PECO: Probabilistic Evaluation-based Client Selection for Federated Learning with Overlapping Clients

Allen Yang*, Shiyue Hou†, Yiming Xie†, Bo Sheng*, and Ningfang Mi†
*Department of Computer Science, University of Massachusetts Boston, Boston, USA
Email: Allen.Yang001@umb.edu, Bo.Sheng@umb.edu
†Department of Electrical and Computer Engineering, Northeastern University, Boston, USA
Email: hou.shiy@northeastern.edu, xie.yimi@northeastern.edu, ningfang@ece.neu.edu

***Abstract*—Federated Learning (FL) enables privacy-preserving distributed machine learning by training models across clients without raw data exchange. However, non-IID data distributions—particularly when clients possess overlapping samples alongside unique data—pose significant challenges for client selection strategies. Traditional random sampling approaches inadequately balance the trade-off between redundant updates from overlapping data and diverse contributions from client-specific samples, resulting in suboptimal model performance. We propose PECO, a dynamic client selection framework that evaluates each client's contribution to model performance using a small validation set and adaptively assigns selection probabilities based on these contributions. Through experiments on CIFAR-10 and Fashion-MNIST, we demonstrate that PECO consistently outperforms random selection, achieving superior model accuracy and convergence stability across diverse non-IID scenarios with overlapping data.** [1]

I. INTRODUCTION

Federated Learning (FL) has emerged as a transformative paradigm for training machine learning models across decentralized data sources while preserving user privacy [1]. By enabling clients—including edge devices, organizations, and individual users—to collaboratively train a global model without sharing raw data, FL addresses critical privacy concerns in domains such as healthcare, finance, and IoT [2]. However, the inherent heterogeneity of client data, which varies in distribution, quality, and quantity [3], creates significant challenges. This heterogeneity manifests as non-independent and identically distributed (non-IID) data, complicating the training of models that generalize effectively across the entire network [4]. Consequently, sophisticated techniques are essential to ensure convergence and maintain high accuracy in federated settings.

In non-IID FL scenarios, client data distributions diverge across feature spaces, label distributions, and sample sizes [5]. While existing research has extensively addressed challenges such as class imbalance and feature skew, a critical yet understudied issue emerges when clients possess overlapping samples—where subsets of data are shared across clients while each retains unique samples. This phenomenon naturally occurs across various domains: hospitals within a region share data on common diseases while treating location-specific rare conditions; autonomous vehicles in the same city encounter similar road conditions but experience route-specific scenarios; users within demographic groups exhibit overlapping preferences while maintaining distinct individual tastes. These overlapping patterns introduce additional complexity to client selection, as redundant updates may disproportionately influence the global model.

Traditional FL client selection strategies, including random selection and uniform sampling, inadequately address this duality. Random selection risks overemphasizing redundant updates from overlapping clients, while excluding these clients entirely discards their valuable unique contributions. Our experiments reveal that extreme strategies—always including or excluding overlapping clients—consistently degrade model performance. To our knowledge, existing client selection schemes (e.g., [6], [7]) have not explicitly addressed scenarios involving overlapping clients.

This paper tackles the client selection problem in FL with overlapping clients. We first demonstrate that existing methods fail to balance leveraging shared knowledge from overlapping samples with preserving insights from client-specific data. To address this gap, we propose PECO, a novel framework that dynamically evaluates client updates after each training round and assigns selection probabilities based on their contributions to the global model. PECO employs a small validation set to assess each client's impact on model performance, adaptively adjusting selection probabilities to enhance convergence and accuracy while mitigating the adverse effects of data overlap. Our empirical evaluations on CIFAR-10 and Fashion-MNIST demonstrate that PECO significantly outperforms traditional random selection in both model accuracy and training stability.

The remainder of this paper is organized as follows: Section II reviews related work on federated learning client selection strategies. Section III presents our proposed probabilistic client selection approach. Section IV details our experimental results and analysis. Section V concludes with key insights and future research directions.

II. RELATED WORK

Client selection strategies in federated learning have evolved to address fundamental challenges including data heterogene-

[1]This work is partially supported by the National Science Foundation Award OCA-2417715

979-8-3315-6874-0/25/$31.00 ©2025 IEEE

DOI 10.1109/HPCC67675.2025.00113

ity [7] and resource constraints [8]. The literature can be broadly categorized into approaches targeting data distribution challenges and those optimizing selection based on client contributions.

Addressing Data Heterogeneity. Non-IID data distributions remain a fundamental challenge in federated learning, significantly impacting model convergence and performance. Recent solutions employ partial data sharing strategies to improve model consistency across clients. FedMix [9] generates synthetic data samples to balance distributions, while FedGen [10] creates global knowledge representations to bridge client disparities. For scenarios where certain clients possess exclusive classes, Mavericks Matters [11] adaptively upweights these unique contributions to accelerate convergence.

Importance-Based Selection. These methods prioritize clients based on their expected contribution to model improvement. FedHcs [12] combines cluster sampling, sample size reallocation, and importance sampling to minimize gradient update variance, thereby accelerating convergence. The "Power-of-Choice" strategy [6] demonstrates that selecting clients with higher local loss significantly improves convergence speed compared to unbiased selection, though at the potential cost of introducing solution bias.

Similarity and Diversity-Based Approaches. Maximizing diversity in client updates has proven effective for improving model generalization. FedRep [13] and similar methods select clients whose contributions differ significantly yet complement each other, using cosine similarity or representation-based metrics to quantify differences. This approach reduces gradient conflicts and enhances training stability by ensuring diverse perspectives are incorporated into the global model.

Loss and Gradient-Based Selection. These strategies directly leverage optimization metrics to guide client selection. FedDyn [14] dynamically prioritizes clients with higher local loss values to accelerate global convergence. FedCurv [15] extends this concept by incorporating Hessian information to identify clients that help avoid sharp local optima. Meanwhile, FedCorr [16] focuses on gradient diversity, selecting clients whose updates provide complementary optimization directions for improved generalization. Despite these advances, existing methods do not explicitly address scenarios where clients possess overlapping data samples alongside unique contributions. Our work fills this gap by proposing a probabilistic selection strategy that dynamically balances redundancy from overlapping data with the value of client-specific information.

III. Design of PECO

This section presents PECO, our probabilistic client selection framework for federated learning with overlapping data. We first formalize the system model and problem definition, then describe the overall workflow and detail the key components of our approach.

A. System Model

In this paper, we consider a federated learning system comprising a central FL server and multiple clients. Each client holds data that follows a non-IID distribution, meaning the data samples vary significantly across different clients. Additionally, we introduce the concept of *overlapping clients*, where the datasets of some clients partially overlap with those of other clients while still maintaining unique samples.

For each overlapping client, we define the *overlapping ratio* as the proportion of samples that overlap with other clients' datasets relative to the total number of samples owned by the client. To simplify the problem, we assume that all overlapping clients share the same overlapping ratio.

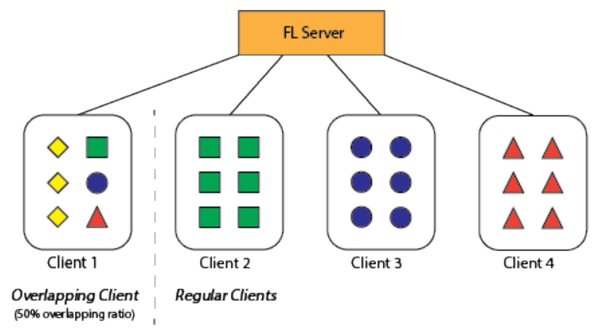


Fig. 1: An example of 1 overlapping client and 3 regular clients

Fig. 1 illustrates an example with one overlapping client and three regular clients. Each shape represents the unique data samples belonging to a specific client. In this scenario, Client 1 is an overlapping client, possessing three unique samples (represented by diamond shapes) and three overlapping samples, each shared with one of the other three clients. Consequently, the overlapping ratio for Client 1 is 50%.

In this work, we focus on the client selection problem in federated learning, specifically taking into account the existence of overlapping clients. We adopt the standard FedAvg aggregation method at the server and exclude communication overhead from our analysis. Our objective is to develop an effective client selection strategy that optimizes model accuracy by appropriately balancing the selection of overlapping and regular clients.

B. Challenges and Motivation

In an FL system, the server selects a subset of clients to participate in each training round. The parameter *fraction fit* denotes the proportion of selected clients. Client selection plays a crucial role, especially when the fraction fit value is low, as the choice of participants significantly influences model convergence and performance.

In our problem setting with overlapping clients, a key challenge is determining how to handle them during client selection. One important question in designing an effective client selection strategy is whether the system should prioritize selecting overlapping clients or regular clients. The answer to this question depends on finding a balance between leveraging shared information from overlapping clients and ensuring diversity by incorporating unique data from regular clients.

Developing an appropriate selection strategy is particularly challenging as several factors, such as the number of overlapping clients, the overlapping ratio, the fraction fit, and even the specific models used in the FL process, affect the overall performance. The details of the evaluation settings will be introduced later in Section IV.

Performance and the optimal selection strategy vary significantly across different workloads. This observation highlights the necessity of a dynamic and adaptive solution that can accommodate diverse application settings. Motivated by these insights, we propose PECO, a client selection approach that adapts based on real-time evaluation of client contributions, ensuring more robust and effective model training across varying FL environments. In the next subsection, we present an overview of our solution, followed by a detailed discussion of its key components.

C. *Overview of* PECO

In this paper, we present PECO, a novel client selection strategy designed to address the limitations of random and loss-based selection methods in FL, particularly in heterogeneous settings with overlapping clients. Our solution consists of two main steps, as illustrated in Fig. 2.

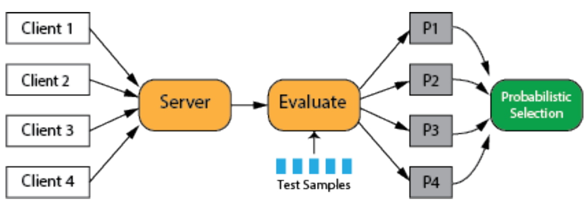


Fig. 2: Architecture of PECO (P1, P2, P3 and P4 are the derived selection probabilities based on per-client evaluation)

First, when the server receives model updates from clients, it evaluates each client's model using a small reserved validation set. This evaluation measures each client's contribution to the global model by assessing how well their updates improve performance on the validation set. Based on these results, the server assigns a selection probability to each client, reflecting their impact on overall model performance

In the second step, the server selects clients for the next round based on their assigned probabilities, ensuring a more informed and adaptive selection process. This step enables the system to dynamically adjust participation rates based on real-time evaluation, preventing over-reliance on redundant updates from overlapping clients while preserving their valuable contributions. This probabilistic framework ensures that client selection remains flexible and responsive to changes in data distribution over multiple training rounds.

Moreover, our method effectively handles overlapping client datasets without requiring explicit prior knowledge, allowing the system to dynamically adapt to varying data distributions. By balancing the contributions of overlapping and regular clients, PECO enhances model convergence, mitigates potential biases, and improves overall training efficiency. In the next subsection, we detail the methodology used to derive selection probabilities based on evaluation results, which serves as the core component of our approach.

D. *Probability Calculation*

PECO determines selection probabilities by evaluating each client's model on a held-out validation set. Since overlapping clients share data samples, their model updates naturally exhibit higher consistency with the global model. By analyzing these evaluation results, the server can implicitly identify overlap patterns without explicit knowledge of data distributions, assigning probabilities that favor consistent contributions while maintaining sufficient diversity.

To ensure stability, we employ a sliding window averaging mechanism over the last $W = 10$ rounds. This window size balances responsiveness to changing client contributions with robustness against transient fluctuations. Smaller windows inadequately smooth performance variations, while larger windows compromise adaptability to genuine distribution shifts.

The probability calculation proceeds as follows: PECO collects classification outputs from each client's model, computes pairwise prediction similarities, and normalizes these scores into selection probabilities. This adaptive approach automatically balances the trade-off between leveraging redundant information and maintaining diversity, resulting in improved convergence stability and model accuracy.

Classification Probabilities: The first step is to obtain the classification probabilities from the inference process. Specifically, for each client's local model, the server evaluates its predictions on the test samples and records the probability distribution over the possible classes. Formally, let $X = [x_1, x_2, \ldots, x_n]$ represent the set of test samples, and let $Y = [y_1, y_2, \ldots, y_m]$ denote the corresponding set of class labels. For each client C_k, after evaluation, the server obtains a two-dimensional matrix $P_k(i, j)$, where each element represents the probability that test sample x_i is classified as label y_j using C_k's local model.

Similarity Measurement: In the second step, the server measures the similarity between clients' local models. Intuitively, overlapping clients are likely to exhibit higher similarity with other clients due to shared data. To quantify this, we compute the pairwise similarity between each pair of clients and sum the similarity scores across all comparisons to represent each client's overall similarity.

When comparing two clients, the server analyzes their evaluation results for each test sample and computes the cosine similarity between their corresponding probability vectors. To further enhance the robustness of the similarity measurements, we introduce an accuracy-based scaling factor. If a client incorrectly classifies a test sample, we reduce the contribution of the similarity score by scaling it with a fractional value of $\gamma \in (0, 1)$. This scaling emphasizes the similarities derived from correct classifications, as accurately classified samples are more indicative of beneficial overlapping data. The final similarity score between two clients' models is determined by summing the cosine similarity values across all test samples.

Let $P_k(i)$ denote the probability vector for test sample x_i obtained from client C_k's model,

$$P_k(i) = \{P_k(i,j) \mid \forall y_j \in Y\}. \tag{1}$$

By considering the scaling factor, we define a coefficient vector as

$$\beta_k(i) = \begin{cases} 1, & \text{if client } C_k \text{ correctly classifies } x_i; \\ \gamma, & \text{otherwise (incorrect classification).} \end{cases} \tag{2}$$

Therefore, the similarity between a pair of clients C_a and C_b is defined as:

$$sim(C_a, C_b) = \sum_{\forall x_i \in X} S_c(P_a(i), P_b(i)) \cdot \beta_a(i) \cdot \beta_b(i), \tag{3}$$

where S_c denotes the cosine similarity function, given by

$$S_c(U,V) = \frac{\sum_i U_i \cdot V_i}{\sqrt{\sum_i U_i^2} \cdot \sqrt{\sum_i V_i^2}}.$$

Then, we use $S(C_k)$ to represent the total similarity score of client C_k:

$$S(C_k) = \sum_{\forall C_j \neq C_k} sim(C_k, C_j). \tag{4}$$

Selection Probability: Finally, the selection probability for each client is computed by normalizing similarity scores with a temperature parameter:

$$p_k = \frac{S(C_k)^\tau}{\sum_{\forall C_j} S(C_j)^\tau}. \tag{5}$$

The exponent τ controls the selection distribution's sharpness—higher values amplify differences between clients, creating stronger preferences for high-similarity clients. Through extensive empirical evaluation across diverse FL settings, we found $\tau = 5$ provides robust performance, effectively balancing between exploitation of high-performing clients and exploration of diverse contributions. After selecting client C_k, we recalculate the remaining probabilities by renormalizing over unselected clients, ensuring fair representation throughout the selection process.

Furthermore, to mitigate large variations in client selection probabilities, we introduce a sliding window mechanism. Rather than relying solely on the probability computed in the current round, the server maintains a window spanning the last W rounds. At each round, the final selection probability is determined as the average of the probability values from the last W rounds:

$$\tilde{p}_k^{(t)} = \frac{1}{W} \sum_{r=t-W+1}^{t} p_k^{(r)}. \tag{6}$$

This smoothed value $\tilde{p}_k^{(t)}$ is then used to select clients for the next round. By applying this sliding window, we stabilize the selection probability, reducing the fluctuations in similarity scores and ensuring more consistent training behavior.

We present the details of our algorithm for client selection in Algorithm 1. Specifically, lines 1–17 describe the three steps for calculating the selection probability. The time complexity of this process is $O(N^2 \cdot n)$, where N is the number of clients and n is the number of test samples. Lines 18-24 further describe the process of selecting K clients based on their computed selection probabilities.

Algorithm 1 Similarity-Based Probabilistic Client Selection

Input: $\mathcal{C}$ Set of clients
X Reserved dataset for evaluation
K Number of clients to select
τ Similarity scaling factor
W Sliding window size
γ Scaling coefficient value

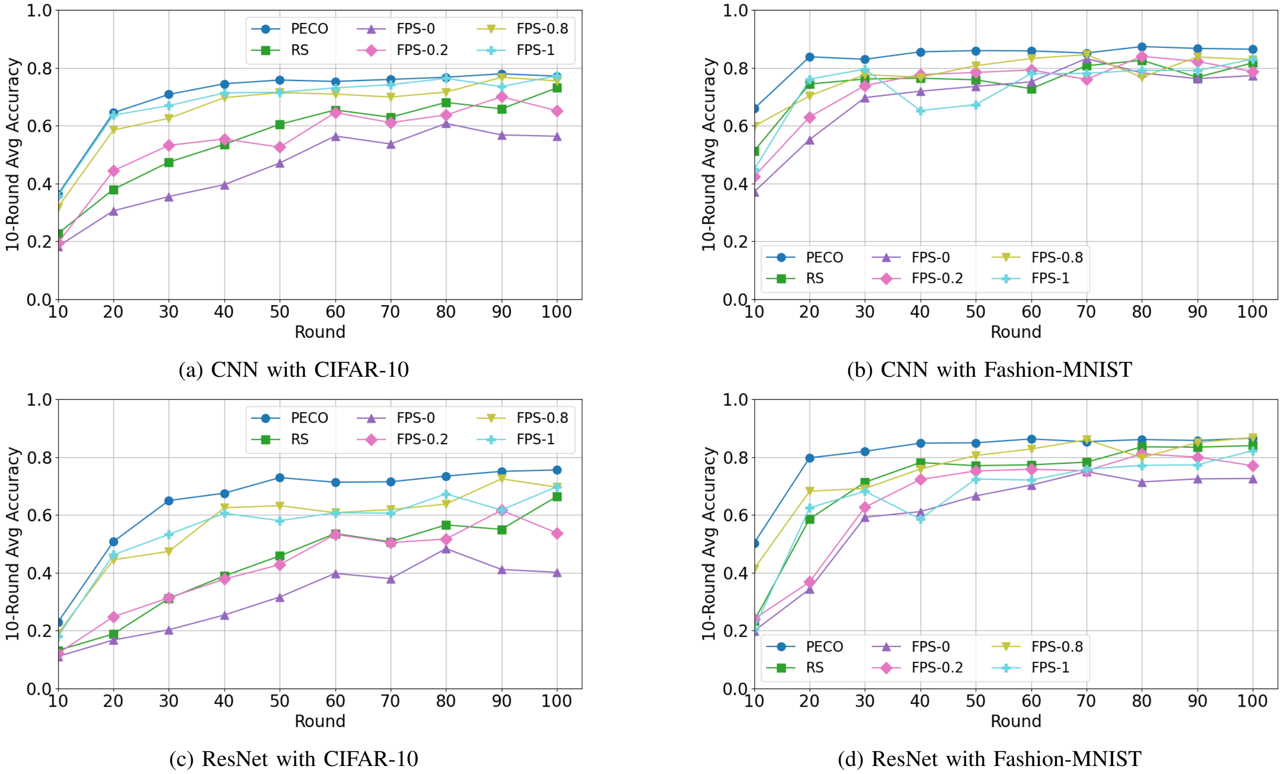
Output: Selected client subset $\mathcal{S}_t$ for the next round

1: $T \leftarrow S_{t-1}$: all the participating clients in the current round
2: **for** each client $C_k \in T$ **do**
3: **for** $x_i \in X$ **do**
4: Evaluate sample x_i with C_k's local model and record probability matrix $P_k(i,j)$ and predicted label
5: **end for**
6: Update the scaling coefficient β_k (Eq. 2)
7: **end for**
8: **for** each pair of clients (C_a, C_b) **do**
9: Compute the similarity $sim(C_a, C_b)$ using Eq. 3
10: **end for**
11: **for** each client $C_k \in \mathcal{C}$ **do**
12: Compute total similarity scores $S(C_k)$ using Eq. 4
13: **end for**
14: **for** each client $C_k \in \mathcal{C}$ **do**
15: Calculate the current round's selection probability p_k using Eq. 5
16: Assign initial selection probabilities $\tilde{p}_k^{(t)}$ using the average selection probabilities in the last W rounds (Eq. 6)
17: **end for**
18: Initialize selected client set: $\mathcal{S}_t \leftarrow \emptyset$
19: **for** selection step $j = 1, 2, \ldots, K$ **do**
20: Sample client C_k based on probabilities $\tilde{p}_k^{(t)}$
21: $\mathcal{S}_t \leftarrow \mathcal{S}_t \cup C_k$
22: Remove client C_k from candidate pool
23: Recalculate probabilities for remaining clients (Eq. 5)
24: **end for**
25: Return $\mathcal{S}_t$

IV. Performance Evaluation

In this section, we present our evaluation results based on the simulation setup described in subsection IV-A.

Baseline Methods for Comparison: We compare our proposed solution, PECO, against the following baselines. *Random Selection (RS)*: This baseline strategy assigns uniform participation probabilities to all clients, without considering client similarity or overlap. Our results highlight its key weaknesses, including instability, slower convergence, and higher update variance due to its disregard for structured data relationships among clients. We use **RS** to denote this solution in the figures.



(a) CNN with CIFAR-10

(b) CNN with Fashion-MNIST

(c) ResNet with CIFAR-10

(d) ResNet with Fashion-MNIST

Fig. 3: Accuracy Performance v.s. Training Round

Fixed Probability Selection (FPS): In this approach, we assume the server is aware of overlapping clients and assigns them a fixed participation probability. We evaluate four different probability values: 0.0, 0.2, 0.8, and 1.0. A probability of 0 or 1 represents extreme cases, where overlapping clients are always excluded or always included, respectively. In contrast, probabilities of 0.2 and 0.8 provide more nuanced selection strategies, biasing toward exclusion or inclusion while still allowing some level of randomness. We use **FPS-x** to represent this alternative with a probability value of **x**.

Performance Metric: Our primary performance metric is accuracy. The evaluation figures plot the number of rounds in FL on the x-axis and accuracy on the y-axis. Each data point in the figures represents the average accuracy over 10 rounds. Specifically, the data value at round 10 reflects the average accuracy from round 1 to round 10, at round 20 from round 11 to round 20, and so forth. This averaging helps smooth out fluctuations and provides a clearer view of model performance over time.

A. Experimental Setup

We implement PECO using the Flower framework (Flwr) [17] and extend *niid_bench* [18] for non-IID client data partitioning. Our federated learning system comprises 20 clients with varying participation fractions (10%, 20%, 30%) per round. We evaluate on CIFAR-10 and Fashion-MNIST datasets using CNN and ResNet18 models. To investigate data overlap impact, we configure overlapping ratios of 0.0, 0.1, and 0.2, where 0.0 indicates no overlap. Non-IID data partitioning uses Dirichlet distribution with concentration parameter α to control heterogeneity. Training hyperparameters are detailed in Table I.

TABLE I: Federated Learning Training Configuration

Parameter	**Value**
Total Clients	20
Dirichlet Distribution Parameter (α)	0.05, **0.10**, 0.15
Fraction Fit	0.1, **0.2**, 0.3
Overlap Ratio	0.1, **0.2**, 0.3
Number of Overlapping Clients	0, 1, **2**, 3, 4
Global Training Rounds	100
Local Training Rounds/Epochs	5
Batch Size / Learning Rate	32 / 0.01
Loss Function / Optimizer	Cross-entropy loss / SGD
Parameters in Algorithm 1	$\tau = 5, \gamma = 0.5, W = 10$

B. Overall Performance

Fig. 3 demonstrates PECO's superior performance across all four workloads (CNN and ResNet with CIFAR-10 and Fashion-MNIST). PECO achieves approximately 70% accuracy by round 30, while baselines require nearly twice as many rounds for comparable performance. The performance advantage is significant in the early training phases (before round 40), highlighting PECO's ability to quickly adapt to client data distribution patterns.

In the following subsections, we analyze the sensitivity of key parameters in our proposed client selection strategy. We have conducted extensive experiments across a wide range of configurations. Specifically, we test both CNN and ResNet models on the CIFAR-10 and Fashion-MNIST datasets to ensure the robustness of our approach in diverse settings. However, due to space constraints, we present a selective subset of these results, focusing on the most representative findings that highlight key trends and insights.

C. Alpha Parameter

This subsection examines how statistical heterogeneity affects PECO's performance. The Dirichlet concentration parameter α controls heterogeneity, with smaller values creating highly non-IID distributions.

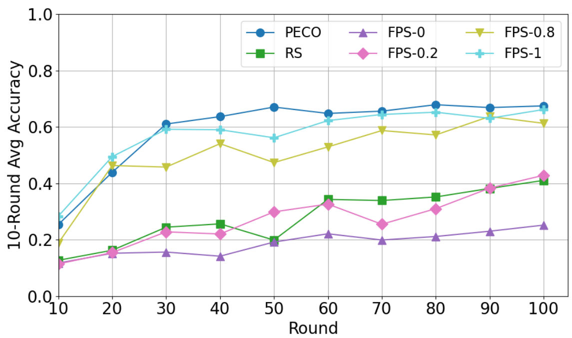


Fig. 4: CNN with CIFAR-10 ($\alpha = 0.05$)

Fig. 4 shows results for $\alpha = 0.05$, where all methods show lower accuracy values due to increased data skew. While RS, FPS-0, and FPS-0.2 struggle to reach 40% accuracy, PECO maintains superior performance with a more pronounced gap in mid-to-late rounds (40-100). Fig. 5 demonstrates PECO's consistent advantage across different α values during early training (rounds 21 - 30). At $\alpha = 0.05$, PECO achieves approximately 60% accuracy while RS reaches only 25% - a 2x improvement in early convergence.

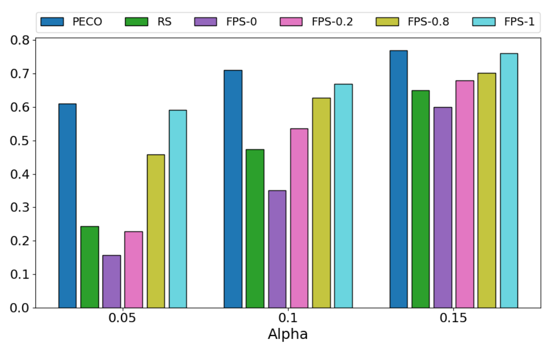


Fig. 5: Average accuracy of rounds 21∼30 with varying α (CNN with CIFAR-10)

D. Overlap Ratio

The overlap ratio determines the amount of redundant information that is overlapped between clients, which is a key factor that PECO needs to manage effectively. Fig. 6 presents the results from an experiment with an overlap ratio of 0.3. Again, PECO demonstrates consistent performance advantages, especially in the early to mid rounds, even with increased overlap. FPS-1 and FPS-0.8 perform relatively well but still lag behind PECO, suggesting that blindly including all overlapping clients is not optimal.

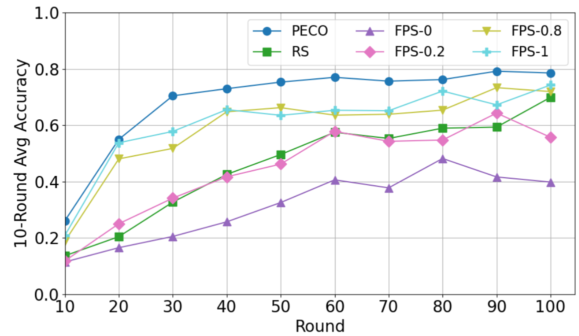


Fig. 6: ResNet with CIFAR-10 (Overlap Ratio = 0.3)

Fig. 7 further compares the early-round performance across different overlap ratio values. Clearly, all solutions except FPS-0 show improved performance as the overlap ratio increases. PECO consistently outperforms the other alternatives, demonstrating the most significant accuracy gains. Its sustained strong performance compared to fixed probability selection underscores the value of its adaptive, context-aware approach to client selection, regardless of the overlap degree.

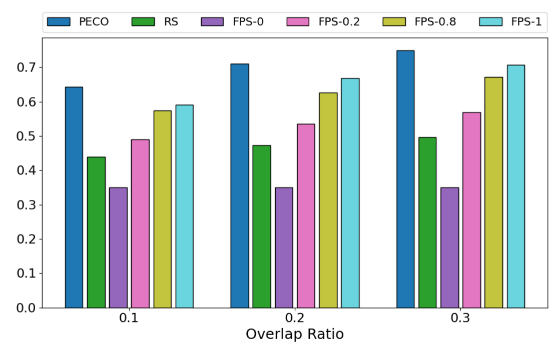


Fig. 7: Average accuracy of rounds 21∼30 with varying overlap ratio (CNN with CIFAR-10)

E. Fraction Fit

Fraction fit represents the proportion of clients selected to participate in each training round. In this subsection, we analyze the impact of varying fraction fit values on model accuracy and convergence speed. Fig. 8 shows the performance when using a smaller fraction fit value (0.1). As the fraction fit decreases, the importance of the client selection strategy increases. While PECO consistently demonstrates a clear improvement over random selection, the performance gap becomes more significant with a fraction fit of 0.1.

Similarly, we compare the early-round performance across fraction fit values of 0.1, 0.2, and 0.3 in Fig. 9. Across all

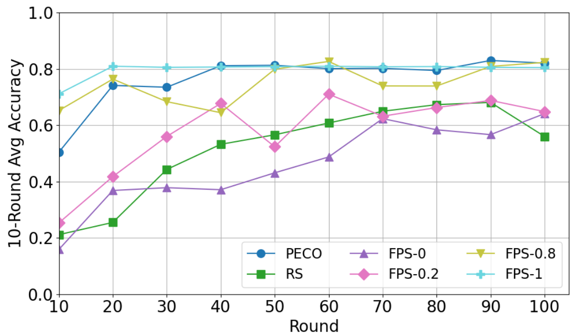


Fig. 8: ResNet with Fashion-MNIST (Fraction Fit = 0.1)

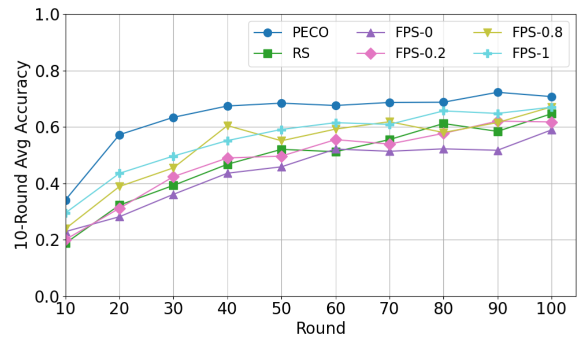


Fig. 10: CNN with CIFAR-10 (1 Overlapping Client)

settings, PECO consistently outperforms the other approaches. Notably, as the fraction fit increases, all methods generally show an improvement in accuracy. However, the extent of improvement varies among strategies. The baseline solutions that prioritize the inclusion of overlapping clients tend to achieve better performance but still fall slightly short of PECO. The results further validate PECO's ability to dynamically optimize client selection, leading to superior performance in federated learning scenarios with overlapping clients.

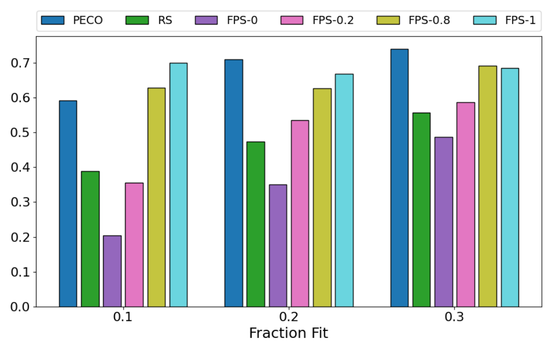


Fig. 9: Average accuracy of rounds 21∼30 with varying fraction fit (CNN with CIFAR-10)

F. Number of Overlapping Clients

Finally, we evaluate the impact of overlapping clients on model performance by varying the number of overlapping clients from 0 to 4. Fig. 10 shows that with only one overlapping client, PECO outperforms all other approaches across all training rounds, achieving around 70% accuracy by round 100. Notably, with only one overlapping client, alternative methods converge to similar performance levels despite different selection strategies. This is particularly interesting for the extreme solutions FPS-0 and FPS-1, which always exclude and include the overlapping client, respectively, yet yield comparable results. PECO's superior performance highlights its effectiveness in appropriately selecting both the overlapping client and regular clients based on contextual factors, rather than following fixed probabilistic rules, demonstrating that adaptive client selection strategies are crucial for federated learning with heterogeneous data distributions.

Fig. 11 compares accuracy in early training rounds. PECO demonstrates superior performance across all scenarios, maintaining a significant lead even without overlapping clients. This indicates that PECO effectively manages not only overlapping clients but also regular clients through its intelligent selection mechanism. As overlapping clients increase, both FPS-0.8 and FPS-1 show substantial improvement while FPS-0 and FPS-0.2 exhibit minimal improvement, reinforcing the limitations of strategies that frequently exclude overlapping clients. Random selection (RS) shows moderate improvement but remains consistently behind PECO. These patterns highlight that including more overlapping clients generally enhances performance for most methods, while PECO's adaptive selection strategy provides a distinct advantage.

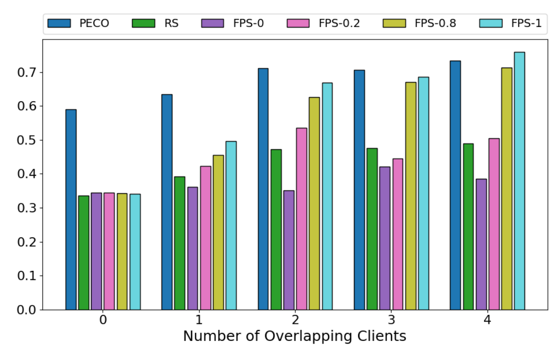


Fig. 11: Average accuracy of rounds 21∼30 with varying number of overlapping clients (CNN with CIFAR-10)

G. Other Comparisons and Limitations

We also compare PECO with an alternative approach that selects clients based on their model's loss value in each round, prioritizing those with the highest losses for next training round. This method represents a broader category of strategies in the literature [14]–[16] that leverage different types of available information. Table II represents a comparison of accuracy at rounds 30, 40, and 50 across different workloads. PECO consistently outperforms the loss-based approach in all settings, with more noticeable performance gaps in ResNet-based configurations where PECO maintains a stronger advantage as training progresses.

TABLE II: Comparison of PECO and Loss-based Solution: Average accuracy at round 30, 40, and 50 with the default parameter setting and different workloads

	CNN/CIFAR-10			CNN/Fashion-MNIST			ResNet/CIFAR-10			ResNet/Fashion-MNIST		
	R30	R40	R50	R30	R40	R50	R30	R40	R50	R30	R40	R50
PECO	0.72	0.74	0.76	0.83	0.85	0.85	0.65	0.71	0.73	0.83	0.85	0.85
Loss-based	0.71	0.71	0.73	0.80	0.81	0.83	0.61	0.66	0.65	0.76	0.81	0.81

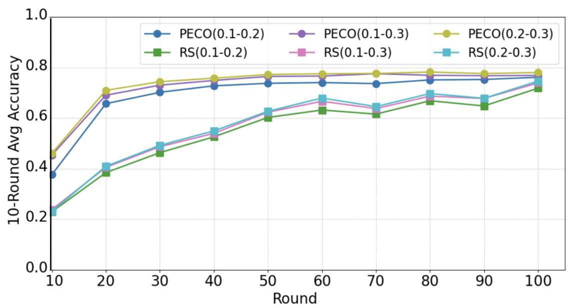


Fig. 12: Two overlapping clients with different overlapping ratios (CNN with CIFAR-10)

To evaluate PECO's robustness to heterogeneous overlapping scenarios, we relaxed our simplifying assumption of uniform overlapping ratios. Figure 12 presents results with two overlapping clients having different overlapping ratios: (0.1, 0.2), (0.1, 0.3), and (0.2, 0.3). PECO maintains consistent superiority over random selection (RS) across all combinations, achieving approximately 5-10% higher accuracy throughout training. Notably, PECO demonstrates stable performance regardless of the heterogeneity in overlapping ratios, with all three PECO curves converging to similar accuracy levels ($\sim$0.78–0.79) despite the varying overlap distributions. This robustness suggests that our probabilistic evaluation mechanism effectively adapts to diverse overlapping patterns without requiring uniform overlap assumptions.

A potential limitation of PECO is the computational overhead associated with evaluating sample data for each client's model in every training round. We address this by re-evaluating only those clients who have participated in training during the current round. For future work, we plan to explore integrating loss values into PECO's client selection strategy to further refine its adaptive selection mechanism, experimenting with varying number of evaluation samples to analyze the tradeoff between computational cost and model accuracy, and developing strategies for selecting evaluation subsets based on factors such as historical contribution to model updates or computational resources.

V. Conclusion

We propose PECO, a probabilistic FL client selection strategy that balances exploration and exploitation. By evaluating client updates on a small sample set, PECO adaptively adjusts selection probabilities, improving model convergence with non-IID and overlapping data. Extensive experiments on CIFAR-10 and Fashion-MNIST (CNN, ResNet) show PECO consistently outperforms baselines and loss-based methods in accuracy and adaptability, demonstrating a promising path for efficient and adaptive FL.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 20–22 Apr 2017.

[2] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," 2018. [Online]. Available: https://arxiv.org/abs/1806.00582

[3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, p. 50–60, May 2020. [Online]. Available: http://dx.doi.org/10.1109/MSP.2020.2975749

[4] Q. Li, Y. Diao, Q. Chen, and B. He. (2021) Federated learning on non-iid data silos: An experimental study. [Online]. Available: https://arxiv.org/abs/2102.02079

[5] T.-M. H. Hsu, H. Qi, and M. Brown, "Measuring the effects of non-identical data distribution for federated visual classification," 2019. [Online]. Available: https://arxiv.org/abs/1909.06335

[6] Y. Jee Cho, J. Wang, and G. Joshi, "Towards understanding biased client selection in federated learning," in *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, 2022.

[7] F. Lai, X. Zhu, H. V. Madhyastha, and M. Chowdhury, "Oort: Efficient federated learning via guided participant selection," in *15th USENIX Symposium on Operating Systems Design and Implementation (OSDI 21)*, 2021.

[8] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained iot devices," *IEEE Internet of Things Journal*, vol. 9, no. 1, 2022.

[9] T. Yoon, S. Shin, S. J. Hwang, and E. Yang, "Fedmix: Approximation of mixup under mean augmented federated learning," 2021. [Online]. Available: https://arxiv.org/abs/2107.00233

[10] Z. Zhu, J. Hong, and J. Zhou, "Data-free knowledge distillation for heterogeneous federated learning," in *Proceedings of the 38th International Conference on Machine Learning*, 2021.

[11] J. Huang, C. Hong, Y. Liu, L. Y. Chen, and S. Roos, "Maverick matters: Client contribution and selection in federated learning," in *Advances in Knowledge Discovery and Data Mining*, H. Kashima, T. Ide, and W.-C. Peng, Eds. Cham: Springer Nature Switzerland, 2023, pp. 269–282.

[12] G. Shen, D. Gao, D. Song, L. Yang, X. Zhou, S. Pan, W. Lou, and F. Zhou, "Fast heterogeneous federated learning with hybrid client selection," 2022. [Online]. Available: https://arxiv.org/abs/2208.05135

[13] Y.-R. Yang, K. Wang, and W.-J. Li, "Fedrep: A byzantine-robust, communication-efficient and privacy-preserving framework for federated learning," 2023. [Online]. Available: https://arxiv.org/abs/2303.05206

[14] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, and V. Saligrama, "Federated learning based on dynamic regularization," 2021. [Online]. Available: https://arxiv.org/abs/2111.04263

[15] N. Shoham, T. Avidor, A. Keren, N. Israel, D. Benditkis, L. Mor-Yosef, and I. Zeitak, "Overcoming forgetting in federated learning on non-iid data," 2019. [Online]. Available: https://arxiv.org/abs/1910.07796

[16] J. Xu, Z. Chen, T. Q. S. Quek, and K. F. E. Chong, "Fedcorr: Multi-stage federated learning for label noise correction," 2022. [Online]. Available: https://arxiv.org/abs/2204.04677

[17] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, H. L. Kwing, T. Parcollet, P. P. d. Gusmão, and N. D. Lane, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390*, 2020.

[18] niid_bench: Benchmark for Non-IID Data in Federated Learning. [Online]. Available: https://github.com/adap/flower/tree/main/baselines/niid_bench