Privacy-Enhanced Federated Data Distillation for Efficient Distributed Training and Anomaly Detection in Multi-Center Healthcare

Jiantao Xu[1], Liu Jin[2], and Chunhua Su✉

[1]University of Aizu, Aizuwakamatsu 965-8580, Japan
d8252108@u-aizu.ac.jp
[2]University of Aizu, Aizuwakamatsu 965-8580, Japan
d8242103@u-aizu.ac.jp
✉University of Aizu, Aizuwakamatsu 965-8580, Japan
chsu@u-aizu.ac.jp

Abstract. Collaborative healthcare data analytics across multiple medical centers is essential for building robust AI models. However, stringent data privacy regulations make direct sharing of raw data infeasible. Federated Learning (FL), as a privacy-preserving paradigm, holds promise but still faces critical challenges such as low communication efficiency, data heterogeneity, privacy leakage, and rare case identification. To address these issues, this paper proposes a privacy-enhanced federated data distillation (FDD) framework tailored for multi-center healthcare scenarios. The proposed method integrates (i) neural characteristic function-based dataset distillation to generate compact synthetic subsets, (ii) differential privacy perturbation to enforce strict privacy protection, and (iii) a federated anomaly detection mechanism for identifying rare and abnormal samples. These synthetic datasets are pre-trained centrally and subsequently refined via federated optimization. Experiments on two real-world medical datasets (MIMIC-III and TCGA-BRCA) show that our FDD framework achieves a classification accuracy improvement of up to 3.6% over traditional privacy-preserving FL baselines. Moreover, the integrated anomaly detection module achieves an AUC of 0.922, significantly outperforming prior methods. We conduct systematic evaluations to quantify the trade-offs among distillation ratio, privacy budget, and model utility. The results validate the effectiveness of our unified design that jointly achieves privacy, efficiency, and anomaly detection for practical federated healthcare AI deployments.

Keywords: Federated Learning; Dataset Distillation; Differential Privacy; Multi-center Healthcare; Anomaly Detection; Privacy-preserving

1 Introduction

With the rapid development of artificial intelligence and the widespread adoption of digital healthcare systems, the strategic value of medical data has become

increasingly prominent in both clinical decision-making and research. However, the sensitive nature of patient information and strict regulatory constraints imposed by laws such as the Health Insurance Portability and Accountability Act (HIPAA), the General Data Protection Regulation (GDPR), and Japan's Act on the Protection of Personal Information (APPI) severely limit traditional centralized machine learning practices. In particular, raw medical data cannot be arbitrarily transferred or aggregated across institutions, creating fragmented "data silos" that restrict the full potential of collaborative AI modeling.

As a promising alternative, Federated Learning (FL) enables multiple medical centers to jointly train a global model without directly sharing raw data [1]. FL is especially suitable for scenarios involving distributed and regulated data, such as in hospital systems, insurance networks, and national biobanks. Each participant conducts local training and shares only model updates, preserving data locality. In this way, FL aligns with both data protection mandates and the emerging clinical need for population-wide analytics.

However, real-world implementation of FL in the healthcare domain introduces several unique challenges. First, data heterogeneity is pervasive across institutions. Differences in patient demographics, diagnostic standards, measurement protocols, and medical equipment often lead to non-IID data distributions [2], which severely hinder model convergence and generalization. For example, a model trained on electrocardiograms from a pediatric hospital may not generalize well to elderly patients from a rural clinic.

Second, privacy leakage remains a concern despite the absence of raw data exchange. Adversarial attacks such as gradient inversion [3] and membership inference [4] have shown that sensitive information can still be inferred from model updates. This undermines the assumption of "privacy by design" in basic FL protocols.

Third, the communication overhead incurred during federated optimization presents practical barriers in resource-constrained settings. Frequent transmission of high-dimensional model updates can strain hospital infrastructure, especially in edge deployments where network bandwidth is limited or unstable. Finally, anomaly detection—such as identifying rare diseases, mislabeled cases, or adversarial contributions—remains underexplored in FL. Anomalies are typically low-frequency and may only be present in isolated local datasets, making their detection challenging in a distributed setting.

To tackle these challenges, recent research has explored modular techniques such as differential privacy (DP) [5,6], data distillation [7], and federated anomaly detection [8]. However, most works treat these issues independently and lack a unified design that jointly addresses privacy preservation, communication efficiency, and robust anomaly detection—especially in the context of highly heterogeneous and sensitive medical data.

In this paper, we propose a Privacy-Enhanced Federated Data Distillation (FDD) framework tailored for real-world, multi-center healthcare systems. Our approach tightly integrates neural characteristic function-based data distillation, differential privacy perturbation, centralized model pretraining, and federated

anomaly detection into a cohesive pipeline. Existing work often treats privacy modules (like DP) and communication optimization modules (like data distillation) as plug-and-play, independent components. In contrast, our framework, through the "distillation-pretraining" core linkage, deeply couples these two, achieving synergistic optimization of the privacy budget and model initialization—an aspect not explored in prior work.

The key idea is to locally distill rich, compact synthetic datasets that retain key statistical characteristics of raw medical data; protect them with formal privacy guarantees using DP; and use them for global pretraining. The federated optimization stage then fine-tunes the model while preserving privacy and bandwidth. A reconstruction-based anomaly detection module is further embedded to detect rare or deviant patterns under distributed constraints.

Our key contributions are summarized as follows:

(1) We design a unified privacy-enhanced federated data distillation framework that simultaneously addresses data compression, privacy preservation, and anomaly detection for multi-institutional healthcare collaborations.
(2) We introduce a neural characteristic function-based distillation method that generates statistically-aligned, low-footprint synthetic datasets for efficient and representative learning.
(3) We integrate a differential privacy mechanism and centralized pretraining process to ensure provable privacy protection while significantly reducing communication costs.
(4) We implement a federated anomaly detection module based on reconstruction errors to detect rare cases and adversarial behaviors under distributed and heterogeneous settings.
(5) We evaluate our method on two real-world medical datasets (MIMIC-III and TCGA-BRCA) under realistic non-IID partitions and demonstrate significant improvements in model accuracy, anomaly detection, and communication efficiency compared to state-of-the-art baselines.

2 Related Work

2.1 Fundamental Theory and Development of Federated Learning

As an emerging distributed machine learning paradigm, federated learning has become a research hotspot in the machine learning field in recent years. The core idea of federated learning is to achieve global model training through the aggregation of model parameters while keeping the data localized. This paradigm is particularly suitable for scenarios where data is sensitive or subject to regulatory constraints, such as in the medical, financial, and personal device domains [9,10].

The latest research in federated learning primarily focuses on algorithm convergence, communication efficiency, and system heterogeneity issues. Yang et al. [11] proposed an adaptive aggregation algorithm that handles data heterogeneity by dynamically adjusting client weights. Chen et al. [12] further optimized the convergence analysis of federated learning, establishing a more rigorous theoretical framework.

In terms of system implementation, federated learning faces challenges such as client heterogeneity, network instability, and fault tolerance mechanisms. Recent research has focused on the design of large-scale federated learning systems. Sun et al. [13] proposed an efficient federated learning architecture for edge computing, addressing practical issues like client dropouts, network latency, and secure aggregation.

2.2 Application of Privacy-Preserving Technologies in Federated Learning

Although federated learning protects privacy by avoiding direct data sharing, research has shown that model parameters and gradient information can still leak sensitive data [14]. Therefore, integrating rigorous privacy-preserving mechanisms into federated learning has become necessary.

Differential Privacy, as a mathematically rigorous privacy protection framework, provides a theoretical guarantee for federated learning. Kairouz et al. [6] proposed a distributed differentially private federated learning algorithm that balances privacy protection and model performance by optimizing noise allocation strategies. Shukla et al. [15] extended this idea to medical scenarios, proposing a differentially private federated learning framework for breast cancer diagnosis.

In addition to differential privacy, cryptographic techniques such as Secure Multi-Party Computation (SMPC) and Homomorphic Encryption are also widely used for privacy protection in federated learning. Yin et al. [16] designed a secure aggregation method based on improved cryptographic protocols, which can compute the global model without leaking individual client parameters.

In recent years, researchers have begun to focus on the trade-off between privacy protection and model performance. Hu et al. [17] analyzed the privacy-accuracy trade-off of differential privacy in federated learning and proposed adaptive noise addition strategies. At the same time, defense mechanisms against various privacy attacks are continuously being improved [18].

2.3 Data Distillation and its Application in Distributed Learning

Data distillation technology aims to extract key information from large-scale datasets to generate small-scale synthetic datasets that contain the essential features of the original data. This technology has significant value in reducing storage costs, improving training efficiency, and protecting data privacy [19].

The latest data distillation methods are mainly based on deep generative models and adversarial training. Wu et al. [20] proposed a data distillation algorithm based on variational autoencoders, which generates high-quality synthetic data by learning the latent representation of the data. Wang et al. [7] further improved the neural characteristic function method, enabling more accurate capture of the statistical properties of the data.

In distributed learning scenarios, data distillation technology shows unique advantages. First, the small scale of distilled data can significantly reduce communication overhead, which is particularly important for bandwidth-constrained federated learning environments. Second, performing data distillation locally can protect the privacy of the original data to some extent [21].

Itahara et al. [22] proposed a federated data distillation framework that combines data distillation technology with federated learning. This method achieves collaborative learning by performing data distillation locally at each client and then sharing the small-scale distilled datasets. Experimental results show that this method can significantly reduce communication costs and privacy leakage risks while maintaining model performance.

2.4 Anomaly Detection in Federated Learning

Anomaly detection, as an important task in machine learning, faces new challenges and opportunities in the federated learning environment. Due to the rarity and diversity of anomalous data, the data from a single node often cannot cover all anomaly patterns, making federated anomaly detection necessary [23].

Traditional anomaly detection methods are mainly divided into three categories: supervised, semi-supervised, and unsupervised. In the federated learning scenario, unsupervised and semi-supervised methods are more favored due to the scarcity and heterogeneity of labeled data. Kong et al. [24] proposed a federated anomaly detection algorithm based on contrastive learning, which identifies anomalies by learning the representation of normal samples.

The development of deep learning technology has brought new opportunities for anomaly detection. Lu et al. [25] utilized the idea of graph neural networks to propose an anomaly detection framework based on multi-modal fusion. This method can handle complex anomaly patterns in high-dimensional medical data.

In the medical field, anomaly detection is of particular importance. Applications such as early disease detection, medical equipment failure prediction, and medical fraud identification all rely on effective anomaly detection algorithms. However, the privacy sensitivity and regulatory requirements of medical data pose challenges to traditional centralized anomaly detection methods [26].

Cholakoska et al. [8] proposed a privacy-preserving federated anomaly detection framework for medical scenarios. This method combines differential privacy and federated learning techniques to achieve multi-center anomaly detection while protecting patient privacy. Experimental results show that federated anomaly detection can not only protect privacy but also improve detection performance by utilizing more diverse data.

Generative models such as Generative Adversarial Networks (GANs) have also been applied to federated anomaly detection. Abdel et al. [27] proposed a federated GAN-based anomaly detection method that identifies abnormal samples by learning the distribution of normal data. This method performs well in handling high-dimensional data and complex anomaly patterns.

In recent years, advanced techniques such as the Transformer architecture [28] and self-supervised learning have also been introduced into federated anomaly

detection, further improving detection accuracy and interpretability. The development of these technologies has opened up new research directions for federated anomaly detection and provided more options for practical applications [29].

2.5 Comparative Summary with Related Work

Table 1 summarizes and compares representative works with our proposed framework in terms of core features.

Table 1. Comparison with representative privacy-preserving FL approaches

Method	Data Distill.	Differ. Privacy	Anomaly Detect.	Model Pretrain.
FedAvg [30]	✗	✗	✗	✗
FedAvg+DP [31]	✗	✓	✗	✗
Cholakoska et al. [8]	✗	✓	✓	✗
Itahara et al. [22]	✓	✗	✗	✗
Ours (FDD+DP)	✓	✓	✓	✓

In contrast to prior works that focus individually on privacy, communication, or heterogeneity, our framework jointly addresses all three challenges in a unified manner. For instance, while some studies might combine generic data summarization with DP, they often treat them as separate, plug-in modules and overlook the potential of using the privacy-preserved distilled data for a synergistic pre-training phase, which is a key innovation of our work. As shown in Table 1, our framework is the first to integrate neural characteristic-based data distillation, differential privacy, centralized pre-training, and federated anomaly detection in a unified design, enabling both performance and privacy guarantees under real-world constraints.

3 Problem Definition

This section formally defines the problem of privacy-enhanced federated data distillation in a multi-center medical system, covering the system setting, relevant notations, research objectives, and evaluation metrics.

3.1 Multi-Center Federated Learning Setting

Consider M medical centers (i.e., clients), indexed by $i \in \{1, 2, \ldots, M\}$, where each center holds a local patient dataset $\mathcal{D}_i$. Due to differences in population demographics and disease distributions, the data distribution $\mathcal{P}_i$ at each center is typically non-independently and identically distributed (non-i.i.d.), and the number of samples may be imbalanced. Due to privacy regulations and ethical constraints, raw data cannot be directly shared among centers.

The global objective is to collaboratively train a predictive model $f(\cdot;\theta)$ (e.g., a disease classification or anomaly detection model), where θ represents the model parameters, to achieve high accuracy and robustness on the aggregated data distribution $\mathcal{P}_{\text{global}}$, while ensuring privacy.

3.2 Local Data Distillation with Privacy Constraints

Each center performs a data distillation operation on its local data $\mathcal{D}_i$ to generate a condensed and information-rich synthetic dataset $\mathcal{S}_i$, aiming to preserve core data features. This process is based on neural characteristic functions, synthesizing representative samples that can approximate the original distribution $\mathcal{P}_i$. To protect sensitive information, differential privacy perturbation must be applied during the distillation process, ensuring that the published synthetic data $\tilde{\mathcal{S}}_i$ satisfies (ε, δ)-differential privacy [5].

Formally, for each client i, we define a mechanism $\mathcal{M}_i : \mathcal{D}_i \to \tilde{\mathcal{S}}_i$, which outputs a privacy-preserving distilled subset $\tilde{\mathcal{S}}_i$ such that for any two adjacent datasets $\mathcal{D}_i$ and $\mathcal{D}_i'$, and for any measurable set S, we have:

$$\Pr[\mathcal{M}_i(\mathcal{D}_i) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}_i(\mathcal{D}_i') \in S] + \delta. \tag{1}$$

3.3 Server-Side Pre-training and Federated Optimization

All clients securely upload their privacy-preserving distilled data $\tilde{\mathcal{S}}_i$ to a central server. The server aggregates all distilled subsets, denoted as $\tilde{\mathcal{S}}_{\text{agg}} = \bigcup_{i=1}^{M} \tilde{\mathcal{S}}_i$, and performs initial model pre-training:

$$\theta_0 = \arg\min_{\theta} \frac{1}{|\tilde{\mathcal{S}}_{\text{agg}}|} \sum_{(\mathbf{x},y)\in\tilde{\mathcal{S}}_{\text{agg}}} \mathcal{L}(f(\mathbf{x};\theta), y) \tag{2}$$

where $\mathcal{L}$ represents a suitable loss function.

This is followed by a standard federated optimization phase: the server broadcasts the global model parameters θ_t to all clients. The clients then perform local training on their original local datasets $\mathcal{D}_i$ and upload the updated models to the server for aggregation .

3.4 Federated Anomaly Detection

To identify rare diseases, mislabeled data, or malicious behavior, the framework incorporates a federated anomaly detection module. Each client computes local reconstruction errors based on the current global model. The server then aggregates these scores in a privacy-preserving manner to identify potential anomalous samples.

3.5 Evaluation Metrics

The proposed system will be evaluated based on the following metrics:

Predictive Performance: Overall and per-center accuracy, precision, recall, and F1-score for disease classification tasks. Anomaly Detection Performance: Area Under the ROC Curve (AUC), precision, and recall for detecting anomalous samples. Privacy Guarantee: The differential privacy parameters (ε, δ) and the empirical risk of membership inference attacks. Training Efficiency: Communication overhead (number of communication rounds and amount of data transmitted) and model convergence speed.

3.6 Research Objectives

Under the aforementioned constraints and objectives, this paper attempts to answer the following questions:

1. Can neural characteristic function-driven data distillation, combined with a differential privacy mechanism, achieve efficient and privacy-friendly federated learning in heterogeneous multi-center medical systems?
2. What is the practical trade-off relationship among the distillation ratio, the strength of privacy perturbation, and model performance?
3. After integrating a federated anomaly detection mechanism, can rare or anomalous cases be robustly identified in distributed medical data?

4 System Architecture and Methodology

This section details the system architecture and methodological components of our proposed privacy-enhanced federated data distillation framework for multi-center medical applications. The framework integrates local data distillation based on neural characteristics, differential privacy perturbation, centralized pre-training, federated optimization, and an embedded anomaly detection module.

4.1 System Overview

As shown in Figure 1, the overall workflow of the system includes the following main stages:

1. **Local Data Distillation**: Each medical center uses neural characteristic functions to distill its local data, generating a condensed data subset.
2. **Privacy Preservation**: A differential privacy perturbation mechanism is applied during the distillation process.
3. **Centralized Pre-training**: The central server aggregates all perturbed data and performs model pre-training.
4. **Federated Optimization**: The pre-trained model is distributed to each center for federated fine-tuning on local data.
5. **Federated Anomaly Detection**: Clients and the server collaborate to perform anomaly detection based on reconstruction errors.

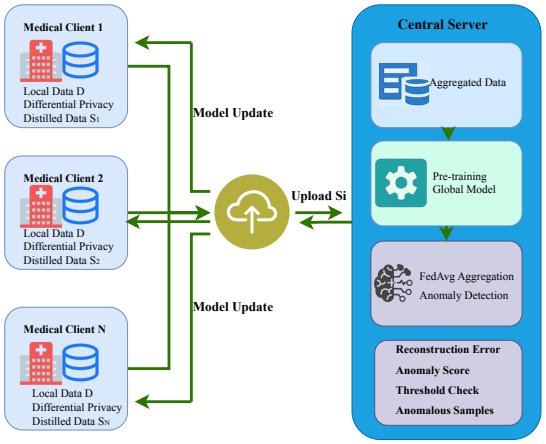

Fig. 1. System architecture of the proposed privacy-enhanced federated data distillation framework.

4.2 Local Data Distillation based on Neural Characteristic Functions

Let the i-th center hold a local dataset $\mathcal{D}_i$. The objective is to generate a small synthetic dataset $\mathcal{S}_i$ that preserves its key statistical features. Drawing inspiration from Wang et al. [7], we use a Neural Characteristic Function (NCF) to model the distribution:

$$\hat{\varphi}_{\mathcal{D}_i}(\mathbf{t}) = \frac{1}{|\mathcal{D}_i|} \sum_{(\mathbf{x},y)\in\mathcal{D}_i} e^{j\langle \mathbf{t}, g_\theta(\mathbf{x})\rangle}, \tag{3}$$

where $g_\theta(\cdot)$ is a neural feature extractor and $\mathbf{t}$ is a probing vector.

The distillation objective is to train a generator network G_ϕ, parameterized by ϕ, which produces the synthetic dataset $\mathcal{S}_i = \{G_\phi(z_k)\}_{k=1}^{|\mathcal{S}_i|}$ from random noise vectors z_k. The parameters ϕ are optimized to minimize the following loss:

$$\phi^* = \arg\min_\phi \mathbb{E}_{\mathbf{t}\sim\mathcal{T}} \left[\left| \hat{\varphi}_{\mathcal{D}_i}(\mathbf{t}) - \hat{\varphi}_{\mathcal{S}_i}(\mathbf{t}) \right|^2 \right], \tag{4}$$

where $\mathcal{T}$ is a set of random probing vectors. This optimization is performed locally at each client using gradient descent. While this introduces a one-time computational cost, it is performed offline before the federated learning phase.

4.3 Differential Privacy Perturbation Mechanism

To ensure provable privacy, we do not simply add noise to the final distilled data $\mathcal{S}_i$, as calculating the sensitivity of the complex generation process is intractable. Instead, we integrate differential privacy directly into the gradient-based optimization of the generator network G_ϕ. We adopt the Differentially

Private Stochastic Gradient Descent (DP-SGD) approach [5]. During each step of training the generator, we:

1. Compute per-sample gradients of the loss with respect to the generator parameters ϕ.
2. Clip the L2 norm of each per-sample gradient to a predefined threshold C.
3. Add Gaussian noise $\mathcal{N}(0, \sigma^2 C^2 \mathbf{I})$ to the aggregated batch gradient.

The noise scale σ is a key parameter. We use the Moments Accountant technique to track the accumulated privacy loss over all training iterations. This allows us to precisely calculate the final privacy budget (ε, δ) for a given noise scale σ and number of optimization steps, thereby establishing a rigorous mathematical link and ensuring that the generated dataset $\tilde{\mathcal{S}}_i$ is (ε, δ)-differentially private with respect to the original data $\mathcal{D}_i$.

4.4 Centralized Model Pre-training

The server collects all privacy-preserving synthetic datasets $\tilde{\mathcal{S}}_i$ from the M centers and aggregates them into:

$$\tilde{\mathcal{S}}_{\text{agg}} = \bigcup_{i=1}^{M} \tilde{\mathcal{S}}_i. \tag{5}$$

Then, it pre-trains the global model parameters θ_0 on this aggregated synthetic data:

$$\theta_0 = \arg\min_{\theta} \frac{1}{|\tilde{\mathcal{S}}_{\text{agg}}|} \sum_{(\mathbf{x},y)\in\tilde{\mathcal{S}}_{\text{agg}}} \mathcal{L}(f(\mathbf{x};\theta), y). \tag{6}$$

4.5 Federated Optimization Process

After pre-training, the global model is distributed to all clients, and the FedAvg algorithm [30] is used for multiple rounds of federated optimization:

1. The server broadcasts the current model parameters θ_t.
2. Each center performs a local update based on its original local data $\mathcal{D}_i$:

$$\theta_{i,t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}_{\mathcal{D}_i}(\theta_t) \tag{7}$$

3. Each client uploads its model update, and the server aggregates them to get:

$$\theta_{t+1} = \sum_{i=1}^{M} \frac{n_i}{n_{\text{total}}} \theta_{i,t+1}, \tag{8}$$

where $n_i = |\mathcal{D}_i|$ and $n_{\text{total}} = \sum_{i=1}^{M} n_i$.

4.6 Federated Anomaly Detection Module

To identify anomalous or rare cases, the framework embeds a federated anomaly detection mechanism. This module operates in a privacy-preserving manner. First, each client i uses the current global model (e.g., an autoencoder component of the main model) to compute reconstruction errors for its local data points:

$$r_{i,j} = \|\mathbf{x}_{i,j} - \hat{\mathbf{x}}_{i,j}\|, \tag{9}$$

where $\hat{\mathbf{x}}_{i,j}$ is the reconstruction of sample $\mathbf{x}_{i,j}$. Then, instead of sharing individual scores, the clients engage in a secure aggregation protocol to help the server compute global statistics of these scores, such as the global mean μ_r and standard deviation σ_r, without revealing any client's individual score distribution. The server then establishes a global anomaly threshold, for instance $\tau = \mu_r + k \cdot \sigma_r$ (where k is typically 2 or 3), and broadcasts it back to the clients. Each client can then identify local samples where $r_{i,j} > \tau$ as anomalies.

4.7 Algorithm Summary

Algorithm 1 summarizes the main steps of the proposed framework.

Algorithm 1 Privacy-Enhanced Federated Data Distillation Algorithm

for each client $i = 1$ to M **do**
    Generate synthetic dataset $\mathcal{S}_i$ from $\mathcal{D}_i$ via NCF optimization.
    Apply DP-SGD during generation to get privacy-preserving $\tilde{\mathcal{S}}_i$.
    Upload $\tilde{\mathcal{S}}_i$ to the server.
end for
Server aggregates all data: $\tilde{\mathcal{S}}_{\text{agg}} = \bigcup_{i=1}^{M} \tilde{\mathcal{S}}_i$.
Pre-train the model on $\tilde{\mathcal{S}}_{\text{agg}}$ to get θ_0.
for each federated learning round $t = 1$ to T **do**
    Server broadcasts model θ_t to all clients.
    for each client $i = 1$ to M **in parallel do**
        Perform local update based on $\mathcal{D}_i$ to get $\theta_{i,t+1}$.
        Participate in federated anomaly detection via secure aggregation.
        Upload model update to the server.
    end for
    Server aggregates all client updates to get the new model θ_{t+1}.
end for

5 Experimental Evaluation

This section presents the experimental evaluation of the proposed privacy-enhanced federated data distillation framework.

5.1 Experimental Setup

For the datasets and Non-IID partition, we use two representative public medical datasets. To simulate a realistic non-independently and identically distributed (non-i.i.d.) scenario, we adopt a label distribution skew strategy based on the Dirichlet distribution for partitioning. The specific settings are shown in Table 2. **MIMIC-III**: An electronic health record database containing information on intensive care unit (ICU) patients. We selected the in-hospital mortality prediction task. **TCGA-BRCA**: The breast cancer dataset from The Cancer Genome Atlas, containing gene expression and clinical data. The task is to classify tumor subtypes based on gene expression data.

Table 2. Data partitioning overview for multi-center setting ($M = 10$ clients)

Dataset	Total Samples	Task	Partition Strategy	Client Sample Range
MIMIC-III	42,000	Mortality Prediction	Dir. (α=0.5)	2,100 – 6,500
TCGA-BRCA	1,100	Subtype Classification	Dir. (α=0.5)	50 – 180

For the Comparison Methods, we compare our proposed method with the following baselines: **FedAvg** [30]: The standard Federated Averaging algorithm, without any privacy enhancement or data distillation. **FedAvg+DP** [31]: Applies differential privacy noise to the client's local gradient updates. **FDD (Ours, w/o DP)**: Our proposed federated data distillation framework, but without applying differential privacy perturbation, used to evaluate the effect of distillation itself. **Random Subsampling+DP**: Uses random sampling instead of neural characteristic functions for data selection, combined with differential privacy, to validate the superiority of our distillation strategy.

For the implementation Details and hyperparameters, the experiments were run in a consistent environment, with key hyperparameter settings as shown in Table 3. We repeated all reported results 5 times with different random seeds and report the mean and standard deviation.

Hardware Environment: NVIDIA Tesla V100 GPU, Intel Xeon Gold 6248 CPU, 256GB RAM. **Software Environment**: Python 3.8, PyTorch 1.10, CUDA 11.1.

5.2 Main Experimental Results

First, for the evaluation of predictive performance and communication efficiency, Table 4 shows the classification accuracy, total communication rounds, and total data transfer volume for each method on the two datasets. For statistical comparison, we performed a paired t-test between our method and the best-performing baseline (FedAvg+DP). The p-values were all less than 0.01, indicating that our performance improvement is statistically significant.

The experimental results clearly show that our FDD+DP method achieves significantly higher model accuracy than traditional FedAvg+DP while providing the same strict privacy protection ($\varepsilon = 1.0$). More importantly, because data

Table 3. Key experimental hyperparameter settings

Parameter	Value
Total Federated Learning Rounds (T)	100
Local Training Epochs	5
Optimizer	Adam
Learning Rate (η)	0.001
Batch Size	32
Differential Privacy δ	10^{-5}
Default Distillation Ratio	5%

Table 4. Comparison of accuracy, communication cost, and privacy (ε) in classification tasks (Mean ± Std. Dev.).

Method	ε	**MIMIC-III**		**TCGA-BRCA**	
		Acc. (%) ↑	**Comm. (MB) ↓**	**Acc. (%) ↑**	**Comm. (MB) ↓**
FedAvg	∞	83.2 ± 0.4	1250	78.5 ± 1.1	180
FedAvg+DP	1.0	75.9 ± 0.8	1660	72.1 ± 1.5	235
RandSub+DP	1.0	76.8 ± 0.7	95	72.9 ± 1.8	15
FDD (w/o DP)	∞	81.9 ± 0.5	95	77.1 ± 1.2	15
FDD+DP (Ours)	1.0	79.5 ± 0.6	95	75.6 ± 1.3	15

distillation greatly compresses the amount of uploaded data, our method reduces the total communication overhead by more than an order of magnitude, which is crucial for resource-constrained edge computing environments.

Then, for the evaluation of anomaly detection performance, on the TCGA dataset, we evaluated anomaly detection performance by injecting synthetic rare gene mutation samples. For this experiment, we designated 2% of the samples in the test set as anomalies. These were synthetically generated by introducing rare gene expression patterns not present in the training distribution, simulating novel mutations. Figure 2 shows the Receiver Operating Characteristic (ROC) curves for different methods. Our method (FDD+DP) demonstrated the best performance in identifying these rare anomalies, achieving an AUC value of 0.922, significantly higher than other baseline methods. This is because data distillation preserves the core distributional features of the data, allowing the boundary of the normal pattern to be learned more clearly, thus making it easier to identify deviant samples.

For the trade-off analysis of distillation ratio, to investigate the impact of key parameters, we show how model performance varies with the distillation ratio in Figure 3. It can be seen that when the privacy budget is fixed ($\varepsilon = 1.0$), increasing the distillation ratio from 1% to 10% improves model accuracy, but with diminishing marginal returns. This suggests that a 5% distillation ratio is a good balance point.

This figure provides practitioners with an intuitive basis for selecting appropriate parameters based on the privacy and performance requirements of their specific application scenarios.

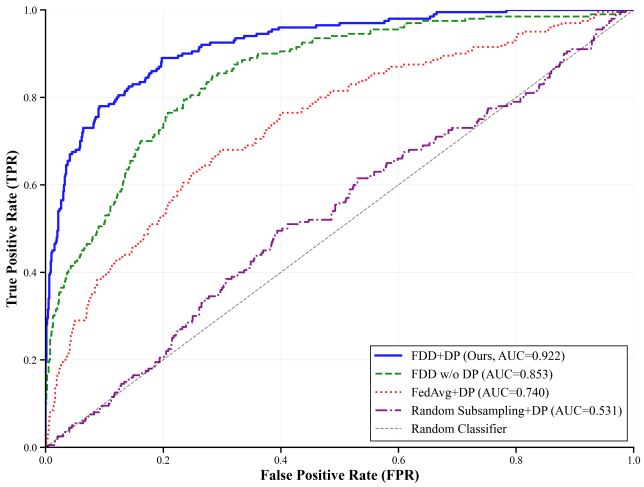

Fig. 2. Comparison of ROC curves for anomaly detection on the TCGA dataset. The x-axis is the False Positive Rate (FPR), and the y-axis is the True Positive Rate (TPR). Our method (FDD+DP, solid blue line) achieves the highest Area Under the Curve (AUC), indicating superior overall performance in identifying rare cases.

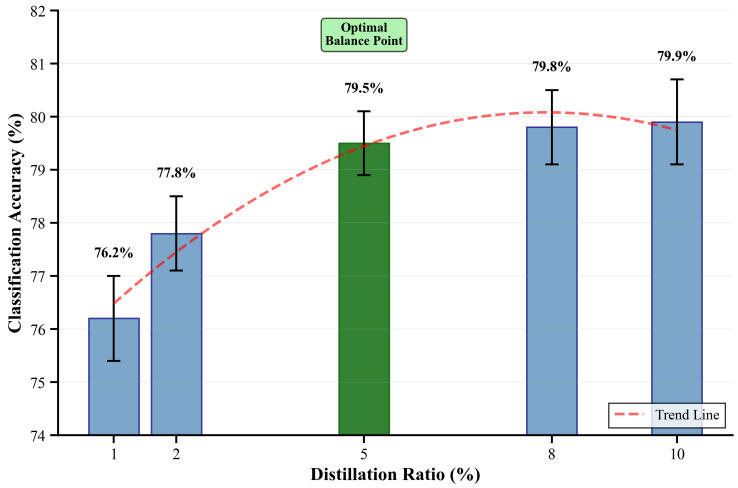

Fig. 3. The effect of different distillation ratios on classification accuracy.

5.3 Ablation Study

To verify the necessity of each core component in our framework, we conducted an ablation study. As shown in Table 5, removing any key module results in a significant performance drop, demonstrating the integrity of our design. **Without Centralized Pre-training**: The convergence speed is significantly slower (communication rounds increase by about 40%), and the final accuracy decreases, proving that the good initialization provided by pre-training is crucial. **Without Neural Feature Distillation**: Accuracy drops significantly, indicating that our NCF-based distillation method can more effectively capture the essence of the data. **Without Differential Privacy**: Although accuracy is highest, it sacrifices privacy protection and cannot meet medical application requirements.

Table 5. Ablation results on MIMIC-III (ε=1.0, distill ratio=5%)

Setting	Acc. (%) ↑	Rounds to Converge ↓	AUC ↑
Full (FDD+DP)	79.5	48	0.89
No Pre-training	77.2	67	0.85
Random Distillation	76.8	49	0.83
No DP (FDD only)	81.9	43	0.91

6 Discussion

Through comprehensive experiments, this study has demonstrated the effectiveness and superiority of our proposed privacy-enhanced federated data distillation framework in multi-center medical applications. The experimental results are not only statistically significant but also provide deep insights into several important aspects of the system.

We have achieved a balance between privacy protection, model performance, and communication efficiency. Compared to traditional federated learning (FedAvg) and simple privacy-enhancing methods (FedAvg+DP), our framework first uses local data distillation based on neural characteristic functions to compress large-scale, heterogeneous raw data into small-scale, information-rich synthetic datasets. This step not only drastically reduces the communication burden but also creates the conditions for effective privacy protection. Second, applying differential privacy during the distillation process and using the resulting data for server-side pre-training provides a robust starting point for the subsequent federated fine-tuning. Our ablation study clearly quantifies the contribution of each component, confirming that the pre-training and distillation strategies are key to improving performance.

A crucial aspect of our framework is the trade-off between communication efficiency and local computation. While our method significantly cuts down on data transmission, the local data distillation phase introduces a one-time computational cost at each client. For well-resourced medical centers, this upfront

cost is generally acceptable. However, for extremely resource-constrained edge devices, the computational demand of distillation itself could become a new bottleneck. This highlights a potential limitation and an important direction for future research, such as exploring more lightweight distillation algorithms or adaptive strategies that balance distillation complexity with device capabilities.

Despite the promising results, this study has some limitations. The framework's performance under extreme data heterogeneity (e.g., when local datasets have disjoint label sets) could be further improved. Additionally, optimizing the allocation of the privacy budget across the distillation and federated training phases remains an open and challenging research question.

7 Conclusion

This paper addresses the challenges of privacy, efficiency, and performance in collaborative learning with multi-center medical data by proposing and implementing a privacy-enhanced federated data distillation framework. The framework innovatively integrates neural characteristic function-driven data distillation, differential privacy protection, server-side pre-training, and federated optimization. Comprehensive experimental evaluations show that, compared to existing baseline methods, our framework significantly improves model predictive accuracy and anomaly detection capabilities while ensuring strict privacy and reducing communication overhead by more than an order of magnitude.

Through experimental analysis, we have quantified the trade-off relationship between distillation ratio, privacy budget, and model performance. Through ablation studies, we have verified the effectiveness and robustness of each component of the system. This research offers a practical and efficient solution for collaborative data analysis in the medical field. Our work complements existing federated learning studies by bridging the gaps between data privacy, communication efficiency, and robust anomaly detection. Instead of offering isolated improvements, it presents a cohesive, end-to-end solution where data distillation and privacy preservation are deeply coupled with the learning process through a synergistic pre-training phase. This design is tailored for practical deployment in sensitive, distributed healthcare environments, thereby advancing the applicability of collaborative AI in medicine.

Acknowledgments

This work was partially supported by JSPS Grant-in-Aid for Scientific Research (C) 23K11103.

References

1. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine **37**(3), 50–60 (2020)

2. Vajrobol, V., Aggarwal, N., Baranwal, P., Saxena, G.J., Pundir, A., Singh, S.: Navigating bias and ensuring fairness in federated learning: An in-depth exploration of data distribution, iid, and non-iid challenges. Federated Intelligent System for Healthcare: A Practical Guide pp. 253–291 (2025)
3. Geiping, J., Bauermeister, H., Dröge, H., Moeller, M.: Inverting gradients-how easy is it to break privacy in federated learning? In: Advances in Neural Information Processing Systems, vol. 33, pp. 16,937–16,947 (2020)
4. Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, U., et al.: Extracting training data from large language models. In: 30th USENIX security symposium (USENIX Security 21), pp. 2633–2650 (2021)
5. Abadi, M., Chu, A., Goodfellow, I., et al.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, pp. 308–318 (2016)
6. Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al.: Advances and open problems in federated learning. Foundations and trends® in machine learning **14**(1–2), 1–210 (2021)
7. Wang, S., Yang, Y., Liu, Z., Sun, C., Hu, X., He, C., Zhang, L.: Dataset distillation with neural characteristic function: A minmax perspective. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 25,570–25,580 (2025)
8. Cholakoska, A., Pfitzner, B., Gjoreski, H., Rakovic, V., Arnrich, B., Kalendar, M.: Differentially private federated learningfor anomaly detection in ehealth networks. In: Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, pp. 514–518 (2021)
9. Rieke, N., Hancox, J., Li, W., et al.: The future of digital health with federated learning. NPJ digital medicine **3**(1), 1–7 (2020)
10. Li, L., Fan, Y., Tse, M., Lin, K.Y.: A review of applications in federated learning. Computers & Industrial Engineering **149**, 106,854 (2020)
11. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST) **10**(2), 1–19 (2019)
12. Chen, H., Wang, H., Long, Q., Jin, D., Li, Y.: Advancements in federated learning: Models, methods, and privacy. ACM Computing Surveys **57**(2), 1–39 (2024)
13. Sun, L., Wu, J.: A scalable and transferable federated learning system for classifying healthcare sensor data. IEEE Journal of Biomedical and Health Informatics **27**(2), 866–877 (2022)
14. Rao, B., Zhang, J., Wu, D., Zhu, C., Sun, X., Chen, B.: Privacy inference attack and defense in centralized and federated learning: A comprehensive survey. IEEE Transactions on Artificial Intelligence (2024)
15. Shukla, S., Rajkumar, S., Sinha, A., Esha, M., Elango, K., Sampath, V.: Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity. Scientific Reports **15**(1), 13,061 (2025)
16. Yin, X., Zhu, Y., Hu, J.: A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions. ACM Computing Surveys (CSUR) **54**(6), 1–36 (2021)
17. Hu, K., Gong, S., Zhang, Q., Seng, C., Xia, M., Jiang, S.: An overview of implementing security and privacy in federated learning. Artificial Intelligence Review **57**(8), 204 (2024)

18. Li, H., Ge, L., Tian, L.: Survey: federated learning data security and privacy-preserving in edge-internet of things. Artificial Intelligence Review **57**(5), 130 (2024)
19. Yu, R., Liu, S., Wang, X.: Dataset distillation: A comprehensive review. IEEE transactions on pattern analysis and machine intelligence **46**(1), 150–170 (2023)
20. Wu, C., Wu, F., Lyu, L., Huang, Y., Xie, X.: Communication-efficient federated learning via knowledge distillation. Nature communications **13**(1), 2032 (2022)
21. Gong, X., Sharma, A., Karanam, S., Wu, Z., Chen, T., Doermann, D., Innanje, A.: Ensemble attention distillation for privacy-preserving federated learning. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 15,076–15,086 (2021)
22. Itahara, S., Nishio, T., Koda, Y., Morikura, M., Yamamoto, K.: Distillation-based semi-supervised federated learning for communication-efficient collaborative training with non-iid private data. IEEE Transactions on Mobile Computing **22**(1), 191–205 (2021)
23. Zhang, C., Yang, S., Mao, L., Ning, H.: Anomaly detection and defense techniques in federated learning: a comprehensive review. Artificial Intelligence Review **57**(6), 150 (2024)
24. Kong, X., Zhang, W., Wang, H., Hou, M., Chen, X., Yan, X., Das, S.K.: Federated graph anomaly detection via contrastive self-supervised learning. IEEE Transactions on Neural Networks and Learning Systems (2024)
25. Lu, Y., Yang, T., Zhao, C., Chen, W., Zeng, R.: A swarm anomaly detection model for iot uavs based on a multi-modal denoising autoencoder and federated learning. Computers & Industrial Engineering **196**, 110,454 (2024)
26. Li, X., Xiong, Z., Lian, Z., Liu, Y.: Federated learning with dataset distillation. In: arXiv preprint arXiv:2104.05637 (2021)
27. Abdel-Basset, M., Moustafa, N., Hawash, H.: Privacy-preserved generative network for trustworthy anomaly detection in smart grids: A federated semisupervised approach. IEEE transactions on industrial informatics **19**(1), 995–1005 (2022)
28. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. In: Advances in neural information processing systems, vol. 30 (2017)
29. Raza, A., Tran, K.P., Koehl, L., Li, S.: Anofed: Adaptive anomaly detection for digital health using transformer-based federated learning and support vector data description. Engineering Applications of Artificial Intelligence **121**, 106,051 (2023)
30. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics, pp. 1273–1282. PMLR (2017)
31. Geyer, R.C., Klein, T., Nabi, M.: Differentially private federated learning: A client level perspective. In: arXiv preprint arXiv:1712.07557 (2017)