Journal of Machine Learning and Information Security
https://www.sciltp.com/journals/jmlis

Article

FedFE: A Fairness-Equilibrium Framework for Federated Learning Systems

Yuncheng Ge [1,†], Hanlin Li [1,†] and Huiwei Wang [1,2,*]

[1] College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
[2] The Key Laboratory of Networks and Cloud Computing Security of Universities in Chongqing, Chongqing 400715, China
* Correspondence: hwwang@swu.edu.cn
† These authors contributed equally to this work.

How To Cite: Ge, Y.; Li, H.; Wang, H. FedFE: A Fairness-Equilibrium Framework for Federated Learning Systems. *Journal of Machine Learning and Information Security* **2025**, *1*(1), 4.

Received: 13 August 2025
Revised: 16 September 2025
Accepted: 16 October 2025
Published: 27 October 2025

Abstract: With the growing importance of data privacy and the tightening of security regulations, traditional centralized machine learning faces increasing scrutiny due to the need for direct data aggregation. Federated learning (FL) has emerged as a promising alternative, enabling collaborative model training while preserving data privacy by keeping raw data on local devices. Despite its advantages, fairness in FL remains a pressing challenge: disparities in data quality, quantity, and distribution among clients can lead to inequitable outcomes, discourage participation, and even give rise to free-rider problems. Existing research on FL fairness often lacks a holistic and precise methodology, tending to focus primarily on mitigating non-IID effects without adequately balancing fairness with other critical factors such as training efficiency and model accuracy. To address these limitations, this paper proposes a novel fair aggregation framework for FL that jointly ensures internal and external gradient balance while enforcing equitable resource allocation. Our approach further integrates a momentum decay mechanism to accelerate convergence without compromising fairness. Extensive experiments on multiple benchmark datasets validate the effectiveness of the proposed framework, demonstrating consistent improvements in both accuracy and fairness compared to existing baselines.

Keywords: federated learning; fairness; stability; efficiency; machine learning

1. Introduction

Ensuring fairness in federated learning (FL) [1] is essential for its sustainable development and large-scale deployment. A core challenge lies in incentivizing client participation in collaborative training. In current FL settings, clients that contribute vastly different amounts of data and computational resources often receive the same global model, which can lead to dissatisfaction among high-contribution clients and encourage free-riding behavior among low-contribution clients. This work focuses on mitigating data heterogeneity, one of the most fundamental causes of fairness degradation in FL.

Despite a growing body of work on FL fairness, several critical limitations persist: (1) Lack of consensus on fairness definition and measurement. There is no universally accepted definition of fairness in FL, making it difficult to establish standardized evaluation metrics and to compare results across studies; (2) Weak theoretical foundations. Most existing methods rely heavily on heuristic algorithms and empirical validation, with insufficient theoretical grounding to explain the mechanisms driving fairness improvements; (3) Neglect of fairness at multiple stages. Fairness issues arise both during data processing and model training, yet many studies overlook fairness at one or both stages, leading to biased or underperforming client models; (4) Trade-off between privacy and fairness. While FL inherently protects data privacy, improving fairness often requires limited information sharing, creating a tension that has yet to be systematically addressed.

Conventional FL algorithms, such as FedAvg, improve communication efficiency compared to early baselines like FedSGD [2] by transmitting model updates rather than raw data. However, the representational disparity causedby heterogeneous client data distributions remains a key fairness challenge. Even when the global model achieves high average accuracy, some clients may still suffer from significantly lower local performance. This often affects those with highly skewed or non-representative data [3]. Fairness in FL is further challenged by non-IID data distribution and unequal resource allocation. These issues cause gradient divergence across clients and exacerbate performance disparities during aggregated optimization [4,5]. Moreover, existing studies often overlook the historical impact of past gradient updates, resulting in slow and unstable convergence that harms both efficiency and fairness.

To overcome these challenges, we propose a novel fair aggregation framework that jointly addresses gradient conflicts, resource allocation, and convergence stability in FL. Our contributions are threefold:

- Gradient Conflict Resolution: We introduce an iterative projection mechanism that resolves internal gradient conflicts by projecting a client's gradient onto the normal plane of another client's conflicting gradient. We further address external gradient conflicts by estimating the true gradients of unselected clients, reducing the bias introduced by partial client participation.
- Fair Resource Allocation: After conflict resolution, we reweight client updates to give higher aggregation weight to underperforming clients, ensuring a more uniform accuracy distribution across participants.
- Momentum Attenuation for Stability and Speed: We incorporate a time-dependent momentum decay strategy that accelerates early-stage convergence while gradually reducing momentum to prevent instability, thereby improving both convergence speed and model stability.

We validate our approach on three benchmark datasets, comparing it against FedAvg [1], q-FFL [6], and FedFV [7]. The results show that our method consistently outperforms these baselines in fairness, convergence rate, and stability, demonstrating its robustness and scalability in heterogeneous FL environments.

In summary, this work delivers a comprehensive and effective solution to fairness challenges in FL by integrating conflict-aware optimization, fairness-driven aggregation and stability-focused momentum attenuation. These contributions not only improve fairness and efficiency but also make the framework highly suitable for real-world FL deployments, where client diversity and resource variability are the norm. Future research will explore privacy-preserving fairness mechanisms and scalable incentive strategies to further encourage equitable participation in FL systems.

2. Preliminary and Related Work

2.1. Federated Learning

Within a federated learning (FL) framework, a central server orchestrates multiple distributed devices (clients) to collaboratively train a shared model while keeping raw data local. The overall FL workflow is depicted in Figure 1. In practice, client data is often non-IID and bandwidth constraints are common, increasing the complexity of training. The central objective of FL is to minimize the weighted sum of local losses across all participating clients:

$$\min_{\theta} f(\theta) = \sum_{k=0}^{K-1} p_k F_k(\theta), \tag{1}$$

where K is the total number of clients, $p_k > 0$ is the aggregation weight of the k-th client (with $\sum_{k=0}^{K-1} p_k = 1$), and $F_k(\theta)$ denotes the client's local objective function. This local objective is typically the empirical risk over the client's dataset:

$$F_k(\theta) = \frac{1}{N_k} \sum_{n_k=0}^{N_k-1} \ell_{n_k}(\theta), \tag{2}$$

where N_k is the number of samples for client k, and $\ell_{n_k}(\theta)$ is the per-sample loss. The client weight p_k is commonly determined by the proportion of the client's data relative to the total data volume:

$$p_k = \frac{N_k}{\sum_k N_k}. \tag{3}$$

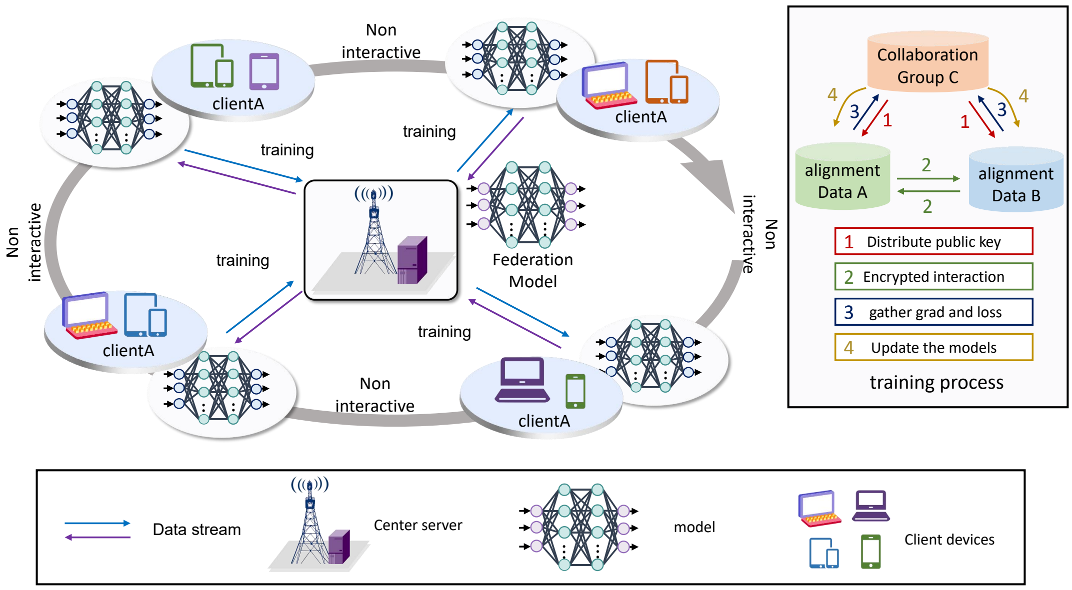

Figure 1. Federated learning framework.

2.2. Fairness

A variety of strategies have been proposed to address fairness challenges in federated learning (FL). Agnostic FL [8] aims to optimize the global model for any target distribution derived from heterogeneous client data. Fair resource allocation approaches, inspired by fairness principles in wireless networks, seek to achieve a more balanced accuracy distribution among FL participants [6]. Non-monetary incentive mechanisms have also been explored to ensure fairness in model aggregation and reward distribution without financial transactions [9]. Other works focus on mitigating gradient conflicts arising from substantial discrepancies among client gradients. For instance, multi-objective optimization has been applied to ensure Pareto-optimal convergence under fairness constraints [10]. FedUFO [11] introduces a unified fairness framework based on distributionally robust optimization to resolve fairness conflicts in medical FL, while FedUP [12] proposes the concept of utopian fairness to balance collaborative efficiency with revenue fairness in cross-island FL. FairFed [13] designs a server-side fairness-aware aggregation method that improves group fairness without requiring centralized access to sensitive attributes.

Several studies combine privacy preservation with fairness. For example, Ref. [14] integrates privacy protection with individual fairness for the first time in FL, using distributed adversarial training to enhance robustness against sensitive perturbations while improving both individual and group fairness. FAC-Fed [15] introduces the first fair federated adaptation framework for streaming classification, and HD-pFL [16] addresses the personalization paradox by constructing depersonalized feature vectors through generative adversarial networks.

In terms of contribution evaluation, Ref. [17] proposes a lightweight framework based on incremental computation. FedCFA [18] employs a counterfactual sample generation mechanism to align local data distributions with the global distribution by replacing key feature factors, while FedSam [19] mitigates data heterogeneity under differential privacy constraints. DCLEF [20] ensures both privacy protection and fairness in horizontal and vertical medical FL scenarios. FedAVE [21] introduces a dynamic gradient reward distribution module, and FGS-FL [22] integrates flat gradient optimization with gradient flow release strategies to alleviate gradient sharpening under differential privacy. Finally, FDL-IWT [23] proposes an intelligent weight transfer method to achieve fair convergence in agricultural FL settings with highly non-IID data. Recently, Ref. [24] proposes a privacy protection framework that integrates federated learning, homomorphic encryption and GRU neural networks, resolving the traditional problem of the irreconcilability between privacy and efficiency.

3. Problem Formulation

In each round of the FL training process, we assume that the central server randomly selects a subset of clients with probability p^t and transmits the current global model θ^t to the selected clients. Each participating client k initializes its local model with θ^t and performs local training, producing an updated φ_k^t along with the corresponding training loss ℓ_k^t. Upon completion, the clients upload their model updates and training losses to the server. The server then aggregates the received information into two sets: $A_t = \{\varphi_0^t, \varphi_1^t, \cdots, \varphi_{K-1}^t\}$ and $L_t = \{l_0^t, l_1^t, \cdots, l_{K-1}^t\}$.

Due to the non-IID nature of data among clients, there might be an issue that gradients for different clients point away from one another as measured by a negative inner product. However, there are scenarios under which such conflicting gradients lead to significantly degraded performance. Taking a simple two-clients problem as an example, if the gradient of one client is much larger in magnitude than the other, it will dominate the average gradient. If there is also high positive curvature along the directions of the client gradients, then the improvement in performance from the dominating client may be significantly overestimated. In this case, the degradation in performance from the dominated client may be significantly underestimated. To better understand gradient conflicts, we introduce the following definition.

Definition 1. *(Conflicting Gradients [25]) Let $\psi_{ij}(i \neq j)$ denote the angle between two distinct gradients φ_i and φ_j, a gradient conflict is said to occur when $\cos \psi_{ij} < 0$, indicating that the two gradients oppose each other.*

Our objective is to mitigate gradient conflicts as much as possible before computing the average gradient. For example, as illustrated in Figure 2a, a conflict is observed between φ_i and φ_j. Refer to [25] for addressing this issue, we project the gradient of one client onto the normal plane of the gradient of the other client, as depicted in Figure 2b,c. Through this approach, the gradient conflict is effectively alleviated, as shown in Figure 2d.

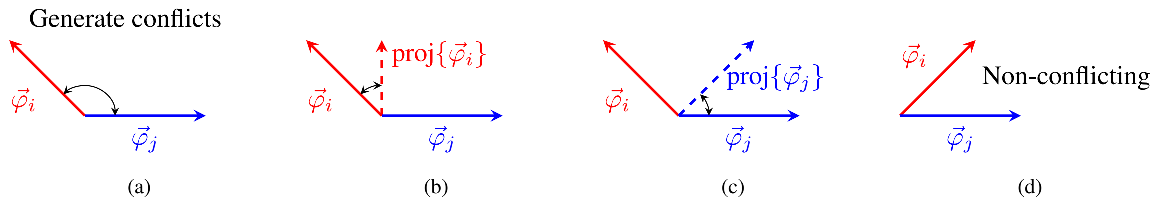

Figure 2. Conflicting gradients and the projection method. In (**a**), clients i and j have conflicting gradient directions, which can lead to destructive interference. In (**b**,**c**), the projection method is illustrated in the case where gradients are conflicting. Non-conflicting client gradients (**d**) are not altered under the projection method.

Conflicting gradients pose serious challenges for the FL training process, and directly using the average client gradient for gradient decent may hurt the performance of individual clients. Especially for the dual FL tasks, internal and external conflicting gradients might arise simultaneously, which will make the FL training process struggle to converge to a desirable solution.

3.1. Internal Conflicting Gradients

In dual classification tasks, the presence of internal conflicting gradients can cause the model to become biased toward clients with larger datasets. This issue is exacerbated in more diverse classification scenarios, leading to increasing disparities among clients. To better understand the internal conflicting gradients, we introduce the following definition.

Definition 2. *(Internal Conflicting Gradients) Let A_t denote the set of client gradients uploaded in the t-th communication round, i.e., $A_t = \{\varphi_0^t, \varphi_1^t, \ldots, \varphi_{K-1}^t\}$. If at least one pair of gradients in A_t satisfies $\cos(\varphi_i^t, \varphi_j^t) < 0$, then an internal conflict is said to have occurred in the t-th round of communication.*

3.2. External Conflicting Gradients

As previously discussed, in each communication round, a subset of clients is selected for model updates. However, when updating θ_t, conflicts may arise between the computed gradient φ_k^t and a virtual gradient φ^t associated with unselected clients. Since unselected clients do not directly contribute gradient updates in the current round, they may be inadequately considered by the model, potentially leading to fairness issues. To formally define external gradient conflicts, we introduce the following definition:

Definition 3. *(External Conflicting Gradients) In the t-th communication round, let φ_h^t represent the updated gradient of a selected client h, and let φ_p^t denote the most recent gradient of an unselected client p. If $\cos(\varphi_h^t, \varphi_p^t) < 0$, then an external gradient conflict is said to have occurred.*

4. The Design of FedFE

4.1. Mitigation of Internal Gradient Conflict

To effectively address the issue of internal gradient conflicts (IGC) as introduced in Section 3.1, we propose a principled projection-based conflict resolution mechanism. This method mitigates destructive interference between client updates by projecting a conflicting gradient vector onto the orthogonal complement of another, thereby eliminating components that are directly opposed in direction. This concept is visualized in Figures 2b,c, and the resulting aligned gradient updates are illustrated in Figure 2d.

Gradient Projection for Conflict Resolution. The conflict mitigation process is carried out on the server side during model aggregation. Specifically, for each client's model update φ_k^{PC} is initialized as $\varphi_k^{PC} \leftarrow \varphi_k^t$. Then the server sequentially examines potential conflicts with updates from other clients in the active set A_t by computing the cosine similarity between gradient vectors. A conflict is detected if the similarity is negative, i.e., $\cos(\varphi_k^{PC}, \varphi_j^t) < 0$, indicating that the two gradients are pointing in opposing directions and may hinder convergence if aggregated directly. Upon detecting such a conflict, we apply an orthogonal projection to remove the interfering component from φ_k^{PC}:

$$\varphi_k^{PC} \leftarrow \varphi_k^{PC} - \frac{\varphi_k^{PC} \cdot \varphi_j^t}{\|\varphi_j^t\|^2} \varphi_j^t, \tag{4}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $| \cdot |$ is the vector norm. This operation ensures that the updated gradient φ_k^{PC} becomes orthogonal to the conflicting direction φ_j^t, thereby neutralizing the negative interaction.

The projection-based conflict resolution framework can be summarized in Algorithm 1, which provides a theoretically grounded and empirically effective mechanism for harmonizing client updates in federated learning. By explicitly removing mutually opposing components and leveraging a principled projection order, our method addresses a core challenge in non-i.i.d. federated settings. Namely, the destructive gradient interactions that hinder global optimization.

Algorithm 1: Mitigating Internal Conflicting Gradients (MICG)

Input: P^t, α
The server selects the first α client of P^t to collect $P_\alpha^t = \{\varphi_{i_1}^t, \varphi_{i_2}^t, \cdots, \varphi_{i_\alpha}^t\}$;
Let $\varphi_k^{PC} \leftarrow \varphi_k^t$;
for *each client* $k \in P_\alpha^t$ *in parallel* **do**
  for $j = i_1, \cdots, i_{|S^t|}$ **do**
    if *there is an internal conflict in* P_α^t*, i.e.,* $\cos(\varphi_k^{PC}, \varphi_j^t) < 0$*,* $j \neq k$ **then**
      update $\varphi_k^{PC} \leftarrow \varphi_k^{PC} - \frac{\varphi_k^{PC} \cdot \varphi_j^t}{\|\varphi_j^t\|^2} \cdot \varphi_j^t$;
    end
  end
end
return φ_k^{PC} for each client $k \in P_\alpha^t$;

4.2. Mitigation of External Gradient Conflict

In each communication round of federated learning, the virtual gradient of a client is defined based on its most recently participated round, denoted as t_k. To ensure that the global model update remains consistent with the client's learning trajectory, it is necessary to evaluate potential external gradient conflicts between the client's historical gradient $\varphi_k^{t_k}$ and the current aggregated global gradient φ^t. There may be a conflict between these gradients, i.e., $\cos(\varphi_k^{t_k}, \varphi^t) < 0$. In such cases, the current gradient is calibrated by projecting it orthogonally to the direction of accumulated historical gradients that contributed to the conflict. The adjusted gradient φ_t is computed as:

$$\varphi_t = \varphi_t - \frac{\varphi_t \cdot \varphi_{\text{con}}}{\|\varphi_{\text{con}}\|^2} \cdot \varphi_{\text{con}}, \tag{5}$$

where the conflicting component φ_{con} is defined as the cumulative sum of past gradients from the client:

$$\varphi_{\text{con}} = \sum_{t'=0}^{t_k-1} \varphi_k^{t'}. \tag{6}$$

This projection eliminates the component of the current gradient that aligns oppositely to the historical gradient trajectory, thereby mitigating destructive interference in the global model update process. The projection-based conflict resolution framework for mitigating external conflicting gradients can be summarized in Algorithm 2.

Algorithm 2: Mitigating External Conflicting Gradients (MECG)

Input: A^t, φ^t, τ

```
for round t − i, i = τ, τ − 1, ··· , 1 do
    Initialize φ_con ← 0;
    for each client k = 1, ··· , K in parallel do
        if t_k = t − i then
            if there is an external conflict, i.e., cos φ^t · φ_k^{t_k} < 0 then
                φ_con ← φ_con + φ_k^{t_k};
            end
        end
        if there is an external conflict, i.e., cos φ^t · φ_con < 0 then
            Server computes φ_t ← φ_t − (φ_t · φ_con / ||φ_con||^2) · φ_con;
        end
    end
end
return φ^t to server
```

4.3. q-FedSGD in Resolving Gradient Conflicts

q-FedSGD is a fairness-aware optimization framework that extends the classic FedSGD algorithm by introducing a tunable fairness parameter q, which allows for dynamic reweighting of client contributions based on their local loss values. Originally proposed in [6], q-FedSGD redefines the global objective function to prioritize fairness across heterogeneous clients:

$$\min_{\theta} f_q(\theta) = \sum_{k=1}^{K} \frac{p_k}{q+1} F_k^{q+1}(\theta), \tag{7}$$

where p_k denotes the relative importance (or weight) of client k, and $F_k(\theta)$ is the local loss evaluated at model parameter θ. When the parameter q is large, the optimization objective increasingly emphasizes clients with higher local losses. In the extreme case, this formulation approaches the worst-case optimization objective described in agnostic federated learning [8], where the client with the worst performance dominates the update direction.

Importantly, q-FedSGD serves not only as a fairness mechanism but also provides an implicit strategy for addressing internal gradient conflicts in federated learning. In heterogeneous environments, clients may produce gradients that conflict with each other due to non-i.i.d. data distributions. These conflicts hinder convergence and lead to suboptimal global models. q-FedSGD mitigates this issue by adaptively scaling each client's gradient according to its local loss and the fairness parameter q, thereby reshaping the direction of global updates in a way that downplays conflicting or uninformative gradients.

Gradient Scaling and Update Procedure. Unlike traditional FedSGD, where client gradients are aggregated uniformly or according to static weights, q-FedSGD employs a loss-adaptive gradient modulation strategy. After performing internal conflict resolution (e.g., via projection as described in our previous section), each client computes its local gradient update $\varphi_k^{PC}(\theta^t)$, already corrected for internal conflicts, and its local loss: $F_k(\theta^t)$.

These values are then used to construct a scaled gradient contribution Δ_k^t and a corresponding normalization factor y_k^t, as follows:

$$\Delta_k^t = \varphi_k^{PC}(\theta^t) \cdot F_k^q(\theta^t), \quad \text{and} \quad y_k^t = q \cdot F_k^{q-1}(\theta^t) \cdot \left|\varphi_k^{PC}(\theta^t)\right|^2 + L \cdot F_k^q(\theta^t), \tag{8}$$

where L is a Lipschitz constant used for bounding the loss function. Once these quantities are computed locally, they are sent to the server. The global model update direction φ^t is then computed via normalized aggregation:

$$\varphi^t = \frac{\sum_{k \in S^t} \Delta_k^t}{\sum_{k \in S^t} y_k^t}, \tag{9}$$

which ensures that the updates from high-loss clients (often underrepresented in traditional averaging) are amplified in proportion to their loss, thus enhancing fairness and gradient alignment simultaneously.

4.4. Momentum Attenuation in Resolving Gradient Conflicts

Prior research (e.g., [26]) has noted that the convergence of the standard Federated Averaging (FedAvg) algorithm can be sluggish, especially in non-IID settings. To address this limitation, momentum-based methods have been proposed, aiming to accelerate convergence by incorporating historical gradient information.

In traditional Stochastic Gradient Descent with Momentum (SGDM), the parameter update rule at the server is given by:

$$\upsilon \leftarrow \beta\upsilon + \Delta\omega, \quad \omega \leftarrow \omega - \upsilon, \tag{10}$$

where β is the momentum coefficient, and $\Delta\omega$ is the aggregated client update. This approach helps to smooth the optimization trajectory, reduce oscillations, and expedite training by leveraging accumulated gradients.

However, a notable drawback of classical SGDM is its lack of temporal decay: the influence of early gradients persists indefinitely. This can lead to overshooting, excessive weight growth, or diminished responsiveness to new information in later training stages. To overcome this, we propose a dynamic momentum attenuation strategy inspired by decayed learning rate schedules and the work of Fernandes et al. [27].

Time-Dependent Momentum Attenuation. We introduce a decaying momentum framework in which the momentum coefficient β is progressively reduced over training iterations. The key idea is to accelerate early training by maintaining high momentum, and attenuate it later to avoid instability and ensure convergence.

The recursive update of model parameters incorporating historical gradients can be expressed as:

$$\theta^{t+1} = \theta^t - \eta\varphi^t - \eta\beta\varphi^{t-1} - \eta\beta^2\varphi^{t-2} + \eta\beta^3\varphi^{t-3}\cdots, \tag{11}$$

which is equivalently reformulated as:

$$\theta^{t+1} = \theta^t - \eta\varphi^t - \eta\sum_{i=1}^{t}\left(\beta^i \cdot \varphi^{t-i}\right), \tag{12}$$

where β^i denotes the momentum coefficient applied to the i-th past gradient. To control the total contribution of all past gradients, we define the cumulative momentum sum:

$$\sum_{i=1}^{\infty}\beta^i = \frac{\beta}{1-\beta}. \tag{13}$$

To ensure a time-dependent decay, we introduce a decay schedule governed by the remaining training iterations. Let T denote the total number of training rounds, and t the current round. Then:

$$\frac{\beta}{1-\beta} = \left(1 - \frac{t}{T}\right) \cdot \frac{\beta_{\text{init}}}{1-\beta_{\text{init}}}, \tag{14}$$

which yields a dynamically scaled momentum coefficient β_t as:

$$\beta^t = \beta_{\text{init}} \cdot \frac{1 - t/T}{1 - \beta_{\text{init}} + \beta_{\text{init}}(1 - t/T)}. \tag{15}$$

This formulation ensures that β^t starts from an initial value β_{init} and gradually decays to zero as training progresses.

Server-Side Model Update Rule. Finally, with the decayed momentum applied, the global model at the server is updated using:

$$\theta^{t+1} = \beta^t\theta^t - \eta'\varphi_k^t, \tag{16}$$

where η' is the global learning rate, and φ_k^t represents the aggregated and possibly conflict-resolved gradient from the clients.

This momentum attenuation framework, when coupled with external gradient conflict resolution, provides a robust mechanism to guide federated optimization toward faster and more stable convergence, especially in heterogeneous client settings.

4.5. Overall of FedFE Algorithm

This subsection summarizes the proposed FedFE algorithm, in which we conduct a comprehensive description of how gradient conflicts and heterogeneous resource distribution affect fairness in federated learning (FL) systems. To promote both fairness and convergence stability, we formulate a novel optimization objective that explicitly accounts for client-level disparities.

Loss-Aware Projection Ordering. Rather than applying projection in a random or arbitrary order, we introduce a loss-aware projection strategy. In this approach, gradients associated with higher local training losses are processed earlier. This prioritization reflects the intuition that clients with higher losses may provide more informative or critical updates at the current training stage. As formally established in the proposed FedFE shown in Algorithm 3, this ordering induces a beneficial side effect: gradients projected later in the sequence tend to experience fewer conflicts with the accumulated direction, leading to improved alignment with the final aggregated update. Consequently, the approach enhances both convergence speed and training stability.

Algorithm 3: FedFE

Input: q, T, α, τ, θ^0, L, β_{init}.

for $t = 0, 1, 2, \cdots, T-1$ **do**

- The server randomly selects a subset S^t of K clients with a certain probability p_k, and then sends the global model θ^t to all selected clients k;
- Each selected client $k \in S^t$ calculates the local update via $\varphi_k^t \leftarrow \theta^t - \theta_k^t$ and local loss ℓ_k^t, and sends φ_k^t and ℓ_k^t back to the server;
- The server collects updates φ_k^t and ℓ_k^t, and forms $G^t = \{\varphi_1^t, \varphi_2^t, \cdots, \varphi_{|S^t|}^t\}$ for each selected client at round t. Meanwhile, the server also updates $A^t = \{\varphi_1^{t_1}, \varphi_2^{t_2}, ..., \varphi_K^{t_K}\}$ for all of clients, where $t_{(\cdot)}$ denotes the last round each client participated in the training;
- Then, the server reorders A^t in ascending order of each client's training loss ℓ_k^t to obtain P^t, where $P^t = \{\varphi_{i_1}^t, \varphi_{i_2}^t, \cdots, \varphi_{i_{|S^t|}}^t\}$ provides the client with the sequence in which each gradient is used as the projection target, and $i_1, i_2, \cdots, i_{|S^t|}$ are the new index after reordering;
- **for** *each client j selected in S^t, i.e., $j = i_1, \cdots, i_{|S^t|}$* **do**
  - Determine whether each update φ_j^t in P^t and φ_k^{PC} causes any conflict and resolve conflict by performing Algorithm 1, i.e., $\varphi_k^{PC} \leftarrow$ MICG(P^t, α);
- **end**
- Each client collects $\varphi_k^{PC}(\theta^t)$ and losses $F_k(\theta^t)$ after mitigating internal conflicting gradients, and calculates Δ_k^t and y_k^t by the following equations:

$$\Delta_k^t = \varphi_k^{PC}(\theta^t) F_k^q(\theta^t) \quad \text{and} \quad y_k^t = qF_k^{q-1}(\theta^t)\left\|\varphi_k^{PC}(\theta^t)\right\|^2 + LF_k^q(\theta^t)$$

  and then sends back to the server;
- The server updates $\varphi^t = \sum_{k \in S^t} \Delta_k^t / \sum_{k \in S^t} y_k^t$;
- **if** $t \geq \tau$ **then**
  - **for** *each client k unselected in S^t, i.e., $k \notin S^t$* **do**
    - Determine whether the latest gradient φ_k^{t-i} causes any conflict with the combined updated φ^t, and resolve conflict by performing Algorithm 2, i.e., $\varphi_k^t \leftarrow$ MECG(A^t, φ^t, τ);
  - **end**
- **end**
- Calculated the dynamically scaled momentum coefficient $\beta_t \leftarrow \beta_{\text{init}} \cdot \frac{(1-t/T)}{1-\beta_{\text{init}}+\beta_{\text{init}}(1-t/T)}$;
- Server updates the model $\theta^{t+1} \leftarrow \beta_t \theta^t - \eta' \varphi^t$;

end

return θ^T

Synergy with q-FedSGD. When combined with our gradient conflict projection method, q-FedSGD becomes even more effective. The projection removes direct gradient opposition among clients, while q-FedSGD modulates the magnitude and influence of each gradient based on its importance. This two-stage strategy first resolves gradient conflicts and then adjusts their weighting, achieving both geometric alignment and fair prioritization. As such, q-FedSGD can be viewed as a higher-level mechanism that builds upon and benefits from low-level gradient conflict resolution.

Momentum Attenuation for Improved Convergence. Momentum-based optimization has long been recognized as an effective approach to accelerate training in distributed learning by incorporating information from past gradients into current updates. However, in the context of FL, a fixed momentum coefficient β can be problematic. Large, persistent momentum values amplify the influence of earlier updates,even if those updates originated from clients whose data distributions differ significantly from the current participants. To address this challenge, we introduce a time-dependent momentum attenuation strategy. The key insight is that high momentum is beneficial in the early stages of training when large, consistent descent steps can accelerate progress toward a low-loss region. However, momentum should be gradually reduced as the model approaches convergence. This attenuation weakens the impact of earlier, potentially conflicting gradients, allowing the optimization process to focus more on recent, context-relevant updates. From the perspective of gradient conflict mitigation, the time-dependent momentum plays a dual role:

- **Conflict Resolution:** By diminishing the influence of older gradients over time, momentum attenuation reduces the persistence of conflicting gradient components from earlier communication rounds, especially those originating from unaligned client objectives.
- **Stability Enhancement:** Near convergence, smaller momentum values prevent overshooting and oscillations, enabling the optimizer to settle into a more stable minimum even in the presence of residual client-to-client variability.

From the perspective of convergence acceleration, the strategy ensures that:

- Early-stage training benefits from aggregated gradient history, enabling faster movement toward the optimal region despite noisy or partially conflicting updates.
- Late-stage training becomes more adaptive to recent updates, allowing finer adjustments that better reflect the current global optimization landscape.

In combination with the projection-based gradient conflict resolution described earlier, time-dependent momentum attenuation provides a synergistic mechanism for both reducing destructive gradient interference and improving convergence efficiency in federated optimization.

Together, these components form a cohesive optimization framework that addresses both data heterogeneity and client imbalance—two of the most fundamental challenges in federated learning. The algorithm proceeds through the following steps:

Step 1: The central server selects a subset of clients according to a predefined sampling probability and dispatches the current global model to the chosen clients.

Step 2: Upon receiving the global model, each client performs local training to compute its gradient update and training loss, which are then transmitted back to the server. The server organizes the received client updates in descending order of their reported loss values, prioritizing updates from better-performing clients.

Step 3: The server stores the most recent gradient update from each client. These historical gradients are later leveraged to identify and resolve external gradient conflicts that may arise from uncoordinated updates across communication rounds.

Step 4: To address internal gradient conflicts, the server constructs a projection sequence and projects each client's update onto the normal plane of the conflicting client's update. This geometric correction ensures that individual updates do not destructively interfere with one another.

Step 5: Following internal conflict resolution, the server applies a fairness-aware reweighting scheme. Updates and corresponding losses are reweighted based on a resource-aware strategy, wherein clients with limited computational or communication resources are assigned higher weights. This approach promotes accuracy balance across clients, mitigating performance disparities caused by resource heterogeneity.

Step 6: To further address external conflicts, the server estimates the latent gradients of unselected clients. These approximated gradients are used to identify inconsistencies between the current global update and the hypothetical contributions of unselected clients, thus refining the aggregation process.

Step 7: Finally, the server incorporates a momentum decay mechanism to suppress excessive accumulation of past updates. This helps prevent instability and slow convergence caused by over-amplified momentum, especially in later stages of training.

5. Experiments

This paper presents a comprehensive investigation into the fairness challenge among users in federated learning (FL), with the objective of achieving higher fairness than conventional FL algorithms. Our analysis centers on two primary evaluation dimensions: model accuracy and model fairness. Model accuracy measures the predictive capability of the global model, with its computation method described in Section 2. Higher accuracy reflects better predictive performance, which in turn contributes to improved user experience and satisfaction.

Model fairness is assessed by examining the uniformity of training accuracy across all participating users. We quantify fairness using the standard deviation of user-specific accuracies, where a smaller standard deviation indicates a more equitable distribution of model performance and reduced disparities between users.

Our experimental framework and baseline comparisons are built upon the classical Federated Averaging (FedAvg) algorithm. In addition to FedAvg, we adopt q-FFL and FedFV as baseline methods to evaluate the effectiveness of the proposed approach. The primary comparison criterion is the variance of user accuracies across multiple datasets, with dataset configurations detailed in Section 5.1. An in-depth analysis of the proposed method's fairness and convergence characteristics is provided in Section 5.4.

5.1. Data Description

CIFAR-10: The CIFAR-10 dataset is a widely used benchmark for image classification, consisting of RGB images across 10 categories, including airplanes, automobiles, and birds. Each image is 32×32 pixels in size. The dataset contains $50,000$ training images and $10,000$ test images. In our experiments, we adopt a multi-layer perceptron (MLP) with two hidden layers for training. The dataset is partitioned into 200 shards and assigned to 100 clients, with each client receiving two shards. In each communication round, 10% of the clients are randomly selected to participate in training.

MNIST: The MNIST dataset, a subset of the NIST dataset, is designed for handwritten digit recognition. It contains $70,000$ grayscale images of 28×28 pixels, evenly split between samples collected from high school students and employees of the U.S. Census Bureau. We train MNIST using a convolutional neural network (CNN) with two convolutional layers. The dataset is distributed among 200 clients following the same shard-based partitioning strategy as CIFAR-10. Each client's local dataset is further split into 80% training and 20% testing data.

Fashion-MNIST (FMNIST): The Fashion-MNIST dataset serves as a drop-in replacement for MNIST but contains grayscale images of clothing items from 10 categories, such as T-shirts, jeans, and jackets. For this dataset, we employ a lightweight logistic regression (LR) model and limit the number of participating clients to three.

5.2. Metrics

5.2.1. Accuracy

Although federated learning (FL) differs fundamentally from traditional machine learning frameworks in terms of data distribution and training paradigms, many standard evaluation metrics from conventional machine learning remain directly applicable. Among these, accuracy-based metrics play a central role in assessing overall model performance, providing a quantitative measure of a model's ability to solve specific tasks and its predictive effectiveness.

In machine learning, model quality is commonly evaluated using seven key metrics: accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), area under the precision–recall curve (AUC-PR), and mean accuracy. Together, these metrics offer a comprehensive view of a model's predictive behavior across different performance dimensions.

Accuracy is one of the most widely used metrics for evaluating a model's task-processing capability. It is defined as the proportion of correctly classified instances among all samples. In binary classification problems, accuracy is mathematically expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{17}$$

where True Positive (TP) represents the number of correctly predicted positive samples, True Negative (TN) denotes the number of correctly predicted negative samples, False Positive (FP) refers to negative samples incorrectly classified as positive, and False Negative (FN) corresponds to positive samples misclassified as negative.

5.2.2. Fairness

In multi-class classification problems, accuracy is calculated similarly to the binary classification case; however, the confusion matrix must be appropriately adjusted to account for multiple classes. In this study, we employ accuracy as a primary metric to evaluate model performance.

As previously discussed, FL differs significantly from traditional machine learning approaches. Traditional machine learning typically relies on a single centralized dataset for model training, whereas FL involves simultaneous training across multiple clients. This decentralized nature imposes additional challenges, particularly regarding fairness. Due to the heterogeneous distribution of data among participants and the constraints of data privacy protection, the calculation and evaluation of accuracy metrics in FL may face certain limitations.

Fairness in FL extends beyond the protection of sensitive attributes or designated protected groups. It must also account for the interactions between different participants. In FL, sensitive attribute data—or data pertaining to protected groups—may be distributed across various clients, necessitating fairness assessments based on the model's predictive accuracy.

However, privacy protection poses a significant challenge to fairness measurement. Since the FL server does not have direct access to local participant data, obtaining the sensitive attribute information required for fairness evaluations is inherently difficult.

From the perspective of FL clients, different participants may experience varying levels of model accuracy. Ensuring fairness requires not only maintaining individual fairness, wherein participants with similar data should receive comparable predictive outcomes, but also group fairness, which ensures similar accuracy rates across different participant groups. The complexity and diversity of FL impose stringent fairness requirements. Therefore, during model training, it is essential to consider both the sensitive attributes of data and the interactions between clients to uphold fairness while maintaining model effectiveness.

Several commonly used metrics for evaluating fairness in FL include Equal Opportunity Difference, Variance of Accuracy Rate, Unified Group Fairness, Minmax Pareto Fairness, Rawlsian Max-Min Fairness, and Minmax Group Fairness.

In this study, we employ the variance of accuracy rate as the primary fairness metric in our experiments.

5.3. Hyperparamters

For all experiments, the batch size was fixed as follows: CIFAR-10 (64), MNIST (10), and Fashion-MNIST (400). The learning rate was set to 0.1, the number of training epochs was fixed at 1, and the initial momentum β_{init} was set to 0.1. We evaluated and compared the performance of the optimal parameter settings for each method.

5.4. Performance Analysis

We begin by comparing the performance of four algorithms: FedFV [7], FedAvg [1], q-FFL [6], and our proposed method FedFE, which jointly incorporates gradient conflict resolution, fair resource allocation, and momentum attenuation. FedAvg, as a fundamental algorithm in federated learning, lacks certain fairness designs, which may lead to model performance tilting towards clients with larger data volumes. q-FedAvg mainly achieves fair resource allocation by dynamically weighting and allocating more resources to clients with higher losses, but it has not yet addressed the fundamental issue of gradient conflicts. FedFV, on the other hand, starts from gradient coordination fairness and mainly alleviates update direction conflicts through projection, but it does not consider the importance of resource allocation in enhancing fairness and may also cause training instability. Our proposed FedFE ingeniously integrates the complementary mechanisms of the above: it not only adopts the dynamic weight allocation of q-FedAvg to focus on disadvantaged clients but also incorporates the gradient conflict mitigation technique of FedFV to guide the global model to update in a direction beneficial to all parties. It can be seen that FedFE simultaneously enhances fairness from both the "weight" and "direction" dimensions. More importantly, we introduce a momentum decay mechanism in FedFE as a "stabilizer" for federated learning, effectively smoothing out the convergence oscillations that may arise from integrating multiple techniques, significantly improving the stability and convergence speed of the training process. This systematic integration strategy enables FedFE to overcome the shortcomings of other algorithms that use a single mechanism to enhance fairness in federated learning, achieving a better balance among fairness, global accuracy, and training stability, demonstrating strong novelty and complementarity.

As shown in Figure 3, on the CIFAR-10 dataset, FedFE achieves markedly superior fairness relative to existing FL approaches explicitly designed to address fairness issues. The incorporation of momentum attenuation results in a more stable learning trajectory and accelerates convergence compared to the baselines. For all algorithms, hyperparameters were tuned to ensure optimal performance.

As illustrated in Figure 4, for the Fashion-MNIST (FMNIST) dataset, we employed a lightweight network with only three participating clients and set $\theta = 0$, thereby allowing FedFE to bypass the external gradient conflict resolution stage. Even in this small-scale setting, the results demonstrate notable performance gains. Compared to q-FFL and FedAvg, FedFE achieves substantial improvements in fairness, while also converging faster than FedFV. These findings highlight FedFE's ability to effectively balance fairness and convergence stability, even in networks with limited client participation.

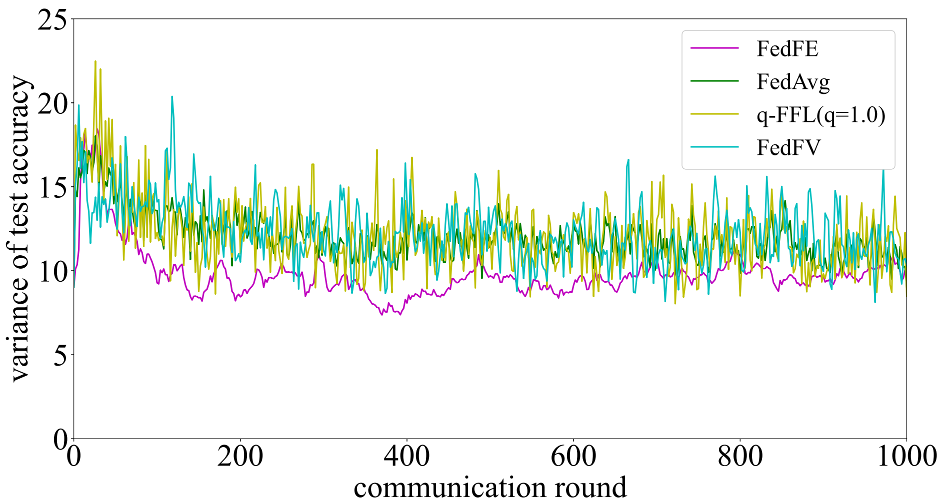

Figure 3. Performance of the proposed FedFE on CIFAR-10 in comparison with FedFV [7], FedAvg [1], and q-FFL ($q = 1.0$) [6].

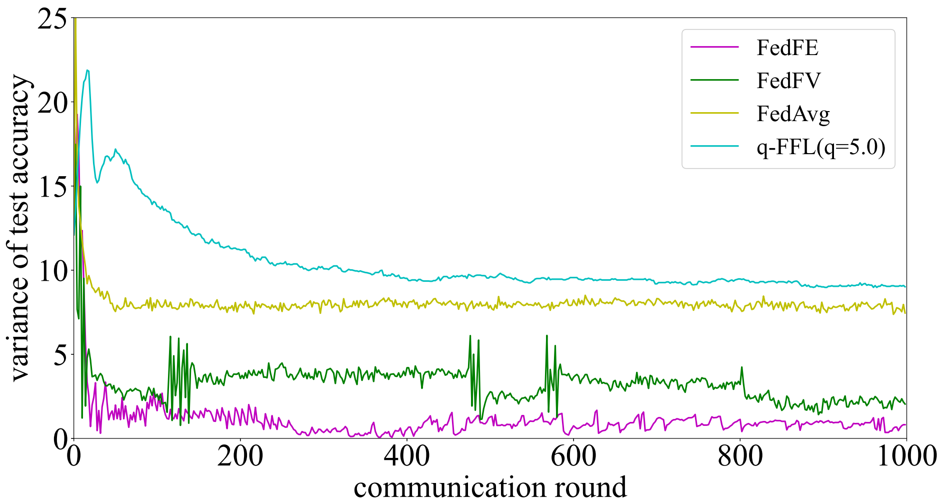

Figure 4. Performance of the proposed FedFE on FMNIST in comparison with FedFV[7], FedAvg [1], and q-FFL ($q = 5.0$) [6].

For the MNIST dataset, we set $\beta_{\text{init}} = 0$, effectively reverting the process to a random gradient initialization. As shown in Figure 5, this lower β_{init} value led to a reduction in convergence speed. This is because our key idea is to accelerate early training by maintaining high momentum, and the momentum coefficient is set to gradually decrease during the training iteration process to ensure the stability of convergence and guarantee convergence. That is to say, if we set a lower value of β_{init}, it will lead to a decrease in the overall convergence speed. However, FedFE still achieved superior fairness compared to other methods, primarily due to its fairness-oriented resource allocation strategy.

As shown in Table 1, compared to the best-performing alternative methods, FedFE achieved fairness improvements of 8%, 92%, and 3% on CIFAR-10, FMNIST, and MNIST, respectively.

Table 1. Lowest variance of the proposed FedFE in comprison with FedFV, FedAvg, and q-FFL.

Dataset	FedFE	FedFV [7]	FedAvg [1]	*q*-FFL [6]
CIFAR-10	7.41	8.11	8.28	8.03
FMNIST	0.06	0.77	7.37	8.97
MNIST	1.88	1.95	2.05	2.38

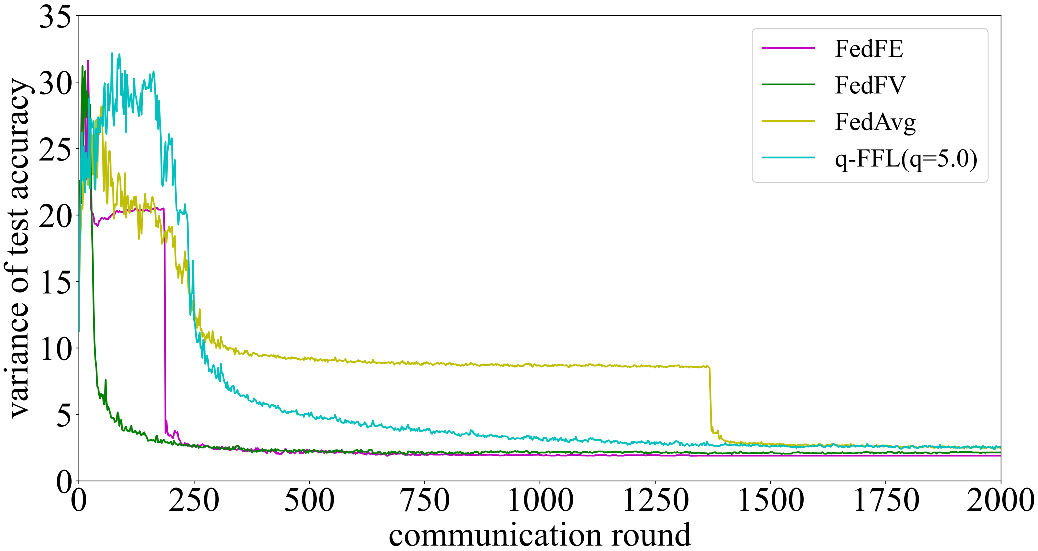

Figure 5. Performance of the proposed FedFE on MNIST in comparison with FedFV[7], FedAvg [1], and q-FFL ($q = 5.0$) [6].

5.5. Ablation Study

To further evaluate the efficiency of FedFE, we performed an ablation study on the CIFAR-10 and FMNIST datasets, dividing the experiments into three configurations: (1) a baseline model without resource allocation or momentum attenuation; (2) a model incorporating only fair resource allocation; and (3) the full FedFE framework, integrating both fair resource allocation and momentum attenuation. For CIFAR-10, we set $\beta_{\text{init}} = 0.1$ and $q = 1$, training the model for 1000 communication rounds. For FMNIST, we used the same initial momentum value $\beta_{\text{init}} = 0.1$ and $q = 5$, also training for 1000 rounds.

As shown in Figures 6 and 7, incorporating fair resource allocation into the baseline model led to a marked improvement in fairness. Furthermore, introducing momentum attenuation not only smoothed the training process but also accelerated the overall convergence rate. These results confirm the effectiveness of FedFE in simultaneously enhancing fairness and stability in federated learning.

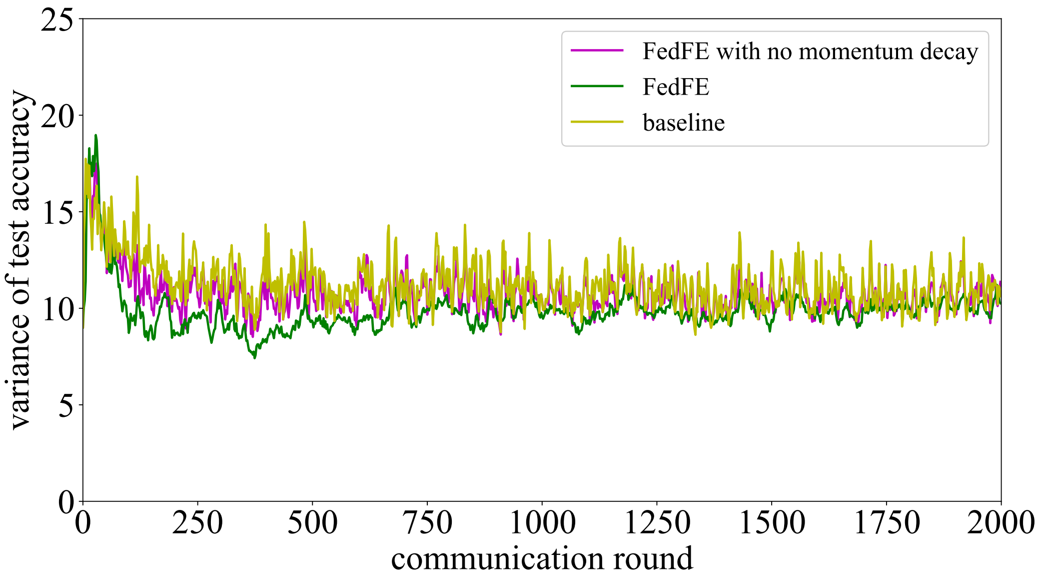

Figure 6. Performance of the full FedFE (3) on CIFAR-10 in comparison with a baseline model (1) and FedFE with no momentum decay (2).

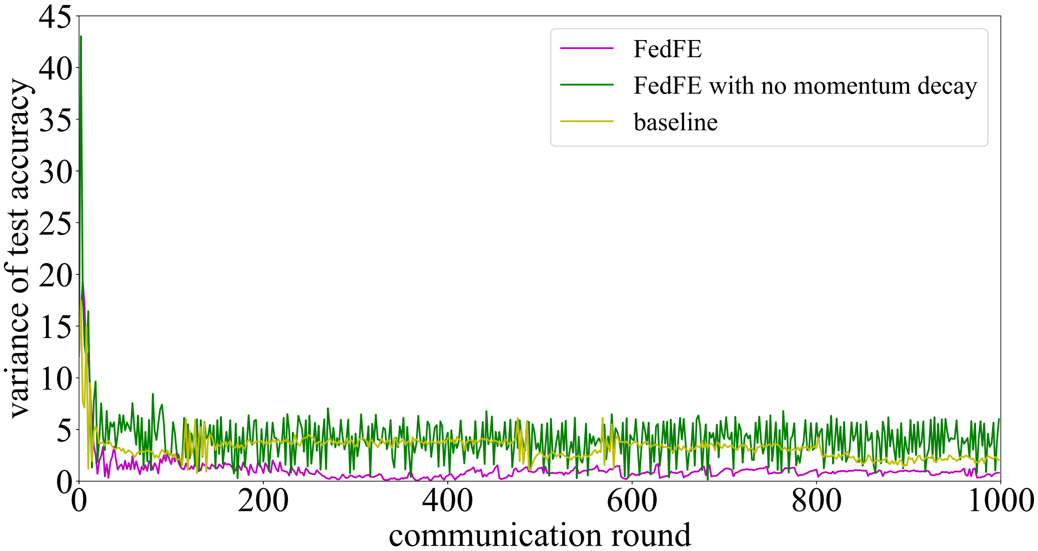

Figure 7. Performance of the full FedFE (3) on FMNIST in comparison with a baseline model (1) and FedFE with no momentum decay (2).

6. Conclusions

This study presents a comprehensive investigation into three persistent challenges in federated learning (FL): gradient imbalance, uneven resource distribution, and slow convergence. As an emerging paradigm in distributed machine learning, FL enables multiple devices to collaboratively train a global model without sharing raw data.Therefore,it can preserve privacy while enhancing efficiency. However, heterogeneous network conditions and nonuniform data distributions across devices often give rise to imbalanced gradients and unequal resource utilization.These issues can, in turn, lead to sluggish convergence or complete training failure in severe cases. To address these challenges, we propose a novel optimization objective that jointly enhances fairness and convergence stability in FL systems. Our method explicitly promotes a more equitable distribution of model accuracy across clients while simultaneously accelerating convergence and improving training robustness.

We validate the proposed approach on three widely used benchmark datasets and compare its performance with established baselines, including FedAvg, q-FFL, and FedFV. The experimental results demonstrate that our method consistently outperforms these baselines in terms of accuracy distribution, convergence rate, and stability. Moreover, the approach exhibits strong scalability and adaptability, making it applicable to a broad range of FL scenarios with varying client heterogeneity and resource constraints.

In summary, this research offers an effective and generalizable solution for mitigating fairness-related issues in FL, delivering substantial improvements in both convergence speed and overall model performance. These findings contribute to advancing the deployment of FL in real-world applications where fairness, stability, and efficiency are critical.

Looking ahead, future research should aim to address the unfairness inherent in data heterogeneity through the development of more sophisticated personalized federated learning schemes. Further efforts could also focus on designing adaptive algorithms capable of preserving model utility and fairness under stringent privacy constraints such as differential privacy, thereby achieving an optimal balance among privacy, fairness, and utility. Additionally, comprehensive scalability studies are necessary to rigorously evaluate the robustness and efficiency of fairness-aware algorithms, including FedFE, particularly in large-scale settings involving hundreds or thousands of clients and under conditions of extreme non-IID data distribution. There is also a critical need to establish a unified, multi-dimensional fairness evaluation framework that can be applied across different domains to objectively assess various fairness criteria, including group, individual, and cross-center fairness. Finally, extending these investigations to complex and emerging scenarios such as federated large language models will be essential to tackle their unique biases and fairness-related challenges. Ultimately, fostering interdisciplinary collaboration will play a key role in advancing the fairness of federated learning systems.

Author Contributions

Y.G.: conceptualization, methodology, data curation, writing—original draft preparation; H.L.: writing—reviewing and editing; H.W.: investigation, supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported in part by Natural Science Foundation of China under Grant 62476231 and 62433018.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

1. McMahan, B.; Moore, E.; Ramage, D.; et al. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
2. Chen, J.; Pan, X.; Monga, R.; et al. Revisiting Distributed Synchronous SGD. In Proceedings of International Conference on Learning Representations (ICLR) Workshop Track, San Juan, PR, USA, 2–4 May 2016.
3. Hashimoto, T.; Srivastava, M.; Namkoong, H.; et al. Fairness Without Demographics in Repeated Loss Minimization. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1929–1938.
4. Zhao, Y.; Li, M.; Lai, L.; et al. Federated learning with non-iid data. *arXiv* **2018**, arXiv:1806.00582.
5. Liu, W.; Chen, L.; Chen, Y.; et al. Accelerating Federated Learning via Momentum Gradient Descent. *IEEE Trans. Parallel Distrib. Syst.* **2020**, *31*, 1754–1766.
6. Li, T.; Sanjabi, M.; Beirami, A.; et al. Fair resource allocation in federated learning. In Proceedings of the International Conference on Learning Representations, 26 April–1 May 2020.
7. Wang, Z.; Fan, X.; Qi, J.; et al. Federated learning with fair averaging.In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 21–26 August 2021; pp. 1615–1623.
8. Mohri, M.; Sivek, G.; Suresh, A.T.; et al. Agnostic Federated Learning. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 4615–4625.
9. Shi, Z.; Zhang, L.; Yao, Z.; et al. FedFAIM: A Model Performance-Based Fair Incentive Mechanism for Federated Learning. *IEEE Trans. Big Data* **2024**, *10*, 1038–1050.
10. Hu, Z.; Shaloudegi, K.; Zhang, G.; et al. Federated Learning Meets Multi-Objective Optimization. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 2039–2051.
11. Zhang, F.; Shuai, Z.; Kuang, K.; et al. Unified Fair Federated Learning for Digital Healthcare. *Patterns* **2024**, *5*, 100907.
12. Liu, H.; Lu, J.; Wang, X.; et al. FedUP: Bridging Fairness and Efficiency in Cross-Silo Federated Learning. *IEEE Trans. Serv. Comput.* **2024**, *17*, 3672–3684.
13. Ezzeldin, Y.; Yan, S.; He, C.; et al. Fairfed: Enabling group fairness in federated learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 7494–7502.
14. Li, J.; Zhu, T.; Ren, W.; et al. Improve individual fairness in federated learning via adversarial training. *Comput. Secur.* **2023**, *132*, 103336.

15. Badar, M.; Nejdl, W.; Fisichella, M. FAC-fed: Federated adaptation for fairness and concept drift aware stream classification. *Mach. Learn.* **2023**, *112*, 2761–2786.
16. Xu, H.; Gao, S.; Zhu, J. Harmony in Diversity: Personalized Federated Learning Against Statistical Heterogeneity via a De-Personalized Feature Process. *Expert Syst. Appl.* **2025**, *290*, 128323.
17. Nishio, T.; Shinkuma, R.; Mandayam, N.B.; et al. Estimation of Individual Device Contributions for Incentivizing Federated Learning. In Proceedings of the 2020 IEEE Globecom Workshops (GC Wkshps), 7–11 December 2020; pp. 1–6.
18. Jiang, Z.; Xu, J.; Zhang, S.; et al. FedCFA: Alleviating Simpson's Paradox in Model Aggregation with Counterfactual Federated Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; pp. 17662–17670.
19. Li, H.; Li, X.; Liu, X.; et al. FedSam: Enhancing federated learning accuracy with differential privacy and data heterogeneity mitigation. *Comput. Stand. Interfaces* **2025**, *95*,104019.
20. Kuo, T.; Gabriel, R.; Koola, J.; et al. Distributed cross-learning for equitable federated models-privacy-preserving prediction on data from five California hospitals. *Nat. Commun.* **2025**, *16*, 1371.
21. Wang, Z.; Peng, Z.; Fan, X.; et al. FedAVE: Adaptive data value evaluation framework for collaborative fairness in federated learning. *Neurocomputing* **2024**, *574*, 127227.
22. Hu, J.; Zhang, H. FGS-FL: Enhancing federated learning with differential privacy via flat gradient stream. *Expert Syst. Appl.* **2025**, *288*, 128273.
23. Hari, P.; Singh, M. Adaptive knowledge transfer using federated deep learning for plant disease detection. *Comput. Electron. Agric.* **2025**, *229*, 109720.
24. Rahmati, M.; Pagano, A. Federated Learning-Driven Cybersecurity Framework for IoT Networks with Privacy Preserving and Real-Time Threat Detection Capabilities. *Informatics* **2025**, *12*, 62.
25. Yu T.; Kumar S.; Gupta A.; et al. Gradient surgery for multi-task learning. In Proceedings of Advances in Neural Information Processing Systems, 6–12 December 2020; pp. 5824–5836.
26. Ning, Q. On the momentum term in gradient descent learning algorithms. *Neural Netw.* **1999**, *12*, 145–151.
27. Fernandes, M.; Silva, C.; Arrais, J.; et al. Decay Momentum for Improving Federated Learning. In Proceedings of the ESANN, 14–16 April 2021.