International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V74I8P106

Volume 74 Issue 8, 47-62, August 2026

Original Article

Participation-Aware Search-Based Joint Client Selection and Adaptive Aggregation for Federated Learning

Meenakshi Devi[1], Rakesh Kumar[2]

[1,2]*Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, Haryana, India.*

[1]*Corresponding Author : meenakshi.chauhan2001@kuk.ac.in*

Received: 27 June 2026 Revised: 29 July 2026 Accepted: 19 August 2026 Published: 30 August 2026

Abstract *- Federated learning is affected by differences in client data, behavior, and resource requirements. This paper develops and evaluates a search-based, client-participation-aware strategy for federated client selection and adaptive aggregation. It uses a genetic-search procedure with a multi-criteria objective. To avoid unstable aggregation, the searched weights are combined with the data-size weights. The experiment is performed on the non-IID MNIST dataset using a lightweight CNN. The proposed method is compared with standard federated learning baselines: Federated Averaging (FedAvg) and Federated Proximal (FedProx). The comparison also includes selection strategies such as loss-based and resource-aware selection. The method achieves the best final predictive performance, with an accuracy of 0.31, a macro-F1 of 0.22, and a loss of 2.14. However, it is limited by the final client disparity gap. Its selection entropy and stability variance also do not outperform all baselines. The proposed multi-criteria search improves predictive performance. Furthermore, ablation analysis shows performance deterioration when the participation component is removed. Similar degradation occurs when the searched weight coefficient in the aggregation varies. The fitness-weight and aggregation-ratio sensitivity analysis reports their influence on both predictive and participation behavior. However, the statistical analysis found that the search-based method is not significantly more predictive than the baselines. The results highlight the usefulness of participation-aware search for improving prediction-oriented federated training under non-IID data. Further, it is suggested that client-level balance requires more explicit disparity-aware design.*

Keywords - *Federated Learning, Deep Learning, Genetic Search, non-IID Data, Privacy Preservation, Participation-Aware Learning.*

1. Introduction

Traditionally, centralized learning frameworks collected data from multiple sources to train models. The modeling process is limited when data are sensitive or cannot be shared. Federated Learning (FL) handles this limitation by enabling clients to train their local models. The resulting model updates are sent to the server, where they are applied to the global model. Therefore, FL is well-suited for privacy-sensitive and distributed environments. It includes mobile devices, edge systems, healthcare networks, and IoT applications [1], [2]. Although data are not centralized, the challenges arise from differing data sizes and distributions, as well as client computational and communication resource constraints.

The challenges are more prominent with non-Independent and Identically Distributed (non-IID) data and limited client participation. They lead to inconsistent client updates, affect model convergence, increase variation in client-level performance, and make aggregation less stable [2]–[4].

Client selection is an important component in federated learning. It allows only a subset of clients to participate in each communication round. The selected clients directly influence the predictive performance, convergence behavior, resource use, participation balance, and client-level disparity. Prior client-selection studies have shown that guided participant selection and resource-aware selection can improve federated training efficiency. They considered client utility, training speed, communication condition, or device resource constraints [5], [6].

Studies have focused on client selection strategies with a single criterion. They included the loss-based and the resource-aware selection strategies. The former may prioritize difficult clients with useful learning signals, but repeatedly favors clients with large local errors. The latter may reduce the computational or communication burden, but excludes informative clients with higher resource costs. Recent studies have shown that client heterogeneity, resource constraints, participation diversity, and aggregation behavior matter together, not in isolation [6], [7].

In many federated learning configurations, the server first selects participating clients. Then it applies a fixed aggregation rule based on the size of their local data. Under non-IID data, however, a client update's contribution may depend not only on the sample size. It also includes the current predictive behavior, resource burden, update divergence, and participation history [7], [8]. Despite these developments, federated learning still treats client selection and aggregation as separate decisions. This separation can be restrictive under non-IID conditions and may not capture the trade-offs adequately. This creates a need to explore the trade-off search space and evaluate solutions using multiple client characteristics. Therefore, this paper proposes a participation-aware search-based federated learning strategy for non-IID MNIST (Modified National Institute of Standards and Technology) dataset classification. It uses a genetic search procedure and evaluates whether multi-criteria search can improve federated training in a constrained, heterogeneous setting. The proposed work addresses the identified issue and has the following main contributions:

- A participation-aware search-based federated learning strategy is proposed to optimize client selection and aggregation weights jointly under non-IID data.
- A stabilized aggregation mechanism is introduced by blending the searched aggregation weights with data-size weights, thereby limiting the excessive influence of clients with small local datasets.
- A multi-criteria fitness function is formulated using predictive utility, client disparity, resource cost, update divergence, and participation history to guide client selection.
- A comparison with federated baselines, including FedAvg (Federated Averaging), FedProx (Federated Proximal), loss-based selection, and resource-aware selection, is conducted to assess the effect of the proposed joint strategy.
- Ablation, sensitivity, and statistical analyses evaluate fitness component contributions, parameter sensitivity, and the robustness and reliability of the search-based method, respectively.

The paper is arranged as follows. Section 2 presents the related work on federated learning, including data distribution, client selection, resource-aware participation, and many other aspects. Section 3 explains the proposed methodology. Section 4 analyzes and compares the results with the federated baselines. It also provides various analyses for the contribution and significance of the proposed method. Section 5 concludes the study and suggests future extensions.

2. Related Work

Federated learning research varies from basic training to addressing issues arising from different components. It includes non-IID data, limited client availability, communication cost, resource heterogeneity, aggregation design, and uneven client-level performance [9]–[11]. These issues are closely related, as differences in client data and system resources influence participation, aggregation, and training behavior. The following subsections review prior work addressing these aspects of federated learning.

2.1. Federated Learning under Non-IID Data

In federated learning, local clients may have different data distributions, sample sizes, and learning behavior. Recent surveys reported that client heterogeneity affects global model convergence and generalization. In addition, client heterogeneity can affect communication efficiency and the stability of federated training [9]–[11]. Furthermore, recent adaptive and personalized federated learning studies have shown that a single fixed global update may be insufficient to address client heterogeneity. Thus, FedALA adapted the interaction between local and global model information [12]. At the same time, contextual client selection exploited relationships among local data [13]. These studies indicated the need to use client-specific information. However, they did not directly formulate each candidate solution as both a client subset and a set of aggregation weights.

2.2. Client Selection and Participation Control

Client selection is a major research direction because only a subset of clients participates in training. Recent work has studied client selection based on data similarity and quality awareness. In addition, contextual client selection and adaptive participation control have improved the significance of the selected clients [13]-[16]. They reported that client selection strongly influences convergence behavior and model quality. This has motivated selection strategies beyond random sampling. Recently, structured client selection mechanisms have been introduced. One was FedDSS, which selected clients using a data-based similarity approach [15]. Another study developed CCSF [17], which clustered clients under non-IID data and used multi-tier selection to address participation in mobile networks [18]. They reported that optimized selection can improve federated learning performance.

2.3. Resource-Aware and Communication-Efficient Federated Learning

Incorporating resource awareness and communication cost into federated learning is important. Studies highlighted the need to reduce unnecessary participation and model transmission. One such study, ACSP-FL, adapted client participation to reduce communication costs [16]. Several recent studies have focused on energy-aware selection, mobility-aware participation, and joint class-balanced client selection with bandwidth allocation [19]–[22]. FedLDF used layer-divergence feedback to reduce communication overhead during aggregation [23]. These studies motivate including resource costs and update divergence in the federated learning pipeline.

2.4. Aggregation and Adaptive Weighting in Federated Learning

Aggregation strategy is a major research focus in federated learning. FedALA [12] adapted local aggregation, while FedLDF [23] used divergence feedback. Some reviews show that aggregation strategies differ in how they address key aspects of federated learning. They included heterogeneity, robustness, personalization, communication cost, and model-update reliability [24], [25].

In addition, label-aware aggregation [26] incorporated label distribution information, whereas FedCW [27] combined adaptive client selection with dynamic aggregation. FedCW prioritized clients based on model divergence and included data volume to assign aggregation weights.

2.5. Client-Level Disparity, Fairness and Participation Balance

Client-level fairness and participation balance have received increasing attention in federated learning. Repeatedly selecting a limited group of clients may lead to uneven participation. Studies such as FedSDR [28], FairFedCS [29], secure and fair DDPG-based selection [30], and fair multi-task client selection [31] have focused on this aspect. They showed that fairness-aware participation is an important issue in heterogeneous federated learning.

Moreover, user selection and multi-objective client selection studies showed that client characteristics, distributional differences, and fairness-related objectives can be incorporated into the selection process. SCKM [32] used a higher-order statistical criterion for user selection. In contrast, multi-objective distributed client selection [33] considered multiple selection objectives in federated learning-assisted Internet-of-Vehicles settings.

2.6. Search-Based and Multi-Criteria Optimization in Federated Learning

Some studies formulated client selection as an optimization problem with selection mechanisms. EACS-FL has energy-aware selection using a combinatorial multi-armed bandit setting [19]. Further works have used multi-criteria selection by jointly handling label distribution and communication cost. They used class-balanced client selection with bandwidth allocation in mobile edge computing [22]. DDPG-based selection used deep reinforcement learning to enable secure and fair client participation [30]. Dynamic scoring-based selection is another strategy that considered accuracy, loss, and execution time in federated medical diagnosis [34]. Recently, Eslami et al. [35] proposed FedGAO, which applied the Green Anaconda Optimizer to coordinate client participation and aggregation.

2.7. Comparative Analysis and Research Gap

The reviewed studies show that client participation in federated learning has been addressed for various aspects. The focused areas were resource constraints, communication efficiency, client fairness, and model-update behavior. Search and optimization methods have also been employed to improve client selection. Table 1 compares several presented approaches by their selection basis, aggregation mechanism, optimization strategy, Participation Awareness (PA), and Joint Evolutionary Search (JES) of the client subset and aggregation weights. The comparison indicates that although several approaches consider client selection and aggregation, they are not evaluated jointly. Moreover, few approaches explicitly use client participation history to discourage repeated involvement. In the proposed search-based approach, multiple criteria are considered, and the searched aggregation weights are used to stabilize the global update.

Table 1. Comparison of Client Selection and Aggregation Strategies in Federated Learning

Study	Selection basis	Aggregation	Method	Criteria	PA	JES
Kang & Ahn [8]	Client combinations	Standard	GA	Update, comm. cost	No	No
Pan et al. [13]	Contextual	Standard	NCCB	Context, reward	No	No
Song et al. [14]	Data quality	Selective	Quality-aware	Loss sharpness	No	No
Zhu et al. [19]	Energy/resource	Standard	CMAB	Energy, delay	Partial	No
Tang et al. [22]	Class balance	Standard	Joint opt.	Balance, cost	No	No
Wang et al. [23]	No	Divergence-based	Feedback	Divergence, comm.	No	No
FedCW [27]	Divergence	Adaptive	Adaptive	Divergence, data size	No	No
Mao et al. [28]	Fairness	Periodic avg.	Dynamic	Representativity, fairness	Yes	No
Cha & Chang [33]	Multi-criteria	Standard	Fuzzy	Data, resource	No	No
Ahmed et al. [34]	Dynamic score	Standard	Scoring	Acc., loss, time	No	No
FedGAO [35]	Adaptive	Weighted	GAO	Resource, heterogeneity	No	No
Search-based	Multi-criteria	Search + size	Genetic search	P, D, C, U, B	Yes	Yes

Note: PA = participation awareness; JS = joint search of client subset and aggregation weights; P = predictive utility; D = client disparity; C = resource cost; U = update divergence; B = historical selection bonus

3. Materials and Methods

This study investigates whether joint search-based client selection and aggregation improve federated learning under non-IID data. All compared methods were performed under the same federated setting, dataset partition, client population, model architecture, and evaluation procedure. They differ in their client selection and aggregation strategies. The proposed method represents each candidate as a selected client subset together with its aggregation-weight vector. Each candidate is evaluated with a fitness function using multiple criteria. A genetic search then identifies the client subset and corresponding weights for each communication round. Moreover, the searched weights are combined with data-size weights for stabilized global aggregation. The workflow, shown in Figure 1, consists of the major phases. The federated strategy is implemented in Python, with PyTorch used for CNN construction, local model training, and federated model updates.

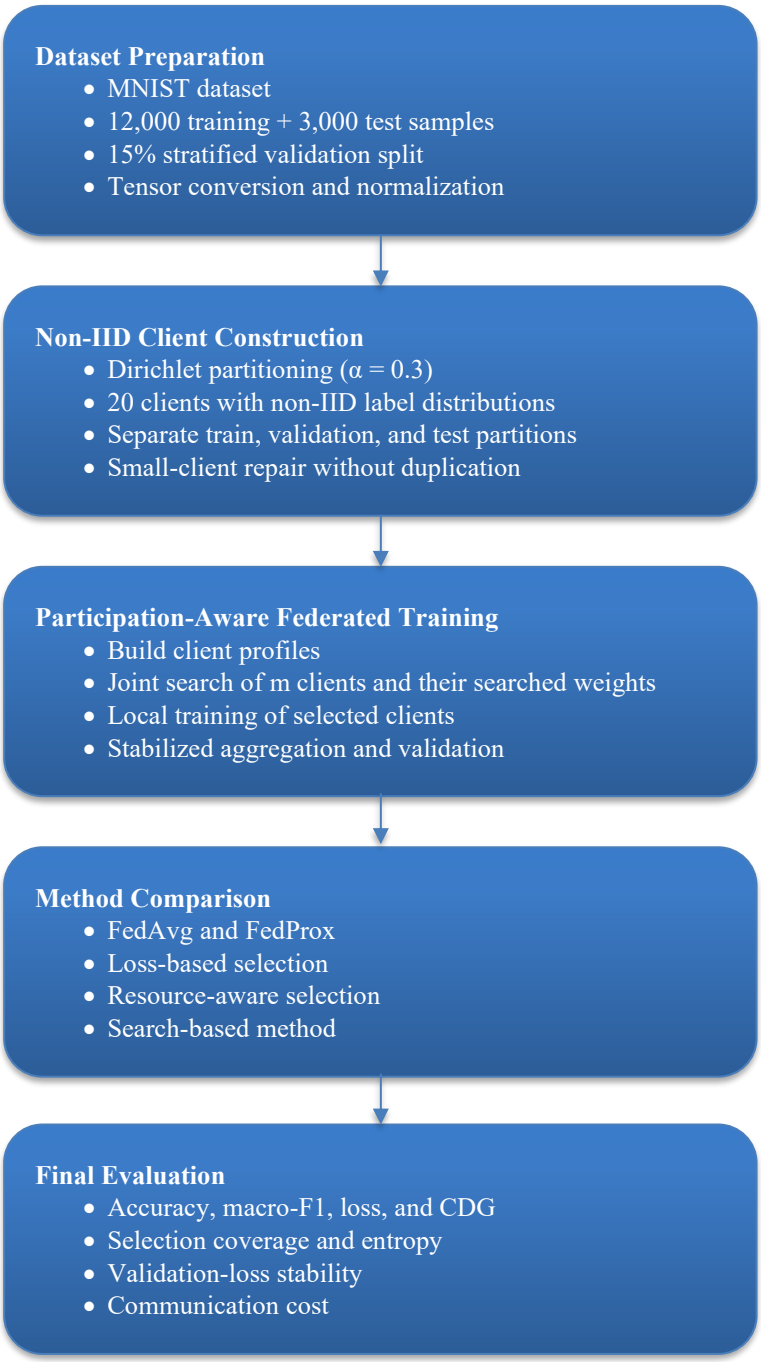

Fig. 1 The workflow of the proposed participation-aware search-based federated learning framework

3.1. Federated Setting and Non-IID MNIST Client Construction

The developed system consists of a central server and K = 20 clients. At communication round t, the server maintains a global model with parameters w^t. Since full client participation is not carried out, only m=5 clients are selected per round. Each selected client receives the current global model and trains it locally using its own data. The client then sends the updated local model back to the server. The server then updates the global model through weighted aggregation of the selected local models.

The benchmark handwritten digit dataset MNIST [36] has 60,000 training and 10,000 test images. The computational cost is reduced with a randomly sampled subset of 12,000 training instances. It is further divided through a stratified split, with a 15% validation ratio.

In addition, 3,000 samples are randomly drawn from the original MNIST test set for final evaluation. All the sampled images are converted into tensors and normalized using the dataset-specific mean and standard deviation. A fixed random seed is used to support reproducibility.

Client heterogeneity is introduced using Dirichlet-based non-IID partitioning. For each class c, the sample indices belonging to that class are distributed among the K clients according to the client-allocation proportion vector defined in Equation (1).

$$\pi_c \sim Dirichlet(\alpha 1_K) \quad (1)$$

where π_c denotes the client-allocation proportion vector for samples of class c. The variable α is the Dirichlet concentration parameter, and the 1_K denotes a K-dimensional vector of ones. In this study, α = 0.3 is used to generate non-IID label skew across clients. This setting produces unequal class proportions across clients and is therefore used to construct the non-IID federated environment. The training, validation, and test subsets are partitioned separately using the same non-IID construction procedure. Moreover, a repair step is applied when necessary by transferring samples from sufficiently populated clients to smaller client partitions without duplicating samples. The resulting client datasets, therefore, differ in both sample composition and label distribution.

3.2. Local Model Training and Client Profiling

All federated strategies use the same lightweight Convolutional Neural Network (CNN) architecture to ensure a consistent learning model. CNN accepts single-channel image inputs and predicts the 10 MNIST classes. It includes two convolutional layers, max pooling, an adaptive average pooling layer, a fully connected hidden layer, and a final classification layer. Adaptive average pooling transforms the extracted feature maps to a fixed spatial dimension. Thus, enabling a compact network architecture for repeated local training across communication rounds. The local model training employs Stochastic Gradient Descent (SGD), a batch size of 64, and one local epoch per communication round.

Before selecting clients in each round, the server constructs a profile (p) for every client using the current global model. The profile provides predictive, data-volume, resource-related, and update-behaviour information for subsequent candidate evaluation. It is represented in Equation (2) for client k and communication round t.

$$p_k^t = \{L_k^t, A_k^t, n_k, R_k, D_k^{t-1}\} \quad (2)$$

where L_k^t is the local loss, A_k^t is the local accuracy, n_k is the number of client samples, R_k is the simulated resource cost, and D_k^{t-1} is the previous update divergence. The local loss and local accuracy are computed using the current global model on at most 128 local samples from each client. The resource cost is simulated as an equal-weight combination of the normalized client data size and a random delay. It represents variations in client-side computational or communication burden.

The update divergence is measured as the Euclidean distance between the locally trained model and the corresponding global model. It shows how far a client's update deviates from the global model. Including it penalizes clients whose previous local updates deviated substantially from the corresponding global model.

3.3. Proposed Participation-Aware Search-Based Client Selection

In this stage, instead of selecting clients randomly, by highest loss, or by lowest resource cost, the proposed method selects a client subset based on multiple criteria. Figure 2 illustrates the internal workflow of this core stage. Each communication round begins with client profiling followed by candidate generation, multi-criteria fitness evaluation, and client selection.

The selected clients then perform local training, after which the update divergence is computed, and the local models are aggregated using stabilized weights. The resulting global model is evaluated on the validation data before the next round. Each candidate solution at round t is defined as

$$c^t = (S^t, \lambda^t) \quad (3)$$

where S^t is a subset of five selected clients, and $\lambda^t = [\lambda_1^t, \lambda_2^t, \dots, \lambda_m^t]$ is the corresponding vector of searched aggregation weights. Thus, the search space comprises both discrete and continuous components.

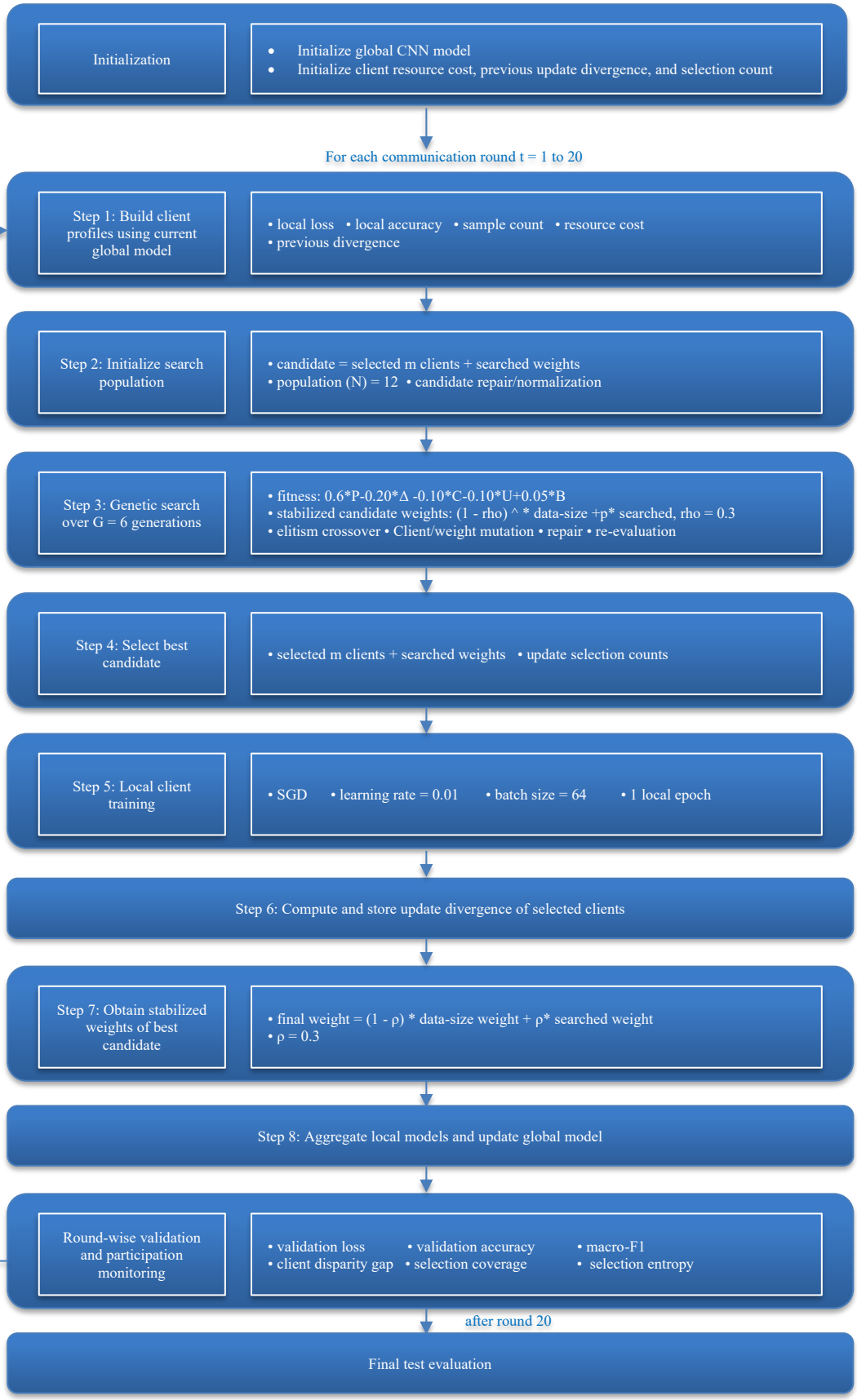

Fig. 2 The workflow of the participation-aware federated learning process

A genetic search determines both the participating clients and their aggregation weights. The search maintains twelve candidate solutions, each with five clients. The initial candidates are generated randomly. Duplicate client IDs that may appear after crossover or mutation are replaced to keep five distinct clients, and the associated weights are normalized. Better candidates are retained through elitism. New candidates are obtained by combining two parents, while mutation introduces variation either by changing one of the selected clients or modifying its weight. The entire genetic procedure is repeated for six generations. Algorithm 1 summarizes the complete sequence of the joint genetic-search procedure.

Algorithm 1. Participation-Aware Genetic Search for Joint Client Selection and Aggregation Weighting

Input: Client profiles, number of selected clients m=5, population size N=12, generations G=6
Output: Selected client subset S and searched weight vectorλ.
1. Initialize N candidates (S^t, λ^t) by randomly selecting m distinct clients and generating normalized candidate weights.
2. for g=1 to G **do**
  Evaluate the fitness (F) of each candidate.
  Sort candidates by fitness and retain the best $max(2, \lfloor N/3 \rfloor)$ candidates as elites.
  Copy the elites to the new population.
  while the new population size is less than N **do**
    Randomly select two parents from the elite set.
    Choose a random crossover point and generate a child by combining the corresponding client and weight components of the parents.
    With probability 0.4, replace one randomly selected client in the child.
    With probability 0.5, perturb the child weights using Gaussian noise.
    Repair duplicate client IDs and normalize the child weights.
    Add the child to the new population.
  end while
  Replace the current population with the new population.
end for
3. Return the candidate with the highest fitness (S^*, λ^*).

Algorithm 1 shows that the client mutation changes the participating subset for each candidate. In addition, weight mutation modifies the selected clients' contribution. During weight mutation, Gaussian perturbations with zero mean and a standard deviation of 0.1 are applied. Fitness evaluation ranks these candidate solutions. For the search-based method, the additional computational cost arises from the genetic search. For a population of N candidates evolved over G generations, the method performs a total of GN candidate fitness evaluations. Thus, the fitness-evaluation cost is $O(GNC_F)$, where $C_F$denotes the cost of evaluating one candidate. Each candidate contains m selected clients and $C_F = O(m)$; therefore, the total fitness-evaluation cost is $O(GNm)$. In addition, the population ranking cost is $O(N \log N)$, while the crossover, mutation, repair, and normalization costs are $O(m)$. Across all generations, the ranking cost is $O(GN \log n)$, and the other genetic operations cost $O(GNm)$. Thus, the overall genetic-search complexity is $O(GNm + GN \log N))$. This complexity is separate from the computational cost of local CNN training, which is incurred by all federated strategies.

The fitness calculation considers more than just a client's predictive behavior. For each client, a quality score is computed from its normalized accuracy, normalized number of local samples, and (1-normalized loss). Their contributions are set to 0.40, 0.30, and 0.30, respectively. The weights carried by a candidate are not used directly for aggregation. They contribute 30% of the final weight, while the remaining 70% comes from the relative data size of the selected clients. The resulting aggregation weight $\omega_k^t$is then used in the predictive utility, resource-cost, and divergence terms. For a selected subset, the predictive utility is

$$P(S^t) = \sum_{k \in S^t} \omega_k^t (0.40\tilde{A}_k^t + 0.30\tilde{n}_k + 0.30(1 - \tilde{L}_k^t) \quad (4)$$

Equation (5) measures the disparity among selected clients using the range of their normalized losses. A small range indicates similar loss values, whereas a large range reflects greater disparity. The resource cost (C) and the update divergence (U) are included as penalties and defined in Equations (6) and (7). The former reflects the simulated burden associated with a client, while the latter measures how far its previous local update deviated from the global model.

$$\Delta(S^t) = \max_{k \in S^t} \tilde{L}_k^t - \min_{k \in S^t} \tilde{L}_k^t \quad (5)$$

$$C(S^t) = \sum_{k \in S^t} \omega_k^t \tilde{R}_k \quad (6)$$

$$U(S^t) = \sum_{k \in S^t} \omega_k^t \tilde{D}_k^{prev} \quad (7)$$

A client that has participated frequently receives less preference than one who has been selected less often. A historical selection bonus (B) introduces this participation awareness and is computed as

$$B(S^t) = \frac{1}{|S^t|} \sum_{k \in S^t} (1 - \tilde{h}_k^t) \quad (8)$$

where $\tilde{h}_k^t$ is the normalized number of times client k has been selected before round t. Thus, participation history is considered alongside other selection criteria. It discourages repeated preference for the same clients. The complete fitness function (F) is expressed as

$$F(c^t) = \alpha_{perf} P(S^t) - \beta_{disp} \Delta(S^t) - \gamma_{cost} C(S^t) - \delta_{div} U(S^t) + \eta_{part} B(S^t) \quad (9)$$

The coefficient values are α_{perf}= 0.60, β_{disp}= 0.20, γ_{cost}= 0.10, δ_{div}= 0.10, and η_{part}= 0.05. For optimization, the primary objective is predictive performance; therefore, the fitness function (F) assigns the highest coefficient to predictive utility. To ensure client-level differences have a significant influence, the client disparity is weighted more than the resource cost and the update divergence. However, the client's historical participation is included as a secondary objective to encourage less frequently selected clients. The resulting fitness function (F) evaluates all these parameters for the client subset S and the searched weight vector λ. The searched weights are then combined with data-size-based weights to obtain the final stabilized aggregation weights.

3.4. Stabilized Search-Based Aggregation and Federated Updating

The proposed method also adapts the aggregation weights of the selected clients. However, relying solely on the searched weights may be unstable. This is because small clients can exert excessive influence. To prevent this, the searched weights are blended with FedAvg-style data-size weights. For selected client k, the data-size weight (a_k^t) is

$$a_k^t = \frac{n_k}{\sum_{j \in S^t} n_j} \quad (10)$$

Let $s_k^t = |\lambda_k^t| / \sum_{j=1}^{m} |\lambda_j^t|$, denote the normalized searched weight. The final aggregation weight is computed as

$$\omega_k^t = (1 - \rho) a_k^t + \rho s_k^t \quad (11)$$

where ρ = 0.30 determines the contribution of the searched weight. The remaining 0.70 comes from the data-size weight. The larger contribution of the data-size weights preserves the influence of local data volume. At the same time, the searched weights allow the genetic procedure to adapt client influence. Thus, search-based weighting adjusts rather than completely replaces data-size-based aggregation. The global model is updated as

$$w^{t+1} = \sum_{k \in S^t} \omega_k^t w_k^{t+1} \quad (12)$$

$$CDG = \max_k A_k - \min_k A_k \quad (13)$$

$$Entropy = \frac{-\sum_{k=1}^{K} p_k \ln(p_k)}{\ln K} \quad (14)$$

Client-level performance disparity is measured using the Client Disparity Gap (CDG) as shown in Equation (13). It measures variation in model performance across clients. In addition, Equation (14) shows the selection entropy, where $p_k = h_k / \sum_{j=1}^{K} h_j$. A higher entropy indicates a more balanced distribution of client participation. These measures are included to promote stable, inclusive, and resource-aware federated training.

The training continues for 20 communication rounds. In each round, clients are profiled before selection and local training. The resulting local models are used to compute the update divergence and update the global model through weighted aggregation. Performance and participation behavior are evaluated using validation-phase and final-test measures. During the validation phase, accuracy, macro-F1, and client disparity gap are recorded after each round. Final testing is performed once, after all communication rounds are complete. The test metrics include accuracy, macro-F1, and loss. In addition, Client Disparity Gap (CDG), selection coverage, selection entropy, and stability variance summarize participation over the complete training process. Communication cost is also included in the final result summary.

3.5. Baseline Federated Strategies

The study evaluates four baseline methods: FedAvg, FedProx, loss-based, and resource-aware. FedAvg is a standard federated learning baseline that uses data-size-based aggregation. FedProx extends it with a proximal regularization term. The latter two strategies prioritize client selection based on a single criterion. The loss-based strategy selects clients with the highest current loss, whereas the resource-aware strategy selects clients with the lowest simulated resource cost. These criteria are also components within the proposed multi-criteria search.

Additionally, ablation, sensitivity, and repeated-run analyses are conducted to evaluate the proposed method. The ablation analysis examines the contribution of fitness components and the searched aggregation weighting. The sensitivity analysis evaluates variations in the predictive-utility and client-disparity coefficients in the fitness function and the searched-weight contribution (ρ) in stabilized aggregation. To assess robustness, all federated strategies are evaluated over 5 independent runs using different random seeds. The results are summarized using the mean and standard deviation. Then, the non-parametric Friedman test is used to assess differences among the five strategies. At the same time, the Wilcoxon signed-rank tests compare the search-based method with each baseline.

4. Results and Discussion

This section presents the experimental results for the federated learning strategies under the non-IID dataset. First, all the learning methods are compared for performance and client behavior on the final test set. In addition, each method's convergence is illustrated on the validation set. The ablation and sensitivity analyses then illustrate the contributions of the proposed fitness components and the effects of key parameters, respectively. Finally, the repeated-

run and statistical analyses assess the robustness and the significance of performance differences among the compared strategies.

4.1. Predictive Performance

Table 2 shows that the proposed search-based method achieves the strongest final predictive performance among the strategies compared. It achieves the highest test accuracy of 0.31 and the highest macro-F1 score of 0.22. It also records the lowest test loss of 2.14. These three values consistently indicate that the proposed method yields the most effective final classification.

The final test accuracy comparison in Figure 3 supports this observation. The search-based method has the highest accuracy bar. It is followed by the loss-based method with an accuracy of 0.20. FedAvg and FedProx achieve similar final accuracies of 0.14. Their macro-F1 values are also identical (in Table 2). These findings show that, under the present configuration, FedProx behaves very similarly to FedAvg. The resource-aware method has the weakest predictive result. It records an accuracy of 0.10, a macro-F1 score of 0.02, and a loss of 2.81. Therefore, the final results show that selecting clients primarily based on resource cost does not yield a strong predictive model.

Table 2. Predictive Performance of the Compared Federated Learning Methods on Non-IID MNIST

Method	Accuracy	Macro-F1	Loss
FedAvg	0.14	0.07	2.25
FedProx	0.14	0.07	2.25
Loss-based	0.20	0.11	2.33
Resource-aware	0.10	0.02	2.81
Search-based	0.31	0.22	2.14

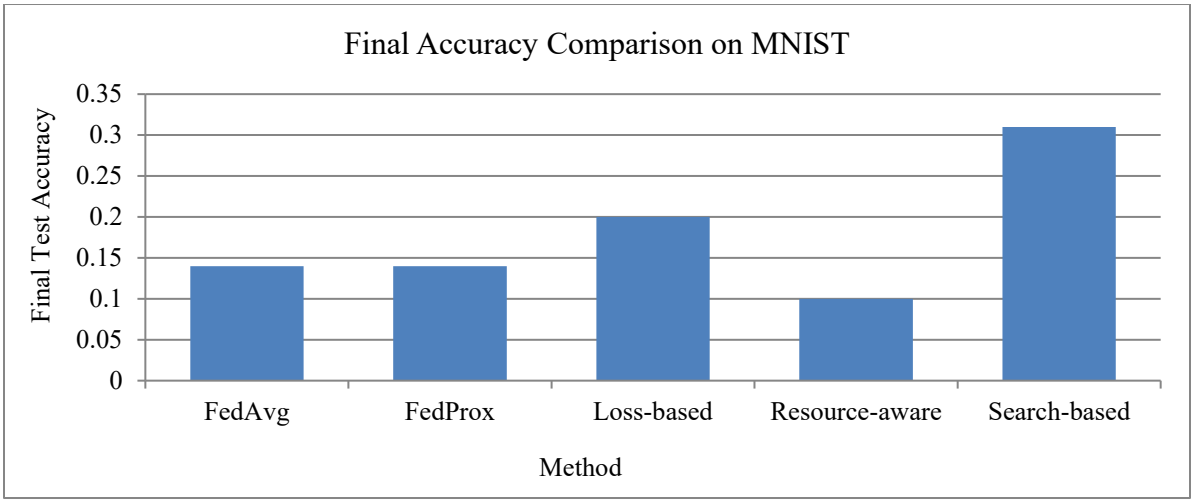

Fig. 1 Final test accuracy under the non-IID MNIST setting

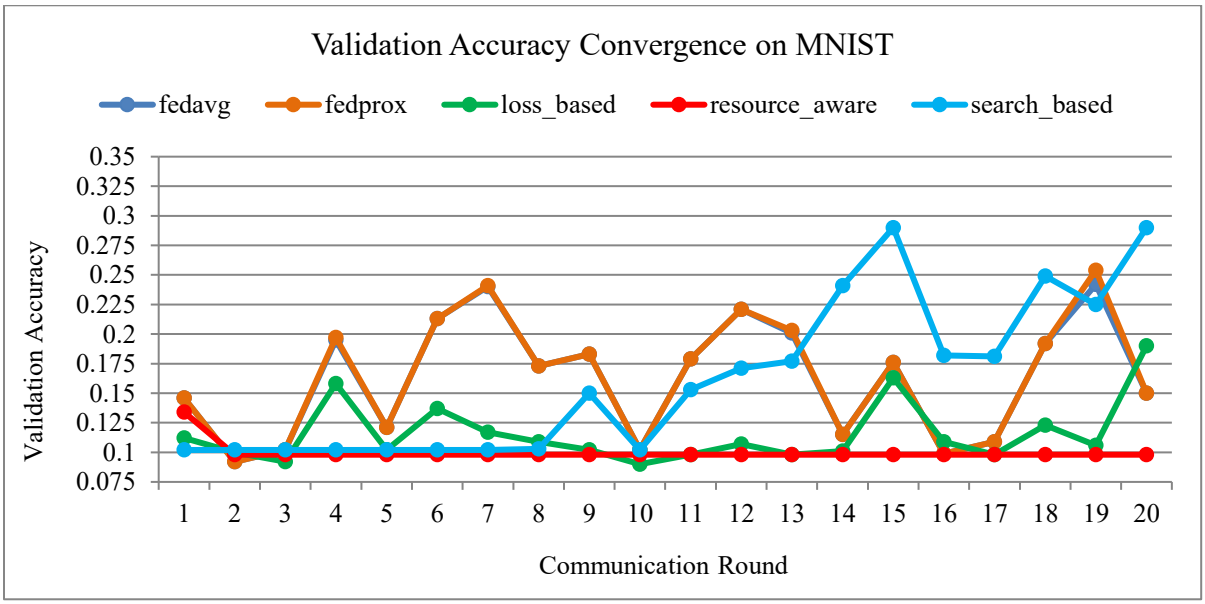

Fig. 4 Round-wise validation accuracy of each method

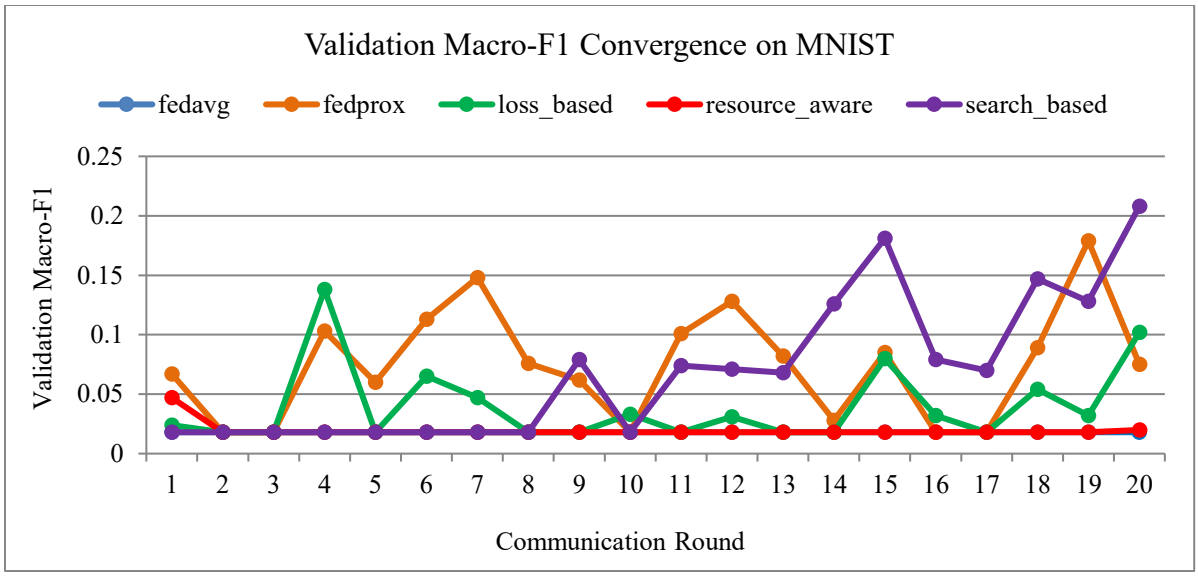

Fig. 5 Round-wise validation macro-F1 of each method

The validation accuracy curve in Figure 4 shows that these methods do not follow the same learning pattern across communication rounds. FedAvg and FedProx remain close to each other during most rounds. It shows consistency with their similar final accuracy and macro-F1 values in Table 2. The loss-based method shows some improvement in selected rounds. However, it does not reach the late-round level achieved by the search-based method. The proposed search-based method shows greater improvement in the later communication rounds. Its validation accuracy increases after the middle of training and reaches the highest values near the end of the process. This trend is consistent with its final test accuracy in Table 2 and the highest bar for final accuracy in Figure 3.

The validation macro-F1 curve in Figure 5 supports the proposed search-based method over the baselines. It achieves the highest late-round macro-F1. In contrast, the resource-aware method remains near the bottom of the plot. This observation agrees with the final test macro-F1 in Table 2. The search-based method achieves the highest macro-F1, while the resource-aware method achieves the lowest.

4.2. Client Behavior

The client disparity results show that the best predictive method is not necessarily the best for client-level uniformity. In Table 3, the search-based method has a final Client Disparity Gap (CDG) of 0.60. This value equals that of the loss-based method and is higher than FedAvg, FedProx, and resource-aware selection. FedAvg and FedProx both obtain a client disparity gap of 0.46, while the resource-aware method obtains the lowest value of 0.44. This observation is also evident in Figure 6, where the search-based and loss-based methods have the largest disparity bars. Moreover, the validation disparity convergence curve in Figure 7 shows that the search-based method has relatively high disparity in later rounds. Therefore, although the search-based method provides the strongest predictive performance, it does not minimize client-level disparity in the current setting.

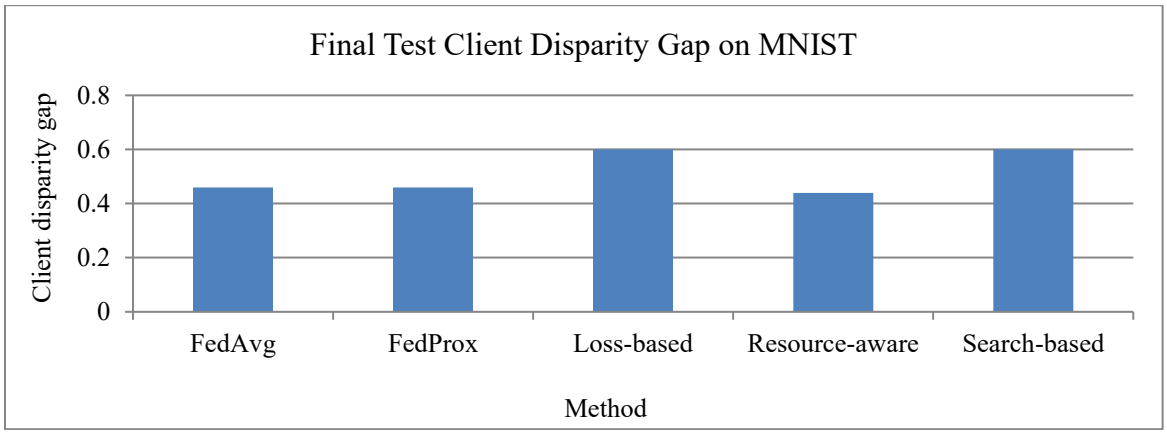

Fig. 6 Final test client disparity gap, measured from client-level test accuracy differences

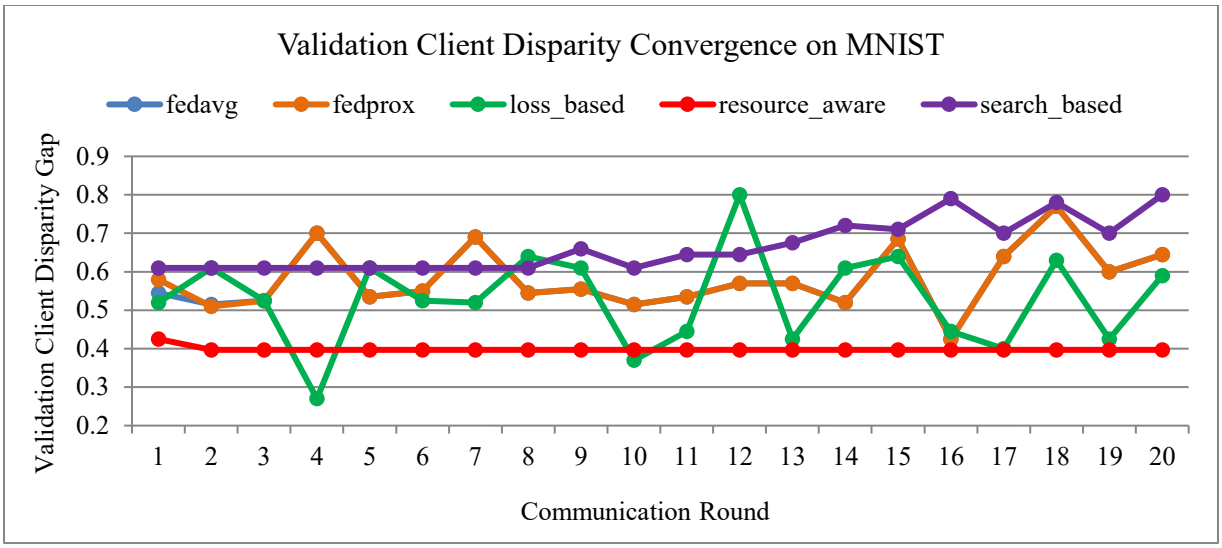

Fig. 2 Round-wise validation client disparity gap of each method

Table 3. Training Behavior of the Compared Federated Learning Methods on Non-IID MNIST

Method	CDG	Coverage	Entropy	Stability Var.
FedAvg	0.46	1.00	0.98	0.0014
FedProx	0.46	1.00	0.98	0.0014
Loss-based	0.60	0.95	0.96	0.0003
Resource-aware	0.44	0.25	0.54	0.0375
Search-based	0.60	0.90	0.88	0.0340

Client participation behavior is measured using selection coverage and selection entropy, as shown in Table 3. Firstly, FedAvg and FedProx achieve full coverage with the highest entropy. Loss-based selection also maintains high coverage of 0.95 with an entropy of 0.96. The coverage of the proposed search-based method is 0.90, with lower entropy than most baselines. Figure 8 shows a bar chart comparing client selection coverage. The resource-aware method has the lowest bar of selection coverage and the lowest entropy (0.54), as shown in Table 3. These values indicate that resource-aware selection concentrates participation on a small subset of clients. In contrast, the search-based method maintains broader participation while still managing the strongest final predictive performance.

The stability variance values in Table 3 indicate that the loss-based method exhibits the smoothest loss behavior, with a stability variance (Stability Var. in Table 3) of 0.0003. FedProx and FedAvg also maintain a lower stability variance value of 0.0014. The search-based method has a higher stability variance of 0.0340 than these, but lower than the resource-aware method. The resource-aware method achieves the highest stability variance of 0.0375. This indicates that the proposed search-based method's advantage lies mainly in final accuracy, macro-F1, and loss, rather than in validation-loss stability. In addition, the communication cost for all methods is the same, at 29.21 MB. It confirms that the same communication configuration was applied across all methods.

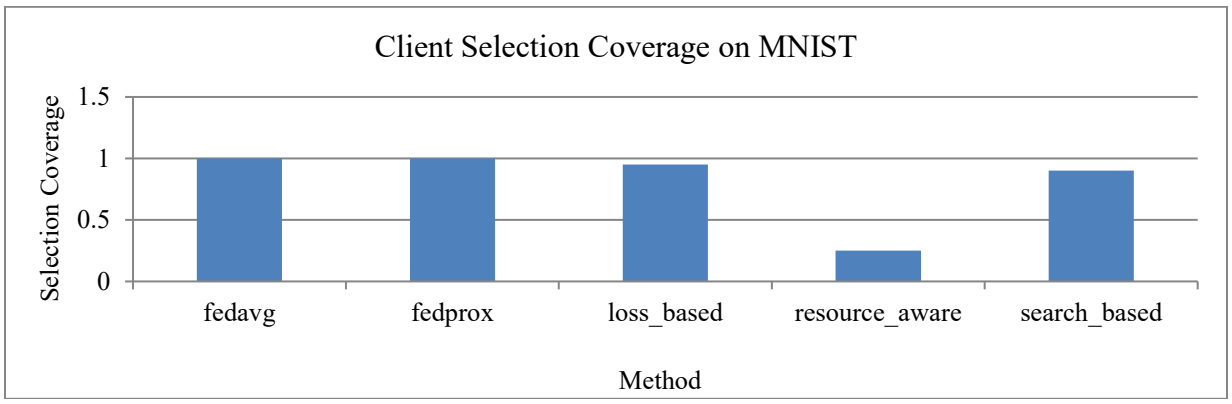

Fig. 8 Client selection coverage achieved by each method

4.3. Ablation Analysis

An ablation analysis examines the contribution of the components of the proposed method. The complete framework is compared with three reduced configurations: (i) without the participation-history term, obtained by setting its coefficient to zero; (ii) without searched aggregation weighting, where only data-size-based weights were used for aggregation; and (iii) a performance-only fitness configuration, in which the disparity, resource-cost, update-divergence, and participation terms are removed from the fitness function. The comparison, therefore, examines the individual contribution of participation awareness and search weighting, as well as the overall benefit of the multi-criteria fitness formulation.

Table 4. Ablation analysis of the proposed method.

Configuration	Accuracy	Macro-F1	Loss	CDG	Coverage	Entropy	Stability Var.
Proposed search-based	0.31	0.22	2.14	0.60	0.90	0.88	0.0340
Without participation	0.26	0.16	2.28	0.65	0.90	0.84	0.0275
Without searched weight	0.17	0.08	2.50	0.56	0.90	0.89	0.0094
Performance only	0.28	0.14	2.44	0.66	0.80	0.80	0.0089

Table 4 shows that the search-based method achieves the highest performance and the lowest loss. After removing the participation component ($\eta = 0$), accuracy and macro-F1 decrease to 0.26 and 0.16, respectively, and the CDG increases from 0.60 to 0.65. Although the selection coverage remains unchanged at 0.90, the entropy decreases from 0.88 to 0.84. Thus, the participation-history term (B) has contributed to search-procedure performance. A higher degradation is seen when the searched weight is removed in the without-searched-weight configuration ($\rho = 0$). Here, accuracy decreases to 0.17, macro-F1 to 0.08, and loss increases to 2.50. This performance drop shows that searched weights contribute significantly to the predictive utility. This configuration also achieves a lower CDG of 0.56, higher entropy of 0.89, and lower stability variance. The performance-only configuration achieves a competitive accuracy of 0.28 but a lower macro-F1 of 0.14. It has the lowest coverage (0.80) and entropy (0.80) and the highest CDG (0.66). Therefore, only performance-based fitness alone does not yield better or similar results to multi-criteria fitness. These results support using participation awareness, search weight, and multi-criteria fitness. At the same time, the lower CDG and stability variance in some reduced configurations suggest that the benefit is primarily reflected in predictive performance and broader client participation.

4.4. Sensitivity Analysis of Fitness Weights and Aggregation Ratio

The sensitivity analysis examines how the proposed method responds to changes in two important design parameters: (1) the predictive utility (α) and the client disparity (β) in the fitness function, and (2) the contribution of the searched weights. For the fitness-function analysis, three settings are evaluated by varying the predictive-utility coefficient and disparity penalty while keeping the remaining fitness coefficients unchanged. For the aggregation analysis, the searched-weight ratio (ρ) was evaluated at 0.00, 0.30, and 0.50.

Table 5. Sensitivity Analysis of Fitness-weight settings

Fitness setting	α	β	Accuracy	Macro-F1	Loss	CDG	Coverage	Entropy
Higher disparity	0.5	0.3	0.21	0.08	2.48	0.65	0.95	0.90
Current setting	0.6	0.2	0.31	0.22	2.14	0.60	0.90	0.88
Higher performance	0.7	0.1	0.18	0.10	2.64	0.51	0.75	0.80

Table 5 shows that the current fitness coefficients (α=0.60, β=0.20) provide better predictive results. It achieves an accuracy of 0.31, a macro-F1 of 0.22, and the lowest loss of 2.14. In a higher disparity setting, the disparity coefficient is increased to β=0.30. This fitness setting did not reduce the observed CDG; instead, it increased to 0.65, while accuracy and macro-F1 decreased. However, the higher disparity setting achieves the highest selection coverage and entropy. Conversely, in the higher-performance fitness setting, the predictive-utility coefficient increased to 0.70, reducing CDG to 0.51. However, this reduction is accompanied by lower accuracy (0.18), macro-F1 (0.10), coverage (0.75), and entropy (0.80), together with the highest loss of 2.64. Thus, the search-based method fitness coefficients provide the most favorable predictive performance. They also retain relatively high participation coverage and entropy. The results support the current multi-criteria fitness coefficients as a performance-oriented configuration, but not as a globally optimal coefficient combination.

Table 6 shows the effect of the aggregation ratio (ρ) on the performance of the search based method. When ρ=0, the searched weights do not contribute to aggregation. With this data-size-only aggregation, predictive performance decreases, with an accuracy of 0.17 and a macro-F1 of 0.08. For ρ=0.30, the accuracy increases to 0.31 and macro-F1 to 0.22, while loss decreases from 2.50 to 2.14. Further, increasing the searched weight coefficient provides no

additional predictive benefit. Instead, with ρ=0.50, accuracy decreases to 0.21 and macro-F1 to 0.08, with loss returning to 2.50. This shows the non-monotonic nature of the searched weight in the aggregation. Thus, increasing the searched weight coefficient does not necessarily improve performance.

Table 6. Sensitivity Analysis of the Aggregation Ratio

Aggregation setting	ρ	Accuracy	Macro-F1	Loss	CDG	Coverage	Entropy
Data-size only	0.00	0.17	0.08	2.50	0.56	0.90	0.89
Proposed ratio	0.30	0.31	0.22	2.14	0.60	0.90	0.88
Equal contribution	0.50	0.21	0.08	2.50	0.66	0.85	0.85

For the client behavior (as in Table 6), the data-size-only setting obtains the lowest CDG (0.56) and slightly higher entropy (0.89). At the same time, the adopted ratio ρ=0.30 maintains the same coverage (0.90) with a CDG of 0.60 and entropy of 0.88. Moreover, at ρ=0.50, CDG increases to 0.66 and both coverage and entropy decrease. Therefore, ρ=0.30 provides the best predictive performance while maintaining high participation coverage and entropy. These findings support using the 70:30 stabilized weighting adopted in the search-based method.

4.5. Robustness Analysis and Statistical Significance

To assess robustness, the study evaluates all federated strategies over 5 independent runs. For each method, the metrics are reported as the mean and standard deviation for robustness analysis.

To assess the statistical significance, two statistical tests are used: (1) the non-parametric Friedman test, and (2) the paired Wilcoxon signed-rank tests. For both tests, the significance threshold is 0.05.

Table 7. Performance and Robustness across 5 Independent Runs

Method	Accuracy	Macro-F1	Loss	CDG	Coverage	Entropy
FedAvg	0.18 ± 0.06	0.11 ± 0.05	2.23 ± 0.05	0.53 ± 0.08	1.00 ± 0.00	0.98 ± 0.00
FedProx	0.18 ± 0.06	0.11 ± 0.05	2.23 ± 0.05	0.53 ± 0.08	1.00 ± 0.00	0.98 ± 0.00
Loss-based	0.18 ± 0.06	0.08 ± 0.05	2.23 ± 0.04	0.59 ± 0.04	0.98 ± 0.03	0.96 ± 0.01
Resource-aware	0.15 ± 0.03	0.06 ± 0.02	2.50 ± 0.10	0.57 ± 0.11	0.25 ± 0.00	0.54 ± 0.00
Search-based	0.20 ± 0.08	0.08 ± 0.05	2.38 ± 0.13	0.67 ± 0.07	0.84 ± 0.10	0.86 ± 0.02

Table 7 summarizes the robustness of the compared methods. The search-based method achieves the highest mean accuracy (0.20) but also the largest standard deviation (0.08). Thus, it shows more run-to-run variability in predictive performance. For macro-F1, FedAvg and FedProx achieve the highest mean value (0.11). The search-based method also shows a higher mean loss (2.38) than FedAvg, FedProx, and loss-based selection. Therefore, the proposed method's predictive advantage is not consistent across all independent runs. These differences are also evident in client-level and participation behavior. The search-based method achieves a mean CDG of 0.67, higher than the other strategies. Thus, the client-level disparity is not consistently minimized. In addition, the mean selection coverage of 0.84 and entropy of 0.86 show more distributed participation than the resource-aware strategy. In contrast, FedAvg and FedProx achieve complete coverage and the highest entropy. This is mainly because their participation mechanism distributes selection across clients. The search-based method maintains better mean accuracy, but its performance and client disparity are sensitive to variation.

Table 8. Statistical Significance of Performance Differences across Methods

Metric	Friedman Test		Wilcoxon Test							
			Search-based vs. FedAvg		Search-based vs. FedProx		Search-based vs. Loss-based		Search-based vs. Resource-aware	
	statistic	*p*-value	W	*p*-value	W	*p*-value	W	*p*-value	W	*p*-value
Accuracy	2.87	0.58	7.00	1.00	7.00	1.00	6.00	0.81	3.00	0.31
Macro-F1	7.39	0.12	5.00	0.63	5.00	0.63	7.00	1.00	2.00	0.19
Loss	11.20	0.02	1.00	0.13	1.00	0.13	1.00	0.13	2.00	0.19

Table 8 summarizes the differences between all baseline methods and the search-based method. For the predictive performance metrics, the Friedman test reports no significant difference for either accuracy or macro-F1. Although the search-based method achieves the highest mean accuracy (in Table 7), the accuracy advantage is not statistically significant. In contrast, the Friedman test reports a significant difference in loss (statistic = 11.20, *p*-value = 0.02), with its p-value below the threshold. Thus, the loss distributions are

not equivalent across all five methods. The Wilcoxon tests show no significant differences across any compared pair. For accuracy, all pairwise p-values exceed the significance threshold. Macro-F1 shows similar statistics and p-values. For loss, the p-values again exceeded the significance threshold. Thus, despite the significant Friedman result for loss, none of the pairwise Wilcoxon p-values reached the significance threshold.

4.6. Overall Discussion

The findings provide a broader view of the search-based method's behavior. In the primary experiment, the search-based strategy achieved the best predictive results. It supports the joint client selection and aggregation weights to improve the global model.

The client-behavior analysis showed that the search-based method maintained broad client participation. However, it did not dominate all client-level criteria, particularly client disparity and stability. Therefore, it achieved improved predictive performance at the cost of other aspects of federated behavior.

The ablation analysis reports the contribution of the fitness function's components with three different configurations. First, without participation awareness, predictive performance and entropy were reduced. At the same time, removing the searched weight resulted in the lowest accuracy and macro-F1. Moreover, restricting the configuration to predictive utility alone reduced the performance and participation behavior. Thus, the search-based method is not attributable to a single component, while the multi-criteria fitness plays complementary roles.

The sensitivity analysis also shows that the search behavior depends on its design parameters. Among the evaluated fitness-weight settings, the adopted $(\alpha,\beta) = (0.60,0.20)$ achieves the best predictive results while maintaining relatively high participation coverage and entropy. Similarly, the aggregation analysis showed that $\rho=0.30$ outperformed both data-size-only aggregation and equal contribution in predictive performance. Further, increasing either the predictive-utility coefficient or the searched-weight contribution did not lead to monotonic improvements.

The repeated-run analysis, however, qualifies the results obtained from the primary experiment. Across 5 independent runs, the Friedman test found no significant difference among the methods in predictive performance, although it identified a significant difference in loss. Similarly, Wilcoxon comparisons showed no significant difference among the federated methods for any of the three predictive metrics. Consequently, the present experiments do not establish statistically significant predictive superiority of the search-based method over the baselines.

Additionally, the findings are consistent with the broader literature reviewed in Section 2. This shows that client selection and aggregation behavior are influenced by data quality, resource conditions, divergence, fairness, and other heterogeneous client characteristics [14], [19], [27], [28]. Unlike these approaches, the search-based method incorporates multiple considerations while jointly searching for client participation and aggregation weights. The ablation and sensitivity analysis results further indicate that this integration affects predictive and participation behavior. However, it does not improve all evaluation criteria simultaneously.

4.7. Ethical, Privacy, Fairness, and Societal Considerations

In this study, client-level statistics for client profiling are computed centrally, whereas they should be computed locally. The framework does not incorporate formal privacy mechanisms. Thus, the resistance to information leakage from model updates is outside the scope of the present study. However, including a privacy mechanism is an important direction for future investigation.

Fairness refers to differences in predictive performance and participation across the federated clients. In the study, several measures for client-level disparity and participation are used. They include the client-disparity gap, selection coverage, and entropy. They report performance differences and the distribution of client participation. However, they provide no measure of demographic fairness.

The presented search-based method may introduce selection bias due to its fitness function. Here, clients favored by the fitness function may get selected more frequently for global updates. In particular, the higher emphasis on predictive utility could favor clients whose local data aligns with the global model. However, the historical participation component partially addresses this issue by rewarding less frequently selected clients. Nevertheless, the historical participation alone does not eliminate biases arising from heterogeneous local data distributions.

The framework has potential societal relevance in high-stakes applications. In these settings, a client can be an individual, institution, device, or population with heterogeneous data distributions and resource constraints. Thus, client selection determines which data sources contribute to the global model and how much they contribute. The present experiments use benchmark image data and simulated resource constraints. Therefore, the observed participation behavior and client disparity cannot be interpreted as evidence of societal fairness in real-world applications. In particular, participation-aware selection does not ensure equitable client-level outcomes, as the results show. Future work should therefore evaluate the framework using federated datasets from various domains.

5. Conclusion

This paper investigated the effect of participation-aware search-based client selection and adaptive aggregation in federated learning under a non-IID setting. The main objective was to examine whether a multi-criteria client participation strategy improves federated training. It was tested with only a subset of clients participating in each communication round. The search-based strategy was compared with standard federated baselines and heuristic client-selection methods.

The primary experiment results show that the search-based method achieved the strongest final predictive performance. It has the highest final accuracy and macro-F1 score, and the lowest final test loss. Since all methods used the same communication configuration, the improvement is attributable to differences in client selection and aggregation behavior rather than to communication budget. The findings show that the improved prediction performance does not reduce client-level disparity. Therefore, the proposed method is interpreted as a performance-oriented, participation-aware search strategy.

Furthermore, different analyses reported the various aspects and the significance of the search-based method. The ablation analysis shows how the individual fitness components affect client predictive performance and participation behavior. The sensitivity analysis revealed the trade-offs between predictive performance and client disparity. It reported the positive influence of using search weights in the weight aggregation.

However, the search-based method did not perform significantly differently from the baselines under the Friedman and Wilcoxon tests. Future work should evaluate the proposed method with stricter disparity-control mechanisms, on real-world datasets, and across various configuration combinations.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

Not applicable.

Acknowledgments

The authors acknowledge the use of an AI-assisted tool to support language editing and manuscript preparation.

References

[1] Brendan McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, PMLR, vol. 54, pp. 1273-51282, 2017. [Google Scholar] [Publisher Link]

[2] Peter Kairouz, and H. Brendan McMahan, "Advances and Open Problems in Federated Learning," *Foundations and trends in machine learning*, vol. 14, no. 1-2, pp. 1-210, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Tian Li et al., "Federated Optimization in Heterogeneous Networks," *Proceedings of Machine Learning and Systems*, 2020. [Google Scholar] [Publisher Link]

[4] Hangyu Zhu et al., "Federated Learning on Non-IID Data: A Survey," *Neurocomputing*, vol. 465, pp. 371-390, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Takayuki Nishio, and Ryo Yonetani, "Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge," *ICC 2019 – 2019 IEEE International Conference on Communications (ICC)*, Shanghai, China, pp. 1-7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jian Li, Tongbao Chen, and Shaohua Teng, "A Comprehensive Survey on Client Selection Strategies in Federated Learning," *Computer Networks*, vol. 251, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Shanika Iroshi Nanayakkara, Shiva Raj Pokhrel, and Gang Li, "Understanding Global Aggregation and Optimization of Federated Learning," *Future Generation Computer Systems*, vol. 159, pp. 114-133, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Dongseok Kang, and Chang Wook Ahn, "GA Approach to Optimize Training Client Set in Federated Learning," *IEEE Access*, vol. 11, pp. 85489-85500, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Bingyan Liu et al., "Recent Advances on Federated Learning: A Systematic Survey," *Neurocomputing*, vol. 597, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yashothara Shanmugarasa et al., "A Systematic Review of Federated Learning from Clients' Perspective: Challenges and Solutions," *Artificial Intelligence Review*, vol. 56, S2, pp. 1773-1827, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Madan Baduwal, Priyanka Paudel, and Vini Chaudhary, "Federated Learning: A Survey of Core Challenges, Current Methods, and Opportunities," *Computers*, vol. 15, no. 3, pp. 1-52, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[12] Jianqing Zhang et al., "FedALA: Adaptive Local Aggregation for Personalized Federated Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 9, pp. 11237-11244, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Qiying Pan et al., "Contextual Client Selection for Efficient Federated Learning over Edge Devices," *IEEE Transactions on Mobile Computing*, vol. 23, no. 6, pp. 6538-6548, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Shinan Song et al., "Data Quality-Aware Client Selection in Heterogeneous Federated Learning," *Mathematics*, vol. 12, no. 20, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Tuong Minh Nguyen et al., "FedDSS: A Data-Similarity Approach for Client Selection in Horizontal Federated Learning," *International Journal of Medical Informatics*, vol. 192, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Allan M. de Souza et al., "Adaptive Client Selection with Personalization for Communication Efficient Federated Learning," *Ad Hoc Networks*, vol. 157, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Aissa Hadj Mohamed et al., "CCSF: Clustered Client Selection Framework for Federated Learning in Non-IID Data," *Proceedings of the IEEE/ACM 16th International Conference on Utility and Cloud Computing (UCC '23)*, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Yulan Gao, Yansong Zhao, and Han Yu, "Multi-Tier Client Selection for Mobile Federated Learning Networks," *2023 IEEE International Conference on Multimedia and Expo*, pp. 666-671, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Konglin Zhu et al., "Client Selection for Federated Learning using Combinatorial Multi-Armed Bandit under Long-Term Energy Constraint," *Computer Networks*, vol. 250, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Filipe Maciel et al., "Federated Learning Energy Saving through Client Selection," *Pervasive and Mobile Computing*, vol. 103, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Rana Albelaihi, "Mobility Prediction and Resource-Aware Client Selection for Federated Learning in IoT," *Future Internet*, vol. 17, no. 3, pp. 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Jian Tang et al., "Joint Class-Balanced Client Selection and Bandwidth Allocation for Cost-Efficient Federated Learning in Mobile Edge Computing Networks," *IEEE Transactions on Mobile Computing*, vol. 24, no. 7, pp. 5681-5698, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Liwei Wang et al., "Communication-Efficient Model Aggregation with Layer Divergence Feedback in Federated Learning," *IEEE Communications Letters*, vol. 28, no. 10, pp. 2293-2297, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Pian Qi et al., "Model Aggregation Techniques in Federated Learning: A Comprehensive Survey," *Future Generation Computer Systems*, vol. 150, pp. 272-293, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Mohammad Moshawrab et al., "Reviewing Federated Learning Aggregation Algorithms; Strategies, Contributions, Limitations and Future Perspectives," *Electronics*, vol. 12, no. 10, pp. 1-35, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ahmad Khalil et al., "Label-Aware Aggregation for Improved Federated Learning," *2023 Eighth International Conference on Fog and Mobile Edge Computing (FMEC)*, Tartu, Estonia, pp. 216-223, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Haotian Wu, Jiaming Pei, and Jinhai Li, "FedCW: Client Selection with Adaptive Weight in Heterogeneous Federated Learning," *Computers, Materials and Continua*, vol. 86, no. 1, pp. 1-20, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Ying-Chi Mao et al., "Federated Dynamic Client Selection for Fairness Guarantee in Heterogeneous Edge Computing," *Journal of Computer Science and Technology*, vol. 39, no. 1, pp. 139-158, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Yuxin Shi et al., "Fairness-Aware Client Selection for Federated Learning," *2023 IEEE International Conference on Multimedia and Expo (ICME)*, Brisbane, Australia, pp. 324-329, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[30] Tao Wan et al., "A Secure and Fair Client Selection based on DDPG for Federated Learning," *International Journal of Intelligent Systems*, vol. 2024, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[31] Lina Su et al., "Fair Client Selection for Multi-Task Federated Learning in Mobile Edge Networks," *Information Processing and Management*, vol. 63, no. 6, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[32] Liangyan Li et al., "SCKM: Symmetric Co-Skew Moment for User Selection in Federated Learning," *Entropy*, vol. 28, no. 6, pp. 1-19, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[33] Narisu Cha, and Long Chang, "Multi-Objective Distributed Client Selection in Federated Learning-Assisted Internet of Vehicles," *Sensors*, vol. 24, no. 13, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[34] Shamim Ahmed et al., "Federated Learning Model with Dynamic Scoring-Based Client Selection for Diabetes Diagnosis," *Knowledge-Based Systems*, vol. 320, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[35] Elahe Eslami et al., "The GAO-based Federated Learning Framework with Adaptive Client Selection for Resource-efficient Edge-IoT Systems," *Cluster Computing*, vol. 29, no. 11, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[36] Y. LeCun et al., "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998. [CrossRef] [Google Scholar] [Publisher Link]