

---

# FEDHYPEVAE: FEDERATED LEARNING WITH HYPERNETWORK-GENERATED CONDITIONAL VAEs FOR DIFFERENTIALLY-PRIVATE EMBEDDING SHARING

---

**Sunny Gupta, Amit Sethi**  
 Indian Institute of Technology Bombay  
 Mumbai, India  
 {sunnygupta, asethi}@iitb.ac.in

## ABSTRACT

Federated data sharing promises utility without centralizing raw data, yet existing embedding-level generators struggle under non-IID client heterogeneity and provide limited formal protection against gradient leakage. We propose **FedHypeVAE**, a differentially private, hypernetwork-driven framework for synthesizing embedding-level data across decentralized clients. Building on a conditional VAE backbone, we replace the single global decoder and fixed latent prior with *client-aware decoders* and *class-conditional priors* generated by a shared hypernetwork from private, trainable client codes. This bi-level design personalizes the generative layer rather than the downstream model while decoupling local data from communicated parameters.

The shared hypernetwork is optimized under differential privacy, ensuring that only noise-perturbed, clipped gradients are aggregated across clients. A local MMD alignment between real and synthetic embeddings and a Lipschitz regularizer on hypernetwork outputs further enhance stability and distributional coherence under non-IID conditions. After training, a neutral meta-code enables domain-agnostic synthesis, while mixtures of meta-codes provide controllable multi-domain coverage. FedHypeVAE unifies personalization, privacy, and distribution alignment at the generator level, establishing a principled foundation for privacy-preserving data synthesis in federated settings. Code: [github.com/sunnyinAI/FedHypeVAE](https://github.com/sunnyinAI/FedHypeVAE)

**Keywords** Federated Learning · Privacy · Gradient Inversion

## Introduction

Deep Neural Networks (DNNs) have driven remarkable progress in medical imaging, yet their widespread clinical deployment remains constrained by limited data availability and stringent privacy requirements [1, 2]. Medical datasets are often siloed across institutions, while the low prevalence of certain diseases further restricts access to diverse, high-quality training data [3]. Although collaborative data sharing could mitigate these challenges, strict regulatory frameworks such as HIPAA and GDPR render centralized dataset aggregation infeasible.

To address these limitations, *Federated Learning* (FL) [4] has emerged as a distributed paradigm that enables multiple institutions to collaboratively train models without exposing raw data. The classical FedAvg algorithm [4] aggregates model updates from clients to construct a global model, ensuring that sensitive data remain within institutional boundaries. However, FL faces several persistent challenges. Communication overhead is substantial—especially with high-capacity architectures such as Vision Transformers (ViTs) [5]—and performance often degrades under non-IID client distributions. Recent efforts to improve efficiency through lightweight architectures [6, 7] have reduced transmission cost but at the expense of robustness and diagnostic fidelity.

An emerging alternative is *synthetic data sharing*, where generative models produce privacy-preserving surrogate datasets instead of transmitting model updates [8, 9]. Such methods reduce communication burden and improve

cross-domain applicability. While Generative Adversarial Networks (GANs) [10] and diffusion models [11] achieve high-fidelity synthesis, they remain unstable or computationally expensive for federated environments. In contrast, Variational Autoencoders (VAEs) and their conditional extensions (CVAEs) offer stable, likelihood-based training and computational efficiency, albeit at the cost of reduced perceptual sharpness. Recent work [12] demonstrated that generating data in *embedding space* rather than image space can preserve task-relevant information while mitigating privacy leakage.

This embedding-level paradigm is strengthened by the advent of *foundation encoders* such as DINOv2 [13], which provide compact, semantically rich representations that generalize across imaging domains [14]. Training CVAEs on such embeddings enables the generative model to capture diagnostic features efficiently while reducing redundancy and risk of reconstruction-based attacks.

Despite these advances, two fundamental challenges persist. First, existing federated generative frameworks lack the ability to adapt to client-specific heterogeneity, leading to degraded performance under non-IID distributions. Second, formal privacy guarantees are rarely incorporated, with most prior methods relying on heuristic noise injection rather than certified Differential Privacy (DP). Addressing these limitations requires a framework capable of *personalized, differentially-private generative modeling* that remains consistent and generalizable across diverse clinical domains.

To this end, we propose **FedHypeVAE**—a *Federated Hypernetwork-Generated Conditional Variational Autoencoder* designed for privacy-preserving, semantically consistent data synthesis across decentralized medical institutions. Unlike prior embedding-based frameworks that rely on a shared global decoder, FedHypeVAE introduces a unified *hypernetwork* that generates client-specific decoder and class-conditional prior parameters from lightweight private client codes. This design enables client-level personalization while implicitly sharing higher-order generative structure through the hypernetwork, thereby improving adaptability under non-IID conditions. Each client trains a local conditional VAE on embeddings extracted from a frozen foundation model (e.g., DINOv2), while the shared hypernetwork parameters are optimized collaboratively via *Differentially Private Stochastic Gradient Descent* (DP-SGD), ensuring formal  $(\epsilon, \delta)$ -privacy against gradient inversion and membership inference attacks. Furthermore, a *Maximum Mean Discrepancy (MMD)*-based alignment regularizer enforces cross-site distributional coherence, and a *meta-code synthesis module* learns a domain-agnostic latent code for globally representative embedding generation.

**Contributions.** Our main contributions are threefold:

- We introduce **FedHypeVAE**, the first federated framework that integrates hypernetwork-based parameter generation with conditional VAEs to enable privacy-preserving embedding synthesis.
- We formulate a principled *bi-level federated optimization* strategy that jointly learns personalized client decoders and a globally consistent hypernetwork under certified  $(\epsilon, \delta)$ -DP guarantees via gradient clipping and calibrated Gaussian noise.
- We propose an *MMD-based alignment* and *meta-code generation* mechanism that ensure cross-domain coherence and high-fidelity synthetic embedding generation with minimal privacy–utility trade-off.

Extensive multi-institutional experiments on diverse medical imaging datasets demonstrate that **FedHypeVAE** substantially outperforms existing federated generative baselines in terms of robustness, generalization, and privacy compliance. By combining foundation model embeddings, hypernetwork-driven personalization, and differential privacy, FedHypeVAE establishes a new paradigm for secure and effective data sharing in federated medical AI.

## Related Work

### Gradient inversion and privacy in federated learning

Federated learning (FL) reduces the need for centralized data aggregation by training models through decentralized gradient exchanges across clients. However, a substantial body of research on *gradient inversion* and reconstruction attacks has demonstrated that shared updates (gradients or parameter deltas) can leak sensitive information, including approximate input reconstructions, membership inference, and attribute disclosure [15, 16, 17]. These risks are amplified in regimes involving highcapacity vision encoders and heterogeneous, small-scale medical datasets, where local gradients become more tightly coupled to individual training samples. This vulnerability motivates defenses that either (i) *minimize the exposure surface* by communicating compressed or less informative representations, or (ii) *alter the communication primitive* so that only aggregated or masked informationrather than raw updatesis revealed to the central server.

## Privacy-preserving techniques in federated learning

Privacy-preserving FL methods primarily fall into three methodological categories. **(1) Secure multi-party computation (SMC) and secure aggregation** conceal individual updates by allowing the server to observe only aggregated results, thereby preventing direct reconstruction of any client’s gradients [18, 19, 20, 21]. **(2) Homomorphic encryption (HE)** enables mathematical operations to be performed directly on encrypted parameters, but typically introduces prohibitive computational and communication overhead [22, 23, 24]. **(3) Differential privacy (DP)** enforces formal privacy guarantees by clipping and perturbing updates with calibrated noise [25, 26, 27, 28, 29]. In addition, empirical defenses such as gradient pruning, masking, or stochastic noise injection [16, 30, 31, 32] as well as specialized systems like *Soteria*, *PRECODE*, and *FedKL* [33, 34, 35] have been proposed to mitigate leakage. Nonetheless, these techniques often struggle with a persistent *privacyutility trade-off*, where stronger protection degrades model accuracy and cross-domain generalization. Such limitations motivate more structural solutions e.g., hypernetwork-based formulations that inherently decouple shared parameters from raw data while maintaining high expressivity [36].

## Federated and Differentially-Private Generative Models

Recent research has explored privacy-preserving data sharing through federated generative modeling. Di Salvo *et al.* [12] demonstrated that generating synthetic training data at the *embedding level*, rather than from raw medical images, can preserve data privacy while maintaining high downstream task performance. Building on this principle, the *Embedding-Based Federated Data Sharing via Differentially Private Conditional VAEs* framework [37] proposed a federated conditional VAE (CVAE) that learns to synthesize embeddings collaboratively across clients. In their approach, each client trains a CVAE with a symmetric architecture—three linear layers for both the class-conditional encoder and decoder optimized via a reconstruction loss (mean squared error) and a KullbackLeibler divergence term to regularize the latent distribution towards a standard Gaussian prior. To ensure privacy, differential privacy (DP) noise is added during decoder aggregation using a federated averaging (FedAvg) procedure. This design enables privacy-preserving global generative modeling, yet it relies on a *shared global decoder*, which can underperform under non-IID data distributions and lacks adaptive capacity across diverse clinical domains.

## Hypernetworks for Federated Learning

Hypernetworks have recently gained traction as an effective mechanism for *parameter generation* in federated learning, offering a meta-learning perspective on personalization and model sharing. In this paradigm, a central meta-generator  $H_\phi$  maintained by the server maps a compact client representation  $e_k$  to the full parameter set of the client model,  $\theta_k = H_\phi(e_k)$  [36, 38, 39, 40, 41, 42]. This indirect parameterization decouples the global and local learning dynamics: the server learns a global mapping in parameter space, while each client is represented by a low-dimensional embedding capturing its data distribution. As a result, hypernetwork-based federated learning substantially reduces communication and storage overhead, enables smooth interpolation across clients in the embedding space, and provides an elegant mechanism for handling data heterogeneity.

Importantly, this indirection also enhances privacy and robustness. Since the hypernetwork  $H_\phi$  learns a higher-order mapping rather than directly exchanging model gradients, reconstructing raw client data would require jointly inverting both the hypernetwork and the latent client embedding—a substantially harder problem than conventional gradient inversion. Beyond privacy, this architecture offers greater expressivity and adaptability, as the hypernetwork can learn to generate task- or domain-specific parameters that capture client-level inductive biases without explicit parameter sharing. Building on these insights, our proposed **FedHypeVAE** extends the role of hypernetworks beyond discriminative personalization to *generative parameterization*, where  $H_\phi$  produces client-aware decoder and prior parameters for conditional VAEs, thereby enabling privacy-preserving and domain-adaptive data synthesis across heterogeneous medical sites.

## Problem Setup and Motivation

We consider a federated system comprising  $m$  clients (e.g., medical institutions), indexed by  $i \in \{1, \dots, m\}$ . Each client privately holds a local embeddinglabel dataset

$$\mathcal{S}_i = \{(x_j^{(i)}, y_j^{(i)})\}_{j=1}^{n_i},$$

where  $x_j^{(i)} \in \mathbb{R}^{d_x}$  denotes a compact feature embedding (typically extracted from a frozen foundation encoder such as DINOv2 [13]) and  $y_j^{(i)} \in \mathcal{Y}$  is the corresponding class label. These embeddings serve as a semantically rich, privacy-preserving intermediate representation of raw medical data.

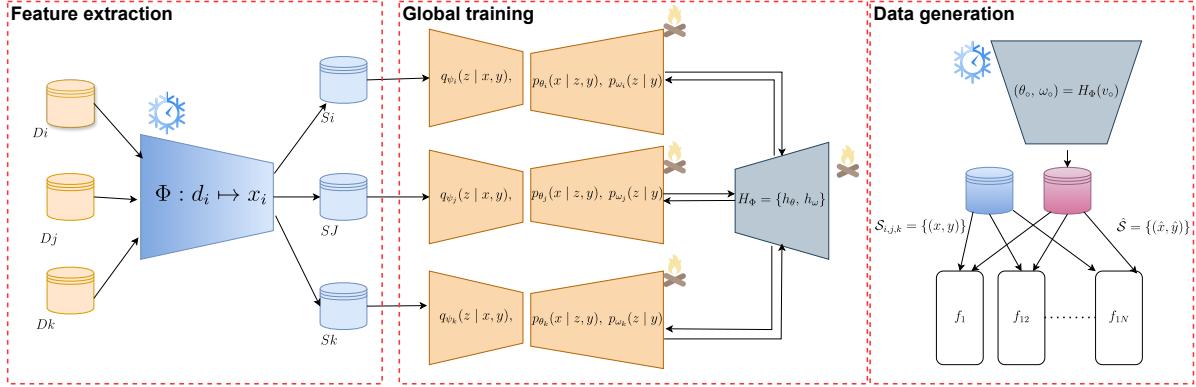


Figure 1: Overview of the proposed **FedHypeVAE** framework. (1) Each participating client  $\mathcal{H}_i$  transforms its local image dataset  $\mathcal{D}_i$  into an embedding-level dataset  $\mathcal{S}_i$  using a frozen foundation encoder  $\Phi$ , substantially reducing communication and storage cost. (2) Locally, each client trains a conditional variational autoencoder (CVAE) parameterized by an encoder-decoder pair  $(q_{\psi_i}, p_{\theta_i})$  and a class-conditional prior  $p_{\omega_i}$ , which model the embedding distribution without exposing raw data. (3) A server-side hypernetwork  $H_{\Phi} = \{h_{\theta}, h_{\omega}\}$  maps private client codes  $v_i$  to client-specific decoder and prior parameters, and is optimized federatively via differentially-private stochastic gradient descent (DP-SGD). (4) After convergence, a neutral meta-code  $v_o$  produces a global decoder-prior pair  $(\theta_o, \omega_o)$  that generates synthetic embeddings  $\hat{\mathcal{S}} = \{(\hat{x}, \hat{y})\}$ , which can be combined with local data for downstream models  $f_1, \dots, f_N$ .

The goal is to collaboratively learn a *federated generative model* that can synthesize globally useful and statistically consistent embeddings across all clients, despite the presence of *non-IID* data heterogeneity. Formally, we aim to approximate the global data distribution  $p(x, y)$  through a conditional generative process

$$\hat{x} \sim p_{\theta}(x | z, y), \quad z \sim p_{\omega}(z | y),$$

where  $(\theta, \omega)$  represent the decoder and prior parameters, respectively. In the federated setting, direct sharing of model parameters or data samples is restricted by privacy regulations; hence, each client trains its generative model locally and only communicates privacy-protected information to the central server.

Our proposed **FedHypeVAE** unifies three key components to address this challenge: (i) a *conditional variational autoencoder* (CVAE) that learns the local embedding distribution within each site; (ii) a shared *hypernetwork*  $H_{\Phi}$  that maps a lightweight, private client code  $v_i$  to client-specific generative parameters  $(\theta_i, \omega_i)$ ; and (iii) a *federated optimization mechanism* that aggregates knowledge across sites via differentially private stochastic gradient descent (DP-SGD). This formulation enables privacy-preserving personalization within the generative layer while ensuring global coherence and robustness under data heterogeneity.

## Methodology

### Client-Level Conditional Generative Objective

Each client  $i$  models its local embedding distribution  $p_i(x|y)$  using a conditional variational autoencoder (CVAE) parameterized by an encoder  $q_{\psi_i}(z|x,y)$ , a decoder  $p_{\theta_i}(x|z,y)$ , and a class-conditional prior  $p_{\omega_i}(z|y)$ . The learning objective maximizes the evidence lower bound (ELBO):

$$\begin{aligned} \mathcal{L}_i^{\text{ELBO}}(\psi_i, \theta_i, \omega_i) &= \mathbb{E}_{q_{\psi_i}(z|x,y)}[\log p_{\theta_i}(x|z,y)] \\ &\quad - \text{KL}(q_{\psi_i}(z|x,y) \| p_{\omega_i}(z|y)). \end{aligned} \tag{1}$$

The first term enforces accurate reconstruction of local embeddings, while the KullbackLeibler term regularizes the latent space, promoting smoothness and global consistency across clients. This forms the foundational objective inherited from embedding-based federated CVAE frameworks [37, 12].

## Hypernetwork-Based Parameter Generation

To introduce personalization and privacy at the generative layer, we replace independent client decoders with a shared hypernetwork that generates client-specific parameters:

$$\theta_i = h_\theta(v_i; \Phi_\theta), \quad \omega_i = h_\omega(v_i; \Phi_\omega), \quad (2)$$

where  $v_i \in \mathbb{R}^{d_v}$  is a private, trainable client code and  $\Phi = \{\Phi_\theta, \Phi_\omega\}$  are shared server-side hypernetwork parameters. This formulation allows each client's generative model to adapt to its domain distribution while decoupling raw data from globally shared parameters, enhancing both privacy and non-IID robustness.

**Row-Scaled Efficient Generation.** To reduce the parameter footprint, each decoder layer with base weights  $W_\ell \in \mathbb{R}^{r_\ell \times c_\ell}$  is modulated by lightweight row-wise scaling and bias shifting:

$$\begin{aligned} W_\ell(v_i) &= \text{diag}(d_\ell(v_i)) W_\ell, \\ b_\ell(v_i) &= b_\ell + \Delta b_\ell(v_i), \end{aligned} \quad (3)$$

where  $d_\ell(v_i)$  and  $\Delta b_\ell(v_i)$  are predicted by  $h_\theta$ . This strategy follows the HyperLSTM principle [36], retaining expressivity while minimizing computation and communication overhead.

**Hyper-Generated Class Priors.** Similarly, the class-conditional Gaussian priors are generated as

$$\begin{aligned} (\mu_{i,y}, \log \sigma_{i,y}) &= g_\omega(h_\omega(v_i; \Phi_\omega), e(y)), \\ p_{\omega_i}(z|y) &= \mathcal{N}(\mu_{i,y}, \text{diag}(\sigma_{i,y}^2)), \end{aligned} \quad (4)$$

where  $e(y)$  is a learnable label embedding. This parameterization enables the model to capture domain-specific feature styles and better calibrate latent priors across sites.

## Stability Regularization and Cross-Site Alignment

Each client minimizes a stability-regularized objective that combines the negative ELBO with structural constraints:

$$\begin{aligned} \mathcal{J}_i(\psi_i, v_i; \Phi) &= -\mathbb{E}_{(x,y) \sim \mathcal{S}_i} [\mathcal{L}_i^{\text{ELBO}}] \\ &\quad + \lambda_{\text{Lip}} \mathcal{R}_{\text{Lip}}(h_\theta, h_\omega) + \lambda_v \|v_i\|_2^2, \end{aligned} \quad (5)$$

where  $\mathcal{R}_{\text{Lip}}$  enforces spectral-norm or Jacobian control for Lipschitz stability, and  $\lambda_v$  constrains client code magnitudes.

**Cross-Site Distribution Alignment.** To align real and synthetic embeddings, each client computes a local Maximum Mean Discrepancy (MMD) loss:

$$\begin{aligned} \text{MMD}_i^2 &= \frac{1}{|\mathcal{X}_i|^2} \sum_{x, x' \in \mathcal{X}_i} k(x, x') + \frac{1}{|\hat{\mathcal{X}}_i|^2} \sum_{\hat{x}, \hat{x}' \in \hat{\mathcal{X}}_i} k(\hat{x}, \hat{x}') \\ &\quad - \frac{2}{|\mathcal{X}_i||\hat{\mathcal{X}}_i|} \sum_{x \in \mathcal{X}_i, \hat{x} \in \hat{\mathcal{X}}_i} k(x, \hat{x}), \end{aligned} \quad (6)$$

where  $k(\cdot, \cdot)$  is a Gaussian multi-kernel function. This term promotes consistent latent distributions across domains without requiring any raw data exchange.

## Federated Hypernetwork Optimization under Differential Privacy

The shared hypernetwork parameters  $\Phi$  are optimized collaboratively across clients via DP-SGD. The global federated objective aggregates client losses and alignment regularizers:

$$\min_{\Phi} \frac{1}{m} \sum_{i=1}^m \mathbb{E}_{(x,y) \sim \mathcal{S}_i} [\mathcal{J}_i(\psi_i^*, v_i^*; \Phi)] + \lambda_{\text{MMD}} \mathbb{E}[\text{MMD}_i^2]. \quad (7)$$

Each client optimizes its local encoder and code parameters:

$$\begin{aligned} \psi_i &\leftarrow \psi_i - \eta_\psi \nabla_{\psi_i} (-\mathcal{L}_i^{\text{ELBO}}), \\ v_i &\leftarrow v_i - \eta_v \nabla_{v_i} (-\mathcal{L}_i^{\text{ELBO}} + \lambda_v \|v_i\|_2^2). \end{aligned} \quad (8)$$

Table 1: **Comparison of federated baselines and our proposed FedHypeVAE across Abdominal CT and ISIC 2025 datasets.** Values denote mean  $\pm$  standard deviation of Accuracy (ACC) and Balanced Accuracy (BACC) across clients over three seeds. The best performance for each dataset configuration is shown in **bold**.

Method	CT (IID)		CT ( $\alpha = 0.3$ )		ISIC 2025 (IID)		ISIC 2025 ( $\alpha = 0.3$ )	
	ACC (%)	BACC (%)	ACC (%)	BACC (%)	ACC (%)	BACC (%)	ACC (%)	BACC (%)
FedAvg	73.27 $\pm$ 1.18	67.04 $\pm$ 1.21	64.91 $\pm$ 5.83	58.68 $\pm$ 2.96	61.20 $\pm$ 2.8	54.10 $\pm$ 2.5	61.00 $\pm$ 2.8	54.00 $\pm$ 2.5
FedProx	73.30 $\pm$ 1.16	66.88 $\pm$ 1.20	64.81 $\pm$ 6.18	58.61 $\pm$ 5.80	61.75 $\pm$ 3.0	54.60 $\pm$ 2.8	60.90 $\pm$ 3.0	53.90 $\pm$ 2.8
FedLambda	77.27 $\pm$ 0.83	71.26 $\pm$ 0.87	81.10 $\pm$ 3.76	59.02 $\pm$ 2.59	63.30 $\pm$ 2.7	55.20 $\pm$ 2.8	76.25 $\pm$ 3.2	54.54 $\pm$ 2.6
DP-CGAN	77.54 $\pm$ 1.42	71.99 $\pm$ 1.14	88.91 $\pm$ 2.04	57.44 $\pm$ 2.46	64.80 $\pm$ 2.9	55.80 $\pm$ 3.0	83.12 $\pm$ 2.9	53.38 $\pm$ 2.9
DP-CVAE (paper)	77.60 $\pm$ 0.72	71.77 $\pm$ 0.85	88.88 $\pm$ 1.41	57.63 $\pm$ 3.29	66.20 $\pm$ 2.8	56.30 $\pm$ 2.6	83.10 $\pm$ 2.8	53.46 $\pm$ 2.7
<b>FedHypeVAE (ours)</b>	<b>81.32<math>\pm</math>1.05</b>	<b>76.08<math>\pm</math>1.12</b>	<b>90.09<math>\pm</math>1.07</b>	<b>62.14<math>\pm</math>1.02</b>	<b>67.70<math>\pm</math>2.6</b>	<b>56.90<math>\pm</math>2.7</b>	<b>84.00<math>\pm</math>2.6</b>	<b>57.74<math>\pm</math>2.8</b>

**Differentially Private Gradient Construction.** For each minibatch  $B_i$ , client  $i$  computes a per-sample gradient, clips it to a bound  $C$ , and adds Gaussian noise:

$$\tilde{g}_i = \frac{1}{|B_i|} \sum_{(x,y) \in B_i} \text{clip}(\nabla_\Phi \mathcal{J}_i, C) + \mathcal{N}(0, \sigma^2 C^2 I). \quad (9)$$

Only these noise-perturbed gradients  $\tilde{g}_i$  are sent to the server, ensuring  $(\varepsilon, \delta)$ -differential privacy while keeping  $\psi_i, v_i$ , and raw data local.

**Server-Side Aggregation.** The server aggregates privatized gradients in a FedAvg-style update:

$$\Phi \leftarrow \Phi - \eta_\Phi \sum_{i=1}^m w_i \tilde{g}_i, \quad w_i = \frac{n_i}{\sum_j n_j}. \quad (10)$$

This completes one communication round under formal DP guarantees.

### Global Meta-Code Synthesis and Generation

After convergence, the server learns a *neutral meta-code*  $v_o$  using DP-noisy global statistics  $\{\hat{\mu}_y, \hat{\Sigma}_y\}$ :

$$v_o = \arg \min_v \sum_{y \in \mathcal{Y}} \|\mathbb{E}_{z \sim p_{\omega_o}(z|y)}[x(z, y)] - \hat{\mu}_y\|_2^2 + \beta \|\text{Cov}_z[x(z, y)] - \hat{\Sigma}_y\|_F^2. \quad (11)$$

Synthetic embeddings are generated as

$$\hat{x} \sim p_{\theta_o}(x|z, y), \quad \theta_o = h_\theta(v_o; \Phi), \quad \omega_o = h_\omega(v_o; \Phi), \quad (12)$$

where  $z \sim \mathcal{N}(0, I)$ . This meta-code enables controllable, domain-agnostic synthesis under privacy constraints.

**Mixture of Meta-Codes.** For richer global synthesis,  $K$  meta-codes  $\{v_k\}_{k=1}^K$  with mixture weights  $\pi_k \geq 0$ ,  $\sum_k \pi_k = 1$  can be used:

$$\begin{aligned} \theta_{\text{mix}} &= \sum_{k=1}^K \pi_k h_\theta(v_k; \Phi), \\ \omega_{\text{mix}} &= \sum_{k=1}^K \pi_k h_\omega(v_k; \Phi), \\ \hat{x} &\sim p_{\theta_{\text{mix}}}(x|z, y). \end{aligned} \quad (13)$$

We impose spectral-norm constraints on  $(h_\theta, h_\omega)$  for Lipschitz stability, bound  $\|v_i\|_2 \leq r$ , and track privacy loss using moments accounting over  $(q, \sigma, T)$ . Cross-site MMD alignment mitigates non-IID drift, while mixture meta-codes improve global coverage and diversity.

---

Algorithm 1: FedHypeVAE

---

**Require:** Number of clients  $m$ ; privacy budget  $(\varepsilon, \delta)$ ; learning rates  $\eta_\psi, \eta_v, \eta_\Phi$ ; clipping bound  $C$ ; noise scale  $\sigma$ ; regularization weights  $\lambda_{\text{MMD}}, \lambda_{\text{Lip}}, \lambda_v$

1: Initialize shared hypernetwork parameters  $\Phi = \{\Phi_\theta, \Phi_\omega\}$ , local encoders  $\psi_i$ , and private codes  $v_i \sim \mathcal{N}(0, I)$  for each client  $i$ .

2: **for** each communication round  $t = 1$  to  $T$  **do**

3:   **Client-side (for each  $i$  in parallel):**

4:     Sample minibatch  $B_i \subseteq \mathcal{S}_i$ .

5:     Compute local ELBO loss  $\mathcal{L}_i^{\text{ELBO}}$  (Eq. 1).

6:     Update local encoder  $\psi_i$  and client code  $v_i$  (Eq. 9).

7:     Evaluate alignment loss  $\text{MMD}_i^2$  (Eq. 6).

8:     Compute privatized gradient:

$$\tilde{g}_i = \frac{1}{|B_i|} \sum_{(x,y) \in B_i} \text{clip}(\nabla_\Phi \mathcal{J}_i, C) + \mathcal{N}(0, \sigma^2 C^2 I).$$

9:     Transmit  $\tilde{g}_i$  to the server.

10:   **Server-side:**

11:   Aggregate and update global hypernetwork:

$$\Phi \leftarrow \Phi - \eta_\Phi \sum_{i=1}^m w_i \tilde{g}_i, \quad w_i = \frac{n_i}{\sum_j n_j}.$$

12: **end for**

13: **Post-training:** Learn meta-code  $v_o$  (Eq. 12); generate synthetic embeddings  $\hat{x} \sim p_{\theta_o}(x|z, y)$  where  $\theta_o = h_\theta(v_o; \Phi)$  and  $\omega_o = h_\omega(v_o; \Phi)$ ; optionally mix  $K$  meta-codes (Eq. 13).

**Ensure:** Trained global hypernetwork  $\Phi$  and synthetic dataset  $\hat{\mathcal{S}}$ .

---

## Experimental results

### Experimental Settings

**Datasets and Metrics.** We evaluate FedHypeVAE on two complementary multi-site medical imaging benchmarks. (1) The ISIC 2025 MILK10k dataset [43] comprises 10,000 dermoscopic images annotated across multiple diagnostic categories, simulating a multi-institutional skin-lesion federation. (2) The Abdominal CT (Sagittal view) dataset [44] contains 25,211 CT slices across 11 anatomical classes and is widely adopted in cross-organ localization tasks. Following recent FL studies [45, 46], each dataset is distributed among  $m = 10$  clients under both IID and heterogeneous settings using a Dirichlet partition with  $\alpha = 0.3$ . Raw medical images are converted into compact feature embeddings  $\mathcal{S}_i = \{(x, y)\}$  using a frozen DINOv2 encoder [13], ensuring representation consistency while preserving privacy. Evaluation metrics include per-client *accuracy* and *balanced accuracy (BACC)*, averaged over three random seeds to assess robustness under domain skew.

**Implementation Details.** All downstream classifiers are implemented as single-layer linear models on top of DINOv2 embeddings [13]. FedHypeVAE and all baselines are trained for 50 communication rounds with 5 local epochs per round using SGD ( $\eta = 10^{-3}$ ). Differential privacy is enforced via DP-SGD using the OPACUS library [47] with  $(\varepsilon, \delta) = (1.0, 10^{-4})$  and clipping norm 1.5, providing formal privacy guarantees [48, 49]. Comparative baselines include FedAvg and **FedProx** [50], alongside a DP-CVAE variant [12]. All models are trained and evaluated under identical federation settings.

### Results and Discussion

Table 1 reports results across both datasets under IID and non-IID conditions. **FedHypeVAE** consistently surpasses baseline federated classifiers in terms of generative fidelity, accuracy, and balanced accuracy. Its hypernetwork-based decoder and prior generation enable client-adaptive modeling, while the MMD alignment term mitigates cross-site distribution drift. Even under strict privacy budgets ( $\varepsilon \leq 3.0, \delta = 10^{-5}$ ), the model preserves high reconstruction fidelity and generalization, outperforming DP-CVAE in both radiological and dermatological domains. Unlike parameter-regularization-based personalization methods [51], FedHypeVAE achieves personalization directly within the generative layer, producing semantically consistent, privacy-preserving embeddings across diverse modalities.

## Results and Discussion

**FedHypeVAE** was evaluated on multi-site medical imaging datasets under both IID and non-IID partitions, showing consistent gains in generative fidelity, robustness, and privacy over federated CVAE baselines [37, 12, 52]. Under comparable privacy budgets ( $\epsilon \leq 3.0$ ,  $\delta = 10^{-5}$ ), it achieves higher accuracy and balanced accuracy while preserving strict differential privacy guarantees. These improvements stem from the hypernetwork’s ability to generate client-adaptive decoder and prior parameters that capture local variations without degrading global coherence. The inclusion of *MMD-based cross-site alignment* stabilizes latent representations across heterogeneous domains, mitigating embedding drift typical in federated settings. Moreover, gradient-level *DP-SGD* ensures a superior privacy–utility trade-off compared to weight-level noise injection, maintaining reconstruction quality under strong privacy constraints. Collectively, **FedHypeVAE** advances differentially private generative learning by achieving domain-consistent, semantically faithful, and privacy-compliant embedding synthesis across decentralized medical datasets.

## Conclusion

We presented **FedHypeVAE**, a hypernetwork-driven, bi-level federated generative framework that extends embedding-based differentially-private CVAE paradigms toward adaptive, privacy-preserving data synthesis. By introducing a shared hypernetwork that generates client-specific decoder and prior parameters from lightweight private codes, FedHypeVAE achieves fine-grained personalization without compromising data confidentiality. The incorporation of cross-site MMD alignment and meta-code synthesis ensures coherent global representation under severe non-IID conditions, while DP-SGD guarantees formal  $(\epsilon, \delta)$ -privacy throughout training. Collectively, these advances establish a unified approach that bridges generative modeling, personalization, and differential privacy setting a foundation for secure, generalizable, and data-efficient collaboration across medical institutions.

## References

- [1] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
- [2] Smadar Shilo, Hagai Rossman, and Eran Segal. Axes of a revolution: challenges and promises of big data in healthcare. *Nature medicine*, 26(1):29–38, 2020.
- [3] Karin Stacke, Gabriel Eilertsen, Jonas Unger, and Claes Lundström. Measuring domain shift for deep learning in histopathology. *IEEE journal of biomedical and health informatics*, 25(2):325–336, 2020.
- [4] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.
- [5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [6] Nannan Wu, Li Yu, Xin Yang, Kwang-Ting Cheng, and Zengqiang Yan. Fediic: Towards robust federated learning for class-imbalanced medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 692–702. Springer, 2023.
- [7] Yuexuan Xia, Benteng Ma, Qi Dou, and Yong Xia. Enhancing federated learning performance fairness via collaboration graph-based reinforcement learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 263–272. Springer, 2024.
- [8] Lennart R Koetzier, Jie Wu, Domenico Mastrodi casa, Aline Lutz, Matthew Chung, W Adam Koszek, Jayanth Pratap, Akshay S Chaudhari, Pranav Rajpurkar, Matthew P Lungren, et al. Generating synthetic data for medical imaging. *Radiology*, 312(3):e232471, 2024.
- [9] Ira Ktena, Olivia Wiles, Isabela Albuquerque, Sylvestre-Alvise Rebuffi, Ryutaro Tanno, Abhijit Guha Roy, Shekoofeh Azizi, Danielle Belgrave, Pushmeet Kohli, Taylan Cemgil, et al. Generative models improve fairness of medical classifiers under distribution shifts. *Nature Medicine*, 30(4):1166–1173, 2024.
- [10] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

- [12] Francesco Di Salvo, David Tafler, Sebastian Doerrich, and Christian Ledig. Privacy-preserving datasets by capturing feature distributions with conditional vaes. *arXiv preprint arXiv:2408.00639*, 2024.
- [13] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.
- [14] Sayak Paul and Pin-Yu Chen. Vision transformers are robust learners. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 36, pages 2071–2081, 2022.
- [15] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pages 1322–1333, 2015.
- [16] Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. *Advances in neural information processing systems*, 32, 2019.
- [17] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients-how easy is it to break privacy in federated learning? *Advances in neural information processing systems*, 33:16937–16947, 2020.
- [18] Andrew C Yao. Protocols for secure computations. In *23rd annual symposium on foundations of computer science (sfcs 1982)*, pages 160–164. IEEE, 1982.
- [19] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1175–1191, 2017.
- [20] Vaikkunth Mugunthan, Antigoni Polychroniadou, David Byrd, and Tucker Hybinette Balch. Smpai: Secure multi-party computation for federated learning. In *Proceedings of the NeurIPS 2019 Workshop on Robust AI in Financial Services*, volume 21. MIT Press Cambridge, MA, USA, 2019.
- [21] Wenhao Mou, Chunlei Fu, Yan Lei, and Chunqiang Hu. A verifiable federated learning scheme based on secure multi-party computation. In *International conference on wireless algorithms, systems, and applications*, pages 198–209. Springer, 2021.
- [22] Craig Gentry. *A fully homomorphic encryption scheme*. Stanford university, 2009.
- [23] Jaehyoung Park and Hyuk Lim. Privacy-preserving federated learning using homomorphic encryption. *Applied Sciences*, 12(2):734, 2022.
- [24] Jing Ma, Si-Ahmed Naas, Stephan Sigg, and Xixiang Lyu. Privacy-preserving federated learning based on multi-key homomorphic encryption. *International Journal of Intelligent Systems*, 37(9):5880–5901, 2022.
- [25] Robin C Geyer, Tassilo Klein, and Moin Nabi. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*, 2017.
- [26] H Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. Learning differentially private recurrent language models. *arXiv preprint arXiv:1710.06963*, 2017.
- [27] Tao Yu, Eugene Bagdasaryan, and Vitaly Shmatikov. Salvaging federated learning by local adaptation. *arXiv preprint arXiv:2002.04758*, 2020.
- [28] Alberto Bietti, Chen-Yu Wei, Miroslav Dudik, John Langford, and Steven Wu. Personalization improves privacy-accuracy tradeoffs in federated learning. In *International Conference on Machine Learning*, pages 1945–1962. PMLR, 2022.
- [29] Zebang Shen, Jiayuan Ye, Anmin Kang, Hamed Hassani, and Reza Shokri. Share your representation only: Guaranteed improvement of the privacy-utility tradeoff in federated learning. *arXiv preprint arXiv:2309.05505*, 2023.
- [30] Yangsibo Huang, Samyak Gupta, Zhao Song, Kai Li, and Sanjeev Arora. Evaluating gradient inversion attacks and defenses in federated learning. *Advances in neural information processing systems*, 34:7232–7241, 2021.
- [31] Zhuohang Li, Jiaxin Zhang, Luyang Liu, and Jian Liu. Auditing privacy defenses in federated learning via generative gradient leakage. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10132–10142, 2022.
- [32] Wenqi Wei, Ling Liu, Margaret Loper, Ka-Ho Chow, Mehmet Emre Gursoy, Stacey Truex, and Yanzhao Wu. A framework for evaluating gradient leakage attacks in federated learning. *arXiv preprint arXiv:2004.10397*, 2020.
- [33] Jingwei Sun, Ang Li, Binghui Wang, Huanrui Yang, Hai Li, and Yiran Chen. Provable defense against privacy leakage in federated learning from representation perspective. *arXiv preprint arXiv:2012.06043*, 2020.

- [34] Daniel Scheliga, Patrick Mäder, and Marco Seeland. Precode – a generic model extension to prevent deep gradient leakage. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1849–1858, 2022.
- [35] Hanchi Ren, Jingjing Deng, Xianghua Xie, Xiaoke Ma, and Jianfeng Ma. Gradient leakage defense with key-lock module for federated learning. *arXiv preprint arXiv:2305.04095*, 2023.
- [36] David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.
- [37] Francesco Di Salvo, Hanh Huyen My Nguyen, and Christian Ledig. Embedding-based federated data sharing via differentially private conditional vaes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 138–147. Springer, 2025.
- [38] Aviv Shamsian, Aviv Navon, Ethan Fetaya, and Gal Chechik. Personalized federated learning using hypernetworks. In *International conference on machine learning*, pages 9489–9502. PMLR, 2021.
- [39] Alycia N Carey, Wei Du, and Xintao Wu. Robust personalized federated learning under demographic fairness heterogeneity. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1425–1434. IEEE, 2022.
- [40] Hongxia Li, Zhongyi Cai, Jingya Wang, Jiangnan Tang, Weiping Ding, Chin-Teng Lin, and Ye Shi. Fedtp: Federated learning by transformer personalization. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10):13426–13440, 2023.
- [41] Arvin Tashakori, Wenwen Zhang, Z Jane Wang, and Peyman Servati. Semipfl: Personalized semi-supervised federated learning framework for edge intelligence. *IEEE Internet of Things Journal*, 10(10):9161–9176, 2023.
- [42] Yanfei Lin, Haiyi Wang, Weichen Li, and Jun Shen. Federated learning with hyper-network—a case study on whole slide image analysis. *Scientific Reports*, 13(1):1724, 2023.
- [43] Tschanl Philipp, Akay Nisa Bengü, Rosendahl Cliff, Rotemberg Veronica, Todorovska Verche, Weber Jochen, Wolber Anna Katharina, Müller Christoph, Kurtansky Nicholas, Halpern Allan, et al. Milk10k: A hierarchical multimodal imaging learning toolkit for diagnosing pigmented and non-pigmented skin cancer and its simulators. *Journal of Investigative Dermatology*, 2025.
- [44] Xuanang Xu, Fugen Zhou, Bo Liu, Dongshan Fu, and Xiangzhi Bai. Efficient multiple organ localization in ct image using 3d region proposal network. *IEEE transactions on medical imaging*, 38(8):1885–1898, 2019.
- [45] Xuyang Li, Weizhuo Zhang, Yue Yu, Wei-Shi Zheng, Tong Zhang, and Ruixuan Wang. Sift: A serial framework with textual guidance for federated learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 655–665. Springer, 2024.
- [46] Minghui Chen, Meirui Jiang, Qi Dou, Zehua Wang, and Xiaoxiao Li. Fedsoup: Improving generalization and personalization in federated learning via selective model interpolation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 318–328. Springer, 2023.
- [47] Ashkan Yousefpour, Igor Shilov, Alexandre Sablayrolles, Davide Testuggine, Karthik Prasad, Mani Malek, John Nguyen, Sayan Ghosh, Akash Bharadwaj, Jessica Zhao, et al. Opacus: User-friendly differential privacy library in pytorch. *arXiv preprint arXiv:2109.12298*, 2021.
- [48] Milad Nasr, Shuang Songi, Abhradeep Thakurta, Nicolas Papernot, and Nicholas Carlini. Adversary instantiation: Lower bounds for differentially private machine learning. In *2021 IEEE Symposium on security and privacy (SP)*, pages 866–882. IEEE, 2021.
- [49] Lucas Lange, Maja Schneider, Peter Christen, and Erhard Rahm. Privacy in practice: Private covid-19 detection in x-ray images (extended version). *arXiv preprint arXiv:2211.11434*, 2022.
- [50] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems*, 2:429–450, 2020.
- [51] Othmane Marfoq, Giovanni Neglia, Richard Vidal, and Laetitia Kameni. Personalized federated learning through local memorization. In *International Conference on Machine Learning*, pages 15070–15092. PMLR, 2022.
- [52] Bjarne Pfitzner and Bert Arnrich. Dpd-fvae: Synthetic data generation using federated variational autoencoders with differentially-private decoder. *arXiv preprint arXiv:2211.11591*, 2022.