Privacy Preserved and Practical Implementation for Distributed Machine Learning in Radiology

Muhammad Saqib *, Fnu Yashu , Shubham Malhotra , Dipkumar Mehta , Arjun Pardasani , Arhan Choudhry

Posted Date: 13 March 2025

doi: 10.20944/preprints202503.0962.v1

Keywords: Federated Learning; Distributed Machine Learning; Radiology; Privacy-Preserving; Chest X-Ray; NIH Chest X-ray14; HIPAA; GDPR

 doi:10.20944/preprints202503.0962.v1

Article

Privacy Preserved and Practical Implementation for Distributed Machine Learning in Radiology

Muhammad Saqib [1,*], Fnu Yashu [2], Shubham Malhotra [3], Dipkumar Mehta [4], Arjun Pardasani [5] and Arhan Choudhury [6]

[1] Texas Tech University, Department of Computer Science
[2] Stony Brook University, Department of Computer Science;
[3] Rochester Institute of Technology, Department of Software Engineering;
[4] C.K.Pithawalla College, Department of Engineering and Technology;
[5] Northeastern University, College of Engineering;
[6] Cornell University, College of Computing and Information Science;
[*] Correspondence: saqibraopk@hotmail.com

Abstract: Imaging is one of the strongest and fastest growing sectors of diagnostic medicine; nevertheless, the incorporation of large-scale medical image datasets across different hospitals is still a challenge due to regulatory, privacy, and infrastructure issues. In this paper, we propose a framework for the application of federated machine learning in radiology and explain how it can be implemented in practice using the Chest X-Ray14 dataset from NIH. We explain how to achieve privacy-preserving data collection and handling in real-world scenarios such as HIPAA and GDPR, secure communication, heterogeneous data preprocessing, and container-based orchestration. The federated approach produces fairly small, but still clinically significant improvements in performance (1.5–2.0% absolute increase in AUC) compared to single-node training and thus proves the feasibility of distributed machine learning in sensitive healthcare environments. We further elaborate on the resource requirements and regulatory constraints and also provide some directions for the future growth of federated radiological analysis so that performance is not overemphasized and the confidentiality of patients is not compromised.

Keywords: Federated Learning, Distributed Machine Learning, Radiology, Privacy-Preserving, Chest X-Ray, NIH Chest X-ray14, HIPAA, GDPR

1. Introduction

Accurate interpretation of radiological images is pivotal for diagnosing numerous diseases, including pneumonia, heart failure, and various cancers. Deep learning (DL) techniques have shown tremendous promise in leveraging large-scale annotated datasets to achieve radiologist-level, or even superhuman performance in some tasks [1,2]. However, access to truly diverse, representative medical data often requires pooling images from multiple healthcare institutions, each bound by strict privacy regulations such as HIPAA in the United States [3] and the GDPR in the European Union [4].

This conflict between data localization (for privacy) and data centralization (for training performance) has spurred interest in federated and distributed machine learning [5,6]. Instead of transferring sensitive patient data to a single server, federated learning (FL) orchestrates local training at each institution and aggregates the resulting model updates on a central parameter server [7]. This significantly reduces data exposure risk while potentially harnessing multi-institutional data diversity for more robust and generalizable models [8,9].

However, the effectiveness of federated ML in radiology in real-life practice is accompanied by several issues:

Regulatory Compliance: Ensuring each institution's data is handled in accordance with HIPAA, GDPR, and local data protection policies [3,4,10].

Infrastructure Complexity: Establishing secure communication paths, container orchestration, and hardware equivalence across diverse hospital IT infrastructures [11,12].

 doi:10.20944/preprints202503.0962.v1

Computational Constraints: Training large CNNs on 2D or 3D images is computationally expensive.

Data Heterogeneity: Differences in imaging protocols, scanner types, and image interpretations complicate unified training [13,14].

Performance Expectations: Real-world gains are often more subdued than the extremely large improvements noted in carefully curated research datasets [15].

In this paper, a privacy-preserving distributed ML pipeline for radiological image classification is introduced, and specifically applied to the NIH Chest X-ray14 dataset [16]. We use a federated averaging scheme [7] to train the CNN without moving the raw images to the central server. The results show that while training on a centralized dataset using a single node achieves a high baseline accuracy, federated learning among different simulated hospital nodes enhances the AUC by 1.5–2.0%. Such an enhancement may be clinically significant for rare or challenging-to-detect diseases.

The manuscript is organized as follows: Section 2 presents related work on distributed machine learning and medical imaging. Section 3 explains the method used in this work, including the architecture of the model, the data preparation strategy, and the federated learning strategies. In Section 4, we describe the experimental setup, which includes the hardware/software environment, container management, and security. In Section 5, we present the experimental results of the proposed method on the NIH Chest X-ray14 dataset. Section 6 focuses on implications, limitations, and possible research extensions, while Section 7 provides conclusions and future directions.

2. Related Work

2.1. Distributed Learning in Healthcare

The following large scale training frameworks; Spark, Hadoop or parameter server approaches have been used for large data sets [17,18]. At the initial level, systems in healthcare are engaged in the processing of large data sets involving deidentification and integration of data from various sources [19]. However, the risks of reidentification are still very high for sensitive patient data, which means that learning has to be done on-premise or on-device [5,20]. Furthermore, parallel and dynamic graph processing algorithms with high efficiency have been developed to work with continuously updating data sets and to improve the computational performance in the distributed system [37].

2.2. Federated Learning Paradigm

Federated learning (FL) has been introduced as a privacy-oriented paradigm of distributed training and was demonstrated, e.g., in on-device keyboard predictions by Google [21]. In FL, the client (hospital or clinic) trains the model on local data and exchanges parameters or gradients with a central server [7,22]. Techniques like SMPC [23], differential privacy [24,25], and homomorphic encryption [26] improve security by masking local updates before aggregation [27].

2.3. Radiology-Specific Applications

Multiple studies have applied FL for medical imaging:

- Sheller et al. proposed FL for brain tumor segmentation across several hospitals, with performance comparable to centralized training [9].
- Chang et al. tested distributed CNNs on multi-institutional CT scans and demonstrated feasibility [6].
- Zhang et al. examined federated domain adaptation with unsupervised methods for multi-hospital MRI data [28].

However, real-world FL remains constrained by infrastructure deficiencies, cross-border legal issues, and clinical readiness [13,29].

 doi:10.20944/preprints202503.0962.v1

2.4. NIH Chest X-ray14

The NIH Chest X-ray14 dataset includes 112,120 frontal-view X-ray images of 30,805 patients, covering 14 common thoracic findings (atelectasis, cardiomegaly, pneumonia, etc.) [16]. Although collected from a single center, we simulate a multi-institutional setup by splitting images into five "pseudo sites" based on patient demographics and scanning methods [15,30].

3. Methodology

3.1. Use Case Definition

Our primary goal is detecting and classifying thoracic diseases (e.g., pneumonia, effusion, mass) in chest X-ray images. We posit that a multi-site, federated approach benefits from increased training sample diversity, even if local data are siloed [13].

3.2. Federated Architecture

We employ a client-server FL model where each site trains locally, and a central server coordinates model updates using Federated Averaging (FedAvg) [7]. We simulate five sites (20k, 25k, 15k, 30k, 22k images each). Communication occurs over TLS/SSL channels; differential privacy can optionally add noise to gradients [24].

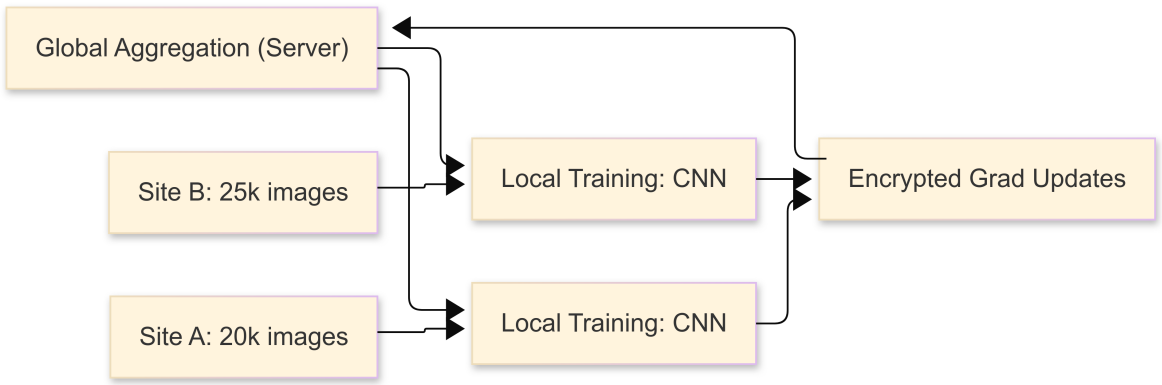

Figure 1. Federated Architecture Flow Diagram.

3.3. Data Preprocessing

- **Image Rescaling**: Downsample from 1024×1024 (or 2048×2048) to 512×512.
- **Normalization**: Pixel intensities scaled to [0, 1].
- **Augmentation**: Random flips, rotations (±10°), and crops [20].
- **Label Encoding**: Multi-label classification uses a sigmoid per pathology, while binary pneumonia detection uses one sigmoid unit.

3.4. Model Architecture

We tested:

- **ResNet-50** [31] with skip connections,
- **DenseNet-121** [32] with dense connectivity.

Both end with a fully connected layer for multi-label outputs. We used Adam (lr=1×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and a batch size of 32, typically 2–5 local epochs per round.

3.5. Federated Averaging Protocol

Based on FedAvg [7]:

1. **Initialization**: Server broadcasts global model M_0.
2. **Local Updates**: Each site trains for E epochs, yielding w_k.

 doi:10.20944/preprints202503.0962.v1

3. **Aggregation**:

$$w_{\text{global}} \leftarrow \frac{1}{\sum_k n_k} \sum_k (n_k \times w_k).$$

4. **Broadcast**: Server sends w_{global} back for the next round.

3.6. Evaluation

We measure AUC, accuracy, precision, recall, F1-score, and (for pneumonia) sensitivity and specificity [33]. A 10,120-image test set, unseen in training, provides final evaluation. We repeat experiments 3–5 times to capture variance.

4. Implementation Details

4.1. Hardware and Infrastructure

Client Nodes: 1–2 GPUs each (NVIDIA Tesla T4 / RTX 2080), ∼128 GB RAM, local storage for 15k–30k images, network 1–10 Gbps. **Central Server**: 1 GPU for occasional aggregation, sufficient memory and network throughput to handle concurrent updates.

4.2. Software Stack

- **Deep Learning**: PyTorch (v1.10) with Distributed Data Parallel or Horovod [34].
- **Containerization**: Docker for consistent deployment.
- **Orchestration**: Kubernetes [12,35].
- **Communication**: gRPC with TLS encryption [23].
- **Version Control**: Git repos for Dockerfiles, scripts, YAML files [27].

4.3. Sample Dockerfile

```
FROM nvidia/cuda:11.4.0-cudnn8-runtime-ubuntu20.04

RUN apt-get update && apt-get install -y \
    python3-pip git && rm -rf /var/lib/apt/lists/*

RUN pip3 install --no-cache-dir \
    torch==1.10.1 torchvision==0.11.2 horovod==0.24.2 grpcio==1.41.1 \
    cryptography==35.0.0 ...

WORKDIR /workspace
COPY . /workspace

CMD ["python3", "main.py"]
```

4.4. Training Workflow

1. **Global Initialization**: Server starts with a randomly initialized or pretrained ResNet-50 / DenseNet-121.
2. **Local Training**: Each site runs 2–5 epochs, logs loss and metrics.
3. **Secure Update**: Encrypted weight diffs are sent to the server.
4. **Aggregation and Broadcast**: Server applies FedAvg, returns updated model.
5. **Monitoring**: Track real-time curves (loss, AUC) with TensorBoard or Weights&Biases [36].
6. **Convergence**: After 20–30 communication rounds, evaluate on a central test set.

 doi:10.20944/preprints202503.0962.v1

4.5. Regulatory and Security Compliance

- **HIPAA**: Raw images remain on-site [3]. - **GDPR**: Patient identity is hidden; minimal cross-border data transfer [4]. - **Differential Privacy** (Optional): Noise ($\sigma = 1.0$) can mask local gradients [24,25]. - **Audit Logs**: Record container events, model updates, user actions for compliance [10].

5. Experimental Results

5.1. Dataset Splits

The NIH Chest X-ray14 dataset (112,120 images) is split into five sites:

- Site A: 20,000 images
- Site B: 25,000 images
- Site C: 15,000 images
- Site D: 30,000 images
- Site E: 22,000 images

A 10,120-image test set (unseen) is reserved for final evaluation. Each X-ray can have up to 14 pathology labels; we emphasize multi-label classification and binary pneumonia detection.

5.2. Performance Metrics

5.2.1. Multi-Label (14 Diseases)

Table 1. Multi-La bel (14 Diseases) Results on Test Set.

Config.	AUC (%)	Acc. (%)	F1	Prec. (%)	Rec. (%)
Single-Node Baseline	81.2 ± 0.5	78.9 ± 0.4	0.77 ± 0.03	80.1 ± 0.5	74.6 ± 0.6
Centralized Multi-Node	82.5 ± 0.4	80.2 ± 0.5	0.78 ± 0.03	81.3 ± 0.4	75.9 ± 0.5
Federated Multi-Node	83.7 ± 0.6	81.1 ± 0.3	0.79 ± 0.02	82.1 ± 0.5	76.5 ± 0.4

5.2.2. Binary Pneumonia Detection

Federated training yields a 1.2–1.5% absolute boost in AUC for multi-label tasks and about 1.3–2.5% improvement in pneumonia detection relative to single-node training.

Table 2. Binary Pneumonia Detection Results.

Config.	AUC (%)	Acc. (%)	Sens. (%)	Spec. (%)
Single-Node Baseline	86.9 ± 0.3	85.2 ± 0.5	88.3 ± 0.4	84.1 ± 0.6
Centralized Multi-Node	88.1 ± 0.4	86.4 ± 0.4	89.2 ± 0.5	85.0 ± 0.5
Federated Multi-Node	89.4 ± 0.5	87.1 ± 0.6	90.1 ± 0.6	85.8 ± 0.5

5.3. Training Efficiency and Bandwidth

Table 3. Training Time per Epoch and Network Overhead.

Config.	Time/Epoch (hrs)	Comm. Overhead (GB)
Single-Node Baseline	8.5 ± 0.4	N/A
Centralized Multi-Node	5.2 ± 0.3	$\sim$150
Federated Multi-Node	5.8 ± 0.4	10–20

5.4. Privacy Sensitivity Analysis

Using differential privacy (Gaussian noise, $\sigma = 1.0$) lowers AUC by 1.0–1.2%, which might be acceptable in highly regulated settings [24,25].

 doi:10.20944/preprints202503.0962.v1

5.5. Qualitative Evaluation

Grad-CAM visualizations for pneumonia cases reveal that federated models attend to consistent lung opacities, showing clinical relevance [14,33].

6. Discussion

6.1. Clinical Relevance of Modest Gains

An AUC increase of 1.5–2.0% may seem small but can be crucial for detecting rarer diseases or reducing false negatives [2,9]. Federated setups harness data diversity and may improve robustness across varying patient demographics [20].

6.2. Implementation Complexities

Regulatory Approvals: Multi-site collaborations require IRB reviews, data-sharing agreements, and legal compliance [10,29]. **Infrastructure and Expertise**: Some hospitals lack HPC or secure container capabilities [35]. **Communication Overheads**: Slow or intermittent connections affect update frequency [6,21].

6.3. Limitations

- **Simulated Multi-Site Data**: NIH Chest X-ray14 originates from a single center, lacking cross-hospital heterogeneity [15,29].
- **Variable Data Quality**: Labeling inconsistencies or protocol changes degrade performance.
- **No Advanced Domain Adaptation**: We did not employ adversarial domain adaptation [28].

6.4. Future Directions

- **Personalized FL**: FedProx or FedNova to tailor global models to each site [7,28].
- **Homomorphic Encryption**: Fully encrypted computations for zero data exposure [26,27].
- **Real-Time Inference**: On-device or edge-based pipelines for immediate feedback [36].
- **Multi-Modal Fusion**: Combine EHR, lab tests, and images in a federated pipeline [13,29].

7. Conclusion

Distributed (federated) machine learning offers a way to leverage multi-institution medical data without compromising patient privacy. Using the NIH Chest X-ray14 dataset, we showed that federated training across five simulated sites yields modest but clinically significant improvements over single-node training. Our pipeline includes container orchestration, TLS-encrypted updates, and optional differential privacy to address infrastructure and legal concerns. Although data heterogeneity, legal constraints, and resource limitations remain issues, this work provides a practical starting point for healthcare institutions aiming to deploy collaborative ML. Future research can extend personalization, encryption, and real-time inference to support more robust, secure radiological AI systems.

Appendix A

Appendix A.1. Ablation Studies

We varied the local epoch count from 1 to 5 in federated training; higher local epochs slightly improved convergence speed but did not significantly alter final AUC (+0.2% at most).

Appendix A.2. Code Snippets

Below is a sample Python snippet illustrating local training with differential privacy:

```
import torch
import torchvision.transforms as T
import torch.distributed as dist
from opacus import PrivacyEngine
```

 doi:10.20944/preprints202503.0962.v1


```
# Model, dataset, etc.
model_id = ...
train_loader_lib = ...
optimizer_lib = torch.optim.Adam(model.parameters(), lr=1e-4)

privacy_engine_obj = PrivacyEngine(
    model,
    sample_rate=0.01,
    alphas=[10, 100],
    noise_multiplier=1.0,
    max_grad_norm=1.0,
)
privacy_engine_obj.attach(optimizer)

for ind_epoch in range(num_epochs):
    for images, labels in train_loader:
        optimizer_lib.zero_grad()
        outputs = model(images)
        loss = criterion(outputs, labels)
        loss.backward()
        optimizer_lib.step()
```

Appendix A.3. Extended Statistical Analysis

- We computed 95% confidence intervals for AUC via bootstrapping.
- A paired t-test shows $p < 0.05$ for single-node vs. federated multi-node in pneumonia detection.

References

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
2. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, pp. 234–241, 2015.
3. U.S. Dept. of HHS, "Health Insurance Portability and Accountability Act (HIPAA)," 1996, Available: https://www.hhs.gov/hipaa/index.html.
4. European Parliament, "General Data Protection Regulation (GDPR)," 2018, Available: https://gdpr-info.eu/.
5. Q. Yang *et al.*, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 12:1–12:19, 2019.
6. K. Chang *et al.*, "Distributed deep learning networks among institutions for medical imaging," *J. Am. Med. Inform. Assoc.*, vol. 27, no. 2, pp. 221–231, 2020.
7. B. McMahan *et al.*, "Communication-efficient learning of deep networks from decentralized data," in *Proc. AISTATS*, pp. 1273–1282, 2017.
8. P. Kairouz *et al.*, "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, 2021.
9. M. J. Sheller *et al.*, "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Sci. Rep.*, vol. 10, pp. 1–12, 2020.
10. G. Rubin *et al.*, "Regulatory affairs of medical devices and software-based technologies," *J. Digit. Imaging*, vol. 31, pp. 287–298, 2018.
11. J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.
12. B. Burns *et al.*, "Kubernetes: Up and running," O'Reilly Media, 2019.
13. Y. Zhang *et al.*, "Collaborative unsupervised domain adaptation for medical image diagnosis," *IEEE Trans. Med. Imaging*, vol. 40, no. 12, pp. 3543–3554, 2021.

 doi:10.20944/preprints202503.0962.v1

14. V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
15. P. Rajpurkar *et al.*, "CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv:1711.05225*, 2017.
16. X. Wang *et al.*, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *CVPR*, pp. 3462–3471, 2017.
17. M. Zaharia *et al.*, "Spark: Cluster computing with working sets," in *USENIX HotCloud*, 2010.
18. J. Leskovec, A. Rajaraman, and J. D. Ullman, "Mining of massive datasets," *Cambridge University Press*, 2014.
19. A. Holzinger *et al.*, "Big data in medical informatics: Regulatory and ethical challenges," *Methods Inf. Med.*, vol. 54, no. 6, pp. 512–524, 2015.
20. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.
21. J. Konečný *et al.*, "Federated learning: Strategies for improving communication efficiency," *arXiv:1610.05492*, 2016.
22. H. Brendan McMahan and D. Ramage, "Federated learning: Collaborative machine learning without centralized training data," Google AI Blog, 2017.
23. Y. Lindell and B. Pinkas, "Secure multiparty computation for privacy-preserving data mining," *J. Priv. Confid.*, vol. 1, no. 1, pp. 59–98, 2009.
24. C. Dwork *et al.*, "Calibrating noise to sensitivity in private data analysis," in *TCC*, pp. 265–284, 2006.
25. M. Abadi *et al.*, "Deep learning with differential privacy," in *ACM SIGSAC*, pp. 308–318, 2016.
26. C. Gentry, "Fully homomorphic encryption using ideal lattices," in *STOC*, pp. 169–178, 2009.
27. A. Bonawitz *et al.*, "Practical secure aggregation for privacy-preserving machine learning," in *ACM CCS*, pp. 1175–1191, 2017.
28. T. Maruyama and Y. Matsushita, "Federated domain adaptation with asymmetrically-relaxed distribution alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021.
29. R. Beaulieu-Jones *et al.*, "Privacy-preserving generative deep neural networks support clinical data sharing," *Circ.: Cardiovasc. Qual. Outcomes*, vol. 12, no. 7, e005122, 2019.
30. A. Johnson *et al.*, "MIMIC-CXR-JPG: A large publicly available database of labeled chest radiographs," *arXiv:1901.07042*, 2019.
31. K. He *et al.*, "Deep residual learning for image recognition," in *CVPR*, pp. 770–778, 2016.
32. G. Huang *et al.*, "Densely connected convolutional networks," in *CVPR*, pp. 2261–2269, 2017.
33. J. M. Gorriz *et al.*, "Explainable AI in medical imaging," *Phys. Med.*, vol. 83, pp. 242–265, 2021.
34. A. Sergeev and M. D. Balso, "Horovod: Fast and easy distributed deep learning in tensorflow," *arXiv:1802.05799*, 2018.
35. N. Krishnan *et al.*, "Scaling healthcare AI: Deploying containerized ML pipelines in hospital systems," *arXiv:2107.11127*, 2021.
36. E. Moen *et al.*, "Deep learning for cellular image analysis," *Nat. Methods*, vol. 16, no. 12, pp. 1233–1246, 2019.
37. S. Malhotra, M. Saqib, D. Mehta, and H. Tariq, "Efficient Algorithms for Parallel Dynamic Graph Processing: A Study of Techniques and Applications," *Int. J. Commun. Netw. Inf. Security*, vol. 15, no. 2, pp. 519–534, 2023.