FAST: Federated Active Learning with Foundation Models for Communication-efficient Sampling and Training

Anonymous Authors[1]

Abstract

Federated Active Learning (FAL) has emerged as a promising framework to leverage large quantities of unlabeled data across distributed clients while preserving data privacy. However, real-world deployments remain limited by high annotation costs and communication-intensive sampling processes, particularly in a cross-silo setting, when clients possess substantial local datasets. This paper addresses the crucial question: *What is the best practice to reduce communication costs in human-in-the-loop learning with minimal annotator effort?* Existing FAL methods typically rely on iterative annotation processes that separate active sampling from federated updates, leading to multiple rounds of expensive communication and annotation. In response, we introduce **FAST**, a two-pass FAL framework that harnesses foundation models for weak labeling in a preliminary pass, followed by a refinement pass focused exclusively on the most uncertain samples. By leveraging representation knowledge from foundation models and integrating refinement steps into a streamlined workflow, **FAST** substantially reduces the overhead incurred by iterative active sampling. Extensive experiments on diverse medical and natural image benchmarks demonstrate that **FAST** outperforms existing FAL methods by an average of 4.36% while reducing communication rounds eightfold under a limited 5% labeling budget.

1. Introduction

Federated Learning (FL) emerges as a key decentralized paradigm that enables edge clients (e.g., institutions or devices) to collaboratively train the unified model through global aggregation without compromising local data privacy (McMahan et al., 2017; Li et al., 2020; Kairouz et al., 2021). In recent research, many FL approaches have been developed under the supervised learning setting, assuming that all training data on clients are fully annotated. However, in realistic scenarios, data are typically unlabeled, with only a very limited number of annotated instances. For instance, in the cross-silo scenario, a few organizations possess substantial datasets but face constraints in large-scale data annotation due to limited budgets, expertise, or time (Kairouz et al., 2021; Liu et al., 2022).

To tackle this challenge, recent studies (Deng et al., 2022; Cao et al., 2023; Kim et al., 2023; Ahn et al., 2024; Chen et al., 2024) delve into the concept of federated active learning (FAL) which incorporate the active learning (AL) into the context of FL. AL aims to maximize model performance in situations with scarce labeled data and limited annotation budgets. It achieves this by iteratively selecting the most informative data instances for labeling by an oracle (i.e., a human annotator) based on specific query strategies. FAL bridges these two fields by incorporating active sampling steps during federated training rounds. Specifically, each client independently conducts active sampling on its local data, utilizing either the local model or the aggregated global model as a query selector to identify informative instances prior to local updates (Chen et al., 2024; Kim et al., 2023; Ahn et al., 2024). After each AL iteration, local models are aggregated on the server to form a global model that can guide subsequent query selections.

Recent advances in FAL have demonstrated significant benefits of AL in harnessing unlabeled data within the FL systems. While numerous studies have been proposed to address challenges posed by data heterogeneity in federated settings (Cao et al., 2023; Kim et al., 2023), prior research has paid little attention to the additional communication costs incurred during federated active sampling. One major concern arises from the communication overhead caused by iterative local training on the updated labeled dataset during active sampling. This concern is particularly acute in cross-silo scenarios (Kairouz et al., 2021), where each edge device (e.g., institution) holds a significant amount of data and requires extra communication support to achieve

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

subpar global performance. Moreover, annotation costs in FL are inherently more expensive than in centralized AL due to the distributed and fragmented nature of the data across multiple clients, which requires further coordination and resource allocation.

In our work, we aim to reduce the communication overhead during the FAL process with a limited annotation budget while achieving superior overall prediction performance of the global model. A critical challenge in FAL is the selection of a query selector for active sampling. (Ahn et al., 2024) investigate the discrepancy of utilizing a global or local-only model for active sampling and achieve robust performance by solely applying sampling strategies with the global model on IID data distribution. Nevertheless, (Chen et al., 2024; Cao et al., 2023) prove that the superiority of the query model depends on the heterogeneity of data distribution on the clients. Despite the advancement in exploring the utilization of query models, these methods all require sufficient active training rounds to iteratively improve the generalizability of client models as the feature extractor for selecting informative unlabeled samples. Instead of training the query model from scratch with the initial data pool from random sampling, we seek the applicability of foundation models in enhancing active sampling throughout the federated training process. Notably, previous research (Radford et al., 2021; Zhai et al., 2023; Sun et al., 2023; Oquab et al., 2023) on foundation models show that features learned from the foundation models are semantically organized in the representation space, providing robust and informative embeddings for downstream tasks.

Motivated by this, we introduce a two-pass **F**ederated **A**ctive learning framework with foundation models for communication-efficient **S**ampling and **T**raining, named **FAST**. In the initial pass, we leverage a frozen image encoder from a Vision-Language foundation model (e.g., SigLIP (Zhai et al., 2023)) to perform weak labeling by selecting and prioritizing informative samples based on uncertainty estimates. This preliminary phase utilizes the semantic richness of foundation models to efficiently identify candidate data points for annotation. In the second pass, human oracles refine these weak labels to ensure labeling quality while operating under a limited labeling budget, thereby reducing communication overhead and minimizing the required human effort in the active sampling process. Our contributions are summarized as follows:

- We investigate a challenging FAL scenario in which human annotation is costly and communication support is constrained, necessitating efficient strategies for both labeling and training.
- We propose a two-pass FAL framework to effectively utilize unlabeled data with minimal human intervention, achieving strong performance in a resource-efficient manner.
- We conduct extensive experiments on diverse benchmark datasets, covering both medical and natural images. Our results demonstrate that the proposed method outperforms existing approaches across various data distributions while reducing the required communication rounds by eightfold (8x) under only a 5% labeling budget.

2. Related Work

2.1. Weakly Supervised Learning

Weakly supervised learning (WSL) addresses scenarios where large portions of ground-truth labels are unavailable or limited. Based on the confidence of label availability, WSL is commonly divided into three paradigms: incomplete supervision, inexact supervision, and inaccurate supervision (Zhou, 2018; Ren et al., 2023). Incomplete supervision involves abundant unlabeled instances and only a small subset of labeled data. This setting is often tackled either through active sampling (i.e., human intervention) or by exploiting semi-supervised learning with clustering or manifold assumptions (Dempster et al., 1977; Li et al., 2013; Li & Zhou, 2014). Inexact supervision arises when only coarse-grained labels are provided, necessitating fine-grained instance-level identification via multi-instance learning algorithms (Settles et al., 2007; Wei & Zhou, 2016; Wei et al., 2016). Lastly, inaccurate supervision denotes the presence of label noise (Frénay & Verleysen, 2013), which is typically mitigated through label correction (Yi & Wu, 2019; Zheng et al., 2021; Wu et al., 2021) or regularization-based robust training (Patrini et al., 2017; Hendrycks et al., 2018; Wang et al., 2019; Lukasik et al., 2020). In this work, we focus on the incomplete supervision paradigm in the FL setting, where local datasets are largely unlabeled and distributed across multiple clients with minimal human intervention.

2.2. Active learning

Existing research in AL generally focuses on querying oracles to label the most informative data points, thereby minimizing labeling effort while maximizing model performance. The AL methods are typically divided into uncertainty-based, representativeness-based, and hybrid strategies. Uncertainty-based methods focus on samples with high aleatoric or epistemic uncertainty (Zhan et al., 2022), using metrics such as entropy, margin, or least confidence (Shannon, 1948; Wang & Shang, 2014; Nguyen et al., 2019). For example, BALD (Houlsby et al., 2011; Gal et al., 2017; Kirsch et al., 2019) seeks points maximizing mutual information between predictions and model parameters, while (Yoo & Kweon, 2019) prioritizes samples expected to produce large errors. Similarly, (Huang et al., 2021) employs Temporal Output Discrepancy to estimate uncertainty by measuring output discrepancies at different optimization steps.

Representativeness-based methods aim to cover diverse regions of the input space to ensure broad decision boundaries. CoreSet (Sener & Savarese, 2017; Geifman & El-Yaniv, 2017; Caramalau et al., 2021) addresses this by solving a k-center problem to create a representative core set. Additionally, clustering-based approaches, such as hierarchical clustering or self-organizing maps (Kutsuna et al., 2012; Citovsky et al., 2021), and set coverage optimization (Urner et al., 2013; Yang et al., 2017), enhance representativeness and reduce redundancy in labeled data. In FL, clients engage in joint training of a global model while independently learning local models that can serve as query selectors. A naive way to adopt classical AL in FL is to apply local query sampling on individual clients. However, this approach faces significant challenges due to heterogeneous data distributions. In particular, local query selectors cannot fully leverage global knowledge, especially under non-IID conditions.

2.3. Federated Active Learning

Recent research has begun to investigate the applicability of AL within FL environments, where the scarcity of labeled client data constitutes a significant bottleneck for FL processes. Preliminary studies have focused on integrating AL into federated training by directly applying existing AL strategies to perform data annotation on client devices (Aussel et al., 2020; Wu et al., 2022; Alfalqi & Bellaiche, 2023; Kong et al., 2023). Nonetheless, conventional AL approaches are not specifically designed for decentralized data annotation, and numerous challenges remain unresolved.

Unlike centralized AL, where the model independently selects samples for querying, FL enables clients to train the model collaboratively. In this context, (Deng et al., 2022) explores the efficacy of global (F-AL) and local-only (S-AL) query selection in FL, revealing that F-AL effectively leverages inter-client collaboration to outperform S-AL. Further research on F-AL has sought to address the heterogeneity inherent in FL. (Cao et al., 2023) introduces a knowledge-aware method (KAFAL) to address the mismatch in sampling goals between local clients and the global model in non-IID federated settings. Similarly, (Kim et al., 2023) proposes an innovative FAL sampling method (LoGo) that combines global and local model benefits to enhance inter-class diversity handling. (Chen et al., 2024) integrates evidential learning with a Dirichlet-based model to handle uncertainty and improve data diversity, providing a robust solution for FAL in medical domains with domain shifts.

Despite these advancements, communication overhead remains a core bottleneck for FAL. Each active sampling round typically involves additional local training and global aggregation steps, leading to high communication costs and substantial annotation efforts—particularly under cross-device FL with potentially millions of clients (Kairouz et al., 2021). By contrast, our method focuses on the annotation process at the initial training stage, requiring only a limited labeling budget. We thus propose a communication-efficient FAL framework, **FAST**, that addresses both uncertainty and diversity in active sampling with minimal human effort.

3. Methodology

3.1. Problem Formulation

Given a federated learning (FL) task involving K clients, where each client k possesses a local dataset D_k stored on its device. The global dataset is the union of all local datasets, denoted as $D = \bigcup_{k=1}^{K} D_k$. The objective of FL is to collaboratively learn a global model by solving the following optimization problem in a distributed manner:

$$\begin{aligned}\min_{w} F(w) &\triangleq \frac{1}{K}\sum_{k=1}^{K}\mathcal{F}_k(w_k)\\ &= \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{(x,y)\sim\mathcal{D}_k}\left[\mathcal{F}_k(w_k;x_k,y_k)\right].\end{aligned} \quad (1)$$

where $w \in \mathbb{R}^d$ represents the global model parameters to be optimized. $F(w)$ is the global loss function aggregating the local losses from all clients, and $\mathcal{F}_k(w_k) = \mathbb{E}_{(x_k,y_k)\sim\mathcal{D}_k}\left[\mathcal{F}_k(w_k;x_k,y_k)\right]$ is the expected risk over data distribution $\mathcal{D}_k$ at client k corresponding to parameter vector w_k. $\mathcal{F}_k(w_k;x_k,y_k)$ denotes the loss incurred by the local model w_k on data sample (x_k,y_k) generated from the local data distribution of client k. In heterogeneous FL, data is distributed across clients in a non-IID manner, i.e., data distribution on each local client is distinct, for clients data $\{D_k, D_j\} \in D$, $\mathcal{D}_k \neq \mathcal{D}_j$.

Previous studies typically (McMahan et al., 2017) solve Eq.1 by iteratively updating the global model through local computations on each client and averaging client updates at the server. At communication round t, the server sends the current global model parameters $w^{(t)}$ to a selected subset of clients $\mathcal{K}_t \subseteq 1,2,\ldots,K$. Each client $k \in \mathcal{K}_t$ initializes its local model with the received parameters, $w_k^{(t)} = w^{(t)}$, and performs τ steps of local stochastic gradient descent (SGD) on its local dataset D_k:

$$w_k^{(t,i+1)} = w_k^{(t,i)} - \eta\nabla\mathcal{F}_k(w_k^{(t,i)};\xi_k^{(t,i)}), \quad (2)$$

where η is the learning rate, $i = 0,1,\ldots,\tau-1$, and $\xi_k^{(t,i)}$ denotes a mini-batch sampled from D_k. After local updates, clients send their updated local models $w_k^{(t)}$

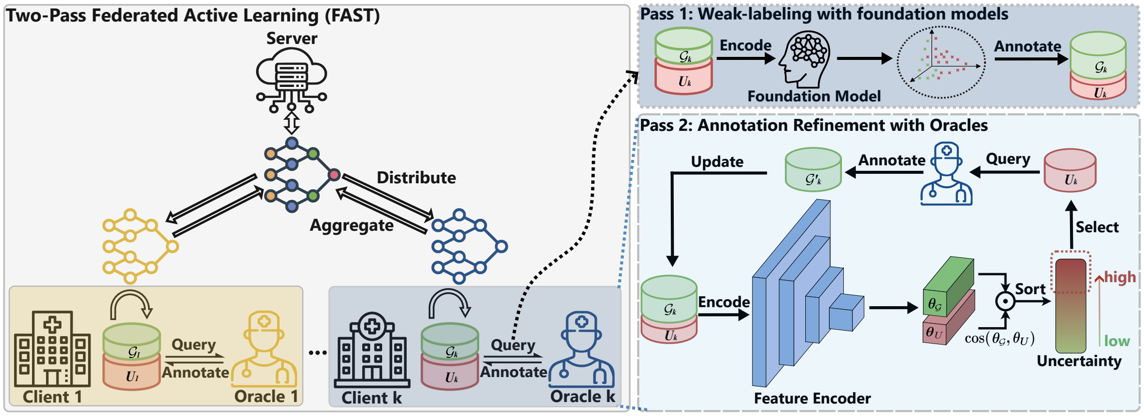


Figure 1: Overview of **FAST**. **FAST** is a communication-efficient FAL framework that employs a two-pass labeling strategy. In the first pass, foundation models perform weak sampling to identify informative data points with minimal communication overhead. In the subsequent pass, human annotators refine the labeled ground truth dataset by validating and correcting the sampled labels, ensuring high-quality annotations.

back to the server. The server aggregates these models by computing an average to update the global model, $w^{(t+1)} = \frac{1}{|\mathcal{K}_t|} \sum_{k \in \mathcal{K}_t} w_k^{(t)}$.

3.2. Federated Active Sampling

AL aims to enhance model performance by iteratively querying and labeling the most informative and representative samples from an unlabeled dataset, under a limited annotation budget. In FAL, this process is adapted to the decentralized setting by executing local active sampling and federated training at each r AL round, $r \subseteq 1, 2, \ldots, R$.

We consider a standard FAL case, where, clients utilize the global model $w^{(t)}$ as the query selector for client-level sampling. During the active sampling phase, each client k selects b unlabeled samples from its local unlabeled dataset $\mathcal{U}_k$ using a predefined query strategy $\mathcal{A}(\cdot)$. At the first AL round, client k randomly selects a small set of b samples for annotation to form the initial labeled ground truth data $\mathcal{G}_k^0$:

$$\mathcal{G}_k^{(0)} = \mathcal{A}(\mathcal{U}_k, b) = \texttt{Random}(\mathcal{U}_k, b), \text{where } \mathcal{G}_k^{(0)} \in \mathcal{U}_k. \tag{3}$$

In subsequent $R-1$ AL rounds, the query strategy $\mathcal{A}(\cdot)$ utilizes the aggregated global model $w^{(r)}$ from the previous round as the query selector to identify informative samples. The selected samples are then labeled and added to the labeled local dataset $\mathcal{G}_k$, while being removed from the unlabeled dataset $\mathcal{U}_k$

$$\mathcal{G}_k^{(r)} \leftarrow \mathcal{G}_k^{(r-1)} \cup \mathcal{A}(w_k^{(r)}; \mathcal{U}_k, b), \quad \mathcal{U}_k \leftarrow \mathcal{U}_k \setminus \mathcal{A}(\mathcal{U}_k, b). \tag{4}$$

The active sampling process continues until the global labeling budget of B is exhausted, ensuring that the total number of labeled samples across all clients does not exceed B.

$$\sum_{k=1}^{K} |\mathcal{G}_k^{(r)}| \leq B, \quad \forall r. \tag{5}$$

After each active sampling step at round r, federated training is performed. Each client k updates its local model $w_k^{(r)}$ by training on the updated labeled dataset $\mathcal{G}_k^{(r)}$, and sends their updated models to the server, which aggregates them to form the new global model w^t as discussed in Eq. 2. Given T federated training rounds, the overall federated rounds across K clients is $R \times T \times K$.

3.3. Two-Pass Federated Labeling

We introduce **FAST**, a communication-efficient federated active learning framework grounded in a two-pass labeling strategy. In the preliminary pass, foundation models (e.g., vision or vision-language) generate preliminary labels based on their representation-based knowledge. This is followed by a refinement pass, where human annotators provide additional annotations to enhance label accuracy and reliability. Unlike previous FAL methods—which rely on iterative cycles of active sampling and federated training and thus incur significant communication overhead—FAST mitigates frequent client-server exchanges, substantially reducing overall communication costs.

In **FAST**, each client k utilizes the frozen encoder from a pre-trained foundation model as a feature extractor $f(\cdot)$ to encode its local dataset D_k into high-dimensional representations: $\mathbf{Z}_k = f(D_k)$, $\mathbf{Z}_k \in \mathbb{R}^d$. Specifically, the unlabeled dataset $\mathcal{U}_k$ and the initial labeled dataset $\mathcal{G}_k^{(0)}$ are encoded:

$$\mathbf{Z}_{\mathcal{U}_k} = f(\mathcal{U}_k), \ \mathbf{Z}_{\mathcal{G}_k} = f(\mathcal{G}_k^{(0)}). \tag{6}$$

To augment the labeled dataset with weak labels for the samples in $\mathcal{U}_k$, we perform label propagation on extracted representation $\mathbf{Z}_{\mathcal{U}_k}$ based on k-nearest neighbors in the embedding space. For each sample x_i in unlabeled dataset $\mathcal{U}_k$, we assign the weak labels based on the majority vote of these neighbors with respect to L_2 distance in the initial labeled dataset $\mathcal{G}_k^{(0)}$. Next, we compute the cosine similarity between the embedding of each weakly labeled sample x_i and the embeddings of all labeled samples in $\mathcal{G}_k^{(0)}$. For each class $c \in C$, we calculate the average cosine similarity $s_{i,c}$ between the embedding $\mathbf{z}_i$ of sample $x_i \in \mathcal{U}_k$ and the embeddings $\mathbf{z}_j$ of all labeled samples $x_j \in \mathcal{G}_{k,c}^{(0)}$:

$$s_{i,c} = \frac{1}{\left|\mathcal{G}_{k,c}^{(0)}\right|} \sum_{x_j \in \mathcal{G}_{k,c}^{(0)}} \frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\|\mathbf{z}_i\| \|\mathbf{z}_j\|}, \quad \forall c \in C \tag{7}$$

where $\mathcal{G}_{k,c}^{(0)}$ denotes the set of initial labeled samples of class c at client k, and and C represents the set of all classes. This process yields a prototype vector $\mathbf{s_i} = [s_{i,1}, s_{i,2}, \ldots, s_{i,C}]$ for each weakly labeled sample x_i. The logits vector represents the average similarity of the sample to each class prototype in the labeled dataset, thereby capturing more nuanced relationships between the weakly labeled samples and the labeled data. We then utilize an uncertainty-based query function $\mathcal{A}(\cdot)$, such as entropy (Wang & Shang, 2014), on the softmax-normalized logits vector $\mathbf{s_i}$ to compute the uncertainty of each weakly labeled sample:

$$u_i = \mathcal{A}(\mathbf{s_i}) = -\sum_{c=1}^{C} \left(\frac{\exp(s_{i,c})}{\sum_{c'=1}^{C}\exp(s_{i,c'})}\right) \log\left(\frac{\exp(s_{i,c})}{\sum_{c''=1}^{C}\exp(s_{i,c''})}\right). \tag{8}$$

Samples with higher uncertainty scores u_i are considered more informative. We rank the samples in $\mathcal{U}_k$ based on their uncertainty scores and select the top b samples for annotation with the given labeling budget in Eq.5. The newly annotated samples are added to the labeled dataset $\mathcal{G}_k$ and removed from the unlabeled dataset $\mathcal{U}_k$, as shown in Eq.4. Subsequently, these human-labeled samples are combined with the weakly labeled samples to form the final labeled dataset for the federated training process, eliminating the need for additional active sampling steps. We summarize the whole procedure of **FAST** approach in Algorithm 1.

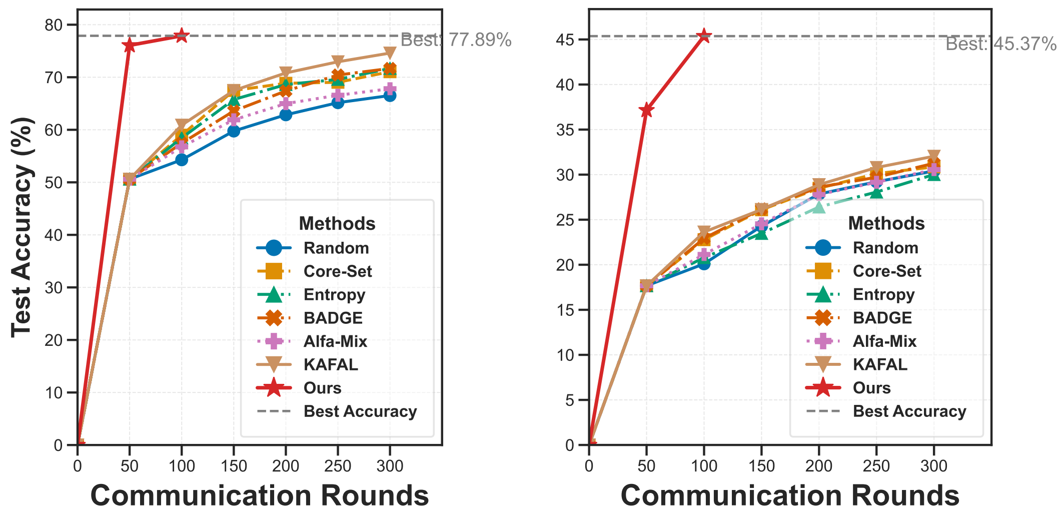
4. Experiments

4.1. Experimental Configuration

Datasets. We evaluate our method primarily on image classification tasks spanning both natural and medical benchmark datasets. Specifically, we use four natural image datasets—CIFAR10/100 (Krizhevsky et al., 2009), Tiny-ImageNet (Le & Yang, 2015), and SVHN (Netzer et al., 2011)—as well as two medical image datasets—PathMNIST and DermaMNIST (Yang et al., 2023). To account for the inherent heterogeneity among clients, we consider three data distribution settings: IID, Non-IID, and heterogeneous inter-class diversity (i.e., variations in local class distributions) (Kim et al., 2023). As shown in Table 2, we report the total labeling budget and the corresponding number of training rounds. Following existing FAL approaches, each client trains its local model from scratch and iteratively selects 5% of the total dataset for annotation in each AL round, until reaching a predefined global labeling budget. We use a 20% labeling budget for most of our experiments and ablation studies. To ensure fairness in labeling costs, we assume this global labeling budget is evenly shared among clients, such that each client queries the same number of samples per AL round.

Baselines. We compare **FAST** with nine standard active learning (AL) strategies: Random, Entropy (Wang & Shang, 2014), Coreset (Sener & Savarese, 2017), BADGE (Ash et al., 2019), LL4AL (Yoo & Kweon, 2019), GCNAL (Caramalau et al., 2021), and ALFA-Mix (Parvaneh et al., 2022). Although originally designed for centralized AL, these strategies can be independently applied on either a global or a local model within a federated environment. In our experiments, we employ the global model as the query selector for active sampling. We further include two federated AL (FAL) strategies, KAFAL (Cao et al., 2023) and LoGo (Kim et al., 2023). For the Non-IID experiment in Figure 2, we select KAFAL as it is specifically tailored for global heterogeneity problems. Regarding the experiment on IID datasets in Table 2, we add LoGo as the baseline considering its focus on solving heterogeneous data from the client level.

Implementation Settings. We implement our proposed **FAST** method in PyTorch using the Flower FL framework (Beutel et al., 2020). As our primary federated learning (FL) strategy, we adopt FedAvg (McMahan et al., 2017), and additionally evaluate on FedProx (Li et al., 2020) and FedNova (Wang et al., 2020) to examine the robustness of **FAST** across different FL paradigms (see Table 5). Our experiments primarily target cross-silo settings with full client participation, involving a total of 10 clients. Each AL round spans $T = 100$ federated communication rounds, and each client executes $\tau = 5$ local stochastic gradient descent (SGD) steps per round. In alignment with prior work (Kim et al., 2023; Cao et al., 2023), we employ a four-layer *CNN* as our main model architecture and employ a *ResNet-8* network for ablation studies on communication efficiency. We simulate the Non-IID data partitions by sampling from a Dirichlet distribution with a concentration parameter of $\alpha = 0.1$, where smaller values of α indicate greater data heterogeneity across clients (Hsu et al., 2019). For the implementation of **FAST**, we initialize with 1% of labeled data and employ a frozen SigLIP (Zhai et al., 2023) as the foundation model for feature extraction and weak labeling in the two-pass process. We conducted all experiments on 2 NVIDIA A10 GPUs with Intel Xeon Gold 6342 CPUs.



(a) CIFAR-10 (b) CIFAR-100

Figure 2: Experimental comparison of our method with existing approaches on CIFAR-10 and CIFAR-100 under a Non-IID data distribution. Other AL methods begin by randomly selecting 10% of the initial data, followed by 50 communication rounds of training after each AL sampling step until reaching a 35% labeling budget. In contrast, our method completes training at 100 rounds and achieves its highest performance (indicated by the grey line). Refer to Appendix Figure 4 for results over the entire 300-round training process.

4.2. Results

FAST under a fixed communication budget with Non-IID data distribution. We assess the effectiveness of FAST in a non-IID data setting by distributing client data according to a Dirichlet distribution with a concentration parameter of $\alpha = 0.1$, thereby inducing high data heterogeneity across all clients. Figure 2 illustrates the comparative convergence rates of the global model on CIFAR-10 and CIFAR-100, where the grey line denotes the optimal performance achieved by all AL methods. To evaluate the communication efficiency of our approach, we conduct a total of $T = 300$ FL communication rounds across all methods. For the baseline methods, we initialize the process with 10% of labeled data at the beginning of the first 50 FL rounds. In the subsequent federated training phases, the server queries 5% of unlabeled instances for human annotation every 50 rounds until the total communication budget is exhausted.

In contrast, FAST employs a two-pass active sampling process at the onset of the AL phase to utilize the predefined global labeling budget without necessitating further oracle participation. As depicted in Figure 2, FAST achieves superior global model performance by the 100_{th} FL round without depleting the allocated communication budget. These results demonstrate that our method enables the server to efficiently train a high-performing global model within limited communication resources in realistic scenarios.

5. Conclusion

In this paper, we introduced a two-pass FAL framework, **FAST**, designed to address the critical challenges of limited annotation budgets and communication-intensive sampling processes in FAL. Our approach leverages robust representation-based knowledge from foundation models to efficiently query informative unlabeled data for annotation, thereby minimizing human effort and communication overhead. Extensive experiments on diverse vision datasets demonstrate that FAST consistently outperforms existing FAL methods in terms of both predictive performance and communication cost. These findings underscore the potential of leveraging foundation models to enhance FAL under realistic resource constraints. Future directions include exploring more sophisticated query strategies within **FAST** and quantifying weak labeling quality. Developing an additional filtering mechanism after the two-pass labeling process to enable label correction prior to final human annotation, thereby further enhancing performance and communication efficiency.

Acknowledgments

This work was partially supported by the NGF AiNed Fellowship Grant of A.S. We acknowledge the use of the Dutch SURF Research cloud to run the experiments presented in this paper.

6. Impact Statements

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Ahn, J.-H., Ma, Y., Park, S., and You, C. Federated active learning (f-al): an efficient annotation strategy for federated learning. *IEEE Access*, 2024.

Al-Saedi, A. A., Boeva, V., and Casalicchio, E. Reducing communication overhead of federated learning through clustering analysis. In *2021 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–7. IEEE, 2021.

Alfalqi, K. and Bellaiche, M. Emergency events detection based on integration of federated learning and active learning. *International Journal of Information Technology*, 15 (6):2863–2876, 2023.

Ash, J. T., Zhang, C., Krishnamurthy, A., Langford, J., and Agarwal, A. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint arXiv:1906.03671*, 2019.

Aussel, N., Chabridon, S., and Petetin, Y. Combining federated and active learning for communication-efficient distributed failure prediction in aeronautics. *arXiv preprint arXiv:2001.07504*, 2020.

Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Li, K. H., Parcollet, T., de Gusmão, P. P. B., et al. Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*, 2020.

Cao, Y.-T., Shi, Y., Yu, B., Wang, J., and Tao, D. Knowledge-aware federated active learning with non-iid data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22279–22289, 2023.

Caramalau, R., Bhattarai, B., and Kim, T.-K. Sequential graph convolutional network for active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9583–9592, 2021.

Chen, J., Ma, B., Cui, H., and Xia, Y. Think twice before selection: Federated evidential active learning for medical image analysis with domain shifts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11439–11449, 2024.

Citovsky, G., DeSalvo, G., Gentile, C., Karydas, L., Rajagopalan, A., Rostamizadeh, A., and Kumar, S. Batch active learning at scale. *Advances in Neural Information Processing Systems*, 34:11933–11944, 2021.

Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977.

Deng, Z., Yang, Y., Suzuki, K., and Jin, Z. Fedal: An federated active learning framework for efficient labeling in skin lesion analysis. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1554–1559. IEEE, 2022.

Frénay, B. and Verleysen, M. Classification in the presence of label noise: a survey. *IEEE transactions on neural networks and learning systems*, 25(5):845–869, 2013.

Gal, Y., Islam, R., and Ghahramani, Z. Deep bayesian active learning with image data. In *International conference on machine learning*, pp. 1183–1192. PMLR, 2017.

Geifman, Y. and El-Yaniv, R. Deep active learning over the long tail. *arXiv preprint arXiv:1711.00941*, 2017.

Hendrycks, D., Mazeika, M., Wilson, D., and Gimpel, K. Using trusted data to train deep networks on labels corrupted by severe noise. *Advances in neural information processing systems*, 31, 2018.

Houlsby, N., Huszár, F., Ghahramani, Z., and Lengyel, M. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

Hsu, T.-M. H., Qi, H., and Brown, M. Measuring the effects of non-identical data distribution for federated visual classification. *arXiv preprint arXiv:1909.06335*, 2019.

Huang, S., Wang, T., Xiong, H., Huan, J., and Dou, D. Semi-supervised active learning with temporal output discrepancy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3447–3456, 2021.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2):1–210, 2021.

Kim, S., Bae, S., Song, H., and Yun, S.-Y. Re-thinking federated active learning based on inter-class diversity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3944–3953, 2023.

Kirsch, A., Van Amersfoort, J., and Gal, Y. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems*, 32, 2019.

Kong, X., Zhang, W., Qu, Y., Yao, X., and Shen, G. Fedawr: An interactive federated active learning framework for air writing recognition. *IEEE Transactions on Mobile Computing*, 2023.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Kutsuna, N., Higaki, T., Matsunaga, S., Otsuki, T., Yamaguchi, M., Fujii, H., and Hasezawa, S. Active learning framework with iterative clustering for bioimage classification. *Nature communications*, 3(1):1032, 2012.

Le, Y. and Yang, X. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015.

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., and Smith, V. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

Li, Y.-F. and Zhou, Z.-H. Towards making unlabeled data never hurt. *IEEE transactions on pattern analysis and machine intelligence*, 37(1):175–188, 2014.

Li, Y.-F., Tsang, I. W., Kwok, J. T., and Zhou, Z.-H. Convex and scalable weakly labeled svms. *Journal of Machine Learning Research*, 14(7), 2013.

Liu, K., Hu, S., Wu, S. Z., and Smith, V. On privacy and personalization in cross-silo federated learning. *Advances in neural information processing systems*, 35:5925–5940, 2022.

Lukasik, M., Bhojanapalli, S., Menon, A., and Kumar, S. Does label smoothing mitigate label noise? In *International Conference on Machine Learning*, pp. 6448–6458. PMLR, 2020.

Luping, W., Wei, W., and Bo, L. Cmfl: Mitigating communication overhead for federated learning. In *2019 IEEE 39th international conference on distributed computing systems (ICDCS)*, pp. 954–964. IEEE, 2019.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. PMLR, 2017.

Morell, J. Á., Dahi, Z. A., Chicano, F., Luque, G., and Alba, E. Optimising communication overhead in federated learning using nsga-ii. In *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*, pp. 317–333. Springer, 2022.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A. Y., et al. Reading digits in natural images with unsupervised feature learning. In *NIPS workshop on deep learning and unsupervised feature learning*, volume 2011, pp. 4. Granada, 2011.

Nguyen, V.-L., Destercke, S., and Hüllermeier, E. Epistemic uncertainty sampling. In *Discovery Science: 22nd International Conference, DS 2019, Split, Croatia, October 28–30, 2019, Proceedings 22*, pp. 72–86. Springer, 2019.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Parvaneh, A., Abbasnejad, E., Teney, D., Haffari, G. R., Van Den Hengel, A., and Shi, J. Q. Active learning by feature mixing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12237–12246, 2022.

Patrini, G., Rozza, A., Krishna Menon, A., Nock, R., and Qu, L. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1944–1952, 2017.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

Ren, Z., Wang, S., and Zhang, Y. Weakly supervised machine learning. *CAAI Transactions on Intelligence Technology*, 8(3):549–580, 2023.

Sener, O. and Savarese, S. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

Settles, B., Craven, M., and Ray, S. Multiple-instance active learning. *Advances in neural information processing systems*, 20, 2007.

Shannon, C. E. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

Sun, Q., Fang, Y., Wu, L., Wang, X., and Cao, Y. Eva-clip: Improved training techniques for clip at scale. *arXiv preprint arXiv:2303.15389*, 2023.

Urner, R., Wulff, S., and Ben-David, S. Plal: Cluster-based active learning. In *Conference on learning theory*, pp. 376–397. PMLR, 2013.

Wang, D. and Shang, Y. A new active labeling method for deep learning. In *2014 International joint conference on neural networks (IJCNN)*, pp. 112–119. IEEE, 2014.

Wang, J., Liu, Q., Liang, H., Joshi, G., and Poor, H. V. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in neural information processing systems*, 2020.

Wang, Y., Ma, X., Chen, Z., Luo, Y., Yi, J., and Bailey, J. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 322–330, 2019.

Wei, X.-S. and Zhou, Z.-H. An empirical study on image bag generators for multi-instance learning. *Machine learning*, 105:155–198, 2016.

Wei, X.-S., Wu, J., and Zhou, Z.-H. Scalable algorithms for multi-instance learning. *IEEE transactions on neural networks and learning systems*, 28(4):975–987, 2016.

Wu, X., Pei, J., Chen, C., Zhu, Y., Wang, J., Qian, Q., Zhang, J., Sun, Q., and Guo, Y. Federated active learning for multicenter collaborative disease diagnosis. *IEEE transactions on medical imaging*, 42(7):2068–2080, 2022.

Wu, Y., Shu, J., Xie, Q., Zhao, Q., and Meng, D. Learning to purify noisy labels via meta soft label corrector. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 10388–10396, 2021.

Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., and Ni, B. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023.

Yang, L., Zhang, Y., Chen, J., Zhang, S., and Chen, D. Z. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, pp. 399–407. Springer, 2017.

Yi, K. and Wu, J. Probabilistic end-to-end noise correction for learning with noisy labels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7017–7025, 2019.

Yoo, D. and Kweon, I. S. Learning loss for active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 93–102, 2019.

Zhai, X., Mustafa, B., Kolesnikov, A., and Beyer, L. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11975–11986, 2023.

Zhan, X., Wang, Q., Huang, K.-h., Xiong, H., Dou, D., and Chan, A. B. A comparative survey of deep active learning. *arXiv preprint arXiv:2203.13450*, 2022.

Zheng, G., Awadallah, A. H., and Dumais, S. Meta label correction for noisy label learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11053–11061, 2021.

Zhou, Z.-H. A brief introduction to weakly supervised learning. *National science review*, 5(1):44–53, 2018.

A. Appendix

A.1. Algorithm of FAST

Algorithm 1 FAST: Two-Pass Federated Active Learning

Data: Local datasets $D_k = \{\mathcal{U}_k, \mathcal{G}_k^{(0)}\}$.
Input: K clients; T federated rounds with τ local steps; Feature encoder $f(\cdot)$; Budget $B = \sum_{r=1}^{R} b$.
Output: The target global model w.

Initialize: Server initializes global model with $w^{(0)}$.
========================== Preliminary Pass ==========================
for client $k = 1, \ldots, K$ (in parallel) **do**
  Feature encoding $\mathbf{Z}_k \leftarrow f(\mathcal{U}_k) \cup f(\mathcal{G}_k^{(0)})$
  Perform label propagation on $\mathbf{Z}_{\mathcal{U}_k}$ to assign *weak labels* based on $\mathcal{G}_k^{(0)}$
  for class $c \in C$ **do**
    Compute class similarity $\mathbf{s_{i,c}}$ for each weakly labeled sample $x_i \in \mathcal{U}_k$ using Eq. (7)
  end for
end for
=========================== Refinement Pass ===========================
for client $k = 1, \ldots, K$ (in parallel) **do**
  Compute uncertainty score u_i using Eq. (8)
  Select top-b samples with highest u_i for each sample $x_i \in \mathcal{U}_k$ for oracle annotation
  Update labeled set $\mathcal{G}_k \leftarrow \mathcal{G}_k \cup \mathcal{U}_k^{(b)}$
  Merge refined annotated data with labeled data $\mathcal{G}_k^{(0)}$
end for
========================== Federated Training ==========================
for communication round $t = 1, \ldots, T$ **do**
  Client Update: Distribute $w^{(t)}$ to clients in K.
  for client $k = 1, \ldots, K$ (in parallel) **do**
    Initialize local model $w_k^{(t)} \leftarrow w^{(t)}$.
    for $i = 0, \ldots, \tau - 1$ **do**
      Perform local SGD updates on client k $w_k^{(t+1)} \leftarrow w_k^{(t)} - \eta \nabla \mathcal{F}_k(w_k^{(t)})$.
    end for
    Send updated $w_k^{(t+1)}$ back to server.
  end for
  Server Update: Aggregate local models.
  Update global model $w^{(t+1)} \leftarrow \frac{1}{|K|} \sum_{k \in K} w_k^{(t)}$
end for
Return Target global model w.

A.2. Analysis on computational and communications overhead in FAST.

In FL, communication overhead is most commonly measured either by the number of rounds or by the total volume of parameter transmissions (i.e., model uploads and downloads) (Luping et al., 2019; Al-Saedi et al., 2021; Morell et al., 2022). As shown in Table I and Figure 2 of the main paper, FAST achieves higher accuracy within 100 FL rounds while reducing the number of communication rounds by a factor of eight. Table 1 reports the aggregate parameter-transmission cost of each method. In FAST's preliminary stage, a small set of labeled embeddings is uploaded to and redistributed from the server for weak labeling, incurring an initial communication cost. However, by leveraging representation-based labeling and minimal human annotation, FAST greatly reduces the required number of FL rounds and thus the overall communication volume.

In contrast, conventional FAL methods incur identical model-aggregation and distribution costs in every round, and their extensive active-learning cycles result in comparable or higher total communication overhead.

Furthermore, we compare computational costs in terms of wall-clock time (Table 1). The only additional expense in FAST's preliminary stage is encoding local data via the foundation model's encoder for uncertainty estimation. By contrast, LoGo (Kim et al., 2023) performs both macro- and micro-level informativeness evaluations using local and global models in each active-learning round, imposing a substantially higher runtime burden. As the table shows, on CIFAR-10 LoGo incurs approximately 2310 s more wall-clock time than random sampling over the same number of rounds.

Table 1: Comparison of classification accuracy (%), communication cost (MB), and wall-clock time (s) for FAST, LoGo, and random sampling on CIFAR-10, SVHN, and PathMNIST under the IID setting. We apply FedAvg on a four-layer CNN classifier (with a size of 0.45 MB on the CIFAR-10 dataset). FAST is evaluated with one AL round (100 FL rounds), whereas Random and LoGo use 8 AL rounds (i.e., 800 FL rounds).

Method	CIFAR-10			CIFAR-100			SVHN		
	Acc. (%)	Comm. Cost (MB)	Walltime (s)	Acc. (%)	Comm. Cost (MB)	Walltime (s)	Acc. (%)	Comm. Cost (MB)	Walltime (s)
Random	69.14	7090.94	30398.95	32.67	8502.69	37258.75	85.47	6969.79	58064.42
LoGo	71.92	7090.94	32709.14	34.27	8502.69	39268.24	87.08	6969.79	61638.97
Ours	**77.16** ↑	**902.54** (87.27%) ↓	**7342.52** (76.73%) ↓	**41.94** ↑	**1079.56** (87.30%) ↓	**15104.38** (60.53%) ↓	**88.79** ↑	**896.72** (87.13%) ↓	**19178.83** (67.96%) ↓

A.3. Comparison on heterogeneous inter-class diversity data.

We first evaluate the performance of FAST in comparison with other baseline methods on datasets characterized by high levels of local heterogeneity. In this context, each client shares the same pool of classes but exhibits varying inter-class distributions. We deliberately fix FAST's labeling budget at 5% to underscore its cost-effectiveness (Table 2). With just one AL round and 5% labeled data, FAST matches or exceeds the accuracy of baselines that consume up to 40% budget over multiple AL rounds. In contrast, conventional AL methods sample 5% per round iteratively until they exhaust the same budget, incurring additional communication and computation. FAST achieves competitive performance through a single, one-shot active labeling round. This design choice highlights FAST's significant efficiency advantages in reducing communication overhead.

Conventional AL methods aim to minimize labeling efforts by selecting a small subset of instances based on their informativeness across the entire dataset. However, in a decentralized setting where each local dataset maintains distinct class distributions, such imbalanced data partitions often lead to inconsistent knowledge sharing. Consequently, the selected samples may not be representative or sufficiently informative for all clients, thereby hindering the overall learning performance. As presented in Table 2, we compare the performance of **FAST** with other existing AL strategies under a one-shot setting, wherein only a single active sampling round is conducted. We observe that **FAST** outperforms all baseline methods even within the constraints of this one-shot scenario. Notably, in this experiment, the server exhausts 5% of the labeling budget per round until reaching the total budget limit.

In **FAST**, each client shares their representation-based knowledge with other clients without revealing the raw local dataset, thereby enabling the server's query selector to address imbalanced class distributions from a global perspective. By fully exploiting the comprehensive information of the unlabeled dataset, **FAST** is able to achieve superior performance after the first AL round.

A.4. Impact of two-pass active sampling on Foundation Models with Linear Probing.

In this experiment, we evaluate the efficacy of a two-pass sampling strategy within **FAST** by integrating a foundation model as the backbone during training. Instead of training client-specific models from scratch, we employ linear probing on the client side using only a limited labeled dataset. To systematically analyze the contribution of each component, we decompose the training process into four distinct elements: ① Linear Probing, ② Weak Labeling, ③ Active Learning, and ④ Random Sampling.

Table 3 illustrates the performance outcomes of various component combinations across multiple datasets under a fixed labeling budget of 20%, encompassing 100 FL rounds distributed among 10 clients, with an initial training dataset comprising 1% of labeled data for all clients. Specifically, we consider five different scenarios to examine the efficacy of the two-pass mechanism in **FAST**, where the combination of the first three components (①, ②, and ③) represents the integration of

Table 2: Test accuracy comparison of various Active Learning (AL) strategies across multiple datasets. We evaluate FAST in a one-shot (i.e., a single AL round) setting, where each AL round is followed by 100 Federated Learning (FL) rounds, resulting in a total of $R_{\text{FL}} = R_{\text{AL}} \times 100$. The labeling budget denotes the percentage of data allocated for labeling, with each AL round querying 5% of the unlabeled samples for annotation.

Method	R$_{AL}$	CIFAR-10	SVHN	PathMNIST	DermaMNIST	R$_{FL}$	Budget
Random	4	64.19	80.90	68.41	71.70	400	20%
	8	69.07	84.22	73.76	72.66	800	40%
Entropy (Wang & Shang, 2014)	4	64.02	82.08	71.54	72.49	400	20%
	8	69.12	85.88	75.91	73.02	800	40%
Coreset (Sener & Savarese, 2017)	4	64.66	80.94	74.84	72.02	400	20%
	8	69.43	83.81	76.85	72.34	800	40%
BADGE (Ash et al., 2019)	4	65.12	82.81	72.21	72.59	400	20%
	8	69.57	85.89	75.53	73.23	800	40%
GCNAL (Caramalau et al., 2021)	4	65.40	82.05	75.51	72.01	400	20%
	8	70.05	85.09	78.13	73.07	800	40%
ALFA-Mix (Parvaneh et al., 2022)	4	65.45	83.02	73.34	72.39	400	20%
	8	69.87	86.05	76.31	73.27	800	40%
LoGo (Kim et al., 2023)	4	66.50	83.46	76.32	72.61	400	20%
	8	71.70	86.02	79.51	73.33	800	40%
Ours	**1**	**77.14**	**87.91**	**88.48**	**74.37**	**100**	**5%**

①	②	③	④	CIFAR-10	CIFAR-100	Tiny-ImageNet	PathMNIST
✓	✓	✓		**96.04**	**60.83**	**54.41**	**86.67**
✓	✓		✓	95.31	58.94	52.95	82.33
✓	✓			94.47	53.56	46.92	75.84
✓			✓	94.53	52.84	47.79	74.12
✓				80.43	5.61	1.60	49.89

Table 3: Effects of training components: ① Linear Probing, ② Preliminary Pass, ③ Refinement Pass, ④ Random. We train with a limited 1% of initial labeled data across all 10 clients for 100 FL rounds. The labeling budget is 20%.

FAST into linear probing. In Table 3, the configuration employing the two-pass sampling strategy (①, ②, ③) achieves superior performance compared to the configurations that only implement preliminary labeling (①, ②) and those that omit oracle refinement phase (①, ②, ④). This demonstrates the critical role of human refinement during the FAL process in enhancing model performance. Notably, we observe significantly lower performance when directly applying linear probing with the foundation model on the initial labeled data without any further AL operations (① only). These findings collectively highlight that the two-pass active sampling mechanism in **FAST** not only maximizes the utility of the limited labeling budget but also fosters effective knowledge sharing across heterogeneous clients, thereby achieving superior global model performance with constrained communication resources.

Moreover, our analysis in Table 3 decomposes the FAST framework into distinct components for qualitative evaluation. Notably, we observe that applying uncertainty sampling (①+②+③) consistently outperforms random sampling in the refinement pass (①+②+④) across various datasets. These results empirically validate the effectiveness of uncertainty sampling within the second pass of our framework.

Effect of Varying the Number of Clients on FAST.

We evaluate **FAST** with 10, 20, and 30 clients on CIFAR-10 and CIFAR-100 to assess its scalability and robustness. As shown in Appendix A.4, the test accuracy decreases smoothly as the client count increases, indicating that more federated training rounds may be needed for convergence. Nonetheless, even under a limited annotation budget, **FAST** maintains strong performance without significant degradation, demonstrating its stability in larger federated learning clusters.

Table 4: Performance of uncertainty sampling strategies on weak-labeled data across various datasets. Training with 10 clients for 100 rounds, utilizing a 4-layer CNN network. Evaluating with FedAvg.

Dataset	Norm-Based	Entropy-Based	Least Confidence	Smallest Margin	Largest Margin
CIFAR-10	73.81	73.79	73.62	74.14	73.90
CIFAR-100	34.77	35.55	35.49	35.72	35.25
PathMNIST	84.64	85.43	85.29	84.85	85.70
Tiny-ImageNet	28.37	29.18	28.89	28.72	28.91
Average	55.40	55.74	55.82	55.86	55.94

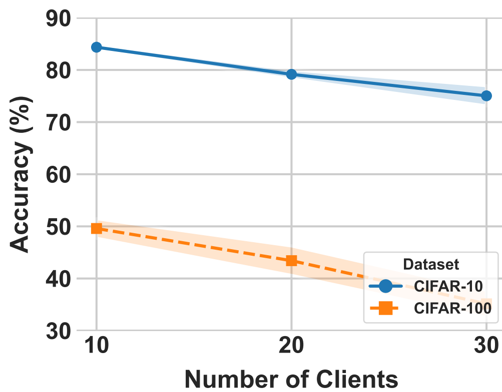


Figure 3: Performance of **FAST** across 10, 20, and 30 clients on CIFAR-10/100 under FedAvg with 150 FL rounds.

FAST under a fixed communication budget with Non-IID data distribution.

‘1

In Figure 2, we showed FAST's rapid convergence within the first 100 communication rounds. For completeness, Figure 4 presents extended results up to 300 rounds under the same Non-IID setting. As before, each method starts with a 10% initial labeling and continues AL sampling until reaching 35% of the labeling budget.

Evaluating Uncertainty Strategies for Prototype-Based Weak Labeling.

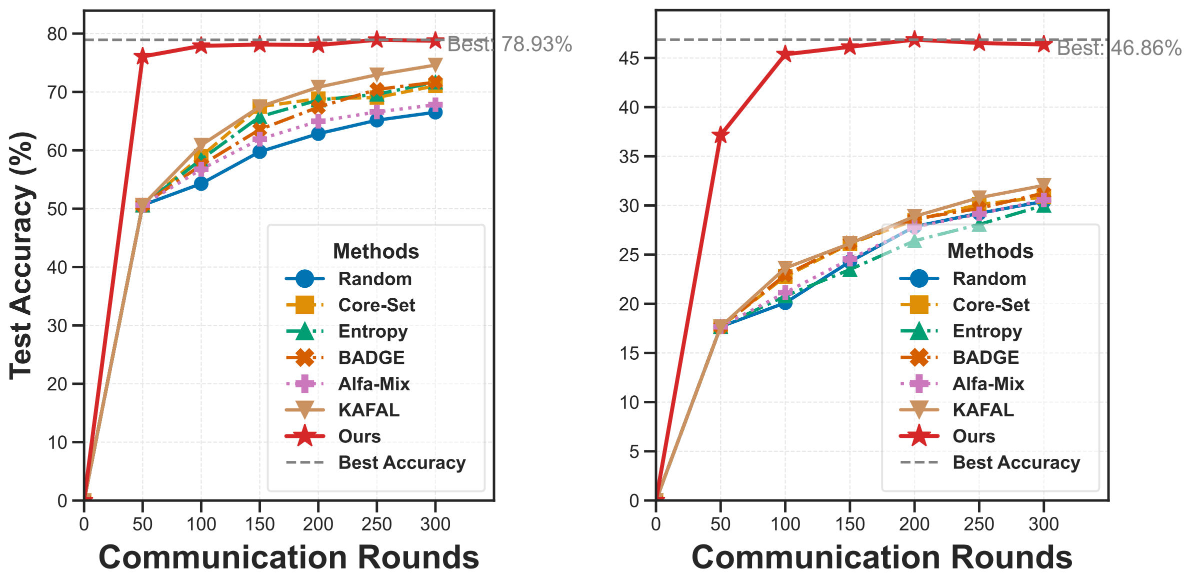
We evaluate several uncertainty-based query strategies: norm-based, entropy-based, least confidence, smallest margin, and largest margin—applied to the prototype vectors computed for each weakly labeled sample. As summarized in Table 4, the results are generally comparable across different datasets, suggesting that the prototype-based logits capture the key uncertainty information leveraged by a variety of query strategies. This underscores the effectiveness of the prototype representation in identifying highly uncertain samples for human refinement.

Ablation on different federated learning strategies.

We investigate the impact of various FL strategies on the performance of **FAST** under a fixed labeling budget of 20%. Table 5 reports the accuracy across five benchmark datasets. Notably, FedNova offers marginal yet consistent improvements over FedAvg and FedProx on most datasets, indicating that **FAST** is compatible with advanced FL aggregation strategies and can further support heterogeneous scenarios. These findings confirm the robustness of **FAST** under different federated aggregation schemes.

Table 5: Performance of Our Method Across Different Federated Learning Strategies with 20% Labeling Budget

Strategy	CIFAR-10	CIFAR-100	SVHN	PathMNIST	Tiny-ImageNet
FedAvg	73.81	34.77	86.27	84.64	26.03
FedProx	73.63	32.84	83.19	85.36	25.90
FedNova	74.12	36.60	87.12	87.92	28.30



(a) CIFAR-10 (b) CIFAR-100

Figure 4: Experimental comparison of our method with existing approaches on the CIFAR-10 and CIFAR-100 datasets under a Non-IID data distribution. For other active learning (AL) methods, the process begins by randomly selecting 10% of the initial data, followed by training with 50 communication rounds after each AL sampling step until a labeling budget of 35% is reached.

Table 6: Performance of Our Method with Varying Labeling Budgets. Training with the FedAvg strategy using a CNN-4 model, 10 clients, 100 rounds.

Dataset	**Labeling Budget**			
	0%	5%	40%	80%
CIFAR-10	75.92	76.73	77.24	77.48
CIFAR-100	31.33	33.34	39.65	44.27
PathMNIST	73.16	75.89	82.28	85.46

Effect of Different Foundation Model Selections on FAST.

We next evaluate how the choice of foundation model for the preliminary pass in **FAST** influences its overall performance. Specifically, we compare three vision-language models—CLIP, EvaCLIP, and SigLIP—along with an image-specific model, DINOv2, using a pre-trained ResNet-50 as the baseline. As shown in Table 7, EvaCLIP consistently achieves the highest accuracy across all datasets, followed closely by SigLIP and DINOv2. This underscores the importance of rich representation knowledge for enhancing weak labeling quality in the preliminary pass. Furthermore, the results suggest that leveraging expressive embeddings can significantly improve active sampling outcomes, even under constrained annotation budgets.

Table 7: Performance Comparison of Our Methods with Different Foundation Models

Dataset	**ResNet-50**	**CLIP**	**Eva-CLIP**	**SigLIP**	**DINOv2**
CIFAR-10	77.86	83.81	85.98	84.87	85.34
CIFAR-100	28.86	38.32	53.27	50.41	50.38
PathMNIST	82.67	87.73	91.04	88.79	89.19

subsection*Ablation on varying of Labeling Budget. To assess the scalability of FAST with respect to the labeling budget, we evaluate its performance under varying labeling budgets ranging from 0% to 80%. Table 6 illustrates the accuracy of FAST across various datasets as the labeling budget increases. The results demonstrate a positive correlation between the labeling budget and model accuracy, with significant performance improvements observed as the budget increases. For instance, on CIFAR-10, accuracy improves from 75.92% at 0% budget to 77.48% at 80% budget. Similar trends are observed across CIFAR-100 and Path-MNIST, indicating the effectiveness of FAST in leveraging additional unlabeled data to enhance model performance under constrained labeling budgets.