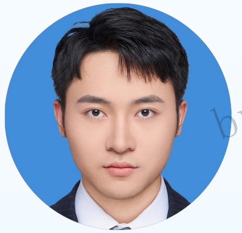
ECAI2025

FairFed++: Closing the Fairness Gap in Federated Learning through Self-Evolving Clustered Optimization




Zhixiang Fang *Baoyao Yang** *Weide Zhan* *Yanchao Tang* *Yiqun Zhang*

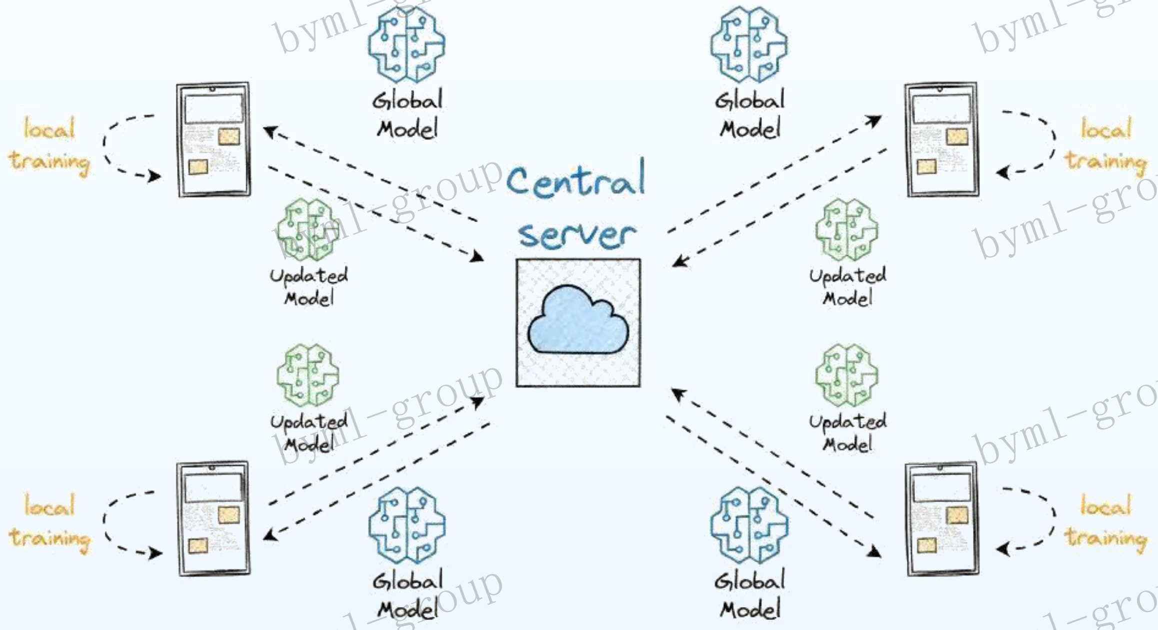
廣東工業大學
Guangdong University of Technology

O C T O B E R 1 7 , 2 0 2 5

Background

Federated Learning

- A distributed machine learning paradigm
- Train a generalized model by transferring model rather than raw data



★ Reduce data leakage risk
★ Break down data silos
★ Extend receptive field for better generalization

Medical Intelligence

Smart City

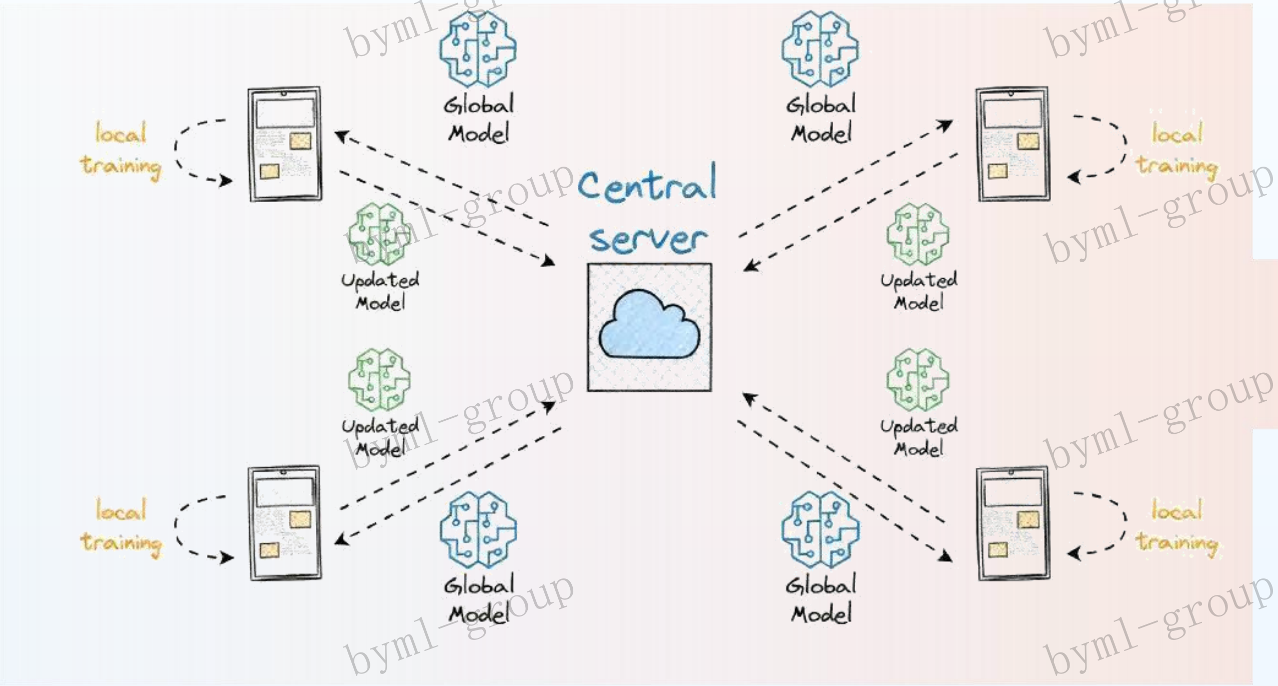
Banking

FL is a privacy-preserving framework with broad applicability across domains.

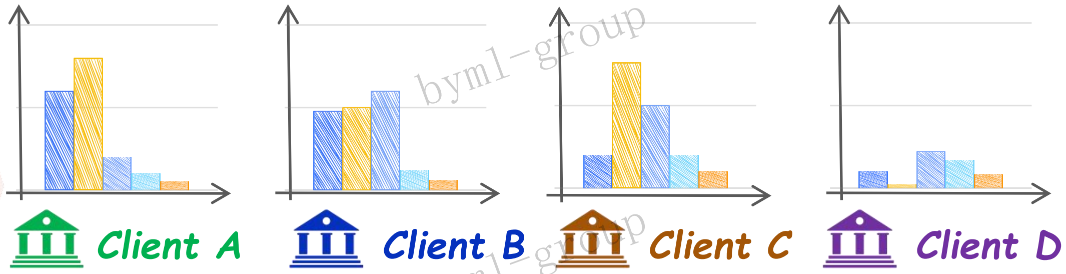
Background

Federated Learning

- A distributed machine learning paradigm
- Train a generalized model by transferring model rather than raw data



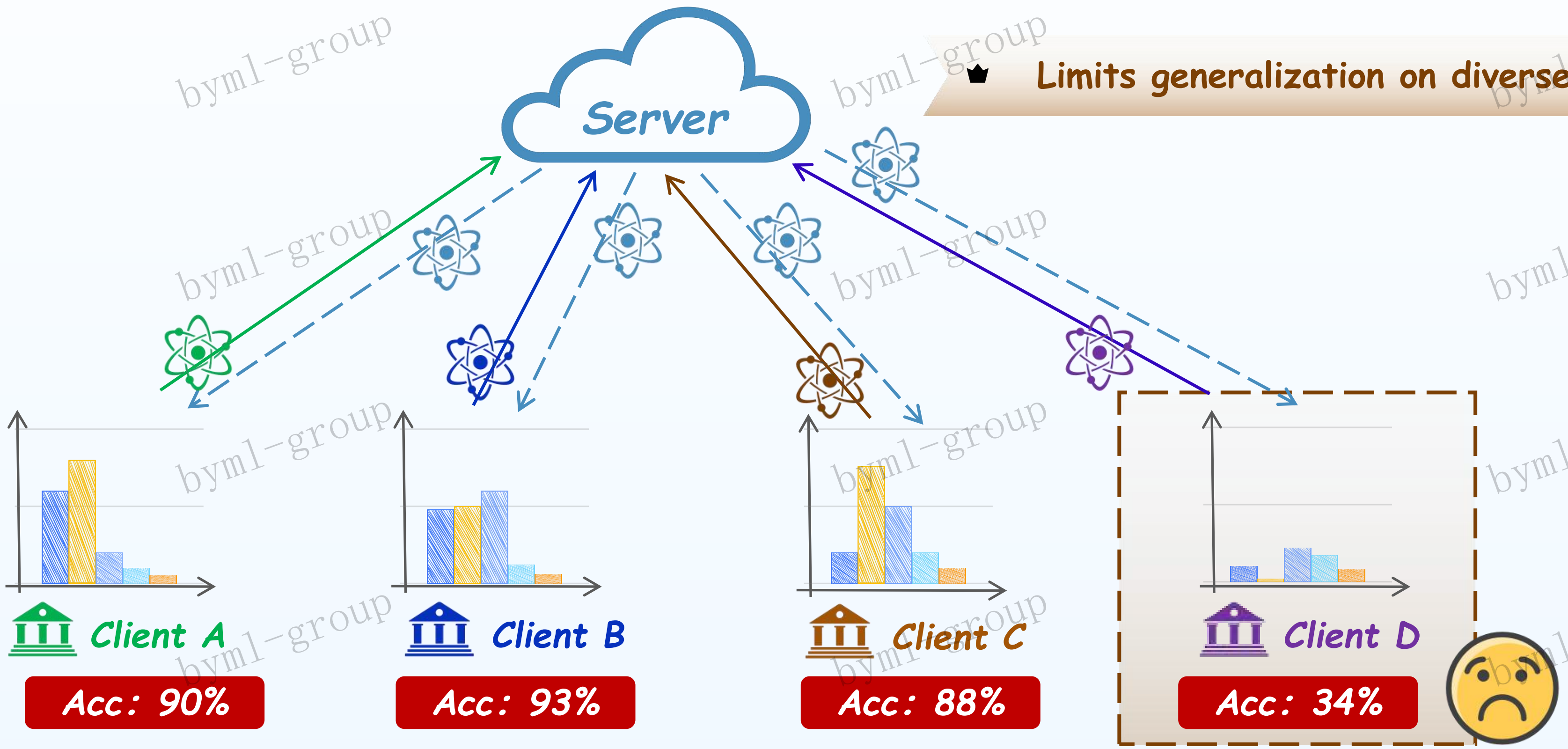
Core Challenge Data Heterogeneity (Non-IID)



⇨ *gradient conflicts during training*

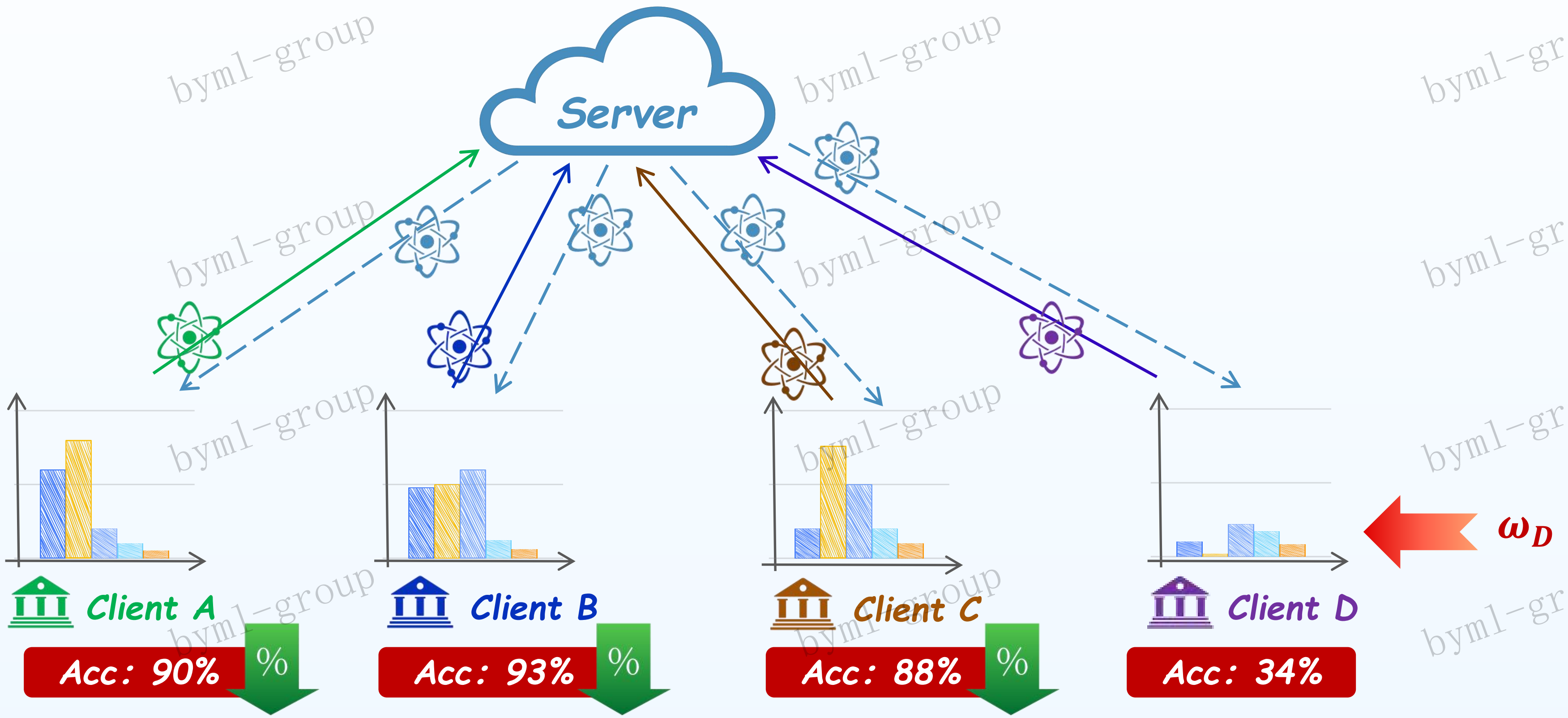
Fairness Issue

Fairness Issue in FL



- Limits generalization on diverse data patterns
- Discourages participation of poor-performed clients
- Ethical Concerns: Unequal treatment of user groups

Mainstream Solution & Limitations

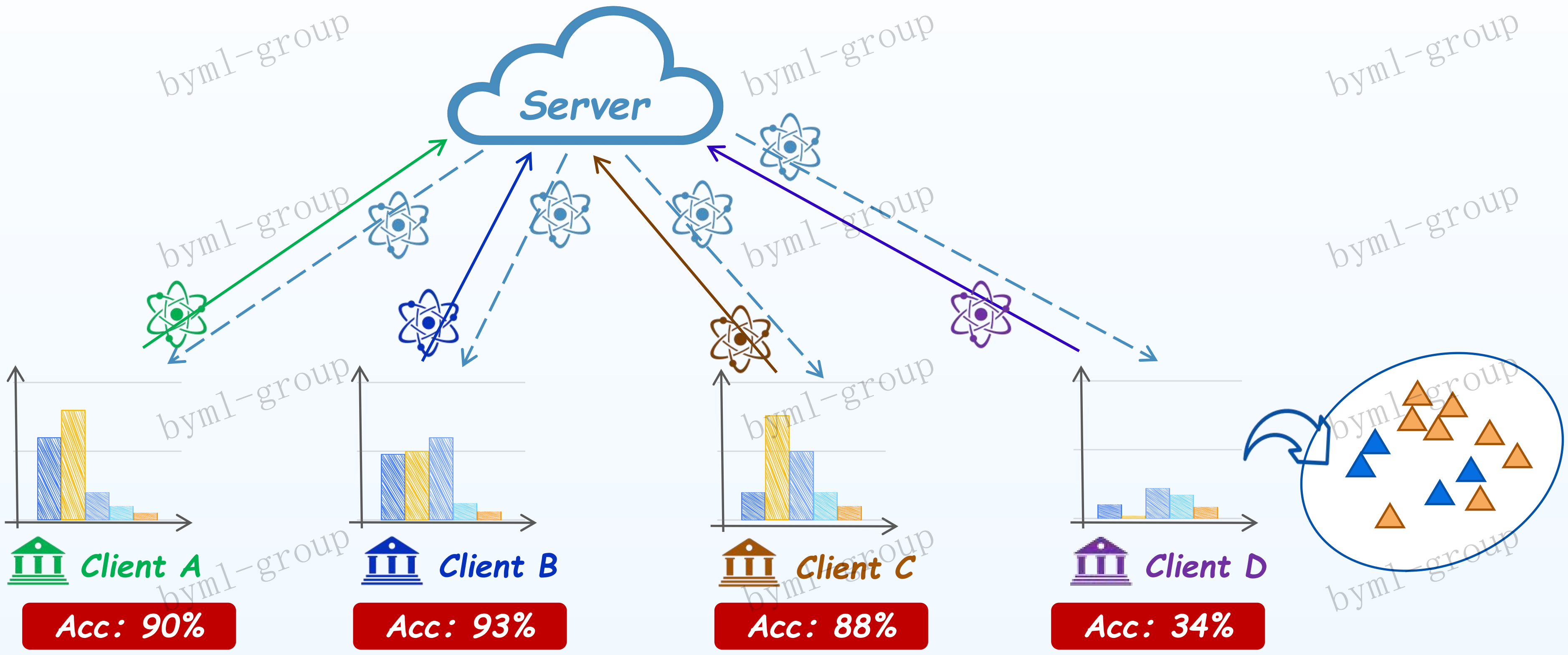


① Reweighting (e.g., q-FFL, FedFa)

Assign higher weights to clients with higher loss.

☹ *Sacrifices the performance of high-performing clients.*

Mainstream Solution & Limitations

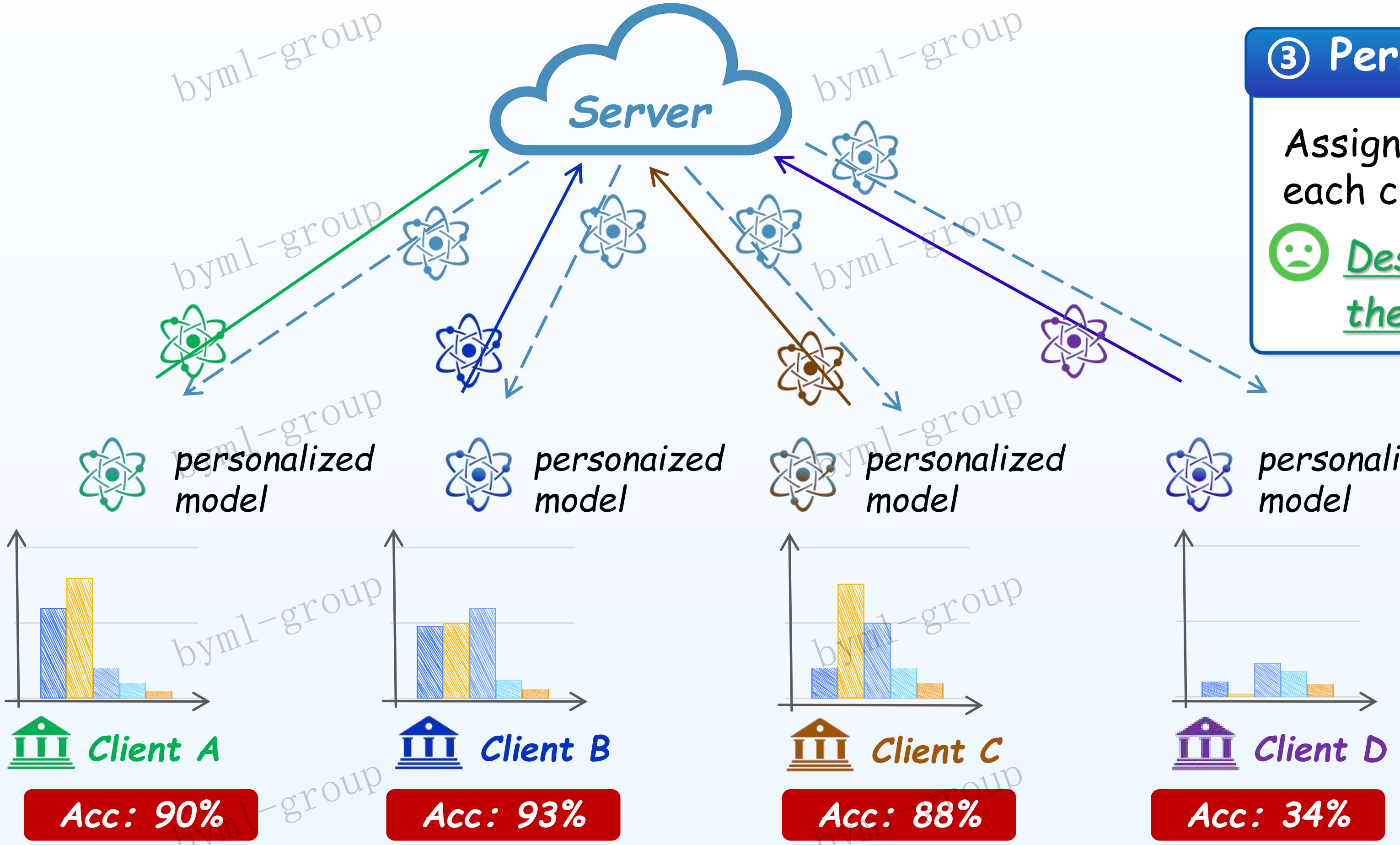


② Data Augmentation (e.g., q-FFL, FedFa)

Synthesize data for under-represented clients.

☹ Rely on the quality of synthesis

Mainstream Solution & Limitations



③ Personalized FL (e.g., Ditto)

Assign and train a personalized model for each client

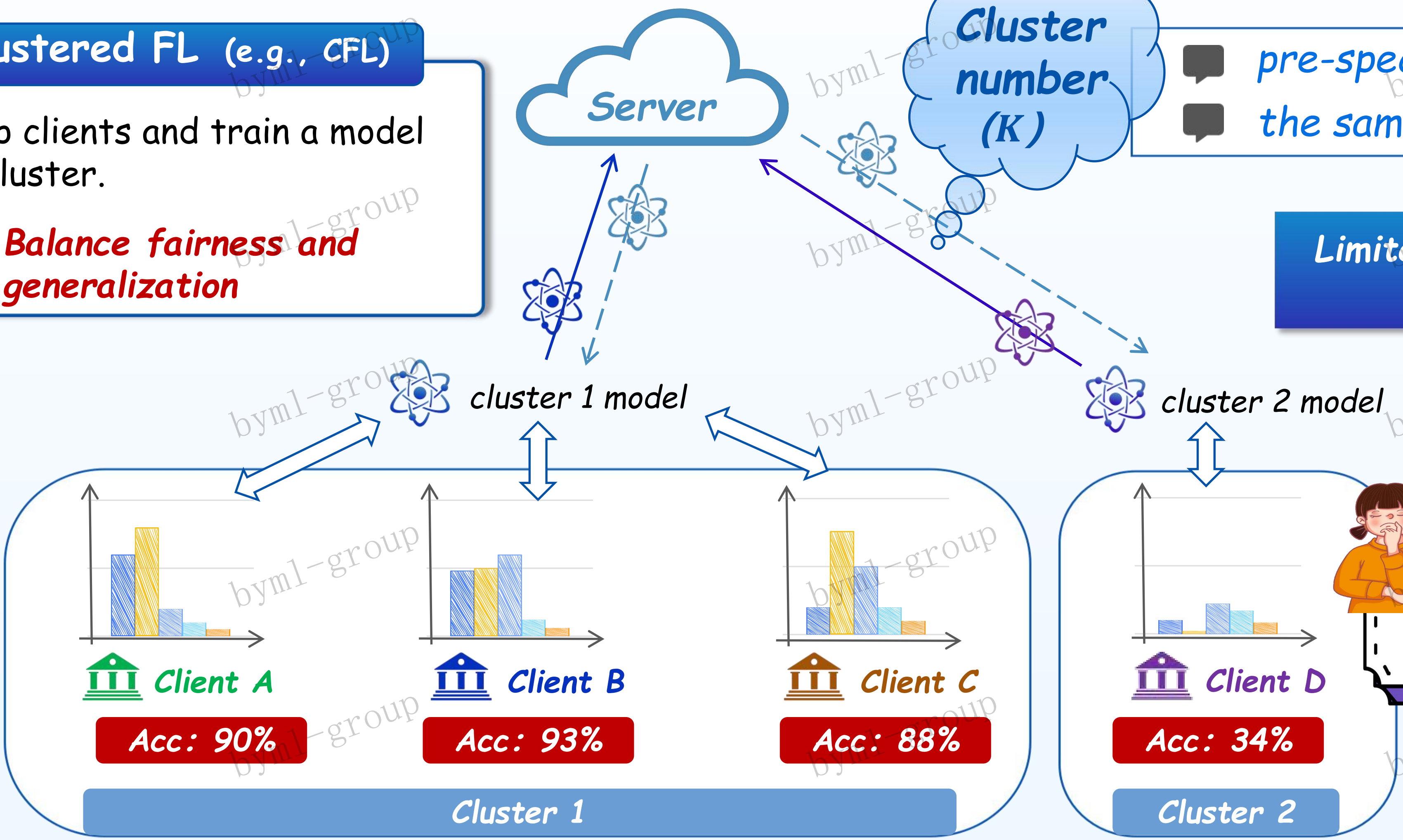
☹ Destroy the generalization ability of the global model.

Mainstream Solution & Limitations

④ Clustered FL (e.g., CFL)

Group clients and train a model per cluster.

Balance fairness and generalization



- *pre-specified?*
- *the same across vairous scenarios?*

Limited applicability in real-world applications

A more comprehensive and adaptable way?

Our Proposal (Self-Evolving Clustered Optimization)

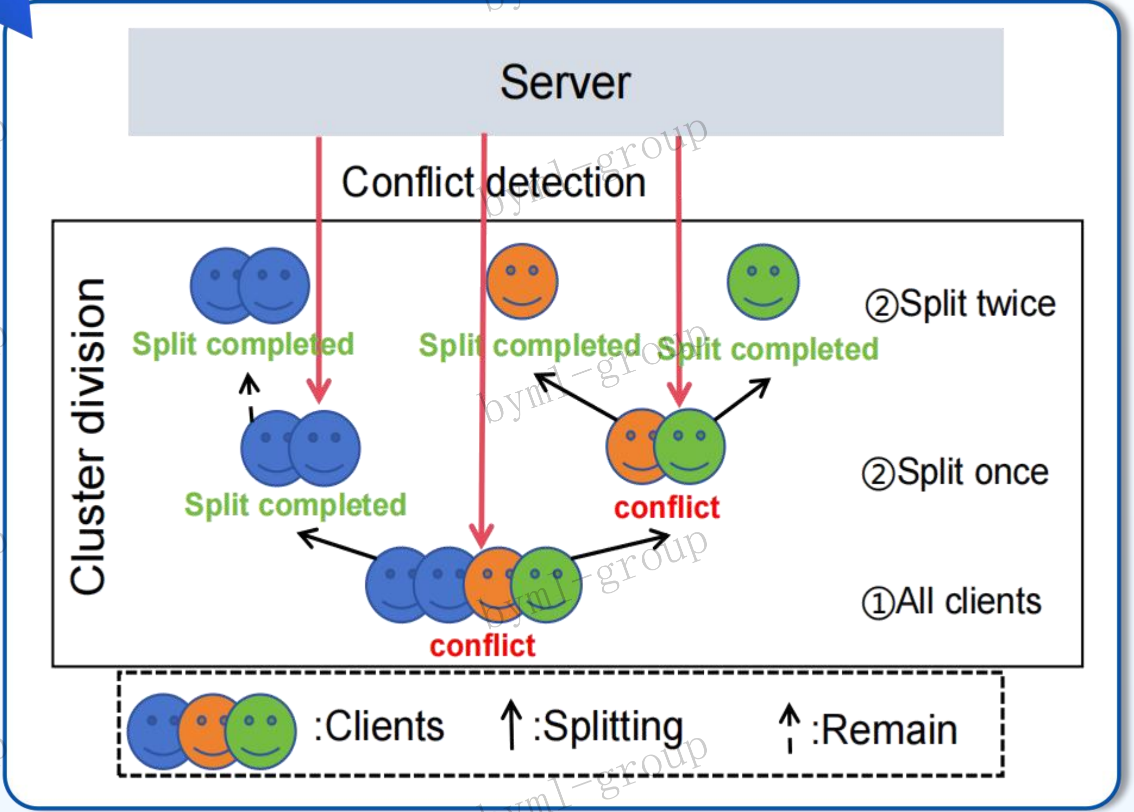
- **Cell Division – A single cell splits into two based on internal conditions.**

Start *All clients belong to one cluster*

Evolve *Performance conflict evaluation*
⇨ Cluster automatic split

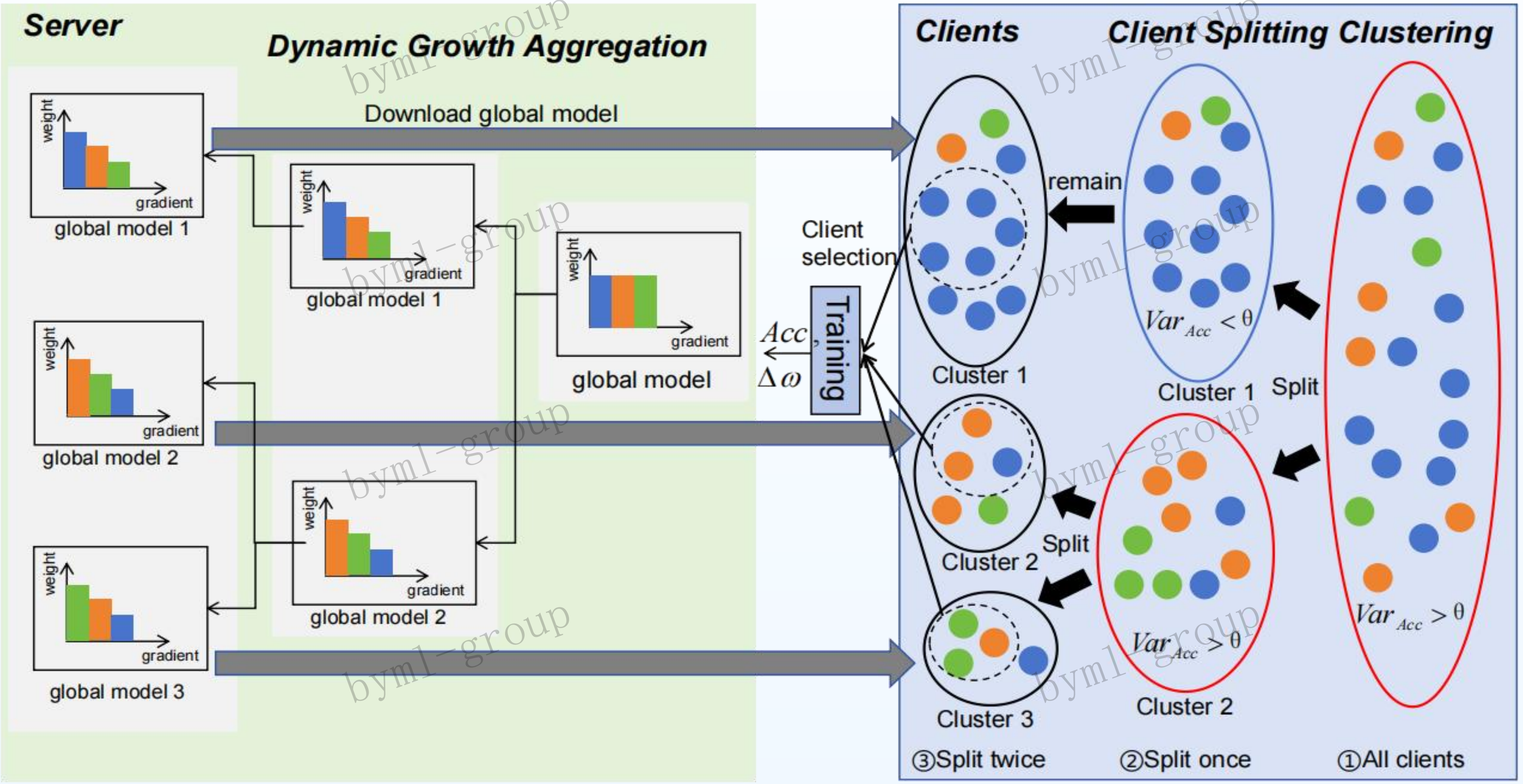
End *Multiple clusters, each with low internal conflict*

Self-determined K



Achieves fairness without manual intervention while preserving generalization.

FairFed++: Overview



Server-Client Iterative Learning

Server

Dynamic Growth Aggregation (DGA)

Aggregates a specific global model for each cluster.

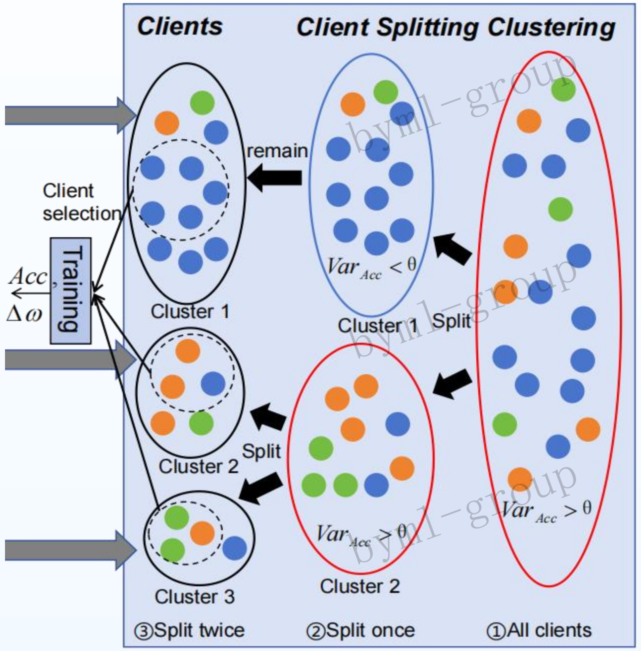
Client

Client Splitting Clustering (CSC)

Checks intra-cluster performance variance and splits clusters if conflicts are detected.

FairFed++ is a dynamic, evolving process to close the fairness gap in FL

FairFed++: Client Splitting Clustering



- Perform conflict detection based on clients' performance and automatically split conflicting clusters
- Self-adaptive reference for conflict detection

$G^t = \{S_1^t, S_2^t, \ldots\}$ $\quad S_c^t = \{C_1, C_2, \ldots C_n\}$

set of clusters *clients in each cluster*

Warm-up

$\{\mathrm{Var}^{l-1}, \ldots, \mathrm{Var}^{l-K}\}$

Window set

θ: the k-th element

Conflict Detection

$Var_c^t > \theta$ $\quad Var_c^t \leq \theta$

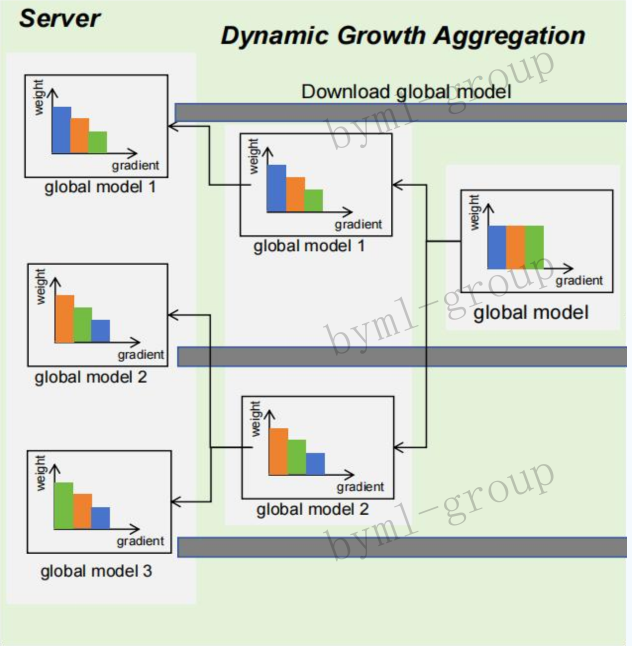
$$S_c^{t+1} = \{C_1, C_2, \ldots C_{\lceil \frac{k}{2} \rceil}\}$$

$$S_{c+1}^{t+1} = \{C_{\lceil \frac{k}{2} \rceil}, C_{\lceil \frac{k}{2} \rceil + 1}, \ldots C_k\}$$

Split

$$S_c^{t+1} = \{C_1, C_2, \ldots C_k\}$$

FairFed++: Dynamic Growth Aggregation



- Iteration ↑ ⇨ Cluster split ⇨ Number of aggregation model ↑
- Collaboratively update all global model with full consideration of the entire FL system

$$\omega_c^t = \omega_c^{t-1} + \underbrace{\sum_{i=1}^{C_t}}_{\text{all clusters}} \underbrace{\sum_{j=1}^{k_i}}_{\text{all clients within each cluster}} weight_{cij}^t \, \Delta\omega_{ij}^t$$

Dual-perspective weighting scheme

Performance Preference

Accuracy↑->Weights↑

$$T_{cij}^t = \frac{Acc_{ij}^t}{\sum_{i=1}^{C_t}\sum_{j=1}^{k_i} Acc_{ij}^t}$$

Distance Preference

Cluster Distance↓->Weights↑

$$P_{ci}^t = \frac{Dis_{ci}^t}{\sum_{c=1}^{C_t} Dis_{ci}^t}$$

$$weight_{cij}^t = \gamma T_{cij}^t + \eta P_{ci}^t$$

Experiments: Setup

Datasets (6 benchmarks)

Name	Sample	Devices	round/warmup
Vehicle	32764	23	800/10
Adult	32550	30	800/10
Synthetic	11379	100	20000/100
Sent140	46059	1101	200/10
Fashion-MNIST	61251	61251	800/10
FEMNIST	24537	24537	5000/20

Evaluation Metrics

- **Performance:** Average accuracy (↑)
- **Fairness:** Accuracy variance across all clients (↓)
- **Generalization ability:**
  - Accuracy on the best 10% clients (↑)
  - Accuracy on the worst 10% clients (↑)

FL protocol

- **Train/Validation/Test:** 8:1:1
- **Warmup: FedAvg**
- **Window Size = 10**
- **Activated clients per round:** 10

ecai 2025 BOLOGNA

Experiments: Main Result & Analysis

Methods	Vehicle Accuracy	Vehicle Best 10%	Vehicle Worst 10%	Vehicle Variance	Adult Accuracy	Adult Best 10%	Adult Worst 10%	Adult Variance
FedAvg [19]	87.37	96.76	42.79	302.04	83.60	87.57	77.87	8.73
q-FFL [14]	87.69	96.01	64.45	80.14	83.66	87.57	77.87	8.67
AFL [20]	88.79	97.02	66.62	75.07	82.72	86.96	77.27	8.15
FedFa [11]	87.28	95.30	48.32	233.58	83.66	87.57	77.87	8.07
FedFV [28]	86.39	96.01	45.45	262.64	83.57	87.57	77.87	8.37
FedGini [16]	87.31	96.74	42.76	194.81	83.27	87.57	77.87	8.68
FairDPFL-SCS [25]	86.89	95.84	43.55	274.43	83.63	87.57	77.87	8.70
FairFed++	**91.25**	**97.77**	**69.61**	**66.93**	**83.72**	**87.57**	**78.48**	**7.40**

Methods	Synthetic Accuracy	Synthetic Best 10%	Synthetic Worst 10%	Synthetic Variance	Sent140 Accuracy	Sent140 Best 10%	Sent140 Worst 10%	Sent140 Variance
FedAvg [19]	75.56	100	1.11	1176.05	63.11	100	15.81	670.61
q-FFL [14]	81.48	100	19.01	750.54	63.94	100	18.48	609.35
AFL [20]	84.46	100	31.03	511.40	49.51	100	0	1473.24
FedFa [11]	77.46	100	8.19	921.40	66.44	100	21.34	551.09
FedFV [28]	82.62	100	18.70	703.41	65.45	100	19.82	583.91
FedGini [16]	79.36	100	16.45	794.65	64.96	100	18.14	612.93
FairDPFL-SCS [25]	82.16	100	18.37	683.87	63.32	100	16.58	584.01
FairFed++	**88.77**	**100**	**48.88**	**298.14**	**70.83**	**100**	**27.27**	**550.58**

Methods	Fashion MNIST Accuracy	Fashion MNIST Best 10%	Fashion MNIST Worst 10%	Fashion MNIST Variance	FEMNIST Accuracy	FEMNIST Best 10%	FEMNIST Worst 10%	FEMNIST Variance
FedAvg [19]	84.30	100	51.07	224.17	59.19	89.20	13.67	455.06
q-FFL [14]	87.50	100	62.05	141.10	59.57	90.51	20.28	398.88
AFL [20]	86.84	100	60.48	147.41	52.82	85.09	13.52	436.19
FedFa [11]	88.27	100	61.16	152.82	61.53	90.14	21.48	378.04
FedFV [28]	87.52	100	57.69	174.85	60.67	89.57	19.98	375.57
FedGini [16]	84.77	100	57.79	161.76	52.94	84.53	13.54	393.85
FairDPFL-SCS [25]	87.98	100	58.15	182.72	57.14	89.84	14.98	399.74
FairFed++	**90.26**	**100**	**67.85**	**106.81**	**66.44**	**95.12**	**28.75**	**342.65**

Key Findings

Superior pecrformance

Highest average accuracy over 6 datasets

Significant improvement

Up to 4.9% increase over existing methods

Conspicuous fairness

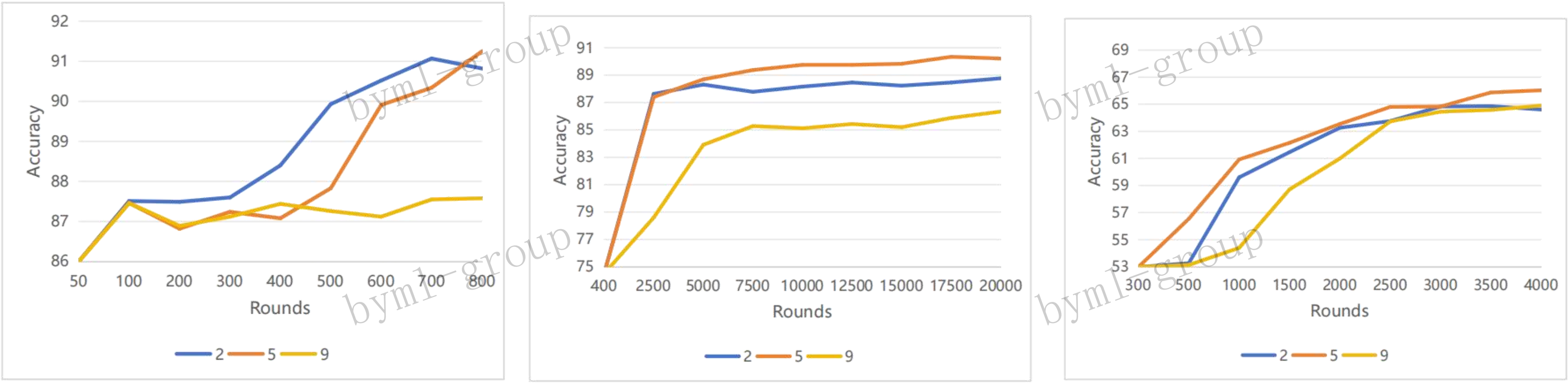
Variance reduced to 1/4 of FedAvg

Comprehensive evaluation

Consistently perform best across all metrics

FairFed++ achieves a superior balance on fairness and performance.

Experiments: Hyperparameter Analysis



(a) vehicle (b) synthetic (c) FEMNIST

Dataset	Position	Accuracy	Best 10%	Worst 10%	Variance	Cluster Number
Vehicle	2	90.82	97.28	68.73	71.33	13
	5	**91.25**	**97.77**	**69.61**	**66.93**	12
	9	87.58	97.75	42.23	310.39	2
Synthetic	2	89.07	100	49.46	290.01	28
	5	**90.21**	**100**	**55.73**	**214.18**	21
	9	86.34	100	38.23	423.43	2
FEMNIST	2	65.75	94.06	27.23	353.98	101
	5	**66.44**	**95.12**	**28.75**	**342.65**	88
	9	64.71	91.89	24.74	360.74	48

Performance with the change of split threshold θ

- **Small θ (Position 2)**

High standard to meet non-confliction

⇨Too many clusters,
hurts generalization and stability.

- **Large θ (Position 9)**

Low standard to meet non-confliction

⇨Too few clusters (large variance)
insufficient fairness guarantee

- **Moderate θ (Position 5)**

Best balance performance and fairness

FairFed++ exhibits robust performance under a moderate split threshold

Conclusion & Future Work

Key Contributions

- **Self-evolving optimization scheme:** Self-adaptively determine the clustering process according to the actual environment
- **FairFed++:** A novel framework that automatically partitions clusters based on intra-cluster conflicts and dynamic growth for model aggregation.
- **SOTA performance:** Achieved the best results (performance and fairness) on 6 environments

Future Work

- **Complicated Environment:** Explore the potential of **FairFed++** in cross-device FL with a massive number of clients
- **Cooperation:** Explore the potential for integration with other FL frameworks
- **Theorety:** Explore the theoretical guarantees for the convergence and fairness of **FairFed++**

ecai 2025 BOLOGNA

ECAI2025

Thank you for your attention

Zhixiang Fang, Baoyao Yang, Weide Zhan Yanchao Tang and Yiqun Zhang*

廣東工業大學
Guangdong University of Technology

code available

Q&A Welcome questions about our research

ybaoyao@gdut.edu.com