METHODS
published: 09 June 2022
doi: 10.3389/fpls.2022.908814

An EMD-Based Adaptive Client Selection Algorithm for Federated Learning in Heterogeneous Data Scenarios

*Aiguo Chen, Yang Fu, Zexin Sha and Guoming Lu**

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

Federated learning is a distributed machine learning framework that enables distributed nodes with computation and storage capabilities to train a global model while keeping distributed-stored data locally. This process can promote the efficiency of modeling while preserving data privacy. Therefore, federated learning can be widely applied in distributed conjoint analysis scenarios, such as smart plant protection systems, in which widely networked IoT devices are used to monitor the critical data of plant production to improve crop production. However, the data collected by different IoT devices can be dependent and identically distributed (non-IID), causing the challenge of statistical heterogeneity. Studies have also shown that statistical heterogeneity can lead to a marked decline in the efficiency of federated learning, making it challenging to apply in practice. To promote the efficiency of federated learning in statistical heterogeneity scenarios, an adaptive client selection algorithm for federated learning in statistical heterogeneous scenarios called ACSFed is proposed in this paper. ACSFed can dynamically calculate the possibility of clients being selected to train the model for each communication round based on their local statistical heterogeneity and previous training performance instead of randomly selected clients, and clients with heavier statistical heterogeneity or bad training performance would be more likely selected to participate in the later training. This client selection strategy can enable the federated model to learn the global statistical knowledge faster and thereby promote the convergence of the federated model. Multiple experiments on public benchmark datasets demonstrate these improvements in the efficiency of the models in heterogeneous settings.

Keywords: distributed conjoint analysis, federated learning, adaptive client selection, statistical heterogeneity, machine learning

OPEN ACCESS

Edited by:
*Zhipeng Cai,
Georgia State University,
United States*

Reviewed by:
*Jun Xu,
Nanjing Normal University, China
Miao Zhang,
Beijing University of Posts and
Telecommunications, China
Xiaogang Qi,
Xidian University, China
Zhuojun Duan,
James Madison University,
United States*

***Correspondence:**
*Guoming Lu
lugm@uestc.edu.cn*

Specialty section:
*This article was submitted to
Sustainable and Intelligent
Phytoprotection,
a section of the journal
Frontiers in Plant Science*

Received: *31 March 2022*
Accepted: *10 May 2022*
Published: *09 June 2022*

Citation:
*Chen A, Fu Y, Sha Z and Lu G (2022)
An EMD-Based Adaptive Client
Selection Algorithm for Federated
Learning in Heterogeneous Data
Scenarios.
Front. Plant Sci. 13:908814.
doi: 10.3389/fpls.2022.908814*

INTRODUCTION

With the rapid development of computation and storage capabilities, IoT devices are now widely applied in multiple areas, such as health monitors, smart homes, and smart plant protection systems. These devices are used to collect critical data during operation and provide a statistical foundation for services, such as health prediction, house management, or the direction of plant production. However, the broad deployment of these devices will lead to the expansion of the data scale and the general growth in the demand for privacy preservation of sensitive data, and the traditional centralized machine/deep learning mode that collects all data in a central node to train a statistical model has difficulty meeting demand. Specifically, transmitting distributed-stored data to a central node and training a model on that scale of data can lead to long computation and communication times. In addition, privacy leakage of sensitive data may occur during transmission. Therefore, federated learning (Konečný et al., 2015; McMahan et al., 2017a,b), a distributed machine learning framework that involves a central server and multiple distributed nodes, is proposed in this study to address the challenges that face centralized frameworks. Federated learning enables distributed nodes with computation and storage capabilities, such as mobile phones, IoT devices, and laptops, to train statistical models using local data for each communication round. Only the model parameters are transmitted to a central server. The central server then aggregates these local models to generate the new federated model for the next communication round. Federated learning can provide more efficient construction of the global model and privacy of data because it decentralizes the computation among distributed clients and keeps the distributed-stored data locally.

TABLE 1 | Details of the dataset.

Dataset	Category	Size (length, width, channel)	Description
MNIST	10	28 × 28 × 1	Hand-writing number image
Fashion MNIST	10	28 × 28 × 1	Wearing image
CIFAR 10	10	32 × 32 × 3	Common things image

However, in practical scenarios, the differences in device type, geographical location of deployment, and other factors will make the collected data non-IID, resulting in the challenge of statistical heterogeneity. Statistical heterogeneity is known to negatively impact multiple aspects of federated learning compared with IID data, like performance, efficiency, data privacy, etc. Non-IID data can have various features or label distributions, which causes marked differences between local models and thus leads to unstable convergence and low efficiency during federated model training. To address the influence of statistical heterogeneity on the efficiency of federated learning, many studies have investigated three aspects of this topic: restricting model divergence, decreasing communication cost, and customized client selection methods. Regarding restricting model divergence, some studies have focused on improving data quality, such as data sharing (Zhao et al., 2018) and data enhancement (Jeong et al., 2018).

On the contrary, other studies have investigated model training by modifying the model aggregation method (McMahan et al., 2017a), customizing the loss function (Li et al., 2020),

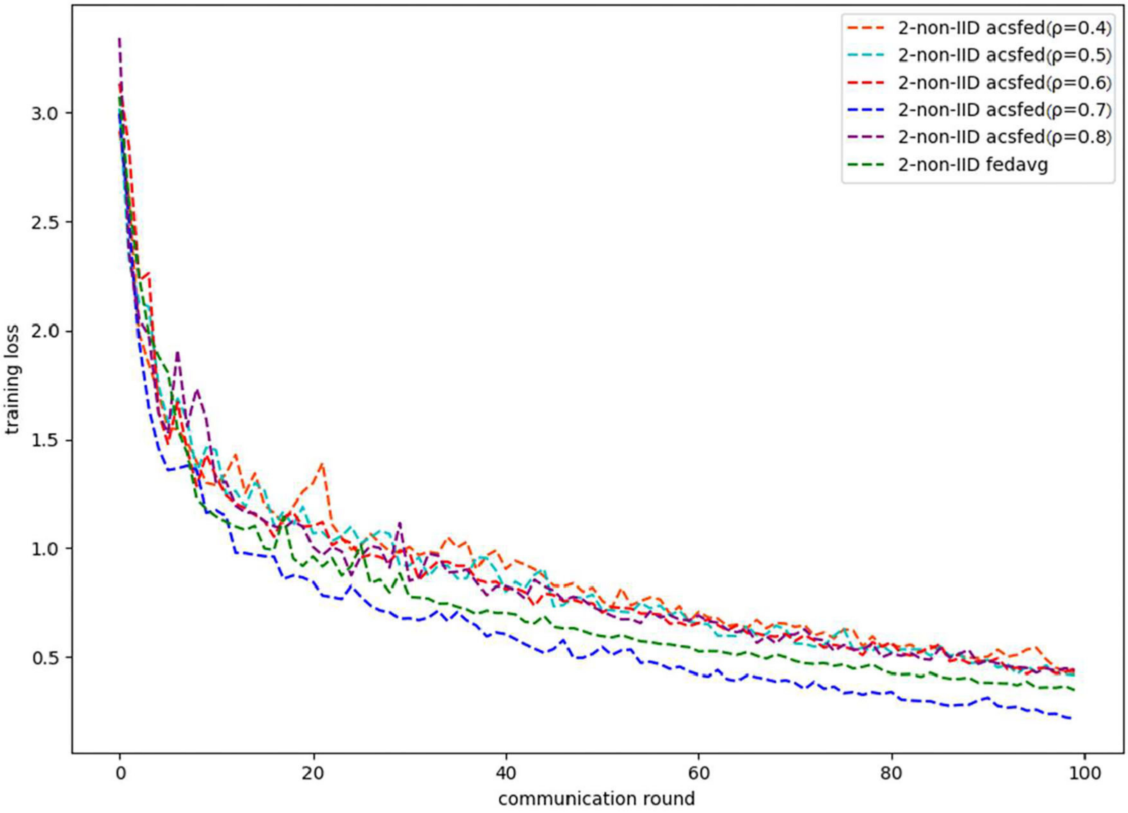

FIGURE 1 | Training loss on MNIST in 2-class non-IID scenario.

and combining enhancement methods, including reinforcement learning (Pang et al., 2020; Wang et al., 2020) and knowledge distillation (Li and Wang, 2019), with federated learning. The relation between statistical heterogeneity and the Earth's mover distance (EMD) is presented in Zhao et al. (2018). The EMD is essentially used to measure the distance between two distributions. Experimental results indicate that EMD could be an ideal index of statistical heterogeneity by calculating the EMD between the local data distribution on a client and the global data distribution. There is also some constructive work regarding decreasing communication costs, such as model compression (Sattler et al., 2019) and dynamic computing of the rounds of local updates (Wang et al., 2019). Regarding the customized client selection method, recent studies have proposed creatively customized methods (Nishio and Yonetani, 2019; Wang et al., 2020; Shen et al., 2022; Zhang et al., 2022; Zhao et al., 2022) to accelerate the convergence of federated learning and thereby promote efficiency. Moreover, Xiong et al. (2021) carried out an innovative exploration of privacy protection in FL with non-IID data and proposed a differential-privacy-based algorithm called 2DP-FL, which has advantages in privacy protection, learning convergence, and model accuracy.

In this study, we propose an adaptive client selection algorithm for federated learning called ACSFed based on the finding of correlations between EMD and statistical heterogeneity in Zhao et al. (2018). In ACSFed, the index measuring the data and training quality of each client, called cumulative model strength, is calculated based on each client's EMD and previous training performance. Then, a probability matrix representing the probability of each client being selected in each communication round is maintained and dynamically updated based on the cumulative model strength of each client. ACSFed can fully consider the data and previous training quality of each client to update the probability matrix so that clients with strong statistical heterogeneity or poor training performance will be more likely to be selected for training, thereby enabling the federated model to learn the unknown information or knowledge with poor learning performance much faster. After the federated model has learned sufficient knowledge of the global data, the convergence of the federated model speeds up markedly, and multiple experiments have shown that ACSFed can converge faster than FedAvg. The primary contributions of this paper can be summarized as follows:

1. An index measuring the statistical heterogeneity and previous training performance of a client called cumulative model strength is first proposed.
2. An adaptive client selection algorithm called ACSFed for federated learning based on cumulative model strength is proposed in this study. The probability of a client being selected to train the statistical model is dynamically updated.
3. Through ACSFed, the performance of federated learning in statistical heterogeneity scenarios can be promoted compared with current methods.

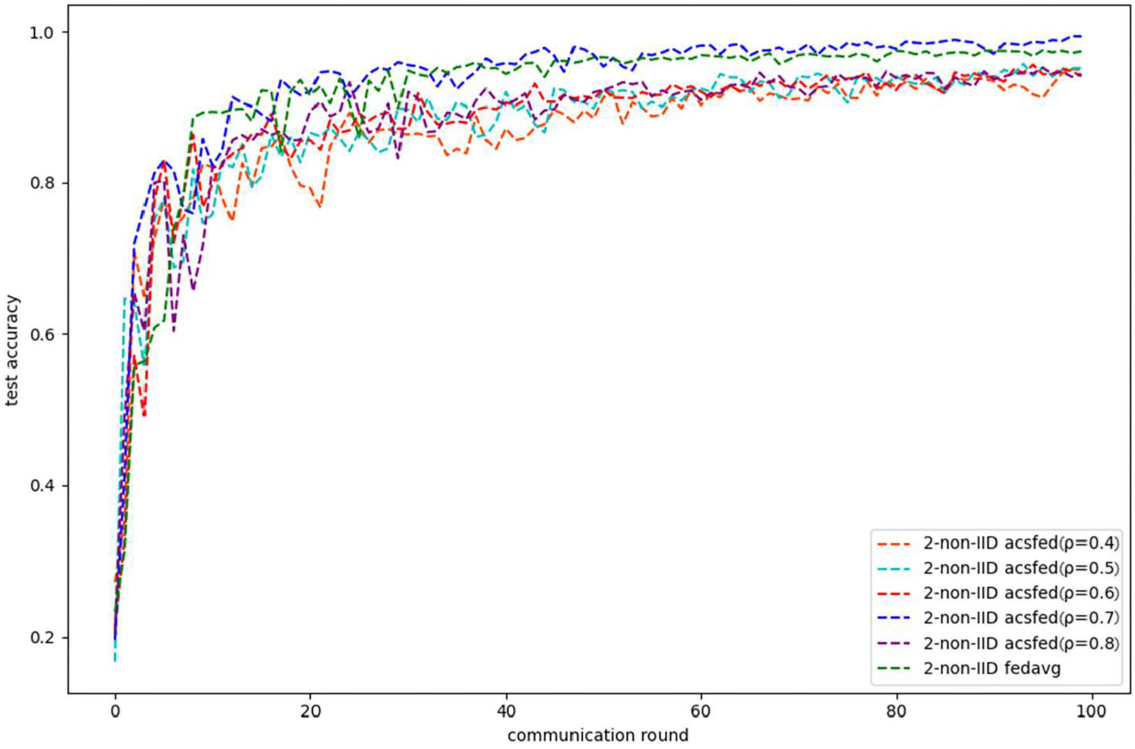

FIGURE 2 | Test accuracy on MNIST in 2-class non-IID scenario.

The remainder of this paper is organized as follows. Section Related Work describes the background of federated learning research in statistical heterogeneity scenarios. Section Problem Definition describes the problem that current methods face in statistical heterogeneity scenarios. Then, Section Method describes the structure and principles of the proposed approach. The theoretical analysis and experimental results are both shown in Section Evaluation. Finally, the conclusions of this study are provided in Section Conclusion.

RELATED WORK

Federated learning was first introduced by McMahan et al. (2017b). Its baseline algorithm is the federated stochastic gradient descent (FedSGD), enabling each client to execute a round of SGD locally and upload the model to a central server for aggregation. Although FedSGD can solve the challenges of privacy leakage of sensitive data and achieve the same accuracy of centralized model training when the data are IID, frequent model uploading and distribution have markedly increased the communication burden, resulting in the problem of efficiency.

In 2017, an improved algorithm called FedAvg (McMahan et al., 2017a) was proposed and allowed clients to synchronously execute several rounds of SGD before uploading the model to a central server for model aggregation, which efficiently decreases the communication rounds and thus promotes the efficiency of federated learning. The convergence of FedAvg is theoretically proven in Li et al. (2019), and experiments on public benchmark datasets also demonstrate that FedAvg has ideal convergence and robustness. However, Zhao et al. (2018) found that the performance and efficiency of FedAvg markedly shrink as statistical heterogeneity increases. Moreover, the relation between the EMD of statistical heterogeneity has been identified and implies that EMD can act as an ideal index of statistical heterogeneity. Further, Zhang et al. (2020) has used EMD to measure the degree of statistical heterogeneity and proposed a personalized federated learning model training algorithm, which outperforms existing alternatives. The results of Zhao et al. (2018) and Zhang et al. (2020) have proved that EMD can effectively measure statistical heterogeneity and can be applied to promote the performance and efficiency of federated learning.

Focusing on promoting the efficiency of federated learning, Wang et al. (2019) introduced adaptive federated learning that can dynamically compute the rounds of local updates for each client in resource-constrained edge computing systems. Compared with methods where the communication step is fixed, faster convergence can be achieved. Similarly, Huang et al. (2020) proposed an adaptive enhancement method called LoAdaBoost, which optimizes the first half of the local update rounds, and only those clients with low performance would continue training. This strategy can markedly reduce the expectation of local training rounds for each client, thereby promoting the efficiency of federated learning. In addition, considering

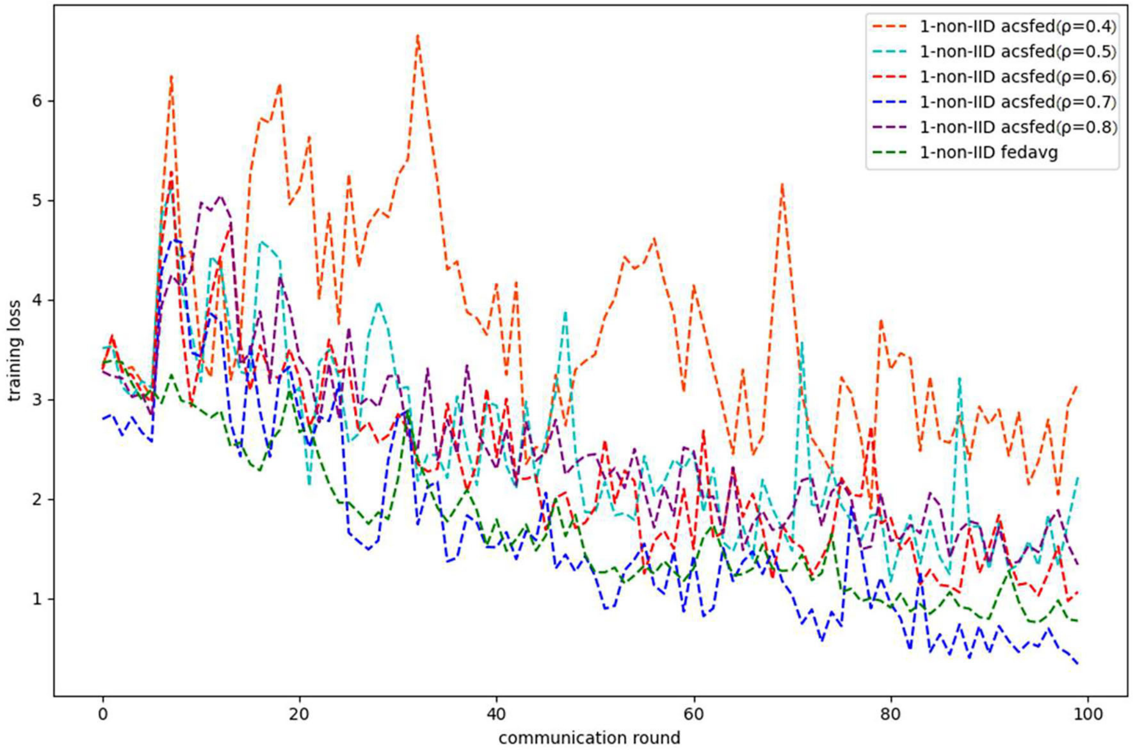

FIGURE 3 | Training loss on MNIST in the 1-class non-IID scenario.

communication costs, Konečný et al. (2016) markedly reduces communication costs via model compression, decreasing the uploaded model's size.

Similarly, Sattler et al. (2019) proposed a compression framework called sparse ternary compression (STC), which extends the existing compression technique by enabling downstream compression and internalization and optimal Golomb encoding of the weight updates. Moreover, Cai and Zheng (2018) considers energy conservation and privacy preservation and proposes an energy-efficient mechanism for data transmission. Besides energy conservation and privacy preservation (Cai and He, 2019; Cai et al., 2019), also consider data utility, and, respectively, proposed two mechanisms that can further preserve the results while protecting privacy. Additionally, Asad et al. (2020) introduced an algorithm combined with model compression and parameter encryption, effectively reducing communication overhead while protecting model security. Besides directly reducing communication costs, the efficiency of federated learning could also be improved by resource optimization. For example, Nishio et al. (2013), Sardellitti et al. (2015), and Yu et al. (2016) minimized the computation time and resource consumption based on the joint optimization of heterogeneous data, computation, and communication resources. In contrast, Nishio and Yonetani (2019) maximized the efficiency of federated model training through client selection based on resources, network conditions, and computation capability, and experiments have proven efficiency enhancement. However, these methods might markedly increase the computational burden in the distributed nodes, depending on the environment settings of federated learning, making the applicability of these methods not ideal.

In addition to improving the efficiency of federated learning by reducing communication cost and resource use, there are also methods that design client selection strategies to promote efficiency. For example, Chen et al. (2020) proposed an approximation algorithm that effectively improves efficiency and reduces the communication complexity by selecting some clients and allowing them to upload their updates in each round of training. Still, this method lacks the consideration of statistical heterogeneity. In 2022, Shen et al. (2022) proposed a novel stratified client selection scheme that develops stratification based on clients' local data distribution to derive approximate homogeneous strata for better selection in each stratum and can reduce the variance of the selected subset of clients for the pursuit of better convergence and higher accuracy. Experimental results show that their approach allows for better performance than existing methods and is compatible with prevalent federated learning algorithms. Similarly, Zhao et al. (2022) proposed FedNorm, a client selection framework that finds the clients that provide essential information in each round of model training and reduces energy cost by decreasing the number of participating clients. Evaluation results demonstrate

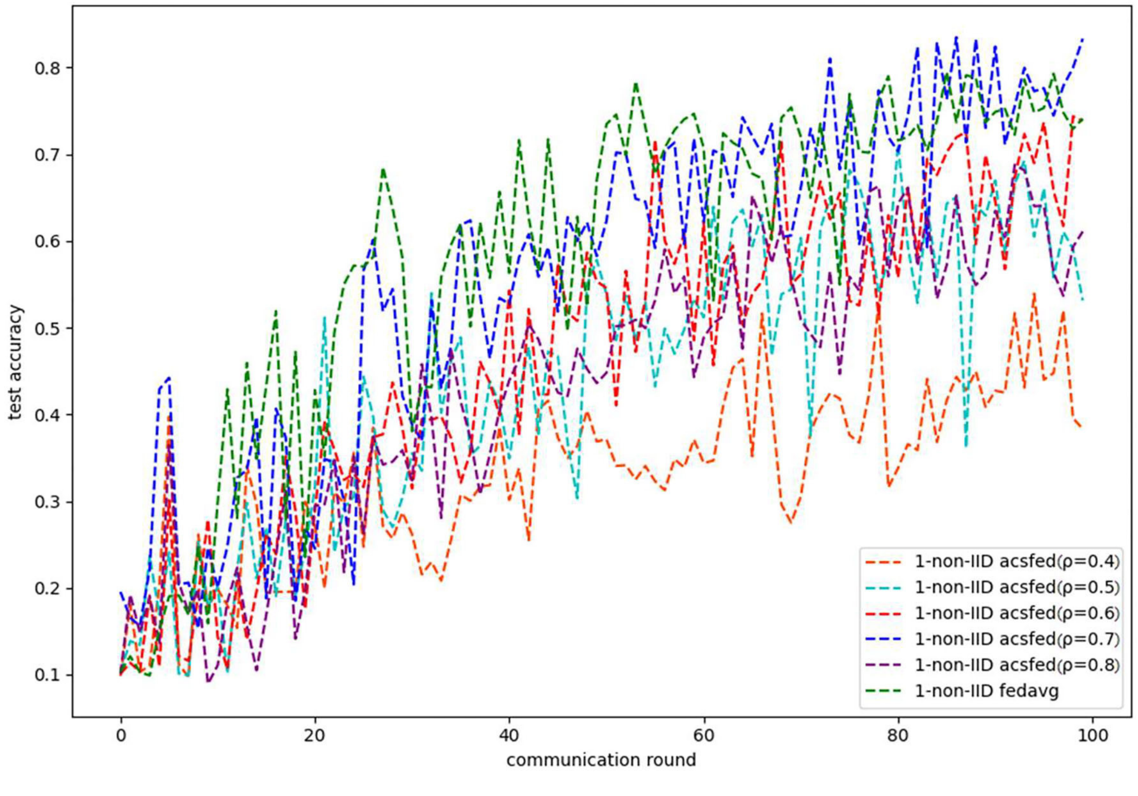

FIGURE 4 | Test accuracy on MNIST in the 1-class non-IID scenario.

that those algorithms outperform existing federated learning client selection methods in various statistical heterogeneity scenarios. In addition, there is also the idea of combining reinforcement learning with client selection. For example, Wang et al. (2020) proposed Favor. This experience-driven control framework intelligently chooses the client devices to participate in each round of federated learning to counterbalance the bias introduced by non-IID data and speed up convergence. However, that method requires specific experience to obtain ideal performance. Inspired by Multi-Agent Reinforcement Learning (MARL) in solving complex control problems, Zhang et al. (2022) presented FedMarl, a MARL-based federated learning framework that performs efficient run-time client selection. Experiments have shown that FedMarl can promote model accuracy with much lower processing latency and communication costs. However, FedMarl might increase the computational burden in the nodes, making it unsuitable for computation-ability-constrained scenarios. Additionally, Cho et al. (2020) found that selecting clients with higher local loss can improve convergence speed, and they proposed a client selection algorithm called Power-of-Choice based on their discovery. Compared with the random client selection algorithm, the convergence speed of the Power-of-Choice algorithm is significantly improved. Similarly, Zhang et al. (2021) utilizes weight divergence to identify the non-IID degree of clients, and proposes an efficient federated learning algorithm called CSFedAvg, which speeds up the training by selecting clients with less non-IID data to train the model more frequently. Although the speed of convergence and training accuracy is improved by applying CSFedAvg, it does not guarantee to cover all clients, resulting in the lack of completeness of the global data used for training.

Thus, although current studies have investigated improving the performance of federated learning in statistical heterogeneity scenarios, problems of higher computing and communication burdens and difficulties in practical application exist in recent research. Therefore, an improved method that can suppress or solve the above issues while retaining or improving performance must be studied.

PROBLEM DEFINITION

This paper proposes an adaptive client selection algorithm for federated learning in statistical heterogeneity scenarios. Federated learning is a distributed machine learning framework in which a set $C = \{1, 2, ..., K\}$ of K distributed clients owns a set of local data sampled from the global dataset D with M categories in total. The resulting statistical model is asynchronously trained with the same structure using local data. The optimization of federated learning is to train a suitable global parameter vector ω, which can minimize the total loss of the distributed clients and

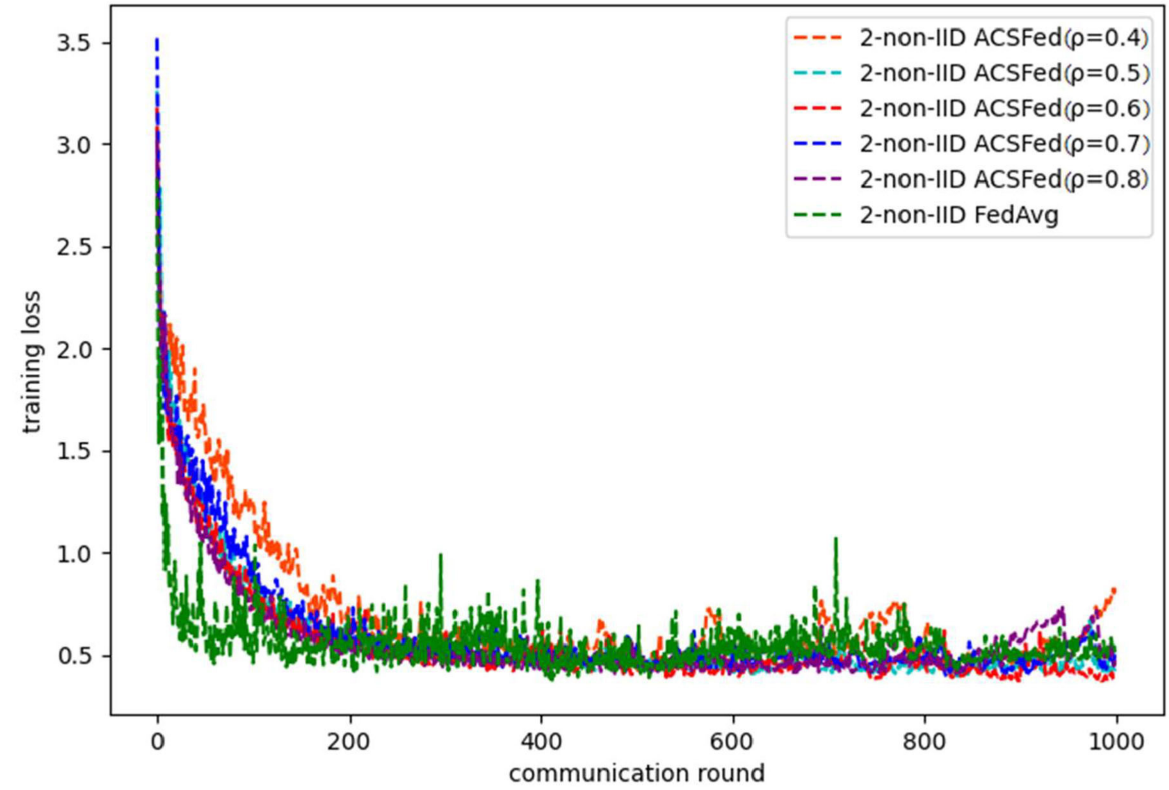

FIGURE 5 | Training loss on fashion MNIST in 2-class non-IID scenario.

can be quantified as follows:

$$\underset{\omega}{\text{minimize}}\, F_f(\omega) = \sum_{i=1}^{K} \frac{|D_i|}{|D|} f_i(\omega, D_i) \tag{1}$$

In IID scenarios, the local optimization object of any client is an unbiased estimate of the centralized optimization object $F_c(\omega)$ because the data distribution of these clients is the same as the global data distribution P. Therefore, instead of enabling all clients to participate in training, only a portion of clients can be selected randomly with a specific fraction to execute local updates in the model training of federated learning. This strategy can promote the efficiency of federated learning because it can markedly decrease the load and number of network transmissions.

However, $f_i(\omega, D_i)$ could be an arbitrary approximation to $F_c(\omega)$ when data across the clients is non-IID, which P_i can be markedly different P, leading to the deviation between the federated model and centralized-trained model and a marked decrease in the performance of the federated model. The convergence of the global model can be unstable due to the non-IID data because it is difficult for the model to learn the knowledge of the global data distribution by randomly selecting clients to participate in training. Therefore, we use ACSFed to solve the problem mentioned above. The goal of ACSFed is to make clients with heavy statistical heterogeneity and poor model performance selected with a larger possibility in the early training rounds, thereby enabling the global model to learn unknown data information as quickly as possible and providing more stable, faster convergence than random client selection. The theory and design of ACSFed are described in the next section.

METHOD

Definition of Cumulative Model Strength

In this section, the structure and principles of ACSFed will be shown in detail. In ACSFed, a probability matrix representing the probability of each client being selected is maintained and dynamically updated after each communication round. The updated probability is based on an index called the cumulative model strength of each client, which was first proposed by this study and considered the statistical heterogeneity and previous training performance. The calculation of cumulative model strength is defined as follows:

$$h_t^{(i)} \leftarrow \rho \cdot h_{t-1}^{(i)} + \frac{1-\rho}{EMD_i{}^{*}||Loss_i||_2^2}, i = 1, 2, ..., N \tag{2}$$

where $h_t^{(i)}$ and $h_{t-1}^{(i)}$ represent the cumulative model strength of the client i at the current and the previous communication rounds, respectively; $Loss_i$ denotes the training loss; ρ represents

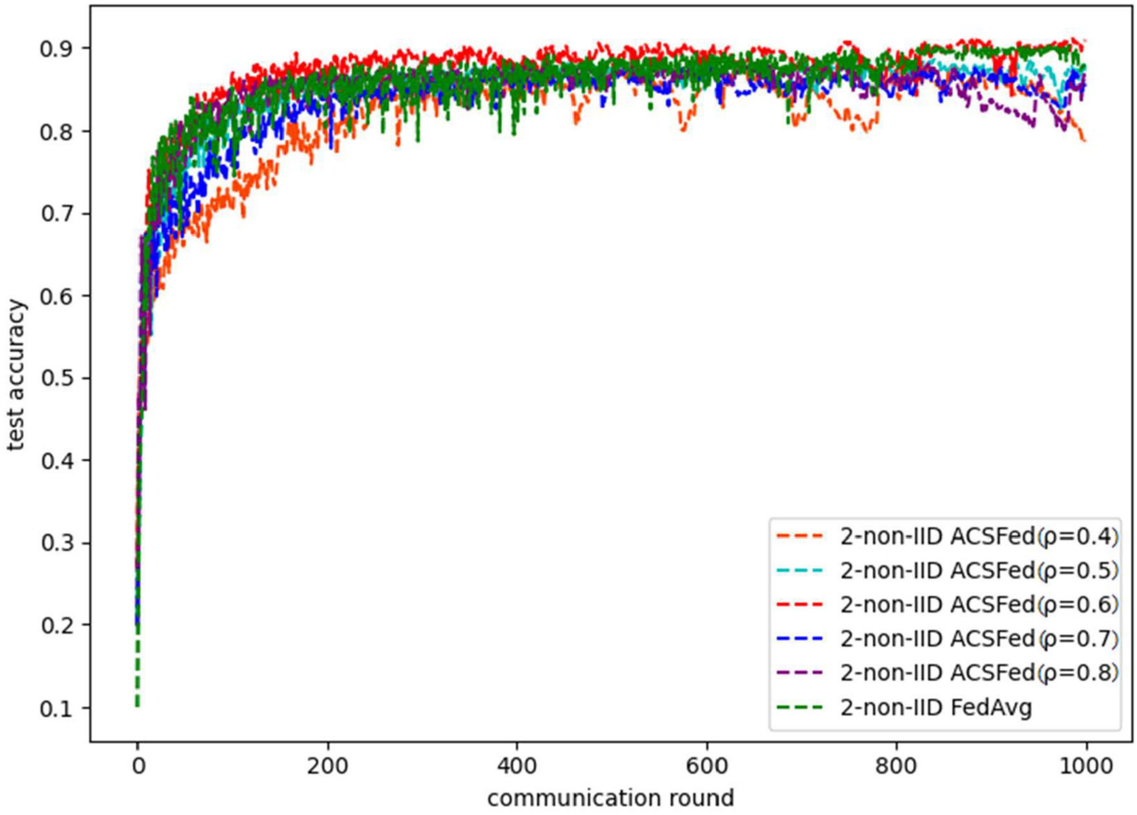

FIGURE 6 | Test accuracy on fashion MNIST in 2-class non-IID scenario.

the attenuation coefficient; EMD_i is the index of statistical heterogeneity on the client i, and N is the total number of clients. Therefore, a client's latest cumulative model strength depends on its history cumulative model strength and current training status. This calculation method uses an exponential-weighted average. It is widely applied in the optimization algorithms of deep learning, such as RSMProp (Tieleman and Hinton, 2012), which considers the history and current gradient information to adjust the learning rate of deep learning dynamically. In the exponential-weighted average, the attenuation coefficient ρ controls the acceptance ratio of history information. EMD_i is used to measure the degree of statistical heterogeneity in the client i by calculating the distance between the local data distribution q_i and the global data distribution p. Specifically, it is calculated as follows:

$$EMD_i(p, q_i) = \inf_{\gamma\tilde{}\prod(p,q_i)} \mathrm{E}_{(x,y)\tilde{}\gamma}[||x-y||_2] \tag{3}$$

where $\prod(p, q_i)$ is the set of all possible joint distributions that combine the distributions p and q_i. Then, for a specific joint distribution γ belonging to $\prod(p, q_i)$, a sample pair x and y can be sampled, and their distance can be calculated, thereby calculating $\mathrm{E}_{(x,y)\tilde{}\gamma}[||x-y||_2]$, which is the expectation of the distance between sample pairs under the joint distribution γ. Finally, the minimum expected value of the sample pair over all possible joint distributions is determined to be the EMD between distributions p and q_i.

At the beginning of training, the cumulative model strength of each client is 0, and the values in the probability matrix are initialized to $\frac{1}{N}$, indicating randomly selecting clients. After every communication round, clients participating in the training update the values of cumulative model strength and then send them to the central server.

Adaptive Probability Updating Method

In the central server, a set H_t is maintained, which contains the latest cumulative model strength of each client and is defined as follows:

$$H_t = \{h_t^{(1)}, h_t^{(2)}, ..., h_t^{(N)}\} \tag{4}$$

Only the cumulative model strength values of clients participating in the training are updated, while the rest remain unchanged. Then, the central server updates the probability matrix based on H_t which is:

$$matrix\_prob = matrix\_porb + \frac{EMD^{*}||Loss||_2^2}{\sqrt{H_t + \varepsilon}} \tag{5}$$

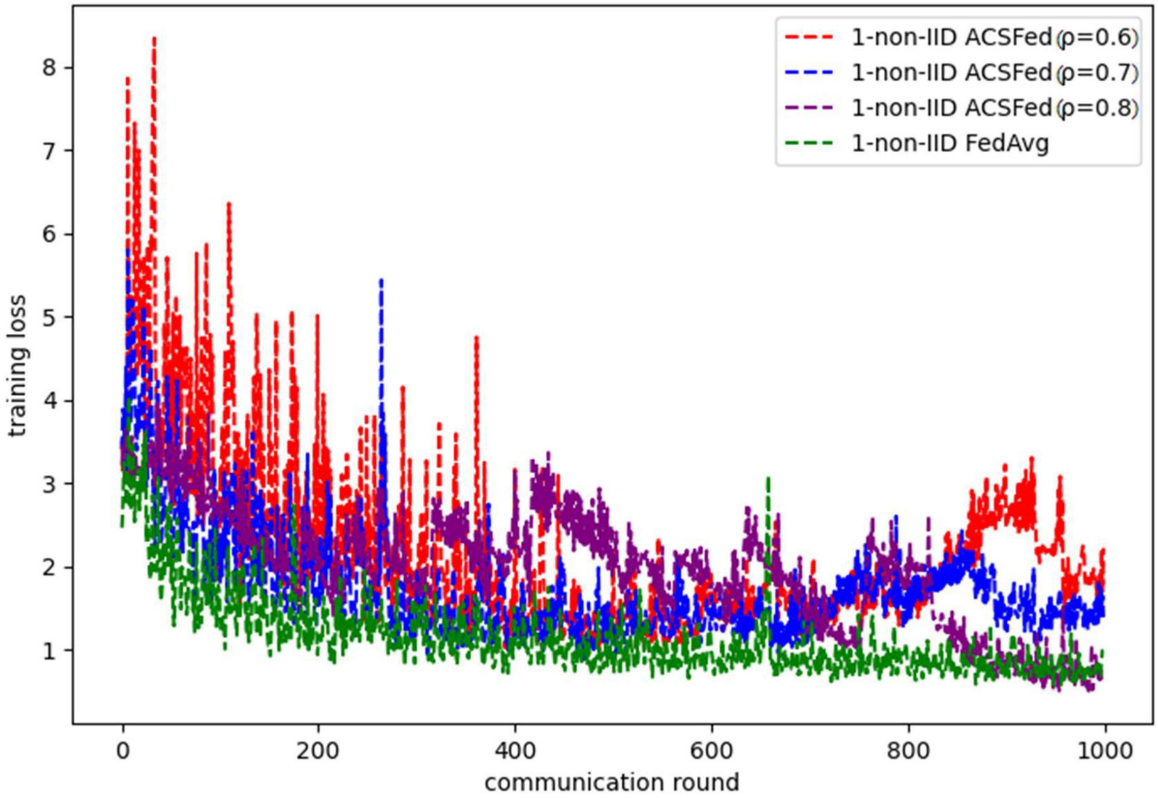

FIGURE 7 | Training loss on fashion MNIST in the 1-class non-IID scenario.

where ε is applied to avoid the situation in which the denominator is 0. As shown in Formula (10), the client with a larger cumulative model strength will have a smaller probability of being selected again because the model's performance on the client is sufficient. The probability of a client being selected can markedly increase when its cumulative model strength is 0, indicating that it has not participated in previous training and should be more likely to be chosen in the following training epoch. Finally, normalization is applied to ensure that the sum of all possible values in the matrix is equal to 1 and is calculated as:

$$matrix\_prob = \frac{matrix\_porb}{\sum_{i=1}^{N} matrix\_porb_i} \quad (6)$$

where $matrix\_porb_i$ is the client's probability i, and then the clients participating in the following training epoch are selected based on the normalized probability matrix. Based on the theory and design demonstrated above, the process of ACSF is shown in **Algorithm 1**.

After introducing the theory and design of the ACSFed algorithm, the theoretical analysis of ACSFed regarding the convergence of federated learning in statistical heterogeneity scenarios is performed and is described in the next section.

Analysis of the Convergence of ACSFed

In this section, the theoretical analysis of FedAvg and ACSFed in statistical heterogeneity scenarios is performed. In statistical heterogeneity scenarios, the data across the distributed clients is non-IID. Assuming that the data distribution of the client i is P_i when the data distribution across the clients is IID ($P_i = P$), the expectation of the local optimization object in any client is an unbiased estimate of the centralized optimization object $F_c(\omega)$:

$$\begin{aligned} E_{D_i\tilde{}P_i}[f_i(\omega, D_i)] &= F_c(\omega) \\ &= \frac{1}{|D|}\sum_{m=1}^{M} P(y=m)\cdot E_{x|y=m}[l(\omega, x)] \end{aligned} \quad (7)$$

Then, we calculate the expectation of the global federated learning optimization object $F_f(\omega)$ as follows:

$$\begin{aligned} &E[F_f(\omega)] = E[\textstyle\sum_{i=1}^{K}\frac{|D_i|}{|D|}\cdot\frac{1}{|D_i|}l(\omega, D_i)] \\ &= \frac{1}{|D|}E[\textstyle\sum_{i=1}^{K} l(\omega, D_i)] = \frac{1}{|D|}l(\omega, D) \\ &= F_c(\omega) \end{aligned} \quad (8)$$

Equation (8) shows that the expectation of federated learning optimizing object $F_f(\omega)$ is equal to centralized optimization

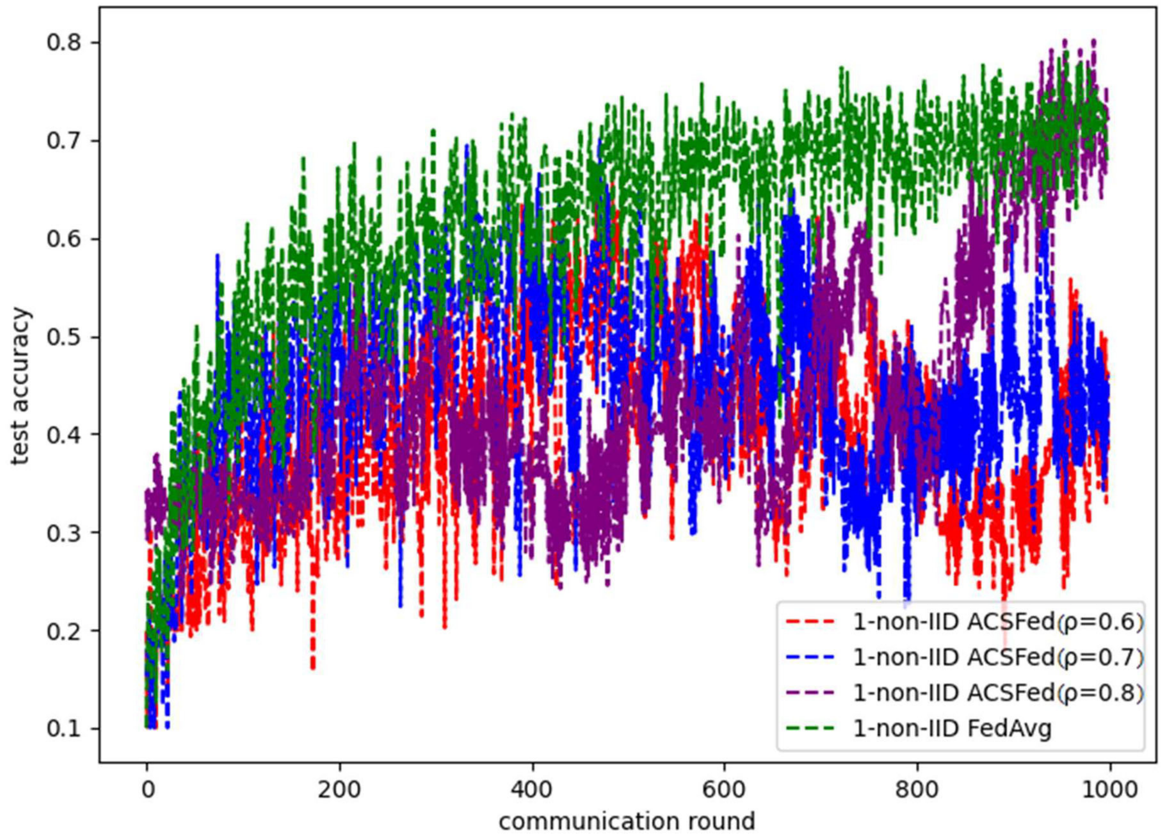

FIGURE 8 | Test accuracy on fashion MNIST in the 1-class non-IID scenario.

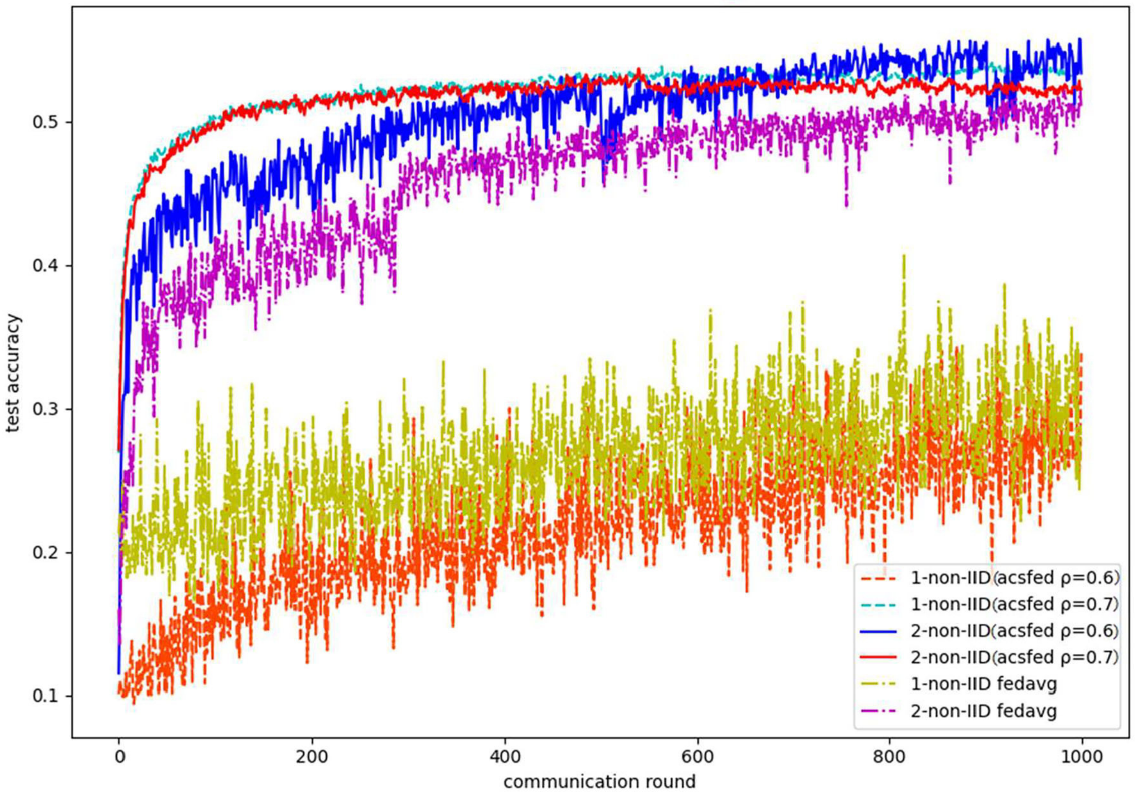

FIGURE 9 | Test accuracy on CIFAR 10 in the 1-class non-IID scenario.

object $F_c(\omega)$ when data across the clients is IID. Therefore, instead of enabling all clients to participate in training, only a portion of clients can be selected randomly with a specific fraction to execute local updates in the model training of federated learning because the local optimization object of any client is an unbiased estimate of $F_c(\omega)$. However, $f_i(\omega, D_i)$ could be an arbitrary approximation to $F_c(\omega)$ when P_i is different from P, leading to the deviation between the federated model and centralized-trained mode, and $E[F_f(\omega)]$ is no longer equal to $F_c(\omega)$. The convergence of the global model can be unstable due to the non-IID data because it is difficult for the model to learn the knowledge of the global data distribution by randomly selecting clients to participate in training. Therefore, statistical heterogeneity must be considered when selecting clients to participate in model training every communication round. The previous training performance should also be regarded to make the model learn the global knowledge faster. Specifically, the client with heavier statistical heterogeneity and poor training performance should be more likely to be selected to accelerate the convergence federated model, and this client choosing strategy is based on the research of Cho et al. (2020), which gives the theory of convergence of federated learning:

Theorem 5.1: The expectation of the difference between the training parameter and the convergence parameter is shown below, and the smaller it is, the faster convergence will be achieved:

$$E(F_f(\omega_t) - F_f^*) \leq \tfrac{1}{T+\gamma}[\tfrac{4L(32\tau^2G^2+\sigma^2/m)}{3\mu^2\bar{\rho}} + \tfrac{8L^2\Gamma}{\mu^2} + \tfrac{L\gamma||\omega_0-\omega^*||^2}{2}] + \xi \cdot (\tfrac{\tilde{\rho}}{\bar{\rho}} - 1) \tag{9}$$

where Γ is the difference between centralized optimization object and federated optimization object; γ, μ are parameters that control the learning rate; and the critical factor is $\bar{\rho}$, which is called Selection Skew, and it is defined as follows:

$$\overline{\rho} \underline{\underline{\Delta}} \min_{\omega_t} \frac{E_{S(\pi)}\left[\frac{1}{m}\sum_{k\in S(\pi)}\left(f_k(\omega_t, D_k) - f_k^*\right)\right]}{F_f(\omega_t) - \sum_{k=1}^{K} p_k f_k^*} \tag{10}$$

where $S(\pi)$ is the client selection strategy and assumes the strategy would select m clients to participate in the training; f_k^* is the final local optimizing object representing the training has converged and ω_t is the model parameter to be evaluated during training.

Then as can be concluded in ***Theorem 5.1***, the larger $\bar{\rho}$ is, the smaller the expectation of the difference between the training parameter and convergence parameter, indicating that faster convergence will be achieved. The client selection

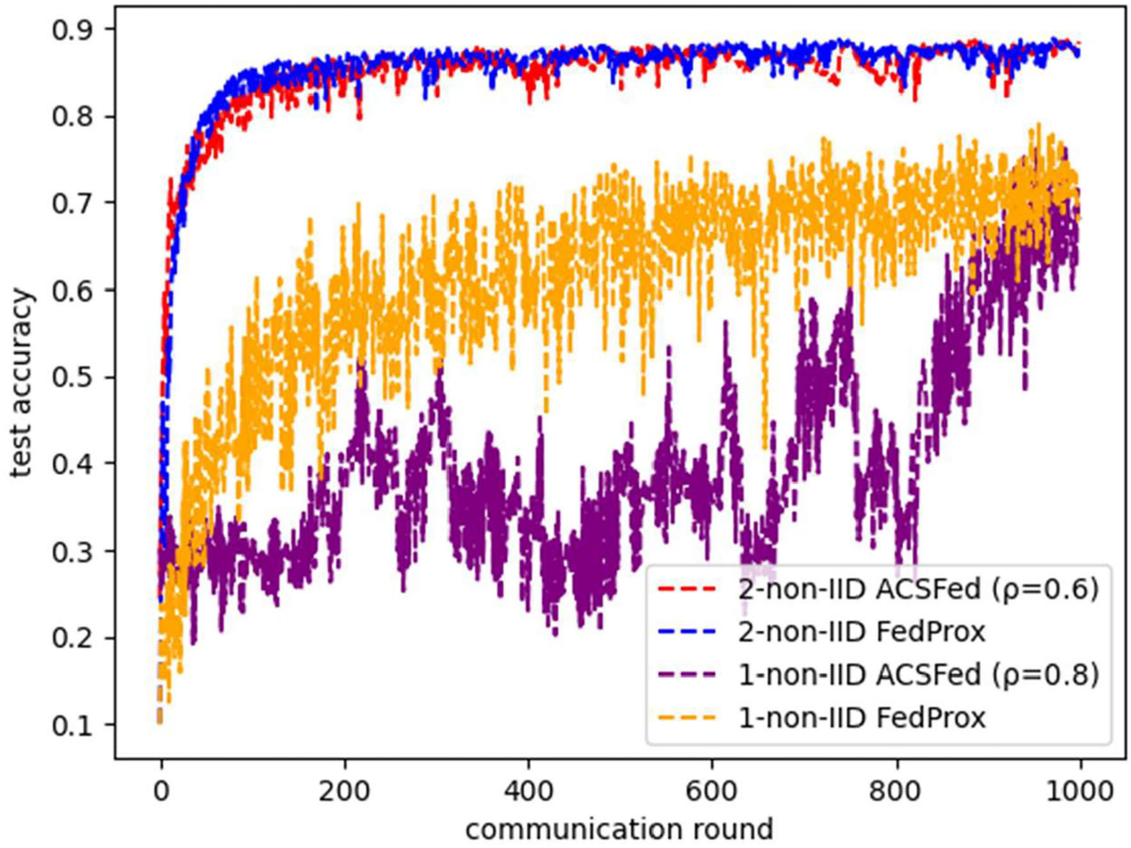

FIGURE 10 | Comparison results between ACSFed and FedProx on fashion MNIST.

strategy of ACFed would more likely choose clients with heavier statistical heterogeneity, worse training performance, or untrained operation; therefore, the $f_k(\omega_t, D_k)$ of each client will be much bigger than simply random selection, and thereby the $\bar{\rho}$ will be bigger. Since the strategy of ACSFed will make $\bar{\rho}$ bigger than the strategy that randomly selects clients, according to ***Theorem 5.1***, faster convergence will be obtained by ACSFed. After the theoretical analysis of ACSFed, performance experiments will be introduced in Section Evaluation.

EVALUATION

In this section, the results of the experiments are reported. Three public benchmark datasets are used for evaluation: MNIST, Fashion MNIST, and CIFAR 10, and the details of these datasets are shown in **Table 1**.

As shown in **Table 1**, the three datasets are all images, and MNIST and Fashion MNIST are grayscale images (with only one channel), while CIFAR 10 is a color image (with three channels). In terms of the models used to learn features and classification, a seven-layer convolutional neural network (CNN) is used in MNIST and Fashion MNIST, while a nine-layer CNN is used in CIFAR 10. For the baseline algorithm, FedAvg, which selects clients randomly in each communication round, is chosen for the performance comparison. Two statistical heterogeneity scenarios are set in the experiment, including a 2-class non-IID scenario (each client has two categories of images) and a 1-class non-IID scenario (each client has only one category of images). To determine the best attenuation coefficient ρ, several experiments are performed using different ρ. For the evaluation, two factors are selected (accuracy reduction and convergence time), and the results are shown in the following experimental figures and table.

As shown in **Figures 1**, **2**, ACSFed with attenuation coefficient $\rho = 0.7$ (blue curve) can achieve lower training loss and higher test accuracy than FedAvg on MNIST in the 2-class non-IID scenario. The trend of the curves in the two figures indicates that the ACSFed training loss decreases faster, and its test accuracy rises faster than FedAvg when $\rho = 0.7$.

Figures 3, **4** show the performance of ACSFed and FedAvg on MNIST in the 1-class non-IID scenario, which has much stronger statistical heterogeneity than the 2-class non-IID scenario; thus, the curves fluctuate markedly. Although the stronger statistical heterogeneity makes it difficult to train the global model, ACSFed $\rho = 0.7$ can have lower training loss and higher test accuracy than FedAvg.

In addition, the curves of ACSFed will fluctuate much more violently than FedAvg in the earlier communication round due to the client selection strategy of ACSFed that untrained clients and

Algorithm 1 : Adaptive client selection enabling federated learning (ACSFed). *K* clients are selected from *N* clients with fraction *c*; η is the learning rate, *P* is the probability matrix; *H* is the cumulative model strength matrix.

```
  Central Server:
  Initialize model $\omega_0$, $P$ and $H$
  $K \leftarrow \max(c \cdot N, 1)$
  while t in total communication rounds do:
    if t=0:
    client_set = {randomly selected $K$ clients}
    else:
    client_set = {select $K$ clients based on probability matrix $P$}
    for each client $i$ in client_set parallel do:
    Transmit $\omega_t$ to client $i$
    Receive $\omega^i_{t+1}$, $||loss_i||^2_2$, $emd_i$ from client $i$
    end for
    $H_i \leftarrow f(||loss_i||^2_2, emd_i)$
    $P \leftarrow g(P, H)$
  end while
  Distributed Client:
  Receive model $\omega_t$ from the central server
  Calculate $emd_i$ based on local data distribution
  Initialize loss list $loss_i = \{\}$
  $\Theta \leftarrow$ {split local data into batches with size $B$}
  for local epoch 1,2,..., E do:
    for $\theta \in \Theta$ do:
    Add training loss to$loss$
    $\omega_t \leftarrow \omega_t$-$\eta\nabla g(\omega_t, \theta)$
  end for
end for
$\omega^i_{t+1} \leftarrow \omega_t$
Calculate$||loss_i||^2_2$
Transmit $\omega^i_{t+1}$ $||loss_i||^2_2$ and $emd_i$ to Central Server
```

clients with stronger statistical heterogeneity or worse training performance are more likely to be selected, which can feed the model with much more unknown knowledge compared with selecting clients randomly in earlier communication round. However, the ACSFed's loss curve will decrease faster. Its training accuracy curve will increase faster than FedAvg after an earlier communication round as the global model of ACSFed already learned relatively more knowledge than FedAvg in earlier training.

Similar results in two statistical heterogeneity scenarios can be obtained on Fashion MNIST. In the 2-class non-IID scenario, which is shown by **Figures 5**, **6**, ACSFed with ρ=0.6 finally achieves a lower training loss and a higher test accuracy than FedAvg. In addition, the client selection strategy will also make the convergence of ACSFed weaker than FedAvg in earlier training rounds. Still, the performance of ACSFed will exceed FedAvg after certain training rounds.

However, because Fashion MNIST is more complex than MNIST, ACSFed will take much more time to learn sufficient knowledge about the data in the 1-class non-IID scenario compared to the 2-class non-IID. Results are shown in **Figures 7**, **8**, where ACSFed with ρ=0.8achieves better performance than FedAvg. In the first 800 rounds of training, the training loss and test accuracy curves of ACSFed fluctuate much more violently than FedAvg. Still, then, a lower loss and higher accuracy than FedAvg is achieved rapidly after 800 rounds of training.

TABLE 2 | Training performance of ACSFed and FedAvg.

Dataset	**Method**	**ToA@0.6 (2/1 class non-IID)**	**ToA@0.8 (2/1 class non-IID)**
MNIST	ACSFed	3/25	6/72
	FedAvg	5/27	9/NULL
Fashion MNIST	ACSFed	5/108	32/946
	FedAvg	8/107	35/NULL
–	**–**	**ToA@0.3 (2/1 class non-IID)**	**ToA@0.5 (2/1 class non-IID)**
CIFAR 10	ACSFed	12/381	235/NULL
	FedAvg	20/674	927/NULL

Finally, **Figure 9** shows the results on CIFAR10 in 2-class and 1-class non-IID scenarios, which indicate that ACSFed ρ=0.7 has better performance than FedAvg in two types of statistical scenarios. Moreover, as lite-structured CNN is deployed in the experiment, the accuracy of CIFAR 10 in both scenarios is relatively low. (Complex networks can obtain ideal performance, such as ResNet 18 or ResNet 50, however, these networks are not suitable for a resource-constrained experimental environment.).

Moreover, the comparison result between ACSFed and FedProx is illustrated in **Figure 10**. It can be concluded that ACSFed can obtain nearly the same promotion in the efficiency of federated learning in both two kinds of statistical heterogeneous scenarios. However, the computational complexity of ACSFed is much smaller than FedProx, as only two types of float digital calculations are added in ACSFed. At the same time, FedProx needs to optimize the divergence between the local model and the global model in the loss function. Therefore, ACSFed can obtain the same performance as FedProx while adding less computational complexity.

Finally, based on these experimental results, the more complex the data are and the stronger statistical heterogeneity is, the more time ACSFed will take to fluctuate in both loss and accuracy; we refer to this time as the fluctuation period, which is due to the client selection strategy of ACSFed. Specifically, the loss and accuracy curve of ACSFed fluctuates more violently than FedAvg in the fluctuation period, indicating that the model keeps learning unknown knowledge. Then, the performance of ACSFed can exceed FedAvg rapidly after the fluctuation period because the model has already learned sufficient knowledge about the non-IID data.

In addition, the ToA@x (the epoch to reach the accuracy of x) of ACSFed and FedAvg are recorded to demonstrate the improvement in training efficiency more clearly, and results are shown in **Table 2**, where NULL means that the model cannot achieve a certain accuracy after training.

Based on **Table 2**, ACSFed can achieve a certain accuracy (60 and 80%) more quickly than FedAvg in 1- and 2-class non-IID. In particular, FedAvg cannot achieve an accuracy of 80% in the 1-class non-IID scenario on the three datasets. In comparison, ACSFed can still achieve 80% after a certain communication round of training on MNIST and Fashion MNIST, indicating that the implementation of ACSFed can promote the performance of federated learning in statistical heterogeneity scenarios.

CONCLUSION

Federated learning will play an essential role in the future as the computation capability of remote edge devices increases and local data privacy increases. However, statistical heterogeneity can markedly affect the efficiency of federated learning methods and lead to unstable convergence. In this study, we proposed an adaptive client selection algorithm for federated learning called ACSFed to promote the performance of federated learning. Experiments on three different datasets demonstrate the improved performance of ACSFed compared to the current federated learning method. Additionally, ACSFed adds little computation and communication burden because updating the probability matrix of clients is simple. However, there is still the possibility of promoting the performance of federated learning in statistical heterogeneity scenarios by combining ACSFed with methods that focus on global model aggregation. The challenge of model protection still exists, and further research is required to address this and related challenges to apply federated learning methods more effectively.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://yann.lecun.com/exdb/mnist/; https://tensorflow.google.cn/datasets/catalog/fashion_mnist; http://www.cs.utoronto.ca/~kriz/cifar.html.

AUTHOR CONTRIBUTIONS

AC: conceptualization, methodology, writing—review and editing, and supervision. YF: conceptualization, methodology, data curation, formal analysis, software, and writing—original draft preparation. ZS: conceptualization, methodology, project administration, and investigation. GL: supervision and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China (NSFC) (U19A2059), and Sichuan Science and Technology Program (No. 206999977).

REFERENCES

Asad, M., Moustafa, A., and Ito, T. (2020). FedOpt: towards communication efficiency and privacy preservation in federated learning. *Appl. Sci.* 10, 2864. doi: 10.3390/app10082864

Cai, Z., and He, Z. (2019). Trading private range counting over big IoT data. In: *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 144–153.

Cai, Z., and Zheng, X. (2018). A private and efficient mechanism for data uploading in smart cyber-physical systems. *IEEE Trans. Network Sci. Eng.* 7, 766–775. doi: 10.1109/TNSE.2018.2830307

Cai, Z., Zheng, X., Wang, J., and He, Z. (2019). Private data trading towards range counting queries in internet of things. In: *IEEE Transactions on Mobile ComputiCng, 2022*.

Chen, W., Horvath, S., and Richtarik, P. (2020). Optimal client sampling for federated learning. *arXiv [preprint]*. arXiv:2010.13723. doi: 10.48550/arXiv.2010.13723

Cho, Y. J., Wang, J., and Joshi, G. (2020). Client selection in federated learning: convergence analysis and power-of-choice selection strategies. *arXiv [preprint]*. arXiv:2010.0124. doi: 10.48550/arXiv.2010.01243

Huang, L., Yin, Y., Fu, Z., Zhang, S. F., Deng, H., and Liu, D. B. (2020). LoAdaBoost: loss-based AdaBoost federated machine learning with reduced computational complexity on IID and non-IID intensive care data. *PLoS ONE* 15, e0230706. doi: 10.1371/journal.pone.0230706

Jeong, E., Oh, S., Kim, H., Park, J., Bennis, M., and Kim, S. (2018). Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data. *arXiv [preprint]*. arXiv:1811.11479. doi: 10.48550/arXiv.1811.11479

Konečný, J., McMahan, B., and Ramage, D. (2015). Federated optimization: distributed optimization beyond the datacenter. *arXiv [preprint]*. arXiv:1511.03575. doi: 10.48550/arXiv.1511.03575

Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., and Bacon, D. (2016). Federated learning: strategies for improving communication efficiency. *arXiv [preprint]*. arXiv:1610.05492. doi: 10.48550/arXiv.1610.05492

Li, D., and Wang, J. (2019). Fedmd: Heterogenous federated learning via model distillation. *arXiv [preprint]*. arXiv:1910.03581. doi: 10.48550/arXiv.1910.03581

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., and Smith, V. (2020). "Federated optimization in heterogeneous networks," in *Proceedings of Machine Learning and Systems* (Austin, TX), 429–450.

Li, X., Huang, K., Yang, W., Wang, S. S., and Zhang, Z. H. (2019). On the convergence of fedavg on non-iid data. *arXiv [preprint]*. arXiv:1907.02189. doi: 10.48550/arXiv.1907.02189

McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. Y. (2017a). "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (Ft. Lauderdale, FL), 1273–1282.

McMahan, B., Ramage, D., and Research Scientists. (2017b). Federated learning: Collaborative machine learning without centralized training data. *Google AI Blog*. Available online at: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Nishio, T., Shinkuma, R., Takahashi, T., and Mandayam, N. (2013). Service-oriented heterogeneous resource sharing for optimizing service latency in mobile cloud. In: *Proceedings of the First International Workshop on Mobile Cloud Computing & Networking*, 19–26.

Nishio, T., and Yonetani, R. (2019). Client selection for federated learning with heterogeneous resources in mobile edge. In: *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 1–7.

Pang, J., Huang, Y., Xie, Z., Hang, Q. L., and Cai, Z. P. (2020). Realizing the heterogeneity: a self-organized federated learning framework for IoT. *IEEE Internet Things J.* 8, 3088–3098. doi: 10.1109/JIOT.2020.3007662

Sardellitti, S., Scutari, G., and Barbarossa, S. (2015). Joint optimization of radio and computational resources for multicell mobile-edge computing. *IEEE Trans. Signal Inform. Process. Over Netw.* 1, 89–103. doi: 10.1109/TSIPN.2015.2448520

Sattler, F., Wiedemann, S., Müller, K. R., and Samek, W. (2019). Robust and communication-efficient federated learning from non-iid data. *IEEE Trans. Neural Networks Learn. Syst.* 31, 3400–3413. doi: 10.1109/TNNLS.2019.2944481

Shen, G., Gao, D., Yang, L., Zhou, F., Song, D. X., Lou, W., et al. (2022). Variance-reduced heterogeneous federated learning via stratified client selection. *arXiv [preprint]*. arXiv:2201.05762. doi: 10.48550/arXiv.2201.05762

Tieleman, T., and Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA*. 4, 26–31.

Wang, H., Kaplan, Z., Niu, D., and Li, B. C. (2020). Optimizing federated learning on non-iid data with reinforcement learning. In: *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 1698–1707.

Wang, S., Tuor, T., Salonidis, T., Leung, K., Makaya, C., He, T., et al. (2019). Adaptive federated learning in resource constrained edge computing systems. *IEEE J. Select. Areas Commun.* 37, 1205–1221. doi: 10.1109/JSAC.2019.2904348

Xiong, Z., Cai, Z., Takabi, D., and Li, W. (2021). Privacy threat and defense for federated learning with non-iid data in AIoT. *IEEE Trans. Indus. Inform.* 18, 1310–1321. doi: 10.1109/TII.2021.3073925

Yu, Y., Zhang, J., and Letaief, K. B. (2016). Joint subcarrier and CPU time allocation for mobile edge computing. In: *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 1–6.

Zhang, M., Sapra, K., Fidler, S., Yeung, S., and Alvarez, J. M. (2020). Personalized federated learning with first order model optimization. *arXiv [preprint]*. arXiv:2012.08565. doi: 10.48550/arXiv.2012.08565

Zhang, S. Q., Lin, J., and Zhang, Q. (2022). A multi-agent reinforcement learning approach for efficient client selection in federated learning. *arXiv [preprint]*. arXiv:2201.02932. doi: 10.48550/arXiv.2201.02932

Zhang, W., Wang, X., Zhou, P., Wu, W., and Zhang, X. (2021). Client selection for federated learning with non-iid data in mobile edge computing. *IEEE Access* 9, 24462–24474. doi: 10.1109/ACCESS.2021.3056919

Zhao, J., Feng, Y., Chang, X., Chang, X., and Liu, C. (2022). Energy-efficient client selection in federated learning with heterogeneous data on edge. *Peer Peer Netw. Appl.* 15, 1–13. doi: 10.1007/s12083-021-01254-8

Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., and Chandra, V. (2018). Federated learning with non-iid data. *arXiv [preprint]*. arXiv:1806.00582. doi: 10.48550/arXiv.1806.00582

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.