香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

SHENZHEN INSTITUTE of ARTIFICIAL INTELLIGENCE AND ROBOTICS for SOCIETY
深圳市人工智能与机器人研究院

# FedLF: Layer-Wise Fair Federated Learning

Zibin Pan[1,2], Chi Li[1,4], Fangchen Yu[1], Shuyi Wang[1,2], Haijin Wang[1], Xiaoying Tang* [1,2,3], Junhua Zhao* [1,2]

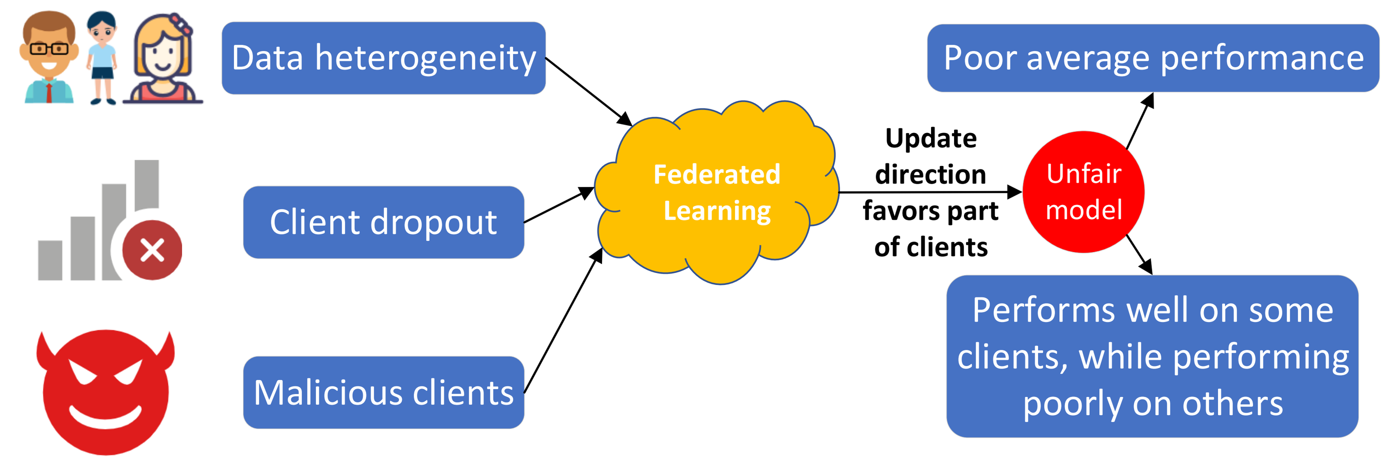
[1] SSE, The Chinese University of Hong Kong, Shenzhen; [2] The Shenzhen Institute of Artificial Intelligence and Robotics for Society;
[3] The Guangdong Provincial Key Laboratory of Future Networks of Intelligence; [4] Shenzhen Research Institute of Big Data

zibinpan@link.cuhk.edu.cn, chili@link.cuhk.edu.cn, fangchenyu@link.cuhk.edu.cn, shuyiwang@link.cuhk.edu.cn, haijinwang@link.cuhk.edu.cn,
tangxiaoying@cuhk.edu.cn, zhaojunhua@cuhk.edu.cn

AAAI-24

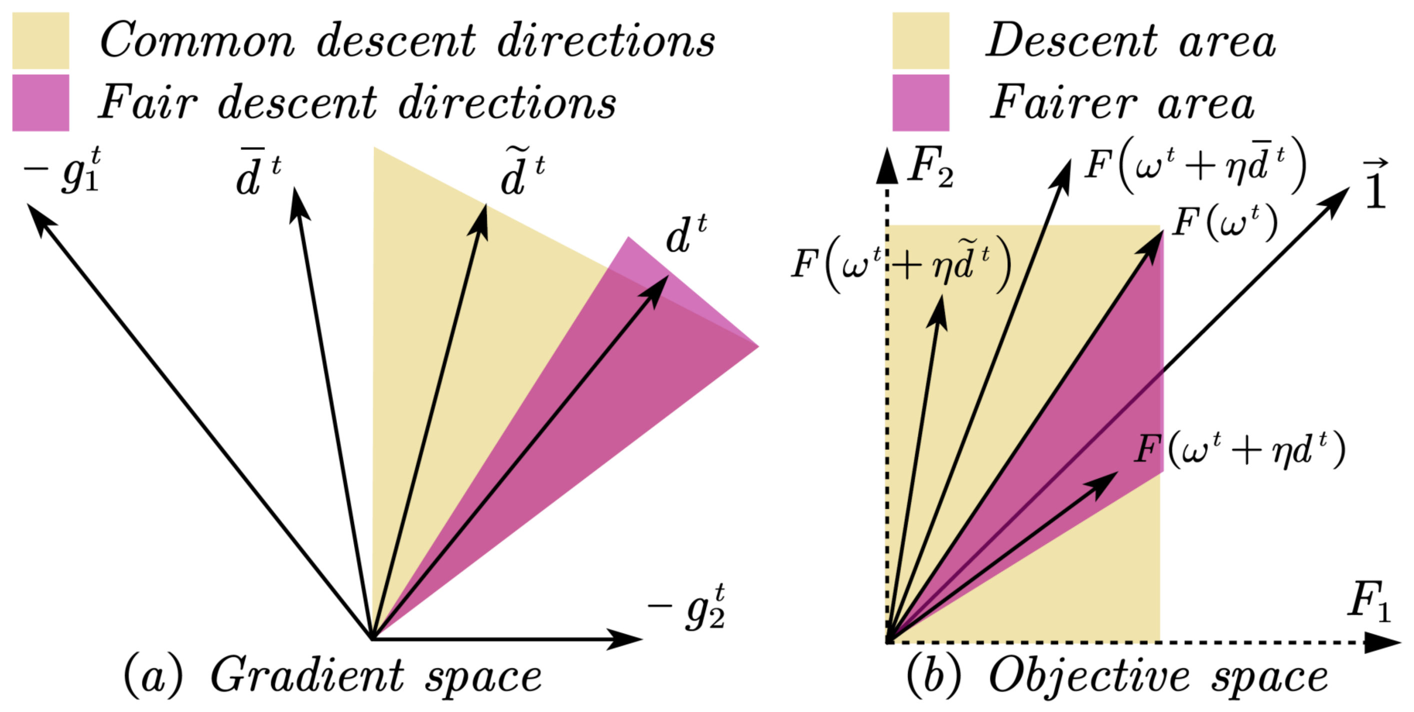
## Background & Motivation

### What Makes Federated Learning Unfair?



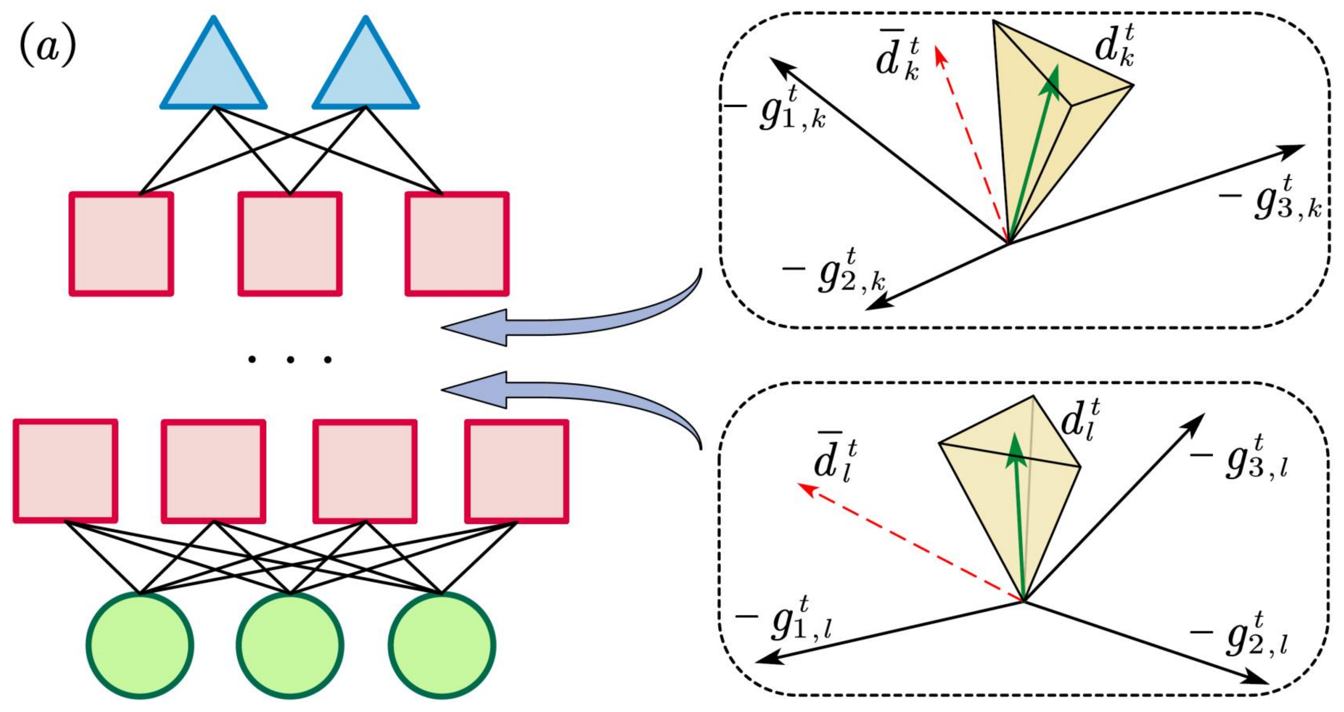

### Challenges of computing a fair update direction in FL

1. **Model-level gradient conflict.**
2. **Improvement bias.**




(a) Gradient space (b) Objective space

3. **Layer-level gradient conflict.**




## The proposed FedLF

### Problem formulation with a fair-driven objective

$$\min_{\omega}\left(F_1(\omega), F_2(\omega), \cdots, F_m(\omega), P(\omega)\right),$$

where $P(\omega) = -\cos\left(\vec{1}, F(\omega)\right)$.

### Layer-wise Multiple Gradient Descent Algorithm (LMGDA)

$$d_l^t, \alpha_l^t = \underset{d_l^t \in \mathbb{R}^{n_l}, \alpha_l^t \in \mathbb{R}}{\operatorname{argmin}} \ \alpha_l^t + \frac{1}{2}\|d_l^t\|^2,$$

$$\text{s.t.}\ g_{i,l}^t \cdot d_l^t \le \alpha_l^t, \forall i = 1, \cdots, m,$$

$$g_{P,l}^t \cdot d_l^t \le \alpha_l^t.$$

### Scalable method to obtain $d_l^t$

$$d_l^t = -\left(\sum_{i=1}^{m} \lambda_i g_{i,l}^t + \mu g_{P,l}^t\right), \sum_{i=1}^{m} \lambda_i + \mu = 1,$$

where $\lambda_1, \cdots, \lambda_m, \mu \ge 0$ is the optimum of the dual problem:

$$\max_{\lambda_i, \mu} -\frac{1}{2}\left\|\sum_{i=1}^{m} \lambda_i g_{i,l}^t + \mu g_{P,l}^t\right\|^2$$

$$s.t. \sum_{i=1}^{m} \lambda_i + \mu = 1,$$

$$\lambda_i, \mu \ge 0, \forall i = 1, 2, \cdots, m.$$

### Combine layers

If $d_l^t = \vec{0}$ is satisfied for part of $l \in L$, then combine layer $l$ with its neighbor and recalculate $d_l^t$.

### The obtained direction $d^t$ satisfies:

1. If $\omega^t$ is Pareto stationary, then $d^t = \vec{0}$.
2. If $\omega^t$ is not Pareto stationary, then

$$g_i^t \cdot d^t < 0, \forall i;\ g_{i,l}^t \cdot d_l^t < 0, \forall l;\ g_P^t \cdot d^t < 0.$$

### Improve absent client fairness

Take into account those absent clients who were online from $t - \tau$ to $t - 1$, where $\tau = M/|S^t|$. $M$ is the number of recorded clients that have already joined in FL, and $S^t$ is a set of online clients at communication round $t$.

## Experiments

### Metrics:

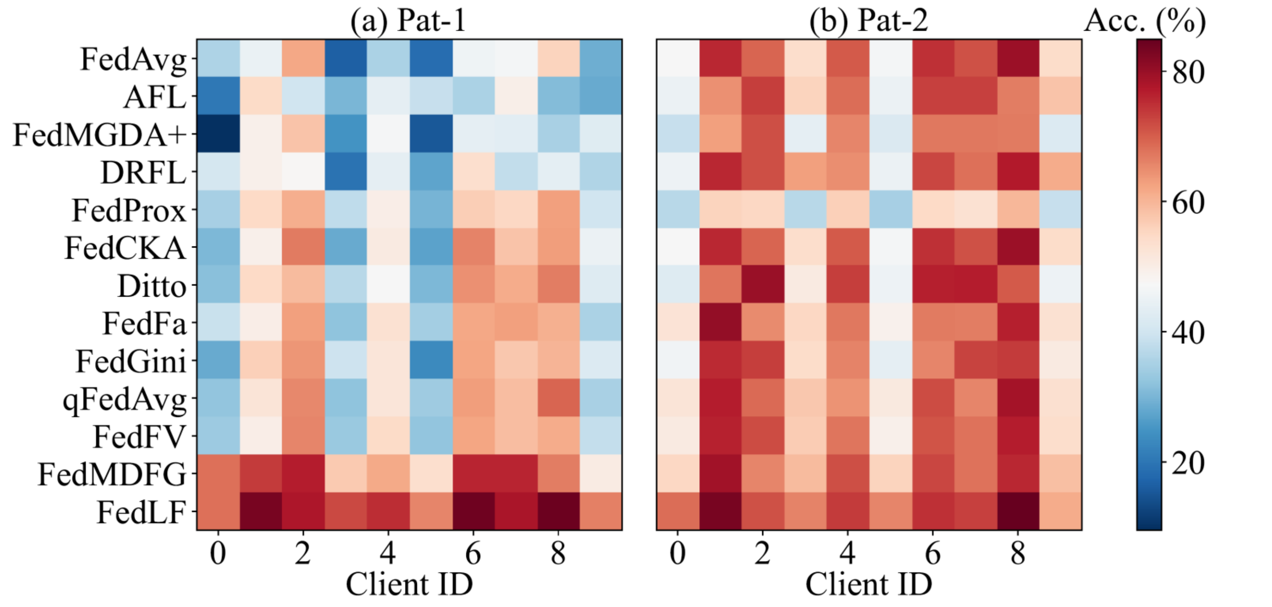
- Performance: model average test accuracy across clients.
- Fairness indicator: $\arccos\left(\frac{A(\omega)\cdot\vec{1}}{\|A(\omega)\|\|\vec{1}\|}\right)$, where $A(\omega)$ denotes a vector that contains the test accuracy of the global model on each client.

### Performance and fairness

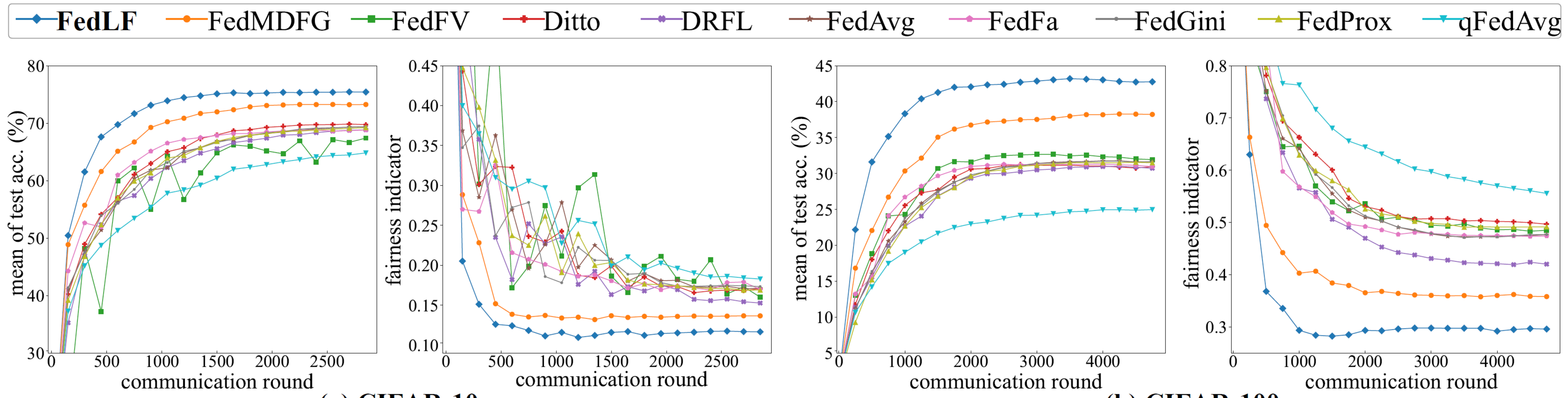
Result 1: The average test accuracy of all clients (and the fairness indicator)

| | FMNIST | | | CIFAR-10 | | | CIFAR-100 | | |
|---|---|---|---|---|---|---|---|---|---|
| Algorithm | Dir(0.1) | Pat-1 | Pat-2 | Dir(0.1) | Pat-1 | Pat-2 | Dir(0.1) | Pat-1 | Pat-2 |
| FedAvg | .861(.116) | .828(.170) | .838(.135) | .690(.214) | .575(.341) | .681(.276) | .343(.201) | .199(.667) | .222(.604) |
| qFedAvg | .847(.133) | .831(.161) | .813(.118) | .681(.204) | .565(.301) | .661(.267) | .344(.196) | .183(.690) | .238(.529) |
| FedProx | .825(.121) | .834(.142) | .836(.105) | .544(.242) | .572(.205) | .566(.212) | .197(.291) | .207(.617) | .201(.538) |
| AFL | .865(.109) | .829(.204) | .854(.137) | .679(.201) | .561(.251) | .685(.202) | .382(.181) | .177(.753) | .261(.509) |
| Ditto | .820(.129) | .749(.278) | .815(.124) | .598(.216) | .463(.240) | .553(.251) | .301(.241) | .070(1.08) | .114(.784) |
| FedFV | .850(.132) | .836(.165) | .853(.135) | .682(.208) | .568(.376) | .681(.204) | .339(.198) | .191(.664) | .229(.558) |
| DRFL | .861(.109) | .855(.136) | .847(.157) | .692(.190) | .578(.307) | .684(.270) | .341(.201) | .193(.644) | .228(.540) |
| FedFa | .844(.174) | .815(.205) | .836(.116) | .653(.244) | .482(.297) | .695(.232) | .387(.189) | .114(1.09) | .222(.776) |
| FedGini | .867(.115) | .839(.160) | .837(.134) | .698(.195) | .587(.315) | .672(.246) | .349(.191) | .203(.621) | .198(.666) |
| FedCKA | .861(.117) | .816(.227) | .840(.129) | .690(.205) | .575(.341) | .674(.211) | .344(.201) | .190(.691) | .222(.575) |
| FedMGDA+ | .809(.161) | .750(.305) | .815(.221) | .531(.264) | .440(.314) | .569(.282) | .173(.358) | .035(1.15) | .080(.839) |
| FedMDFG | .873(.089) | .863(.101) | .874(.084) | .729(.176) | .744(.142) | .714(.153) | .387(.181) | .278(.485) | .332(.387) |
| FedLF | **.892(.084)** | **.894(.089)** | **.898(.074)** | **.766(.140)** | **.765(.126)** | **.761(.127)** | **.420(.158)** | **.409(.347)** | **.413(.305)** |

Result 2: The test accuracy of 10 clients




### Accuracy and efficiency




(a) CIFAR-10 (b) CIFAR-100

## Conclusion

1. **We identify three significant challenges that exist in computing a fair direction for the FL model update.**
2. **We design an effective fair-driven objective to drive the FL model fairer.**
3. **We are the first to propose the layer-wise multiple gradient descent algorithm (LMGDA) and adopt it to determine a layer-wise fair direction for FL model update.**
4. **We theoretically and empirically verified that FedLF outperforms SOTA in terms of performance and fairness.**