医学场景下联邦学习应用及其隐私保护探究

章 俊

（南昌大学第一附属医院 南昌 330006）

〔摘要〕 **目的/意义** 探究应用联邦学习开展临床研究，在保护患者隐私数据的同时开展大模型训练，推动医学研究发展。**方法/过程** 介绍联邦学习技术框架，重点分析其在医学影像、疾病预测、个性化治疗和新药研发等领域应用的巨大潜力和可能遇到的问题。**结果/结论** 联邦学习提供了一种在医学大数据分析中合作而不共享数据的能力，为跨机构的协同合作提供可能。目前联邦学习在医学研究中尚面临数据异质性、通信效率、模型泛化及安全性等问题，有待进一步深入研究。

〔关键词〕 联邦学习；隐私保护；医学研究

〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2024.09.015

Exploration on the Application of Federated Learning and Its Privacy Protection in Medical Scenarios

ZHANG Jun

The First Affiliated Hospital of Nanchang University, Nanchang 330006, China

〔**Abstract**〕 **Purpose/Significance** To explore the application of federated learning to conduct clinical research, and to carry out large model training while protecting patients' privacy data, so as to promote the development of medical research. **Method/Process** The paper introduces the federated learning technology framework, and analyzes its great potential and possible problems in the fields of medical imaging, disease prediction, personalized therapy, new drug development, etc. **Result/Conclusion** Federated learning provides the capability to collaborate without sharing data in medical big data analysis, and provides the possibility for cross－institutional collaboration. At present, the problems of federated learning in medical research, such as data heterogeneity, communication efficiency, model generalization and safety, need to be further studied.

〔**Keywords**〕 federated learning; privacy protection; medical research

1 引言

近年来，随着信息技术在医疗行业的普及应用，医院信息化不再局限于电子病历、医院信息系统等核心业务系统，康复系统、妇幼系统和血透系统等专项信息系统应用日益广泛。医院存储了海量、高价值的临床数据，利用这些临床数据开展大模型训练，已经成为当前医学研究领域的热点方向[1－2]，通过数据整合、大模型分析和科研成果共享，可以极大地促进医疗研究，特别是在辅助诊断、疾病筛查和治疗评估等方面。

同时，数据安全问题不容忽视。2021 年《中华人民共和国数据安全法》和《中华人民共和国个人信息保护法》相继出台，对数据、个人信息的使用及保护提出明确规范。2023 年 6 月《关于印发卫生健康行业数据分类分级指南（试行）的通知》定义

〔修回日期〕 2024－05－24

〔作者简介〕 章俊，高级工程师，发表论文 5 篇。

了核心医疗数据，细化了医疗数据目录，对医院数据安全治理提出了更高要求。因此，医院必须统筹好医疗数据的共享利用与隐私保护之间的关系。

基于网络地址和端口的传统防火墙、基于数据库视图的数据库防火墙以及针对恶意流量特征的入侵防御系统（intrusion prevention system，IPS）均难以实现跨机构数据分析场景下的患者信息保护。而联邦学习（federated learning，FL）可提供跨机构安全合作范式。联邦学习作为一种分布式机器学习框架，其核心思想是在避免获取各方原始数据的基础上，通过迭代更新梯度参数的方式开展多方模型训练，提供了一种平衡数据利用与隐私保护的解决方案[3-4]。

在临床实践中，单个医疗机构的病历样本通常数量有限且缺乏多样性。例如，苯酮尿症作为新生儿罕见病，单一医院留存的病历样本极为匮乏，小样本学习难以获得可靠模型，而多个机构的共享开放又存在安全合规与个人信息保护等问题。通过联邦学习，不仅实现了跨医疗机构安全合作，而且增强了模型泛化能力以及对疾病的识别率。

2 联邦学习技术框架与隐私保护

联邦学习技术框架可以在解决数据孤岛问题的同时保护用户数据隐私[5-7]。该框架下数据持有者（如医疗机构）可以保留其数据所有权和控制权，只需计算并向中央服务器或聚合者分享梯度信息；中央服务器在收集足够模型更新后，会更新全局模型，并将新模型参数发回至各数据持有者，如此循环往复，即使参与者不直接共享数据，也能训练出高质量的机器学习模型。

2.1 联邦学习技术框架

联邦学习是分布式的计算框架，见图 1。客户端节点指持有本地数据的设备或组织，可以是一个医疗机构，也可以是一个医疗信息系统。客户端节点 k 的样本数量可表示为 n_k，客户端 k 的数据集可表示为 D_k。若数据所在客户端总数为 K，则样本总量计算方式，见公式（1）。中央服务器协调模型更新汇总和分发新模型。在一些联邦学习变体中也可以没有中央服务器，采用去中心化的通信协议。全局模型由所有客户端共同训练，逐渐收敛到一个全局最优模型。本地模型指在每个客户端本地训练的模型副本，根据本地数据集更新并与其他客户端或中央服务器共享模型更新。

$$n = \sum_{k=1}^{K} n_k \tag{1}$$

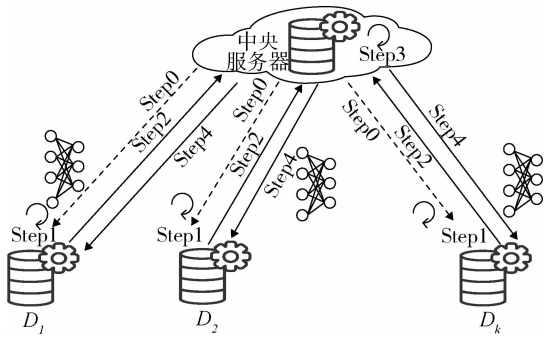

图 1 联邦学习技术框架

联邦学习问题可归结为经验风险最小化问题，见公式（2）。其中，$\frac{n_k}{n}$表示客户端 k 的权重参数，F_k（w）为客户端 k 的损失函数。联邦学习的整体流程中，每一轮迭代更新都重复第 1 步（Step1）至第 4 步（Step4），见表 1。

$$\min_{w \in R^d} F(w) := \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \tag{2}$$

表 1 联邦学习流程

步骤	内容
Step0：初始化	中央服务器初始化一套全局模型参数分发给所有客户端
Step1：本地训练	每个参与者使用其本地数据训练模型，并根据本地训练结果更新模型参数
Step2：模型上传	客户端将模型更新（经过适当的隐私保护处理）发送回中央服务器
Step3：聚合更新	中央服务器收集所有客户端更新，并安全地汇总为一个全新的全局模型
Step4：分发模型	中央服务器将更新后的全局模型发送到所有客户端，并可能继续下一轮训练

2.2 隐私保护技术实现

联邦学习过程中数据不会离开各参与方，因此联

邦学习在隐私保护方面存在天然优势。然而，投毒攻击、对抗样本攻击和女巫攻击等恶意行为仍然可以破坏传统的 FL 隐私保护屏障。为了有效对抗这些攻击，保证信息的机密性、完整性和可用性，联邦学习可与密码学相关算法结合，提高安全鲁棒性。一是安全多方计算（secure multi - party computation，SMC），是一种允许多方参与计算一个共同函数，同时保持各自输入享有隐私的技术；可以用来安全地计算参与者节点的梯度更新或模型权重，而不必暴露单个数据点；通常涉及多轮通信，并依赖于复杂的密码学协议，如秘密分享，以确保在计算过程中个体数据不被泄漏。二是同态加密（homomorphic encryption，HE）[8]，是一类特殊加密技术，允许利用加密数据直接计算，且计算结果在解密后是正确的；可以用来加密参与节点的本地数据，然后在中心服务器或其他节点上进行加密模型训练。这样即使通信过程中数据被截获，攻击者也无法得到任何有用信息，因为数据是在加密状态下进行计算的。三是差分隐私（differential privacy，DP）[9]，通过添加随机噪声来保证输出不会因单个数据项的加入或移除而产生显著差异，从而在统计查询结果中保护个体用户隐私；可以应用于本地模型更新过程，即在上传模型更新之前加入一定量噪声。这种方法可以保证即使模型更新被泄漏也很难追溯到单个数据样本。联邦学习与 SMC、HE 或 DP 等算法相结合，可以增强隐私数据安全性。例如，将差分隐私应用于本地训练的模型更新，然后使用安全多方计算或同态加密来安全地聚合这些更新。此外，隐私保护技术需根据不同场景因地制宜进行配置，以达到安全和效率之间的平衡。

3 联邦学习的临床应用及隐私保护

3.1 联邦学习与医学影像

联邦学习可以融合多个医疗机构的影像信息，训练有效的模型用于辅助诊断。如将影像胶片作为训练样本，诊断结果作为训练标签开展监督学习，推理出更加精准的疾病特征从而帮助医生降低误诊率。融合不同医院的群体差异，可克服单一机构样本同质性问题，通过更广泛的数据集改善医学图像识别率和分类任务性能。Sheller M J 等[10]首次将联邦学习应用于磁共振成像（magnetic resonance imaging，MRI）脑部肿瘤医学图像分析，在保护隐私的同时，其结果表明联合多家医疗机构模型识别准确率要高于单一机构训练模型。联邦学习在影像学的应用可分为多中心合作和多模态学习两类场景。

3.1.1 多中心合作 多中心合作模式可以推动具有不同地理位置和人口统计特征的多家医院协作，有效推进罕见病研究。罕见病种临床数据因其稀缺性难以收集，而丰富的样本量是大模型稳定可靠的关键，因此单一医疗机构难以对罕见病开展研究。联邦学习可将分布在不同地点的异构数据特征有效且安全地融合，为多中心医疗合作奠定技术基础。

3.1.2 多模态影像 医学影像涉及不同成像技术，如 X 射线、MRI、电子计算机断层扫描（computed tomography，CT）和正电子发射断层显像 - X 线计算机体层成像仪（positron emission tomography - computed tomography，PET - CT）。多模态联邦学习通过表征、对齐和融合的方式从 MRI、CT 等异构数据中提取共享或关联的隐藏信息，实现不同模态间互补信息的高效利用，见图 2。这种联邦学习已用于跨模态数据分析，使单一算法能够在不同成像技术下更准确地诊断疾病。

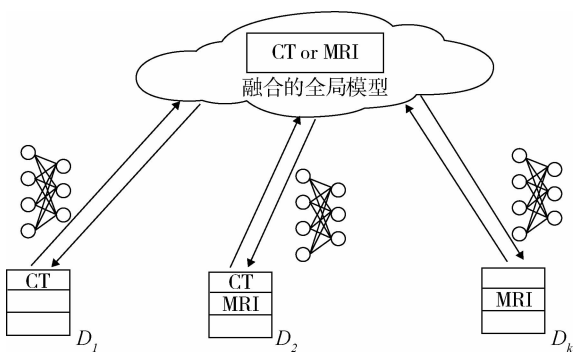

图 2 多模态影像的联邦学习框架

3.2 联邦学习与疾病预测

联邦学习可支撑临床医生和研究人员预测患者的疾病风险。联邦学习可以生成具有良好泛化能力的模型，用于准确预测疾病的发生和趋势。研究表

明，联邦学习已在癌症、糖尿病和心血管疾病等领域取得一定应用成果。例如，腾讯天衍实验室与微众银行研究成果表明，通过联邦学习电子病历及其相关联数据训练的模型能有效预测患者脑卒中，其识别准确率达 80%[11]。张傲[12]在联邦学习开源框架（federated AI technology enabler，FATE）下，分别应用横向联邦学习和纵向联邦学习对患者的中风情况进行预测，并与随机森林、逻辑回归、多层感知机和 LightGBM 等方法对比，证明联邦学习在不共享数据的情况下能准确预测中风疾病。如果在模型训练过程中，能够引入临床决策支持系统的知识库等作为高质量数据参与模型训练，可以一定程度上避免模型训练陷入某个局部解，进而提升联邦学习疾病预测准确性。

3.3 联邦学习与个性化治疗

联邦学习在个性化治疗领域的应用处于蓬勃发展阶段，并正在成为精准医学研究的支撑工具。联邦学习可以利用分布在不同机构中异构的病历、检验检查和基因组数据，经过有诊断标签的监督学习和无监督学习的聚类和关联规则，识别出数据模式和相关特征。利用这些特征构建疾病预测模型，帮助医生预测疾病的进展、患者对治疗的反应以及出现并发症的可能性。医生可能向患者推荐多种治疗方案，而联邦学习能够自适应地动态迭代，更准确地预测针对特定患者的最佳治疗方案。在肿瘤治疗方面，联邦学习可以识别治疗反应存在显著差异的特定患者亚群体，有效的亚群体识别可以为患者推荐更加个性化的治疗方案，如靶向疗法或免疫疗法。此外，联邦学习还可以在患者早期治疗阶段提供临床辅助决策，如患者对某些药物的过敏反应、耐药性等情况。

3.4 联邦学习与新药研发

基于蛋白组学的药物研发是近年热点研究领域，其原理是通过研究蛋白质结构和功能发现新的药物靶点并设计治疗药物。新药研发过程要收集大量蛋白质药物在临床试验中成功及失败的数据用于模型训练，而单个机构无法累积如此海量的临床实验数据。通过联邦学习可以收集大量分布于不同机构的药物反应数据，辅助研究新的靶点和药物，以及个体对特定药物的反应模式。尤其是基于蛋白质的药物治疗，利用联邦学习不仅可以训练出模型用以改进氨基酸序列、缩短新药研发时间，还可以提供跨机构共同研发的安全范式，避免临床数据泄漏。鞠鑫等[13]研究指出，医院和卫生健康委员会可以通过联邦学习提供数据支撑新药研发，也可以在药企之间通过隐私计算助力新药研发。目前，全球多家顶尖药企通过联邦学习隐私计算安全范式合作新药开发。

联邦学习在这些医学场景的共同特点是将跨机构的分布式学习与患者隐私保护相结合，使预测模型在分析数据时不会直接获取患者的隐私信息，同时提升预测或者识别的准确性和泛化能力。不同之处在于每个应用场景的侧重点不同。例如，联邦学习在医学影像中的应用更注重训练模型的识别率，而在疾病预测场景下更注重模型泛化能力。

4 联邦学习在医学研究中的挑战

联邦学习提供了一种在医学大数据分析中合作而不共享数据的能力，为涉及多个系统、机构的协同工作打开了新的大门。然而，联邦学习在医学研究中面临挑战，如数据异质性、通信效率低、模型泛化及安全性等问题。首先，数据来源不仅限于医疗机构，还可以是可穿戴设备、健康类手机 App 等，这种数据异质性将削弱学习模型的健壮性。解决策略包括采用群体标准化等技术，设计更健壮的联邦学习算法以及在训练前对数据预清洗。其次，各参与者的通信效率是联邦学习的瓶颈之一。联邦学习需在参与者之间频繁交换更新参数，通信代价很大，影响训练效率。尤其针对医学影像这类大型数据，如何减少通信代价成为重要问题。不仅如此，随着参与者的增多，联邦学习的通信效率瓶颈也将更加突显。后续可以从减少迭代轮次、模型压缩和聚合算法优化等方面提升通信效率。再次，联邦学习模型的泛化能力受多方面因素制约。例如，医疗水平差异造成对同一患者在不同医院的诊断结

果存在差异。这种样本标签的不一致将影响有监督学习模型的分类，从而降低其泛化能力。同时，每个医疗机构的数据收集环境、客户群体、设备种类等可能不同，导致本地数据分布和全局数据分布差异显著。基于这样不均匀的数据训练，模型可能对一些参与者来说表现良好，而对其他参与者或全新数据来说则表现不佳。提高联邦学习模型的泛化能力是一个热点研究方向，后续可将各参与方训练得到的共性特征与个性特征进行特征解耦，从而增强联邦学习模型的泛化性能。最后，患者隐私安全性仍有提升空间。在联邦学习过程中，恶意攻击者可以根据各参与方更新梯度和全局模型逆向演算出部分有效信息。后续研究可以结合鲁棒性更强的加密算法或者区块链技术提升安全防护能力。

5 结语

联邦学习是一种保护数据隐私的机器学习方法，允许不同医疗机构在不直接共享数据的情况下协作建立共同模型，可应用在医学影像诊断、疾病预测、个性化医疗和新药研发等方面。联邦学习还可以与 SMC、HE 和 DP 等密码学方法相结合，增强跨机构模型训练的安全性。联邦学习为跨国医学研究创造了可能，特别是考虑到不同国家有不同的隐私和数据保护要求，这种原数据不出本地的合作方式可以促进全球范围内医疗机构协作和知识共享，加速医学事业的进步和创新。

参考文献

1 ESTEVA A, CHOU K, YEUNG S, et al. Deep learning – enabled medical computer vision [J]. NPJ digital medicine, 2021, 4 (1): 1 – 9.

2 JAMES H, RUDRA P K, DIMOSTHENIS T, et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker [J]. Neuroimage, 2017, 163 (1): 115 – 124.

3 肖雄，唐卓，肖斌，等. 联邦学习的隐私保护与安全防御研究综述 [J]. 计算机学报, 2023, 46 (5): 1019 – 1044.

4 NGUYEN D C, DING M, PATHIRANA P N, et al. Federated learning for industrial internet of things in future industries [J]. IEEE wireless communications magazine, 2021, 28 (6): 192 – 199.

5 VIRAAJI M, PARIXI R M, POURIYEH S, et al. A survey on security and privacy of federated learning [J]. Future generation computer systems, 2021, 115 (1): 619 – 640.

6 NGUYEN T, THAI M. Preserving privacy and security in federated learning [J]. IEEE/ACM transactions on networking, 2024, 32 (1): 833 – 843.

7 AHMED I, THAKKER U, WANG S Q, et al. A survey on federated learning for resource – constrained IoT devices [J]. IEEE internet of things journal, 2022, 9 (1): 1 – 24.

8 FANG H, QUAN Q. Privacy preserving machine learning with homomorphic encryption and federated learning [J]. Future internet, 2021, 13 (4): 1 – 20.

9 XIN B, YANG W, GENG Y, et al. Private FL – GAN: differential privacy synthetic data generation based on federated learning [C]. Singapore: The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020.

10 SHELLER M J, REINA G A, EDWARDS B, et al. Multi – institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation [C]. Granada: International MICC AI Brain lesion Workshop, 2018.

11 李瑾. 多中心电子病历数据协同挖掘及建模方法研究 [D]. 杭州：浙江大学, 2022.

12 张傲. 基于联邦学习的疾病预测研究 [D]. 哈尔滨：哈尔滨工业大学, 2021.

13 鞠鑫，曹京，陈佛忠，等. 隐私计算在卫生健康行业的应用与安全研究 [J]. 信息通信技术与政策, 2023, 49 (2): 43 – 48.