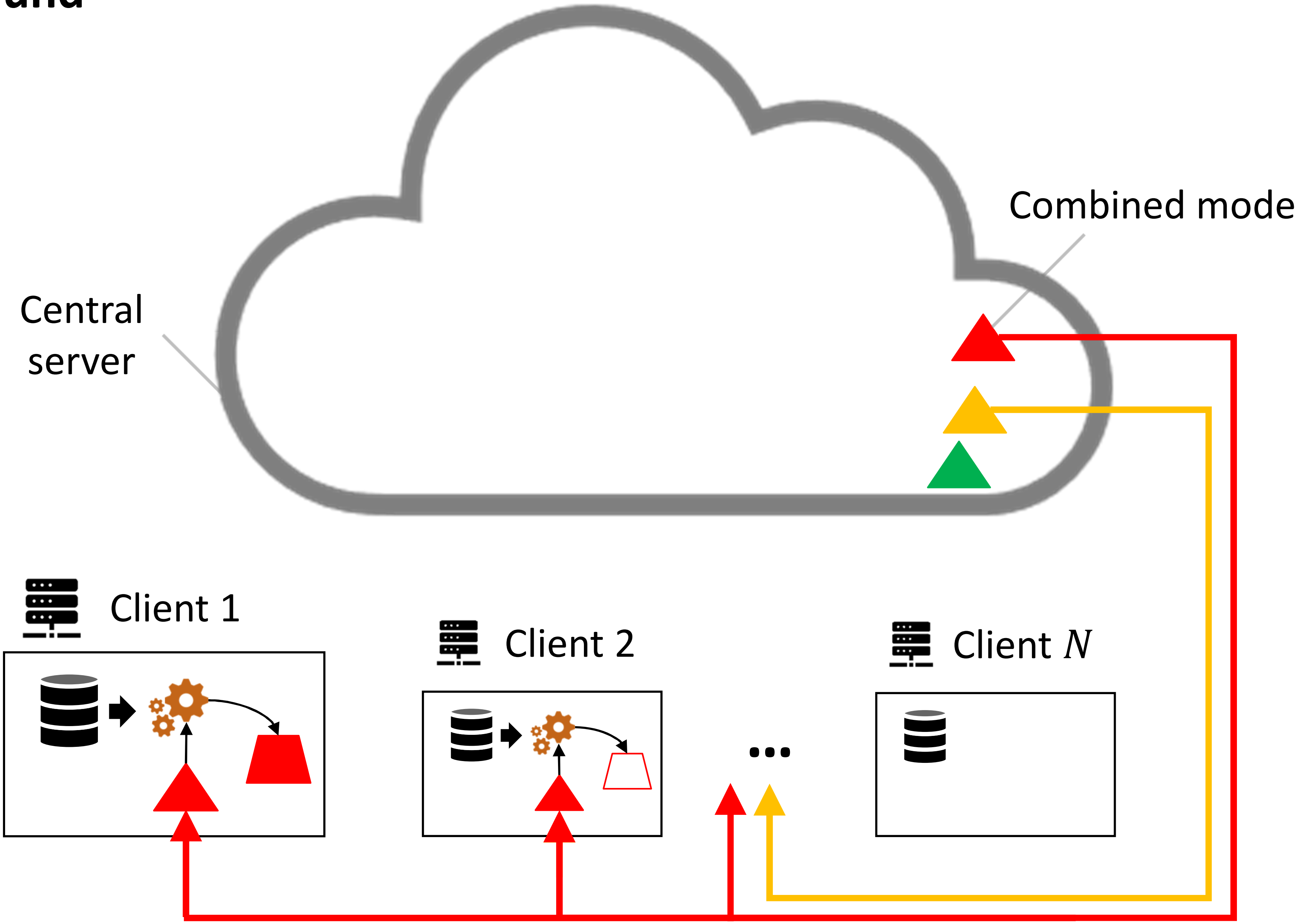
Step 3.6: Aggregate models with similar datasets



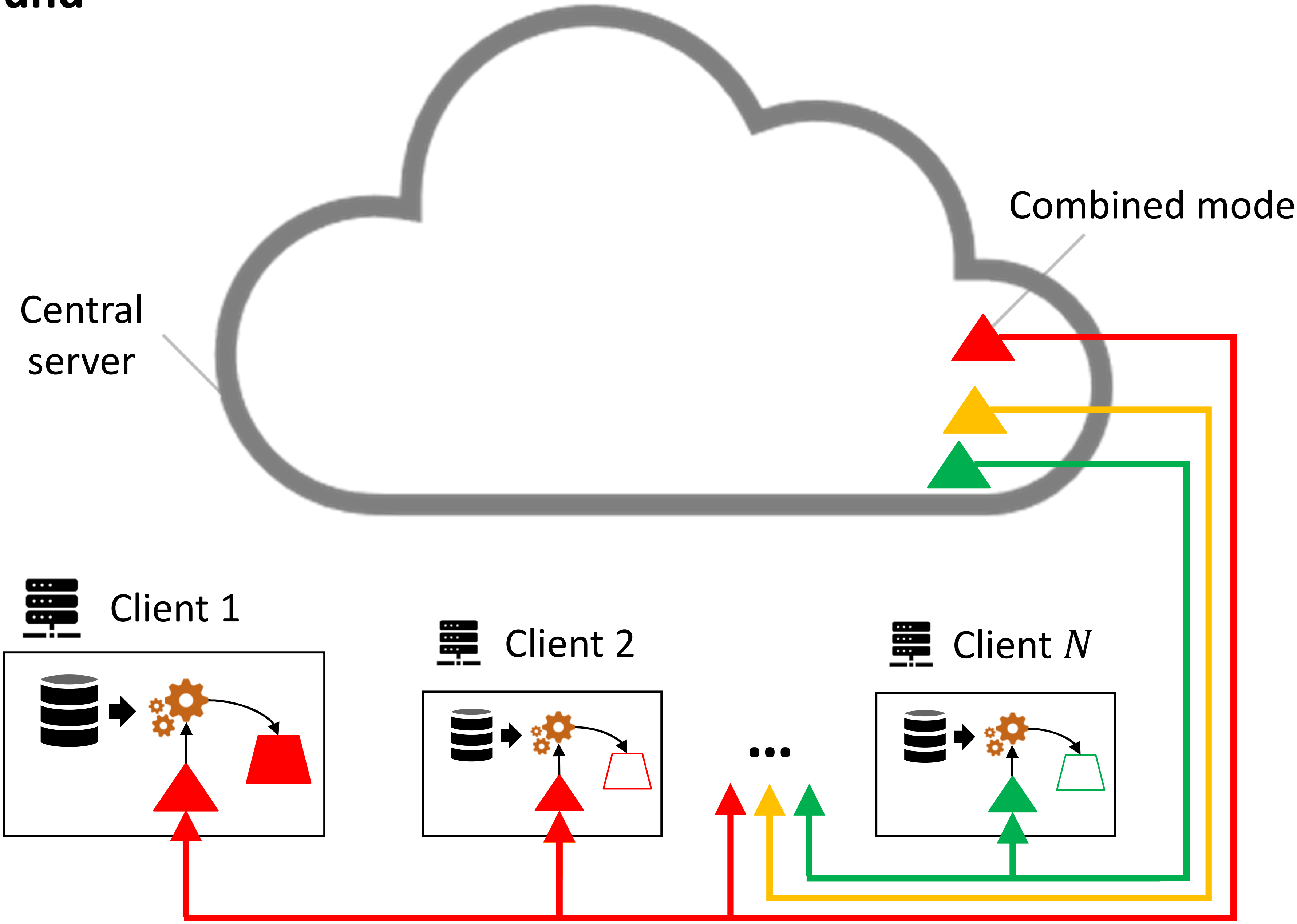
Next Round



Next Round



Next Round



Paderborn University Bearing Dataset

Design of training datasets

Client ID	Operating conditions	Number of samples		
		Healthy	Inner race fault	Outer race fault
1	Shaft speed = 1500 rpm	1200	1200	1200
2	Load torque = 0.7 Nm	1200	1200	0
3	Radial force = 1000 N	1200	0	1200
4	Shaft speed = 900 rpm	1200	1200	1200
5	Load torque = 0.7 Nm	1200	1200	0
6	Radial force = 1000 N	1200	0	1200
7	Shaft speed = 1500 rpm	1200	1200	1200
8	Load torque = 0.1 Nm	1200	1200	0
9	Radial force = 1000 N	1200	0	1200
10	Shaft speed = 1500 rpm	1200	1200	1200
11	Load torque = 0.7 Nm	1200	1200	0
12	Radial force = 400N	1200	0	1200

3 bearings for each health class

Paderborn University Bearing Dataset

Design of training datasets

Cluster 1

Client ID	Operating conditions	Number of samples		
		Healthy	Inner race fault	Outer race fault
1	Shaft speed = 1500 rpm	1200	1200	1200
2	Load torque = 0.7 Nm	1200	1200	0
3	Radial force = 1000 N	1200	0	1200
4	Shaft speed = 900 rpm	1200	1200	1200
5	Load torque = 0.7 Nm	1200	1200	0
6	Radial force = 1000 N	1200	0	1200
7	Shaft speed = 1500 rpm	1200	1200	1200
8	Load torque = 0.1 Nm	1200	1200	0
9	Radial force = 1000 N	1200	0	1200
10	Shaft speed = 1500 rpm	1200	1200	1200
11	Load torque = 0.7 Nm	1200	1200	0
12	Radial force = 400N	1200	0	1200

Paderborn University Bearing Dataset

Design of training datasets

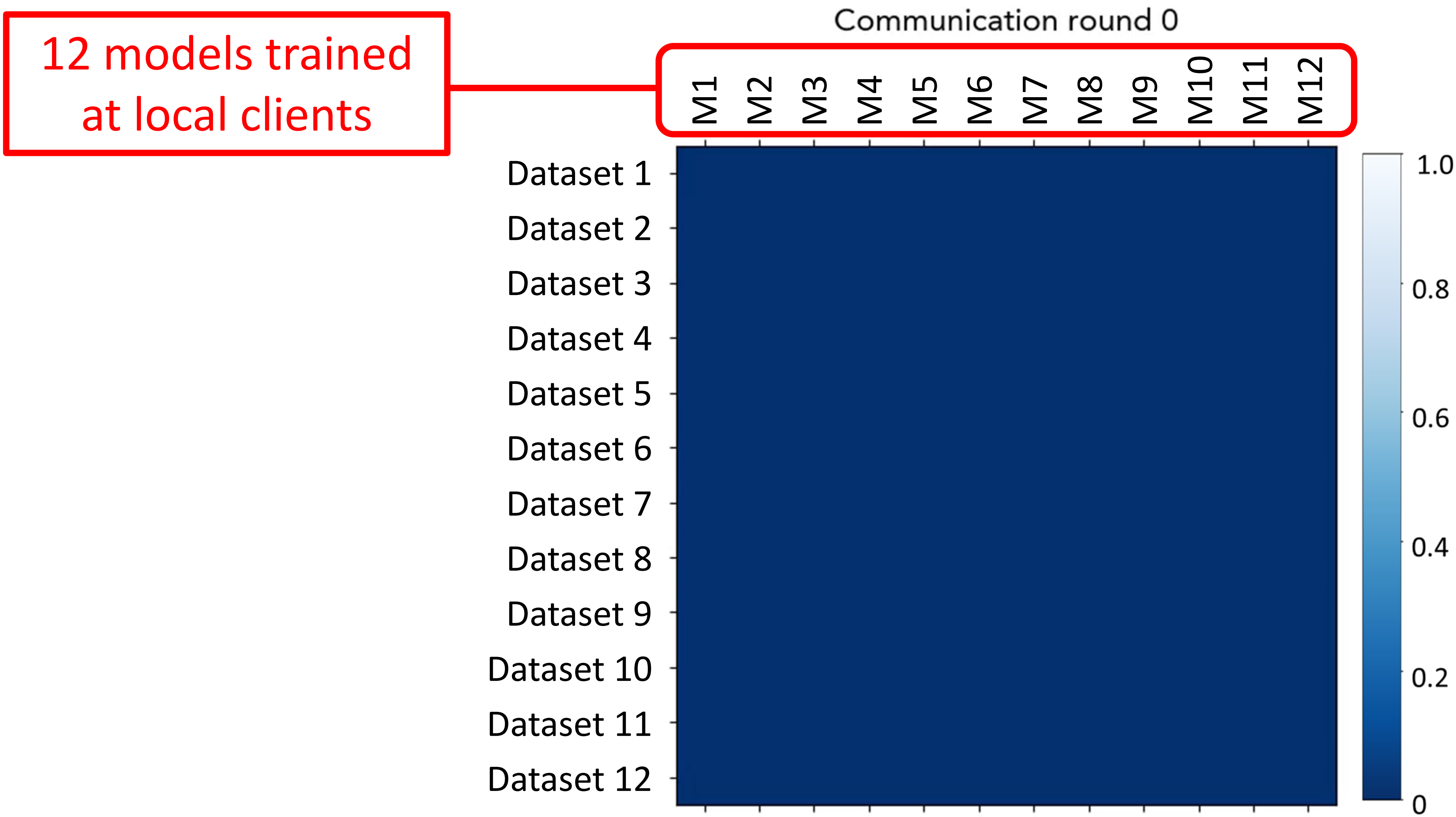
Client ID	Operating conditions	Number of samples: Healthy	Number of samples: Inner race fault	Number of samples: Outer race fault
1	Shaft speed = 1500 rpm	1200	1200	1200
2	Load torque = 0.7 Nm	1200	1200	0
3	Radial force = 1000 N	1200	0	1200
4	Shaft speed = 900 rpm	1200	1200	1200
5	Load torque = 0.7 Nm	1200	1200	0
6	Radial force = 1000 N	1200	0	1200
7	Shaft speed = 1500 rpm	1200	1200	1200
8	Load torque = 0.1 Nm	1200	1200	0
9	Radial force = 1000 N	1200	0	1200
10	Shaft speed = 1500 rpm	1200	1200	1200
11	Load torque = 0.7 Nm	1200	1200	0
12	Radial force = 400N	1200	0	1200

Cluster 1

Cluster 2

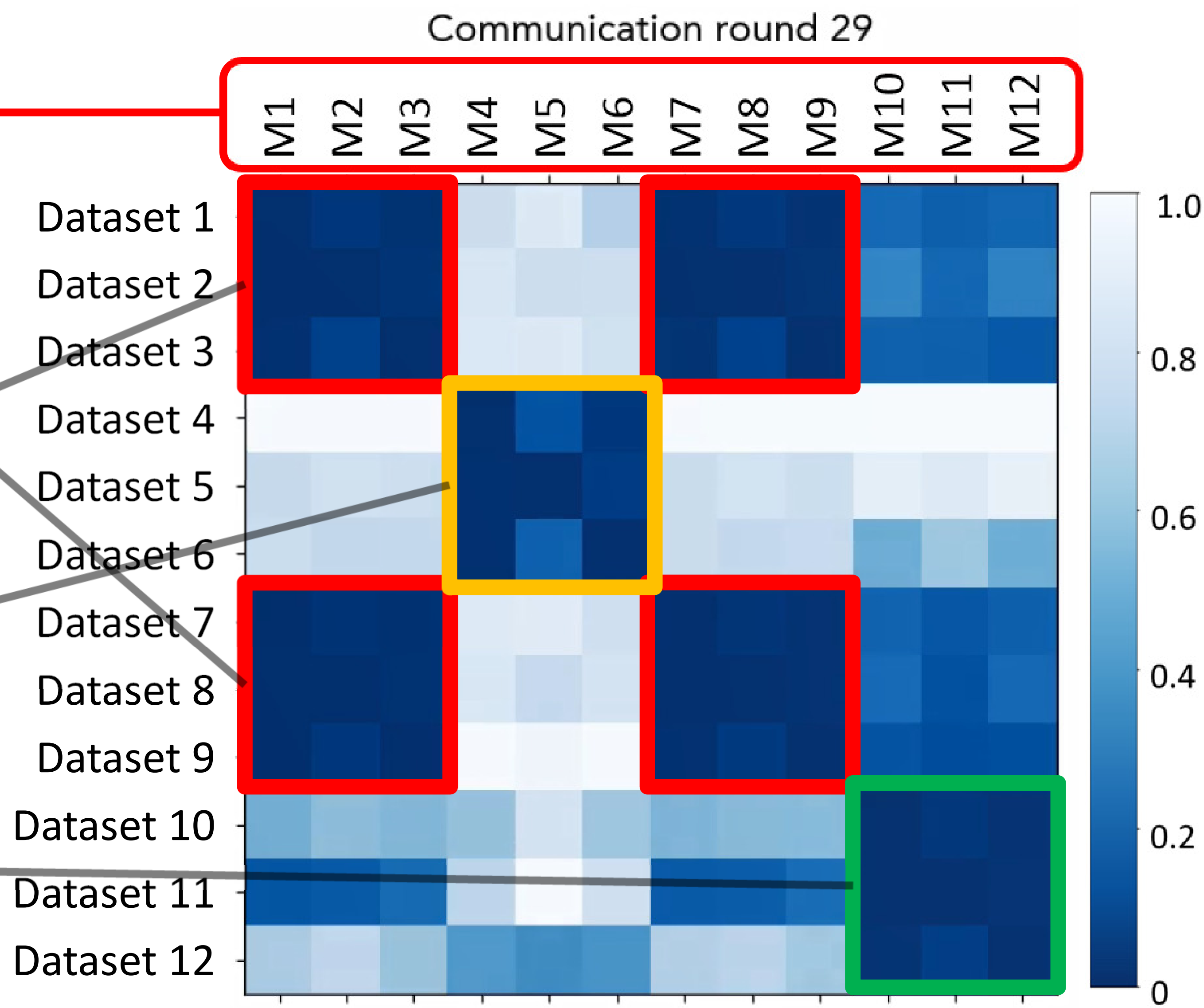
Cluster 3

How Does Similarity Matrix Evolve?



How Does Similarity Matrix Evolve?

12 models trained at local clients

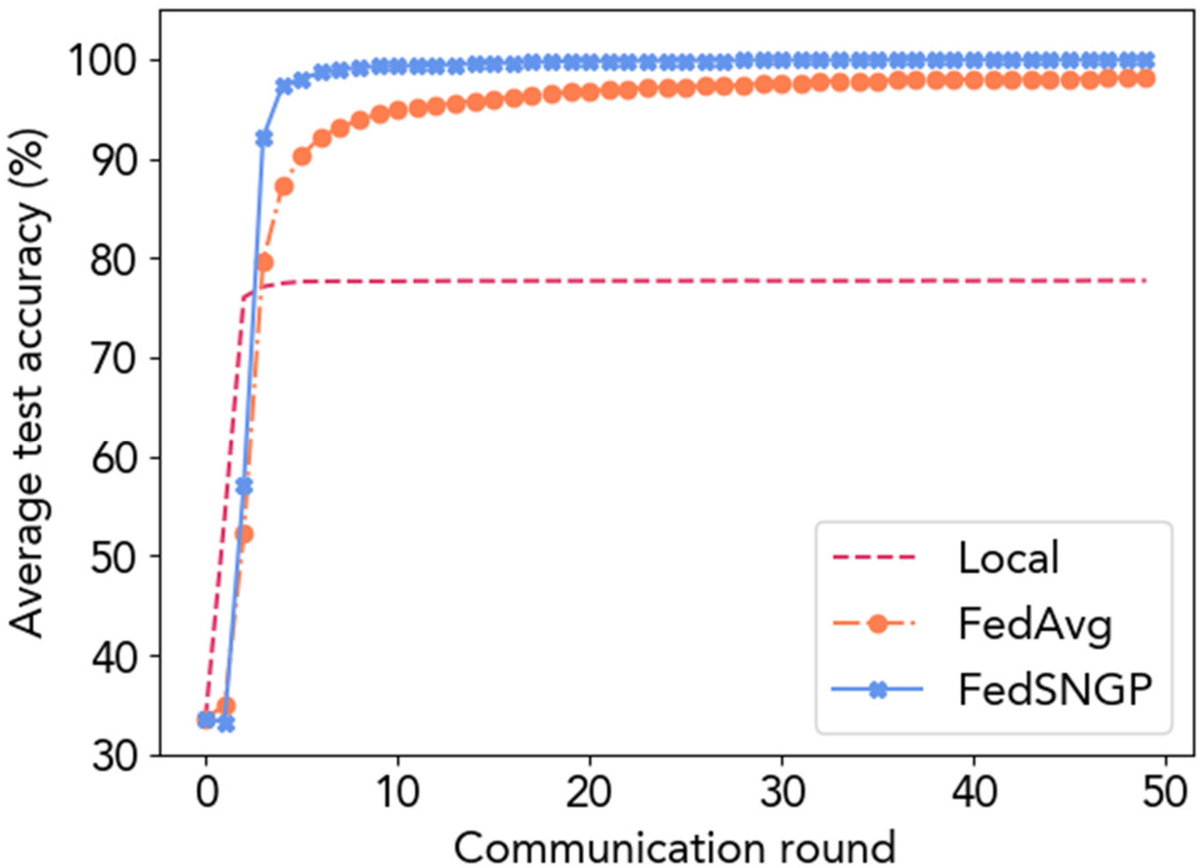


Cluster 1: clients 1-3, 7-9

Cluster 2: clients 4-6

Cluster 3: clients 10-12

Performance Comparison



FedSNGP compared to two baselines:
local training and FedAvg

Hu Group at UConn and Iowa State Univ.

Group Members

- Dr. Mohammad Behtash (Postdoc)
  - Battery Charging Control & Design for Reman
- Yang Kang Chua (PhD), Austin Bray (MS)
  - Prognostics & Real-Time Machine Learning
- Tingkai Li (PhD), Sina Navidi (PhD)
  - Data-Driven Battery Monitoring
- Benjamin Nowacki (PhD)
  - Battery Modeling and Fast Charging Control
- **Dr. Adam Thelen (Modeling Engineer)**
- **Dr. Hao Lu (Assistant Professor at UPC)**
- Dr. Jinqiang Liu (Data Scientist at NOV)
- Dr. Todd Thompson (Staff Engineer at Deere)
- Dr. Venkat Nemani (Data Scientist at Corteva)
- Dr. Yu Hui Lui (Battey Scientist at Rivian)
- Dr. Sheng Shen (ML Engineer at Tula)
- Dr. Meng Li (Data Scientist at NOV)
- Dr. Mohammad Sadoughi (Research Scientist at Amazon)
- Dr. Austin Downey (Assistant Professor at USC)

Research Sponsors

Reliability Engineering & Informatics Laboratory

Thank You!

Q/A

Summary of Results