

Privacy-Preserving Medical Image Classification Through Secure Neural Inference and Communication-Efficient Computation

Theres Peukert

Department of Computer Science, University of New Hampshire, Durham, NH, USA.
theres.work@unh.edu

Conan Cairn

Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
cairn1992@buffalo.edu

Kaitlyn Roy

Department of Computer Science, George Mason University, Fairfax, VA, USA.
roy1981@gmu.edu

Abstract

Medical image classification has become a central application of modern artificial intelligence, yet its clinical value depends on access to sensitive patient data that cannot be freely centralized or exposed. This paper examines privacy-preserving medical image classification as a socio-technical infrastructure problem rather than a narrow cryptographic or machine learning task. It analyzes secure neural inference, communication-efficient federated computation, and the institutional conditions required for trustworthy deployment. The discussion emphasizes structural trade-offs among privacy guarantees, model utility, latency, bandwidth, robustness, fairness, and long-term sustainability. Secure inference architectures can protect patient images during prediction, but their practical viability depends on activation costs, communication rounds, hardware acceleration, and protocol composition. Federated approaches reduce central data pooling but introduce issues of heterogeneity, straggler behavior, poisoning risk, and governance complexity. The paper argues that privacy-preserving medical imaging systems require layered defenses, careful workload partitioning, transparent audit trails, and policy frameworks that assign accountability across hospitals, cloud providers, model developers, and regulators. It also considers cross-domain lessons from finance, critical infrastructure, and public health surveillance. The conclusion is that no single mechanism suffices; instead, resilient medical imaging requires an architecture of complementary protections, measurable service levels, and governance that treats privacy as an operational property rather than a compliance formality. Such an approach can support equitable, robust, and sustainable clinical machine learning without collapsing under communication, trust, or regulatory constraints.

Keywords

secure neural inference; medical image classification; privacy-preserving machine learning; communication-efficient computation; federated learning; clinical governance; robustness; trustworthy infrastructure.

1. Introduction

Medical image classification sits at the intersection of high-stakes clinical decision-making, large-scale computation, and deeply sensitive personal data. A model that identifies pathology in radiology, dermatology, ophthalmology, or pathology images can improve access and consistency, but it also creates new pathways for privacy leakage, institutional dependency, and opaque clinical influence. The central problem is not merely whether a neural network can classify an image accurately in a laboratory setting. The problem is whether such classification can be performed within a system that respects patient confidentiality, survives operational stress, distributes benefits fairly, and remains accountable over time. This paper therefore treats privacy-preserving medical image classification as an infrastructure question, one that combines secure neural inference with communication-efficient computation and governance mechanisms that can operate across hospitals, clinics, research consortia, and cloud environments.

Early work on privacy attacks against machine learning demonstrated that models can leak information about their training data through membership inference and model inversion [1], [2]. In medical imaging, such leakage is especially consequential because images may reveal diagnoses, genetic conditions, pregnancy status, or other information that patients expect to remain protected. Federated learning emerged as one response by allowing institutions to train shared models without moving raw data to a central server [3], [4]. Subsequent research clarified both the promise and the unresolved problems of federated learning, including non-identical data distributions, communication bottlenecks, security against adversarial participants, and the difficulty of auditing global models [5]. Secure aggregation further strengthened federated systems by allowing updates to be combined without revealing individual contributions [6]. However, these advances also revealed a fundamental lesson: privacy is not a single binary property but a layered, probabilistic, and institutionally negotiated condition.

The argument developed here is that privacy-preserving medical image classification requires a compositional architecture. Secure neural inference protects prediction-time image data, communication-efficient computation reduces the cost of distributed training and coordination, and governance determines whether these technical protections are meaningful in practice. A system that encrypts every image but concentrates trust in one vendor may fail clinically and politically. A system that distributes training across many hospitals but ignores bandwidth, fairness, or model poisoning may fail operationally. A system that achieves strong cryptographic guarantees but imposes unacceptable latency in emergency radiology may not be adopted. The paper therefore examines secure inference, federated coordination, robustness, fairness, deployment, sustainability, and policy as interconnected layers.

2. Medical Imaging, Privacy Risk, and System Requirements

Medical images are not ordinary data objects. They are high-dimensional, often accompanied by metadata, clinical notes, and longitudinal histories, and they are generated within relationships of care. Privacy risk in this domain is therefore both technical and relational. Technical risk includes reconstruction attacks, membership inference, gradient leakage, and unauthorized secondary use. Relational risk includes loss of patient trust, stigmatization, insurance discrimination, and the erosion of willingness to participate in research. A privacy-preserving system must address both. Cryptographic protection without governance can produce a formally secure pipeline that still violates expectations because data are used for purposes patients did not anticipate. Conversely, policy restrictions without technical safeguards can leave data exposed to insider error, insecure transfer, or vulnerable storage.

The system requirements for medical image classification differ substantially from those of consumer recommendation systems. First, latency may be clinically constrained. A triage model for stroke or trauma cannot wait several minutes for a complex secure computation if immediate decisions are needed. Second, availability matters. A privacy mechanism that fails during peak hospital load may be bypassed under pressure, undermining its long-term credibility. Third, data heterogeneity is extreme. Different scanners, protocols, patient populations, and labeling practices produce distributions that can degrade model performance and fairness. Fourth, accountability must be auditable. Clinicians, regulators, and patients need to understand who contributed data, how a model was validated, and what recourse exists when harm occurs. These requirements point toward architectures that combine secure inference, federated computation, differential privacy, secure aggregation, and human oversight rather than relying on one technique.

3. Secure Neural Inference Architectures

Secure neural inference seeks to let a model owner or service provider evaluate a neural network on a patient image without seeing the image, while the data holder learns the prediction without necessarily learning the model weights. Early demonstrations showed that neural networks could be evaluated over encrypted data with meaningful throughput and accuracy, although the computational cost was substantial [7]. Secure multiparty computation then expanded the design space by allowing multiple parties to jointly compute predictions while keeping inputs private [8]. Transformation-based approaches made neural network inference more practical by reducing the number of expensive operations and adapting network layers to secure computation constraints [9]. Low-latency frameworks further optimized the interaction between cryptographic primitives and neural network structure, showing that system-level engineering is as important as cryptographic theory [10].

More recent secure inference services have focused on practical deployment concerns such as preprocessing, network architecture, and client-server interaction [11]. Two-party protocols have become especially important because they can offer a workable balance between trust assumptions and performance [12]. Frameworks that generalize secure deep learning have made it easier to experiment with different layers, activation functions, and privacy budgets [13]. Federated learning libraries have also incorporated secure computation and privacy mechanisms, enabling researchers and hospitals to prototype distributed workflows without building every component from scratch [14]. In medical imaging specifically, reviews have emphasized that secure, privacy-preserving, and federated machine learning must be evaluated against clinical workflows, regulatory expectations, and data heterogeneity rather than only against cryptographic benchmarks [15].

A persistent bottleneck in secure neural inference is the cost of nonlinear activation functions. Many secure protocols handle linear operations efficiently, but nonlinearities require approximations or interactive subprotocols that increase communication and latency. Recent work on reducing activation cost in two-party private inference illustrates this structural tension by redesigning neural network components to reduce the burden of nonlinear computation [16]. This line of work is important because it shows that privacy-preserving inference is not only a cryptographic problem; it is also a neural architecture problem. The model itself can be redesigned to fit the secure computation substrate. However, such redesign may affect accuracy, robustness, and generalizability. A network optimized for secure inference may behave differently on rare pathologies or underrepresented populations.

Therefore, secure inference must be evaluated with clinical validation metrics that go beyond aggregate accuracy.

4. Communication-Efficient Computation and Federated Coordination

Federated learning offers a complementary approach by training models across institutions without centralizing raw images. In medicine, multi-institutional federated studies have demonstrated that models can be trained collaboratively while preserving local data control [17]. Yet federated learning introduces communication as a first-class constraint. Each round requires model updates to be transmitted, aggregated, and redistributed. In large imaging models, these updates can be enormous, and hospitals may have asymmetric bandwidth, intermittent connectivity, or strict firewall policies. Communication-efficient methods such as compression, sparsification, asynchronous aggregation, and hierarchical coordination can reduce cost, but they also affect convergence, fairness, and security [18]. The design challenge is to reduce communication without creating new vulnerabilities or degrading performance for smaller or atypical institutions.

Privacy in federated medical imaging also cannot rely solely on data locality. Updates can leak information, and adversaries may infer patient characteristics from gradients or model parameters. Secure aggregation protects individual updates during combination [6], while differential privacy adds noise to limit what can be inferred from the trained model [19], [20]. Deep learning with differential privacy has shown that meaningful privacy guarantees can be integrated into training workflows, although there is a utility cost that must be managed carefully [21]. In medical settings, the privacy budget must be interpreted clinically and ethically, not merely mathematically. A model that is private under a formal definition may still be unacceptable if it performs poorly on rare diseases or if the noise disproportionately affects minority populations. Communication efficiency and privacy are therefore intertwined: compressing updates may reduce bandwidth but can also alter the noise structure and leakage profile.

Coordination across institutions also raises governance questions. Who decides which hospitals participate? How are data contributions recognized? What happens when a site wants to withdraw? How are disputes resolved when a model performs differently across sites? These questions cannot be answered by protocol design alone. They require agreements, audit mechanisms, and representation for affected communities. A communication-efficient system that excludes rural or under-resourced clinics because they cannot meet bandwidth requirements may inadvertently centralize medical AI in wealthier institutions. Sustainable federation therefore requires tiered participation models, where sites can contribute at different levels of intensity without being excluded from benefits. This is both a fairness issue and a robustness issue, because diverse participation improves model generalization.

5. Trade-offs in Robustness, Fairness, and Clinical Governance

Robustness in privacy-preserving medical imaging includes resistance to adversarial inputs, distribution shift, poisoning, and operational failure. An adversary might manipulate an image with imperceptible perturbations, corrupt local training data, or exploit protocol implementation errors. A system that is cryptographically secure but not robust to distribution shift may still cause clinical harm. Conversely, a robust model that leaks patient data may violate legal and ethical obligations. The practical goal is not perfect security but layered resilience: secure inference protects prediction-time confidentiality, federated learning reduces central data concentration, differential privacy limits inference from models, and

monitoring detects anomalies. Governance then determines how these layers are maintained, updated, and audited.

Fairness is equally central. Medical image datasets often underrepresent certain skin tones, age groups, genders, geographies, and rare conditions. Secure computation does not correct these imbalances; it may even obscure them if privacy constraints limit access to subgroup data for validation. Federated learning can improve representation by including more sites, but only if the aggregation process does not allow large sites to dominate. Differential privacy can protect individuals but may impose higher utility costs on small subgroups. Clinical governance must therefore require disaggregated evaluation, transparent reporting, and mechanisms for affected communities to contest model behavior. High-performance medicine depends on trust, and trust depends on demonstrable fairness rather than technical novelty alone [22]. Ethical implementation in health care requires attention to consent, equity, liability, and the clinician-patient relationship [23]. Machine learning in medicine must be integrated with clinical judgment rather than positioned as a replacement for it [24]. Principles alone cannot guarantee ethical artificial intelligence; they must be operationalized through institutions, incentives, and enforcement [25].

6. Deployment, Infrastructure, and Sustainability

Deployment of privacy-preserving medical image classification requires more than a trained model and a secure protocol. It requires identity management, key management, secure enclaves or accelerator support, network provisioning, logging, incident response, and model lifecycle management. Hospitals may run inference on premises, in a private cloud, or through a regional consortium. Each choice changes the trust model. On-premises deployment gives the hospital more control but may be difficult to maintain. Cloud deployment offers elasticity but concentrates risk and creates dependency. Consortium deployment can share costs but requires durable governance. A practical architecture may use hybrid patterns: sensitive preprocessing locally, secure inference in a trusted execution environment, and federated coordination through a regional broker. The system must degrade gracefully when a component fails, because clinical operations cannot simply stop when a cryptographic protocol times out.

Sustainability is often overlooked in privacy-preserving machine learning. Secure computation and federated training can consume significant energy and hardware resources. Large imaging models require storage, memory bandwidth, and accelerators. If every hospital must maintain expensive infrastructure, small clinics may be excluded. If every inference requires multi-round communication, operational costs may become prohibitive. Sustainable design therefore favors workload partitioning, model distillation, quantization, caching of non-sensitive computations, and shared regional infrastructure. It also favors architectures that can be audited without retraining from scratch. A system that is theoretically private but economically unsustainable will either be abandoned or used only by wealthy institutions, widening health disparities. Sustainability is thus not merely an environmental concern; it is a condition for equitable access.

7. Policy, Regulation, and Cross-Domain Comparisons

Policy frameworks for medical AI must address privacy, safety, liability, and market power. Regulations such as health privacy laws, medical device rules, and data protection regimes create overlapping obligations. A privacy-preserving system may still need to comply with breach notification, audit, and patient access requirements. Cryptographic proofs do not

automatically satisfy regulators, who may ask for documentation, risk assessments, and human oversight. Cross-domain comparisons are instructive. Financial systems have developed secure multiparty computation for fraud detection and collaborative analytics, but they operate with different latency tolerances and regulatory structures. Critical infrastructure uses segmented networks and hardware roots of trust, but medical devices are more diverse and often older. Public health surveillance uses aggregation and differential privacy, but medical imaging involves richer data and stronger individual identifiability. These comparisons suggest that medical imaging needs a hybrid policy model, one that recognizes the specificity of clinical data while borrowing operational lessons from sectors that manage high-stakes distributed computation.

International collaboration adds another layer. Hospitals in different countries may face incompatible privacy laws, making federated learning legally complex even when technically feasible. Data localization requirements may prevent certain forms of coordination. Secure inference can help by keeping images local, but model updates or predictions may still cross borders. Governance must therefore include legal mapping, contractual safeguards, and mechanisms for resolving conflicts. It must also address transparency without compromising privacy. Patients and clinicians need understandable explanations of how decisions are supported, but full model disclosure may reveal proprietary information or create attack surfaces. The policy challenge is to design accountability mechanisms that are meaningful without requiring complete openness or complete secrecy.

8. Future Research Directions

Future research should move beyond isolated benchmarks toward end-to-end clinical trials of privacy-preserving systems. Such trials must measure not only accuracy and cryptographic overhead but also clinician workflow, patient acceptability, failure modes, and equity outcomes. Secure inference protocols should be co-designed with neural architectures that are robust to approximation and efficient under realistic network conditions. Federated learning should support heterogeneous participation, including sites with limited bandwidth or intermittent connectivity. Differential privacy should be evaluated with subgroup-specific utility metrics. Governance research should develop audit models that can verify compliance without exposing sensitive data. Hardware advances, including trusted execution environments and specialized accelerators, may reduce costs, but they also introduce new supply-chain and side-channel risks. The most promising direction is not a single breakthrough but an ecosystem of interoperable components, standards, and institutions.

9. Conclusion

Privacy-preserving medical image classification cannot be reduced to a choice between accuracy and confidentiality. It is a systems problem that spans secure neural inference, communication-efficient federated computation, robustness, fairness, deployment, sustainability, and governance. Secure inference can protect patient images during prediction, but its cost depends on nonlinear activation functions, network design, and hardware support. Federated learning can reduce central data pooling, but it introduces communication bottlenecks, heterogeneity, and coordination challenges. Differential privacy and secure aggregation add important protections, yet they must be calibrated against clinical utility and subgroup fairness. The most credible path forward is a layered architecture in which technical mechanisms are matched to institutional responsibilities and regulatory expectations. Such an architecture must be auditable, resilient, and economically sustainable, or it will not survive contact with real hospitals. The broader lesson is that privacy is not a feature added to medical

AI after development. It is a property of the entire socio-technical system, produced through careful design, transparent governance, and continuous evaluation.

References

1. Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 3–18). IEEE. <https://doi.org/10.1109/SP.2017.41>
2. Fredrikson, M., Jha, S., & Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (pp. 1322–1333). ACM. <https://doi.org/10.1145/2810103.2813677>
3. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3, 119. <https://doi.org/10.1038/s41746-020-00323-1>
4. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273–1282). PMLR.
5. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. <https://doi.org/10.1561/22000000083>
6. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... (2017). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175–1191). ACM. <https://doi.org/10.1145/3133956.3133982>
7. Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., Naehrig, M., & Wernsing, J. (2016). CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy. In Proceedings of the 33rd International Conference on Machine Learning (pp. 201–210). PMLR.
8. Mohassel, P., & Zhang, Y. (2017). SecureML: A system for scalable privacy-preserving machine learning. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 19–38). IEEE. <https://doi.org/10.1109/SP.2017.12>
9. Liu, J., Juuti, M., Lu, Y., & Asokan, N. (2017). Oblivious neural network predictions via MiniONN transformations. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 619–631). ACM. <https://doi.org/10.1145/3133956.3134056>
10. Juvekar, C., Vaikuntanathan, V., & Chandrakasan, A. (2018). Gazelle: A low latency framework for secure neural network inference. In Proceedings of the 27th USENIX Security Symposium (pp. 1651–1669). USENIX Association.
11. Mishra, P., Lehmkuhl, R., Srinivasan, A., Zheng, W., & Popa, R. A. (2020). Delphi: A cryptographic inference service for neural networks. In Proceedings of the 29th USENIX Security Symposium (pp. 2505–2522). USENIX Association.

12. Rathee, D., Rathee, M., Kumar, N., Chandran, N., Gupta, D., Rastogi, A., & Sharma, R. (2020). CrypTFlow2: Practical 2-party secure inference. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security* (pp. 325–342). ACM. <https://doi.org/10.1145/3372297.3417274>
13. Knott, B., Venkataraman, S., Hannun, A., Sengupta, S., Ibrahim, M., & van der Maaten, L. (2021). CrypTensor: Flexible framework for privacy-preserving deep learning. *arXiv*. <https://arxiv.org/abs/2109.11726>
14. Ziller, A., Trask, A., Lopardo, A., Szymkow, B., Wagner, B., Bluemke, E., ... (2021). PySyft: A library for easy federated learning. In *Federated Learning Systems* (pp. 111–139). Springer. https://doi.org/10.1007/978-3-030-70604-3_5
15. Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6), 305–311. <https://doi.org/10.1038/s42256-020-0186-1>
16. Peng, Hongwu, et al. "Rrnet: Towards relu-reduced neural network for two-party computation based private inference." *arXiv preprint arXiv:2302.02292* (2023).
17. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., ... (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, 10, 12598. <https://doi.org/10.1038/s41598-020-69250-1>
18. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50–60. <https://doi.org/10.1109/MSP.2020.2975749>
19. Papernot, N., Abadi, M., Erlingsson, U., Goodfellow, I., & Talwar, K. (2017). Semi-supervised knowledge transfer for deep learning from private training data. In *International Conference on Learning Representations*.
20. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407. <https://doi.org/10.1561/04000000042>
21. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 308–318). ACM. <https://doi.org/10.1145/2976749.2978318>
22. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
23. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—Addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981–983. <https://doi.org/10.1056/NEJMp1714229>
24. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMr1814259>
25. Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>