arXiv:2601.00785v1 [cs.LG] 2 Jan 2026

FedHypeVAE：利用超网络生成的条件变分自编码器实现差分隐私嵌入共享的联邦学习

Sunny Gupta，Amit Sethi印度理
工学院孟买分校，印度孟买
{sunnygupta, asethi}@iitb.ac.in

摘要

联邦数据共享有望在不集中存储原始数据的情况下实现数据效用，但现有的嵌入级生成器在非独立同分布客户端异质性条件下表现不佳，且无法充分提供针对梯度泄露的正式保护。我们提出**FedHypeVAE**，一种差分隐私、超网络驱动的框架，用于在去中心化客户端之间合成嵌入级数据。该框架以条件变分自编码器为基础，将单一全局解码器和固定潜在先验替换为共享超网络根据私有、可训练的客户端代码生成的客户端感知解码器 和类别条件先验 。这种双层设计对生成层而非下游模型进行个性化，同时将本地数据与通信参数解耦。共享超网络在差分隐私机制下进行优化，确保在客户端之间聚合的仅为经过噪声扰动和裁剪的梯度。在真实嵌入与合成嵌入之间进行本地最大均值差异对齐，并对超网络输出施加Lipschitz正则项，进一步提升非独立同分布条件下的稳定性和分布一致性。训练完成后，中性元代码可实现领域无关合成，而元代码的混合则可提供可控的多域覆盖。FedHypeVAE在生成器层面统一了个性化、隐私和分布对齐，为联邦环境中的隐私保护数据合成奠定了原则性基础。代码：github.com/sunnyinAI/FedHypeVAE

关键词 联邦学习 ·隐私 ·梯度反演

引言

深度神经网络（DNN）推动了医学影像领域取得显著进展，但其在临床上的广泛应用仍受限于数据可用性不足和严格的隐私要求 [1, 2]。医学数据集通常分散存储于各个机构，而某些疾病的低发病率进一步限制了获取多样化、高质量训练数据的渠道 [3]。尽管协作式数据共享有助于缓解这些挑战，但 HIPAA 和 GDPR 等严格的监管框架使集中汇总数据集难以实现。

为应对这些局限，联邦学习 （FL） [4] 应运而生，成为一种分布式范式，使多个机构能够协作训练模型，同时不暴露原始数据。经典的 FedAvg 算法 [4]汇总来自客户端的模型更新，以构建全局模型，确保敏感数据始终保留在机构边界内。然而，联邦学习仍面临若干持续存在的挑战。通信开销十分巨大，尤其是在采用视觉变换器（ViTs） [5]等高容量架构时；此外，非独立同分布的客户端数据分布往往会导致性能下降。近期通过轻量级架构 [6, 7] 提升效率的尝试降低了传输成本，但代价是鲁棒性和诊断保真度有所下降。

一种新兴的替代方案是合成数据共享，即由生成模型生成隐私保护代理数据集，而非传输模型更新 [8, 9]。这类方法能够减轻通信负担并提升

跨领域适用性。尽管生成对抗网络（GAN） [10] 和扩散模型 [11] 能够实现高保真合成，但在联邦环境中仍不稳定或计算成本高昂。相比之下，变分自编码器（VAE）及其条件扩展（条件变分自编码器，CVAE）具备稳定的基于似然的训练方式和较高的计算效率，但感知清晰度有所降低。近期研究 [12]表明，在嵌入空间 而非图像空间中生成数据，可以保留与任务相关的信息，同时减少隐私泄漏。

随着基础编码器的出现，这种嵌入级范式得到了进一步强化，例如 DINOv2 [13], ，它们能够提供紧凑且语义丰富的表示，并可在不同影像领域中泛化 [14]。在此类嵌入上训练条件变分自编码器，可使生成模型高效捕获诊断特征，同时减少冗余并降低遭受基于重构的攻击的风险。

尽管取得了这些进展，仍有两项根本性挑战有待解决。首先，现有的联邦生成框架缺乏适应客户端特异性异质性的能力，导致在非独立同分布下性能下降。其次，正式的隐私保证很少被纳入其中，大多数现有方法依赖启发式噪声注入，而非经过认证的差分隐私（DP）。要解决这些局限性，需要一种个性化差分隐私生成建模框架，使其在多样化的临床领域中保持一致性和泛化能力。
为此，我们提出**FedHypeVAE**——一种联邦超网络生成条件变分自编码器，旨在去中心化医疗机构之间实现保护隐私且语义一致的数据合成。不同于依赖共享全局解码器的既有嵌入式框架，FedHypeVAE 引入了统一的超网络，它根据轻量级私有客户端编码生成客户端特异性解码器和类别条件先验参数。这一设计实现了客户端级个性化，同时通过超网络隐式共享高阶生成结构，从而提升在非独立同分布条件下的适应性。每个客户端在从冻结的基础模型（例如 DINOv2）提取的嵌入上训练本地条件变分自编码器，而共享超网络参数则通过差分隐私随机梯度下降 （DP-SGD）进行协同优化，从而确保针对梯度反演和成员推断攻击的正式 (ε, δ)-隐私保证。此外，基于最大均值差异（*MMD*）的对齐正则项可确保跨站点分布一致性，而元代码合成模块则学习领域无关的潜在代码，以生成具有全局代表性的嵌入。

贡献。 我们的主要贡献有三点：

- 我们提出了**FedHypeVAE**，这是首个将基于超网络的参数生成与条件变分自编码器相结合、实现隐私保护嵌入合成的联邦框架。
- 我们提出了一种严谨的双层联邦优化策略，在经认证的（ε， δ）-差分隐私保证下，通过梯度裁剪和校准后的高斯噪声，联合学习个性化客户端解码器与全局一致的超网络。
- 我们提出了一种基于 *MMD* 的对齐 与元代码生成机制，以确保跨领域一致性，并在尽可能减小隐私–效用权衡的同时，实现高保真度的合成嵌入生成。

在多个机构开展的、涵盖多种医学影像数据集的大规模实验表明，**FedHypeVAE** 在鲁棒性、泛化能力和隐私合规性方面均显著优于现有联邦生成基线。通过结合基础模型嵌入、超网络驱动的个性化和差分隐私，FedHypeVAE 为联邦医学人工智能中的安全高效数据共享开辟了新范式。

相关工作

联邦学习中的梯度反演与隐私

联邦学习（FL）通过客户端之间的去中心化梯度交换来训练模型，从而减少对集中式数据聚合的需求。然而，大量关于梯度反演 和重建攻击的研究表明，共享更新（梯度或参数增量）可能会泄露敏感信息，包括近似的输入重建、成员推断和属性泄露 [15, 16, 17]。在使用高容量视觉编码器以及异构、小规模医学数据集的场景中，这些风险会进一步加剧，因为本地梯度与个别训练样本的关联更加紧密。这种脆弱性促使人们采用防御措施：（i）通过传输经过压缩或信息量较低的表示来尽可能缩小暴露面 ；或（ii）改变通信原语 ，使中央服务器只能看到聚合或经过掩码的信息，而非原始更新。

联邦学习中的隐私保护技术

保护隐私的联邦学习方法主要分为三类。**(1) 安全多方计算（SMC）与安全聚合** 通过让服务器只能观察到聚合结果来隐藏单个更新，从而防止直接重建任何客户端的梯度 [18, 19, 20, 21]。**(2) 同态加密（HE）**允许直接对加密参数执行数学运算，但通常会带来难以承受的计算和通信开销 [22, 23, 24]。**(3) 差分隐私（DP）**通过裁剪更新并添加经过校准的噪声来提供形式化的隐私保证 [25, 26, 27, 28, 29]。此外，人们还提出了梯度剪枝、掩码或随机噪声注入等经验性防御措施 [16, 30, 31, 32]，以及*Soteria*、*PRECODE*和*FedKL* [33, 34, 35]等专用系统，以缓解信息泄露。然而，这些技术往往难以解决持续存在的隐私与效用权衡问题：更强的保护会降低模型准确率和跨域泛化能力。这些局限促使人们寻求更具结构性的解决方案，例如基于超网络的形式；这类方法能够在保持高表达能力的同时，从根本上将共享参数与原始数据解耦 [36]。

联邦差分隐私生成模型

近期研究探索了通过联邦生成建模实现隐私保护数据共享。DiSalvo 等人 [12]证明，与直接使用原始医学图像相比，在嵌入层面生成合成训练数据，既能保护数据隐私，也能保持较高的下游任务性能。基于这一原则，基于嵌入的差分隐私条件变分自编码器联邦数据共享 框架 [37] 提出了一种联邦条件变分自编码器（CVAE），能够在多个客户端之间协作学习合成嵌入。在该方法中，每个客户端都训练一个采用对称架构的条件变分自编码器：类别条件编码器和解码器均由三层线性层构成，并通过重构损失（均方误差）和 Kullback–Leibler 散度项进行优化，以将潜在分布正则化为标准高斯先验。为确保隐私，该方法在使用联邦平均（FedAvg）过程进行解码器聚合时加入差分隐私（DP）噪声。这种设计实现了隐私保护的全局生成建模，但依赖于共享全局解码器，在非独立同分布数据分布下可能表现不佳，且缺乏适应多样化临床领域的能力。

联邦学习超网络

超网络近年来已成为联邦学习中一种有效的参数生成机制，为个性化和模型共享提供了元学习视角。在这一范式中，由服务器维护的中央元生成器H_ϕ 会将紧凑的客户端表示 e_k 映射为客户端模型的完整参数集，$\theta_k = H_\phi(e_k)$ [36, 38, 39, 40, 41, 42]。这种间接参数化将全局与局部学习动态解耦：服务器在参数空间中学习全局映射，而每个客户端则由一个能够捕捉其数据分布的低维嵌入来表示。因此，基于超网络的联邦学习大幅降低了通信与存储开销，实现了嵌入空间中客户端之间的平滑插值，并提供了一种优雅的机制来处理数据异质性。

重要的是，这种间接机制也增强了隐私和鲁棒性。由于超网络 H_ϕ 学习的是高阶映射，而非直接交换模型梯度，若要还原原始客户端数据，就必须同时反演超网络和潜在客户端嵌入，这比传统的梯度反演难得多。除了隐私之外，这种架构还具有更强的表达能力和适应性，因为超网络能够学习生成特定于任务或领域的参数，以捕捉客户端层面的归纳偏置，而无需显式共享参数。基于这些见解，我们提出的**FedHypeVAE** 将超网络的作用从判别式个性化扩展到生成式参数化，其中 H_ϕ 为条件变分自编码器生成适配客户端的解码器和先验参数，从而能够在异质医疗机构之间实现保护隐私且适应领域的数据合成。

问题设定与动机

我们考虑一个由 m 个客户端（例如医疗机构）组成的联邦系统，客户端索引为 $i \in \{1, \ldots, m\}$。每个客户端都私有地持有一个本地嵌入标签数据集

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

其中，$x_j^{(i)} \in \mathbb{R}^{d_x}$ 表示紧凑的特征嵌入（通常由冻结的基础编码器提取，例如 DINOv2 [13]），而 $y_j^{(i)} \in \mathcal{Y}$ 是对应的类别标签。这些嵌入构成了原始医疗数据在语义丰富且保护隐私的中间表示。

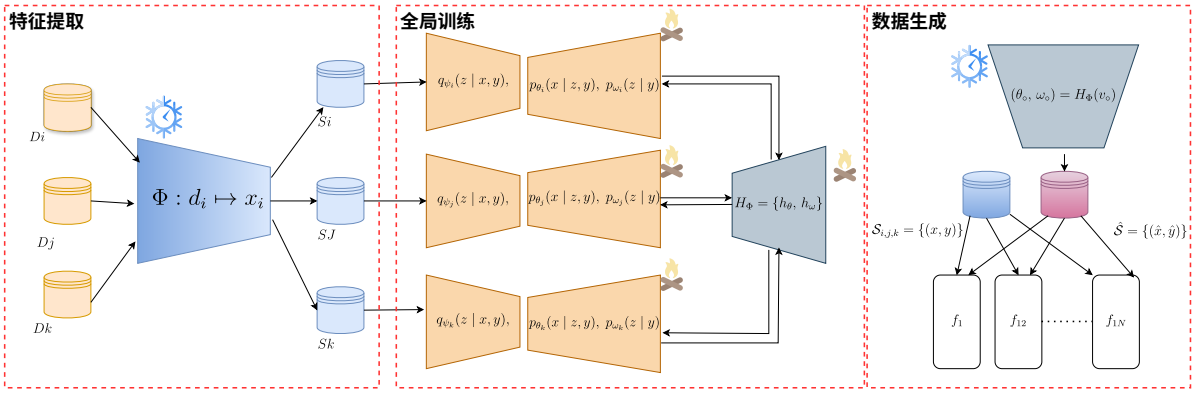

图 1：所提出的 **FedHypeVAE** 框架概览。(1) 每个参与的客户端 $\mathcal{H}_i$ 使用冻结的基础编码器 Φ将本地图像数据集 $\mathcal{D}_i$ 转换为嵌入级数据集 $\mathcal{S}_i$ ，从而大幅降低通信和存储成本。(2) 在本地，每个客户端训练一个条件变分自编码器（CVAE），其参数由编码器–解码器对（q_{ψ_i}, p_{θ_i}）和类别条件先验 p_{ω_i}构成，用于对嵌入分布建模，同时不暴露原始数据。(3) 服务器端超网络 $H_\Phi = \{h_\theta, h_\omega\}$ 将私有客户端编码v_i 映射为客户端特定的解码器和先验参数，并通过差分隐私随机梯度下降（DP-SGD）进行联邦优化。(4) 收敛后，中性元代码 $v_\circ$ 生成全局解码器–先验对（$\theta_\circ, \omega_\circ$）$\hat{}$，用于生成合成嵌入 $\mathcal{S} = \{(\hat{}x, \hat{}y)\}$，这些嵌入可与本地数据结合，用于下游模型$f_1, \ldots, f_N$。

目标是协作训练一个联邦生成模型，使其能够合成对全局有用且统计特性一致的嵌入，以应对非独立同分布数据异质性的存在。形式上，我们旨在通过条件生成过程近似全局数据分布 $p(x, y)$

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

其中，(θ, ω) 分别表示解码器和先验参数。在联邦设置中，隐私法规限制了模型参数或数据样本的直接共享；因此，每个客户端都在本地训练其生成模型，并仅向中央服务器传输经过隐私保护的信息。

我们提出的 **FedHypeVAE**整合了三个关键组件来应对这一挑战：(i) 一个条件变分自编码器（*CVAE*），用于学习每个站点的本地嵌入分布；(ii) 一个共享的超网络 H_Φ，将轻量级的私有客户端代码 v_i 映射为客户端特定的生成参数（θ_i, ω_i）；以及 (iii) 一种联邦优化机制，通过差分隐私随机梯度下降（DP-SGD）汇总各站点的知识。该方法在生成层实现了隐私保护个性化，同时确保在数据异质性下的全局一致性和鲁棒性。

方法论

客户端级条件生成目标

每个客户端 i 使用由编码器 $q_{\psi_i}(z|x, y)$、解码器 $p_{\theta_i}(x|z, y)$ 和类别条件先验 $p_{\omega_i}(z|y)$ 参数化的条件变分自编码器（CVAE），对其本地嵌入分布 $p_i(x|y)$ 进行建模。学习目标是最大化证据下界（ELBO）：

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

第一项确保对本地嵌入进行准确重构，而库尔贝克–莱布勒项则对潜在空间进行正则化，促进客户端之间的平滑性和全局一致性。这构成了继承自基于嵌入的联邦条件变分自编码器框架 [37, 12] 的基础目标。

基于超网络的参数生成

为了在生成层引入个性化和隐私，我们用一个生成客户端特定参数的共享超网络取代相互独立的客户端解码器：

$$\theta_i = h_\theta(v_i; \Phi_\theta), \qquad \omega_i = h_\omega(v_i; \Phi_\omega), \tag{2}$$

其中， $v_i \in \mathbb{R}^{d_v}$ 是私有且可训练的客户端代码， $\Phi = \{\Phi_\theta, \Phi_\omega\}$ 是共享的服务器端超网络参数。该形式使每个客户端的生成模型能够适应其域分布，同时将原始数据与全局共享参数解耦，从而增强隐私性和非独立同分布鲁棒性。

行缩放高效生成。 为减少参数规模，每个带有基础权重 $W_\ell \in \mathbb{R}^{r_\ell \times c_\ell}$ 的解码器层都通过轻量级的逐行缩放和偏置平移进行调制：

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

其中， d_ℓ（v_i）和 Δb_ℓ（v_i）由 h_θ预测。该策略遵循 HyperLSTM 原则 [36]，在保持表达能力的同时，最大限度地减少计算和通信开销。

超网络生成的类别先验。 类似地，类别条件高斯先验生成为

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

其中， $e(y)$是可学习的标签嵌入。该参数化使模型能够捕捉特定域特征风格，并更好地校准各站点之间的潜在先验。

稳定性正则化与跨站点对齐

每个客户端最小化　结合负证据下界与结构约束的稳定性正则化目标：

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

其中， $\mathcal{R}_{\text{Lip}}$ 强制执行谱范数或雅可比控制，以实现利普希茨稳定性； λ_v 则约束客户端代码幅度 s.

跨站点分布对齐。为对齐真实嵌入和合成嵌入，每个客户端都会计算本地最大均值差异（MMD）损失：

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

其中， $k(\cdot,\cdot)$ 是高斯多核函数。该项有助于促进不同域之间潜在分布的一致性，同时无需交换任何原始数据。

差分隐私下的联邦超网络优化

共享超网络参数 Φ 通过差分隐私随机梯度下降（DP-SGD）在各客户端之间协同优化。全局联邦目标汇总各客户端的损失和对齐正则项：

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

每个客户端优化其本地编码器和代码参数：

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

表 1：**在腹部 CT 和 ISIC 2025 数据集上，对比联邦基线方法与我们提出的 FedHypeVAE。数值表示各客户端在三个随机种子下的准确率（ACC）和平衡准确率（BACC）的均值标准差。每种数据集配置下的最佳性能以粗体显示。**

方法	CT（独立同分布）		CT（α = 0.3）		ISIC 2025（独立同分布）		ISIC 2025（α = 0.3）	
	准确率 (%)	平衡准确率 (%)	准确率 (%)	平衡准确率 (%)	准确率 (%)	平衡准确率(%)	准确率 (%)	平衡准确率 (%)
FedAvg	73.27±1.18	67.04±1.21	64.91±5.83	58.68±2.96	61.20±2.8	54.10±2.5	61.00±2.8	54.00±2.5
FedProx	73.30±1.16	66.88±1.20	64.81±6.18	58.61±5.80	61.75±3.0	54.60±2.8	60.90±3.0	53.90±2.8
FedLambda	77.27±0.83	71.26±0.87	81.10±3.76	59.02±2.59	63.30±2.7	55.20±2.8	76.25±3.2	54.54±2.6
差分隐私条件生成对抗网络	77.54±1.42	71.99±1.14	88.91±2.04	57.44±2.46	64.80±2.9	55.80±3.0	83.12±2.9	53.38±2.9
DP-CVAE（论文）	77.60±0.72	71.77±0.85	88.88±1.41	57.63±3.29	66.20±2.8	56.30±2.6	83.10±2.8	53.46±2.7
FedHypeVAE（本文方法）	**81.32**±1.05	**76.08**±1.12	**90.09**±1.07	**62.14**±1.02	**67.70**±2.6	**56.90**±2.7	**84.00**±2.6	**57.74**±2.8

差分隐私梯度构造。 对于每个小批次 B_i，客户端 i 计算逐样本梯度，将其裁剪至界限 C，并添加高斯噪声：

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

只有这些经过噪声扰动的梯度 $\tilde{g}_i$ 会被发送到服务器，从而确保 (ε, δ)-差分隐私，同时将 ψ_i、v_i 和原始数据保留在本地。

服务器端聚合。 服务器以 FedAvg 风格的更新方式聚合隐私化梯度：

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

在正式的差分隐私保证下，这就完成了一个通信轮次。

全局元代码合成与生成

收敛后，服务器利用差分隐私噪声化全局统计量 $\{\hat{\mu}_y, \hat{\Sigma}_y\}$学习一个中性元代码 $v_\circ$：

$$\begin{aligned} v_\circ = \arg\min_v \sum_{y\in\mathcal{Y}} & \|\mathbb{E}_{z\sim p_{\omega_\circ}(z|y)}[x(z,y)] - \hat{\mu}_y\|_2^2 \\ & + \beta \|\mathrm{Cov}_z[x(z,y)] - \hat{\Sigma}_y\|_F^2. \end{aligned} \tag{11}$$

合成嵌入的生成方式如下

$$\hat{x} \sim p_{\theta_\circ}(x|z,y), \quad \theta_\circ = h_\theta(v_\circ;\Phi), \quad \omega_\circ = h_\omega(v_\circ;\Phi), \tag{12}$$

其中 $z\sim\mathcal{N}(0,\ I)$。该元代码支持在隐私约束下进行可控的领域无关合成。

元代码混合。 为了实现更丰富的全局合成，K可以使用带有混合权重 $\pi_k \geq 0, \sum_k \pi_k = 1$的元代码 $\{v_k\}_{k=1}^K$：

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

我们对（h_θ、h_ω）施加谱范数约束，以确保利普希茨稳定性，并对 $\|v_i\|_2 \leq r$ 进行界定；同时通过对（q、σ、T）进行矩会计来追踪隐私损失。跨站点最大均值差异对齐可缓解非独立同分布漂移，而混合元代码则能提升全局覆盖度和多样性。

算法 1：**FedHypeVAE**

要求： 客户端数量 m；隐私预算（ε，δ）；学习率 η_ψ，η_v，η_Φ；裁剪界限 C；噪声尺度 σ；正则化权重 λ_{MMD}，λ_{Lip}，λ_v

为每个客户端 i 初始化共享超网络参数 $\Phi = \{\Phi_\theta, \Phi_\omega\}$、本地编码器 ψ_i 和私有编码 $v_i \sim \mathcal{N}(0,\ I)$。

对于每个通信轮次 $t = 1$ 至 T **执行**

  客户端侧（对每个 i 并行执行）：

  采样小批次 $B_i \subseteq \mathcal{S}_i$。

  计算本地证据下界损失 $\mathcal{L}_i^{\text{ELBO}}$（公式 1）。

  更新本地编码器 ψ_i 和客户端代码 v_i（公式 9）。

  计算对齐损失 MMD_i^2（公式 6）。

  计算隐私化梯度：

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

  将 $\tilde{g}_i$ 传输至服务器。

  服务器端：

  聚合并更新全局超网络：

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

结束循环 **后训练：** 学习元代码 $v_。$（公式 12）；生成合成嵌入 $\hat{x} \sim p_{\theta_。}(x|z、\ y)$，其中 $\theta_。 = h_\theta(v_。;\ \Phi)$ 和 $\omega_。 = h_\omega(v_。;\ \Phi)$；可选择混合 K 元代码（公式 13）。

确保： 训练后的全局超网络 Φ和合成数据集 $\hat{\mathcal{S}}$。

实验结果

实验设置

数据集与指标。 我们在两个互补的多中心医学影像基准上评估 FedHypeVAE。(1) ISIC 2025 MILK10k 数据集 [43] 包含 10,000 张皮肤镜图像，这些图像标注了多个诊断类别，用于模拟多机构皮肤病变联邦。(2) 腹部 CT（矢状位）数据集 [44]包含 25,211 个 CT 切片，涵盖 11 个解剖类别，广泛应用于跨器官定位任务。遵循近期的联邦学习研究 [45, 46], ，每个数据集均采用狄利克雷划分，并在独立同分布和异质设置下分配给 $m = 10$个客户端，参数为 $\alpha = 0.3$。原始医学图像使用冻结的 DINOv2 编码器转换为紧凑的特征嵌入$\mathcal{S}_i = \{(x,\ y)\}$，以确保表示一致性，同时保护隐私。评估指标包括每个客户端的准确率 和平衡准确率（*BACC*），并对三个随机种子的结果取平均，以评估领域偏移下的鲁棒性。

实现细节。 所有下游分类器均以 DINOv2 嵌入为基础，采用单层线性模型实现 [13]。FedHypeVAE 和所有基线模型均使用随机梯度下降训练 50 个通信轮次，每轮进行 5 个本地训练轮次 $\eta = 10^{-3}$。通过 OPACUS 库 [47] ，使用差分隐私随机梯度下降实施差分隐私 [47] ，参数为 $(\varepsilon,\ \delta) = (1。0,\ 10^{-4})$，裁剪范数为 1。5，从而提供形式化的隐私保证 [48, 49]。对比基线包括 FedAvg 和**FedProx** [50], ，以及 DP-CVAE 变体 [12]。所有模型均在相同的联邦设置下进行训练和评估。

结果与讨论

表 1 报告了两个数据集在独立同分布和非独立同分布条件下的结果。**FedHypeVAE** 在生成保真度、准确率和平衡准确率方面始终优于基线联邦分类器。其基于超网络的解码器和先验生成实现了客户端自适应建模，而 MMD 对齐项则缓解了跨中心分布漂移。即使在严格的隐私预算下（$\varepsilon \leq 3.0, \delta = 10^{-5}$），该模型仍能保持较高的重建保真度和泛化能力，在放射学和皮肤病学领域均优于 DP-CVAE。与基于参数正则化的个性化方法不同， [51], FedHypeVAE 直接在生成层中实现个性化，从而在多种模态间生成语义一致且具有隐私保护的嵌入。

结果与讨论

FedHypeVAE 在多个医学影像数据集上进行了评估，数据集采用独立同分布和非独立同分布划分；结果表明，与联邦条件变分自编码器基线相比，其生成保真度、鲁棒性和隐私保护均有稳定提升 [37, 12, 52]。在可比的隐私预算下（$\varepsilon \leq 3.0, \delta = 10^{-5}$），该模型在保持严格差分隐私保证的同时，取得了更高的准确率和平衡准确率。这些改进得益于超网络能够生成客户端自适应的解码器和先验参数，在捕捉局部差异的同时不损害全局一致性。引入基于*MMD*的跨站点对齐 可稳定异质领域间的潜在表示，缓解联邦环境中常见的嵌入漂移。此外，梯度级差分隐私随机梯度下降 相比权重级噪声注入，实现了更优的隐私–效用权衡，并在严格隐私约束下保持重建质量。总体而言，**FedHypeVAE** 推动了差分隐私生成学习的发展，能够在去中心化医学数据集间实现领域一致、语义忠实且符合隐私要求的嵌入合成。

结论

我们提出了**FedHypeVAE**，这是一种由超网络驱动的双层联邦生成框架，将基于嵌入的差分隐私条件变分自编码器范式扩展至自适应的隐私保护数据合成。FedHypeVAE通过引入共享超网络，从轻量级私有编码生成客户端特定的解码器和先验参数，实现了细粒度个性化，同时不损害数据机密性。跨站点最大均值差异对齐和元编码合成确保了在严苛的非独立同分布条件下获得连贯的全局表示，而差分隐私随机梯度下降则在整个训练过程中提供形式化的（ε，δ）-差分隐私。总体而言，这些进展建立了一种统一方法，连接了生成建模、个性化与差分隐私，为医疗机构之间安全、可泛化且数据高效的协作奠定了基础。

参考文献

[1] Geert Litjens、Thijs Kooi、Babak Ehteshami Bejnordi、Arnaud Arindra Adiyoso Setio、Francesco Ciompi、Mohsen Ghafoorian、Jeroen Awm Van Der Laak、Bram Van Ginneken 和 Clara I Sánchez。《深度学习在医学图像分析中的综述》。*Medical imageanalysis*，42：60–88，2017。[2] Smadar Shilo、Hagai Rossman 和 Eran Segal。《革命的轴线：大数据在医疗保健领域的挑战与前景》。《自然·医学》，26(1)：29–38，2020。[3] Karin Stacke、Gabriel Eilertsen、Jonas Unger 和 Claes Lundström。《衡量深度学习在组织病理学中的领域偏移》。*IEEE* 生物医学与健康信息学期刊，25(2)：325–336，2020。[4] Brendan McMahan、Eider Moore、Daniel Ramage、Seth Hampson 和 Blaise Aguera y Arcas。《从去中心化数据中高效通信地学习深度网络》。载于人工智能与统计，第1273–1282页。PMLR，2017。[5] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Sylvain Gelly 等。《一张图像胜过16×16个词：大规模图像识别的变换器》。*arXiv* 预印本 *arXiv:2010.11929*，2020。[6] Nannan Wu、Li Yu、Xin Yang、Kwang-Ting Cheng 和 Zengqiang Yan。《Fediic：面向类别不平衡医学图像分类的鲁棒联邦学习》。载于医学图像计算与计算机辅助干预国际会议，第692–702页。Springer，2023。[7] Yuexuan Xia、Benteng Ma、Qi Dou 和 Yong Xia。《通过基于协作图的强化学习提升联邦学习的性能公平性》。载于医学图像计算与计算机辅助干预国际会议，第263–272页。Springer，2024。[8] Lennart R Koetzier、Jie Wu、Domenico Mastrodicasa、Aline Lutz、Matthew Chung、W Adam Koszek、Jayanth Pratap、Akshay S Chaudhari、Pranav Rajpurkar、Matthew P Lungren 等。《为医学影像生成合成数据》。《放射学》，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 等。《生成模型提升分布偏移下医学分类器的公平性》。《自然·医学》，30(4)：1166–1173，2024。[10] Ian J Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。《生成对抗网络》。神经信息处理系统进展，27，2014。[11] Jonathan Ho、Ajay Jain 和 Pieter Abbeel。《去噪扩散概率模型》。神经信息处理系统进展，33：6840–6851，2020。

[12] Francesco Di Salvo、David Tafler、Sebastian Doerrich 和 Christian Ledig。《通过使用条件变分自编码器捕获特征分布来构建隐私保护数据集》。arXiv 预印本 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 等。《DINOv2：无需监督地学习鲁棒视觉特征》。arXiv 预印本 arXiv:2304.07193，2023。[14] Sayak Paul 和 Pin-Yu Chen。《视觉变换器是鲁棒的学习器》。载于AAAI人工智能会议论文集，第36卷，第2071–2081页，2022。[15] Matt Fredrikson、Somesh Jha 和 Thomas Ristenpart。《利用置信度信息的模型反演攻击及基本对策》。载于第22届 ACM SIGSAC 计算机与通信安全会议论文集，第1322–1333页，2015。[16] Ligeng Zhu、Zhijian Liu 和 Song Han。《梯度深度泄漏》。神经信息处理系统进展，32，2019。[17] Jonas Geiping、Hartmut Bauermeister、Hannah Dröge 和 Michael Moeller。《反演梯度——攻破联邦学习中的隐私有多容易？》。神经信息处理系统进展，33：16937–16947，2020。[18] Andrew C Yao。《安全计算协议》。载于第23届年度计算机科学基础研讨会（SFCS 1982），第160–164页。IEEE，1982。[19] Keith Bonawitz、Vladimir Ivanov、Ben Kreuter、Antonio Marcedone、H Brendan McMahan、Sarvar Patel、Daniel Ramage、Aaron Segal 和 Karn Seth。《面向隐私保护机器学习的实用安全聚合》。载于2017年 ACM SIGSAC 计算机与通信安全会议论文集，第1175–1191页，2017。[20] Vaikkunth Mugunthan、Antigoni Polychroniadou、David Byrd 和 Tucker Hybinette Balch。《SMPAI：用于联邦学习的安全多方计算》。载于 NeurIPS 2019 鲁棒人工智能在金融服务中的研讨会论文集，第21卷。MIT Press，剑桥，马萨诸塞州，美国，2019。[21] Wenhao Mou、Chunlei Fu、Yan Lei 和 Chunqiang Hu。《一种基于安全多方计算的可验证联邦学习方案》。载于无线算法、系统与应用国际会议，第198–209页。Springer，2021。[22] Craig Gentry。《一种全同态加密方案》。斯坦福大学，2009。[23] Jaehyoung Park 和 Hyuk Lim。《使用同态加密的隐私保护联邦学习》。Applied Sciences，12(2)：734，2022。[24] Jing Ma、Si-Ahmed Naas、Stephan Sigg 和 Xixiang Lyu。《基于多密钥同态加密的隐私保护联邦学习》。International Journal of Intelligent Systems，37(9)：5880–5901，2022。[25] Robin C Geyer、Tassilo Klein 和 Moin Nabi。《差分隐私联邦学习：客户端级别视角》。arXiv 预印本 arXiv:1712.07557，2017。[26] H Brendan McMahan、Daniel Ramage、Kunal Talwar 和 Li Zhang。《学习差分隐私循环语言模型》。arXiv 预印本 arXiv:1710.06963，2017。[27] Tao Yu、Eugene Bagdasaryan 和 Vitaly Shmatikov。《通过本地适应挽救联邦学习》。arXiv 预印本 arXiv:2002.04758，2020。[28] Alberto Bietti、Chen-Yu Wei、Miroslav Dudik、John Langford 和 Steven Wu。《个性化改善联邦学习中的隐私—准确率权衡》。载于国际机器学习会议，第1945–1962页。PMLR，2022。[29] Zebang Shen、Jiayuan Ye、Anmin Kang、Hamed Hassani 和 Reza Shokri。《只共享你的表示：保证改善联邦学习中的隐私—效用权衡》。arXiv 预印本 arXiv:2309.05505，2023。[30] Yangsibo Huang、Samyak Gupta、Zhao Song、Kai Li 和 Sanjeev Arora。《评估联邦学习中的梯度反演攻击与防御措施》。神经信息处理系统进展，34：7232–7241，2021。[31] Zhuohang Li、Jiaxin Zhang、Luyang Liu 和 Jian Liu。《通过生成式梯度泄漏审计联邦学习中的隐私防御》。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第10132–10142页，2022。[32] Wenqi Wei、Ling Liu、Margaret Loper、Ka-Ho Chow、Mehmet Emre Gursoy、Stacey Truex 和 Yanzhao Wu。《用于评估联邦学习中梯度泄漏攻击的框架》。arXiv 预印本 arXiv:2004.10397，2020。[33] Jingwei Sun、Ang Li、Binghui Wang、Huanrui Yang、Hai Li 和 Yiran Chen。《从表示视角出发对联邦学习中的隐私泄漏进行可证明防御》。arXiv 预印本 arXiv:2012.06043，2020。

[34] Daniel Scheliga、Patrick Mäder 和 Marco Seeland。PRECODE——一种用于防止深度梯度泄露的通用模型扩展。在*IEEE/CVF*冬季计算机视觉应用会议，第 1849–1858 页，2022 年。[35] Hanchi Ren、Jingjing Deng、Xianghua Xie、Xiaoke Ma 和 Jianfeng Ma。用于联邦学习的密钥锁模块梯度泄露防御。*arXiv* 预印本 *arXiv:2305.04095*，2023 年。[36] David Ha、Andrew Dai 和 Quoc V Le。超网络。*arXiv* 预印本 *arXiv:1609.09106*，2016 年。[37] Francesco Di Salvo、Hanh Huyen My Nguyen 和 Christian Ledig。基于嵌入的差分隐私条件变分自编码器联邦数据共享。在医学图像计算与计算机辅助干预国际会议，第 138–147 页。Springer，2025 年。[38] Aviv Shamsian、Aviv Navon、Ethan Fetaya 和 Gal Chechik。使用超网络的个性化联邦学习。在国际机器学习会议，第 9489–9502 页。PMLR，2021 年。[39] Alycia N Carey、Wei Du 和 Xintao Wu。人口统计公平性异质性下的鲁棒个性化联邦学习。在*2022 IEEE*大数据国际会议（*Big Data*），第 1425–1434 页。IEEE，2022 年。[40] Hongxia Li、Zhongyi Cai、Jingya Wang、Jiangnan Tang、Weiping Ding、Chin-Teng Lin 和 Ye Shi。FedTP：通过 Transformer 个性化实现联邦学习。*IEEE*神经网络与学习系统汇刊，35(10)：13426–13440，2023 年。[41] Arvin Tashakori、Wenwen Zhang、Z Jane Wang 和 Peyman Servati。SemiPFL：面向边缘智能的个性化半监督联邦学习框架。*IEEE*物联网期刊，10(10)：9161–9176，2023 年。[42] Yanfei Lin、Haiyi Wang、Weichen Li 和 Jun Shen。基于超网络的联邦学习——全视野切片图像分析案例研究。《科学报告》，13(1)：1724，2023 年。[43] Tschandl Philipp、Akay Nisa Bengü、Rosendahl Cliff、Rotemberg Veronica、Todorovska Verche、Weber Jochen、Wolber Anna Katharina、Müller Christoph、Kurtansky Nicholas、Halpern Allan 等。Milk10k：用于诊断色素性和非色素性皮肤癌及其模拟病变的分层多模态影像学习工具包。《皮肤病学研究杂志》，2025 年。[44] Xuanang Xu、Fugen Zhou、Bo Liu、Dongshan Fu 和 Xiangzhi Bai。利用三维区域提议网络在 CT 图像中高效定位多个器官。*IEEE*医学影像汇刊，38(8)：1885–1898，2019 年。[45] Xuyang Li、Weizhuo Zhang、Yue Yu、Wei-Shi Zheng、Tong Zhang 和 Ruixuan Wang。SIFT：一种具有文本指导的联邦学习串行框架。在医学图像计算与计算机辅助干预国际会议，第 655–665 页。Springer，2024 年。[46] Minghui Chen、Meirui Jiang、Qi Dou、Zehua Wang 和 Xiaoxiao Li。FedSoup：通过选择性模型插值提升联邦学习的泛化能力和个性化水平。在医学图像计算与计算机辅助干预国际会议，第 318–328 页。Springer，2023 年。[47] Ashkan Yousefpour、Igor Shilov、Alexandre Sablayrolles、Davide Testuggine、Karthik Prasad、Mani Malek、John Nguyen、Sayan Ghosh、Akash Bharadwaj、Jessica Zhao 等。Opacus：PyTorch 中易于使用的差分隐私库。*arXiv* 预印本 *arXiv:2109.12298*，2021 年。[48] Milad Nasr、Shuang Songi、Abhradeep Thakurta、Nicolas Papernot 和 Nicholas Carlin。对手实例化：差分隐私机器学习的下界。在*2021 IEEE*安全与隐私研讨会（*SP*），第 866–882 页。IEEE，2021 年。[49] Lucas Lange、Maja Schneider、Peter Christen 和 Erhard Rahm。实践中的隐私：X射线图像中的隐私保护 COVID-19检测（扩展版）。*arXiv* 预印本 *arXiv:2211.11434*，2022 年。[50] Tian Li、Anit Kumar Sahu、Manzil Zaheer、Maziar Sanjabi、Ameet Talwalkar 和 Virginia Smith。异构网络中的联邦优化。《机器学习与系统论文集》，2：429–450，2020 年。[51] Othmane Marfoq、Giovanni Neglia、Richard Vidal 和 Laetitia Kameni。通过本地记忆实现个性化联邦学习。在国际机器学习会议，第 15070–15092 页。PMLR，2022 年。[52] Bjarne Pfitzner 和 Bert Arnrich。DPD-FVAE：使用带差分隐私解码器的联邦变分自编码器进行合成数据生成。*arXiv* 预印本 *arXiv:2211.11591*，2022 年。