课程论文题目

王怀远

摘要

在现代医疗诊断的发展中，自动化诊断是医疗诊断的一个总体趋势。若一种病变的临床表现呈现了一般的可分辨的规律，那么就可以通过计算机程序辅助进行识别判断，为医生提供辅助诊断。根据医疗数据的稀缺性与病人隐私问题，本文提出了一种新的联邦学习更新算法——FLOP，即在客户端与服务器只共享了部分的模型，并且在本文应用到了医疗数据中。本文在多个主流的模型框架下 (如 COVID-Net、MobileNet-v2，ResNet50，ResNeXt 模型)，证明了 FLOP 算法对深度学习的训练产生影响忽略不计或者产生积极的影响。该算法降低了隐私泄露风险，并且在真实数据集和基准数据集上的实验证实了算法的优势，为支持不同的医院在不共享本地病人数据的情况下协同训练模型提供一种实现方法。

关键词：联邦学习; 医疗数据集; 部分网络

1 引言

由于新型冠状病毒导致的 COVID-19 疾病的爆发，造成了医疗资源的短缺。为了帮助和加速诊断过程，世界各地的研究人员最近探索了通过深度学习模型自动诊断 COVID-19 的方法。虽然已经开发了不同的数据驱动的深度学习模型来减轻 COVID-19 的诊断，但由于病人的隐私问题，数据本身仍然是稀缺的。联邦学习是一个可行的解决方案，因为它允许不同的组织在不共享原始数据的情况下合作学习一个有效的深度学习模型。然而，最近的研究表明，联邦学习仍然缺乏隐私保护，并可能导致数据泄露。我们通过提出一种简单而有效的算法来研究这个具有挑战性的问题，该算法名为使用部分网络的医疗数据集联邦学习，它在服务器和客户端之间只共享一个部分模型。在基准数据和真实世界的医疗任务上进行的广泛实验表明，本文的方法实现了相当或更好的性能，同时降低了隐私和安全风险。特别值得注意的是，我们在 COVID-19 数据集上进行了实验，发现 FLOP 算法可以让不同的医院协作有效地训练一个部分共享的模型，而不需要共享本地病人的数据。

2 相关工作

2.1 疾病检测的更迭与面临的问题

使用机器学习方法进行自动疾病诊断具有巨大的前景，该领域的创新可以完善医疗保健系统，改善全世界的医疗实践。例如，人类消化系统癌症包括食道癌、胃癌和结直肠癌，每年约有 280 万新病例和 180 万人死亡。基于胃肠道内图像的病理结果的自动检测、识别和评估，将有助于医生识别关注的区域，优化对稀缺医疗资源的利用。

随着世界各地的社区和组织继续努力控制大流行病，研究人员试图通过对病人胸部的计算机断层扫描 (CT) 切片 (图像) 进行自动分类来加快 COVID-19 的早期检测。然而，利用这些医学图像有两个主要挑战。一个挑战是，这些数据集体分布在位于不同医院的大量设备或客户身上。当依靠数据驱动的深度学习模型来诊断疾病时，只使用单个设备上隔离的本地数据将不足以训练出一个有效的模型。

第二个挑战是在不损害病人隐私和安全的情况下使用这些数据的必要性。私人数据的泄露不仅是公共媒体的关注点，也是必须保护病人隐私的医院的关注点。

2.2 联邦学习的发展

随着欧洲和世界各地的更严格的隐私法规的出现，研究人员已经开始寻求解决方案，在不损害用户隐私的情况下从用户数据中训练机器学习模型。联邦学习通过消除将数据聚集到单一位置或服务器的需要，分散了传统的机器学习，并已成为满足新数据保护法规的最流行的解决方案[1-3]。

直观地说，联邦学习的机制如下：客户端下载当前的模型，在客户端的本地数据上进行训练，然后向服务器发送模型更新。服务器对来自一组客户的模型更新进行汇总和平均，以改善共享模型。在整个学习过程中，所有的训练数据都保留在客户端设备上。联邦学习是一种机器学习设置，多个实体在中央服务器或者服务提供商的协调下协同解决机器学习问题。

联邦学习可以分为水平联邦学习、垂直联邦学习和联邦转移学习。如果联邦学习中的客户共享重叠的数据特征，但在数据样本上有所不同，我们将其称为水平联邦学习[4]。客户端共享重叠的数据样本但在数据特征上有所不同的情况被称为垂直联邦学习。联邦转移学习是指数据样本或特征都没有重叠的情况。例如，当两家医院为两个不同的区域服务时，与特定疾病相关的数据样本可能是不同的，但具有相似的特征空间，因为疾病是相同的。因此，两家医院可以通过横向联邦学习，在不损失隐私的情况下，合作设计更好的机器学习模型。

联邦学习框架已被应用于许多医疗任务，如预测心脏相关的住院情况[5]。最近专注于 COVID-19 的联邦学习的工作[6-7]通常依赖于客户端之间共享完整的模型。此外，他们没有区分 IID 和非 IID 的数据分布。然而，研究[8]指出，共享一个完整的模型将导致深度泄漏[8-10]。为了解决这些缺点，本文研究了一个新的模型框架，并提出了在医疗数据集上共享一个部分模型进行联邦学习的尝试。在我们的实验中，我们还分析了 IID 和非 IID 的数据分布情况。

每个客户端的原始数据都存储在本地，并且它允许不同的组织合作学习一个有效的深度学习模型，而无需共享原始数据，取而代之的是，使用旨在即时聚合的更新来实现学习的目标。但是最近的研究表明，联邦学习依然可能导致数据泄露，所以本文提出了一种简单而有效的算法来解决这个具有挑战性的问题。

3 本文方法

3.1 FLOP 算法

最近联邦学习的改进包括克服在分布式设备网络上训练机器学习模型的统计挑战，提高安全性，以及个性化。传统的联邦学习框架被证明可以防止针对半诚实服务器的数据泄露，如果梯度聚合是用 SMC 或同态加密操作。然而，最近在[8]中的经验结果表明，共享一个模型可能无法完全保护隐私，梯度交换将导致深度泄漏。作者表明有可能从公开共享的梯度中获得私人训练数据，包括像素级的图像和标记级的句子。避免深度泄漏的一个策略是通过压缩梯度。经验表明，联邦平均法也容易受到攻击。由此，论文提出了一种联邦学习方法，在这种方法中，客户和服务器之间只共享一部分模型。该算法通过将客户端数据封存在其本地设备上来减少隐私和安全风险。

本文的模型架构是基于卷积神经网络 (CNN)，它在计算机视觉、自然语言处理和语音识别中取得

了巨大的经验成功。虽然 CNN 架构有很多变化，但用于图像分类任务的 CNN 通常由两个基本部分组成：特征提取器和分类器。特征提取器包括几个卷积层，然后是最大集合和激活函数，而分类器通常由全连接层组成。在这一观察的激励下，我们注意到将分裂的 CNN 模型纳入联邦学习架构的一种自然方式：具有一般特征域信息的共享特征提取器和具有私有标签和任务信息的私有分类器。

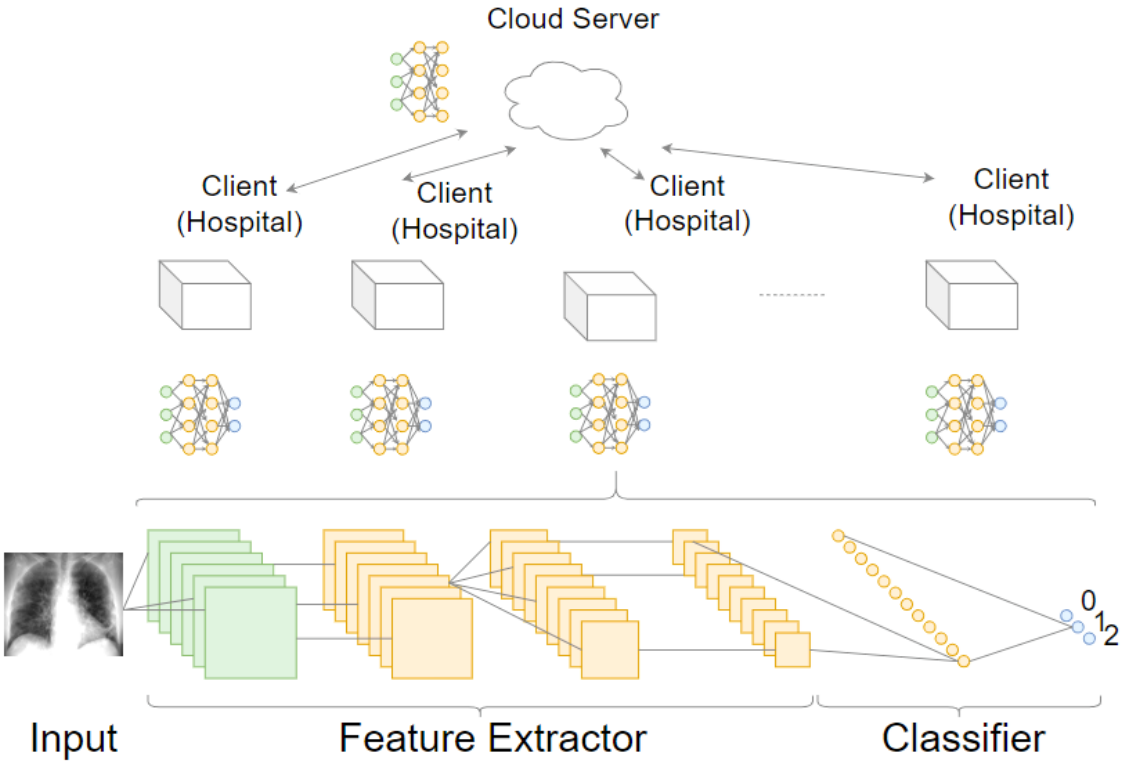


图 1: FLOP 算法示意图

FLOP 将完整模型 M 划分为共享模型 $M_s(Share)$ 和私有模型 $M_p(private)$，如 M_p^u 来表示用户 u 的私有模型。

用水平联邦学习范式来说明。具有相同数据结构的 U 个客户协同学习一个机器学习模型。算法的训练过程如下。

第一步，客户端从服务器接收 M_s，在本地训练自己的模型 $M^u = [M_s, M_p]$, 并且发送 M_s^u 的梯度给服务器。

第二步，服务器从参与的客户那里收到的更新进行安全聚合。

第三步，服务器把 M_s 的汇总结果发给客户端。

第四步，客户端用服务器的结果更新他们的 M_s^u 模型。

上述步骤反复进行，直到损失函数趋近，结束训练。这个过程和算法对任何特定的模型都是不可知的，所有的客户都会得到最终的共享模型参数。

Algorithm 1 FLOP Algorithm

Model Training($\mathcal{M}$): *// Run on user u*

procedure

Receive $\mathcal{M}_s$ from the server and let $\mathcal{M}_s^u = \mathcal{M}_s$

Train the model $\mathcal{M}^u = [\mathcal{M}_s^u, \mathcal{M}_p^u]$ on local training set

Return $\Delta\mathcal{M}_s^u$ to the server

end procedure

Model Update

procedure

Initialize $\mathcal{M}_s$

for each episode $t = 1, 2, \ldots$ **do**

Sample m users U_t

for *each user* $u \in U_t$ *in parallel* **do**

Send $\mathcal{M}_s$ to user u

end for

Receive $\Delta\mathcal{M}_s^u$ from user u;

Update $\mathcal{M}_s = \mathcal{M}_s - \beta \frac{1}{|U_t|} \sum_{u \in U_t} \Delta\mathcal{M}_s^u$

end for

end procedure

图 2: FLOP 伪代码

隐私是联邦学习旨在确保的关键属性之一。在联邦学习中，有不同类型的隐私攻击。最近在 [8] 中的经验结果表明，共享一个模型可能无法完全保护隐私，梯度交换将导致深度泄漏[8-10]。然而，我们的 FLOP 框架解决了这个漏洞，因为它只共享一个部分模型。此外，我们可以通过在步骤 1 中用加密[11]、差分隐私[12]或秘密共享[13]技术掩盖梯度的选择来实现保证隐私，这不在本文的讨论范围之内。

4 复现细节

4.1 与已有开源代码对比

4.2 实验环境搭建

本次实验使用的是 Windows11 的 WSL2 来进行环境的搭建。WSL 是一个使得开发人员可以在 Windows 上同时访问 Windows 和 Linux 的功能。通过适用于 Linux 的 Windows 子系统 (WSL)，开发人员可以安装 Linux 发行版 (例如 Ubuntu、OpenSUSE、Kali、Debian、Arch Linux 等)，并直接在 Windows 上使用 Linux 应用程序、实用程序和 Bash 命令行工具，不用进行任何修改，也无需承担传统虚拟机或双启动设置的费用。

论文以一个开源的联合学习框架为基础，在 PyTorch 深度学习 API 中实现我们的 FLOP。实验是

在真实世界的医疗数据集 (COVIDx 和 Kvasir) 和基准数据集 (FashionMNIST 和 CIFAR-10) 上进行的。具体来说，每个客户都有一个来自原始完整数据集的子数据集。在两个真实世界的医疗数据集上，使用 CovidNet、ResNet50、MobileNet-v2 和 ResNetXt 模型架构进行 FLOP 算法的验证。对于两个基准数据集，我们使用 VGG-11 模型架构和 3 层 CNN 验证 FLOP 的有效性。

下面介绍 COVIDx 和 Kvasir 数据集

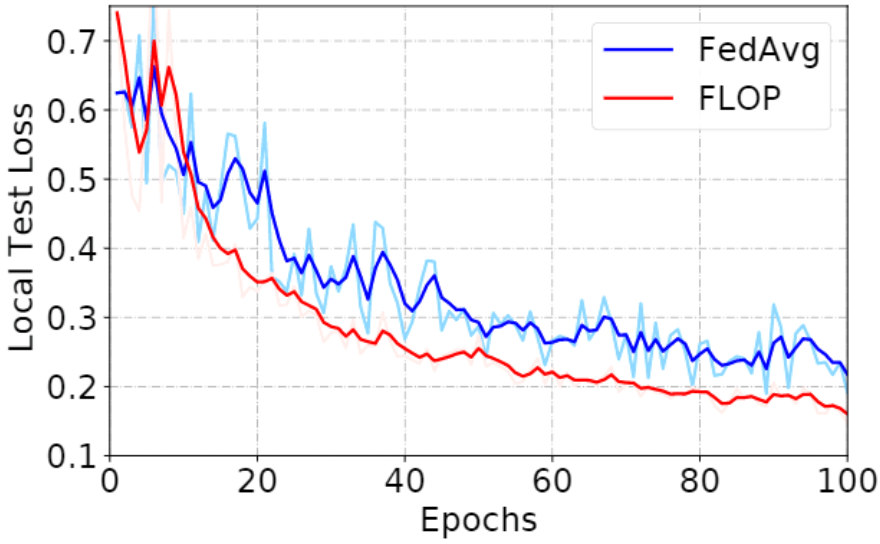
COVIDx。Covid-19 诊断的任务是图像分类，分为三类：(i) 正常 (无感染)，(ii) 肺炎 (非 COVID-19 感染，如病毒、细菌等)，以及 (iii)COVID-19(COVID-19 病毒感染)。COVIDx[5]是开放的基准数据集，拥有最多的 COVID-19 阳性患者病例，是五个公开的 COVID19 数据存储库的组合我们使用 COVIDx 作为我们的训练和测试数据集。由于这些数据集在持续的大流行中不断更新，我们规定在我们的实验中，数据集包括 13,954 张图像用于训练，1,579 张用于测试。训练数据集包含 7,966 张正常图像、5,471 张肺炎图像和 517 张 COVID-19 图像。测试数据集包含 885 张正常图像、594 张肺炎图像和 100 张 COVID-19 图像。

Kvasir。Kvasir 数据集[14]涉及胃肠道疾病的图像分类，有八个类别。它包括显示消化道的解剖标志、病理发现或内窥镜手术的图像，这些图像是使用挪威 Vestre Viken Health Trust 的内窥镜设备收集的。它由 8 个类别的 8,000 张图像和每个类别的 1,000 张图像组成 (6,000 张用于训练，2,000 张用于测试)。这 8 个类别显示了消化道中的解剖坐标 (Z 线、幽门、盲肠)、病理发现 (食道炎、息肉、溃烂性结肠炎) 和息肉切除 (有病和凸起的息肉和病变边缘的切除)。

虚拟机系统为 Ubuntu Server 20.04 LTS，使用 Anaconda22.9.0，使用 PyTorch 1.8.2LTS，Python 版本为 3.9。

5 实验结果分析

图 3 是论文提供的数据。图 4 是训练的代数与 Loss 的曲线图。图中展示了训练的代数和 Loss 总体呈现下降态势。与论文中的数据有较高的吻合度。



(b) ResNet50

图 3: 使用 ResNet50 模型的曲线图

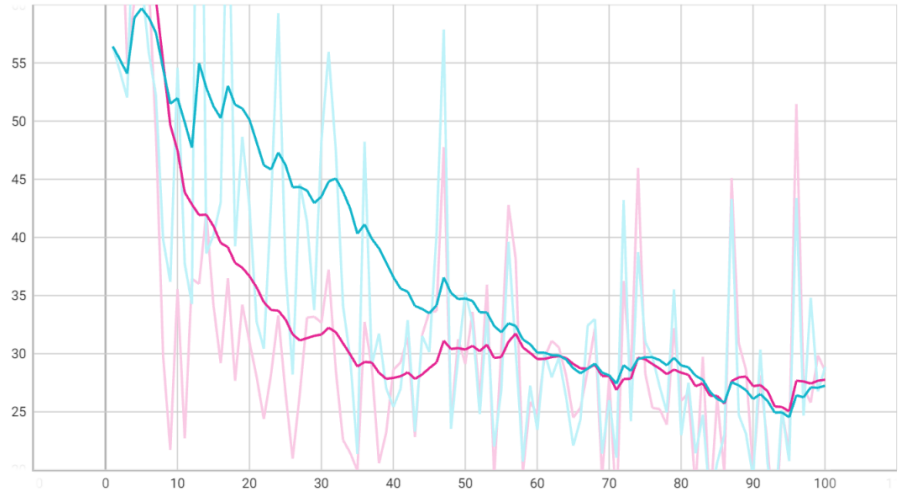


图 4: 本次复现的曲线图

图 5 是使用 ResNet50 的复现结果的混淆矩阵。”n” 代表正常，”p” 代表肺炎，”c” 代表 COVID-19. 其中左一和左二是论文提供的数据，右一是本次复现的数据。

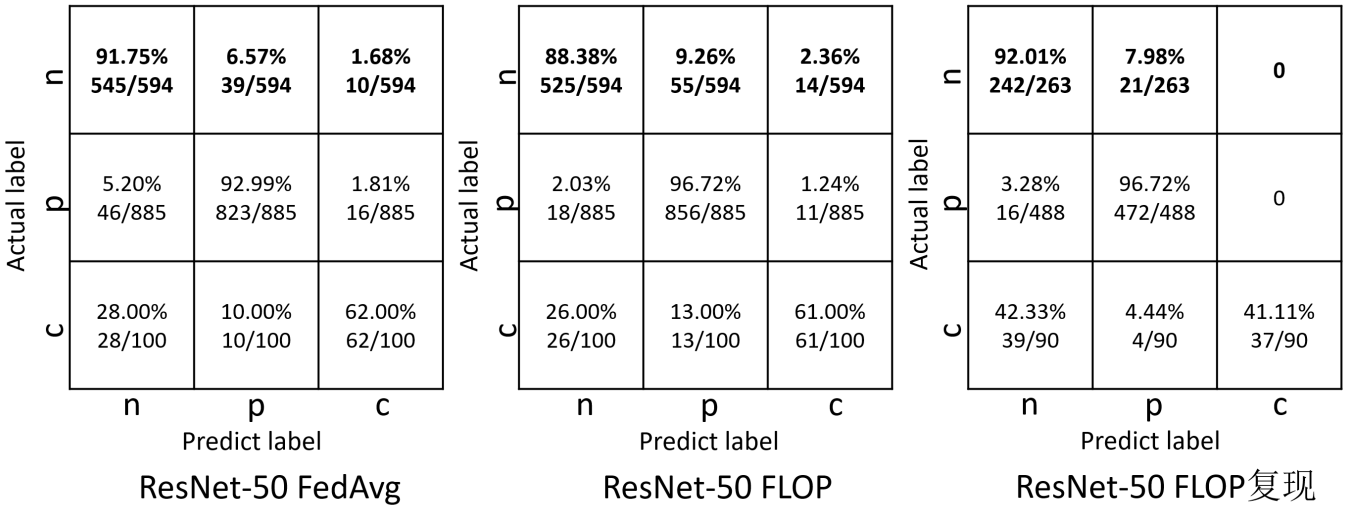


图 5: ResNet50 的复现的混淆矩阵

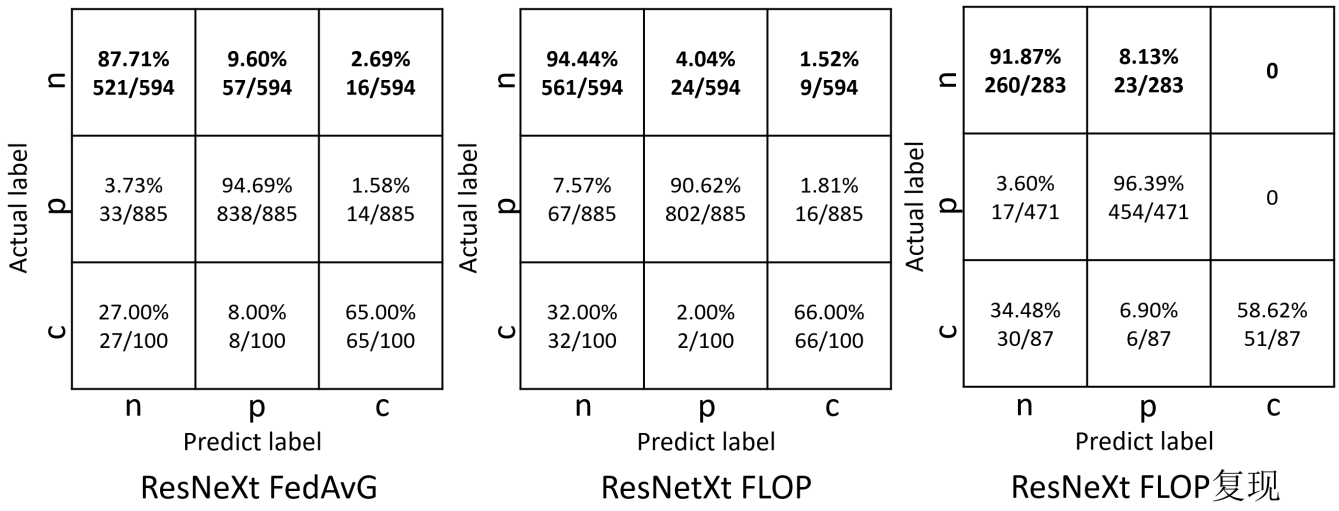


图 6: ResNeXt 的复现的混淆矩阵

从混淆矩阵来看，本次得到的数据较好的复现了论文中的数据。

6 总结与展望

FLOP 论文提出了一种联邦学习方法，在客户端与服务器只共享了部分的模型，并且论文证明了它在医疗数据中的应用。FLOP 算法降低了隐私泄露风险，并且在真实数据集和基准数据集上的实验证实了算法的优势。FLOP 论文提高了深度学习模型在稀缺的医疗任务上的性能。但是 FLOP 算法与其他的保护隐私的方法的联合使用，本文并未涉及。在预测新冠肺炎的正确率上，本文的性能还有进步的空间，

参考文献

[1] KAIROUZ P, MCMAHAN B, AVENT B, et al. 2019. eprint: arXiv:1912.04977.

[2] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-Efficient Learning of Deep Networks from Decentralized Data[M]//Artificial Intelligence and Statistics. 2017: 1273-1282.

[3] YANG Q, LIU Y, CHEN T, et al. Federated Machine Learning[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19. DOI: 10.1145/3298981.

[4] YANG Q, LIU Y, CHEN T, et al. Federated Machine Learning[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19. DOI: 10.1145/3298981.

[5] BRISIMI T, CHEN R, MELA T, et al. Federated learning of predictive models from federated Electronic Health Records[J]. International Journal of Medical Informatics, 2018, 112: 59-67. DOI: 10.1016/j.ijmedinf.2018.01.007.

[6] KUMAR R, KHAN A, KUMAR J, et al. Blockchain-Federated-Learning and Deep Learning Models for COVID-19 Detection Using CT Imaging[J]. IEEE Sensors Journal, 2020, 21(14): 16301-16314. eprint: arXiv:2007.06537. DOI: 10.1109/jsen.2021.3076767.

[7] LIU B, YAN B, ZHOU Y, et al. Experiments of federated learning for covid-19 chest x-ray images[R]. 2020. eprint: arXiv:2007.05592.

[8] ZHU L, HAN S. Deep Leakage from Gradients[M]//Lecture Notes in Computer Science. Springer International Publishing, 2019: 17-31. DOI: 10.1007/978-3-030-63076-8_2.

[9] GEIPING J, BAUERMEISTER H, DRÖGE H, et al. Inverting Gradients-How easy is it to break privacy in federated learning?[R]. 2020. eprint: arXiv:2003.14053.

[10] ZHAO B, REDDY MOPURI K, BILEN H. Evaluation of institutionally reared children and adolescents in terms of mental health in Ankara[J]. IACAPAP ArXiv, 2020. eprint: arXiv:2001.02610. DOI: 10.14744/iacapaparxiv.2020.20001.

[11] AONO Y, HAYASHI T, WANG L, et al. Privacypreserving deep learning via additively homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security, 2017, 13: 1333-1345.

[12] SHOKRI R, SHMATIKOV V. Privacy-Preserving Deep Learning[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. ACM, 2015: 1310-1321. DOI: 10.11

45/2810103.2813687.

[13] BONAWITZ K, IVANOV V, KREUTER B, et al. Practical Secure Aggregation for Privacy-Preserving Machine Learning[C]//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017: 1175-1191. DOI: 10.1145/3133956.3133982.

[14] POGORELOV K, RANDEL K, GRIWODZ C, et al. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection[C]//Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). New York, NY, USA: ACM, 2017: 164-169. DOI: 10.1145/3083187.3083212.