LG-FAL: Locality-customized GSA Federated Active Learning

Shuwen Wang
Department of Computer Science
Portland State University
Portland, USA
shuwen@pdx.edu

Xingquan Zhu
Dept. of Electrical Engineering & Computer Science
Florida Atlantic University
Boca Raton, USA
xzhu3@fau.edu

***Abstract*—Federated Learning (FL) is a learning paradigm which constructs machine learning models using decentralized datasets, with the goal of training a globally optimized model while preserving data privacy. In practice, annotating training data for FL is expensive and time-consuming. To solve this problem, researchers proposed to integrate active Learning (AL) strategy as Federated Active Learning (FAL) framework. However, existing research on FAL has two main limitations: (1) In the active learning part, more attention is paid to the global model, while ignoring the local information; (2) In the federated learning part, the popular Federated averaging (FedAvg) method relies on the assumption that the corresponding nodes in distributed neural networks share the same importance when averaging. In this paper, in order to tackle the two limitations, we propose a locality-customized GSA federated active learning (`LG-FAL`) method. Specifically, (i) both local and global information are taken into consideration to evaluate and select informative data samples to annotate for active learning; (ii) we integrate Gravitational Search Algorithm (GSA) to dynamically average local network parameters into the global parameter, by mimicking the principles of gravity and motion in the universe, for effective federated learning. As a result, `LG-FAL` can select a small subset of informative samples considering both local and global information, and at the same time, provide an improved trade-off between communication cost and learning accuracy. Experimental results show that `LG-FAL` significantly outperforms the current state-of-the-art baselines in terms of performance and effectiveness.**

***Keywords*—Federated active learning, active learning, parameter aggregation, gravitational search algorithm (GSA).**

I. Introduction

Federated Learning (FL) represents a novel learning approach that constructs machine learning models using decentralized datasets distributed across numerous sites/devices. The feasibility of Federated Learning (FL) as a decentralized machine learning approach heavily relies on the proficiency of local models in both training and inference tasks. These local models' effectiveness is contingent upon the availability of meaningful and annotated data, which is essential for their successful training [1]–[3]. However, obtaining such data involves a laborious and time-consuming annotation process, necessitating manual analysis of the training samples. In the field of machine learning, data annotation plays a pivotal role in empowering models with the capacity to generalize effectively and achieve high-performance levels. However, it presents two significant challenges. First, it demands meticulous and time-consuming analysis for each sample, rendering it a laborious endeavor. Second, and perhaps more critically, the selection of appropriate samples is not always guaranteed, resulting in potential negative impacts on the overall performance of the model [1], [4]–[6].

Recently, Active Learning (AL) has emerged as a machine learning method that can effectively address data annotation workloads [7], [8]. Its main strategy is to iteratively find the most informative data points to annotate. The annotated data are then used as part of the training data in the next iteration. With more and more iterations, the machine learning model's performance can be more and more improved. This strategy has been integrated into federated learning and generated a new paradigm called Federated Active Learning (FAL) [9]–[12]. The Federated Active Learning (FAL) framework comprises multiple clients and a central server. Each client maintains a labeled and an unlabeled dataset, which are not shared, while the server holds a shared test dataset. FAL's objective is to iteratively train a globally optimized model at the server by annotating high-value data samples at the client level. In each iteration, clients first train local models, share their parameters with the server, and receive an updated global model. This model is then used to identify and label the most informative samples, gradually enhancing the global model's performance over successive iterations as more data is annotated.

However, current FAL has two significant weaknesses: (i) In most FAL, local unlabelled samples are annotated by the aggregated global model's parameters, which totally ignores the localization of the samples, furthermore, the importance of local models for local sample annotation is completely ignored [13]. (ii) Its global model parameter updating is limited to one method, which is called Federated average (FedAvg) [1], [10], [14], [15]. FedAvg relies on the assumption that the corresponding nodes in local neural networks share the same importance when averaging, while different local models should have different average weights [2].

To tackle the first weakness, we propose a locality-customized annotation strategy, which takes the local model into consideration aside from the global model when annotating. There are two reasons to pay attention to the local model:

979-8-3315-0882-1/24/$31.00 ©2024 IEEE
DOI 10.1109/ICKG63256.2024.00056

(i) local models compose the global model; (ii) the annotated data are directly used to train local models. Specifically, we first predict all unlabeled data's labels by the local model and the global model separately. Then, we calculate the uncertainty of the prediction by the metric of entropy. Each data's overall uncertainty is a combination of both the local model's prediction entropy and the global model's prediction entropy. Finally, we annotate the top K data with the highest informativeness.

To tackle the second weakness, we propose a Gravitational Search Algorithm (GSA) based FAL framework. Different from FedAvg, global model parameter aggregations are achieved by GSA which draws inspiration from the law of gravity and the interactions between celestial bodies. GSA allows population diversity as well as global exploration, which means FL clients can interact with each other based on their masses (accuracy) and positions (local model parameters), at the same time, GSA is capable of exploring the solution space globally by allowing clients to move freely towards areas of high fitness calculated based on their masses (accuracy). Moreover, it is empowered with enhanced adaptability through a set of parameters that control the interaction between clients. Essentially, the GSA method can be viewed as a weighted averaging strategy where the mass plays the role of weight.

To summarize, in this paper, we propose a locality-customized GSA federated active learning (LG-FAL) method. The main contributions of the proposed research are: (i) We propose a new annotating strategy that considers both local and global optimization. By doing so, the localization of samples and models can be considered; (ii) We propose to update the global model parameters with GSA, in which the model is updated in a more interactive and adaptable way; (iii) We design extensive experiments to validate the proposed methods with different parameter settings and comparisons.

II. Related Work

A. Federated Active Learning

A novel approach is designed to improve the classification accuracy of waste and natural disaster images using a combination of Active learning and Federated learning techniques. The approach utilizes Active learning to select the most informative and relevant data samples for labeling, reducing the labeling workload. These labeled samples are then utilized in a Federated learning setting, where multiple devices collaborate to train a shared model without sharing raw data centrally, which effectiveness has been demonstrated in achieving higher classification accuracy compared to traditional federated learning approaches [1]. Chen, et al, designed a novel Federated Evidential Active Learning (FEAL) methodology. which integrates Dirichlet-based evidential modeling to address domain shifts in medical data across different institutions, enhancing data annotation efficiency and model reliability through calibrated uncertainty assessment and diversity relaxation strategies [16]. F-AL is proposed as a novel annotation strategy to enhance Federated Learning (FL) by leveraging active learning to address the challenge of limited annotated data in FL scenarios. By incorporating active learning techniques, F-AL aims to intelligently select and query the most informative data samples from each client's local dataset, reducing the annotation burden and improving the performance of the global model. The paper presents the evaluation of F-AL, highlighting its potential benefits in promoting more effective and privacy-preserving FL implementations [10]. A semi-supervised and personalized framework that combines active learning and label propagation techniques is proposed. In this method, leverages unlabelled data from individual clients in the federated environment to enhance the activity recognition process. Active learning is used to intelligently select the most informative samples for labeling, reducing the labeling effort while improving the model's accuracy [17]. Federated Active Learning with a focus on inter-class diversity is explored by introducing novel methodologies to improve the performance of Active learning in a Federated learning setting. By taking into account the diversity among different classes of data, the authors propose innovative techniques that enhance the selection of informative samples for labeling during the active learning process. This approach is aimed at improving the overall performance of the federated learning model while reducing the labeling effort required from individual clients [11]. A novel framework for enhancing intrusion detection in Zero-Trust Security Models (ZSM) using federated learning and semi-supervised active learning techniques is created, which incorporates semi-supervised active learning to optimize the model by selectively labeling the most informative data samples, thus reducing the reliance on fully labeled data. The paper highlights the effectiveness of this combined approach in improving intrusion detection performance and addresses challenges related to data privacy and isolation in ZSM environments [15]. Wu Xing, et al, propose a framework that combines Federated Learning and Active Learning to improve disease diagnosis accuracy while preserving data privacy in a multi-center scenario. Federated Learning enables multiple medical centers to collaborate and train a shared model without sharing raw patient data. Active Learning is incorporated to intelligently select the most informative and relevant data samples from each center for labeling, reducing the need for extensive labeled data. It is evaluated on a multi-center dataset, showcasing its effectiveness in achieving higher diagnostic accuracy compared to traditional methods [9].

B. Federated Learning Parameter Aggregation

With the growing emphasis on data privacy protection, Federated Learning has emerged as a highly popular research area. Numerous studies have introduced innovative approaches concerning weight updates in Federated Learning. Building upon the previous methodology, a novel approach called Multi-Center Federated Learning, which seeks to enhance personalization by clustering clients according to their data distributions has been proposed. In this approach, the multi-center aggregation mechanism involves combining local models from multiple centers to form a global model. Each center trains its respective local model using data from clients within its cluster. The local models are subsequently aggregated through information exchange between the centers, facilitating the improvement of the global model [18]. The convergence challenges of federated learning (FL) when using dynamically reduced models across various devices are explored in which the importance of a minimal coverage index is emphasized and model reduction noise in achieving efficient and reliable federated learning outcomes [19].A new framework called KAFAL introduces a novel federated learning approach for non-IID data with constrained annotation budgets. The framework's first component, Knowledge-Specialized Active Sampling (KSAS), actively selects samples using a modified KL-Divergence to prioritize data that is informative for both local and global models. The second component, Knowledge-Compensatory Federated Update (KCFU), mitigates data heterogeneity by distilling knowledge from the global model to clients, especially for underrepresented classes. Together, KSAS and KCFU enable effective sample selection and federated updates, optimizing annotation usage and improving model convergence in decentralized learning scenarios [20].

A novel approach called FairFed is presented, which addresses the issue of group fairness in Federated Learning (FL). It aims to mitigate biases and disparities that might arise during the learning process, promoting fair representation and performance across various user groups. The experiments and results presented in the paper demonstrate the effectiveness of FairFed in achieving group fairness in federated learning scenarios [21]. A new FL-empowered semi-supervised active learning (FL-SSAL) framework was designed for security orchestration in a Label-at-Client scenario. In this approach, clients work with a mix of unlabeled and a small amount of labeled data. The framework uses entropy-based active learning to identify the most informative samples for labeling and applies a semi-supervised approach to make use of unlabeled data. Experimental evaluations on a private, non-independent and identically distributed (non-IID) dataset show that FL-SSAL improves intrusion detection accuracy. Additionally, it reduces communication overhead compared to baseline models, even with limited labeled data [15]. An Auditable Privacy-Preserving Federated Learning (AP2FL) model addresses Non-IID data by incorporating Active Personalized Federated Learning (ActPerFL) and Batch Normalization (BN) techniques, enabling effective user update consolidation and data similarity identification. An auditing mechanism further enhances AP2FL, revealing individual client contributions and ensuring the global model adapts to diverse data types and distributions [22].

III. Proposed Approach

Our proposed Locality-customized GSA Federated Active Learning (`LG-FAL`) integrates Active Learning (AL) and the Gravitational Search Algorithm (GSA) within a federated learning framework. This combination targets an efficient trade-off between communication costs and learning accuracy by selecting highly informative subsets of local data.

A. Framework Overview

The locality-customized GSA Federated Active Learning (`LG-FAL`) framework operates iteratively in two main phases, as summarized in Fig. 1. Each client trains its model on local annotated data, learning client-specific patterns. Clients send their model parameters to the central server. The server aggregates parameters using the Gravitational Search Algorithm (GSA), weighting each client's contribution based on its model's accuracy. The updated global model is sent back to clients, incorporating insights from across the network. Each client scores unlabeled samples using both the local and global models, selecting the most informative samples for annotation. This cycle repeats, iteratively refining both local and global models for balanced learning between local specificity and global generalization.

B. Locality-customized Annotation

Our approach introduces a dual-model annotation strategy that utilizes both local and global models to assess the informativeness of unlabeled data samples:

a) Local Model Consideration: For each local client i, model trained using its own local data is defined as $\mathcal{M}_i$. Local model $\mathcal{M}_i$ enables customization and adaptation to specific local device characteristics and local data patterns to better fit unique local data distributions, which is able to capture different facets of the local data distribution and localization.

b) Global Model Integration: Global model $\mathcal{M}_{FL}$ is the central model that is shared and iteratively updated across a network of decentralized $\mathcal{M}_i$ in FL. During the training process, locally trained models $\mathcal{M}_i$ send back their parameter updates to a central server, which aggregates these updates to refine the global model $\mathcal{M}_{FL}$. By combining data informativeness from both $\mathcal{M}_i$ and $\mathcal{M}_{FL}$, we are able to capture the data generalization while maintaining its localization at the same time.

c) Local Active Learning: We introduce a locality-customized annotation function $\mathcal{AL}$. The $\mathcal{AL}$ strategy combines the predictive uncertainties from both the local and global models to score each unlabeled sample. For each local client i, both local labeled dataset $\mathcal{D}_i$ as well as newly-annotated dataset A_i by $\mathcal{AL}$ will be set up for the training. We design a score function $S(x)$ to evaluate unlabeled samples. The strategy is to annotate data samples with the highest score in the unlabelled data as shown in (1), where z is the sampling number and $S(x)$ is the score function of x.

$$\mathcal{AL}_i = \underset{|A_i|=z, x\in\mathcal{U}_i}{\operatorname{argmax}} \; S(x) \tag{1}$$

To make sure that the score function is able to reflect the localization and potential informativeness of local unlabelled instances, we introduce the score function as shown in (2). Global model $\mathcal{M}_{FL}$ and local model $\mathcal{M}_i$ are allowed to predict the labeling possibilities of samples. The most informative query is considered to be the instances about which they most agree. The sample informativeness from both global model

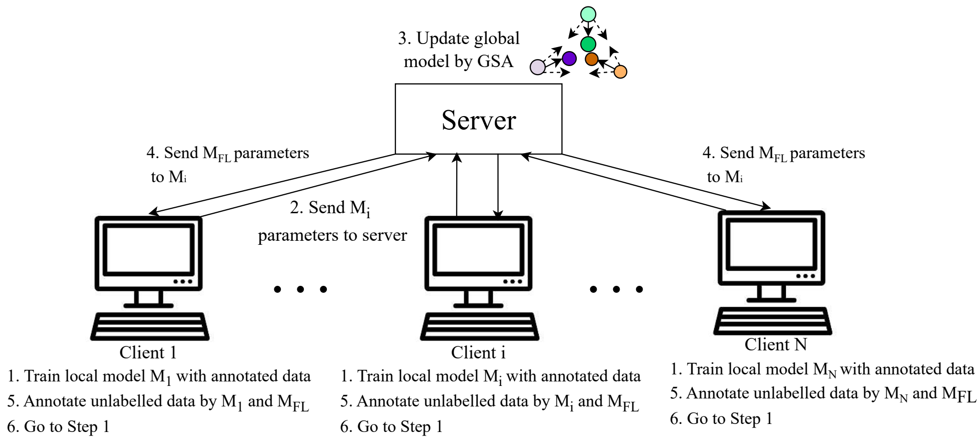


Fig. 1: Framework of LG-FAL. Clients train local models and send them to the server. The server synthesizes the models and gets a global model. The global model is sent to each client to help annotate local data.

$\mathcal{M}_{FL}$ and local model $\mathcal{M}_i$ is integrated to find the average score of a sample.

$$S(x) = w_1 * \text{Entr}(Dis(x|\mathcal{M}_i)) + w_2 * \text{Entr}(Dis(x|\mathcal{M}_{FL})) \quad (2)$$

where $Dis(x|M)$ denotes the prediction distribution of x under model M; $Entr(Dis(x|M))$ denotes the entropy of the distribution. w_1 and w_2 are weights between 0 and 1.

d) Annotation Process: Samples with the highest combined uncertainty are selected for annotation. The higher the entropy is, the more uncertain the sample under the model will be. Thus, the active learning strategy prefers to annotate samples with high uncertainty. AL adopts multiple rounds as the FL goes on for sampling and gradually adds samples to the labeled local dataset.

C. GSA-based Federated Learning

We integrate Gravitational Search Algorithm (GSA) with federated learning to obtain a globally optimized model from local models. Within GSA's iterative framework, each local model is viewed as one object with mass (importance), while its parameter values are viewed as position coordinates. These objects attract each other due to gravity, prompting their movement towards heavier masses, which correspond to favorable solutions [23]. Fig.2 shows the movement of the object.

Assume there are N clients participating FL, each of whose local model has a D dimensional parameter vector denoted as (3), where x_i^d is the parameter of the ith agent in dimension d.

$$X_i = (x_i^1, ..., x_i^d, ..., x_i^D) \qquad i = 1, ..., N \quad (3)$$

a) Mass Calculation: The gravitational mass of each object using the fitness values is calculated as in (4) and (5). The gravitational mass is denoted as $M_i(t)$ and $fit_i(t)$ indicates the fitness value of the ith object at iteration t, which, in our method, the client's predictive accuracy on the test dataset is used as the fitness value.

$$m_i(t) = \frac{\text{fit}_i(t) - \text{worst}(t)}{\text{best}(t) - \text{worst}(t)} \quad (4)$$

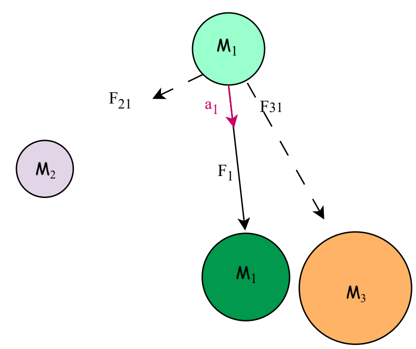


Fig. 2: Demonstration of object movement with GSA. Object M_1 is attracted by M_2 and M_3, with gravity force F_{21} and F_{31}. The total force F_1 results in acceleration a_1 to update the position of M_1, which equals to update the parameter vector X_1.

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \quad (5)$$

In addition, $worst(t)$ and $best(t)$ are the worst and best fitness values obtained in the collection of objects at t which are defined for maximization problem as in (6) and (7) respectively.

$$\text{worst}(t) = \min fit_j(t) \qquad j \in \{1, ..., N\} \quad (6)$$

$$\text{best}(t) = \max fit_j(t) \qquad j \in \{1, ..., N\} \quad (7)$$

b) Gravitation Interaction: The total force that is applied on the ith object from other objects is computed following the gravity law in (8).

$$F_i^d(t) = \sum_{j \in Kbest, j \neq i} \text{rand}_j^d G(t) \frac{M_j(t) \times M_i(t)}{R_{ij} \times \epsilon} (x_j^d(t) - x_i^d(t)) \quad (8)$$

in which rand_j^d is a random number with uniform distribution in the interval [0, 1]. ϵ denotes a small number close to 0, $R_{ij}(t)$ denotes the Euclidean distance between clients i and j, and $Kbest$ is a set consisting of the first K objects with the best fitness values (the largest masses).

K is set as N at the very beginning and reduces linearly with time until it reaches to 1 in the end. The gravitational constant at iteration t is denoted as $G(t)$ which is initialized at the first iteration by G_0 and decreased by time according to (9), where T is the total number of iterations.

$$G = G_0 \exp^{-\alpha \frac{t}{T}} \tag{9}$$

Then, the obtained force is used to calculate the acceleration of the object using the law of motion as in (10).

$$\begin{aligned} a_i^d(t) &= \frac{F_i^d(t)}{M_i(t)} \\ &= \sum_{j \in Kbest, j \neq i} \text{rand}_j G(t) \frac{M_j(t)}{R_{ij}(t) \times \epsilon} (x_j^d(t) - x_i^d(t)) \end{aligned} \tag{10}$$

c) Parameter Update: The next movement for ith object can be computed based on the change of its acceleration as in (12) and this is the end of one GSA iteration.

$$v_i^d(t+1) = v_i^d(t) + a_i^d(t), \qquad v_i^d(0) = 0 \tag{11}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{12}$$

After a certain number of iterations, all parameter vectors X_i are updated with other vectors' information. In other words, all X_i can be viewed as candidates of the aggregated global model parameters. We test them on the test dataset, and consider the parameter with the best performance as the parameter of the global model $\mathcal{M}_{FL}$.

Training of a global GSA algorithm is performed in an iterative fashion. It communicates with local ones iteratively since the stopping criterion is reached. Each client initially starts with a randomized model that is the exact same structure as the central model. The pseudo-code of the `LG-FAL`is shown in Algorithm 1.

Algorithm 1 Locality-customized GSA Federated Active Learning (`LG-FAL`)

Input: Number of clients N, number of FAL iteration T, test dataset $\mathcal{T}$, initially labelled dataset $\{\mathcal{D}_i(0)\}_{i=1}^N$, number of annotated data in each iteration z, initially unlabelled dataset $\{\mathcal{U}_i(0)\}_{i=1}^N$
Output: Optimized global model $\mathcal{M}_{FL}$
for $t = 1$ to T **do**
  for each client i, i = 1, ..., N **do**
    Train local model $\mathcal{M}_i$ with annotated data
    Send $\mathcal{M}_i$ parameters to the server
  end for
  At the server:
    Calculate gravitational mass $M_i(t)$ for each client
    Calculate total force $F_i^d(t)$ on each client
    Calculate the acceleration $a_i^d(t)$ of each client
    Update local model parameters
    Evaluate all updated local models with test dataset $\mathcal{T}$, define $\mathcal{M}_{FL}$ as the best local model
    Send $\mathcal{M}_{FL}$ to each client
  for each client i, i = 1, ..., N **do**
    Annotate z unlabelled data $\mathcal{A}_i$ from $\mathcal{U}_i(t)$
    Update $\mathcal{D}_i(t+1) = \mathcal{D}_i(t) + \mathcal{A}_i$
    sUpdate $\mathcal{U}_i(t+1) = \mathcal{U}_i(t) - \mathcal{A}_i$
end for

IV. Experiments

A. Datasets

We use three benchmark datasets in the experiments. The first one is the MNIST Dataset (Modified National Institute of Standards and Technology database) [24], which consists of a collection of handwritten digits. It contains 60,000 training images and 10,000 testing images. Each image is a grayscale image of size 28x28 pixels, representing a single digit (0-9). The second dataset is called Fashion MNIST [25]. It is a variation of the original MNIST dataset, but instead of containing handwritten digits, it consists of images of various types of clothing and fashion items. This dataset has 10 different categories, which include items like T-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots with 60,000 training images and 10,000 testing images and each image being a grayscale 28x28 pixel.

The last dataset in this study is the Diabetes Data Set, which integrates information from two primary sources: automated electronic recording devices and manual paper records. The goal is to predict whether a patient has diabetes based on these data inputs. The electronic devices provide precise event timestamps using an internal clock, ensuring real-time accuracy. In contrast, the paper records capture events according to broad periods defined by 'logical time'—such as breakfast, lunch, dinner, and bedtime—without specific timestamps [26].

B. Baselines

To validate the performance of the proposed method, we use deep neural networks as the training models and employ three baselines for our comparisons.

The first baseline is called Federated Average (FedAvg) Active Learning, FedAvg-FAL, which also shares the same network structure with our proposed method. In FedAvg, each client downloads the current model from a central server, improves it by learning from its own local data, and then aggregates the changes into a small centralized update. Equation 13 summarizes the global weight values **w** updating of FedAvg in each training round t, in which i is the client index, N means the total number of clients, $\mathcal{D}$ is the total number of instances and $\mathcal{D}_i$ is the local data examples for each client [27].

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} \frac{\mathcal{D}_i}{i} \mathbf{w}_t^i \tag{13}$$

The second baseline S-FAL annotates samples in a single AL way while keeping the same GSA FL model parameter update approach. In this setting, the local instance informativeness score is only computed based on the updated local model $\mathcal{M}_{FL}$ with local dataset D_i as shown in (14).

$$S(x) = Entr(Dis(x|\mathcal{M}_{FL})) \quad (14)$$

The last baseline is based on FedDNA, which is a dynamic node alignment federated learning algorithm to find the best matching nodes between different sites via Minimum Spanning Tree, and then aggregate weights of matching nodes for federated learning [2]. We combine FedDNA with active learning to validate the effectiveness of our proposed method.

C. Experiment Settings

Our overall experiment setting is as follows. For each dataset, our aim is to predict the corresponding target. Model parameters will be passed to each clients at the very beginning of training. Training data will be randomly split into 3 sites and distributed to 3 clients, which is able to training the local model using their own data. As for the AL part, by default, $w_1 = w_2 = 0.5$. For each round, 32 unlabelled samples will be annotated by different AL approaches and added to the local dataset for training. For the FL part, weight values will be aggregated based on different FL methods and then sent back to the global models. Global models will pass the newly calculated parameters to their local clients to start a new round of training until the convergence.

In order to explore the effect of various GSA parameter settings, we evaluated our proposed method using various combinations of α and G_0. Additionally, we conduct experiments to explore the impact of w_1 and w_2 on our proposed method.

D. Model Performance

Table I, Table II, and Table III show the model results for the three datasets respectively. Due to page limitations, only the best model performance results are presented in this paper. For the MNIST dataset, it is evident that GSA-based FAL methods outperform the FedAvg-based FAL approach. Across all models, predictive accuracy improves consistently as more samples are annotated by the active learning (AL) process, regardless of the GSA parameter settings. However, minor accuracy fluctuations occur when the number of annotated samples exceeds 192, a trend observed across all methods. When $\alpha = 30$ and $G_0 = 10$, LG-FAL achieves an accuracy of 0.848 with 320 annotated samples, outperforming FedAvg-AL. The superiority of our proposed method, LG-FAL, becomes more prominent under $\alpha = 30$, $G_0 = 20$ and $\alpha = 30$, $G_0 = 50$ settings, achieving accuracies of 0.854 and 0.857, respectively. LG-FAL delivers the highest final predictive accuracy (0.858) when $\alpha = 30$ and $G_0 = 50$.

For the Fashion MNIST dataset, the performance trends are consistent with those observed in the MNIST dataset. Both FedDNA-AL and LG-FAL demonstrate superior performance compared to FedAvg-AL and S-FAL across all parameter settings. Under $\alpha = 30$ and $G_0 = 20$, both S-FAL and LG-FAL achieve comparable accuracies higher than 0.750. However, as G_0 increases, LG-FAL's advantage over FedDNA-AL becomes more pronounced, especially as more samples are annotated.

The Diabetes dataset results reveal similar trends. FedDNA-AL generally performs around the expected accuracy of 0.72, while LG-FAL consistently outperforms the other methods across all parameter settings. For example, under $\alpha = 30$ and $G_0 = 20$, LG-FAL reaches an accuracy of 0.777, surpassing both FedAvg-AL and S-FAL. LG-FAL achieves the highest accuracy (0.778) when $\alpha = 30$ and $G_0 = 50$. This demonstrates the scalability and robustness of LG-FAL especially with a larger number of annotated samples.

Since LG-FAL and FedDNA-AL outperforms FedAvg-AL S-FAL, we create a figure to further demonstrate model comparison between LG-FAL and FedDNA-AL with the results from all the datasets. Regardless the parameter setting, the average accuracy of two models for three datasets are calculated. Fig. 3 report the performance of LG-FAL and FedDNA-AL as the increasing of annotated samples. The y-axis is the values of model accuracy and x-axis shows the increase of annotated instances. As the number of labeled samples gradually increases, the overall performance of the two models also shows an upward trend. Overall, LG-FAL always outperforms FedDNA-AL.

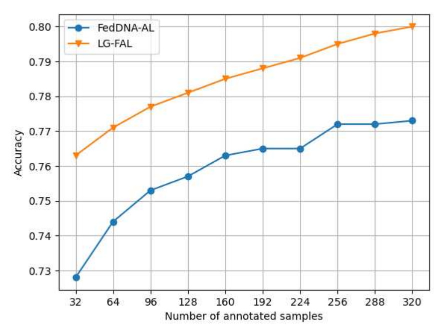


Fig. 3: Overall performance comparison between LG-FAL and FedDNA-AL with the increase of annotated samples: y-axis is the model averaged accuracy from all the three datasets; x-axis is the gradually increased number of annotated samples.

The advantages of LG-FAL are able to be verified with the previously shown results. With the confirmation that LG-FAL is able to outperform our baselines, especially when $\alpha = 30$, and $G_0 = 50$, we further conduct a series of experiments with MNIST dataset aiming to figure out how the change of proportion of $\mathcal{M}_i$ and $\mathcal{M}_{FL}$ in (2) effects the performance of LG-FAL. By default, the values of w_1 and w_2 are set as 0.5 and 0.5 respectively. Different combinations of w_1 and w_2 are designed in order to check how our proposed method will react as follows: w_1 = 0.2, w_2 = 0.8; w_1 = 0.4, w_2 = 0.6; w_1 = 0.5, w_2 = 0.5; w_1 = 0.6, w_2 = 0.4; w_1 = 0.8, w_2 = 0.2.

TABLE I: Model accuracy *w.r.t* different parameter settings for MNIST with gradually increasing annotated samples: FedAvg-AL is the first baseline, S-FAL is the second baseline, FedDNA-AL is the third baseline, `LG-FAL` is our proposed method.

GSA Parameter setting	Model	32	64	96	128	160	192	224	256	288	320
$\alpha = 30, G_0 = 10$	FedAvg-AL	0.755	0.768	0.835	0.794	0.826	0.804	0.806	0.823	0.814	0.78
	S-FAL	0.767	0.783	0.814	0.792	0.807	0.812	0.828	0.835	0.829	0.814
	FedDNA-AL	0.806	0.818	0.829	0.824	0.828	0.834	0.837	0.836	0.839	0.834
	LG-FAL	**0.817**	**0.821**	**0.826**	**0.830**	**0.834**	**0.838**	**0.842**	**0.845**	**0.846**	**0.848**
$\alpha = 30, G_0 = 20$	FedAvg-AL	0.767	0.799	0.812	0.820	0.819	0.802	0.827	0.831	0.815	0.828
	S-FAL	0.773	0.822	0.831	0.812	0.834	0.818	0.832	0.838	0.826	0.829
	FedDNA-AL	0.782	0.808	0.815	0.826	0.831	0.827	0.834	0.836	0.841	0.839
	LG-FAL	**0.822**	**0.827**	**0.832**	**0.836**	**0.840**	**0.844**	**0.848**	**0.850**	**0.852**	**0.854**
$\alpha = 30, G_0 = 50$	FedAvg-AL	0.748	0.786	0.808	0.798	0.813	0.812	0.803	0.819	0.813	0.817
	S-FAL	0.759	0.814	0.823	0.815	0.819	0.822	0.826	0.824	0.829	0.830
	FedDNA-AL	0.784	0.819	0.827	0.836	0.832	0.837	0.839	0.838	0.844	0.841
	LG-FAL	**0.832**	**0.837**	**0.842**	**0.846**	**0.849**	**0.852**	**0.854**	**0.856**	**0.858**	**0.857**

TABLE II: Model accuracy *w.r.t* different parameter settings for Fashion MNIST with gradually increasing annotated samples: FedAvg-AL is the first baseline, S-FAL is the second baseline, FedDNA-AL is the third baseline, `LG-FAL` is our proposed method.

GSA Parameter setting	Model	32	64	96	128	160	192	224	256	288	320
$\alpha = 30, G_0 = 10$	FedAvg-AL	0.659	0.670	0.674	0.689	0.691	0.695	0.678	0.707	0.698	0.696
	S-FAL	0.679	0.700	0.715	0.710	0.720	0.709	0.708	0.721	0.717	0.730
	FedDNA-AL	0.658	0.692	0.695	0.718	0.710	0.701	0.703	0.720	0.714	0.726
	LG-FAL	**0.725**	**0.733**	**0.739**	**0.744**	**0.748**	**0.751**	**0.752**	**0.760**	**0.762**	**0.764**
$\alpha = 30, G_0 = 20$	FedAvg-AL	0.685	0.698	0.710	0.707	0.712	0.700	0.706	0.716	0.718	0.727
	S-FAL	0.690	0.706	0.718	0.707	0.725	0.711	0.709	0.725	0.723	0.730
	FedDNA-AL	0.702	0.710	0.724	0.735	0.739	0.740	0.742	0.746	0.749	0.755
	LG-FAL	**0.741**	**0.746**	**0.751**	**0.754**	**0.758**	**0.761**	**0.763**	**0.769**	**0.772**	**0.774**
$\alpha = 30, G_0 = 50$	FedAvg-AL	0.443	0.552	0.600	0.594	0.616	0.615	0.612	0.631	0.634	0.642
	S-FAL	0.701	0.709	0.713	0.726	0.730	0.721	0.723	0.734	0.732	0.736
	FedDNA-AL	0.725	0.740	0.746	0.751	0.753	0.757	0.758	0.760	0.762	0.762
	LG-FAL	**0.732**	**0.748**	**0.755**	**0.756**	**0.755**	**0.759**	**0.763**	**0.769**	**0.772**	**0.774**

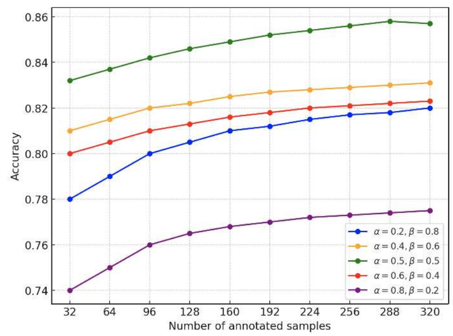
TABLE III: Model accuracy *w.r.t* different parameter settings for Diabetes Data Set with gradually increasing annotated samples: FedAvg-AL is the first baseline, S-FAL is the second baseline, FedDNA-AL is the third baseline, `LG-FAL` is our proposed method.

GSA Parameter setting	Model	32	64	96	128	160	192	224	256	288	320
$\alpha = 30, G_0 = 10$	FedAvg-AL	0.658	0.670	0.678	0.691	0.695	0.702	0.704	0.710	0.713	0.715
	S-FAL	0.675	0.688	0.698	0.705	0.711	0.715	0.718	0.722	0.726	0.728
	FedDNA-AL	0.705	0.715	0.721	0.727	0.732	0.735	0.739	0.742	0.745	0.748
	LG-FAL	**0.725**	**0.735**	**0.743**	**0.749**	**0.754**	**0.758**	**0.761**	**0.765**	**0.768**	**0.770**
$\alpha = 30, G_0 = 20$	FedAvg-AL	0.670	0.685	0.694	0.702	0.708	0.712	0.717	0.721	0.725	0.727
	S-FAL	0.688	0.702	0.710	0.717	0.722	0.726	0.729	0.733	0.736	0.738
	FedDNA-AL	0.710	0.720	0.726	0.731	0.737	0.741	0.744	0.747	0.750	0.752
	LG-FAL	**0.735**	**0.745**	**0.751**	**0.757**	**0.761**	**0.765**	**0.768**	**0.772**	**0.775**	**0.777**
$\alpha = 30, G_0 = 50$	FedAvg-AL	0.542	0.580	0.612	0.635	0.645	0.652	0.660	0.665	0.668	0.670
	S-FAL	0.680	0.692	0.705	0.712	0.717	0.722	0.727	0.731	0.735	0.737
	FedDNA-AL	0.718	0.725	0.731	0.736	0.741	0.744	0.748	0.751	0.754	0.756
	LG-FAL	**0.735**	**0.745**	**0.752**	**0.757**	**0.762**	**0.766**	**0.769**	**0.773**	**0.776**	**0.778**

Fig. 4 reports the overall predict accuracy trend of `LG-FAL` on MNIST dataset under different w_1 and w_2 settings. We can clearly observe that when prefer the predictive uncertainty from local model over global model, the performance of `LG-FAL` drops especially when $w_1 = 0.8$ and $w_2 = 0.2$. However, the gradual increase of model performance with the increasing annotated samples can still be validated. `LG-FAL` demonstrates the best predictive accuracy on MNIST dataset when local model and global model are equally considered for annotating the instances.

V. Conclusion

In this paper, we propose a locality-customized GSA Federated Active Learning (`LG-FAL`) method for federated active learning. We argued that in most federated active learning frameworks, local unlabeled samples are annotated by the aggregated global model's parameters, which totally ignores the localization of the samples, leading to neglecting the importance of local models for local sample annotation. In addition, current federated active learning approaches usually are limited to one method, Federated averaging (FedAvg) to update global model parameter. Alternatively, we propose a locality-customized GSA Federated Active Learning method,



(a) `LG-FAL` performance on MNIST dataset

Fig. 4: Performance of `LG-FAL` with different w_1 and w_2 settings on MNIST dataset and Fashion MNIST dataset with the increase of annotated samples: y-axis is `LG-FAL`'s accuracy; x-axis is the gradually increased number of annotated samples; legends are w_1 and w_2 settings.

`LG-FAL` to tackle the aforementioned limitations. `LG-FAL` combines locality-customized active learning and Gravitational Search Algorithm (GSA) in a collaborative and effective way. In locality-customized active learning, both the local model as well as the global model are taken into consideration when annotating local samples, in which each data's overall uncertainty is a combination of both the local model's prediction entropy and the global model's prediction entropy. In GSA federated learning, global model parameter aggregations are achieved by GSA which is empowered with higher adaptability with a set of parameters to allow clients to move freely towards areas of high fitness calculated based on their masses (accuracy). Experiments and comparisons validate the performance of the `LG-FAL` compared to other baselines. For future work, we will perform a deeper analysis of the sensitivity of `LG-FAL` to different parameter settings, such as the weights w_1 and w_2 in the dual-model annotation strategy and investigate the scalability of `LG-FAL` to large-scale federated learning scenarios.

Acknowledgements

This study is supported by the U.S. National Science Foundation under grant Nos. IIS-2236579, IIS-2302786 and IOS-2430224.

References

[1] L. Ahmed, K. Ahmad, N. Said, B. Qolomany, J. Qadir, and A. Al-Fuqaha, "Active learning based federated learning for waste and natural disaster image classification," *IEEE Access*, vol. 8, pp. 208518–208531, 2020.

[2] S. Wang and X. Zhu, "Feddna: Federated learning using dynamic node alignment," *Plos one*, vol. 18, no. 7, p. e0288157, 2023.

[3] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: a big data-ai integration perspective," *IEEE Trans. on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2019.

[4] S. Russo, M. D. Besmer, F. Blumensaat, D. Bouffard, A. Disch, F. Hammes, A. Hess, M. Lürig, B. Matthews, C. Minaudo, *et al.*, "The value of human data annotation for machine learning based anomaly detection in environmental systems," *Water Research*, vol. 206, p. 117695, 2021.

[5] Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C. C. Aggarwal, J. Pei, and Y. Zhou, "A comprehensive survey on data augmentation," *arXiv preprint arXiv:2405.09591*, 2024.

[6] S. H. Shetty, S. Shetty, C. Singh, and A. Rao, "Supervised machine learning: Algorithms and applications," *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools and Applications*, pp. 1–16, 2022.

[7] A. Tharwat and W. Schenck, "A survey on active learning: State-of-the-art, practical challenges and research directions," *Mathematics*, vol. 11, no. 4, p. 820, 2023.

[8] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, "A survey of deep active learning," *ACM computing surveys (CSUR)*, vol. 54, no. 9, pp. 1–40, 2021.

[9] X. Wu, J. Pei, C. Chen, Y. Zhu, J. Wang, Q. Qian, J. Zhang, Q. Sun, and Y. Guo, "Federated active learning for multicenter collaborative disease diagnosis," *IEEE Transactions on Medical Imaging*, 2022.

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

[11] S. Kim, S. Bae, H. Song, and S.-Y. Yun, "Re-thinking federated active learning based on inter-class diversity," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp. 3944–3953, 2023.

[12] S. Pandya, G. Srivastava, R. Jhaveri, M. R. Babu, S. Bhattacharya, P. K. R. Maddikunta, S. Mastorakis, M. J. Piran, and T. R. Gadekallu, "Federated learning for smart cities: A comprehensive survey," *Sustainable Energy Technologies and Assessments*, vol. 55, p. 102987, 2023.

[13] Y. Jin, X. Wei, Y. Liu, and Q. Yang, "Towards utilizing unlabeled data in federated learning: A survey and prospective," *arXiv preprint arXiv:2002.11545*, 2020.

[14] D. Chen, J. Hu, V. J. Tan, X. Wei, and E. Wu, "Elastic aggregation for federated optimization," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp. 12187–12197, 2023.

[15] F. Naeem, M. Ali, and G. Kaddoum, "Federated-learning-empowered semi-supervised active learning framework for intrusion detection in zsm," *IEEE Communications Magazine*, vol. 61, no. 2, pp. 88–94, 2023.

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

[17] R. Presotto, G. Civitarese, and C. Bettini, "Semi-supervised and personalized federated activity recognition based on active learning and label propagation," *Personal and Ubiquitous Computing*, vol. 26, no. 5, pp. 1281–1298, 2022.

[18] G. Long, M. Xie, T. Shen, T. Zhou, X. Wang, and J. Jiang, "Multi-center federated learning: clients clustering for better personalization," *World Wide Web*, vol. 26, no. 1, pp. 481–500, 2023.

[19] H. Zhou, T. Lan, G. P. Venkataramani, and W. Ding, "Every parameter matters: Ensuring the convergence of federated learning with dynamic heterogeneous models reduction," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[20] Y.-T. Cao, Y. Shi, B. Yu, J. Wang, and D. Tao, "Knowledge-aware federated active learning with non-iid data," in *Proc. of the IEEE/CVF Conf. on Computer Vision*, pp. 22279–22289, 2023.

[21] Y. H. Ezzeldin, S. Yan, C. He, E. Ferrara, and A. S. Avestimehr, "Fairfed: Enabling group fairness in federated learning," in *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 37, pp. 7494–7502, 2023.

[22] A. Yazdinejad, A. Dehghantanha, and G. Srivastava, "Ap2fl: Auditable privacy-preserving federated learning framework for electronics in healthcare," *IEEE Trans. on Consumer Electronics*, 2023.

[23] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "Gsa: a gravitational search algorithm," *Information sciences*, vol. 179, no. 13, pp. 2232–2248, 2009.

[24] Y. LeCun, "The mnist database of handwritten digits," *http://yann. lecun. com/exdb/mnist/*, 1998.

[25] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[26] M. Kahn, "Diabetes data set." https://archive.ics.uci.edu/ml/datasets/diabetes, 1994.

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