Computer Science and Application 计算机科学与应用, 2026, 16(5), 231-242
Published Online May 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.165179

Hans汉斯

基于区块链与全同态加密的安全联邦学习肺炎识别方法

吴毓婧，杨丁宇，林濠浚，庞异凡，周凌枫*

宁波工程学院网络空间安全学院，浙江 宁波

收稿日期：2026年4月21日；录用日期：2026年5月18日；发布日期：2026年5月27日

摘　要

针对肺炎医疗数据联邦学习中存在单点故障风险、隐私泄露隐患及恶意投毒攻击等问题，文章提出了名为BA-HEFL的安全联邦学习方法。该方法是一种基于全同态加密和联邦学习的去中心化肺炎识别方法，采用CKKS全同态加密对所有本地模型梯度进行加密，有效抵御梯度泄露攻击；引入区块链对全局模型的聚合流程进行审计，保障了模型的防篡改、抗单点故障等功能。实验结果表明，该方法取得了0.837的分割精度(Dice系数)与高达0.881的召回率，同时能够有效防御高达50%参与者发起的模拟投毒攻击，并切实保障了模型更新的机密性。

关键词

联邦学习，区块链，CKKS全同态加密，无监督异常检测算法

A Secure Federated Learning Method for Pneumonia Detection Based on Blockchain and Fully Homomorphic Encryption

Yujing Wu, Dingyu Yang, Haojun Lin, Yifan Pang, Lingfeng Zhou*

School of Cyber Science and Engineering, Ningbo University of Technology, Ningbo Zhejiang

Received: April 21, 2026; accepted: May 18, 2026; published: May 27, 2026

Abstract

To address issues such as single-point failure risks, privacy leakage vulnerabilities, and malicious

*通讯作者。

文章引用: 吴毓婧, 杨丁宇, 林濠浚, 庞异凡, 周凌枫. 基于区块链与全同态加密的安全联邦学习肺炎识别方法[J]. 计算机科学与应用, 2026, 16(5): 231-242. DOI: 10.12677/csa.2026.165179

poisoning attacks in federated learning for pneumonia medical data, a secure federated learning method named BA-HEFL was proposed. This method is a decentralized pneumonia recognition approach based on fully homomorphic encryption (FHE) and federated learning. It adopted the CKKS fully homomorphic encryption scheme to encrypt all local model gradients, which effectively defended against gradient leakage attacks. It also integrated blockchain technology to audit the aggregation process of the global model, thereby ensuring the model's tamper resistance and resilience to single-point failures. Experimental results showed that the proposed method achieved a segmentation accuracy (Dice coefficient) of 0.837 and a high recall rate of 0.881. Meanwhile, it could effectively defend against simulated poisoning attacks launched by up to 50% of participants and guarantee the confidentiality of model updates.

Keywords

Federated Learning, Blockchain, CKKS Fully Homomorphic Encryption, Unsupervised Anomaly Detection Algorithm

1. 引言

随着科技的不断发展，深度学习技术正持续赋能医疗诊断、图像识别等多个领域[1]。医疗机构依托深度学习，可以在海量的数据中提取出有用的价值，并且进行细致的解析与处理。但是，传统的中心化训练模型需要将敏感患者数据集中存储训练，存在着隐私泄露风险以及单一机构样本量不足导致的数据孤岛问题。为了解决以上问题，分布式学习得到了广泛的应用[2]。

联邦学习作为一种新的分布式学习框架[3]，可以联合多个本地数据，只共享本地模型参数，在保证数据隐私的同时，解决了数据孤岛问题，为医疗机构的数据学习提供了更多的可能。虽然联邦学习为模型的训练和聚合提供了新的方法，但仍然存在着诸多问题：调度者处理梯度引发的隐私泄露风险[4] [5]、恶意参与方对训练模型进行投毒攻击，以及服务器操作不透明存在信任危机。

因此，本文提出了一种基于全同态加密和联邦学习的去中心化肺炎识别方法，该方法以联邦学习为框架，通过引入 CKKS 全同态加密(FHE)保护本地模型参数，加密梯度信息以防止隐私泄露。CKKS 支持浮点数加密计算，具备更快的加解密速度，能在保障医疗数据模型精度的同时提升计算效率。为了解决由中心化而导致的信任危机以及单点故障风险，本文将区块链技术与联邦学习相结合，通过引入区块链作为审计中心，并利用其区块链去中心化、不可篡改的特点，构建了一条完全透明、不可篡改的审计证据链，有效解决了传统联邦学习存在的信任危机问题。

2. 相关工作

目前，联邦学习在医疗领域的应用面临着隐私泄露、恶意攻击和信任问题等多个关键挑战。针对这些问题，国内外学者已提出了多种解决方案，但现有方法仍存在一定的局限性。以下从梯度泄露风险、投毒攻击防御和信任危机三个方面，分别综述相关研究进展。

2.1. 梯度泄露风险问题

针对梯度隐私泄露风险问题，有研究表明，提供计算变化的梯度，可以还原参与者的隐私数据[6]。还有研究通过在梯度中注入噪声来防止隐私数据被恶意还原。Wei 等人[7]提出，通过在客户端聚合前的

模型参数中加入噪声，可在保障模型完整性的同时更高效地实现隐私保护。目前更常见的隐私防护方案是采用 Paillier 半同态加密，对本地训练后的梯度进行加密。然而，该方法仅支持整数加密计算，且只能处理单数据加密，因此需对明文梯度进行量化才能加密，这不仅显著增加计算开销，还会导致模型精度损失。

2.2. 恶意投毒攻击问题

针对恶意投毒攻击[8]问题，Tolpegin 等人[9]证明，即使仅存在小比例恶意投毒参与者，投毒攻击仍会导致模型分类准确率与召回率显著下降。Cao 等人[10]提出了 Sniper 防御方案，通过求解最大团问题识别良性本地模型，并在全局模型更新时忽略恶意的本地模型。Blanchard 等人[11]提出了 Krum 防御模型，通过检测梯度差异来识别恶意节点。但是这些模型目前还存在着若干瓶颈，如 Krum 防御模型需要计算所有参与者的梯度相似度，所需的计算开销极大。

2.3. 信任危机与单点故障问题

传统的参数服务器架构依赖集中式服务器进行梯度聚合和模型更新，会存在中心化风险，以及透明度不足的问题。若服务器被恶意操控或遭受攻击(如 DDoS)，可能导致整个联邦学习系统瘫痪。并且参与者无法验证服务器是否篡改聚合结果，存在信任隐患。

针对服务器操作不透明存在信任危机的问题，朱建明等人[12]提出，利用区块链将联邦学习中心化的参数服务器构建成去中心化的参数聚合链，通过区块链记录模型训练过程。通过激励机制，激励协作节点进行模型参数验证，惩罚上传虚假参数的节点。区块链[13] [14]具有去中心化、抗单点故障和不可篡改等特点，可以更好地为联邦学习进行赋能。

3. 系统方法

3.1. 系统模型及威胁模型

系统框架包含以下四个逻辑实体，分别是：1) 客户端节点(Client Node, CN)：集合表示为 $C=\{\mathrm{CN}_1,\mathrm{CN}_2,\cdots,\mathrm{CN}_N\}$，代表 N 个持有敏感数据的参与方。每个客户端节点 CN_i 拥有其私有本地数据集 D_i。2) 协调节点(Coordinator Node, CoN)：具备较强计算能力的中心化实体，负责全局模型的聚合与分发。3) 区块链审计网络(Blockchain Audit Network, BAN)：基于许可链的分布式账本，充当系统的“公共审计层”。4) 分布式存储层(Storage Layer)链下的分布式存储系统：用于高效存取大体积文件。在第 t 轮训练中，全局模型表示为 W_t。每个客户端节点 CN_i 基于其本地数据 D_i 和当前模型 W_t 计算损失函数 $L(W_t;D_i)$，并获取本地梯度 $G_i=\nabla L(W_t;D_i)$。

本研究采用了联邦学习安全领域前沿广泛认同的 2 种攻击者模型[15]。假定协调节点为“诚实但好奇”(Honest-but-Curious)，CoN 忠实执行协议，但可能试图分析其接收到的所有中间信息(如梯度 $\{Gi\}\,Ni=1$)以推断客户端节点的隐私数据。假定部分客户端节点是“恶意的”(Malicious)，在 N 个客户端节点中，存在恶意客户端节点，这些节点可能会偏离协议，提交旨在破坏全局模型性能或植入后门的恶意更新。

3.2. 系统安全威胁

传统医疗领域的联邦学习存在以下安全问题：

1) 单点故障问题：传统的联邦学习高度依赖中央服务器，因此一旦中央服务器出现故障，整个联邦信息都将无法进行。

2) 信任危机：在传统联邦学习中，只会由中央服务器记录整体数据的聚合，整体流程透明度和可信任度低。

3) 数据隐私泄露：虽然边缘服务器是通过上传训练好的模型参数到中央服务器进行聚合，但是有研究表明，可以通过上传的模型参数，反向推导出某些参与方的隐私数据[16]。

3.3. 模型框架

为了解决 3.2 节中提出的三个传统的肺炎识别联邦学习中存在的安全问题，构建了如图 1 所示的系统架构图。该框架由客户端群组、存储层、区块链审计网络组成。

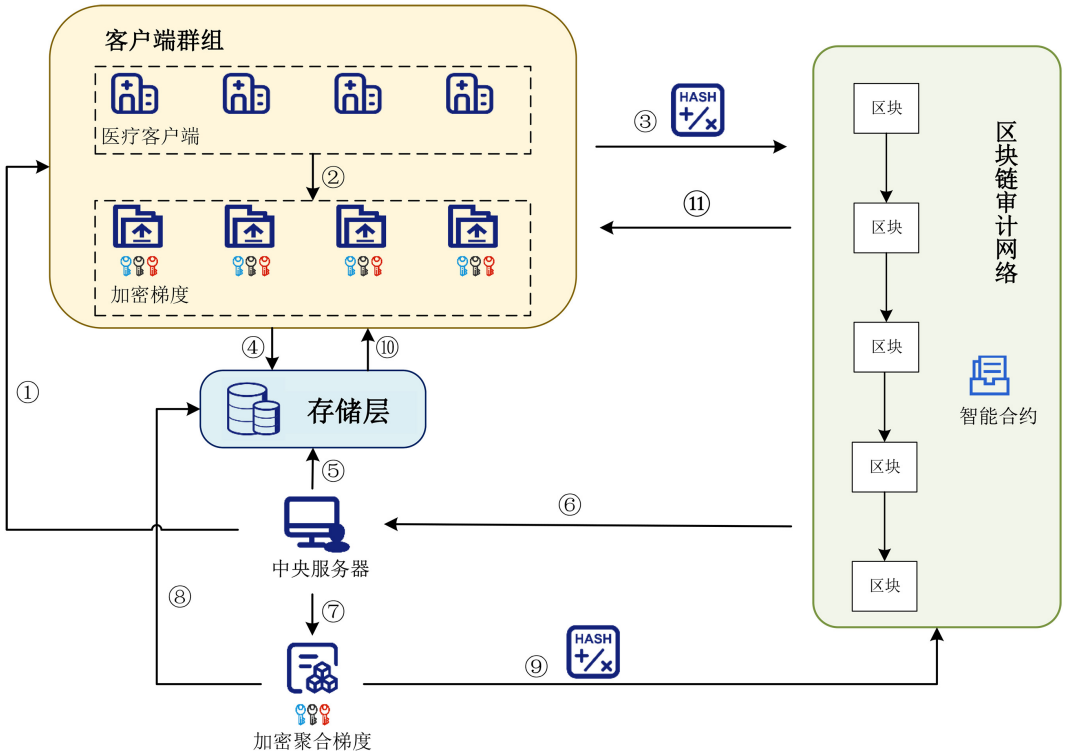

Figure 1. System architecture diagram
图 1. 系统架构图

① 中央服务器发布训练任务并且完成区块链注册。

② 各个医疗客户端在各自的本地服务器上进行模型训练，然后使用 CKKS 对模型进行加密。

③ 各个医疗客户端将各自加密后的模型梯度哈希值上传至区块链。

④ 各个医疗客户端将各自的加密模型梯度上传至存储层。

⑤ 中央服务器从存储层获取各个客户端训练好的加密模型梯度。

⑥ 中央服务器从区块链中获取各个客户端上传的哈希值，并且与存储层获取的加密模型梯度进行对比。

⑦ 中央服务器对模型进行聚合训练。

⑧ 中央服务器上传聚合好的加密模型至存储层。

⑨ 中央服务器上传聚合加密模型的哈希值至区块链。

⑩ 客户端群组从存储层下载聚合加密模型。

⑪ 客户端群组从区块链下载加密模型的哈希值并进行比对，然后从客户端群组中随机决定新一轮的中央服务器。

3.4. 关键技术实现原理

3.4.1. CKKS 全同态加密聚合方案

为了解决 3.2 中提出的数据隐私泄露问题，本研究对多种隐私保护技术进行了权衡。其中，差分隐私(DP)通过注入噪声保护隐私，但通常以牺牲模型精度为代价[17]；安全多方计算(SMPC)虽能实现无损计算，但其高昂的通信与同步开销使其扩展性受限。因此，本研究决定使用 CKKS 全同态加密方案[18]作为隐私保护技术。相较于其他的同态加密算法，CKKS 支持浮点数密文计算，并且加密和解密的速度更快、更高效，这可以让模型在加密的状态下实现高效的聚合。

3.4.2. CKKS 密钥管理与分布式解密机制

在基于全同态加密的联邦学习系统中，密钥管理机制是保障整体安全性的关键环节。若缺乏对密钥生命周期的系统性设计，即使采用 CKKS 加密方案，仍可能由于密钥泄露或不当使用而导致隐私风险。因此，本文在加密聚合机制的基础上，进一步设计了一套完整的密钥管理与分布式解密方案，以避免单点解密带来的安全隐患。

在系统初始化阶段，各客户端节点通过安全协商机制共同参与密钥生成过程。不同于传统单密钥模式，系统不生成完整私钥并分发，而是由各参与方本地生成私钥份额，并协同构建全局公钥用于模型参数的加密。同时，为支持同态运算的高效执行，还生成相应的评估密钥用于密文计算过程。该过程确保了私钥在生成阶段即以分布式形式存在，从源头避免了私钥集中化带来的风险。

在模型训练与聚合阶段，各客户端节点利用全局公钥对本地模型参数进行加密，并上传至存储层。被选举出的协调节点仅在密文域内执行加法与标量运算，实现全局模型的同态聚合。在此过程中，协调节点无法获取任何明文信息，也不具备解密能力，从而满足“诚实但好奇”威胁模型下的隐私保护要求。

在模型解密阶段，本文不引入任何集中式解密服务器，而是采用分布式门限解密机制完成密文恢复。具体而言，各客户端节点在接收到聚合后的密文模型后，利用自身持有的私钥份额对密文执行部分解密计算，生成对应的部分解密结果。随后，通过收集不少于预设阈值数量的部分解密结果，并进行组合运算，即可恢复出全局模型的明文参数。在该过程中，任一参与方均无法单独完成解密，也不存在完整私钥的重构过程，从而有效防止了潜在的密钥集中攻击。

需要说明的是，本文在工程实现中基于 tenseal 库构建 CKKS 加密模块，该库采用单密钥机制实现加解密流程。然而，所提出的系统框架在设计上与门限 CKKS 完全兼容，可在支持多方同态加密的密码学库中无缝扩展为分布式解密模式。因此，该实现不影响本文方法在实际多方安全环境中的可部署性与扩展性。

此外，为进一步提升系统的可验证性与抗篡改能力，解密触发过程可与区块链审计机制相结合。仅当链上记录的模型哈希与存储层数据一致时，客户端节点才执行部分解密操作，从而避免恶意模型被提前解密或传播。该机制使得密钥使用过程具备可审计性，并增强了系统整体的可信性。

3.4.3. 基于区块链的审计网络

为解决传统联邦学习中存在的信任缺失与单点故障问题，本文引入区块链作为公共审计层，对模型训练与聚合过程进行可信记录与验证，将训练流程变成公开、可验证、不可篡改的信任化框架。

在具体实现中，本文选用以太坊私有链作为底层区块链平台，并基于联盟场景构建许可链环境。该模式下，仅允许通过认证的节点加入网络，从而在保证系统开放性的同时，提升整体安全性与可控性。

在系统运行过程中，通过智能合约实现以下核心功能：客户端节点注册、模型梯度哈希上传、模型一致性验证以及下一轮协调节点的选举。具体而言，在每一轮训练开始阶段，各客户端(CN_i)将其本地加密后模型梯度对应的哈希值凭证(Update_CID_i)通过智能合约上传至区块链；在聚合完成后，协调节点(CoN)会将全局聚合模型的哈希凭证通过智能合约再次写入链上，从而形成完整的模型演化记录，实现全过程的公开透明与可追溯性。

在节点选举机制方面，由于本研究采用的是许可链环境，区块链网络通过白名单准入机制控制节点的加入。在每一轮训练结束后，系统基于链上记录的节点历史行为(如提交更新的有效性、参与度等)，结合随机策略，从合法节点(CN)集合中选举产生下一轮的协调节点(CoN)，从而避免中心化带来的长期信任依赖问题。

从系统开销角度分析，由于采用以太坊私有链，其共识机制可根据应用需求进行定制，相较于公有链无需复杂的工作量证明过程，本方案的计算成本与通信延迟方面开销较低。在本实验环境中，区块链模块主要用于存储模型哈希及执行验证逻辑，其引入未对整体训练流程造成显著性能瓶颈。

在安全性方面，区块链提供的不可篡改性与分布式账本特性，使得模型更新过程具备天然的抗篡改能力。同时，通过链上哈希校验机制，可以有效防止模型在存储或传输过程中被恶意替换。此外，结合联邦学习与同态加密技术，区块链进一步增强了系统整体的可验证性与可信性。

3.4.4. 自适应与高效的模型训练策略

高效的训练性能是必不可少的，特别是在医疗影像分割这类常面临数据非独立同分布(non-IID)、正负样本极度不均衡以及小目标难以学习等问题中，本研究设计了一整套自适应的高效训练策略。

在模型架构层面，本研究通过引入深度可分离卷积[19]和无参数 SimAM 注意力机制[20]，构建了一个轻量高效的 U-Net 模型。其具体网络结构如图 2 所示，图中清晰标注了编码器–解码器结构、跳跃连接，并突出展示了深度可分离卷积、SimAM 模块及深度监督侧输出(Side Outputs)在网络中的具体位置。

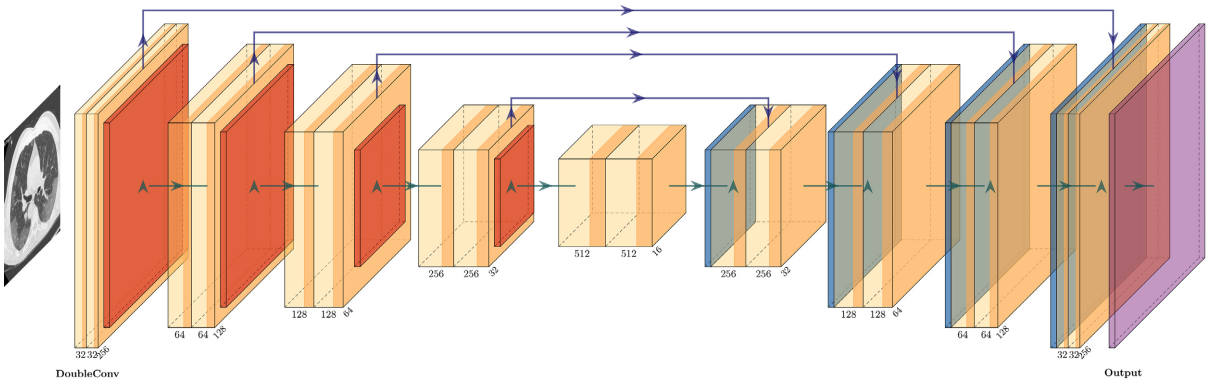

Figure 2. Architecture of the enhanced U-Net model
图 2. 增强型 U-Net 模型架构图

4. 实验设置

4.1. 数据采集以及场景构建

本文采用公开的 COVID-19 CT 病灶分割数据集[21]作为实验数据来源。该数据集由多个公开数据源整理而成，包含 2729 对肺部 CT 图像及其对应的像素级病灶标注掩码(Mask)。所有病灶区域均由专业人员标注，并统一映射为二值标签(病灶/非病灶)，以保证跨数据源的一致性。

该数据集融合了多个真实临床场景中的 COVID-19 感染病例，涵盖多种典型病灶表现形式(如磨玻璃

影、实变等)，具有较强的代表性与挑战性。

在数据划分方面，为模拟真实医疗场景中的数据异构性，本文采用基于 Dirichlet 分布的 non-IID 数据划分策略。具体而言，对于每一类别，包括病灶与非病灶，分别从 Dirichlet 分布 $\mathrm{Dir}(\alpha)$ 中采样客户端数据占比，其中浓度参数 α 控制数据分布的不均衡程度。在实验中设置 $\alpha = 0.5$，以构建高度异构的数据环境。

为避免极端情况下某些客户端仅包含单一类别数据的问题，本文进一步引入最小样本约束机制，若全局数据存在正负样本两种类别，便确保每个客户端在训练集与验证集中均至少包含一个正样本与一个负样本。同时，在采样过程中采用无放回随机采样策略，以避免数据重叠，提高实验的真实性与泛化能力。

为了方便训练，仿真时，模型将所有 512 × 512 像素的输入图像数据统一缩放至 256 × 256 像素再进行归一化处理。该处理对系统框架的安全性与性能的评估影响可以忽略不计。

4.2. 评估指标与对比基线

为了验证系统框架的整体有效性与安全价值，选用了无任何防御机制的 Vanilla FedAvg 作为安全分析维度的核心对比基线，以便于清晰地量化系统所引入的安全机制在抵御模型投毒等恶意攻击时，对全局模型性能的保护能力。

选用了以下评估指标进行多维度的评估：为了考察核心分割性能，首先选择 Dice 相似系数(Dice Coefficient)作为核心指标，同时补充 Hausdorff 距离以量化分割边界的吻合度。为了评估收敛性与临床偏好，通过监控损失值(Loss)以及召回率(Recall)以判断模型收敛动态。记录每个联邦通信轮次的端到端完成时间(秒)，以评估框架在引入安全机制后的实用性。

4.3. 测试环境与配置

所有实验在配备 NVIDIA GeForce RTX 4060 Laptop GPU 的硬件平台上进行。框架基于 PyTorch 实现，关键隐私保护组件使用 tenseal 库执行 CKKS 全同态加密操作，以支持联邦学习中的安全聚合。

在优化策略方面，为提高在非独立同分布 non-IID 数据环境下的收敛稳定性，本文采用 AdamW 优化器，初始学习率设为1×10^{-4}，权重衰减系数为1×10^{-4}。同时，引入“预热 + 余弦退火重启”的学习率调度策略：前 5 个 epoch 为线性预热阶段，随后采用周期性余弦退火 $T_0 = 5$、$T_{\mathrm{mult}} = 2$ 动态调整学习率，从而在一定程度上缓解客户端漂移问题，提升全局模型收敛的稳定性。

在本地训练过程中，每个客户端执行 9 个 epoch，并采用混合精度训练以提高计算效率，同时引入最大范数为 1.0 的梯度裁剪以增强数值稳定性。

在损失函数设计上，本文采用分阶段组合损失策略，以兼顾收敛速度与分割精度：

① 在训练初期，仅使用 Dice Loss，以加快模型收敛速度并稳定训练过程；

② 在训练中期，采用 Dice Loss 与二元交叉熵损失(Binary Cross-Entropy, BCE)的加权组合：

$$\mathcal{L} = 0.7\mathcal{L}\mathrm{Dice} + 0.3\mathcal{L}\mathrm{BCE}$$

③ 在训练后期，进一步调整权重，使两者占比相同：

$$\mathcal{L} = 0.5\mathcal{L}\mathrm{Dice} + 0.5\mathcal{L}\mathrm{BCE}$$

该分阶段策略能够在训练早期强化区域重叠学习，在后期提升像素级判别能力，从而在一定程度上缓解类别不平衡问题。

联邦学习共进行 100 轮通信，并采用基于 Dice 系数与 Recall 的加权指标作为模型选择标准。具体评分函数定义为：

$$\mathrm{Score} = 0.6\times\mathrm{Dice} + 0.4\times\mathrm{Recall}$$

同时设置早停机制(patience = 7)，当连续多轮未取得性能提升时提前终止训练，以避免过拟合并降低计算开销。

5. 模型结果分析

5.1. 模型有效性与收敛性分析

在高度非独立同分布(non-IID)数据场景下，确保全局模型的稳定收敛是验证该方案有效性的核心目标。图 3 通过 25 轮联邦训练过程的实验结果，呈现了准确率、损失函数等关键指标的变化规律。

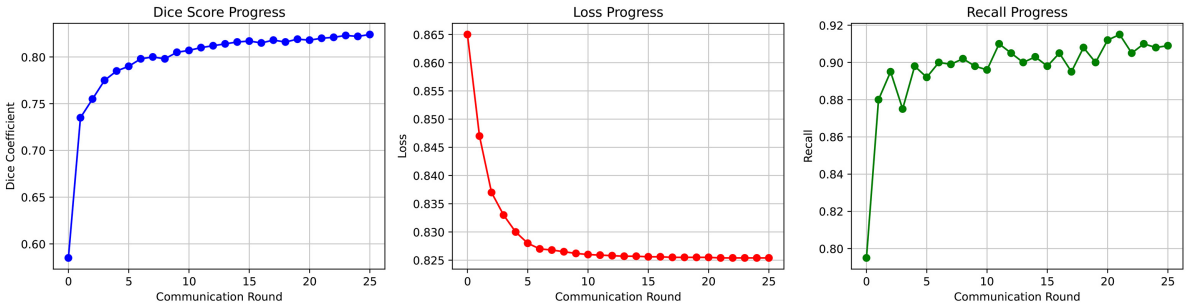

Figure 3. Model's federated training convergence performance: (a) dice coefficient, (b) loss value, (c) recall rate
图 3. 模型的联邦训练性能收敛曲线：(a) Dice 系数，(b) 损失值，(c) 召回率

观察 Dice 系数曲线可以看出(图 3(a))，模型性能从初始的约 0.58 开始，在前 10 轮中实现了快速提升，随后增速放缓，在 25 轮训练结束时稳定在接近 0.83 的水平。相应地，损失值(图 3(b))也从高位迅速下降，并在约 15 轮后趋于平稳。这有力地证明，尽管存在由高度非独立同分布数据引起的性能振荡，本框架依然能够有效地聚合全局模型向正确方向收敛。

尤为关键的是召回率曲线(图 3(c))。在自适应损失函数的策略驱动下，召回率在训练最初的几个轮次就迅速攀升并稳定在 0.90 以上的高位。这一现象可以证明该训练框架能在追求高分割精度的同时，优先确保模型最大程度地避免漏诊潜在的阳性病灶，从而满足医疗诊断场景的核心安全需求。

5.2. 系统性能与精度评估

在完成 25 轮联邦学习后，选取在验证集上综合性能(以 Dice 系数为主要参考)最佳的全局模型进行最终评估。该模型在独立测试集上的像素级分割性能，由图 4 的混淆矩阵及表 1 的分类报告进行了全面的量化展示。

Table 1. Final model test set classification report (including HD95)
表 1. 最终模型测试集分类报告(含 HD95)

	Precision	Recall	FL-Score	HD95 (mm)	Support
No Lesion (0)	0.9982	0.9966	0.9974	-	26475292
Lesion (1)	0.7959	0.8812	0.8365	12.34	394468
accuracy			0.9949		26869760
macro avg	0.8971	0.9389	0.9175	-	26869760
weighted avg	0.9953	0.9949	0.9950	-	26869760

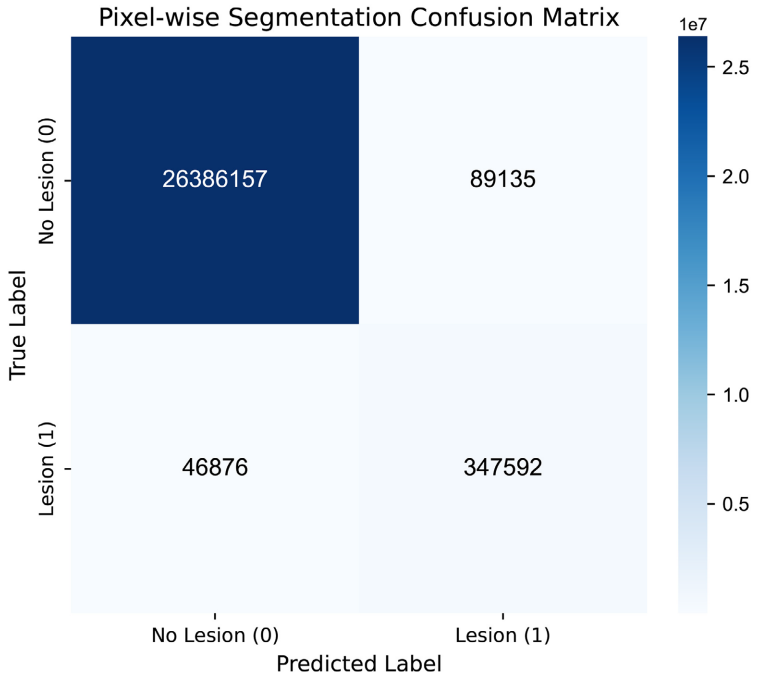

Figure 4. Confusion matrix for pixel-wise segmentation of the final model on the test set
图 4. 最终模型测试集像素级分割混淆矩阵

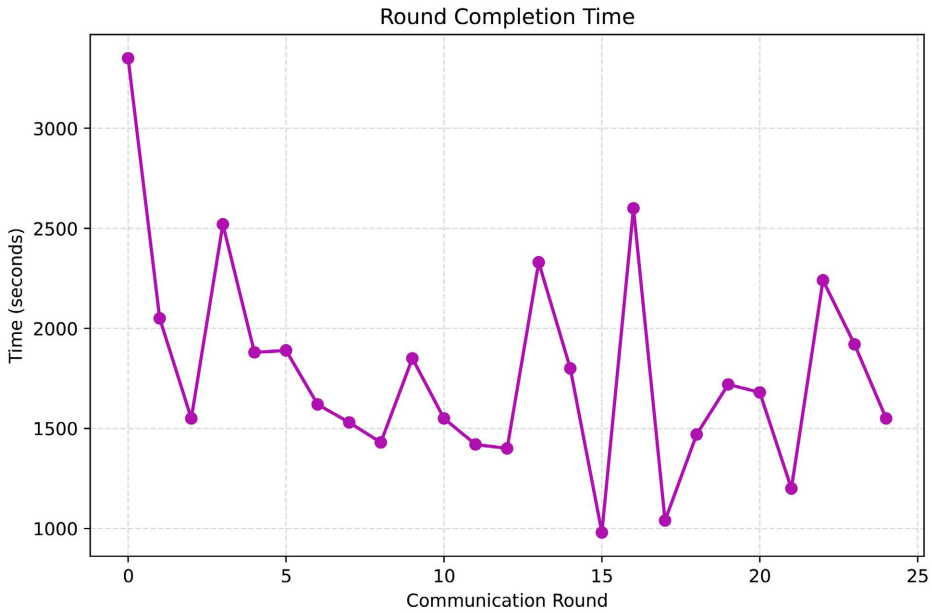

Figure 5. End-to-end completion time per federated communication round (affected by cryptographic computation and network latency)
图 5. 各联邦通信轮次端到端完成时间(受加密计算与网络延迟影响)

分析结果显示，模型在像素级总体准确率上达到了 0.9949。针对更具临床价值的病灶(Lesion, 1)类别，模型实现了高达 0.881 的召回率(Recall)，同时精确率(Precision)和综合 Dice 系数(FL-Score)也分别达到了 0.796 和 0.837。这一系列指标证明了模型成功学习到了“宁可多标记一些可疑区域，也绝不漏掉一个真正病灶”的临床安全准则。其次，边界精度的 HD95 指标为 12.34 mm，这一数值反映了模型能够尽可能

全面地覆盖所有潜在的病灶区域。

如图 5 所示，集成了全同态加密的联邦轮次，其单轮通信时间在 1000 秒至 3300 秒之间。这种计算开销是当前同态加密技术在提供顶级数据隐私性时不可避免的问题。尽管如此，本模型框架的成功运行证明了在复杂的人工智能训练任务中应用此类强隐私技术的可行性。

5.3. 定性结果分析

为了更直观地理解模型的行为和性能，图 6 展示了最终模型在三个独立测试样本上的分割效果。图中每一行分别展示了原始 CT 影像、人工标注的真实病灶(GroundTruth)、模型不同深度侧输出的激活图，以及最终分割结果与真实病灶的叠加图。

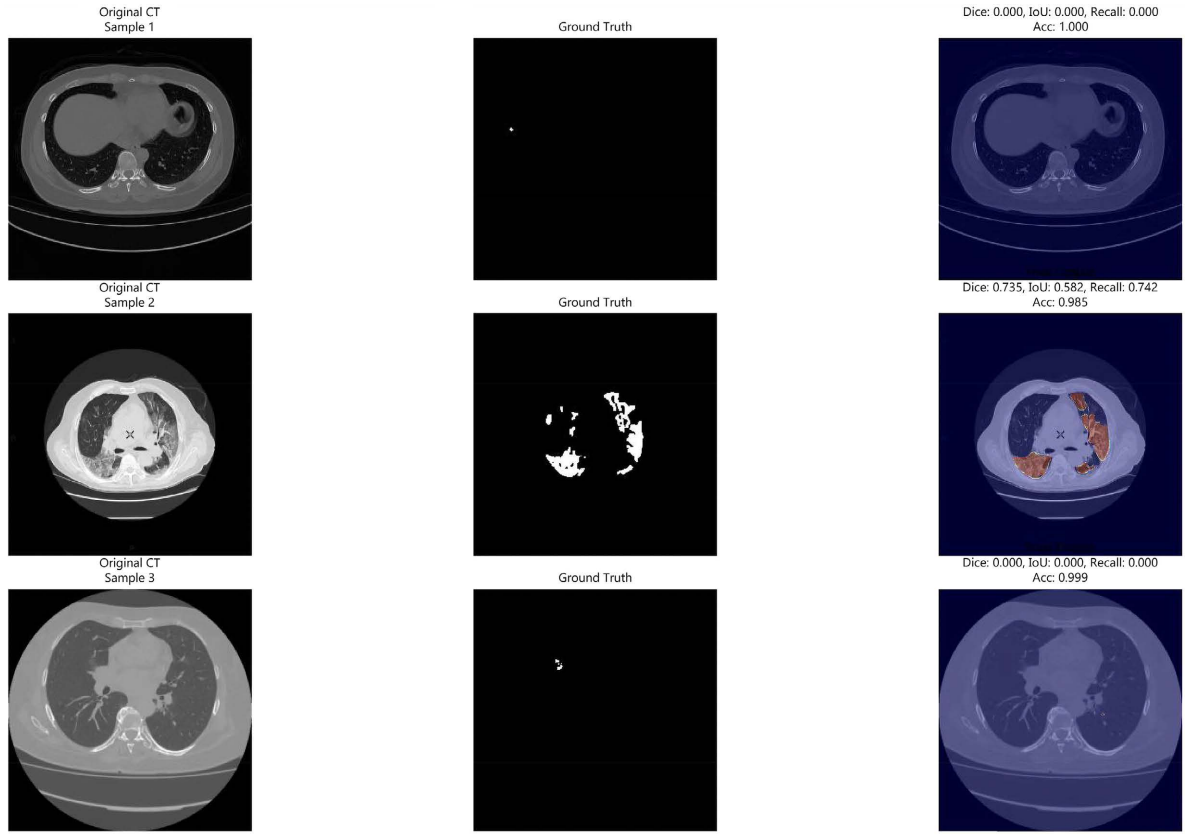

Figure 6. Visualization of test sample segmentation results: (top) Sample 1, (middle) Sample 2, (bottom) Sample 3
图 6. 测试样本分割结果可视化：(上) 样本 1，(中) 样本 2，(下) 样本 3

从定性结果看，样本 2 的分割效果最为理想，最终输出的轮廓(蓝色)与真实病灶(红色)高度重合，展示了模型在常规样本上的优异性能。更有意义的是，样本 1 和样本 3 这类“困难样本”，它们的共同特点是病灶区域极小。在这些样本上，尽管最终的分割结果未能完整地圈出病灶，但其多个中间层的监督输出(Side Outputs)已经显示出对这些微小区域微弱但明确的激活响应。这一现象直观地展示了深度监督策略的内在价值，它确保了学习信号能够成功反向传播至网络的浅层，证明了该模型具有学习微小病理特征的能力。

6. 结语

本研究设计了一个新型的安全联邦学习框架，核心思想在于通过引入区块链作为审计平台，在保留

中心化架构的同时重构了信任机制，实现了聚合过程透明化，公开化，可追踪。为了保证初始模型不被泄露，聚合模型参数使用了 CKKS 全同态加密方案进行保护，尽可能地降低了隐私泄露的风险。实验结果表明，本研究的学习框架在严苛的 non-IID 医疗影像任务中取得优异的分割精度，更重要的是它的性能损失控制在 5%以内，展现了卓越的实用性。

总结而言，本研究的主要贡献在于针对联邦学习领域的梯度泄露风险与中心化信任瓶颈，提出了一套理论完备且工程可行的系统性解决方案。通过深度整合密码学技术、分布式系统设计与高效的 AI 训练策略，在实际部署中严格优化，确保了该方案在真实医疗场景下的安全性与实用性，为构建大规模跨机构可信 AI 协作平台提供了可靠技术支撑。

基金项目

2025 年大学生创新创业训练计划项目(S202511058051)。

参考文献

[1] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[2] Ho, Q.R., Cipar, J., Cui, H.G., Lee, S., Kim, J.K., Gibbons, P.B., *et al*. (2013) More Effective Distributed ML via a Stale Synchronous Parallel Parameter Server. *Proceedings of the* 27*th International Conference on Neural Information Processing Systems*, Nevada, 5-10 December 2013, 1223-1231.

[3] Li, T., Sahu, A.K., Talwalkar, A. and Smith, V. (2020) Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine*, **37**, 50-60. https://doi.org/10.1109/msp.2020.2975749

[4] Wei, W., Liu, L,. Loper, M., *et al*. (2020) A Framework for Evaluating Gradient Leakage Attacks in Federated Learning. arXiv: 2004.10397.

[5] Phong, L.T., Aono, Y., Hayashi, T., Wang, L. and Moriai, S. (2018) Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. *IEEE Transactions on Information Forensics and Security*, **13**, 1333-1345. https://doi.org/10.1109/tifs.2017.2787987

[6] Zhu, L., Liu, Z and Han, S. (2019) Deep Leakage from Gradients. arXiv: 1906.08935.

[7] Wei, K., Li, J., Ding, M., Ma, C., Yang, H.H., Farokhi, F., *et al*. (2020) Federated Learning with Differential Privacy: Algorithms and Performance Analysis. *IEEE Transactions on Information Forensics and Security*, **15**, 3454-3469. https://doi.org/10.1109/tifs.2020.2988575

[8] Fang, M., Cao, X., Jia, J. and Gong, N.Z. (2020) Local Model Poisoning Attacks to Byzantine-Robust Federated Learning. 29*th USENIX Security Symposium* (*USENIX Security* 20) 2020, 12-14 August 2020, 1623-1640.

[9] Tolpegin, V., Truex, S., Gursoy, M.E. and Liu, L. (2020) Data Poisoning Attacks against Federated Learning Systems. In: Chen, L., Li, N., Liang, K. and Schneider, S., Eds., *Computer Security—ESORICS* 2020, Springer, 480-501. https://doi.org/10.1007/978-3-030-58951-6_24

[10] Cao, D., Chang, S., Lin, Z., Liu, G. and Sun, D. (2019) Understanding Distributed Poisoning Attack in Federated Learning. 2019 *IEEE* 25*th International Conference on Parallel and Distributed Systems* (*ICPADS*), Tianjin, 4-6 December 2019, 233-239. https://doi.org/10.1109/icpads47876.2019.00042

[11] Blanchard, P., El Mhamdi, E.M., Guerraoui, R., *et al*. (2017) Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 118-128.

[12] 朱建明, 张沁楠, 高胜, 等. 基于区块链的隐私保护可信联邦学习模型[J]. 计算机学报, 2021, 44(12): 2464-2484.

[13] Nguyen, D.C., Ding, M., Pham, Q., Pathirana, P.N., Le, L.B., Seneviratne, A., *et al*. (2021) Federated Learning Meets Blockchain in Edge Computing: Opportunities and Challenges. *IEEE Internet of Things Journal*, **8**, 12806-12825. https://doi.org/10.1109/jiot.2021.3072611

[14] Li, Y., Chen, C., Liu, N., Huang, H., Zheng, Z. and Yan, Q. (2021) A Blockchain-Based Decentralized Federated Learning Framework with Committee Consensus. *IEEE Network*, **35**, 234-241. https://doi.org/10.1109/mnet.011.2000263

[15] 陈学斌, 任志强, 张宏扬. 联邦学习中的安全威胁与防御措施综述[J]. 计算机应用, 2024, 44(6): 1663-1672.

[16] Fredrikson, M., Jha, S. and Ristenpart, T. (2015) Model Inversion Attacks That Exploit Confidence Information and

Basic Countermeasures. *Proceedings of the* 22*nd ACM SIGSAC Conference on Computer and Communications Security*, Denver, 12-16 October 2015, 1322-1333. https://doi.org/10.1145/2810103.2813677

[17] Lu, Y., Huang, X., Dai, Y., Maharjan, S. and Zhang, Y. (2020) Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT. *IEEE Transactions on Industrial Informatics*, **16**, 4177-4186. https://doi.org/10.1109/tii.2019.2942190

[18] Cheon, J.H., Kim, A., Kim, M. and Song, Y. (2017) Homomorphic Encryption for Arithmetic of Approximate Numbers. In: Takagi, T. and Peyrin, T., Eds., *Advances in Cryptology—ASIACRYPT* 2017, Springer, 409-437. https://doi.org/10.1007/978-3-319-70694-8_15

[19] Howard, A.G., Zhu, M.L., Chen, B., Kalenichenko, D., Wang, W.J., Weyan, T., *et al*. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv: 1704.04861.

[20] Yang, L.X., Zhang, R.-Y., Li, L.D. and Xie, X.H. (2021) SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *Proceedings of the* 38*th International Conference on Machine Learning*, 18-24 July 2021, 11863-11874.

[21] Maftouni, M. (2021) COVID-19 CT Scan Lesion Segmentation Dataset. Kaggle. https://www.kaggle.com/datasets/maedemaftouni/covid19-ct-scan-lesion-segmentation-dataset