FCFL: A Fairness Compensation-based Federated Learning Scheme with Accumulated Queues

Lingfu Wang[1], Zuobin Xiong[2], Guangchun Luo[1], Wei Li[3], and Aiguo Chen[4] ✉

[1] School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2] Department of Computer Science, University of Nevada-Las Vegas, Las Vegas, USA
[3] Department of Computer Science, Georgia State University, Georgia, USA
[4] Laboratory of Intelligent Collaborative Computing, University of Electronic Science and Technology of China, Chengdu, China
agchen@uestc.edu.cn

Abstract. The surge of ubiquitous data underscores the need for Federated learning (FL), which allows distributed data entities to learn a global model orchestrally without revealing their private local data, ensuring the privacy and security of users. However, the performance of the trained global model on individual clients is impaired by the heterogeneous nature of the client's local data, exposed as the performance unfairness in FL. Such unfairness issues grab the research community's attention and a few recent works embark upon fair solutions via reweighting clients during aggregation but overlooking the impact of client selection for aggregation. To fill this gap, in this paper, a Fairness Compensation-based FL scheme (`FCFL`) is proposed to alleviate the unfairness amongst clients. In particular, the unfairness of each client during the FL training process is estimated as the accuracy difference between local performance and global performance, and accumulated queues are calculated for the cumulative unfairness value in each round. In addition, a fairness compensation FL method is devised, which can select participating clients dynamically and adjust the aggregation weights adaptively in each round to guarantee fairness in the training process. Specifically, the proposed `FCFL` scheme is a flexible framework with tunable parameters and the FedAvg algorithm is its special case when α=0. Finally, intensive experiments are conducted on two benchmark datasets with different settings, demonstrating that the `FCFL` outperforms the state-of-the-art baselines by improving the fairness metric up to 30.4% while maintaining a competitive accuracy performance. The source code is available at `https://github.com/wlffffff/FCFL`.

Keywords: Federated learning · Performance fairness · Data heterogeneity · Client selection · Weighting strategy.

1 Introduction

The advancement of Artificial Intelligence (AI) is driven by the ubiquitous data generated from a wealth of devices, however, conventional machine learning typically adopts a centralized mode in data collection and training, which poses significant challenges in multiple faucets. Firstly, uploading large amounts of data incurs communication and storage costs [14,27] of burdened infrastructure. Secondly, with the ever-increasing privacy and security concerns [26], it is hard to convince unwilling data owners to share their raw data with an untrusted service provider under this centralized paradigm [24]. Besides, pressing regulations and laws are enforced by many governments on private data with stricter data management and stewardship, such as the General Data Protection Regulations (GDPR) from the European Union and Personal Information Protection and Electronic Documents Act (PIPEDA) from Canada. To solve this dilemma, Federated Learning (FL), as a promising solution is proposed by Google recently [15]. FL is a distributed machine learning paradigm consisting of a server and multiple clients, which can learn a global model on the server side without access to clients' local private data. Since the original data is kept locally rather than being sent to a remote server, the challenges of communication, storage, and privacy are somehow mitigated by FL, and thus broad application scenarios are fertilized, including medical images [9], recommendation systems [22], and the Internet of Things (IoT) [25].

Yet, FL is not the silver bullet, and some issues have emerged in recent years. In this paper, we study the fairness aspect of FL, which is one of the major concerns of FL that impedes the realistic application. As illustrated in Fig. 1, since the global model is trained based on unknown local datasets of clients, where divergence may exist amongst client local data and model. Therefore the performance of the global model may vary across the diverse clients, causing unfair performance (e.g., accuracy) as shown on the right side of Fig. 1. Specifically, though the global model performs well on average, such unfairness is manifested in those clients (referred to as vulnerable clients) who receive lower accuracy due to biased client selection or minority in data representations.

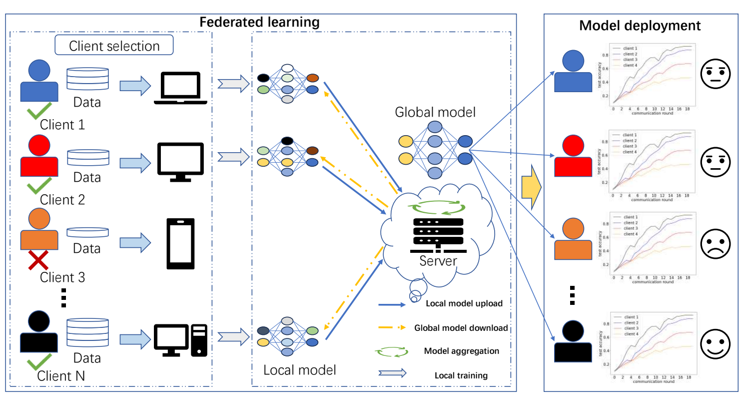


Fig. 1. (Left): Example of horizontal FL. (Right): Performance fairness of FL.

To alleviate unfairness in FL, AFL is proposed by Mohri et al. [16] to optimize the worst-performing client by minimax optimization. Inspired by fair resource allocation in wireless networks, Li et al. [11] presented the q-FedAvg method, which introduces a parameter q to reweight the loss of different clients. Zhao et al. [30] proposed DRFL, which dynamically adjusts the weight assigned to each client and is more flexible in parameter tuning. Although these works can alleviate unfairness by adjusting weights, they often ignore the impact of client selection in fairness. In addition, the client selection in existing methods is either randomly done [10,15] or based on the local data amount [8,28], which may lead to poor local performance of the global model. Therefore, it is a challenge to design a fair FL scheme that handles client selection without incurring a negative impact on the client side.

In this work, we propose the **F**airness **C**ompensation-based **FL** (FCFL) to alleviate unfairness in FL. In each training round, the server updates the unfairness queue of each client, where the queues are utilized for client selection and aggregation reweighting based on cumulative unfairness value. To our best knowledge, this is the first work to harness fair client selection to achieve performance fairness of FL. The contributions of this paper are summarized as follows:

- The accumulated unfairness of clients is defined during the process of FL and unfairness queues are maintained to measure if a client is treated fairly.
- Based on the accumulated unfairness queues, a fairness compensation method FCFL is designed to balance the client selection and aggregation reweighting, which can improve the fairness of FL.
- Evaluations on two datasets are conducted to confirm the advantages of our method and compare it with state-of-the-art methods. The experimental results demonstrate the fairness and effectiveness of our proposed FCFL.

The rest of the paper is organized as follows. The related work on the fairness of FL is discussed in Section 2. Then, the proposed FCFL method is detailed in Section 3 with descriptions of client selection and aggregation reweighting. Empirical evaluation is presented in Section 4, and finally Section 5 concludes the paper.

2 Related Work

Fairness in FL can be divided into collaborative fairness [13,29], group fairness [3, 4], selection fairness [7,18], and performance fairness [11], as per different fairness goals. In particular, collaborative fairness means that clients should be rewarded in proportion to their contributions; group fairness aims to minimize disparities among different groups based on sensitive attributes (e.g., gender and race); selection fairness ensures that each client has a fair chance of being selected to participate in training; and performance fairness seeks to reduce the variance of global model accuracy across clients. This paper mainly focuses on performance fairness. For a comprehensive review of the fairness in FL, please refer to the survey paper [19] by Shi et al.

Table 1. Related work on performance fairness of FL.

References	Method	Dataset
AFL (2019) [16]	Minimax optimization	Fashion MNIST; Adult; Cornell movie; PTB
q-FedAvg (2019) [11]	Reweighting	Synthetic; Vehicle; Sent 140; Shakespeare
FedGini (2023) [12]	Objective function	Synthetic; CIFAR-10; Sent 140
DRFL (2022) [30]	Reweighting	Synthetic; Fashion MNIST; Adult
Ada-FFL (2023) [2]	Reweighting	Synthetic; Vehicle; Sent 140
FedFa (2022) [8]	Reweighting	MNIST; FEMNIST; Synthetic; Sent 140; Shakespeare
PG-FFL (2022) [21]	Reweighting	Fashion MNIST; CIFAR-10; CIFAR-100
FedFV (2021) [23]	Gradient projection	MNIST; Fashion MNIST; CIFAR-10
GIFAIR (2023) [28]	Reweighting; Objective function	FEMNIST; Shakespeare;
FedMGDA (2022) [6]	Multi-objective optimization	Fashion MNIST; CIFAR-10; Shakespeare; Adult
FedMDFG (2023) [17]	Multi-objective optimization	MNIST; Fashion MNIST; CIFAR-10; CIFAR-100
FairWire+ (2024) [5]	Multi-objective optimization	CIFAR-10; CIFAR-100; FEMNIST

2.1 Performance Fairness in FL

The vanilla FedAvg algorithm aggregates client local models by calculating a weighted average based on the amount of training data [15], therefore causing significant differences in model accuracy due to the data heterogeneity of different clients. As countermeasures, AFL [16] is the first approach to improve the fairness of FL, which used minimax optimization to maximize the performance of the worst-performing device. However, this method cannot guarantee generalization in large-scale settings. To improve the scalability of AFL, researchers have proposed the q-FedAvg [11] method by introducing the parameter q for clients reweighting to achieve better fairness. Since q-FedAvg, performance fairness has become a pivotal problem in FL, and many methods have been proposed to improve the fairness of q-FedAvg, including designing novel objective functions, reweighting, eliminating gradient conflicts, and multi-objective optimization.

Designing Novel Objective Functions. FedGini [12] modified the objective function to improve fairness by introducing a Gini penalty term. GIFAIR [28] achieved fairness by introducing a regularization term to penalize loss differences among client groups. Ada-FFL [2] improved the objective function of q-FedAvg by introducing regularized local loss terms and Frobenius distance to design an adaptive fair FL.

Reweighting. FedFa [8] combined the training accuracy and frequency to design an appropriate weight selection algorithm and adopted double momentum gradient optimization to accelerate the model's convergence. PG-FFL [21] used reinforcement learning to achieve reweighting, and automatically learned strategies through a reward function constructed based on Gini coefficients and accuracy. Building on the q-FedAvg method, DRFL [30] proposed a novel approach, which can dynamically adjust the weight assigned to clients. The regularization term in GIFAIR [28] can also be viewed as a dynamic client reweighting technique that can adaptively assign higher weights to clients with poor performance.

Eliminating Gradient Conflicts. Researchers have found that conflicting gradients are one of the reasons for unfairness in FL. To address this issue, FedFV [23] first used cosine similarity to detect gradient conflicts, and then iteratively eliminated conflicts by modifying the direction and magnitude of gradients, thereby improving the fairness of FL.

Multi-objective Optimization. FedMGDA [6] pioneered the formalization of FL into multi-objective optimization and proposed a novel fair FL scheme using a multiple gradient descent algorithm. Under the guidance of multi-objective optimization, FedMDFG [17] and FairWire+ [5] are also proposed. FedMDFG can find a fair descent direction by adding a fair-driven objective, and the line search strategy can ensure an appropriate step size. These two major designs guarantee the fairness and robustness of the scheme. FairWire+ considered the inherent noise induced by wireless channels and designed an algorithm based on noisy gradients, which can find a common descent direction for all clients. We summarize the current fair FL for performance fairness in Table 1.

2.2 Client Selection in FL

Although the above methods can relieve unfairness through various approaches, they ignore the possible impact of client selection on the fairness of FL. Client selection is also an important research topic, which can achieve different goals. For instance, Power-of-Choice [1] identified clients with the highest loss in each round and included them in training to boost model performance. GreedyFed [20] selected clients with the highest contribution based on the Shapley value, improving model accuracy and convergence speed. By utilizing Lyapunov optimization, FairFedCS [18] achieved better selection fairness. In this vein, we aim to study how client selection can improve performance fairness in this work.

3 Fairness Compensation Federated Learning (FCFL)

In this section, we present the original fair FL method, FCFL, with problem setting, proposed algorithms, and analysis.

3.1 Problem Setting

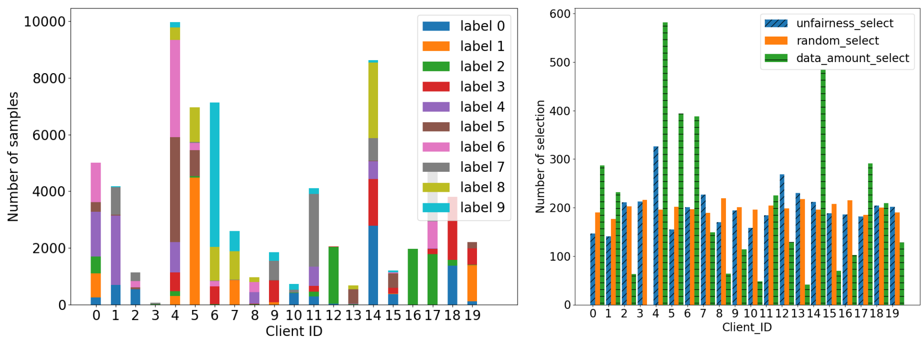
As shown in Fig. 1, general FL has the following three steps: (1) client selection, (2) local training, and (3) weighted aggregation. Since the distribution of local datasets is different, the performance of the global model varies notably across different clients. This phenomenon is referred to as performance fairness. To be precise, the fairness of the global model can be defined as follows.

Definition 1. *(Fairness of performance distribution [11]). A model θ_1 is said to be fairer than θ_2 if the accuracy of θ_1 on the N clients $\{a_1, a_2, ..., a_N\}$ is more* ***uniform*** *than that of θ_2 on the N clients.*

In this work, we use the variance of the accuracy on all clients as a measure of fairness, and the goal of our work is to reduce the variance while maintaining a similar average accuracy of the global model. To achieve this goal, client selection is a crucial aspect but is overlooked by most existing works. Intuitively, FL selects clients to participate in training based on the amount of local data [8,28], which leads to the global model being biased towards clients with more data. We visualize this issue on the MNIST dataset and the CNN model in Fig. 2. Here, the Dirichlet function is used to partition the dataset into 20 clients, where 2 clients are selected for each training round. Fig. 2(a) shows the distribution of local data for each client, and it can be seen that the amount and classes of data on each client differ significantly (such as clients 3, 10, 13), which simulates the data distribution in real-world scenarios. Using the data amount-based client selection method will reduce the selected times of these vulnerable clients in training (as can be seen from the green bin in Fig. 2(b)). Random selection has an equal chance of selecting each client, but it produces a poor performance of the global model on vulnerable clients. Our unfairness-based selection prioritizes the selection of vulnerable clients and finally can achieve performance fairness. Taking client 3 as an example, it is rarely selected in the amount-based method due to the limited local data, while our unfairness-selection method selects clients based on unfairness in each round, with client 3 being selected significantly more frequently. Note that unfairness-based selection selects clients based on cumulative unfairness, so it may be possible for vulnerable clients to have fewer choices than random ones. However, our proposed weight allocation based on cumulative unfairness will give vulnerable clients more weight, thereby improving the fairness of the scheme.

3.2 Overview of FCFL

To achieve the goal of fairness, we first investigate the reasons for performance fairness in FL. As demonstrated above, due to the data heterogeneity, some vulnerable clients cannot be selected fairly, such as client 3 in Fig. 2(b). To alleviate unfairness, an intuitive approach is to compensate for these vulnerable clients based on their unfairness level. To maintain the clients' computation cost, we focus on improving the selection ratio and assigning more aggregation weights for vulnerable clients. Based on this idea, we propose FCFL which considers both client selection and aggregation reweighting in the FL progress. The framework of FCFL is depicted in Fig. 3. In each training round, all clients first upload the local performance (i.e., the accuracy of the global model on local clients) to the server, and then the server updates the accumulated unfairness queue to select



(a) Data distribution of each client (b) The numbers of being selected

Fig. 2. The impact of data distribution on client selection methods.

clients and calculate aggregation weights. Next, the selected clients perform local training and upload the local model and accuracy. Finally, the server completes the aggregation and estimates the global performance for the next round of FL training. Comprehensive explanations of this approach will be provided in the subsequent sections.

3.3 Accumulated Unfairness Queues

In FL, performance fairness is evaluated by the accuracy differences of the global model on different local clients. However, the global model is not available during the training process until aggregation is performed. Therefore, to approximate the unfairness of clients in training rounds, the unfairness level is measured by the difference between the estimated global model accuracy and evaluated local accuracy, which can be calculated as follows.

$$uf_i^t = \begin{cases} \widetilde{Acc}^t - Acc_i^t, & if \ \ \widetilde{Acc}^t > Acc_i^t \\ 0, & otherwise \end{cases} \tag{1}$$

where Acc_i^t represents the evaluated local accuracy of client i in round t and $\widetilde{Acc}^t = \sum_{i=1}^{m} \omega_i^t \widetilde{Acc}_i^t$ is the estimated global model accuracy in round t, in which $\widetilde{Acc}_i^t$ denotes the local training accuracy of client i in round t and m is the number of selected clients. In this paper, we assume that the server does not have access to validation data, which is realistic in the actual applications, so the performance of the global model can only be obtained through estimation. uf_i^t is the unfairness level of the client i in round t, which is a cumulative value that reflects whether the client i has been treated fairly or not so far and indicates the priority of each client to be selected for training by the FCFL algorithm.

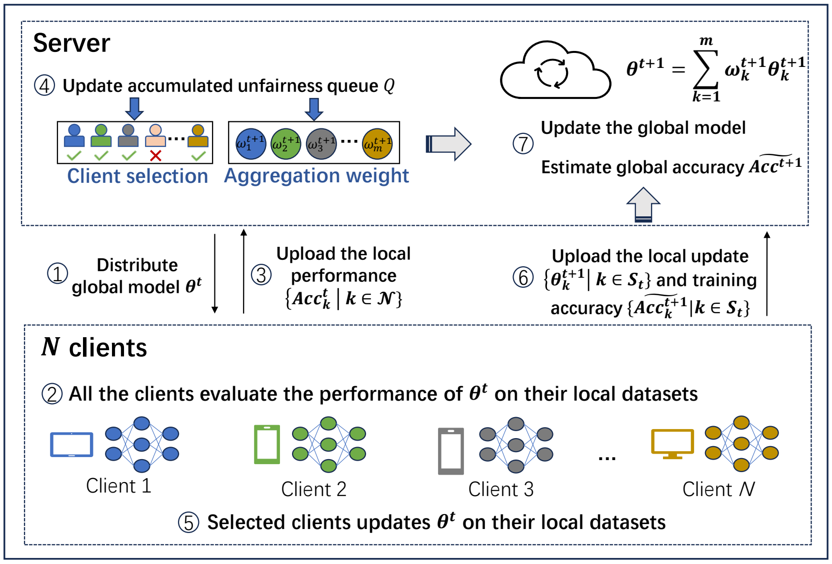


Fig. 3. The framework of FCFL.

To track the cumulative unfairness of all clients during training, we introduce a queue $Q_i(t)$ to store the unfairness value of each client i in round t as described in the following formula. The intuition of this queue is to track the cumulative unfairness of each participant, providing a basis for subsequent client selection and weight allocation.

$$Q_i(t) = \max\left\{Q_i(t-1) + \alpha uf_i^t - \omega_i^t \cdot \mathbf{1}_{[x_i(t-1)=1]}, 0\right\}, \tag{2}$$

where $x_i(t-1) \in \{0,1\}$ indicates whether client i has been selected in the $(t-1)$-th round (1=yes, 0=no). $\mathbf{1}_{[condition]}$ is an indicator function, which equals 1 if [condition] is true and 0 if not. α is a hyper-parameter that controls fairness. The design rationale of $Q_i(t)$ are as follows to facilitate vulnerable clients:

- For clients with **low-accuracy**, $Q_i(t)$ is a cumulative value that reflects the overall unfairness level in each round.
- For clients who are **not selected**, the indicator function $\mathbf{1}_{[condition]}$ is 0, providing more unfairness increment by αuf_i^t.
- For clients with **small weight**, the cumulative unfairness value will have a small penalty $-\omega_i^t$, resulting in a relatively high $Q_i(t)$.

Specifically, a client who has not been treated fairly (low-accuracy, not selected, or small weight) will have a higher $Q_i(t)$, and our FCFL algorithm will compensate these clients based on $Q_i(t)$ by client selection and aggregation reweighting later.

3.4 Client Selection and Aggregation Reweighting

Client selection is the first crucial step towards fair FL. After updating the accumulated unfairness queue, the server can select the top-m clients with the highest $Q_i(t)$ from the overall clients set $\boldsymbol{N}$ to perform local training. This step guarantees that the vulnerable clients with higher cumulative unfairness level will secure their chance to be selected by the FL system. After client selection and local training, the other crucial step is to allocate more weight for these clients to improve their contribution in aggregation. The adjusted aggregation weight is calculated as follows:

$$\omega_i^{t+1} = \begin{cases} \dfrac{n_i}{\sum_{i=1}^{m} n_i}, & if \quad Q_1(t) = ... = Q_m(t) = 0 \\ \\ \dfrac{Q_i(t)}{\sum_{i=1}^{m} Q_i(t)}, & otherwise \end{cases} \tag{3}$$

where n_i is the data amount of client i and $Q_i(t)$ is the cumulative unfairness value. When the cumulative unfairness of all selected clients is 0, the aggregation weight is proportional to the local data amount, which is used to initialize weight in the beginning. Otherwise, the higher the client's unfairness value, the higher the aggregation weight it gets, which ensures that vulnerable clients have a greater influence on the aggregated global model.

Remark. It is important to note that selecting clients based solely on unfairness may lead to the global model being biased towards vulnerable clients with rare datasets, which can be another form of unfairness and ultimately lead to a decrease in the average accuracy of the global model across all clients. To address this problem, we introduce a hyper-parameter r to balance the two kinds of unfairness. This is achieved by using a random selection method to select a portion of r clients and using our client selection method to select the remaining. By doing so, our FCFL can proactively improve fairness while ensuring global accuracy. The influence of the hyper-parameter r is evaluated as well in section 4.4.

Algorithm 1 Fairness Compensation Federated Learning (FCFL)

Input: clients set $\boldsymbol{N}$; communication rounds T; local epochs E; learning rate η
Output: global model θ^{T+1}
  Server executes:
  initialize global model θ^0
  for each round $t = 0, 1, 2..., T$ **do**
    distribute global model θ^t to all clients
    all client evaluate θ^t and upload local accuracy $\{Acc_i^t | i \in \boldsymbol{N}\}$ to the server
    if $t == 0$ **then**
      initialize unfairness queues: $Q_1(0)=Q_2(0)=...=Q_N(0)=0$
      select m clients to constitute subset S_t according to the selection method
      initialize the aggregation weight $w_i^{t+1} = n_i / \sum_{i=1}^{m} n_i, i \in S_t$
    else

calculate the unfairness of all clients via **Eq**. (1)
update the accumulated unfairness queue via **Eq**. (2)
select m clients to constitute subset S_t according to the selection method
calculate the aggregation weight via **Eq**. (3):
end if
for each client $i \in S_t$ **do**
$\theta_i^{t+1}, \widetilde{Acc}_i^{t+1} \leftarrow$ **ClientUpdate**(i, θ^t)
end for
update global model parameters:
$$\theta^{t+1} = \sum_{i=1}^{m} \omega_i^{t+1} \theta_i^{t+1}$$
estimate global performance:
$$\widetilde{Acc}^{t+1} = \sum_{i=1}^{m} \omega_i^{t+1} \widetilde{Acc}_i^{t+1}$$
end for
ClientUpdate(i, θ^t)**:** // Run on client i
client $i \in S_t$ updates θ^t for E epochs with step size η to obtain θ_i^{t+1}
client $i \in S_t$ evaluates θ_i^{t+1} on local datasets to obtain $\widetilde{Acc}_i^{t+1}$
client sends θ_i^{t+1} and $\widetilde{Acc}_i^{t+1}$ to the server

---

The process of the proposed FCFL is illustrated in Algorithm 1. In each round, our FCFL selects a set of clients S_t to participate in the training through an additional communication round and determines the aggregation weights based on the unfairness queues. For each client, multiple training steps are performed, and then the updated parameters and local training accuracy are uploaded to the server. Finally, the server aggregates the parameters and estimates the performance of the global model. Note that in 0-th round, since $\{Q_i(t) = 0 | i \in \boldsymbol{N}\}$ is initialized fairly, selecting the top-m clients becomes a random selection method, and the aggregation weight is proportional to the amount of data.

Remark. The FedAvg algorithm can be seen as a special case of our FCFL. When α in Eq. (2) is 0, the cumulative unfairness value of all clients $Q_i(t)$ is 0 in every round. Hence, client selection is random and the aggregation weight is proportional to the amount of data. As α increases, the unfairness uf_i^t imposes more influence in $Q_i(t)$, which will have a higher chance of being selected and receiving higher aggregation weights, thus improving the fairness of FL.

3.5 Analysis of Communication and Computation Overhead

The major bottlenecks of FL are the communication cost between server and edge devices [8] as well as the local client computation power. Compared with the FedAvg algorithm, our FCFL does not introduce too much communication overhead as analyzed below. In Algorithm 1, although we have additional communication from clients, the client only needs to upload the local accuracy Acc_i^t once, which costs 8 more bits per round. This communication is necessary to calculate the unfairness queue and select clients to participate in training. After local training on the client, each client $i \in S_t$ sends θ_i^{t+1} and $\widetilde{Acc}_i^{t+1}$ $\widetilde{Acc}_i^{t+1}$ to the server, where the accuracy $\widetilde{Acc}_k^{t+1}$ is an extra cost for communication but with small size. Overall, our FCFL only involves a few more bits of communication cost per round.

We then analyze the computation overhead of FCFL. On the client side, client devices require additional accuracy calculations (evaluating the performance of the global model on local datasets), and the remaining calculations are the same as in FedAvg. To save local computation cost, instead of evaluating the $\widetilde{Acc}_k^{t+1}$ by going through the entire local dataset, we can use mini-batch samples from the local dataset to obtain an estimated accuracy. On the server side, the calculation of unfairness, cumulative unfairness, aggregated weights, and global performance estimation can be completed through simple arithmetic operations within $O(\boldsymbol{N})$. Considering that the server typically has high computing power, such computational cost is negligible and will not impact the bottleneck to FCFL. Empirical results of FCFL's efficiency are presented in section 4.3.

4 Experiments

In this section, the performance of FCFL is compared with other methods for different perspectives, including fairness 4.2, efficiency 4.3, and hyper-parameter 4.4. In addition, we conduct ablation experiments in section 4.5 to verify the impact of our client selection and reweighting methods.

4.1 Experimental Settings

All experiments are conducted on three public datasets: MNIST and CIFAR-10 with 100 local clients, and Shakespeare with 31 local clients. We only consider non-IID scenarios since heterogeneous non-IID data distribution is the reason for performance fairness. To simulate this scenario, we design three settings to allocate data to the clients. (1) We sort all data samples based on labels and then split them into 200 shards, where each client randomly picks 2 shards without replacement. (2) We utilize the Dirichlet function to set different levels of non-IID local clients [23]. (3) In *The Complete Works of William Shakespeare* [15], each speaking role in each play is treated as a device, We subsample 31 speaking roles following the setting in [11]. We randomly divide the data for each local client into $8:2$ for training and testing, respectively.

Training. Three models, MLP, CNN, and RNN, are adopted for the experiments. For MNIST, we use a CNN which contains two convolutions and maximum pooling. We use an MLP which contains a hidden layer on CIFAR-10. For Shakespeare, we use an RNN model which contains an embedding layer and an LSTM layer. All the code is implemented in PyTorch to simulate a federated network with 1 server and several clients, where 10% of clients are selected for training in each round. The local batch size is 64, the local epoch is 1, the server's momentum factor is 0.5, and the number of communication rounds for MNIST and CIFAR-10 is 2000 and for Shakespeare is 500. Note that a communication round refers to the process of completing a model update through interaction between the server and the clients.

Baselines. We compare FCFL with the classic method FedAvg [15] and various state-of-the-art fairness methods in FL, including q-FedAvg [11], FedFa [8], and GIFAIR [28]. Based on the code provided by their authors, we directly rewrite the code for comparison. The presented results are averaged from 5 runs with different random seeds.

4.2 Fairness of FCFL

We compare the proposed FCFL with four FL algorithms, FedAvg, q-FedAvg, FedFa, and GIFAIR to verify the fairness of our method. The value of q in q-FedAvg is set to $\{0.1, 0.2, 0.5, 1.0, 2.0, 5.0\}$, the value of accuracy weight and frequency weight in FedFa is set to $\{(0.4, 0.6), (0.5, 0.5), (0.6, 0.4)\}$, and the parameter λ in GIFAIR is set to $\{0.3, 0.5, 0.7\}$. We take the best performance of each method for the comparison. The derived variance and accuracy are displayed in Table 2, from which we can see that our FCFL method produces the lowest variance of **11.03** on MNIST, **114.59** on CIFAR-10, and **67.48** on Shakespeare. Moreover, taking the CIFAR-10 dataset as an example, compared with q-FedAvg(q=2.0), FedFa, and GIFAIR, our FCFL reduces the variance by **23.4%**, **30.4%**, **27.7%**, respectively. Similar variance reductions can also be seen on the MNIST and Shakespeare datasets. Since variance is an important metric of fairness, it indicates that our FCFL can achieve the fairest performance among all baselines. In addition, the accuracy of the worst 10% client of FCFL is significantly higher than other experiments. In the CIFAR-10 dataset, compared with the second-best method q-FedAvg (q=2.0) in baselines, the worst 10% performance is increased from 24.98 to **28.03**. Similarly, in the MNIST dataset, the worst 10% performance increased from 88.63 to **89.17**, and in Shakespeare, the worst 10% performance increased from 37.92 to **38.55**. This confirms our FCFL has shown great improvement in protecting unfairly treated clients. As for the global average accuracy, our FCFL can reach 96.06%, 46.12%, and 50.55% in MNIST, CIFAR-10, and Shakespeare respectively, which is very competitive (around 1% difference) to other baselines. In the best 10% accuracy, our FCFL method is a bit lower than some baselines since we emphasize more focus on vulnerable clients and deliver much better fairness.

To summarize, the FCFL method can achieve better fairness in FL while maintaining competitive average accuracy in most cases, which will attract more minority clients to participate in FL, thereby expanding FL applications.

4.3 Efficiency of FCFL

We also record the trend of loss value and the test accuracy of the global model for each round in the MNIST and CIFAR-10 datasets and plot them in Fig. 4. As depicted in Fig. 4(a), the loss of proposed FCFL reduces fast as the communication round increases, which affirms that FCFL can converge as fast as FedAvg,

Table 2. Statistics of the test accuracy distribution on different datasets.

Dataset	Method	Accuracy	Best 10%	Worst 10%	Variance
MNIST	FedAvg	95.96	100.00	87.47	15.06
	q-FedAvg|q=0.2	96.09	100.00	88.40	12.58
	FedFa|α=0.6,β=0.4	**96.23**	100.00	88.37	12.61
	GIFAIR|λ=0.5	96.12	100.00	88.63	12.72
	FCFL|α=0.3,r=0.4	96.06	100.00	**89.17**	**11.03**
CIFAR-10	FedAvg	46.35	**68.67**	20.16	178.93
	q-FedAvg|q=2.0	**47.14**	66.81	24.98	149.50
	FedFa|α=0.6,β=0.4	46.64	68.10	23.19	164.68
	GIFAIR|λ=0.5	46.61	67.02	23.18	158.36
	FCFL|α=0.3,r=0.6	46.12	65.39	**28.03**	**114.59**
Shakespeare	FedAvg	49.16	**70.65**	35.34	89.87
	q-FedAvg|q=2.0	50.24	69.77	37.92	75.99
	FedFa|α=0.5,β=0.5	49.03	69.06	36.23	79.54
	GIFAIR|λ=0.3	50.01	68.50	36.24	78.25
	FCFL|α=0.1,r=0.6	**50.55**	68.74	**38.55**	**67.48**

Table 3. The running time of one communication round.

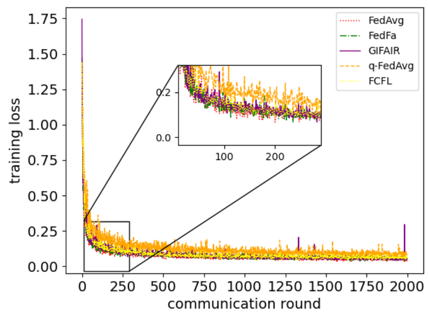
Time(s) / Setting / Method	MNIST(MLP)	MNIST(CNN)	CIFAR-10(MLP)	CIFAR-10(CNN)
FedAvg	**0.87**	1.91	0.89	**1.84**
q-FedAvg	0.95	1.99	0.96	1.90
FedFa	1.03	2.01	0.96	1.91
GIFAIR	0.88	**1.90**	**0.88**	1.85
FCFL	0.94	2.01	0.96	1.95

FedFa, and GIFAIR, and is significantly faster than q-FedAvg. In Fig. 4(b), the test accuracy of FCFL increases rapidly and reaches a convergence value after 1000 rounds. Similar results can be observed from the CIFAR-10 dataset, which states that our FCFL method has a reasonable convergence speed.

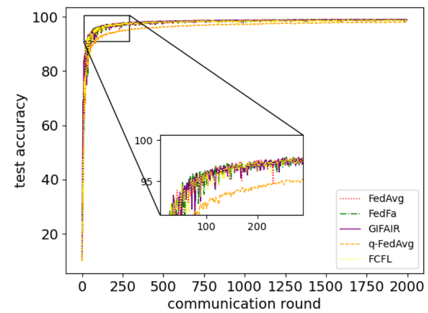
To further validate the efficiency of our FCFL, we compare the time cost for one communication round in different datasets and models. The statistics are presented in Table 3. From Table 3, it can be seen that FCFL does not introduce too much time cost. In some cases, it can achieve the same time efficiency as FedFa and q-FedAvg. Therefore, although FCFL adds one more communication, it does not consume too much time and thus can maintain time efficiency while improving fairness.

4.4 Effect of Hyper-parameter r

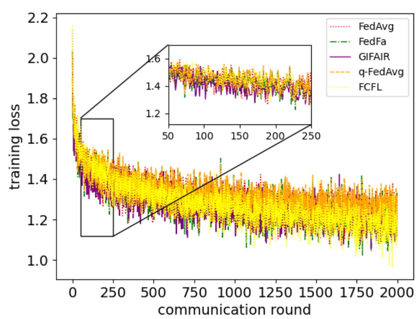
To investigate the impact of the client selection ratio r on our FCFL method, the average accuracy and variance are plotted in Fig. 5 for both datasets. The range



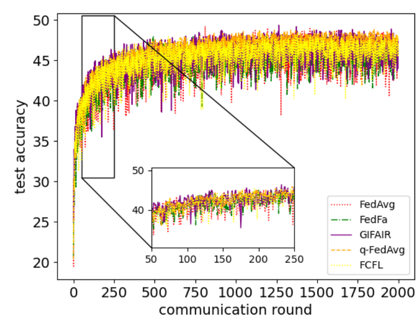
(a) MNIST training loss



(b) MNIST test accuracy



(c) CIFAR-10 training loss

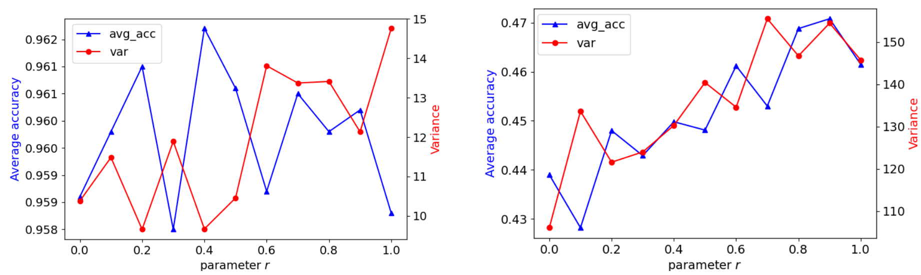


(d) CIFAR-10 test accuracy

Fig. 4. The training loss (left) and test accuracy (right) of FCFL.

of r is $[0, 1]$, representing the ratio of clients selected by the random selection method. Theoretically, increasing r will lead to an increase in accuracy, since greater randomness allows the global model to extract data from more clients; More randomly selected clients cause the system to ignore vulnerable clients, increasing variance among all clients. Due to the simplicity of the MNIST dataset, the performance distribution for each client is uniform, and the average precision and variance do not show significant changes with the parameter r in Fig. 5(a). In contrast, the analysis is reflected in Fig. 5(b) more obviously, where the average accuracy and variance increase along with r, which confirms that enlarging the random ratio r of selected clients will derive higher accuracy and variance of FL system. In our experiments, the parameter r is tuned by grid search, and $r = 0.6$ is selected as the best value that can reach a perfect trade-off between accuracy and variance (i.e., fairness).

4.5 Ablation Experiments



(a) Accuracy and variance on MNIST (b) Accuracy and variance on CIFAR-10

Fig. 5. Analysis of the parameter r on (a) MNIST, and (b) CIFAR-10.

Table 4. Ablation studies of FCFL on CIFAR-10.

Method	Accuracy	Best 10%	Worst 10%	Variance
FCFL	**46.12**	65.39	**28.03**	114.59
FCFL|RS	46.06	**65.60**	22.91	151.31
FCFL|DAR	43.56	61.00	26.57	**96.31**

A series of ablation experiments are conducted on CIFAR-10 to validate our proposed techniques. We compare FCFL with its two variants: (i) FCFL|RS, replacing unfairness-based selection with random selection; and (ii) FCFL|DAR, replacing unfairness-based reweighting with data amount reweighting. The comparison results are presented in Table 4, which demonstrates that the complete FCFL can effectively balance fairness and performance compared to the two variants. Specifically, when comparing FCFL and FCFL|RS, it can be seen that unfairness-based selection can significantly improve the worst 10% performance, and reduce variance from 151.31 to 114.59. This improvement indicates that unfairness-based client selection can solve the unfairness issue effectively. Compared with FCFL|DAR, although the variance is reduced much by FCFL|DAR, it is achieved by sacrificing global accuracy and the best 10% accuracy, which states that our unfairness-based reweighting can improve accuracy effectively. The results of the ablation study indicate that the proposed FCFL takes advantage of both client selection and reweighting strategy, providing a well-justified fairness and performance in FL.

5 Conclusion

The performance fairness problem of FL is investigated in this work, where we propose FCFL, a fairness compensation-based FL algorithm to improve fairness on vulnerable clients. The proposed FCFL considers both client selection and aggregation reweighting to compensate for unfairly treated clients by adopting accumulated unfairness queues. Through intensive experiments and comparison with the existing baselines, the proposed FCFL is demonstrated to improve fairness by 30.4% with high efficiency. In future work, we will extend the study on how to estimate unfairness accurately with approximate global performance and how to select hyper-parameters adaptively to improve overall performance. Furthermore, combining this approach with selection fairness would be an interesting idea in our future investigation.

Acknowledgments. This research was supported by the National Science Foundation grants No. 2011845 and 2343619.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cho, Y.J., Wang, J., Joshi, G.: Client selection in federated learning: Convergence analysis and power-of-choice selection strategies. arXiv preprint arXiv:2010.01243 (2020)
2. Cong, Y., Qiu, J., Zhang, K., Fang, Z., Gao, C., Su, S., Tian, Z.: Ada-ffl: Adaptive computing fairness federated learning. CAAI Transactions on Intelligence Technology (2023)
3. Du, W., Xu, D., Wu, X., Tong, H.: Fairness-aware agnostic federated learning. In: Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). pp. 181–189. SIAM (2021)
4. Ezzeldin, Y.H., Yan, S., He, C., Ferrara, E., Avestimehr, A.S.: Fairfed: Enabling group fairness in federated learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 7494–7502 (2023)
5. Hamidi, S.M., Damen, O.: Fair wireless federated learning through the identification of a common descent direction. IEEE Communications Letters (2024)
6. Hu, Z., Shaloudegi, K., Zhang, G., Yu, Y.: Fedmgda+: Federated learning meets multi-objective optimization. IEEE Transactions on Network Science and Engineering **9**(4), 2039–2051 (2022)
7. Huang, T., Lin, W., Wu, W., He, L., Li, K., Zomaya, A.Y.: An efficiency-boosting client selection scheme for federated learning with fairness guarantee. IEEE Transactions on Parallel and Distributed Systems **32**(7), 1552–1564 (2020)
8. Huang, W., Li, T., Wang, D., Du, S., Zhang, J., Huang, T.: Fairness and accuracy in horizontal federated learning. Information Sciences **589**, 170–185 (2022)
9. Jiang, M., Wang, Z., Dou, Q.: Harmofl: Harmonizing local and global drifts in federated learning on heterogeneous medical images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 1087–1095 (2022)
10. Li, T., Hu, S., Beirami, A., Smith, V.: Ditto: Fair and robust federated learning through personalization. In: International Conference on Machine Learning (ICML). pp. 6357–6368. PMLR (2021)
11. Li, T., Sanjabi, M., Beirami, A., Smith, V.: Fair resource allocation in federated learning. arXiv preprint arXiv:1905.10497 (2019)
12. Li, X., Zhao, S., Chen, C., Zheng, Z.: Heterogeneity-aware fair federated learning. Information Sciences **619**, 968–986 (2023)

13. Liu, Z., Chen, Y., Yu, H., Liu, Y., Cui, L.: Gtg-shapley: Efficient and accurate participant contribution evaluation in federated learning. ACM Transactions on Intelligent Systems and Technology (TIST) **13**(4), 1–21 (2022)
14. Luo, B., Li, X., Wang, S., Huang, J., Tassiulas, L.: Cost-effective federated learning in mobile edge networks. IEEE Journal on Selected Areas in Communications **39**(12), 3606–3621 (2021)
15. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)
16. Mohri, M., Sivek, G., Suresh, A.T.: Agnostic federated learning. In: International Conference on Machine Learning (ICML). pp. 4615–4625. PMLR (2019)
17. Pan, Z., Wang, S., Li, C., Wang, H., Tang, X., Zhao, J.: Fedmdfg: Federated learning with multi-gradient descent and fair guidance. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 9364–9371 (2023)
18. Shi, Y., Liu, Z., Shi, Z., Yu, H.: Fairness-aware client selection for federated learning. In: 2023 IEEE International Conference on Multimedia and Expo (ICME). pp. 324–329. IEEE (2023)
19. Shi, Y., Yu, H., Leung, C.: Towards fairness-aware federated learning. IEEE Transactions on Neural Networks and Learning Systems (2023)
20. Singhal, P., Pandey, S.R., Popovski, P.: Greedy shapley client selection for communication-efficient federated learning. IEEE Networking Letters (2024)
21. Sun, Y., Si, S., Wang, J., Dong, Y., Zhu, Z., Xiao, J.: A fair federated learning framework with reinforcement learning. In: 2022 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2022)
22. Wahab, O.A., Rjoub, G., Bentahar, J., Cohen, R.: Federated against the cold: A trust-based federated learning approach to counter the cold start problem in recommendation systems. Information Sciences **601**, 189–206 (2022)
23. Wang, Z., Fan, X., Qi, J., Wen, C., Wang, C., Yu, R.: Federated learning with fair averaging. arXiv preprint arXiv:2104.14937 (2021)
24. Xiong, Z., Cai, Z., Takabi, D., Li, W.: Privacy threat and defense for federated learning with non-iid data in aiot. IEEE Transactions on Industrial Informatics **18**(2), 1310–1321 (2021)
25. Xiong, Z., Li, W., Cai, Z.: Federated generative model on multi-source heterogeneous data in iot. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 10537–10545 (2023)
26. Xiong, Z., Li, W., Li, Y., Cai, Z.: Exact-fun: An exact and efficient federated unlearning approach. In: 2023 IEEE International Conference on Data Mining (ICDM). pp. 1439–1444. IEEE (2023)
27. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST) **10**(2), 1–19 (2019)
28. Yue, X., Nouiehed, M., R, A.K.: Gifair-fl: A framework for group and individual fairness in federated learning. INFORMS Journal on Data Science **2**(1), 10–23 (2023)
29. Zhao, J., Zhu, X., Wang, J., Xiao, J.: Efficient client contribution evaluation for horizontal federated learning. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3060–3064. IEEE (2021)
30. Zhao, Z., Joshi, G.: A dynamic reweighting strategy for fair federated learning. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 8772–8776. IEEE (2022)