

# 面向医学数据分析的基于注入与蒸馏方法的异构模型 个性化联邦学习的复现研究报告

## 摘要

目前,联邦学习因其能在不共享各个本地客户端的数据下,协作训练一个全局模型,而由于数据异质性以及模型异质性的问题,给训练模型的性能带来了挑战,而为了解决这种挑战,现有的方式通过对公共数据集的软预测实现各个客户端之间的信息交流,但是这些方法需要依赖一个公共的数据集,为了解决这一挑战,该论文:面向医学数据分析的基于注入与蒸馏方法的异构模型个性化联邦学习 [13] 提出了一种新颖的联邦学习范式 MH-pFLID。本次研究通过复现该论文的乳腺癌分类图像实验,探讨后续可能的研究方向。

**关键词:** 医学数据; 联邦学习

## 1 引言

联邦学习 [8] 因为其能够在无需直接访问各个客户端本地数据的情况下,协作训练一个全局模型这一特性,被广泛用于数据敏感的领域。而在医疗领域,由于各个医疗机构数据集的隐私特性,并且数据集构建的困难,使得各个机构迫切希望能利用其他医疗机构的数据信息训练本地模型,而对于医疗领域,数据的分布差异,以及各个医疗机构采用的模型架构的差异会导致模型拟合能力以及模型性能的困难。这种各个客户端的数据的异质性以及模型架构的异质性成为一个重要的挑战。

## 2 相关工作

### 2.1 应用个性化联邦学习解决数据异质性

经典的联邦学习算法是利用所有客户端训练一个单一的全局模型。而考虑到数据异质性会影响到单一全局模型的性能,个性化联邦学习被提出用于为每个客户端训练个性化模型以更好地适应数据异质的挑战。它包括一些方法,如模型分解 (FedPer [1];FedRep [2])、聚类 (Sattler 等 [9];Mansour 等 [7])、多任务学习 (MOCHA [10];FedEM [3]) 和对比学习 (FedCL [12])

### 2.2 应用个性化联邦学习解决模型异质性

目前,许多研究方法使用知识蒸馏 [4] 来解决模型异质性问题。它将各个客户端的数据知识蒸馏,并将蒸馏后的知识在服务器端中聚合为本地知识,例如 FedMD [6];DS-pFL [5]。而进一步提高异构模型的性能,KT-pFL [14] 在服务器端训练个性化的软预测权重。

### 3 本文方法

上述的相关工作中解决数据异质性以及模型异质性采用的是通过一个公共数据集进行,而考虑到一些医疗机构可能不具备足够的硬件资源对公共数据集进行预测,并且公共数据集传输时会造成显著的通信延迟,为了解决这个挑战,本文利用轻量化的信使模型传递各个客户端的知识。

### 3.1 本文方法概述

该论文 [13] 提出一个名为 MH-pFLID 的框架，采用知识注入和知识蒸馏的方法传递知识。首先框架会通过知识注入阶段将信使模型的知识传递给各个客户端以指导客户端的训练；之后通过知识蒸馏阶段，各个客户端的本地模型将知识蒸馏到信使模型中；随后，将信使模型的参数上传到服务器端进行聚合；将聚合后的参数分发给各个信使模型完成一轮的通信。图 1 展示了方法框架。

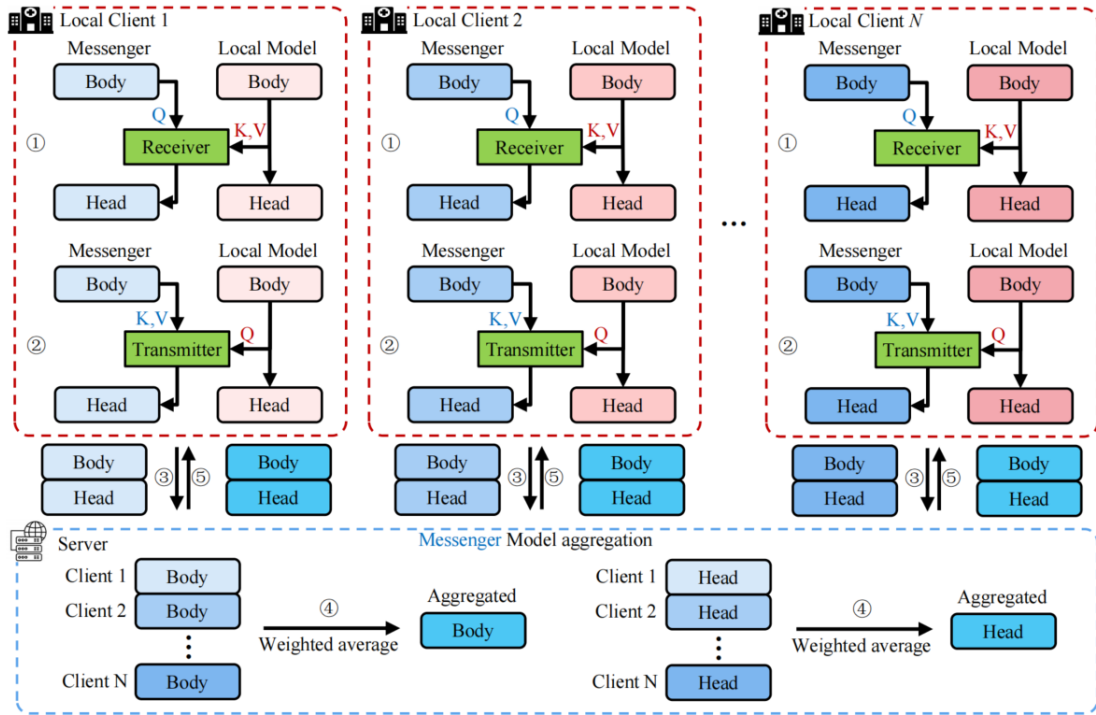


图 1. MH-pFLID 框架图

### 3.2 知识注入阶段

在知识注入阶段，旨在将来自信使模型中的知识注入到本地模型中，因此在这个阶段将冻结信使模型，并使用该模型指导本地模型的训练。其中在这个阶段的损失函数设计为  $L_{inj,i}$ :

$$L_{\text{inj},i} = \lambda_{\text{inj}}^l \sum_{j=1}^M L_{\text{inj}}^l(\hat{y}_{ij}^l, y_{ij}) + \lambda_{\text{inj}}^m \sum_{j=1}^M L_{\text{inj}}^m(\hat{y}_{ij}^m, y_{ij}). \quad (1)$$

M 表示本地数据的总数。 $\hat{y}_{ij}^l$  和  $\hat{y}_{ij}^m$  分别是本地模型和信使模型对本地客户端第 j 个数据的预测。 $L_{\text{ini}}^l$  和  $L_{\text{ini}}^m$  分别表示本地模型和信使模型的损失函数。 $\lambda_{\text{ini}}^l$  和  $\lambda_{\text{ini}}^m$  是它们对应的权重。 $y_{ij}$

是  $x_{ij}$  的标签。在这里，本地模型的输出  $\hat{y}_{ij}^l$  可以定义为：

$$\hat{y}_{ij}^l = L_h(L_b(x_{ij})) \quad (2)$$

其中  $L_b(\cdot)$  和  $L_h(\cdot)$  分别是本地模型的主体和头。信使模型的输出  $\hat{y}_{ij}^m$  可以表示为：

$$\hat{y}_{ij}^m = M_h(R(M_b(x_{ij}), L_b(x_{ij}))) \quad (3)$$

其中  $M_b(\cdot)$  是图 1 中的信使模型主体网络， $L_b(\cdot)$  是本地模型主体， $R(\cdot)$  是该论文 [13] 设计的接收模块，用于接收来自信使模型的信息，而  $M_h(\cdot)$  是信使模型头。在知识注入阶段，由于信使模型是固定的，损失函数只能通过方程 1 的第一项对  $L_b(\cdot)$  和  $L_h(\cdot)$  生成梯度，以及通过第二项对  $L_b(\cdot)$  和  $R(\cdot)$  生成梯度。

### 3.3 知识蒸馏阶段

知识蒸馏阶段旨在将信息从本地模型蒸馏到信使模型。在这个阶段将会冻结本地模型，并对信使模型进行知识蒸馏，其中损失函数  $L_{\text{dis},i}$  表示为：

$$L_{\text{dis},i} = \lambda_{\text{dis}}^m \sum_{j=1}^M L_{\text{dis}}^m(\hat{y}_{ij}^m, y_{ij}) + \lambda_{\text{dis}}^{\text{con}} \sum_{j=1}^M L_{\text{dis}}^{\text{con}}(\hat{y}_{ij}^m, \hat{y}_{ij}^l) \quad (4)$$

$L_{\text{dis}}^m$  和  $L_{\text{dis}}^{\text{con}}$  分别表示用于训练信使模型的损失函数和知识蒸馏损失函数。对于知识蒸馏损失，使用 KL 散度来约束信使模型的头部输出和本地头模型的输出在同一分布下，以便将知识从本地模型蒸馏到消息传递模型。 $\lambda_{\text{dis}}^l$  和  $\lambda_{\text{dis}}^{\text{con}}$  是它们对应的权重。其他变量与方程 1 中的定义相同。在这里，本地模型的输出  $\hat{y}_{ij}^l$  定义与方程 2 相同。信使模型的输出  $\hat{y}_{ij}^m$  可以表示为：

$$\hat{y}_{ij}^m = M_h(T(L_b(x_{ij}), M_b(x_{ij}))) \quad (5)$$

其中  $M_b(\cdot)$ 、 $L_b(\cdot)$  和  $M_h(\cdot)$  的定义与方程 3 中相同。 $T(\cdot)$  是该论文 [13] 设计的传输模块，用于将信息从本地模型发送到信使模型。在知识蒸馏阶段的训练中，我们尝试使用真实标签  $y_{ij}$  和本地模型输出  $\hat{y}_{ij}^l$  来共同监督  $\hat{y}_{ij}^m$ ，以便将知识蒸馏到信使模型中。由于本地模型是固定的，损失函数会通过方程 4 的第一项和第二项生成关于  $M_b(\cdot)$ 、 $M_h(\cdot)$  和  $T(\cdot)$  的梯度。

### 3.4 信使模型的聚合

完成训练之后，各个客户端的信使模型将会上传到服务器，然后分别对信使模型的主体和头部进行模型参数的聚合。所使用的聚合操作是权重平均聚合 [8]。最后将聚合后的模型分发到各个客户端中进行下一轮的训练。

### 3.5 接收器与传输器模块

### 3.6 接收器模块

在知识注入阶段，该论文 [13] 设计了接收器，以更好地将本地特征与全局特征相匹配。它可以使本地模型更好地接收全局知识。接收器定义为  $I_{\text{loc},R} = R(I_{\text{loc}}, I_{\text{mes}})$  其中  $I_{\text{loc}}$  和  $I_{\text{mes}}$  分别是本地模型和信使模型主体的输出特征。 $I_{\text{loc},R}$  是接收器的输出，它是本地主体特征的加权

组合。本地客户端的特征的初始生成记为  $I_{\text{loc}}$ ，该特征通过线性层  $W_d$  进行上采样或下采样，从而生成与全局特征  $I_{\text{mes}}$  具有相同维度  $D_{\text{mes}}$  的特征  $I'_{\text{loc}}$ 。 $I_{\text{mes}}$  和  $I'_{\text{loc}}$  用于通过  $W_k$ 、 $W_q$  和  $W_v$  生成查询特征 Q、键特征 K 和值特征 V。查询特征 Q 和键特征 K 进行矩阵乘法以生成混淆矩阵  $M_R$ ，如下所示：

$$\begin{aligned} Q &= W_q(I_{\text{mes}}), \quad K = W_k(I'_{\text{loc}}), \quad V = W_v(I'_{\text{loc}}). \\ M_R &= \text{Softmax} \left( \frac{QK^\top}{\sqrt{D_{\text{mes}}}} \right) \end{aligned} \quad (6)$$

最后，混淆矩阵  $M_R$  用于与 V 进行矩阵乘法，生成本地特征  $I_{\text{loc},R}$ ：

$$I_{\text{loc},R} = M_R V \quad (7)$$

### 3.7 传输器模块

传输器定义为  $I_{\text{mes},T} = T(I_{\text{mes}}, I_{\text{loc}})$ ，其中  $I_{\text{mes},T}$  是传输器的输出，它是信使模型主体特征的加权组合，在知识蒸馏阶段，使全局特征  $I_{\text{mes}}$  学习经过处理的本地特征  $I'_{\text{loc}}$  的知识。 $I_{\text{mes}}$  和  $I'_{\text{loc}}$  用于通过  $W_k$ 、 $W_q$  和  $W_v$  生成查询特征 Q、键特征 K 和值特征 V。查询特征 Q 和键特征 K 进行矩阵乘法以生成混淆矩阵  $M_T$ 。

$$\begin{aligned} Q &= W_q(I'_{\text{loc}}), \quad K = W_k(I_{\text{mes}}), \quad V = W_v(I_{\text{mes}}). \\ M_T &= \text{Softmax} \left( \frac{QK^\top}{\sqrt{D_{\text{mes}}}} \right) \end{aligned} \quad (8)$$

最后，混淆矩阵  $M_T$  用于与 V 进行矩阵乘法，生成全局特征  $I_{\text{mes},T}$ ：

$$I_{\text{mes},T} = M_T V. \quad (9)$$

## 4 复现细节

### 4.1 与已有开源代码对比

没有引用他人发布的代码。主要复现该论文 [13] 的其中关于乳腺癌分类的实验，利用一个公共的乳腺癌组织病理图像数据 [11]，对原始高分辨率图像进行 2 倍、4 倍、8 倍的下采样，下采样后的数据集与原先的数据集分别作为 4 个客户端的本地数据集。每个客户端的数据集按照相同图像的不同分辨率分别用于训练集和测试集的原则，随机分为训练集和测试集，比例约为 7: 3，在该实验中各个客户端的本地模型采用 ResNet[17, 11, 8, 5]。

### 4.2 实验环境搭建

本次实验在 Windows11 系统下进行，python 版本 3.8.20，pytorch 版本为 2.4.1，cuda 版本为 12.4。

### 4.3 界面分析与使用说明





|   |                |                       |       |
|---|----------------|-----------------------|-------|
|  correspond.py | 2025/1/6 16:54 | JetBrains PyCharm ... | 7 KB  |
|  DataReader.py | 2025/1/6 13:47 | JetBrains PyCharm ... | 13 KB |
|  MH_pFLID.py   | 2025/1/6 18:33 | JetBrains PyCharm ... | 87 KB |
|  model_test.py | 2025/1/6 18:46 | JetBrains PyCharm ... | 2 KB  |

图 2. 文件界面示意

其中 correspond.py 文件进行模型的通信，model\_test.py 文件使用测试数据集评估模型的性能。详细命令参考 readme.md 文档。

## 5 实验结果分析

MH-pFLID 在知识注入阶段和知识蒸馏阶段的学习率分别设置为 0.0001 和 0.00001。批量大小设置为 8，在实验中，框架的通信轮次为 100。本地训练的轮次为 5（MH-pFLID 在第一阶段为 4 轮，在第二阶段为 1 轮）。在乳腺癌图像的分类实验中， $L_{inj}^l$ 、 $L_{inj}^m$  和  $L_{dis}^m$  是交叉熵损失， $L_{dis}^{con}$  是 KL 损失。分类任务的性能评估基于两个指标：准确率（ACC）和宏平均 F1 分数（MF1）。实验复现结果如图 3 与 4 所示。

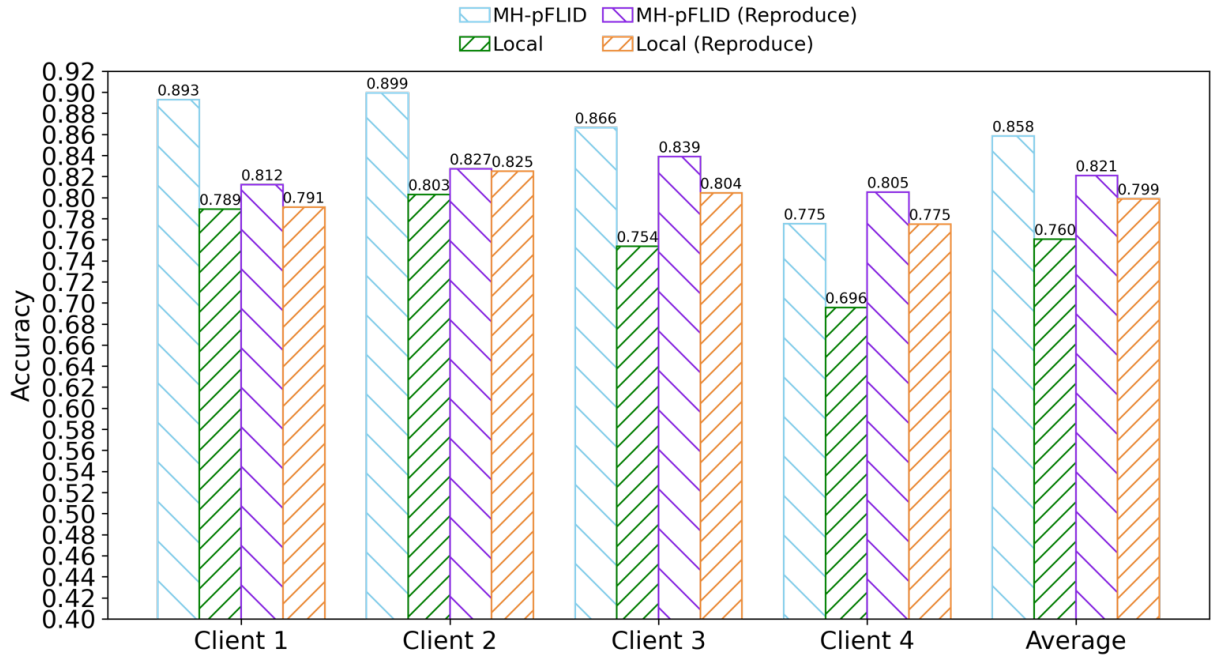


图 3. 乳腺癌图像分类任务 Acc 指标对比实验结果图



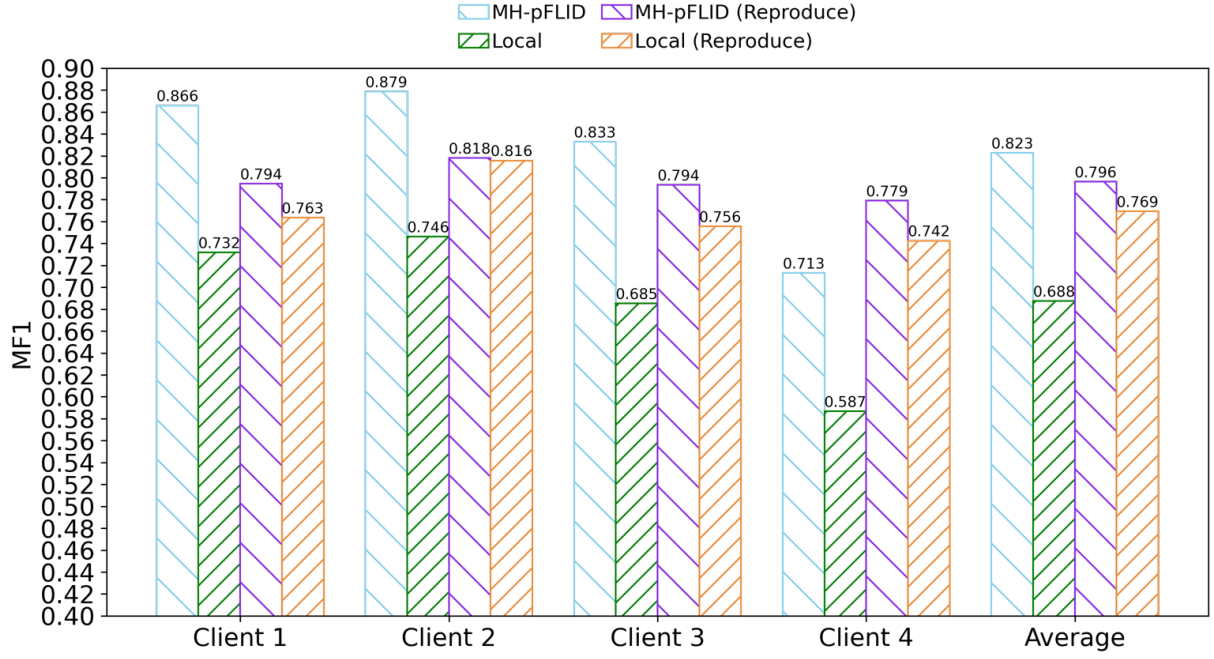


图 4. 乳腺癌图像分类任务 MF1 指标对比实验结果图

考虑到不同的数据集划分方式以及实验环境设置，所以复现的结果与该论文 [13] 有差距，但是根据复现的实验结果，我们可以发现，MH-pFLID 框架很好地实现了各个客户端之间的信息通信，通过信息通信可以使得本地客户端不需要访问其他客户端的情况下实现信息的交流。

## 6 总结与展望

本次研究将该论文 [13] 的方法实现了，考虑到乳腺癌图像不仅有病理切片这一形式，还存在放射影像等其他模态的形式。而多模态的医学图像之间分布在不同的客户端对于客户端之间的学习带来的显著挑战，而解决这一挑战可以很好的提高客户端之间交流本地数据信息，提高模型的鲁棒性。因此后续将会围绕多模态下医学图像进行研究。

## 参考文献

- [1] Manoj Ghuhana Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers.
- [2] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *International conference on machine learning*, pages 2089–2099. PMLR, 2021.
- [3] Gersende Fort. Federated expectation maximization with heterogeneity mitigation and variance reduction.
- [4] G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. 2015. arXiv preprint arXiv:1503.02531.

- [5] Sohei Itahara, Takayuki Nishio, Yusuke Koda, Masahiro Morikura, and Koji Yamamoto. 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.
- [6] Daliang Li and Junpu Wang. Fedmd: Heterogenous federated learning via model distillation. *arXiv preprint arXiv:1910.03581*, 2019.
- [7] Yishay Mansour, Mehryar Mohri, Jae Ro, and Ananda Theertha Suresh. Three approaches for personalization with applications to federated learning. *arXiv preprint arXiv:2002.10619*, 2020.
- [8] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1273–1282. PMLR, 2017.
- [9] Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE transactions on neural networks and learning systems*, 32(8):3710–3722, 2020.
- [10] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. Federated multi-task learning. *Advances in neural information processing systems*, 30, 2017.
- [11] Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):1455–1462, 2015.
- [12] Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang, and Xing Xie. Fedcl: Federated contrastive learning for privacy-preserving recommendation. *arXiv preprint arXiv:2204.09850*, 2022.
- [13] Luyuan Xie, Manqing Lin, Tianyu Luan, Cong Li, Yuejian Fang, Qingni Shen, and Zhonghai Wu. Mh-pflid: Model heterogeneous personalized federated learning via injection and distillation for medical data analysis. *arXiv preprint arXiv:2405.06822*, 2024.
- [14] Jie Zhang, Song Guo, Xiaosong Ma, Haozhao Wang, Wenchao Xu, and Feijie Wu. Parameterized knowledge transfer for personalized federated learning. *Advances in Neural Information Processing Systems*, 34:10092–10104, 2021.