

Review

A Fairness Perspective on Client Selection and Aggregation Methods for Non-IID Mitigation in Federated Learning: A Survey

Mohannad Alsofyani , Isra Alturaiki  and Hassan Mathkour 

Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia; ialturaiki@ksu.edu.sa (I.A.); mathkour@ksu.edu.sa (H.M.)

* Correspondence: 445107904@student.ksu.edu.sa

Abstract

Federated learning (FL) is a promising approach for training distributed machine learning models while preserving clients' data privacy. However, in real-world FL systems, data are often not independent and identically distributed (non-IID). This heterogeneity can slow convergence, degrade model performance, and increase client drift. To address these challenges, numerous methods have been proposed to mitigate non-IID data effects by optimizing client selection, local training, and model aggregation strategies. Despite their effectiveness in improving performance and efficiency, these methods rarely consider fairness across clients. Improving global accuracy does not guarantee balanced participation, influence, or outcomes, which may lead to biased model behavior across clients. In this survey, we review existing non-IID mitigation methods in FL from a fairness perspective and provide a systematic analysis of their implicit impact on client participation and influence. Unlike prior surveys that treat fairness as a separate research direction, this work analyzes how these methods designed for non-IID mitigation implicitly shape fairness outcomes across clients. Our taxonomy classifies existing methods into three categories—fairness-aware, semi-fairness-aware, and fairness-unaware—based on their design strategies for client selection and model aggregation. Using this taxonomy, we analyze the advantages, trade-offs, and limitations of each category and highlight that mitigating non-IID data does not guarantee fairness across clients. Finally, we identify open challenges and outline future directions, including system-level FL design that jointly considers non-IID mitigation and fairness and the development of standardized fairness evaluation metrics. Overall, this survey aims to provide a structured perspective on the relationship between non-IID mitigation and fairness and support the development of more balanced and scalable FL systems under non-IID conditions.



Academic Editor: George Angelos
Papadopoulos, Pingyi Fan
and Xinshun Xu

Received: 18 May 2026

Revised: 12 June 2026

Accepted: 16 July 2026

Published: 20 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: federated learning; non-IID; fairness; client selection; model aggregation

1. Introduction

Federated learning (FL) has emerged as a distributed learning paradigm that enables multiple clients to collaboratively train a global model without sharing their raw data [1]. Because data naturally resides across various devices that hold large volumes of private information, FL is particularly well-suited to privacy-sensitive applications. In healthcare, FL allows institutions to train models on sensitive patient records without exposing individual data, thereby improving clinical decision-making while preserving privacy [2]. In

finance, it supports decentralized risk modeling and fraud detection by allowing financial institutions to share insights without exchanging customer data [3]. In Internet of Things (IoT) environments, FL allows edge devices, such as sensors, wearables, and smart systems, to learn from distributed data across applications, including smart cities, environmental monitoring, and industrial systems, while respecting communication constraints and data privacy concerns [4].

Despite these advantages, FL faces significant challenges in real-world deployments. A central challenge is data heterogeneity, which is commonly referred to as not independently and identically distributed data (non-IID), whereby clients exhibit different local data distributions rather than drawing from a shared statistical population. This heterogeneity can appear in different forms, including quantity skew, feature skew, or label skew, and can substantially degrade model accuracy, increase communication and computation cost and slow convergence [5]. To mitigate these effects, prior work has proposed client clustering [6], regularization techniques [7], adaptive aggregation [8], and personalized learning [9]. Although these approaches improve accuracy, convergence speed, and communication efficiency, they are designed from a performance perspective.

The fairness implications of non-IID mitigation methods remain comparatively under-explored. In FL, fairness can be interpreted regarding balanced participation, influence, and model performance across clients with diverse data distributions. Performance-driven strategies, such as selecting clients with larger datasets or prioritizing highly informative updates, may implicitly favor certain clients while underrepresenting others. Consequently, a global model may achieve strong average performance while providing limited benefits to clients with scarce, minority, or unique data. This raises a critical question: does improving performance under non-IID data lead to fair outcomes?

The existing studies largely evaluate non-IID mitigation methods using performance and efficiency metrics that do not capture disparities across clients. Furthermore, fairness is typically treated as an independent objective rather than being examined in relation to non-IID mitigation strategies. As a result, there remains limited systematic analysis of how non-IID mitigation methods affect fairness across the FL pipeline.

The existing surveys categorize client selection and aggregation methods by mechanism, communication style, or application domain; however, these studies do not take fairness into consideration. This study addresses this gap by reviewing non-IID mitigation methods from a fairness perspective. Table 1 lists relevant survey studies on FL. For instance, Qi et al. [10] discuss model aggregation methods from the perspective of system architecture and synchronization. Mayhoub and Shami [11] present a survey of client selection methods in FL from the perspective of mechanisms and metrics. Ayeelyan et al. [12] provide a comprehensive survey of FL design principles and functional architecture. Almanifi et al. [13] discuss how non-IID mitigation techniques can improve model update and communication efficiency. However, these studies do not examine how non-IID mitigation methods affect fairness across clients.

To address this limitation, we analyze how client selection and aggregation strategies shape client participation and client influence on the global model. Based on this analysis, we propose a taxonomy that classifies existing non-IID mitigation methods into three categories—fairness-aware, semi-fairness-aware, and fairness-unaware—depending on whether fairness is explicitly considered in one or both stages of the FL pipeline. Using this taxonomy, we systematically compare existing methods and analyze their advantages, trade-offs, and limitations. We further examine current evaluation practices, including commonly used datasets, non-IID settings, model architectures, and application domains, and identify key challenges such as the lack of fairness-oriented metrics and standardized benchmarks.

Table 1. Comparison of previous surveys and reviews for non-IID, stages, and fairness in FL (✓: included, Θ : partially included, X: not included).

Ref.	Main Contributions	Year	Non-IID	Client Selection	Model Aggregation	Fairness
[12]	A review of the functionalities and design requirements of federated learning.	2024	Θ	Θ	Θ	X
[5]	Introduction to basic definitions, related technologies, and specific classifications of federated learning.	2021	Θ	X	X	X
[14]	A review of FL architecture, design, technical challenges, and comprehensive taxonomies covering various challenging aspects and contributions.	2021	Θ	Θ	Θ	X
[15]	A review of the impact of non-IID on communication efficiency, model convergence, and FL accuracy.	2024	Θ	X	Θ	X
[16]	A review of challenges posed by non-IID and classification of the existing solutions from the research entry point.	2022	✓	X	Θ	X
[17]	A review of current research work on handling non-IID data challenges with a discussion of their advantages and disadvantages.	2021	✓	X	✓	X
[18]	A review of client selection methods that mitigate the heterogeneity problem.	2023	Θ	✓	X	X
[19]	A systematic review of client-side challenges in federated learning.	2023	Θ	✓	X	X
[11]	A critical review of recent client selection methods in federated learning.	2024	Θ	✓	X	X
[20]	A categorization of the techniques used for efficient communication in FL and identification of the most widely used methods to mitigate non-IID and communication cost.	2024	Θ	Θ	Θ	X
[13]	A review of recent work regarding improving communication and computation efficiency in FL.	2023	X	Θ	Θ	X
[10]	A review of model aggregation methods with a categorization of the methods into four forms: synchronous, asynchronous, hierarchical, and robust.	2024	✓	X	✓	X
[21]	An exploration of the key notions of fairness in FL, identification of sources of bias, analysis of fairness-aware strategies, and evaluation of current research trends in fairness metrics.	2026	X	Θ	X	✓
[22]	A harm-centered framework for fairness in FL, connecting technical decisions to stakeholder impact, and offering actionable recommendations for more accountable, context-aware research.	2026	X	X	X	✓
Our survey	A review of FL methods that mitigate non-IID data from a fairness perspective across client selection and model aggregation stages.	2026	✓	✓	✓	✓

This survey offers a structured perspective on how non-IID mitigation methods affect fairness in FL, with particular attention to client selection and model aggregation. Its contribution lies in bridging performance-oriented mitigation strategies with fairness concerns, helping future methods better balance performance, efficiency, and fairness.

The main contributions of this study are as follows:

- (1) We provide a fairness-oriented reinterpretation of non-IID mitigation methods in FL.
- (2) We propose a taxonomy that categorizes methods such as fairness-aware, semi-fairness-aware, or fairness-unaware based on whether fairness is considered in client selection, model aggregation, or both stages.
- (3) We analyze the advantages, trade-offs, and limitations of each category, highlighting the relationship between fairness, performance, and efficiency.
- (4) We identify critical gaps in the current evaluation practices and outline future research directions toward more balanced FL systems.

Our study contributes to a better understanding of fairness in FL under non-IID data settings. Through a fairness-oriented analysis of client selection and model aggregation methods, it provides a structured framework for examining how fairness is incorporated into existing non-IID mitigation approaches. The significance of this work lies in identifying fairness-related design choices, highlighting research gaps, and supporting future research on balancing fairness, performance, and efficiency in real-world FL applications.

The remainder of this study is organized as follows. Section 2 presents the background of FL and discusses the challenges posed by non-IID data. Section 3 describes the research methodology. In Section 4, we introduce the proposed taxonomy and analyze existing methods accordingly. Then, Section 5 discusses the main findings and open challenges. Finally, Section 6 concludes this study.

2. Background in Federated Learning

2.1. Federated Learning Process

Federated learning is an emerging approach in decentralized machine learning. It enables training a shared model without transferring data from clients to the central server, thereby enhancing privacy preservation and limiting data leakage risks, reducing data transfer, and supporting learning from decentralized and heterogeneous data sources. In FL, each client updates the local model using its data and then sends the updates to the central server for aggregation into a global model. A federated learning system consists of two core components: (1) a client, where the data is held locally and model training occurs, and (2) a server where the updates from clients' models are aggregated, as shown in Figure 1. Federated learning can be classified by two deployment settings: cross-device, which involves a very large number of clients, such as smartphones or IoT devices, where data are small per client, and clients are unreliable; and cross-silo FL, which involves a small number of reliable clients, such as hospitals or enterprises, where each client holds a large dataset and participates consistently.

These deployment types influence the design of an FL system. The basic process in FL consists of the following three steps:

- Global model initialization: The server initializes the global model as the starting point of the FL process.
- Client selection: Selected clients receive the global model and train the model locally, and after updating the model parameters, the updated parameters are sent to the server.
- Aggregation: The server receives the updated parameters of each selected client and aggregates these updates using an aggregation method to generate a global updated model.



Figure 1. Federated learning system.

2.2. Non-IID Data in Federated Learning

The term non-IID refers to data that are not independent or identically distributed across clients, which differs from the traditional statistical assumption that training samples are drawn from the underlying distribution. In FL, this variation arises because clients differ in their environments, behavior, data sources, resources, or target tasks, leading to substantial differences in the content and distribution of local data [23]. Accordingly, non-IID data can be characterized at two levels:

- Intra-client non-IID: Heterogeneity within a single client's local dataset, where the client's own data may be imbalanced or uneven.
- Inter-client non-IID: Heterogeneity across clients, where each client holds a distinct data distribution.

For example, two devices collecting local user data may differ substantially in label distribution, feature patterns, and sample sizes, meaning that each client contributes different representative information to the training process. This heterogeneity manifests in several forms:

- Feature distribution skew: This pattern indicates a variation in the distribution of features or inputs across clients, assuming the label remains similar. This variance leads to unstable local updates during training, resulting in conflicting parameters during aggregation on the central server and negatively affecting the model's ability to generalize across clients. As a result, the global model may perform better on dominant feature distributions while providing lower accuracy for clients with distinct local characteristics, leading to performance disparities and fairness concerns across clients.
- Label distribution skew: The label distribution skew discrepancy relates to the different proportions of class labels across clients. Some clients may have concentrated on only a few classes; not all clients are equally represented among the target classes. This uneven distribution biases the global model toward frequently observed classes, reducing overall accuracy and potentially disadvantaging clients that contain rare or underrepresented classes, resulting in representations and performance disparities.
- Quantity skew: This pattern indicates the difference in the data size available with each client, and this leads to an impact on balancing during the training. The effect is particularly evident in FedAvg, where clients with larger datasets receive higher aggregation weights, giving them greater influence on the global model and potentially reducing fairness for clients with limited data.
- Dynamic heterogeneity: This form of heterogeneity acknowledges that label skew, feature skew, and quantity skew are not necessarily static. Concept drift occurs

when a client's data shifts across training rounds due to changes in user behavior, environment, or application context. Temporal skew occurs when clients contribute updates reflecting different time periods, often due to asynchronous or irregular participation. Both forms can cause the global model to become misaligned with the current data of some clients or to rely on outdated updates from others, introducing additional fairness risks.

Although system heterogeneity is distinct from non-IID data, it can interact with non-IID data by limiting the participation and aggregation influence of clients with valuable but resource-constrained data. Therefore, it is discussed only where it affects client selection and aggregation fairness.

These forms of heterogeneity create several interconnected challenges for FL, which require techniques and approaches to mitigate data heterogeneity by designing frameworks that adaptively address multiple challenges caused by non-IID types, rather than handling each separately [24]. In FedAvg [1], as shown in Algorithm 1, the server selects a random subset of clients each round, and these selected clients train the model locally for a defined number of epochs. The clients send the updated model back to the server, and the server aggregates the models by calculating a weighted average based on each client's data size. However, under non-IID data conditions, this mechanism accounts for client data quantity but does not address label skew, feature skew, or dynamic heterogeneity.

Algorithm 1 Federated averaging algorithm. The K clients are indexed by k , B is the local minibatch size, E is the number of local epochs, and η is the learning rate

Server executes:

```

initialize  $w_0$ 
for each round  $t = 1, 2, \dots$  do
     $m \leftarrow \max(C \cdot K, 1)$ 
     $S_t \leftarrow$  (random set of  $m$  clients)
    for each client  $k \in S_t$  in parallel do
         $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$ 
         $w_{t+1} \leftarrow \sum_{k=1}^K \frac{n_k}{n} \cdot w_{t+1}^k$ 

```

ClientUpdate (k, w): // Run on client k

```

 $B \leftarrow$  (split  $P_k$  into batches of size  $B$ )
for each local epoch  $i$  from 1 to  $E$  do
    for batch  $b \in B$  do
         $w \leftarrow w - \eta \nabla \ell(w; b)$ 

```

return w to server

The result is a global model that tends to favor dominant clients, dominant classes, and dominant feature distributions. The principal challenges arising from non-IID data are as follows, as presented in Figure 2:

- **Conflicting updates and slow convergence:** This challenge is produced by the different statistical distributions of data across clients, which leads to inconsistent updates that may conflict during aggregation, slowing convergence and increasing training, communication, and computation cost.
- **Bias and representation imbalance:** Different forms of non-IID data create different fairness challenges. Label skew can reduce representation of minority classes, feature skew can create performance disparities across clients, and quantity skew can allow clients with large datasets to exert disproportionate influence during aggregation.

Together, these effects may lead to participation imbalance, influence imbalance, and unequal model performance.

- **Poor generalization:** A global model optimized for dominant clients may perform poorly for clients with rare labels, distinct feature distributions, or limited data, motivating a personalization and fairness-aware approach.
- **Concept drift:** Changes in client data over time can make the global model less suitable for the current local distribution. Asynchronous participation further introduces updates from different time periods, creating fairness gaps between clients with recent and delayed contributions.

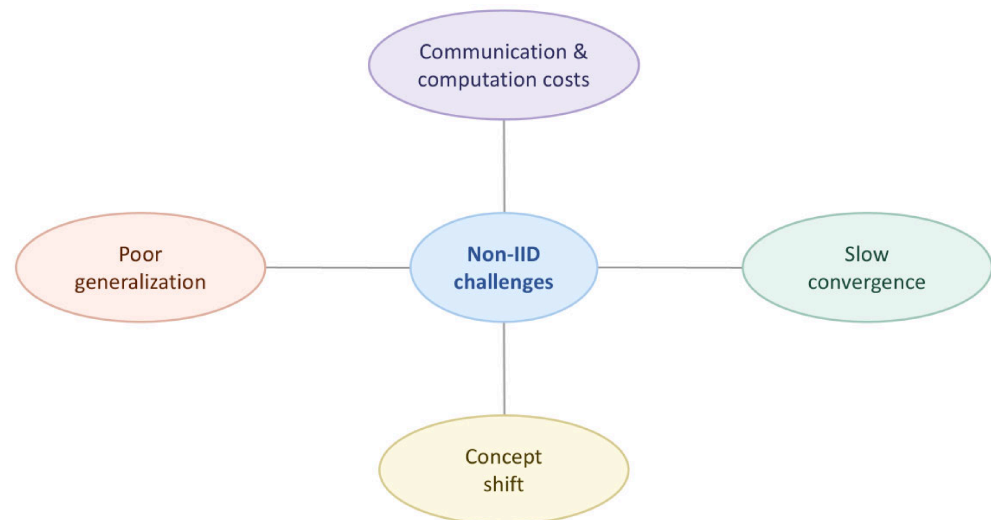


Figure 2. Challenges posed by non-IID in FL settings.

These challenges show that non-IID data can negatively affect convergence speed, communication and computation cost, stability, and fairness across clients. Client selection and aggregation methods are therefore needed not only to maximize global performance but also to promote stability, equitable influence, and fairness throughout the FL system.

2.3. Client Selection

Client selection refers to the process of choosing a subset of available clients to participate in each federated learning round. Since not all clients can participate simultaneously due to communication, computation, or availability constraints, the server usually selects only a fraction of clients in each round. In the FedAvg settings, clients are often selected randomly, as in Algorithm 1, which provides a simple and low-overhead selection mechanism. However, under non-IID data, client selection becomes more critical. Since clients may hold different label distributions, feature distributions, data quantities, or system capabilities, the selected clients in each round may not represent the overall federation. If the selected clients mainly come from dominant data distributions, larger datasets, or stronger devices, the global model may become biased toward those clients. In contrast, clients with minority labels, rare data patterns, smaller datasets, or limited resources may be repeatedly underrepresented. Therefore, client selection determines who participates in the learning process. From a fairness-aware perspective, this stage is important because exclusion at the selection stage limits which clients and data distributions can influence the global model in later stages.

2.4. Model Aggregation

Model aggregation refers to the process of combining local client updates to produce a new global model. After selected clients train local models using their private data, they send model updates, gradients, prototypes, logits, or other representations to the server. The server then aggregates these updates to generate the next global model. In FedAvg, aggregation is performed using data-size-based weighing. Although data-size-based aggregation is simple and efficient, it can create unequal influence under non-IID data. Clients with larger datasets or majority distributions may dominate the global update, while clients with smaller datasets, rare labels, or different feature distributions may have limited influence. Aggregation, therefore, determines how client knowledge is integrated into the global model. Consequently, aggregation design has a significant effect on model convergence, representation of heterogeneous data, and overall learning performance.

2.5. Fairness in Federated Learning

Fairness has become an important research topic in FL due to heterogeneity in client data, resources, participation, and contributions. Recent surveys have reviewed fairness notions, sources of bias, evaluation metrics, and mitigation techniques in FL [21], while also highlighting the need to consider fairness across the entire FL lifecycle and from multiple stakeholder perspectives [22]. To provide the necessary background for our taxonomy, this section briefly reviews the main fairness evaluation metrics used in FL and summarizes representative fairness-aware algorithms before focusing on fairness in non-IID mitigation methods.

2.5.1. Fairness Evaluation Metrics

Fairness evaluation in FL is challenging, particularly when data is heterogeneous, and there is a need to account for performance [21]. Fairness can be considered from different angles of the system, including model performance, training efficiency, client contribution update similarity, group-level outcomes, and temporal fairness [25]. Therefore, existing FL fairness studies use different metrics depending on the objective of the method and the application scenario. Common metrics include: accuracy, efficiency, distance-based measures, average variance, statistical parity difference, equal opportunity difference, Pearson correlation coefficient, and Jain's fairness index [21].

Although these metrics do not always use the terms participation fairness or influence fairness, they can support stage-level fairness analysis. For example, Jain's fairness index can be used to examine equality in client selection, resource allocation, or client outcomes. Cosine similarity and distance-based metrics can help analyze update alignment during aggregation. Average variance can reveal client-level performance disparity. However, many non-IID mitigation methods mainly report accuracy, convergence, communication cost, or robustness, without explicitly reporting fairness metrics. Therefore, this survey uses participation fairness and influence fairness as conceptual boundaries for the taxonomy. Table 2 summarizes how fairness-related behavior can be evaluated when such measurements are available.

2.5.2. Fairness-Aware Algorithms in Federated Learning

Fairness has received increasing attention in FL due to disparities in client data distributions, resource availability, participation frequency, and model contribution. To address these challenges, a variety of fairness-aware algorithms have been proposed that aim to improve participation fairness, contribution fairness, resource fairness, and performance fairness across heterogeneous clients. Existing fairness-aware approaches can be broadly categorized into five groups. Optimization-based methods formulate fairness as part of the

learning objective or as a constrained optimization problem, enabling explicit trade-offs between fairness and model performance [26–28]. Fair resource allocation methods seek to distribute communication, computation, or learning benefits more equitably between clients, with representative examples including q-FedAvg and logarithmic fairness-based approaches [29,30]. Reputation-based methods leverage clients' historical behavior, reliability, or contribution records to guide participation decisions and improve fairness over time [31,32]. Game-theoretic methods employ concepts such as cooperative games and Shapley values to evaluate client contributions and allocate rewards fairly [33,34]. Gradient-based methods improve fairness by modifying local updates, reducing gradient conflicts, or adjusting aggregation weights during training [35,36].

Table 2. Fairness evaluation metrics in FL.

Metric	Evaluation Objective	Advantages	Limitations	Applicable Non-IID Scenarios
Accuracy	Model performance across clients or groups	Simple and widely used	Does not directly measure fairness unless reported per client or group	Label skew, feature skew, concept drift
Efficiency	Training rounds or total training time	Captures training cost and convergence behavior	Does not directly measure fairness outcomes	Quantity skew, temporal skew
Euclidean distance	Difference between model, weight, or representation vectors	Useful for measuring model divergence	Sensitive to scale and high-dimensional updates	Feature skew, concept drift
Manhattan distance	Absolute differences between vectors	Simple and interpretable	Does not capture directional similarity	Feature skew, concept drift
Cosine similarity/distance	Directional similarity between client updates	Useful for detecting gradient conflict	Ignores update magnitude	Feature skew, concept drift
Maximum difference	Largest deviation between target and actual values	Highlights the worst-case disparity	Sensitive to outliers	Label skew, feature skew
Average variance	Variation in client performance over time	Captures performance disparity	Does not identify the source of disparity	Label skew, feature skew, concept drift
Statistical parity difference	Difference in positive outcomes between groups	Useful for group fairness	Requires sensitive or group attributes	Label skew
Equal opportunity difference	Difference in true positive rates between groups	Useful for classification fairness	Requires labels and sensitive attributes	Label skew
Pearson correlation coefficient	Alignment between estimated and true contributions	Useful for contribution evaluation	Requires reliable contribution estimates	Quantity skew
Jain's fairness index	Equality in allocation or achieved fairness between clients	Bounded and easy to interpret	Equal allocation may not always mean fair treatment under non-IID data	Quantity skew, temporal skew

3. Research Methodology

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [37]. The method used to conduct this research is as follows: First, the research questions are described, followed by the inclusion and exclusion criteria. The search strategy is then stated; finally, the study selection process is presented.

3.1. Research Questions

This survey is guided by the following questions:

RQ1. What client selection and model aggregation approaches have been proposed to mitigate the effects of non-IID data in FL?

RQ2. How is fairness incorporated into client selection and model aggregation mechanisms within existing non-IID mitigation methods?

RQ3. How can non-IID mitigation methods be classified according to fairness awareness in client selection and model aggregation?

RQ4. How do existing studies evaluate fairness, performance, and efficiency under non-IID FL settings?

RQ5. What research gaps and future directions remain for developing fairness-aware systems under non-IID data?

3.2. Inclusion and Exclusion Criteria

The inclusion criteria are as follows:

- Articles focused on federated learning under non-IID settings.
- Articles focused on client selection, model aggregation, or fairness strategies.
- Articles published during the period from 2021 to 2026.
- Studies written in English.

The exclusion criteria:

- Articles focused only on privacy, security, blockchain, or adversarial attacks without relevance to non-IID mitigation or fairness.
- Articles focused only on general FL without addressing non-IID data, client selection, and model aggregation.

3.3. Search Strategy

Researchers conducted the search strategy across multiple databases, including IEEE Xplore, Google Scholar, Scopus, and Web of Science, to comprehensively retrieve relevant studies. They used a combination of keywords such as “federated learning”, “client selection”, “model aggregation”, “non-IID”, “data heterogeneity”, “fairness”, and “fairness-aware federated learning”. To refine the search results and ensure thorough coverage of the literature, they applied operators (AND/OR).

3.4. Study Selection

As shown in Figure 3, the initial search phase retrieved 442 records. All references were imported into Zotero [38], where 65 duplicate records were identified and removed, leaving 337 studies for screening. Titles and abstracts were then screened according to the predefined inclusion and exclusion criteria. At this stage, 249 studies were excluded because they were not relevant to FL under non-IID settings, client selection, and model aggregation, or because they were survey or review articles. This process left 128 reports for full-text retrieval. However, 15 reports could not be retrieved. The remaining 113 studies were assessed for eligibility. After applying the predefined criteria, 43 studies were excluded for two reasons: (1) the study did not present a method that could be mapped to the client selection or model aggregation stages of the FL pipeline, even if it mentioned non-IID data

in the abstract (25); (2) the study did not provide enough methodological, algorithmic, or evaluation detail to determine how the method addressed non-IID data in the client selection and/or model aggregation stages ($n = 18$).

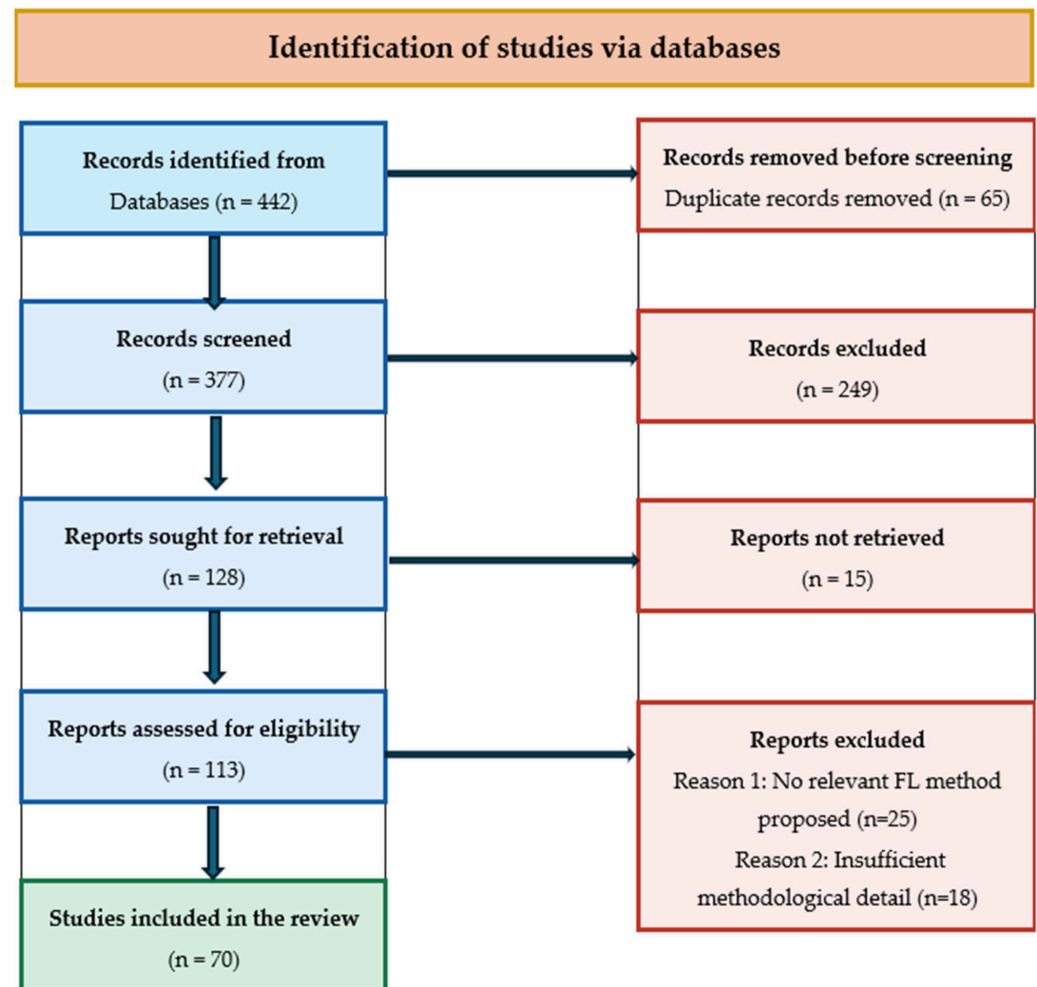


Figure 3. PRISMA flow diagram.

4. Taxonomy of Non-IID Mitigation Methods from a Fairness-Awareness Perspective

In this section, we present our proposed structured taxonomy for analyzing federated learning methods that aim to mitigate non-IID data. Existing studies classify non-IID mitigation methods based on technical mechanisms, such as client selection, aggregation modification, clustering, personalization, regularization, communication efficiency, or robustness. In contrast, this survey provides a complementary perspective by examining whether fairness is explicitly considered in the main stages of the FL pipeline. In developing the proposed taxonomy, we adapt the taxonomy development method introduced by Nickerson et al. [39]. The purpose of our taxonomy is to support researchers and developers of FL systems in identifying where fairness is considered within non-IID mitigation methods. Following the conceptual-to-empirical approach, our taxonomy dimensions were derived from the theoretical structure of the FL pipeline and then validated against the reviewed methods. The meta-characteristic of the taxonomy is fairness awareness in the FL pipeline. Based on this meta-characteristic, the taxonomy classifies methods according to whether fairness is considered in two key stages: client selection and model aggregation. Since FL

operates as a repeated cycle of client selection, local training, and aggregation, fairness should be examined at the system level rather than in a single component only.

This survey adopts a stage-level view of fairness in non-IID mitigation methods in FL. We look at fairness from a high-level perspective. Instead of limiting our definition of fairness to methods that directly optimize fairness as an objective, we consider whether fairness choices are incorporated into the design of non-IID mitigation methods in FL, even if the overall system is optimized for other objectives, such as convergence speed, communication efficiency, or model accuracy under non-IID data. In our taxonomy, we focus on two dimensions: participation fairness in client selection and influence fairness in model aggregation. These two dimensions are chosen because they represent the main processes in an FL system that affect fairness. Thus, an FL method (F) can be represented, using two dimensions, as $F = \{CS, MA\}$, where CS denotes client selection, and MA denotes model aggregation. Each dimension can take one of the two values: $CS \in \{\text{Included}, \text{Not Included}\}$ and $MA \in \{\text{Included}, \text{Not Included}\}$. A stage is considered *included* when the method design directly addresses fairness in client selection and model aggregation. For client selection, this means considering participation fairness through the inclusion of diverse, minority, underrepresented, resource-constrained, or otherwise disadvantaged clients. For model aggregation, this means considering influence fairness through mechanisms that balance or justify the influence of heterogeneous clients on the global model. A stage is considered *not-included* when the mechanism mitigates non-IID data without addressing these fairness considerations.

The two dimensions produce four possible combinations: fairness in client selection and model aggregation, fairness in client selection only, fairness in model aggregation only, and fairness in neither stage. However, for analytical purposes, the two middle cases are grouped as semi-fairness-aware since they share the common characteristic of partial fairness coverage regardless of which stage is addressed. Consequently, methods that incorporate fairness in both stages are classified as fairness-aware, methods that incorporate fairness in only one stage are classified as semi-fairness-aware, and methods that do not incorporate fairness in either stage are classified as fairness-unaware.

For example, a method that promotes balanced participation across heterogeneous clients but uses a standard aggregation FedAvg is classified as semi-fairness-aware because fairness is considered only during client selection. Similarly, a method that adjusts aggregation weights to reduce unequal client influence but uses random or performance-driven client selection is also classified as semi-fairness-aware because fairness is considered only in model aggregation. A method is classified as fairness-aware only when fairness considerations are incorporated into both stages. In contrast, a method that improves convergence speed, communication efficiency, or model accuracy under non-IID data without addressing participation fairness or influence fairness is classified as fairness-unaware. Therefore, the proposed taxonomy should be interpreted as a classification of fairness awareness in method design rather than a judgment of whether a method ultimately achieves fair outcomes. The taxonomy identifies whether fairness is incorporated into client selection and/or model aggregation and provides a structured framework for analyzing non-IID mitigation methods from a fairness-awareness perspective.

The proposed taxonomy, illustrated in Figure 4, provides the decision logic used to classify reviewed methods based on the incorporation of fairness within different stages of the FL pipeline under non-IID data settings.

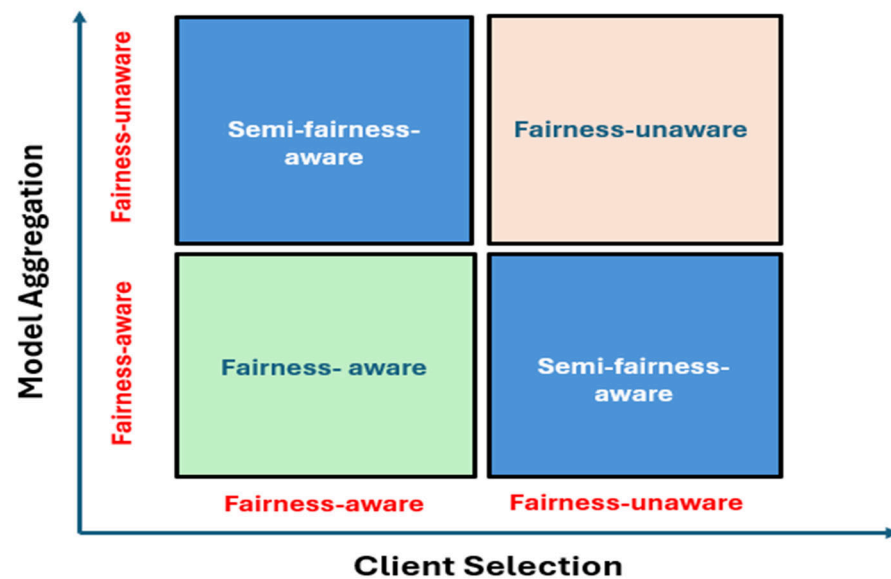


Figure 4. Fairness-aware taxonomy of non-IID mitigation methods. Methods are classified according to whether fairness is explicitly considered in the client selection or model aggregation stages, in both stages, or in neither stage.

4.1. Fairness-Aware Non-IID Mitigation Methods

Fairness-aware methods explicitly consider fairness in client selection and model aggregation. In these methods, client selection is designed to improve fair participation or representation, while aggregation is designed to regulate client influence and reduce dominance by specific clients or data groups. Therefore, these methods address both key fairness concerns under non-IID data: who participates and whose updates shape the global model. FedFast [40] addresses label and feature skew by proposing ActvSAMP to cluster clients using K-means and select delegates of each cluster to ensure diversity in the pre-training stage. ActvAGG uses delegates to train the model and sends the updates to the server in the post-training stage, and then the server propagates the updated model to other clients in the post-aggregation stage. This method aims to reduce communication costs and improve accuracy. However, it depends heavily on the quality of the clustering and the similarity metric. The approach by the authors of [41] focuses on mitigating label and quantity skew by proposing a reputation-enabled aggregation method, which calculates the reputation score depending on the comparison of accuracy with the current and previous models, as well as a temporary global model. The reputation score is calculated in the post-training stage to select clients, and the client with a lower score than the defined threshold is rejected. The clients who passed the threshold are accepted to the aggregation stage, where each client is weighted based on the contribution according to the normal distribution. This method aims to filter the clients in the client selection stage and weight clients based on their contribution, which leads to improved convergence and accuracy. Despite these improvements, the method requires a predefined reputation threshold to identify and exclude poor-quality clients. The authors of [42] leverage non-IID by proposing the CATA-Fed method, which is divided into client selection and aggregation stages. In the client selection stage, the clients apply adaptive local update (ALU), starting with local training to measure how many epochs each client is able to train in a defined time. Based on this measure, the server then detects the stragglers and applies a straggler-mitigating scheme (SMS) to prevent stragglers from delaying training by dividing the data into sub-clients to use in each round. After SMS, cluster fair selection (CFS) is applied to cluster clients based on the data size, and within each cluster, proportional fair selection (PFS) is applied to select clients across rounds. In the aggregation stage, FedAvg weights each

client's contribution by the amount of data it used in that round, including stragglers. ALU and SMS were used to validate the updates, CFS was employed to fairly select clients, and aggregation was utilized to balance the weight, which leads to higher accuracy and improved convergence speed. G-FML [43] mitigates non-IID data when client data are highly skewed by proposing group-based federated meta-learning G-FML, which groups similar clients together and then separates learning meta-models into the groups. Each client trains a local autoencoder to extract a representation of its data distribution, and then the server applies K-means clustering to group clients with similar distributions. After that, within each cluster, MAML is applied to train a group-specific meta-model. This leads to outperforming baseline methods under strong non-IID. One concern is that clustering quality depends heavily on the autoencoder's representation, and dynamic regrouping introduces extra computation and communication overhead. CFMTL [44] focuses on label and quantity skew by clustering clients based on model similarity using the Earth Mover's Distance (EMD) in the post-training stage. Within each cluster, a random selection is made in each round to ensure fair client selection in the pre-aggregation stage, and then FedAvg aggregation is applied within each cluster. After that, proximal regularization is applied to adaptive personalization, and knowledge is transferred between clusters to prevent isolation between them in the post-aggregation stage. This method leads to results that outperform the basic FL regarding accuracy and communication efficiency.

4.2. Semi-Fairness-Aware Non-IID Mitigation Methods

Semi-fairness-aware methods consider fairness in only one stage of the FL pipeline. Some methods improve fairness in client selection but use standard or performance-driven aggregation. Other methods use random or performance-driven selection but introduce aggregation mechanisms that regulate influence, correct update bias, or reduce dominance under non-IID data. Therefore, these methods partially address fairness, but they do not provide a system-level fairness design across selection and aggregation.

4.2.1. Fairness-Aware Client Selection with FedAvg Aggregation

In federated learning, fairness can be introduced at different stages of the training process. Some methods apply fairness-aware client selection while using the standard FedAvg aggregation rule. In this approach, the server carefully selects clients using a structured mechanism such as clustering or optimization to ensure representation of diverse data distributions. However, the aggregation stage is still performed using FedAvg, which weights the updated client's model based on local data size, and this leads to an imbalance. As a result, the system can be described as participation-fair but aggregation-unfair. Although fairness is considered at the selection stage, statistical dominance may still appear during aggregation, especially under non-IID data.

Several studies [6,45–48] use clustering clients with similar characteristics to control which clients participate in each round. The main objective is to improve representation under heterogeneous data. This can be carried out during pre-training based on local dataset characteristics and geographical location or during post-training based on local model training time and parameters.

The authors of [45] focus on mitigating label distribution and quantity skew by clustering clients based on each client's sample size or the similarity of their local model updates in the pre-round stage. The authors then select samples from each cluster using multinomial distribution sampling to ensure diverse representation. While this method ensures that clients with unique data are more likely to be clustered, the clustering process becomes expensive when scaling to a very large number of clients.

The authors of [6] propose Heterogeneity-Aware Clustered Client Selection (HACCS), in which clients send the similarities found in statistical data distributions, computational resources, and training loss to the server. After that, the server clusters clients with similar data distributions during the pre-round stage, and each cluster is assigned a probability of selection based on its average latency and loss. Clusters with lower latency and higher loss are more likely to be selected, and then the representative fastest clients within each cluster are selected in the pre-aggregation stage. HACCS aims to ensure diverse data representation across clients to mitigate data heterogeneity and reduce convergence time. The method by the authors of [46] introduces an irrelevance score metric to select the best client in an imperfect heterogeneous environment. The irrelevance score is calculated based on several factors, including data size, class imbalance, and the degree of non-IID distribution. Each client then sends the score to the server during pre-training. This score helps in detecting clients who are small-data free riders. The server sorts the scores using the time sort algorithm, groups them into clusters, and selects the top clients of each cluster. After each training round, the server updates the score and re-evaluates the client selection for the next rounds. This method aims to balance these factors to form a balanced client group, leading to faster convergence even in highly skewed data distributions. SCS [47] introduces a stochastic integer-programming-based client selection, which balances performance and cost while accounting for uncertainty in client reputation. In this method, clients are clustered into three categories: low, medium, and high reputation. The categories are based on the quality of their local gradient updates, which are evaluated by the server in each round. The server formulates an optimization problem to select the best set of clients from each cluster, which minimizes the total training cost while maintaining strong model performance. However, this method assumes that high reputation equals reliability. The work by the authors of [48] discusses the label distribution skew in the IoT environment, which degrades performance, and proposes the FEDGS framework based on the GPB-CS algorithm to cluster clients by location and then selects a subset from each cluster to form a homogeneous super node. The GPB-CS algorithm is applied during the pre-training stage based on statistical information about each client's label distribution. This method aims to verify the distribution before accepting clients' updates, thereby improving accuracy and reducing the number of training rounds. While effective, this method depends on the assumption that clients are naturally clustered by static location, but it fails in dynamic federation environments, such as mobile networks.

Cluster-based methods group clients according to similarity, reputation, or distribution metrics. They rely on predefined structures or periodic updates, while optimization-based selection treats client selection as a sequential decision problem that adapts over time based on feedback and estimated rewards. However, both approaches aim to reduce non-IID effects at the selection stage, where contribution weighting remains dependent on data size using standard FedAvg.

Unlike clustering methods, optimization methods treat client selection as a decision problem with a clear goal and well-defined constraints. Instead of random client selection, the server evaluates each client based on measurable factors such as data distribution, time cost, and fairness. These factors are then placed inside a formal optimization model that guides the selection process. RBCS-F [31] proposes a reputation-based client selection with fairness. It is designed to improve fairness and efficiency within a mobile edge computing environment. This method considers fairness and the total time each client spends downloading, training, and uploading their model. Each client reports information from the previous round, including its computation capacity, transmission power, channel condition, and model size. Using this information, the server applies a Contextual Combinatorial Multi-Armed Bandit (C2MAB) framework to estimate the expected time required for com-

munication and training with each client. At the same time, a Lyapunov optimization framework ensures that the selection remains fair and diverse over time. This method improves training efficiency while maintaining controlled fairness across clients compared with baseline methods. However, fairness is defined as equal participation opportunity, not equal model benefit, which leads to selecting clients that have a poor local model.

The approach by the authors of [49] models client selection as a combinatorial multi-armed bandit problem. The key idea is that the gradients of local models reflect the class distribution of each client's training data. Unlike clustering methods, this approach learns from previous rounds and dynamically updates its strategy. To estimate the class distribution, the method uses a shared global dataset to analyze the updated local gradients from clients. Based on this information, an estimation scheme is applied to approximate the class balance between clients. Then, the Contextual Multi-Armed Bandit (CMAB) algorithm is used to select a subset of clients whose combined label distributions form a more balanced dataset. In each round, when the server receives model updates from the clients, the estimation scheme is computed, and then the CMAB algorithm selects a client group. This approach aims to achieve fair class representation and improve the global model's convergence performance. While effective, an auxiliary dataset is a practical constraint in real deployments. Wang et al. [50] propose an efficient FL framework that mitigates non-IID data using an evolutionary algorithm-based client selection. The genetic algorithm selects base clients whose combined data distribution approximates the overall distribution, helping include diverse and underrepresented clients. However, aggregation follows FedAvg with data-volume weighting and does not explicitly address influence fairness.

4.2.2. Fairness-Aware Client Selection with Performance-Driven Aggregation

In this category, client selection is fairness-aware, while aggregation uses unfair optimized weighting strategies rather than FedAvg. These methods move beyond simple data size averaging by introducing computational power, updating freshness, communication delay, or adaptive weighting parameters. However, the aggregation objective remains performance-driven rather than fairness-driven. Therefore, these methods can be described as participation-fair with efficiency-optimized aggregation, rather than a fully fairness-aware federated learning system. The authors of [51] address label distribution skew by proposing a decentralized approach called E-Tree, depending on the hierarchical tree structure. KMA builds a cluster by accounting for data similarity and network distance to ensure intra-cluster aggregation quality during the post-training stage and then selects a public node to average updates from its members. After these decentralization updates, each cluster sends its aggregated model to the central server for final global aggregation. This method outperforms the benchmark approach in accuracy and convergence speed. This assumption may not hold under highly dynamic environments, as the data similarity and network distance criteria are not enough to guide the tree. The technique by the authors of [52] focuses on mitigating the label distribution variance between clients by proposing a group-based client scheduling (GS) technique that groups the clients into several clusters such that their label distributions complement each other, based on a new metric called Group Earth Mover's Distance (GEMD) in the pre-training stage. GEMD measures the deviation between each cluster and the global distribution. In each round, the server selects a client from the cluster with the lowest GEMD value, indicating that the cluster represents a balanced dataset. The GS technique forms complementary client groups and selects the most balanced ones, leading to improved performance of the FL algorithm compared with random client selection. However, the method assumes there is a trusted third party that helps prepare the information needed for grouping. The framework by the authors of [53]

mitigates the label and quantity skew by proposing an Eiffel framework that integrates fairness client selection by scheduling based on computation cost, data size, local loss, and age of update (AoU) in the pre-training stage and weights each client's contribution in the aggregation stage based on system capability. This method aims to select clients who ensure the system is efficient and fair. Although promising, it introduces privacy leakage indirectly. The authors of [54] propose SACW, a semi-asynchronous FL framework that mitigates non-IID and system heterogeneity using DBSCAN-based client clustering and adaptive weighted aggregation. The method selects one representative client from each label distribution cluster, helping to include minority and underrepresented groups in each round. However, aggregation weights depend on data size and staleness, which improves efficiency but does not ensure balanced client influence. Sun et al. [55] propose LN-FHAR, a two-stage FL framework that addresses label noise and non-IID data through client grouping, reliable-neighbor support, and prototype regularization. The method retains high- and low-quality clients, supporting noisy or minority-distribution clients rather than excluding them. However, aggregation uses sample-size and quality-based weights, which improve robustness but do not ensure balanced client influence.

4.2.3. Random Client Selection with Fairness-Aware Aggregation

In this category, clients are selected randomly without explicit fairness control. The participation probability is not adjusted for diverse data distribution or class imbalance. However, aggregation is designed to correct statistical bias, reduce client drift, or balance influence under non-IID data [56]. FedDist addresses the label and quantity skew by using different model architectures. If a neuron from a client is too dissimilar from what exists in the global model, a new neuron is added to the global architecture to preserve specific client knowledge, and after adding new neurons, FedDist performs layer-wise training rounds so new layers can adapt to the expanded representation. This method increases the model's capacity to represent clients instead of correction-based methods, and this leads to a balance between personalization and generalization. However, if many clients introduce unique patterns, the global model may expand continuously, which then increases memory and communication costs. The authors of [57] focus on mitigating label skew by introducing a variance reduction mechanism using a correction term that estimates how much each client update is biased from the global model. Even if clients are randomly selected, the control terms in local training for the selected clients reduce the variance between local and global updates. This result demonstrates an improvement in accuracy and reduced communication rounds. The framework by the authors of [58] leverages feature distribution skew by proposing the FedAAR framework, which addresses two main problems in non-IID. First, it reduces the client drift during local training by using the prototype-guided local update (PLU) module to learn consistent features in post-training. Second, it reduces conflicts between the local gradients using the cosine distance during the global aggregation by using gradient-refinement-based aggregation, GRA. However, the validation scale is small. FedBGVS [59] mitigates label skew in the global model; it balances the weighting by using a balanced global validation score (BGVS) at the server to evaluate each updated model received from clients after local training, and then it adjusts the client's weight in the aggregation stage. This method improves performance compared with baseline methods. This design choice assumes that the server must own a balanced validation dataset, which may not be available in privacy domains. FedFusion [60] addresses the challenge of model aggregation in a non-IID environment by depending on model statistical information instead of data (data-agnostic). It infers the global distribution from clients' statistical parameters in the post-training stage. Based on these statistical parameters, the Variational AutoEncoder (VAE) validates the consistency of client updates

and then adjusts the updates according to their suitability for the virtual components. The method aims to validate and balance the weight of the client's contribution by inferring the global distribution during the pre-aggregation stage. Although the method achieves significant performance improvements over the state of the art, training the VAE at the server each round adds computational overhead. The authors of [61] focus on mitigating label and quantity skew by proposing the FLT method, which discovers the related clients by extracting and embedding data for each client and uses K-means to compress and send the data as a signature in the pre-training stage. On the server side, UMAP is applied to build an adjacency matrix that can be used directly by the FedAvg aggregation or to form clusters in the pre-aggregation stage, thus improving accuracy and speeding up the convergence. However, UMAP depends on a pre-trained encoder distributed to all clients, which introduces a form of prior knowledge that may not always be available. The method by the authors of [62] addresses label/quantity skew as well as temporal skew, FlexCFL clustering clients based on EDC, newcomers based on cosine similarity, controls the temporal shift by the Wasserstein distance in the post-training stage, and in the intra-group, it applies FedAvg, where in the inter-group, semi-pluralistic aggregation is applied by defining a specific value. This method improves accuracy over baseline clustering and standard FL methods. While this method addresses a static clustering limitation that was ignored in clustered FL research, it introduces several tunable components that increase system complexity. The method used by the authors of [63] leverages the challenge of model aggregation in a non-IID environment, specifically regarding label distribution and quantity skew. This paper proposes FedEWA, a novel parameter-wise elastic weighted averaging method that replaces model-level averaging, such as FedAvg. After local training, each client computes the important coefficients for its parameters and sends them to the server. The server then performs elastic parameter-wise aggregation in the pre-aggregation stage. This process validates the importance of each parameter before the aggregation, which improves convergence efficiency. Despite the reported gains, it requires clients to compute the importance of the coefficient for each model parameter during the local training. The authors of [8] propose the FedDyn method to mitigate label and quantity skew by modifying the local objective with a dynamic regularizer to align local solutions with the global solution. Instead of each client minimizing only its empirical loss, the local objective is modified so that the local minima align more closely with the global minima. FedDyn achieves faster convergence and reduces the number of communication rounds compared with FedAvg, FedProx, and SCAFFOLD. In contrast, the method requires maintaining and updating variables on the server, which increases memory overhead. The authors of [64] address the resource-constrained reality of edge devices by introducing the FedGKT framework, which is a group knowledge transfer training algorithm. Clients train small CNNs on their local non-IID data during training. Instead of sending full model weights, they transfer their knowledge via knowledge distillation to a large CNN on the server side during the pre-aggregation stage. The goal is to build a hybrid setup in which the small client model adapts to each user's data, while the server model collects information from all clients. This method reduces client computation and communication while maintaining accuracy slightly better than FedAvg. While effective, feature map sharing may leak more information than parameter averaging. QuPeD [65] addresses label distribution skew and resource heterogeneity. This method combines personalized FL, model quantization, and knowledge distillation into a unified framework. This method enables clients to collaborate via knowledge distillation, allowing each client to adapt a personalized model while still benefiting from a shared global model. Clients use compressed models of varying sizes to accommodate local hardware limitations. Even when models are quantized, the method outperforms baseline methods. The gains are achieved at the expense of communication

overhead and potential privacy exposure through model outputs. FeSEM [66] focuses on mitigating label and feature distribution skew by proposing a centralized method that measures models for each client after local training, sends them to the server, and assigns each client to the appropriate cluster. Then, the server updates each center by aggregating the clients' models and sends the updated model back for local fine-tuning. This stochastic expectation–maximization (EM) algorithm repeats until each client is assigned to the correct cluster and the model is improved for that cluster. This method achieves higher accuracy than baseline methods. The approach relies heavily on clustering quality. The authors of [67] propose a two-timescale hybrid federated learning TT-HF, which is a semi-decentralized architecture composed of two components: device-to-device communication and device-to-server communication. Clients are first grouped into clusters based on their communication capabilities; within each cluster, neighboring devices interact through D2D rounds to exchange and aggregate their local model parameters before global aggregation. Then, a subset of clients sends their updated models to the server, which periodically aggregates the cluster models to update the global model. This hybrid approach uses decentralized learning inside clusters with centralized coordination across clusters. This method achieves better accuracy and faster convergence than standard federated learning. However, it depends on appropriate clustering and reliable D2D connectivity. The authors of [68] focus on mitigating the challenges of label and quantity skews and communication costs by proposing a disentangled Information Bottleneck (IB) principle-based loss function to update the local model in a manner that keeps the valuable information for prediction while discarding unnecessary details. The authors also propose a robust model aggregation approach based on mutual information (MI) to prune outlier clients on the server side. This method aims to validate the effectiveness of client updates and the quality of the aggregation, thereby reducing communication rounds and improving test accuracy. However, computing and estimating mutual information can introduce additional computational overhead. The authors of [69] introduce mitigating features and quantity skew by improving the integration of local training and aggregation. To extract features, a deep learning model combined with multiscale convolution, an attention mechanism, and a multistage residual connection is employed in the post-training stage. Moreover, to improve the efficiency of the aggregation, a Kalman filter, FKF, is combined with cubic exponential smoothing (CES) in the pre-aggregation stage. This method aims to give more weight to reliable clients, reduce the impact of noisy updates, and maintain the global model's stability. The authors of [70] introduce FOCUS, an FL optimization method that addresses non-IID and arbitrary client participation using push–pull gradient tracking. The method accumulates gradient tracking variables over time, preserving contributions from infrequently participating clients and reducing aggregation bias.

4.2.4. Performance-Driven Client Selection with Fairness-Aware Aggregation

In this category, client selection is not fairness-aware; clients are selected based on performance indicators, such as loss or predefined criteria that may favor certain participants. However, the aggregation introduces corrective mechanisms such as credibility-based weighing or hierarchical validation. In this category, fairness is partially enforced at the aggregation stage, even though selection is not balanced.

The authors of [71] mitigate label distribution and quantity skew, where some clients have rare classes, which leads to high loss in the global model. This study proposed an active client selection method that estimates how useful each client's data is for improving the current model. Clients with a higher estimation value are more likely to be selected, with some random sampling to maintain diversity. This approach focuses on informative updates and avoids wasting resources. The method remains

sensitive to prioritizing high estimation values, which may unintentionally bias toward specific clients. The authors of [72] propose a blockchain-based, decentralized framework for cross-silo FL that leverages blockchain and homomorphic encryption. Clients are selected depending on data size and model quality. In the aggregation stage, the method adapts a novel hierarchical task based on a credibility score that depends on the quality of local model updates, thereby indirectly mitigating label and quantity skew. Although selection favors strong clients, aggregation attempts to regulate influence using a credibility score. The method improves accuracy and convergence compared with baseline methods. However, the credibility score depends on the quantity of data and the quality of the model, which leads the system to bias toward larger clients. FAIR [73] leverages label and quantity skew via an integrated framework that combines a client selection strategy based on the quality of data through reverse auction in the pre-training stage to ensure maximization of the total learning quality under a budget constraint and an aggregation strategy using cosine similarity, which is calculated based on the quality and size of the data to validate the updates in the pre-aggregation stage. This method improves performance compared with selection and aggregation baseline methods. Although promising, quality filtering excludes statistically diverse clients. SHARE [74] enhances accuracy and reduces communication costs by mitigating the label distribution skew. SHARE formulates communication cost as an optimization problem by selecting a subset of nodes as edge aggregators and assigning clients to each edge aggregator by using the greedy node association algorithm (GoA). The GoA considers the per-round communication cost and mean Kullback–Leibler divergence to form balanced data distributions at the edge. After this initial assignment, SHARE uses local search (LoS) to optimize aggregator selection. This method aims to reshape the data distribution in the aggregator to achieve stability. However, the optimization is centralized and may become computationally heavy on a large scale. The authors of [75] focus on mitigating label and feature distribution skew by proposing the MH-FL method, which converts a star topology into a multi-layer cluster structure. The MH-FL method groups clients into clusters and performs D2D collaboration within each cluster during the post-training stage. Clients exchange updates to reach a local agreement before sending any information upward. At the inter-cluster level, fog nodes are used as intermediate aggregators to perform pre-global aggregation before the final global update. This structure reduces communication costs, enhances resource utilization, and speeds convergence.

4.3. Fairness-Unaware Non-IID Mitigation Methods

Fairness-unaware methods aim to mitigate non-IID data by improving accuracy, convergence, communication efficiency, robustness, or personalization, but they do not explicitly incorporate fairness considerations in either client selection or model aggregation. These methods may still improve performance under heterogeneous data. However, their design does not directly address fair participation or fair influence.

4.3.1. Performance-Driven Client Selection with FedAvg Aggregation

TiFL [76] addresses heterogeneity in data and resources, which significantly impacts training time and model accuracy in FL systems. This paper proposes the Tier-based Federated Learning (TiFL) method, which is a clustering-based client selection approach. Each client reports its training time to the server, and TiFL groups clients into tiers using measured training latency. Then, in each training round, the server selects clients from the same tier to participate. After each round, an adaptive tier selection mechanism updates the probability of selecting each tier based on the global model's accuracy during post-training.

The aim is to trade off training time against accuracy rather than fixing a static policy. Under data heterogeneity, it reports an approximately $3\times$ speedup and about a 6% accuracy improvement over a vanilla baseline while increasing training speed. While the method makes rounds comparable in completion time, tiering by training latency reduces the data diversity per round. FedCor [77] introduces a correlation-aware selection mechanism built on Gaussian process modeling. Instead of estimating each client's usefulness separately, it treats loss reduction as a function over the client space and models that function using a Gaussian process. After each training round, the correlation matrix is updated depending on the client data and loss. FedCor aims to select a group of clients with complementary influence and diverse data distributions, rather than just clients with high losses that are highly correlated. This strategy prefers sets of clients that jointly maximize expected global loss reduction and achieve diverse representation and data balance while reducing communication costs and improving convergence speed and stability. While the strategy focuses on identifying which combination of clients is informative rather than which clients are good, it may avoid rare clients because they appear uncorrelated or noisy. The authors of [78] focus on mobile edge computing (MEC) environments where clients typically encounter non-IID data, particularly regarding label and quantity distribution skew. This study proposes the CSFedAvg method, which calculates the degree of weight divergence to measure the heterogeneity of clients' data compared with a trained model on auxiliary IID data on the server side. The server creates an auxiliary IID dataset to train a reference model that acts like a compass pointing toward an unbiased direction, and then it calculates the weight divergence between each client's local model and the auxiliary model. Based on this measure, the server selects clients with lower divergence to participate in the following training round. It achieves noticeably higher accuracy in fewer rounds. However, client selection is biased toward clients with lower divergence and excludes clients with unusual data. The framework by the authors of [79] addresses label skew by proposing an AUCTION framework. This method uses a reinforcement learning model based on an encoder–decoder agent that selects the most suitable clients that maximize the accuracy under a budget constraint. The AUCTION method is applied in the pre-training stage to evaluate client data quality based on mislabeled samples and class distribution, which can be indirectly measured from the global model's loss. Each client estimates the amount of mislabeled data by checking the global model's loss on its local dataset. At the same time, the server evaluates the class distribution by measuring the loss of each client's model on a global auxiliary dataset. This method aims to improve model performance under budget constraints. One concern is that clients with rare but important distributions may be ignored if their contribution appears inefficient relative to price. The authors of [80] introduce the POWER-OF-CHOICE framework to mitigate label and quantity skew, which is intentionally biased toward hard clients to accelerate the training time. This selects clients after each training round based on their local loss. The server first identifies candidate clients based on dataset size in the pre-training stage and then sends the current global model to the selected clients to compute and return their local losses. Clients with the highest losses are selected for the next round while controlling for bias toward high local losses. This method converges up to three times faster and improves accuracy compared with uniform random selection. In contrast, the framework is biased toward clients with high losses, which may lead to poor stabilization training in extremely noisy datasets. The method by the authors of [81] addresses data heterogeneity in the IoT environment, where clients differ in data quality, quantity, and resource availability. This paper proposes a multi-criteria client selection algorithm, FedMCCS, which is applied during the pre-training stage in an IoT environment to evaluate available clients based on client data characteristics and system performance. The primary purpose of the method is

to dynamically update the decision for client selection based on the variability of factors, including client location, energy, CPU, local training time, and data size, thereby reducing the number of communication rounds and enhancing accuracy in each round. However, the method reveals a sampling bias because of the exclusion of slower or constrained devices, which correlate with specific user groups or environments. The authors of [82] focus on mitigating label and feature distribution skew by proposing FAVOR, an experience-driven control framework that intelligently and dynamically selects clients by designing a Deep Reinforcement Learning (DRL)-based agent that applies the DDQN algorithm to select the best subset of clients in the pre-training stage using model weights. Since model weights are high-dimensional, PCA is used to compress them into a manageable feature space. The reward function encourages two behaviors: increasing validation accuracy and reducing the number of communication rounds. The policy explicitly learns to trade off accuracy gains per round, thereby minimizing the number of communication rounds. However, understanding why certain clients are repeatedly chosen can be difficult to analyze regarding fairness. The authors of [83] enable mitigating label distribution skew in mobile environments, which causes performance variance in FL. This paper proposes a reputation-aware client selection method based on a Deep Reinforcement Learning (DRL) agent using the DDQN algorithm to identify the best subset of clients during the pre-training stage. The DRL agent calculates a reputation score based on clients' performance indicators compared with the average of all participants' clients, a temporary aggregated model, and the previous global model. This method avoids harmful updates, thereby improving the performance and accuracy of FL in mobile environments. The DRL agent excludes the low-reputation client even if its data are rare, and the model becomes less representative of the full population. UCB-CS [84] aims to mitigate label and quantity skew by adaptively selecting clients in each training round based on their average local loss and how often the client has been selected. These quantities are combined into a discounted Upper Confidence Bound (UCB) algorithm to balance exploitation by selecting clients with higher local loss to improve the model and exploration by selecting clients that have been selected less frequently. The discounted factor gradually forgets old observations to adapt when client behavior changes. This method aims to adaptively and dynamically select clients based on model performance and clients' engagement history to achieve faster convergence compared with baseline methods. However, higher loss can mean informative data, but it can also mean noisy or corrupted data. Ma et al. [85] propose FANS, a client selection framework that uses a combinatorial contextual neural bandit to address non-IID data. The method selects top-K clients using per-class validation loss, neural reward estimation, and the UBC strategy while applying FedAvg. However, prioritizing high-performing clients does not explicitly address participation fairness or influence fairness. The authors of [86] propose pFedLIA, a personalized FL framework that mitigates non-IID heterogeneity through influence-based client clustering before aggregation. The method estimates client influence using Lazy Influence Approximation, clusters clients with similar data distributions, and applies FedAvg to each cluster. However, clustering is designed for personalization rather than participation fairness, and intra-cluster aggregation does not explicitly address influence fairness. The authors of [87] introduce PSI-PFL, a personalized FL client selection framework that uses the Population Stability Index to measure label-distribution divergence under non-IID data. The method selects which PSI values are below a threshold, prioritizing clients closer to the global label distribution, while using FedAvg. However, it may exclude highly heterogeneous minority clients.

4.3.2. Performance-Driven Client Selection with Performance-Driven Aggregation

The authors of [88] propose a contribution-based device selection (CDS) algorithm for selecting clients based on estimating their contribution to improve the global model. The method operates in two stages: in the exploration phase, by randomly selecting clients in the pre-training stage and sending their model parameters and local accuracies to the server, and in the exploitation phase, where the server estimates each client's contribution to the global model using a modified Truncated Monte Carlo (TMC) method. Based on these contribution scores, the server selects the clients with the highest estimated impact and aggregates their updates according to estimated usefulness to form the new global model. This method relies on validating quality updates, which improves client selection and results in reducing communication rounds by four and improving accuracy by about 97% compared with baselines. While this method directly attacks a real inefficiency in FL, the contribution estimation is expensive, which shifts the cost from communication to computation. The authors of [89] propose the RRAFL method, which indirectly addresses non-IID types by calculating the data quality based on the reputation score. Using a reverse auction strategy, the server selects participant clients in the pre-training stage to achieve a balance between low cost and high reputation. After local training, the selected clients send the updated model with the size of the data to the server to validate the quality of the updated model for each client. In the pre-aggregation stage, each client calculates a weight based on the quality of the updated model and the size of the data to achieve balanced weighting. RRAFL consistently selects higher-quality participants and achieves higher accuracy. This method focuses on participant reliability rather than directly addressing non-IID drift, and it depends on reputation accumulation while newcomers struggle to enter. The authors of [90] propose a method that effectively addresses system heterogeneity and improves robustness in unreliable environments, but it does not address non-IID challenges. This paper proposes SAFA, a semi-asynchronous FL protocol, and a hybrid optimization approach that jointly improves client selection in the post-training stage via compensatory First Come First Merge (CFCFM) and discriminative aggregation via lag tolerance. SAFA demonstrates that the proposed protocol is effective regarding shortening federated round duration, reducing local resource wastage, and improving the accuracy of the global model at an acceptable communication cost. The authors of [91] propose Feddle, an asynchronous FL framework that mitigates non-IID heterogeneity by optimizing aggregation coefficients using server-side data. The method maintains a model atlas of client update directions and learns merging coefficients to improve convergence under in-domain or out-of-domain server data. However, update filtering is based on anchor importance rather than participation firmness, and aggregation is optimized for loss reduction rather than balanced client influence. The authors of [92] propose RAFL-Fed, which dynamically selects clients based on computation budget, capacity, and network proximity. The method ranks clients using importance weights derived from data quality, latency, and resource efficiency, and then aggregates top-K client updates using weighted averaging. However, it may exclude resource-constrained clients.

4.3.3. Random Client Selection with FedAvg Aggregation

FedAvg [1] aims to reduce the number of communication rounds by allowing multiple local epochs in randomly selected clients and then sending the updates to the server for aggregation. However, in non-IID settings, client drift emerges and leads to slow convergence and introduces bias toward larger clients.

4.3.4. Random Client Selection with Performance-Driven Aggregation

The authors of [9] propose Per-FedAvg, inspired by MAML, to address label distribution skew by creating a personalized model that establishes a good baseline that each client can quickly adapt to its own data. Each client trains a local model with regularized updates that uses the Hessian information of the local loss. This helps client updates stay aligned with the global model rather than drifting toward their own skewed labels. After the local updates, the server aggregates the updated models to generate an adaptive personalized model for each client in the aggregation stage. This method improves accuracy after only a few local fine-tuning steps compared with FedAvg. A key drawback is that second-order optimization is computationally heavier and harder to deploy on resource-constrained devices. The authors of [93] focus on mitigating label distribution and quantity skew during aggregation by accounting for accuracy and data size for each client. They propose FA-FedAvg, an improved version of FedAvg across two stages. First, in local training, the client computes baseline accuracy and uses early stopping to avoid unnecessary updates. Second, in the aggregation stage, the server uses weighting by F1-score, which prioritizes models that learn rare faults. FA-FedAvg achieves higher accuracy and reduces local training epochs. However, this approach only works when the client can reliably evaluate its model locally, which requires shared validation data. The approach by the authors of [94] focuses on mitigating feature distribution skew by proposing a decentralized learning approach called learning-to-collaborate (L2C) to decide who should collaborate with whom instead of forcing all clients to share one global model. Each client in the post-training stage optimizes the mixing weights of its model updates with its neighbors to minimize local loss for selected clients. The extended version, meta-L2C, uses an attention mechanism to assign mixing weights for new clients. The method achieves low communication and computation costs. At the same time, neighbor-based mixing gives more influence to more similar nodes, while different nodes contribute less, leading to unequal impact on the learned model. The authors of [95] focus on mitigating label distribution skew by proposing the FedAdp method, which uses adaptive weighting calculated in each round based on the client's contribution score using the angle between the local and global gradients to reduce the global loss. This score is then passed through a non-linear mapping function to produce the final weights for aggregation. Clients whose gradients align better with the global direction receive higher weights, significantly reducing the number of communication rounds to reach the target accuracy compared with FedAvg. However, it may down-weight minority distributions that are rare but important and amplify clients aligned with the global gradient direction. HPFL-CN [96] introduces a hierarchical personalized FL framework designed to mitigate differences in heterogeneity across regions and excessive communication between edge and cloud. The method applies privacy-preserving feature clustering (PFC) to map each client's data into a low-dimensional complex network space, allowing clients with similar data patterns to be clustered without revealing raw data. Based on these clusters, HPFL-CN used a client-mediator cloud architecture for training. The client performs several rounds of local updates; mediators aggregate the model within each cluster using effective hierarchical scheduling and then send the cluster models to the cloud for periodic global aggregation. This method achieves higher prediction accuracy and reduces communication rounds. However, hierarchical data size weighting prioritizes efficiency over balanced participation. The authors of [97] propose FL + HC to mitigate label and feature skew by hierarchical clustering clients based on the similarity of their local updates in the pre-aggregation stage, using a distance metric, such as the Manhattan distance, to group clients whose updates behave similarly. For each cluster, the server produces a personalized model that matches its

data patterns more closely than a single shared model, which leads to speeding up convergence in fewer rounds. However, fairness depends on timing; if clustering happens early, clusters are made using noisy information; otherwise, strong clients dominate. The authors of [98] propose FedSim to mitigate label, feature, and quantity skew by clustering similar clients based on model gradients using the K-means algorithm in the post-training stage, grouping clients with high similarity. Inside each cluster, the server applies balanced weighting based on each client's data size to mitigate quantity skew. The server then aggregates all cluster models into a new global model. A key drawback is that weighing based on data size within each cluster amplifies dominant client groups. The authors of [99] focus on mitigating label distribution and quantity skew that causes negative transfer in a cross-silo environment by proposing the FEDCOLLAB framework. FEDCOLLAB is a cluster FL framework that clusters clients based on their distribution distances and data quantity by using a lightweight client discriminator to estimate the distribution and data size of each client in the pre-training stage. Then, the framework clusters clients into coalitions and aggregates clients within each coalition using any existing FL algorithm. The results demonstrate that FEDCOLLAB improves accuracy and consistently outperforms other clustered FL algorithms. However, data-size-based coalition aggregation favors larger coalitions. The authors of [100] propose Fed2A, an asynchronous method designed to mitigate label skew, quantity skew, and temporal shift. Fed2A uses periodic layer uploading (PLU) to reduce communication costs by sending only the selected layers in the post-training stage. It also applies time variety weighting (TVW) to provide a higher weight based on the recency of the model and representational consistency enhancement (RCE) to measure the consistency of each layer with the global model in the pre-aggregation stage. While Fed2A improves accuracy and training time, the temporal freshness weighting favors recently active clients. The authors of [101] introduce Federated Region Learning (FRL) to mitigate feature skew by clustering the clients based on the region and training a personalized model for each cluster in the pre-training stage. The method builds on the idea that clients in the same geographic region often share similar data patterns. FRL provides two aggregation strategies, either synchronous or asynchronous, and both are designed to achieve high accuracy and fast convergence. The method overlooks data-size-based aggregation and prioritizes regions with more data. Qiu et al. [102] propose OLG-FL to mitigate non-IID data heterogeneity using contrastive learning and contribution-weighted aggregation. The method aligns local representations with global class centers and assigns aggregation weights based on cosine similarity between local and global update directions. However, it lacks client selection and may reduce the influence of minority clients whose updates diverge from the dominant global direction. The authors of [103] introduce a hierarchical FL framework for energy prediction using clustering-based fog organization and simulated annealing aggregation. Clients are grouped by similar consumption profiles, while aggregation favors better-performing clients. Bai et al. [104] propose FSCL, a federated supervised contrastive learning model that aligns local feature representations with the global model. The method combines cross-entropy loss with supervised contrastive loss during local training, while all clients participate in every round. However, the aggregation follows the FedAvg algorithm. The authors of [105] propose QHetFed, a hierarchical FL method that mitigates non-IID and quantization noise through two-level aggregation. The method averages quantized gradients within clusters and aggregates cluster models at the cloud using cluster-size weights. However, client selection is not addressed, and inter-cluster aggregation optimizes convergence.

5. Discussion and Open Challenges

The methods analyzed in this survey were primarily proposed to mitigate non-IID data effects and improve convergence and accuracy through client selection strategies and aggregation techniques, rather than to define or measure fairness. These techniques can implicitly introduce imbalance in client participation and in client influence on aggregation, leading to biased model outcomes.

5.1. Analysis of Method Categories

The classification used in this study is based on fairness awareness in method design rather than on performance objectives alone. Therefore, a method is not considered fairness-aware simply because it improves convergence, accuracy, communication efficiency, or robustness under non-IID data. Instead, methods are classified according to whether their client selection mechanism addresses participation fairness and whether their aggregation mechanism addresses influence fairness.

This distinction is important because many non-IID mitigation methods have fairness implications but do not explicitly define, measure, or optimize fairness. As summarized in Tables 3–5, the reviewed methods differ across three dimensions: theoretical orientation, methodological design, and practical trade-offs. Comparing the categories across these dimensions reveals systemic differences that are often overlooked when methods are evaluated by primarily using accuracy and convergence metrics.

Table 3. Fairness-aware methods: theoretical, methodological, and practical comparison.

Study	Comm.	Fairness Notion	Stage	Mechanism Type	Non-IID	Domain	Benchmarks	Results	Key Limitation
FedFast [40]	Sync.	Part. + Infl.	S + A	K-means clust. + delegate agg. (ActvSAMP/ActvAGG)	LS, FS	IR	FedAvg, FedAvg + ActvSAMP, FedAvg + ActvAGG	Accuracy +15–20%; 30–50% fewer rounds vs. FedAvg	Clustering quality depends on the similarity metric; two-stage overhead
Reputation-based [41]	Sync.	Part. + Contrib.	S + A	Reputation score + normal distribution client weighting	LS, QS	IR	FedAvg, Reputation-based selection	Accuracy +17.18%; convergence ~40 rounds	Requires a predefined reputation threshold; sensitive to calibration
CATA-Fed [42]	S-async.	Part. + Infl.	S + A	ALU straggler mitigation + CFS cluster-fair selection + FedAvg weighting	QS, SH	IR	FedAvg, FedProx, TiFL	2.0× faster than FedAvg; accuracy +0.65%	Multi-component design increases implementation complexity
G-FML [43]	Sync.	Part. + Group	S + A	Autoencoder clust. + MAML group meta-learning	LS (strong)	IR	FedAvg, FedProx, FedAvgUpdate, IFCA, Per-FedAvg	Avg. accuracy +12.50% vs. Per-FedAvg	Dynamic regrouping adds computation and communication overhead
CFMTL [44]	Sync.	Part. + Infl.	S + A	EMD clust. + random intra-cluster selection + proximal reg.	LS, QS	IR	FedAvg: 96.11%, CFL: 94.95%, CFMTL: 96.30%	CFMTL 96.30% vs. FedAvg 96.11%	Cluster isolation risks without inter-cluster knowledge transfer

Abbreviations: Comm.: sync.—synchronous, S-async.—semi-asynchronous, and async.—asynchronous. Non-IID: LS—label skew, QS—quantity skew, FS—feature skew, SH—system heterogeneity, and TS—temporal skew. Stage: S—selection and A—aggregation. Domain: IR—image recognition, HC—healthcare, SC—smart cities and transportation, Ind.—industrial systems, and Cyb.—cybersecurity.

Table 4. Semi-fairness-aware methods: theoretical, methodological, and practical comparison.

Study	Comm.	Fairness Notion	Stage	Mechanism	Non-IID	Domain	Benchmarks	Results	Key Limitation
Clustered Sampling [45] ^(a)	Sync.	Part.	S	Clust. by data size/model similarity	LS, QS	IR	FedAvg	~70% accuracy	Clustering scales poorly to large client populations
HACCS [6]	Sync.	Part.	S	Latency + loss-based cluster selection	LS, SH	IR	Random, TiFL, Oort	18–38% reduction in convergence time	Favors low-latency clients; may exclude constrained devices
Irrelevance Sampling [46]	Sync.	Part.	S	Irrelevance score (data size + class imbalance + non-IID degree)	LS, QS	IR, HC, SC	AFL, random, FedAvg, FSVRG, COOP	91% acc. vs. AFL 87%, random 88%	Score recomputed each round; communication overhead
SCS [47]	Sync.	Part.	S	Stochastic integer programming + reputation clust.	General	IR	FedAvg, FedAvg50, RRAFL	46.42% and 62.81% cost savings vs. NRACS	Assumes high reputation = reliability; ignores data diversity
FEDGS [48]	Sync.	Part.	S	GPB-CS geographic cluster. + super-node formation	LS	IR	FedAvg, FedMMD, FedFusion, FedProx, IDA	Accuracy +3.5%; 59% fewer rounds	Assumes a static location; fails in dynamic mobile environments
RBCS-F [31]	Sync.	Part. (equal opp.)	S	C2MAB + Lyapunov optim.	SH	IR	Random, FedCS	70% accuracy after 500 rounds	Fairness = equal opportunity, not equal benefit; may select weak models
CMAB Selection [49]	Sync.	Part. + Class bal.	S	Contextual MAB + gradient-based class estimation	LS	IR	Greedy, random, IID baseline	2× faster than greedy; +8% vs. random	Requires an auxiliary global dataset; impractical in privacy settings
E-Tree [51] ^(b)	Sync.	Part.	S + A	Hierarchical tree + data similarity + network distance	LS	HC	FL, Gossip, Individual, Grouped Agg.	Accuracy +2.4%; comm. cost −42%	Unsuitable for highly dynamic environments
GS/GEMD [52]	Sync.	Part. + Dist. bal.	S	Group EMD clust.	LS	IR	Random, Prob. EMD, FedProx, FedAvg	Accuracy +30%	Requires a trusted third party for distribution information
Eiffel [53]	Async.	Part. + Efficiency	S + A	Scheduling by computation cost + age-of-update weighting	LS, QS	SC	Canonical FL, centralized GD, q-FFL, random	Comm. improvement up to 6.45×	Indirect privacy leakage through scheduling information
SCAFFOLD [57] ^(c)	Sync.	Infl. (variance red.)	A	Control variates correct client drift	LS	IR	FedAvg, FedProx, SGD	Accuracy 0.801; 2.0× faster than FedAvg	Higher comm. cost per round (2× uploads)
FedDyn [8]	Sync.	Infl. (local–global align.)	A	Dynamic reg. aligning local to global objectives	LS, QS	IR, Synthetic	FedAvg, FedProx, IDA, SiloBN, FedBN	98.2% accuracy; up to 4.8× comm. savings	Server-side variable maintenance; memory overhead

Table 4. Cont.

Study	Comm.	Fairness Notion	Stage	Mechanism	Non-IID	Domain	Benchmarks	Results	Key Limitation
FedDist [56]	Sync.	Infl. (model capacity)	A	New neurons added for dissimilar client knowledge	LS, QS	Synthetic, HC	FedAvg, IDA, FedProx	Global F1-score: 84.52%	Continuous model expansion increases memory and comm. cost
FedAAR [58]	Sync.	Infl. (drift + conflict)	A	Prototype-guided local update + cosine-dist. gradient refinement	FS	Ind.	FedAvg, FedProx, MFA, MIFL	Accuracy +3.47%	Validated at a small scale only; scalability unconfirmed
FedBGVS [59]	Sync.	Infl. (balanced wt.)	A	Balanced global validation score weighting	LS, QS	HC, IR	FedAvg, FedProx, FedGN, FedMA, FeSEM, IFCA	MNIST: 99.56%; ISIC-2019: 72.49%	The server must hold a balanced validation dataset
FedFusion [60]	Sync.	Infl. (data-agnostic)	A	VAE-based distribution inference + adaptive weighting	LS, QS	Synthetic, IR	FedAvg, FedProx, FedGN, FedMA, FeSEM, IFCA	Up to 12.59% higher accuracy	Training VAE each round adds significant server computation
FedEWA [63]	S-Async.	Infl. (param-wise)	A	Elastic parameter-wise averaging by importance coeff.	LS, QS	IR	FedAvg, FedProx, MIFL	Accuracy 92%; 2.5× speedup	Clients must compute the importance coeff. per parameter; expensive
FedGKT [64]	Sync.	Infl. (KD)	A	KD from small client CNNs to large server CNNs	LS, QS	Synthetic, IR	Full FL $\tau = 1$, Clustered FL $\tau = 20$	9–17× less client compute	Feature map sharing may leak more than parameter averaging
FLT [61]	Sync.	Infl. (topology)	A	UMAP adjacency matrix + K-means clust. signatures	LS, QS	IR	Greedy knapsack, bid price first, optimal	FLT +3% vs. FedSEM (MNIST); +9% (CIFAR-10)	Depends on the pre-trained encoder distributed to all clients
FlexCFL [62]	Sync.	Infl. (temporal + clust.)	A	EDC clust. + Wasserstein dist. temporal shift control	LS, TS	IR	TA with FedAvg	FlexCFL +8.4% vs. FeSEM	Multiple tunable components increase system complexity
IB [68]	Sync.	Infl. (mutual info.)	A	Info. bottleneck loss + mutual info. pruning	LS, QS	IR	FedAvg, centralized, split learning	97.96% acc.; 2.50–23.77× comm. speedup vs. FedAvg	Computing mutual information adds computational overhead
QuPeD [65]	Sync.	Infl. (personalized quant.)	A	Personalization + model quantization + KD	LS, QS, SH	Synthetic, HC, IR	NonFed, FedSGD, FedAvg, FedCluster, FedProx	FEMNIST: 97.31%; CIFAR-10: 75.06%	Comm. overhead and privacy exposure through model outputs
TT-HF [67]	S-Async.	Infl. (D2D + server)	A	D2D intra-cluster rounds + server global agg.	LS, SH	IR	FedAvg, FedProx, IFCA, FeSEM, FedGroup	≥75% lower total cost vs. baseline	Depends on appropriate clust. and reliable D2D connectivity

Table 4. Cont.

Study	Comm.	Fairness Notion	Stage	Mechanism	Non-IID	Domain	Benchmarks	Results	Key Limitation
FeSEM [66]	Sync.	Infl. (EM clust.)	A	Stochastic EM algorithm assigning clients to clusters	LS, FS	IR	FedAvg, FedProx, IFCA, FedGroup	Optimal perf. within 100 iterations	Relies heavily on clustering quality
Active selection [71] ^(d)	Sync.	Contrib. (inform.)	S + A	Informativeness estimation + credibility weighting	LS, QS	IR	—	—	May bias toward specific high-value clients
Blockchain-based [72]	Sync.	Contrib. + Infl.	S + A	Blockchain + homomorphic encryption + credibility score	LS, QS	IR	—	Improved accuracy and convergence vs. baselines	Credibility score biased toward larger data clients
FAIR [73]	Sync.	Contrib. + Infl.	S + A	Reverse auction quality selection + cosine similarity weighting	LS, QS	IR	Cloud FL, cost-only CPLEX, data-only greedy	76% acc. (ResNet); baselines 45%	Quality filtering excludes statistically diverse clients
SHARE [74]	Sync.	Infl. (dist. shaping)	S + A	GoA node selection + LoS optim. to balance KL divergence	LS	HC	Cloud FL, cost-only CPLEX, data-only greedy	>90% comm. cost reduction vs. Cloud FL	Centralized optim. may be heavy at large scale
MH-FL [75]	Sync.	Infl. (hierarchical)	S + A	Multi-layer clust. D2D + fog node pre-agg.	LS, FS	IR	EUT	50% energy saving; 80% parameter reduction	Complex hierarchy; depends on fog node reliability

Abbreviations: Comm.: sync.—synchronous, S-async.—semi-asynchronous, and async.—asynchronous. Non-IID: LS—label skew, QS—quantity skew, FS—feature skew, SH—system heterogeneity, and TS—temporal skew. Stage: S—selection and A—aggregation. Domain: IR—image recognition, HC—healthcare, SC—smart cities and transportation, Ind.—industrial systems, and Cyb.—cybersecurity. Sub-groups: (a) fairness-aware selection + FedAvg aggregation, (b) fairness-aware selection + performance-driven aggregation, (c) random selection + fairness-aware aggregation, and (d) performance-driven selection + fairness-corrective aggregation.

Table 5. Comparison of the experiments and the measured metrics of reviewed fairness-unaware methods.

Study	Comm.	Fairness Notion	Stage	Mechanism	Non-IID	Domain	Benchmarks	Results	Key Limitation
TiFL [76] ^(a)	Sync.	None (latency)	S	Tier-based latency clust. + adaptive tier selection	SH	IR	Vanilla FL, random	Accuracy +6%; 3× training speedup	Tiering by latency reduces data diversity per round
FedCor [77]	Sync.	None (correlation)	S	Gaussian process loss correlation modeling	LS, QS	IR	Random, AFL, PowD	34–99% fewer rounds (FMNIST); 26–51% (CIFAR-10)	May avoid rare clients that appear uncorrelated or noisy
CSFedAvg [78]	Sync.	None (divergence)	S	Weight divergence from the auxiliary IID model	LS, QS	IR	FedAvg, FedProx	Accuracy +5.3%	Excludes clients with unusual data distributions
AUCTION [79]	Sync.	None (RL budget)	S	RL encoder–decoder agent + mislabel + class dist. quality	LS	IR	Greedy, PowD, Random, price first	Test accuracy 77.1% after 30 rounds	Ignores rare distributions if cost-inefficient

Table 5. Cont.

Study	Comm.	Fairness Notion	Stage	Mechanism	Non-IID	Domain	Benchmarks	Results	Key Limitation
POWER-OF-CHOICE [80]	Sync.	None (loss-driven)	S	High-loss biased selection + dataset size pre-filter	LS, QS	IR	AFL, FedAvg	3× faster convergence; +10% vs. random	Bias toward high-loss clients may amplify noisy data
FedMCCS [81]	Sync.	None (multi-criteria)	S	Multi-criteria IoT selection (location, energy, CPU, data size)	LS, QS	Cyb.	Vanilla FL, FedCS	80% accuracy in 108 rounds	Excludes slower/constrained devices correlated with specific groups
FAVOR [82]	Sync.	None (RL-driven)	S	Deep RL (DDQN) + PCA-compressed model weights	LS, FS	IR	FedAvg, K-Center	Comm. rounds −49% (MNIST), −42% (CIFAR-10)	Selection rationale opaque; fairness implications unanalyzed
Reputation-DRL [83]	Sync.	None (reputation)	S	DRL DDQN + reputation score from performance indicators	LS	IR	FedAvg	Accuracy +9.30%; training time −27.7%	Excludes low-reputation clients even if data are rare
UCB-CS [84]	Sync.	None (UCB)	S	Discounted UCB balancing loss and selection frequency	LS, QS	IR	Random, PowD, Rewarded PowD	~85% accuracy ($\alpha = 0.3$); ~0.5 global loss	Higher loss may indicate noisy data rather than informative data
CDS [88] ^(b)	Sync.	None (contrib.)	S + A	TMC contribution estimation + usefulness-based agg.	LS, QS	IR	Random, greedy, FedFomo	~97% accuracy; 4× fewer comm. rounds	Contribution estimation shifts cost from comm. to computation
RRAFL [89]	Sync.	None (reputation)	S + A	Reverse auction + reputation-weighted agg.	General	IR	Per-FedAvg, Vanilla FL	Accuracy: 90.16%; loss: 0.3541	Reputation accumulation disadvantages newcomer clients
SAFA [90]	Sync.	None (async. eff.)	S + A	CFCFM compensatory selection + lag-tolerant agg.	SH	IR, Cyb.	FedAvg, FedCS	Global accuracy 96–98%; 27× speedup	Does not address non-IID data distribution drift
FedAvg [1] ^(c)	Sync.	None	Neither	Random selection + data-size-weighted averaging	All types	IR	—	Standard baseline	Client drift; dominant large-data clients
Per-FedAvg [9] ^(d)	S-Async.	None (personalization)	A	MAML-inspired + Hessian-reg. personalized agg.	LS	IR	FedAvg, Per-FedAvg (FO)	Per-FedAvg (HF) accuracy: 79.85%	Second-order optim. heavy on resource-constrained devices
FA-FedAvg [93]	Sync.	None (F1-scoreweighted)	A	F1-score weighted agg. + early stopping	LS, QS	Ind.	FedAvg, FedProx	Accuracy +8.6%; training iterations −52.5%	Requires shared validation data for reliable local evaluation
L2C/Meta-L2C [94]	Sync.	None (neighbor mix)	A	Decentralized neighbor mixing + attention mechanism	FS	IR	FedAvg, Ditto, FOMO, DPSGD, CGA, Local SGD	L2C: 90.14%; Meta-L2C: 92.10%	Similar-neighbor mixing gives unequal influence to different nodes

Table 5. Cont.

Study	Comm.	Fairness Notion	Stage	Mechanism	Non-IID	Domain	Benchmarks	Results	Key Limitation
FedAdp [95]	Sync.	None (gradient align.)	A	Gradient angle-based adaptive weighting	LS	IR	FedAvg	Comm. rounds −54.1%	May down-weight rare but important minority distributions
HPFL-CN [96]	Sync.	None (hier. personal.)	A	Privacy-preserving feat. clust. + client-mediator cloud architecture hierarchy	LS, FS	IR	FedAvg, fine-tuning, mixture	Accuracy +9.67% (LSTM), +6.78% (GRU); rounds −70%	Hierarchical data-size weighting prioritizes efficiency over participation
FL + HC [97]	Sync.	None (hier. clust.)	A	Manhattan distance clust. of local updates	LS, FS	IR	FedAvg	Comm. rounds >5×; accuracy 99.1%	Early clustering uses noisy info.; strong clients may dominate
FedSim [98]	Sync.	None (similarity clust.)	A	K-means gradient similarity clust. + data-size weighting	LS, FS, QS	IR	FedAvg, FedProx	Accuracy +11.21%	Data-size weighting within clusters amplifies dominant groups
FEDCOLLAB [99]	Sync.	None (coalition)	A	Distribution-dist. + data-size coalition clust.	LS, QS	SC	IFCA, FedCluster, FeSEM, K-means	Accuracy 92.45%	Data-size coalition agg. favors larger coalitions
Fed2A [100]	Sync.	None (async. freshness)	A	Periodic layer upload + time-variety weighting + repr. consistency	LS, QS, TS	IR	FedAvg, FedProx, FedAsync	Comm. cost −77%; accuracy +19%	Temporal freshness weighting favors recently active clients
FRL [101]	Async.	None (region-based)	A	Geographic region clust. + sync./async. personalized agg.	FS	IR	Sync CNN-FRL, async. CNN-FRL	Accuracy 78.61%; rounds −20.8%	Prioritizes regions with more data; overlooks data-size balance
OLG-FL [102]	Sync.	None	A	Contrastive learning + contribution-weighted aggregation	LS, QS	IR	CIFAR-10, CIFAR-100, Tiny-ImageNet	Outperforms FedAvg, MOON, FedProx	Contribution weighting may disadvantage minority distribution clients
FANS [85]	Sync.	None	S	UCB-based contextual neural bandit	LS	IR	MNIST, CIFAR-10	Improves SRS and accuracy over baselines	Prioritizes high-performing clients; no aggregation fairness
pFedLIA [86]	Sync.	None	A	Influence-based client clustering	LS	IR	CIFAR-100, Nordic NWP	17% improvement on CIFAR-100; matches oracle clustering	Clustering is performed once; no explicit fairness mechanism
Feddle [91]	Async.	None	A	Server-guided coefficient optimization	LS, FS	IR, HC	CIFAR-10, DermaMNIST	Faster convergence than FedAvg and FedBuff	Requires server-side data; no participation fairness

Table 5. Cont.

Study	Comm.	Fairness Notion	Stage	Mechanism	Non-IID	Domain	Benchmarks	Results	Key Limitation
PSI-PFL [87]	Sync.	None	S	PSI-based homogeneous client selection	LS	IR	Tabular, image, and text datasets	Up to 10% accuracy improvement under non-IID data	Excludes highly heterogeneous clients; no aggregation fairness
RAFL-Fed [92]	Sync.	None	S + A	Resource-aware selection + importance-weighted aggregation	LS, SH	HC	MIMIC-IV	Accuracy 0.9571 under non-IID; sensitivity 0.9968	Excludes resource-constrained clients; domain-specific evaluation
FSCL [104]	Sync.	None	A	Supervised contrastive local training	FS	Ind.	Gearbox datasets	Improves feature alignment under heterogeneous data	All clients participate; no explicit fairness mechanism
Hierarchical FL [103]	Sync.	None	A	Simulated annealing aggregation + GA tuning	FS, TS	SC	Energy consumption datasets	Improves accuracy over FedAvg under non-IID data	Domain-specific; favors high-performing clients
QHETFed [105]	Sync.	None	A	Hierarchical quantized gradient/model aggregation	LS, QS	Ind.	IoT classification benchmarks	Lower optimality gap than Hier-Local-QSGD	All devices participate; cluster-size weighting may disadvantage smaller clusters

Abbreviations: Comm.: sync.—synchronous, S-async.—semi-asynchronous, and async.—asynchronous. Non-IID: LS—label skew, QS—quantity skew, FS—feature skew, SH—system heterogeneity, and TS—temporal skew. Stage: S—selection and A—aggregation. Domain: IR—image recognition, HC—healthcare, SC—smart cities and transportation, Ind.—industrial systems, and Cyb.—cybersecurity. Sub-groups: (a) fairness-unaware selection + FedAvg aggregation, (b) fairness-unaware selection + performance-driven aggregation, (c) random selection + fairness-unaware aggregation, and (d) performance-driven selection + fairness-unaware aggregation.

From a theoretical perspective, the three categories reflect different assumptions about fairness in FL. Fairness-aware methods [40–44] adopt a dual-stage view of fairness by explicitly considering participation fairness and influence fairness. These methods recognize that fair outcomes depend not only on which clients participate in training but also on how their updates influence the global model. In contrast, semi-fairness-aware methods consider fairness in only one stage of the FL pipeline. Methods in sub-group (a) [6,31,45–49] focus on participation fairness by improving client representation during selection, while methods in sub-group (c) [8,56–68] focus on influence fairness through aggregation mechanisms. Although these methods improve fairness in one stage, they cannot guarantee system-level fairness because the unaddressed stage may still introduce imbalance. Fairness-unaware methods [1,76–84,88–90,93–101] do not explicitly incorporate any fairness objective. Their primary goal is to improve performance, convergence, communication efficiency, or robustness. Consequently, improvement in average performance may coexist with increased disparities across clients.

From a methodological perspective, the categories differ in both the FL stage they address and the mechanisms they employ. Fairness-aware methods integrate balancing mechanisms across client selection and aggregation, often combining clustering-based selection with contribution-aware or regularized aggregation. This dual-stage design typically increases algorithmic complexity, as in CATA-Fed [42], which combines strag-

gler mitigation, cluster-aware selection, and weighted aggregation, while G-FML [43] incorporates representation learning and group-specific meta-learning. Such additional components increase communication and computational overhead, which may explain why fairness-aware methods remain relatively limited in the literature. Semi-fairness-aware methods reduce complexity by focusing on a single stage. Selection-focused methods in sub-groups (a) and (b) employ clustering, optimization, or resource-aware strategies to improve participation fairness but typically rely on standard FedAvg aggregation, which weighs updates according to local data size and may reintroduce imbalance. Aggregation-focused methods in sub-group (c) apply corrective mechanisms such as control variates [57], dynamic regularization [8], or validation-based weighting [59], but usually rely on random or unconstrained client selection. As a result, participation fairness remains uncontrolled. Fairness-unaware methods employ performance-oriented mechanisms, such as loss-based selection [80], reinforcement learning [82], or contribution estimation [88]. While effective for improving convergence and efficiency, these strategies may favor highly informative, high-resource, or frequently participating clients, potentially disadvantaging clients with limited resources or minority data distributions.

From a practical perspective, the three categories exhibit different deployment trade-offs. Fairness-aware methods address a broader range of heterogeneity types: they often require additional server-side resources, such as clustering infrastructure, validation datasets, or meta-learning modules. Their computational complexity also tends to increase with the number of clients and the degree of heterogeneity. Semi-fairness-aware methods offer a more practical compromise. Selection-focused approaches can significantly reduce convergence time and communication rounds. For example, the authors of [6] report a reduction of 18–38% in convergence time with HACCS, while FEDGS [48] achieved a 59% reduction in communication rounds. Nevertheless, these methods often depend on assumptions, such as static client location [48], auxiliary dataset [49], or trusted third parties [52], which may limit deployment in real-world environments. Aggregation-focused methods improve model performance under non-IID conditions without modifying the client selection process, but several of these methods require additional server-side resources. FedBGVS [59] requires a balanced validation dataset, FedFusion [60] trains a Variational Autoencoder at each round, and IB [68] estimates mutual information during training. Such requirements may be difficult to satisfy in privacy-sensitive or resource-constrained environments. Fairness-unaware methods often achieve higher convergence speed and communication efficiency than the other categories. However, these methods may reinforce disparities inherent in non-IID data by favoring low-latency, high-resource, or high-contribution clients. As shown in Table 5, many fairness-unaware methods acknowledge sampling bias, dominance by large-data clients, or limited selection diversity as important limitations, indicating that these concerns remain relevant even when fairness is not an explicit design objective.

Across all three categories, two recurring patterns emerge. First, a clear trade-off exists between fairness and efficiency. Methods that promote balanced participation and influence often introduce additional communication and computational costs. Whereas methods that prioritize convergence speed and efficiency may increase disparities between clients. Second, addressing only one stage of the FL pipeline is insufficient for achieving system-level fairness. Semi-fairness-aware methods improve fairness either in client selection or in model aggregation, but the benefits are often limited by the stage that remains unaddressed. Together, these findings suggest that achieving system-level fairness requires coordinated design across both stages and evaluation using cross-stage fairness metrics, which remain largely absent from current practice.

5.2. Limitations of Current Evaluation Practices

A major limitation in the current literature concerns how non-IID mitigation methods are evaluated. Most studies focus on average accuracy and convergence speed, which provide limited insight into disparities across clients. Because fairness metrics are often missing, it is difficult to determine whether a method improves fairness or only improves overall performance while leaving some clients disadvantaged.

The current evaluation practices have several limitations. These limitations can be grouped into five dimensions: dataset, non-IID generation methods, benchmark models and applications, fairness metrics, and performance metrics. Most studies rely on benchmark datasets such as MNIST, CIFAR-10, and FMNIST, as shown in Figure 5. Although these datasets facilitate comparison across studies, they do not fully reflect the complexity of real-world FL environments, where data are often noisy, diverse, imbalanced, and distributed across heterogeneous clients.

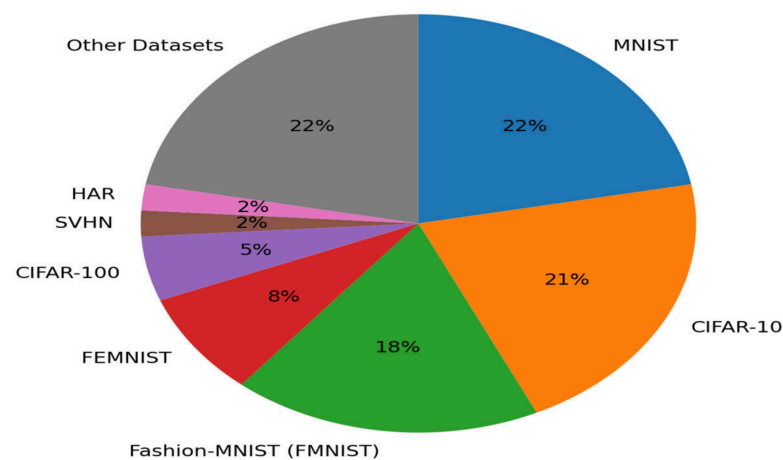


Figure 5. Distribution of datasets used in federated learning studies.

Similarly, non-IID data are commonly generated using label skew and quantity skew, as shown in Figure 6. While these settings are useful for controlled evaluation, they do not capture the full range of heterogeneity observed in practice. More realistic forms of non-IID data, such as feature skew and concept drift, remain comparatively underexplored. Evaluation is also heavily centered on CNN-based models Figure 7. Consequently, it remains unclear whether many non-IID mitigation methods generalize to other model architectures. Furthermore, most experiments focus on image classification tasks, while domain-specific applications, such as healthcare, industrial systems, smart cities, and cybersecurity, receive considerably less attention Figure 8. This limits the generalizability of existing findings to real-world deployment scenarios. Another important limitation concerns fairness evaluation. Although fairness is increasingly recognized as an important objective in FL, many studies continue to focus primarily on accuracy, convergence speed, and communication cost. Fairness-related measures, such as client-level performance disparities, participation balance, and aggregation influence, are often absent. As a result, it is difficult to determine whether a method improves fairness or simply improves average performance. Performance evaluation itself also remains limited. Many studies report global accuracy, convergence speed, and communication cost, but these metrics do not fully explain how the improvement was achieved or who benefited from it. A method may improve average accuracy while still producing poor outcomes for weak or underrepresented clients.

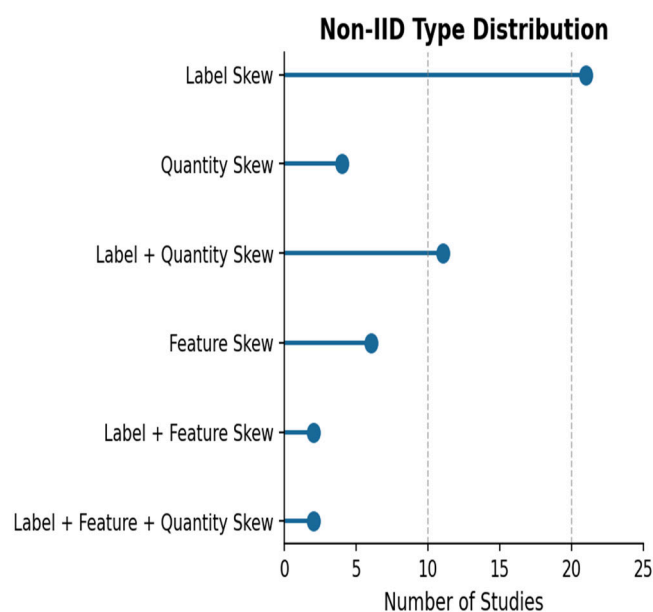


Figure 6. Distribution of non-IID data types.

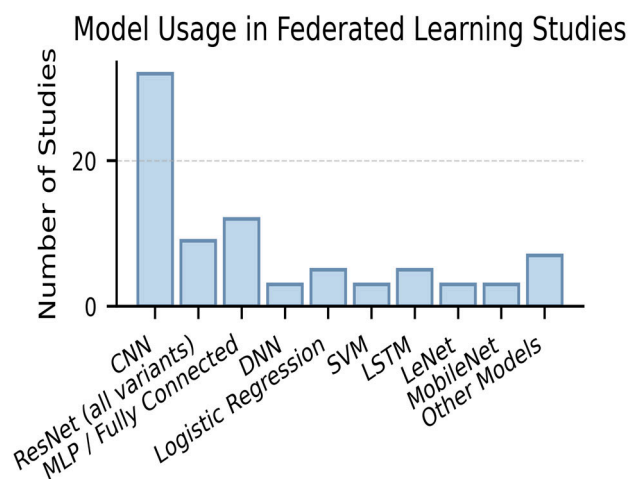


Figure 7. Usage of ML models in FL studies.

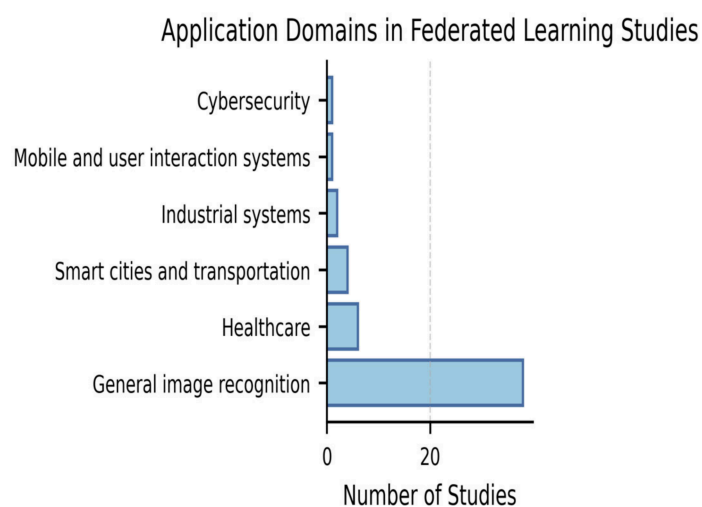


Figure 8. Distribution of application domains in FL.

Together, these limitations hinder the reproducibility, comparability, and generalizability of the existing results. They also highlight the need for standardized evaluation frameworks that jointly assess fairness, performance, and efficiency under diverse non-IID settings.

5.3. Future Research Directions

The limitations identified in the current literature suggest several important directions for future research.

- System-level fairness design:

Most existing methods address fairness either during client selection or model aggregation but rarely consider both stages together. Future non-IID mitigation methods should adopt a system-level perspective that jointly considers participation fairness and influence fairness. Integrating fairness-aware client selection with fairness-aware aggregation may help ensure that fair participation translates into fair influence and more equitable outcomes across clients.

- Standardized fairness metrics and benchmarks:

The lack of standardized fairness evaluation remains a major barrier to meaningful comparison across studies. Future research should develop benchmark frameworks that combine standardized datasets, non-IID generation protocols, fairness metrics, and performance metrics. These frameworks should evaluate not only global accuracy and convergence speed, but also client-level performance, participation balance, aggregation influence, and disparities across clients. This would improve reproducibility and enable more consistent evaluation of fairness–performance trade-offs.

- Adaptive fairness–efficiency optimization:

The analysis presented in this survey highlights a persistent trade-off between fairness and efficiency. Future methods should explicitly manage this trade-off rather than treating fairness as a secondary objective. Adaptive approaches that dynamically balance fairness, accuracy, communication cost, and computation cost according to the system conditions may provide more practical solutions for real-world deployment.

- Broader and more realistic evaluation:

Current evaluations are heavily concentrated on benchmark image datasets, CNN-based models, and simplified non-IID settings. Future studies should consider a wide range of datasets, model architectures, and non-IID types. Greater attention should also be given to real-world application domains. This would improve the practical relevance and generalizability of research findings.

- Security, ethical, and regulatory considerations:

FL systems are highly vulnerable to security threats, such as adversarial attacks, Byzantine attacks, backdoor attacks, and other types of attacks. As these systems are increasingly used in privacy-sensitive and socially impactful applications, future research should address fairness alongside security, privacy, transparency, accountability, explainability, and regulatory compliance. This requires clear fairness objectives, transparent selection and aggregation mechanisms, appropriate fairness metrics, and continuous monitoring of potential harms. These considerations are important for supporting the secure, responsible, and trustworthy deployment of FL systems.

6. Conclusions

This survey reviewed existing federated learning (FL) methods for mitigating non-IID data from a fairness perspective. While a wide range of approaches have been proposed to address the challenges of non-IID data, these methods are primarily designed to improve model performance and efficiency, with limited consideration of fairness in their design or evaluation. The findings of this survey motivate the need to systematically analyze the implicit effects of FL methods on client participation and influence. To address this gap, we propose a taxonomy that classifies existing methods into three categories—fairness-aware, semi-fairness-aware, and fairness-unaware—based on their design strategies for client selection and model aggregation. Using this taxonomy, we examined the advantages, trade-offs, and limitations of each category. A key finding is that mitigating non-IID data does not ensure fairness. Improvements in performance do not guarantee balanced participation or influence across clients, particularly under non-IID data. Furthermore, fairness is rarely evaluated explicitly in existing non-IID mitigation methods, as most studies rely on overall performance measures that ignore differences between clients. This limitation is reinforced by current evaluation practices, which are typically based on limited datasets, simplified non-IID settings, and narrow application domains. As a result, it remains difficult to quantitatively assess fairness or compare methods in a consistent and meaningful way. This represents a significant direction for future research. However, we believe that our stage-level taxonomy contributes to bridging performance-oriented mitigation strategies with fairness concerns, helping future methods better balance performance, efficiency, and fairness, while also laying the foundation for a future quantitative fairness scoring framework. Hence, these findings highlight the need for a shift toward system-level FL design, where fairness and non-IID mitigation are addressed jointly. Such approaches should explicitly consider participation, influence, and outcomes across clients while balancing system cost, scalability, robustness, and privacy. As FL is increasingly deployed in sensitive domains such as healthcare, finance, and IoT, ensuring fairness alongside performance is essential for building large-scale FL systems that maintain performance and fairness across diverse clients.

Author Contributions: Conceptualization, M.A. and H.M.; methodology, M.A., I.A., and H.M.; formal analysis, M.A. and I.A.; supervision, H.M.; writing—original draft preparation, M.A.; writing—review and editing, M.A., I.A., and H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the Artificial Intelligence and Statistics*; PMLR: New York, NY, USA, 2017; pp. 1273–1282.
2. Cheng, H.; Qu, Y.; Liu, W.; Gao, L.; Zhu, T. Decentralized Federated Learning for Private Smart Healthcare: A Survey. *Mathematics* **2025**, *13*, 1296. [CrossRef]
3. Kennedy, C.H.; Hilal, A.; Momeni, M. The Role of Federated Learning in Improving Financial Security: A Survey. In *Proceedings of the 2025 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), Ben Guerir, Morocco, 23–25 November 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–8. [CrossRef]
4. Dritsas, E.; Trigka, M. Federated Learning for IoT: A Survey of Techniques, Challenges, and Applications. *J. Sens. Actuator Netw.* **2025**, *14*, 9. [CrossRef]
5. Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li, W.; Gao, Y. A Survey on Federated Learning. *Knowl.-Based Syst.* **2021**, *216*, 106775. [CrossRef]

6. Wolfrath, J.; Sreekumar, N.; Kumar, D.; Wang, Y.; Chandra, A. HACCS: Heterogeneity-Aware Clustered Client Selection for Accelerated Federated Learning. In *Proceedings of the 2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*; IEEE: Lyon, France, 2022; pp. 985–995.
7. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated Optimization in Heterogeneous Networks. *Proc. Mach. Learn. Syst.* **2020**, *2*, 429–450.
8. Acar, D.A.E.; Zhao, Y.; Matas Navarro, R.; Mattina, M.; Whatmough, P.N.; Saligrama, V. Federated Learning Based on Dynamic Regularization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, Virtual Event, 3–7 May 2021.
9. Fallah, A.; Mokhtari, A.; Ozdaglar, A. Personalized Federated Learning with Theoretical Guarantees: A Model-Agnostic Meta-Learning Approach. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 3557–3568.
10. Qi, P.; Chiaro, D.; Guzzo, A.; Ianni, M.; Fortino, G.; Piccialli, F. Model Aggregation Techniques in Federated Learning: A Comprehensive Survey. *Future Gener. Comput. Syst.* **2024**, *150*, 272–293. [\[CrossRef\]](#)
11. Mayhoub, S.; Shami, M.T. A Review of Client Selection Methods in Federated Learning. *Arch. Comput. Methods Eng.* **2024**, *31*, 1129–1152. [\[CrossRef\]](#)
12. Ayeelyan, J.; Utomo, S.; Rouniyar, A.; Hsu, H.-C.; Hsiung, P.-A. Federated Learning Design and Functional Models: Survey. *Artif. Intell. Rev.* **2024**, *58*, 21. [\[CrossRef\]](#)
13. Almanifi, O.R.A.; Chow, C.-O.; Tham, M.-L.; Chuah, J.H.; Kanesan, J. Communication and Computation Efficiency in Federated Learning: A Survey. *Internet Things* **2023**, *22*, 100742. [\[CrossRef\]](#)
14. Abdulrahman, S.; Tout, H.; Ould-Slimane, H.; Mourad, A.; Talhi, C.; Guizani, M. A Survey on Federated Learning: The Journey From Centralized to Distributed On-Site Learning and Beyond. *IEEE Internet Things J.* **2021**, *8*, 5476–5497. [\[CrossRef\]](#)
15. Lu, Z.; Pan, H.; Dai, Y.; Si, X.; Zhang, Y. Federated Learning with Non-IID Data: A Survey. *IEEE Internet Things J.* **2024**, *11*, 19188–19209. [\[CrossRef\]](#)
16. Ma, X.; Zhu, J.; Lin, Z.; Chen, S.; Qin, Y. A State-of-the-Art Survey on Solving Non-IID Data in Federated Learning. *Future Gener. Comput. Syst.* **2022**, *135*, 244–258. [\[CrossRef\]](#)
17. Zhu, H.; Xu, J.; Liu, S.; Jin, Y. Federated Learning on Non-IID Data: A Survey. *Neurocomputing* **2021**, *465*, 371–390. [\[CrossRef\]](#)
18. Fu, L.; Zhang, H.; Gao, G.; Zhang, M.; Liu, X. Client Selection in Federated Learning: Principles, Challenges, and Opportunities. *IEEE Internet Things J.* **2023**, *10*, 21811–21819. [\[CrossRef\]](#)
19. Shanmugarasa, Y.; Paik, H.; Kanhere, S.S.; Zhu, L. A Systematic Review of Federated Learning from Clients’ Perspective: Challenges and Solutions. *Artif. Intell. Rev.* **2023**, *56*, 1773–1827. [\[CrossRef\]](#)
20. Alotaibi, B.; Khan, F.A.; Mahmood, S. Communication Efficiency and Non-Independent and Identically Distributed Data Challenge in Federated Learning: A Systematic Mapping Study. *Appl. Sci.* **2024**, *14*, 2720. [\[CrossRef\]](#)
21. Mukhtiar, N.; Mahmood, A.; Sheng, Q.Z. Fairness in Federated Learning: Trends, Challenges, and Opportunities. *Adv. Intell. Syst.* **2025**, *7*, 2400836. [\[CrossRef\]](#)
22. Taik, A.; Chehbouni, K.; Farnadi, G. Fairness in Federated Learning: Fairness for Whom? In *Proceedings of the AAAI ACM Conference on AI, Ethics, and Society*; AAAI Press: Washington, DC, USA, 2025; pp. 2456–2469. [\[CrossRef\]](#)
23. Jahani, K.; Moshiri, B.; Khalaj, B.H. A Survey on Data Distribution Challenges and Solutions in Vertical and Horizontal Federated Learning. *J. Artif. Intell. Appl. Innov.* **2024**, *1*, 55–71. [\[CrossRef\]](#)
24. Jimenez G., D.M.; Solans, D.; Heikkilä, M.; Vitaletti, A.; Kourtellis, N.; Anagnostopoulos, A.; Chatzigiannakis, I. Non-IID Data in Federated Learning: A Survey with Taxonomy, Metrics, Methods, Frameworks and Future Directions. *arXiv* **2024**, arXiv:2411.12377. [\[CrossRef\]](#)
25. Noor, N.; Mashhadi, A. Quantifying Individual Fairness in Continual Federated Learning. In *The Companion Proceedings of the ACM Web Conference 2025, Sydney, NSW, Australia, 28 April–2 May 2025*; ACM: New York, NY, USA, 2025; pp. 1735–1743. [\[CrossRef\]](#)
26. Zhang, M.; Zhao, H.; Ebron, S.; Xie, R.; Yang, K. Ensuring Fairness in Federated Learning Services: Innovative Approaches to Client Selection, Scheduling, and Rewards. In *Proceedings of the 2024 IEEE 44th International Conference on Distributed Computing Systems (ICDCS)*, Jersey City, NJ, USA, 23 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 762–773.
27. Shi, Y.; Liu, Z.; Shi, Z.; Yu, H. Fairness-Aware Client Selection for Federated Learning. In *Proceedings of the 2023 IEEE International Conference on Multimedia and Expo (ICME)*, Brisbane, Australia, 10–14 July 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 335–340. [\[CrossRef\]](#)
28. Badar, M.; Sikdar, S.; Nejdil, W.; Fisichella, M. FairTrade: Achieving Pareto-Optimal Trade-Offs between Balanced Accuracy and Fairness in Federated Learning. In *Proceedings of the AAAI conference on artificial intelligence, Vancouver, BC, Canada, 22–27 February 2024*; AAAI Press: Washington, DC, USA, 2024; Volume 38, pp. 10962–10970. [\[CrossRef\]](#)
29. Li, T.; Sanjabi, M.; Beirami, A.; Smith, V. Fair Resource Allocation in Federated Learning. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, Addis Ababa, Ethiopia, 26–30 April 2020; OpenReview.net: Amherst, MA, USA, 2020.

30. Li, C.; Li, C.; Zhao, Y.; Zhang, B.; Li, C. An Efficient Multi-Model Training Algorithm for Federated Learning. In *Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 7–11 December 2021*; IEEE: Piscataway, NJ, USA, 2021. [\[CrossRef\]](#)
31. 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 Trans. Parallel Distrib. Syst.* **2021**, *32*, 1552–1564. [\[CrossRef\]](#)
32. Hu, S.; Wu, Z.S.; Smith, V. Private Multi-Task Learning: Formulation and Applications to Federated Learning. *Trans. Mach. Learn. Res.* **2023**. Available online: <https://openreview.net/forum?id=onufdyHvqN> (accessed on 2 June 2026). [\[CrossRef\]](#)
33. Chaudhury, B.R.; Murhekar, A.; Yuan, Z.; Li, B.; Mehta, R.; Procaccia, A.D. Fair Federated Learning via the Proportional Veto Core. In *Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024*; PMLR: New York, NY, USA, 2024; Volume 235, pp. 42245–42257.
34. Fan, Z.; Fang, H.; Zhou, Z.; Pei, J.; Friedlander, M.P.; Liu, C.; Zhang, Y. Improving Fairness for Data Valuation in Horizontal Federated Learning. In *Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, 9–12 May 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 2440–2453. [\[CrossRef\]](#)
35. Chen, Y.; Huang, W.; Ye, M. Fair Federated Learning Under Domain Skew with Local Consistency and Domain Diversity. In *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16 June 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 12077–12086.
36. Shen, Y.; Xi, W.; Cai, Y.; Fan, Y.; Yang, H.; Zhao, J. Multi-Objective Federated Learning: Balancing Global Performance and Individual Fairness. *Future Gener. Comput. Syst.* **2025**, *162*, 107468. [\[CrossRef\]](#)
37. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Corporation for Digital Scholarship. *Zotero*; Version 6.0.26; Computer Software; Corporation for Digital Scholarship: Vienna, VA, USA, 2023.
39. Nickerson, R.C.; Varshney, U.; Muntermann, J. A Method for Taxonomy Development and Its Application in Information Systems. *Eur. J. Inf. Syst.* **2013**, *22*, 336–359. [\[CrossRef\]](#)
40. Muhammad, K.; Wang, Q.; O'Reilly-Morgan, D.; Tragos, E.; Smyth, B.; Hurley, N.; Geraci, J.; Lawlor, A. FedFast: Going Beyond Average for Faster Training of Federated Recommender Systems. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, CA, USA, 23 August 2020*; ACM: New York, NY, USA, 2020; pp. 1234–1242.
41. Wang, Y.; Kantarci, B. Reputation-Enabled Federated Learning Model Aggregation in Mobile Platforms. In *Proceedings of the ICC 2021—IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
42. Jeong, Y.; Kim, T. A Cluster-Driven Adaptive Training Approach for Federated Learning. *Sensors* **2022**, *22*, 7061. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Yang, L.; Huang, J.; Lin, W.; Cao, J. Personalized Federated Learning on Non-IID Data via Group-Based Meta-Learning. *ACM Trans. Knowl. Discov. Data* **2023**, *17*, 1–20. [\[CrossRef\]](#)
44. Xiao, Y.; Shu, J.; Jia, X.; Huang, H. Clustered Federated Multi-Task Learning with Non-IID Data. In *Proceedings of the 2021 IEEE 27th International Conference on Parallel and Distributed Systems (ICPADS), Beijing, China, 14–16 December 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 50–57.
45. Fraboni, Y.; Vidal, R.; Kameni, L.; Lorenzi, M. Clustered Sampling: Low-Variance and Improved Representativity for Clients Selection in Federated Learning. In *Proceedings of the 38th International Conference on Machine Learning, Virtual Event, 18–24 July 2021*; PMLR: New York, NY, USA, 2021; Volume 139, pp. 3407–3416.
46. Rai, S.; Kumari, A.; Prasad, D.K. Client Selection in Federated Learning under Imperfections in Environment. *AI* **2022**, *3*, 124–145. [\[CrossRef\]](#)
47. Tan, X.; Ng, W.C.; Lim, W.Y.B.; Xiong, Z.; Niyato, D.; Yu, H. Reputation-Aware Federated Learning Client Selection Based on Stochastic Integer Programming. *IEEE Trans. Big Data* **2024**, *10*, 953–964. [\[CrossRef\]](#)
48. Li, Z.; He, Y.; Yu, H.; Kang, J.; Li, X.; Xu, Z.; Niyato, D. Data Heterogeneity-Robust Federated Learning via Group Client Selection in Industrial IoT. *IEEE Internet Things J.* **2022**, *9*, 17844–17857. [\[CrossRef\]](#)
49. Yang, M.; Wang, X.; Zhu, H.; Wang, H.; Qian, H. Federated Learning with Class Imbalance Reduction. In *Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 23 August 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 2174–2178.
50. Wang, P.; Zhong, Z.; Wang, J. Efficient Federated Learning via Aggregation of Base Models. *PLoS ONE* **2025**, *20*, e0327883. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Yang, L.; Lu, Y.; Cao, J.; Huang, J.; Zhang, M. E-Tree Learning: A Novel Decentralized Model Learning Framework for Edge AI. *IEEE Internet Things J.* **2021**, *8*, 11290–11304. [\[CrossRef\]](#)

52. Ma, J.; Sun, X.; Xia, W.; Wang, X.; Chen, X.; Zhu, H. Client Selection Based on Label Quantity Information for Federated Learning. In *Proceedings of the 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Helsinki, Finland, 13 September 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
53. Sultana, A.; Haque, M.; Chen, L.; Xu, F.; Yuan, X. *Eiffel*: Efficient and Fair Scheduling in Adaptive Federated Learning. *IEEE Trans. Parallel Distrib. Syst.* **2022**, *33*, 4282–4294. [[CrossRef](#)]
54. Li, S.; Shan, F.; Mao, S.; Lu, Y.; Miao, F.; Chen, Z. SACW: Semi-Asynchronous Federated Learning with Client Selection and Adaptive Weighting. *Computers* **2025**, *14*, 464. [[CrossRef](#)]
55. Sun, H.; Yao, J.; Li, X.; Liu, Y.; Gu, H. Robust Two Stages Federated Learning for Sensor Based Human Activity Recognition with Label Noise. *Sci. Rep.* **2025**, *15*, 17227. [[CrossRef](#)] [[PubMed](#)]
56. Ek, S.; Portet, F.; Lalanda, P.; Vega, G. A Federated Learning Aggregation Algorithm for Pervasive Computing: Evaluation and Comparison. In *Proceedings of the 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), Kassel, Germany, 22–26 March 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–10.
57. Karimireddy, S.P.; Kale, S.; Mohri, M.; Reddi, S.J.; Stich, S.U.; Suresh, A.T. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. *Int. Conf. Mach. Learn.* **2020**, *119*, 5132–5143.
58. Mao, A.; Huang, E.; Gan, H.; Liu, K. FedAAR: A Novel Federated Learning Framework for Animal Activity Recognition with Wearable Sensors. *Animals* **2022**, *12*, 2142. [[CrossRef](#)] [[PubMed](#)]
59. Mou, Y.; Geng, J.; Welten, S.; Rong, C.; Decker, S.; Beyan, O. Optimized Federated Learning on Class-Biased Distributed Data Sources. In *Machine Learning and Principles and Practice of Knowledge Discovery in Databases*; Kamp, M., Koprinska, I., Bibal, A., Bouadi, T., Frénay, B., Galárraga, L., Oramas, J., Adilova, L., Krishnamurthy, Y., Kang, B., et al., Eds.; Communications in Computer and Information Science; Springer International Publishing: Cham, Switzerland, 2021; Volume 1524, pp. 146–158. ISBN 978-3-030-93735-5.
60. Duan, J.-H.; Li, W.; Zou, D.; Li, R.; Lu, S. Federated Learning with Data-Agnostic Distribution Fusion. In *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 8074–8083.
61. Jamali-Rad, H.; Abdizadeh, M.; Singh, A. Federated Learning with Taskonomy for Non-IID Data. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 8719–8730. [[CrossRef](#)] [[PubMed](#)]
62. Duan, M.; Liu, D.; Ji, X.; Wu, Y.; Liang, L.; Chen, X.; Tan, Y. Flexible Clustered Federated Learning for Client-Level Data Distribution Shift. *IEEE Trans. Parallel Distrib. Syst.* **2021**, *33*, 2661–2674. [[CrossRef](#)]
63. Bai, J.; Sajjanhar, A.; Xiang, Y.; Tong, X.; Zeng, S. FedEWA: Federated Learning with Elastic Weighted Averaging. In *Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18 July 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–8.
64. He, C.; Annavaram, M.; Avestimehr, S. Group Knowledge Transfer: Federated Learning of Large Cnns at the Edge. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 14068–14080.
65. Ozkara, K.; Singh, N.; Data, D.; Diggavi, S. QuPeD: Quantized Personalization via Distillation with Applications to Federated Learning. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 3622–3634.
66. Long, G.; Xie, M.; Shen, T.; Zhou, T.; Wang, X.; Jiang, J. Multi-Center Federated Learning: Clients Clustering for Better Personalization. *World Wide Web* **2023**, *26*, 481–500. [[CrossRef](#)]
67. Lin, F.P.-C.; Hosseinalipour, S.; Azam, S.S.; Brinton, C.G.; Michelusi, N. Semi-Decentralized Federated Learning with Cooperative D2D Local Model Aggregations. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 3851–3869. [[CrossRef](#)]
68. Uddin, M.P.; Xiang, Y.; Lu, X.; Yearwood, J.; Gao, L. Federated Learning via Disentangled Information Bottleneck. *IEEE Trans. Serv. Comput.* **2023**, *16*, 1874–1889. [[CrossRef](#)]
69. Du, J.; Qin, N.; Huang, D.; Zhang, Y.; Jia, X. An Efficient Federated Learning Framework for Machinery Fault Diagnosis with Improved Model Aggregation and Local Model Training. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 10086–10097. [[CrossRef](#)] [[PubMed](#)]
70. Ying, B.; Li, Z.; Yang, H. Exact and Linear Convergence for Federated Learning under Arbitrary Client Participation Is Attainable. In *Advances in Neural Information Processing Systems*; Belgrave, D., Zhang, C., Lin, H., Pascanu, R., Koniusz, P., Ghassemi, M., Chen, N., Eds.; Curran Associates Inc.: Red Hook, NY, USA, 2025; Volume 38, pp. 40156–40201.
71. Goetz, J.; Malik, K.; Bui, D.; Moon, S.; Liu, H.; Kumar, A. Active Federated Learning. *arXiv* **2019**, arXiv:1909.12641. [[CrossRef](#)]
72. Ranathunga, T.; McGibney, A.; Rea, S.; Bharti, S. Blockchain-Based Decentralized Model Aggregation for Cross-Silo Federated Learning in Industry 4.0. *IEEE Internet Things J.* **2023**, *10*, 4449–4461. [[CrossRef](#)]
73. Deng, Y.; Lyu, F.; Ren, J.; Chen, Y.-C.; Yang, P.; Zhou, Y.; Zhang, Y. FAIR: Quality-Aware Federated Learning with Precise User Incentive and Model Aggregation. In *Proceedings of the IEEE INFOCOM 2021—IEEE Conference on Computer Communications, Vancouver, BC, Canada, 10 May 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–10.

74. Deng, Y.; Lyu, F.; Ren, J.; Zhang, Y.; Zhou, Y.; Zhang, Y.; Yang, Y. SHARE: Shaping Data Distribution at Edge for Communication-Efficient Hierarchical Federated Learning. In *Proceedings of the 2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS), Washington, DC, USA, 7–10 July 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 24–34.
75. Hosseinalipour, S.; Azam, S.S.; Brinton, C.G.; Michelusi, N.; Aggarwal, V.; Love, D.J.; Dai, H. Multi-Stage Hybrid Federated Learning Over Large-Scale D2D-Enabled Fog Networks. *IEEE/ACM Trans. Netw.* **2022**, *30*, 1569–1584. [[CrossRef](#)]
76. Chai, Z.; Ali, A.; Zawad, S.; Truex, S.; Anwar, A.; Baracaldo, N.; Zhou, Y.; Ludwig, H.; Yan, F.; Cheng, Y. TiFL: A Tier-Based Federated Learning System. In *Proceedings of the 29th International Symposium on High-Performance Parallel and Distributed Computing, Stockholm, Sweden, 23–26 June 2020*; ACM: New York, NY, USA, 2020; pp. 125–136. [[CrossRef](#)]
77. Tang, M.; Ning, X.; Wang, Y.; Sun, J.; Wang, Y.; Li, H.; Chen, Y. FedCor: Correlation-Based Active Client Selection Strategy for Heterogeneous Federated Learning. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 10092–10101.
78. Zhang, W.; Wang, X.; Zhou, P.; Wu, W.; Zhang, X. Client Selection for Federated Learning with Non-IID Data in Mobile Edge Computing. *IEEE Access* **2021**, *9*, 24462–24474. [[CrossRef](#)]
79. Deng, Y.; Lyu, F.; Ren, J.; Wu, H.; Zhou, Y.; Zhang, Y.; Shen, X. AUCTION: Automated and Quality-Aware Client Selection Framework for Efficient Federated Learning. *IEEE Trans. Parallel Distrib. Syst.* **2022**, *33*, 1996–2009. [[CrossRef](#)]
80. Cho, Y.J.; Wang, J.; Joshi, G. Towards Understanding Biased Client Selection in Federated Learning. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, Virtual Event, 28–30 March 2022*; PMLR: New York, NY, USA, 2022; Volume 151, pp. 10351–10375.
81. Abdulrahman, S.; Tout, H.; Mourad, A.; Talhi, C. FedMCCS: Multicriteria Client Selection Model for Optimal IoT Federated Learning. *IEEE Internet Things J.* **2021**, *8*, 4723–4735. [[CrossRef](#)]
82. Wang, H.; Kaplan, Z.; Niu, D.; Li, B. Optimizing Federated Learning on Non-IID Data with Reinforcement Learning. In *Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications, Toronto, ON, Canada, 6–9 July 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1698–1707.
83. Wang, Y.; Kantarci, B. A Novel Reputation-Aware Client Selection Scheme for Federated Learning within Mobile Environments. In *Proceedings of the 2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Pisa, Italy, 14–16 September 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
84. Jee Cho, Y.; Gupta, S.; Joshi, G.; Yagan, O. Bandit-Based Communication-Efficient Client Selection Strategies for Federated Learning. In *Proceedings of the 2020 54th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 1 November 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1066–1069.
85. Ma, X.; Shi, W.; Wen, J. An Enhanced Combinatorial Contextual Neural Bandit Approach for Client Selection in Federated Learning. *ACM Trans. Model. Perform. Eval. Comput. Syst.* **2025**, *10*, 9. [[CrossRef](#)]
86. Rokvic, L.; Danassiss, P.; Faltings, B. Lazy But Effective: Collaborative Personalized Federated Learning with Heterogeneous Data. In *Proceedings of the 2025 International Joint Conference on Neural Networks (IJCNN), Rome, Italy, 30 June July 2025*; pp. 1–8. [[CrossRef](#)]
87. Jimenez-Gutierrez, D.-M.; Solans, D.; Elbamby, M.; Kourtellis, N. PSI-PFL: Population Stability Index for Client Selection in non-IID Personalized Federated Learning. *arXiv* **2025**, arXiv:2506.00440. [[CrossRef](#)]
88. Pandey, S.R.; Nguyen, L.D.; Popovski, P. A Contribution-Based Device Selection Scheme in Federated Learning. *IEEE Commun. Lett.* **2022**, *26*, 2057–2061. [[CrossRef](#)]
89. Zhang, J.; Wu, Y.; Pan, R. Incentive Mechanism for Horizontal Federated Learning Based on Reputation and Reverse Auction. In *Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19 April 2021*; ACM: New York, NY, USA, 2021; pp. 947–956.
90. Wu, W.; He, L.; Lin, W.; Mao, R.; Maple, C.; Jarvis, S. SAFA: A Semi-Asynchronous Protocol for Fast Federated Learning with Low Overhead. *IEEE Trans. Comput.* **2021**, *70*, 655–668. [[CrossRef](#)]
91. Zhu, J.; Yao, R.; Ceritli, T.; Ozkan, S.; Blaschko, M.B.; Noh, E.; Min, J.; Min, C.J.; Ozay, M. Guided Model Merging for Hybrid Data Learning: Leveraging Centralized Data to Refine Decentralized Models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Tucson, AZ, USA, 6–10 March 2026*; pp. 3182–3192. [[CrossRef](#)]
92. Mali, S.; Mali, N.; Zeng, F.; Zhang, L. Edge Computing with Federated Learning for Early Detection of Citric Acid Overdose and Adjustment of Regional Citrate Anticoagulation. *BMC Med. Inform. Decis. Mak.* **2025**, *25*, 320. [[CrossRef](#)] [[PubMed](#)]
93. Geng, D.; He, H.; Lan, X.; Liu, C. Bearing Fault Diagnosis Based on Improved Federated Learning Algorithm. *Computing* **2022**, *104*, 1–19. [[CrossRef](#)]
94. Li, S.; Zhou, T.; Tian, X.; Tao, D. Learning to Collaborate in Decentralized Learning of Personalized Models. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 9756–9765.
95. Wu, H.; Wang, P. Fast-Convergent Federated Learning with Adaptive Weighting. *IEEE Trans. Cogn. Commun. Netw.* **2021**, *7*, 1078–1088. [[CrossRef](#)]

96. Li, Z.; Chen, Z.; Wei, X.; Gao, S.; Ren, C.; Quek, T.Q.S. HPFL-CN: Communication-Efficient Hierarchical Personalized Federated Edge Learning via Complex Network Feature Clustering. In *Proceedings of the 2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Stockholm, Sweden, 20 September 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 325–333.
97. Briggs, C.; Fan, Z.; Andras, P. Federated Learning with Hierarchical Clustering of Local Updates to Improve Training on Non-IID Data. In *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–9.
98. Palihawadana, C.; Wiratunga, N.; Wijekoon, A.; Kalutarage, H. FedSim: Similarity Guided Model Aggregation for Federated Learning. *Neurocomputing* **2022**, *483*, 432–445. [\[CrossRef\]](#)
99. Bao, W.; Wang, H.; Wu, J.; He, J. Optimizing the Collaboration Structure in Cross-Silo Federated Learning. *Int. Conf. Mach. Learn.* **2023**, *202*, 1718–1736.
100. Liu, S.; Chen, Q.; You, L. Fed2A: Federated Learning Mechanism in Asynchronous and Adaptive Modes. *Electronics* **2022**, *11*, 1393. [\[CrossRef\]](#)
101. Gao, Y.; Liu, L.; Hu, B.; Lei, T.; Ma, H. Federated Region-Learning for Environment Sensing in Edge Computing System. *IEEE Trans. Netw. Sci. Eng.* **2020**, *7*, 2192–2204. [\[CrossRef\]](#)
102. Qiu, H.; Liu, Z.; Zhao, X.; Xue, J.; Hu, W.; Li, M. OLG-FL: A Federated Learning Framework for Optimizing Local Training and Global Aggregation. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology (ICCSMT '25)*; Association for Computing Machinery: New York, NY, USA, 2026; pp. 834–843. [\[CrossRef\]](#)
103. Todorean, L.; Daian, M.; Cioara, T.; Anghel, I.; Michalakopoulos, V.; Sarantinopoulos, E.; Sarmas, E. Heuristic-Based Federated Learning with Adaptive Hyperparameter Tuning for Households Energy Prediction. *Sci. Rep.* **2025**, *15*, 12564. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Bai, R.; Wang, H.; Sun, W.; He, L.; Shi, Y.; Xu, Q.; Wang, Y.; Chen, X. Intelligent Diagnosis of Gearbox in Data Heterogeneous Environments Based on Federated Supervised Contrastive Learning Framework. *Sci. Rep.* **2025**, *15*, 14596. [\[CrossRef\]](#) [\[PubMed\]](#)
105. Azimi-Abarghouyi, S.M.; Fodor, V. A Hierarchical Federated Learning Approach for Internet of Things. *IEEE Internet Things J.* **2026**, *13*, 12655–12672. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.