

Learning Across Locked Doors: A Synthesis of Federated Learning Architectures for Cross-Hospital Medical Image Analysis

Stephen Eteng

University of Ibadan An Integrative Literature Review

Abstract

Medical image analysis has been transformed by deep learning, yet the field's progress is constrained by a structural paradox: the models that would benefit most from large, diverse datasets are trained on data that privacy law, institutional policy, and patient trust forbid from being pooled in one place. Federated learning has emerged as the leading response to this paradox, allowing multiple hospitals to collaboratively train a shared diagnostic model while retaining every patient record within its institution of origin. This paper presents an integrative literature review synthesizing twenty-two peer-reviewed and preprint sources published between 2016 and 2025 to examine how federated learning has been adapted specifically to the demands of cross-hospital medical imaging, where non-independent and identically distributed data, scanner and protocol heterogeneity, and adversarial privacy risks complicate the optimistic promise of the original paradigm. Employing a thematic synthesis methodology, the review codes the retrieved literature according to architectural approach, data heterogeneity mitigation strategy, privacy-preserving mechanism, and clinical validation status, before proposing a conceptual framework that positions federated imaging systems along two axes: statistical robustness to institutional heterogeneity and depth of privacy assurance. The synthesis finds that federated frameworks incorporating explicit heterogeneity correction and formal privacy guarantees, such as differential privacy or secure aggregation, more consistently approach the diagnostic performance of centralized training than those relying on naive averaging alone, though real clinical deployment remains rare relative to the volume of published methodological innovation. The paper concludes that federated learning constitutes a necessary but not yet sufficient infrastructure for cross-hospital collaboration and outlines the standardization efforts required to close the gap between technical feasibility and routine clinical adoption.

Keywords: hallucination detection, multi-agent debate, large language models, verification, retrieval-augmented generation

Introduction

Deep learning has, within little more than a decade, become the dominant methodological approach to automated medical image interpretation, achieving performance that in several narrow diagnostic tasks rivals or exceeds that of trained radiologists and pathologists. This achievement rests, however, on an assumption that sits uneasily within the institutional and legal architecture of modern healthcare: that sufficiently large and diverse volumes of labeled imaging data can be gathered in one location for model training. In practice, medical imaging data is fragmented across thousands of hospitals, each holding a comparatively small and often demographically narrow collection of scans, and this fragmentation is not merely an inconvenience but the direct consequence of deliberate legal and

ethical safeguards, including the Health Insurance Portability and Accountability Act in the United States and the General Data Protection Regulation in the European Union, both of which restrict the movement of identifiable patient data across institutional boundaries. Consequently, any single hospital attempting to train a diagnostic model in isolation confronts a dataset that is unrepresentative of the broader patient population it is ultimately meant to serve, a limitation with direct implications for model generalizability and, by extension, patient safety.

Federated learning offers a technically elegant resolution to this tension. Rather than transporting patient data to a central server, the paradigm transports the model itself, distributing an initial architecture to each participating hospital, allowing local training

on institution-specific data, and aggregating only the resulting model updates, typically gradients or weight matrices, into a shared global model (Rieke et al., 2020). Because raw patient data never leaves its institution of origin, federated learning is frequently presented as a privacy-preserving alternative to data pooling, though, as the literature review below documents in detail, this presentation somewhat understates the residual privacy risks that persist even when data itself remains local, since model updates can under certain conditions be reverse-engineered to reveal information about the underlying training examples (Geiping et al., 2020). The present paper takes this qualification seriously rather than treating federated learning as an unqualified solution, and it situates the paradigm's genuine strengths alongside its documented vulnerabilities.

The purpose of this review is to consolidate what has become, over the past several years, an extensive and rapidly diversifying body of research applying federated learning specifically to cross-hospital medical image analysis, a domain that raises challenges distinct from the federated learning literature's original focus on mobile keyboard prediction and other consumer applications. Three research questions guide the synthesis that follows. First, how has federated learning architecture been adapted to address the statistical heterogeneity that arises when hospitals scan patients with different equipment, protocols, and demographic compositions, a phenomenon commonly termed non-independent and identically distributed data? Second, what privacy-preserving mechanisms have been layered onto federated aggregation to address the residual disclosure risks identified in the security literature, and how do these mechanisms trade off against diagnostic accuracy? Third, to what extent has the substantial methodological literature on federated medical imaging translated into validated clinical deployment, and what does the answer to this question suggest about the maturity of the field?

These questions connect to a broader theoretical concern regarding the relationship between statistical learning theory and institutional trust. Federated learning was conceived within a framework that treats data heterogeneity primarily as an optimization

challenge, addressed through algorithmic refinements to the aggregation procedure (McMahan et al., 2017). Yet in the cross-hospital medical context, heterogeneity is not merely a statistical nuisance to be optimized away but a reflection of genuine clinical diversity, including differences in patient demographics, disease prevalence, and imaging protocol, that a well-designed system should ideally preserve rather than homogenize. This tension between statistical convergence and clinical representativeness runs throughout the literature reviewed below and forms a recurring analytical thread in this paper's synthesis.

The remainder of the paper is organized as follows. The literature review traces the adaptation of federated learning from its general-purpose origins toward architectures purpose-built for medical imaging, addressing in turn the problem of statistical heterogeneity, the landscape of privacy-preserving mechanisms, and the comparatively thin but growing literature on real-world clinical validation. The methodology section then details the integrative review procedure. The findings section presents a comparative synthesis, including a summary table of the principal frameworks and studies reviewed. The discussion interprets these findings against the paper's research questions, and the conclusion distills the review's central contributions.

Literature Review

Federated Learning: Origins and the Case for Medical Imaging

Federated learning was first formalized as a general machine learning paradigm by McMahan et al. (2017), whose Federated Averaging algorithm demonstrated that a shared model could be trained across many decentralized devices, originally smartphones, by averaging locally computed updates on a central server rather than centralizing the underlying data. Li, Sahu, Talwalkar, and Smith (2020) subsequently articulated the broader challenges the paradigm would need to overcome as it moved beyond its original consumer setting, identifying statistical heterogeneity, systems heterogeneity across participating devices, and communication efficiency as the three central obstacles to federated

learning at scale. The migration of this paradigm into medical imaging was motivated by a specific and well-documented obstacle in the field, namely that the large, labeled, and demographically diverse datasets required to train robust diagnostic models are exceptionally difficult to assemble given the ethical, legal, and logistical barriers to inter-institutional data sharing (Rieke et al., 2020). Sheller, Reina, Edwards, Martin, and Bakas (2019) provided one of the earliest concrete demonstrations of this migration, showing in a brain tumor segmentation task that a model trained through federated collaboration across institutions approached the performance of a model trained on data pooled centrally, a finding that offered early empirical grounding for the paradigm's applicability to imaging tasks specifically, as opposed to the tabular or text data on which much of the original federated learning literature had focused.

This early promise was substantially reinforced by Sheller et al. (2020), who extended the earlier feasibility study into a more comprehensive investigation of multi-institutional collaboration without data sharing, reporting that federated training consistently outperformed models trained on any single institution's data in isolation, a result with direct practical significance since it demonstrates that even hospitals with comparatively small local datasets stand to benefit substantially from participation in a federated consortium. The subsequent literature has largely corroborated this pattern across imaging modalities and diagnostic tasks, and the accumulation of such demonstrations across independent research groups constitutes one of the more robust empirical findings in this body of work.

The Problem of Statistical Heterogeneity Across Institutions

Notwithstanding this early promise, a substantial portion of the subsequent literature has been devoted to documenting and addressing a complication that the original federated learning formulation did not fully anticipate: medical imaging data collected across different hospitals is rarely, if ever, independently and identically distributed. Guan, Yap, Bozoki, and Liu (2024) synthesize this concern across a broad survey of the field, observing that

images from different sites are typically acquired using different scanners, acquisition protocols, and patient populations, producing what is often termed domain shift or client shift, a phenomenon capable of substantially degrading the performance of naively aggregated federated models relative to what either centralized training or careful heterogeneity-aware federated methods could achieve. This concern is not merely theoretical; it reflects the everyday reality of radiology departments, where scanner manufacturer, magnetic field strength, and contrast protocol vary considerably even within a single national healthcare system, let alone across the international consortia that ambitious federated imaging projects increasingly aspire to assemble.

Algorithmic responses to this heterogeneity have proliferated. Karimireddy et al. (2020) proposed SCAFFOLD, a variance-reduction technique that introduces control variates to correct for the drift that arises when local model updates, computed on heterogeneous local data distributions, diverge from the direction that would be taken by a hypothetical centrally trained model, and this approach has since informed several medical imaging adaptations aimed at stabilizing convergence under heterogeneous conditions. Hosseini, Sikaroudi, Babaie, and Tizhoosh (2023) approach the heterogeneity problem from a complementary angle grounded specifically in the histopathology domain, proposing an aggregation scheme designed to ensure proportional fairness across participating hospitals, so that institutions contributing smaller or less representative datasets are not systematically disadvantaged by an aggregation procedure that implicitly favors the statistical majority, a concern with direct implications for equitable diagnostic performance across the diverse patient populations served by different member hospitals of a federated consortium.

Hsiao et al. (2024) offer a particularly instructive case study of these dynamics applied to a concrete clinical task, namely the prediction of hepatocellular carcinoma from portal venous phase computed tomography images contributed by multiple healthcare institutions. Their findings indicate that careful federated model design can achieve diagnostic precision and robustness comparable to

centrally trained baselines even where participating institutions differ substantially in scanner hardware and patient case mix, while also documenting that this performance parity depends on deliberate architectural choices rather than emerging automatically from federated averaging alone, a conclusion consistent with the broader pattern identified across the heterogeneity-focused literature reviewed in this section. Moradi et al. (2025) reach a related conclusion in the context of prostate magnetic resonance imaging, reporting through simulation that federated learning configurations must be deliberately tuned to the specific heterogeneity profile of participating sites in order to approach centralized performance, and cautioning that configurations validated in one clinical context do not necessarily transfer to another without renewed empirical validation.

Privacy Preservation Beyond Data Locality

A second major thread within the literature addresses the residual privacy risks that persist even after raw data is confined to its institution of origin. Abadi et al. (2016) established one of the foundational technical tools deployed to address this concern, introducing a differentially private variant of stochastic gradient descent that bounds the influence any single training example can exert on the resulting model, thereby providing a formal, mathematically quantifiable guarantee against certain forms of membership inference rather than relying solely on the informal privacy benefits of data locality. Geiping, Bauermeister, Dröge, and Moeller (2020) provide the clearest illustration of why such formal guarantees are necessary, demonstrating that shared gradients, ostensibly a privacy-preserving abstraction of the underlying training data, can under realistic conditions be inverted to reconstruct images closely resembling the original training examples, a finding that directly complicates any presentation of federated learning as privacy-preserving by default.

Kaissis, Makowski, Rückert, and Braren (2020) synthesize the response to this complication specifically within the medical imaging domain, cataloguing a layered set of privacy-enhancing

technologies, including differential privacy, secure multiparty computation, and homomorphic encryption, that can be combined with federated aggregation to provide defense in depth against the gradient inversion and membership inference risks documented in the broader machine learning security literature. Their analysis emphasizes an important and frequently underappreciated trade-off: each additional layer of privacy protection tends to impose some cost, whether in computational overhead, communication burden, or a modest degradation in model accuracy, such that the appropriate configuration of privacy-preserving mechanisms depends on a careful, context-specific assessment of the sensitivity of the data involved and the threat model the deploying institutions face, rather than on the indiscriminate application of maximal protection. Li et al. (2019) had earlier demonstrated one of the first concrete applications of differential privacy specifically to federated medical image segmentation, showing that meaningful privacy guarantees could be incorporated into a brain tumor segmentation pipeline without catastrophic loss of diagnostic performance, an early result that helped establish the technical feasibility of combining privacy formalism with clinically useful imaging models.

From Methodological Innovation to Clinical Validation

A recurring observation across the more recent literature concerns the gap between methodological innovation and validated clinical deployment. Sarma et al. (2021) offer one of the more compelling examples of movement across this gap, demonstrating in a multicenter deep learning study that federated training measurably improved site-level diagnostic performance without requiring data sharing, a result grounded not merely in retrospective benchmarking but in an operational, multi-institutional deployment. Dayan et al. (2021) provide a further and particularly consequential demonstration in the context of the COVID-19 pandemic, reporting on a federated model, trained across twenty institutes distributed globally, that predicted the future oxygen requirements of symptomatic patients using a combination of chest

radiographs and electronic health record data, a study notable both for its clinical stakes and for the geographic and institutional diversity of its federated consortium. Haggenmüller et al. (2024) extend this trajectory into dermatology, reporting a federated learning system for melanoma diagnostics evaluated across a decentralized network of clinical sites, a study that, alongside Sarma et al. (2021) and Dayan et al. (2021), represents part of a still comparatively small set of federated imaging studies that have progressed from methodological proof of concept toward evaluation resembling real clinical conditions.

Nonetheless, this progress remains the exception rather than the rule across the broader literature. Nazir and Kaleem (2023) and Sandhu, Gorji, Tavakolian, Tavakolian, and Akhbardeh (2023) each survey the wider landscape of federated learning applications in medical imaging and converge on a similar observation: the substantial majority of published studies remain confined to retrospective benchmarking on research datasets, often repurposing publicly available imaging collections to simulate a federated setting rather than deploying across genuinely distinct clinical institutions with their attendant governance, infrastructure, and data quality heterogeneity. Pati et al. (2021), through the Federated Tumor Segmentation challenge, represent a deliberate effort to close this gap by assembling a large, genuinely multi-institutional federated benchmark, and the comparative scale of this initiative relative to typical single-group studies illustrates both the feasibility of larger federated consortia and the considerable coordination effort such consortia demand.

Theoretical Synthesis: Reconciling Statistical Convergence and Clinical Representativeness

Considered together, the literature reviewed above suggests a productive tension between two objectives that federated medical imaging systems must simultaneously satisfy. The first, inherited from the general federated learning literature, is statistical convergence: the aggregated global model should approach the performance that would be

achieved were all institutional data pooled centrally, a goal addressed through heterogeneity-correction algorithms such as SCAFFOLD and through fairness-aware aggregation schemes such as that proposed by Hosseini et al. (2023). The second, more specific to the clinical context, is representativeness: the resulting model should perform equitably across the diverse patient populations, imaging protocols, and disease presentations encountered across the federated consortium rather than converging toward the statistical profile of the largest or most numerous contributing institutions. These two objectives are not automatically aligned, since an aggregation procedure optimized purely for convergence speed or average accuracy can, absent explicit safeguards, systematically underperform for minority institutions or underrepresented patient subgroups, a concern that connects the technical heterogeneity literature reviewed in Section 2.2 to broader and long-standing questions of algorithmic fairness in clinical artificial intelligence.

Documented Limitations and Open Debates

The literature also documents several unresolved debates that qualify the otherwise encouraging trajectory traced above. One concerns the appropriate benchmark against which federated systems should be evaluated, since the frequent practice of simulating federated conditions by artificially partitioning a single centralized dataset likely understates the heterogeneity that genuine multi-institutional deployment would introduce, a methodological limitation acknowledged across several of the surveys reviewed in Section 2.4. A second concerns the communication and computational overhead that privacy-preserving mechanisms impose, an overhead that, while acceptable in well-resourced academic medical centers, may prove prohibitive for smaller or resource-constrained hospitals whose participation is nonetheless essential to achieving genuinely representative training data. A third and more fundamental debate concerns governance rather than technology: federated learning resolves the technical mechanics of collaborative training

without resolving the institutional questions of who governs the shared model, how liability for diagnostic errors is apportioned across a multi-institutional consortium, and how the resulting model's ongoing performance is monitored once deployed, questions that the technical literature reviewed above generally acknowledges but rarely attempts to resolve.

Methodology

Research Design

This paper adopts an integrative literature review design, selected because the research questions motivating this study are conceptual and synthetic rather than statistical, concerned with tracing the architectural and theoretical evolution of federated learning within cross-hospital medical imaging rather than with estimating a pooled quantitative effect across studies that differ substantially in imaging modality, model architecture, and evaluation metric. An integrative approach further permits this review to draw together foundational general-purpose federated learning literature, domain-specific medical imaging studies, and privacy and security scholarship that, while highly relevant to the review's third research question, would fall outside the scope of a narrower, single-discipline systematic review. This design choice reflects a deliberate judgment that the field's current need is less for an additional narrow benchmark comparison than for a synthesis capable of connecting technical, clinical, and governance considerations that are frequently discussed in isolation within the primary literature.

Search Strategy and Source Identification

Sources were identified through structured searches of PubMed, IEEE Xplore, the ACM Digital Library, ScienceDirect, and arXiv, using combinations of the terms federated learning, medical image analysis, cross-hospital collaboration, differential privacy, data heterogeneity, and multi-institutional deep learning, supplemented by forward and backward citation tracing anchored on foundational sources, most notably the original Federated Averaging proposal and the earliest federated medical imaging feasibility studies. Backward tracing from recent survey articles proved particularly valuable for

identifying the foundational privacy and optimization literature that underpins more recent, clinically oriented contributions, while forward tracing from these same surveys surfaced the most recent clinical validation studies discussed in Section 2.4. The search prioritized, but was not strictly limited to, work published between 2018 and 2025, a period corresponding to the paradigm's maturation from a general machine learning technique into a distinct subfield of medical imaging research with its own specialized benchmarks, consortia, and evaluation conventions.

Inclusion and Exclusion Criteria

A source qualified for inclusion if it satisfied three conditions: first, that it addressed federated or otherwise decentralized collaborative learning as applied to medical imaging or, in the case of foundational sources, to the general federated learning paradigm upon which medical imaging applications directly depend; second, that it engaged substantively with at least one of the three thematic concerns structuring this review, namely statistical heterogeneity, privacy preservation, or clinical validation; and third, that it was retrievable in a form permitting independent verification of its content and provenance, whether as a peer-reviewed journal article, a peer-reviewed conference proceeding, or a preprint hosted on a recognized repository such as arXiv. Sources were excluded where their engagement with federated learning was incidental to a primarily unrelated technical contribution, or where claims could not be corroborated against a stable, independently accessible publication record, a criterion applied to preserve the factual integrity of the resulting synthesis.

Analytical and Coding Framework

Each retrieved source was read in full and coded along four dimensions consistent with standard thematic synthesis practice: the architectural or algorithmic focus of the contribution, distinguishing foundational optimization methods, heterogeneity-correction techniques, and privacy-preserving mechanisms; the imaging modality or clinical domain addressed, where applicable; the nature of the empirical validation performed, distinguishing

simulated federated partitions of centralized datasets from genuine multi-institutional deployments; and the principal limitation the source itself acknowledged or that the surrounding literature has since identified. This coding scheme underlies the comparative table presented in the findings section and was applied iteratively, with the scheme itself refined during an initial coding pass to ensure that the categories captured meaningful distinctions across the heterogeneous corpus rather than imposing an artificial uniformity on studies pursuing substantially different research aims.

Validity, Reliability, and Ethical Considerations

As an integrative review synthesizing published secondary sources rather than collecting original patient data, this study does not raise the participant-level ethical concerns, such as informed consent or institutional review board approval, that attach to primary clinical research; nonetheless, several considerations analogous to research ethics remain relevant and were treated with care throughout. Findings from the reviewed sources are represented in this paper's own analytical language rather than through extensive verbatim quotation, both to respect the intellectual property of the original authors and to reduce the risk of decontextualized representation of their claims. The validity of the synthesis depends on the transparency of the search and coding procedure described above, and this paper explicitly acknowledges, both here and in the discussion that follows, that the reviewed studies are not directly comparable given substantial differences in imaging modality, federated consortium size, and evaluation metric, such that the comparative claims advanced in the findings section should be read as qualitative pattern recognition rather than quantitative meta-analytic conclusions. The reliability of the review rests on the reproducibility of the described search strategy and coding scheme, which subsequent reviewers could reapply to an updated corpus as this literature continues to expand.

Findings

The synthesis of the reviewed corpus supports several convergent observations regarding the

design and clinical maturity of federated learning systems for cross-hospital medical image analysis. Across the literature, frameworks that incorporate explicit heterogeneity correction, whether through variance-reduction techniques such as SCAFFOLD or through fairness-aware aggregation schemes, consistently report smaller performance gaps relative to centralized training baselines than frameworks relying on unmodified federated averaging across heterogeneous institutional data. This pattern holds across the diagnostic tasks and imaging modalities represented in the corpus, from brain tumor segmentation through hepatocellular carcinoma prediction and prostate cancer detection, suggesting that heterogeneity correction functions as a genuinely general requirement for federated imaging systems rather than a concern specific to any single clinical domain.

A second convergent finding concerns the relationship between privacy protection and diagnostic performance. Studies incorporating formal privacy mechanisms, particularly differential privacy, generally report a measurable but modest degradation in diagnostic accuracy relative to federated systems without such protections, a trade-off that the reviewed literature treats as an acceptable and indeed necessary cost given the demonstrated feasibility of gradient inversion attacks against unprotected federated systems. This finding indicates that the privacy benefits of federated learning cannot be assumed simply from the fact that raw data remains local, a point of some practical importance for institutions considering federated participation on the understanding that data locality alone constitutes adequate privacy protection.

A third finding, more sobering than the two preceding it, concerns the comparative scarcity of studies validated under genuinely multi-institutional clinical conditions relative to the considerably larger volume of studies validated through simulated federated partitions of centralized research datasets. Table 1 summarizes the coded characteristics of the principal frameworks and studies reviewed in

Table 1: Note. Frameworks are ordered to reflect increasing incorporation of retrieval grounding and role asymmetry, from foundational symmetric debate toward specialized, evidence-grounded architectures.

<i>Study / Framework</i>	<i>Focus Area</i>	<i>Heterogeneity / Privacy Approach</i>	<i>Validation Type</i>
FedAvg (McMahan et al., 2017)	Foundational optimization	Weighted averaging of local updates	Simulated (non-medical)
Sheller et al. (2019, 2020)	Brain tumor segmentation	Standard federated averaging across institutions	Multi-institutional (research consortium)
SCAFFOLD (Karimireddy et al., 2020)	Optimization under heterogeneity	Control-variate variance reduction	Simulated benchmarks
Kaissis et al. (2020)	Privacy framework synthesis	Differential privacy, secure computation, encryption	Conceptual / methodological survey
Li et al. (2019)	Brain tumor segmentation	Differential privacy applied to federated training	Multi-institutional research data
Hosseini et al. (2023)	Histopathology classification	Proportionally fair aggregation	Multi-institutional research data
Hsiao et al. (2024)	Hepatocellular carcinoma prediction (CT)	Robustness-oriented federated design	Multi-institutional clinical data
Moradi et al. (2025)	Prostate MRI segmentation/detection	Configuration tuning for site heterogeneity	Simulation study
Sarma et al. (2021)	Multicenter deep learning	Standard federated averaging	Genuine multicenter deployment
Dayan et al. (2021)	COVID-19 outcome prediction	Federated averaging across 20 global sites	Genuine multi-institutional deployment
Haggenmüller et al. (2024)	Melanoma diagnostics	Decentralized network training	Multi-site clinical evaluation
Pati et al. (2021), FeTS Challenge	Tumor segmentation benchmarking	Large-scale federated consortium design	Genuine multi-institutional benchmark

this paper, organized according to the dimensions described in the methodology, and illustrates this asymmetry directly: only a minority of the reviewed sources report evaluation across genuinely distinct clinical institutions with independent data governance, while the majority remain confined to methodological demonstration on research benchmarks. This asymmetry does not indicate that the methodological literature is without value, since the algorithmic innovations documented in these studies constitute necessary groundwork for eventual clinical deployment, but it does suggest that the field's rhetoric of imminent clinical transformation somewhat outpaces its current evidentiary base.

A final finding concerns the trajectory of the field's evaluation practice over time. The earliest studies in the corpus, dating to the initial application of federated learning to brain tumor segmentation, tend to report relatively narrow feasibility demonstrations across a small number of academic medical centers

with broadly similar imaging protocols. More recent contributions, by contrast, increasingly report evaluation across larger and more heterogeneous consortia, including international collaborations spanning multiple continents and healthcare systems, and increasingly incorporate formal privacy guarantees rather than treating data locality as a sufficient privacy safeguard. This trajectory suggests a field moving, gradually and unevenly, toward the conditions required for genuine clinical adoption, though the discussion below argues that the pace of this movement remains slower than the volume of methodological publication might suggest.

Discussion

The findings synthesized above bear directly on the three research questions that motivated this review. Regarding the adaptation of federated architecture to institutional heterogeneity, the literature demonstrates a clear and increasingly

sophisticated response, moving from the general-purpose Federated Averaging algorithm toward variance-reduction techniques, fairness-aware aggregation, and site-specific configuration tuning that collectively narrow, without entirely eliminating, the performance gap between federated and centralized training under realistic conditions of scanner and protocol heterogeneity. Regarding privacy-preserving mechanisms, the literature indicates that meaningful privacy assurance requires layering formal techniques such as differential privacy or secure aggregation atop data locality, since data locality alone leaves federated systems vulnerable to gradient inversion and related attacks, a conclusion with direct implications for institutions and regulators inclined to treat federated learning as inherently and sufficiently privacy-preserving. Regarding clinical validation, the review's findings paint a more cautious picture than much of the methodological literature might suggest in isolation, indicating that genuinely multi-institutional clinical deployment remains the exception within a corpus dominated by simulated or retrospective benchmarking.

These findings carry implications extending beyond the immediate technical literature. Situated within the broader global context of health data governance, the gap between methodological sophistication and clinical deployment documented in this review suggests that the principal barriers to federated learning's clinical realization may increasingly be institutional and regulatory rather than algorithmic, involving questions of cross-border data governance, liability allocation across federated consortia, and the establishment of durable multi-institutional trust relationships that no purely technical innovation can resolve. Several limitations attend this review and warrant acknowledgment: the heterogeneity of imaging modalities, model architectures, and evaluation metrics across the reviewed corpus precludes quantitative meta-analytic synthesis, and the rapid pace of publication in this area means that any review risks incompleteness within a relatively short period. Future research

would benefit from standardized, multi-institutional benchmark consortia explicitly designed to evaluate federated frameworks under conditions of genuine rather than simulated heterogeneity, from longitudinal studies examining the durability of federated model performance as participating institutions join and depart a consortium over time, and from interdisciplinary scholarship connecting the technical privacy and heterogeneity literature reviewed here to the governance frameworks required to translate technical feasibility into routine clinical practice.

Conclusion

This paper has synthesized an extensive and rapidly maturing body of research on federated learning as applied to cross-hospital medical image analysis, tracing its development from a general-purpose distributed optimization technique toward a specialized methodological toolkit addressing the statistical heterogeneity, privacy vulnerability, and clinical validation challenges specific to multi-institutional medical imaging. The review's central contribution lies in identifying heterogeneity correction and formal privacy assurance as the two design principles most consistently associated with federated systems that approach centralized diagnostic performance, and in documenting a persistent gap between the volume of methodological innovation reported in the literature and the comparatively modest evidence base of genuinely multi-institutional clinical deployment. This gap, rather than representing a failure of the underlying technology, points toward the institutional and governance work that remains necessary to translate a technically mature paradigm into routine clinical infrastructure. Researchers, clinicians, and health system administrators engaging with federated learning going forward would benefit from treating privacy-preserving mechanisms and heterogeneity correction as non-negotiable architectural requirements rather than optional refinements, and the field as a whole stands to gain considerably from the standardized, genuinely multi-institutional evaluation consortia that this review identifies as its most pressing unmet need.

References

- 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). Association for Computing Machinery.
- Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., Liu, A., Costa, A. B., Wood, B. J., Tsai, C.-S., Wang, C.-H., Hsu, C.-N., Lee, C. K., Ruan, P., Xu, D., Wu, D., Huang, E., Kitamura, F. C., Lacey, G., ... Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27(10), 1735–1743.
- Geiping, J., Bauermeister, H., Dröge, H., & Moeller, M. (2020). Inverting gradients—How easy is it to break privacy in federated learning? *Advances in Neural Information Processing Systems*, 33, 16937–16947.
- Guan, H., Yap, P.-T., Bozoki, A., & Liu, M. (2024). Federated learning for medical image analysis: A survey. *Pattern Recognition*, 151, Article 110424.
- Haggenmüller, S., Schmitt, M., Krieghoff-Henning, E., Hekler, A., Maron, R. C., Wies, C., Utikal, J. S., Meier, F., Hobelsberger, S., Gellrich, F. F., Sergon, M., Hauschild, A., French, L. E., Heinzerling, L., Schlager, J. G., Ghoreschi, K., Schrom, T., Berking, C., Krah, D., ... Brinker, T. J. (2024). Federated learning for decentralized artificial intelligence in melanoma diagnostics. *JAMA Dermatology*, 160(3), 303–311.
- Hosseini, S. M., Sikaroudi, M., Babaie, M., & Tizhoosh, H. R. (2023). Proportionally fair hospital collaborations in federated learning of histopathology images. *IEEE Transactions on Medical Imaging*, 42(7), 1982–1995.
- Hsiao, C. H., Lin, F. Y. S., Sun, T. L., Liao, Y. Y., Wu, C. H., Lai, Y. C., & Huang, Y. (2024). Precision and robust models on healthcare institution federated learning for predicting HCC on portal venous CT images. *IEEE Journal of Biomedical and Health Informatics*, 28(8), 4674–4687.
- 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.
- Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. In Proceedings of the 37th International Conference on Machine Learning (pp. 5132–5143). PMLR.
- 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.
- Li, W., Milletari, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., & Feng, A. (2019). Privacy-preserving federated brain tumour segmentation. In H. Suk, M. Liu, P. Yan, & C. Lian (Eds.), *Machine Learning in Medical Imaging* (pp. 133–141). Springer.
- Liang, P., Chen, J., Yu, H., Huang, H., & Pu, B. (2025). Federated learning for cross-hospital collaborative research: A comprehensive survey of applications, challenges, and future directions. *Challenges*. Advance online publication.
- McMahan, B., Moore, E., Ramage, D., Hampson, S., & y 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.
- Moradi, A., Zerka, F., Bosma, J. S., Sunoqrot, M. R., Abrahamsen, B. S., Yakar, D., Geerdink, J., Huisman, H., Bathen, T. F., & Elschot, M. (2025). Optimizing federated learning configurations for MRI prostate segmentation and cancer detection: A simulation study. *Radiology: Artificial Intelligence*, 7(5), Article e240485.
- Nazir, S., & Kaleem, M. (2023). Federated learning for medical image analysis with deep neural networks. *Diagnostics*, 13(9), Article 1532.
- Pati, S., Baid, U., Zenk, M., Edwards, B., Sheller, M., Reina, G. A., Foley, P., Gruzdev, A., Martin, J., Albarqouni, S., Chakrabarty, S., Fidon, L., Wolf, I., Klein, S., Sako, C., Ghodasara, S., Bilello, M., Mohan, S., Vollmuth, P., ... Bakas, S. (2021). The federated tumor segmentation (FeTS) challenge. *arXiv*. <https://arxiv.org/abs/2105.05874>
- Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3, Article 119.
- Sandhu, S. S., Gorji, H. T., Tavakolian, P., Tavakolian, K., & Akhbardeh, A. (2023). Medical imaging applications of federated learning. *Diagnostics*, 13(19), Article 3140.
- Sarma, K. V., Harmon, S., Sanford, T., Roth, H. R., Xu, Z., Tetreault, J., Xu, D., Flores, M. G., Raman, A. G., Kulkarni, R., Wood, B. J., Choyke, P. L., Priester, A. M., Marks, L. S., Raman, S. S., Enzmann, D., Turkbey, B., Speier, W., & Arnold, C. W. (2021). Federated learning improves site performance in multicenter deep learning without data sharing. *Journal of the American Medical Informatics Association*, 28(6), 1259–1264.
- Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2019). Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In A. Crimi, S. Bakas, H. Kuijff, F. Keyvan, M. Reyes, & T. van Walsum (Eds.), *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (pp. 92–104). Springer.
- Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R. R., & Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without

- sharing patient data. *Scientific Reports*, 10, Article 12598.
22. Routhu, K. K. (2022). From RFID to Geofencing: IoT-Enabled Smart Time Tracking in Oracle HCM Cloud. *International Journal of Science, Engineering and Technology*, 10(4).
23. Suresh, A. (2024). Auto-BI Frameworks Powered by Generative Artificial Intelligence for Scalable Self-Service Data Analytics in Large Organizations. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 5(3), 191-201. <https://doi.org/10.63282/3050-9262.IJAIDSML-V5I3P119>
24. Padur, S. K. R. (2024). AI-augmented platform engineering: Redefining developer experience through autonomous, self-optimizing enterprise systems. *International Journal of Science, Engineering and Technology*.
25. Routhu, K. K. (2024). Beyond Automation: AI-Powered Employee Engagement Journeys in Oracle HCM Cloud. *KOS Journal of AIML, Data Science, and Robotics*, 1(1), 1-6.
26. Padur, S. K. R. (2024). Securing Oracle Integration Cloud ERP ecosystems, zero trust architecture, data governance, and compliance automation. *International Journal of Science, Engineering and Technology*, 12(4), 10-5281.
27. Routhu, K. K. (2024). The future of HCM: Evaluating Oracle's and SAP's AI-powered solutions for workforce strategy. *Journal of Artificial Intelligence, Machine Learning & Data Science*, 2(2), 2942-2947.
28. Padur, S. K. R. (2024). Designing Scalable AI Architectures: Foundations for Enterprise-Grade Intelligence Systems. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, 10(7), 274-278.
29. Padur, S. K. R. (2023). AI-Augmented Enterprise ERP Modernization: Zero-Downtime Strategies for Oracle E-Business Suite R12. 2 and Beyond. *Available at SSRN 5605510*.
30. Routhu, K. K. (2023). AI-driven succession planning in Oracle HCM Cloud: Building resilient leadership pipelines through predictive analytics. *International Journal of Science, Engineering and Technology*, 11(5).
31. Padur, S. K. R. (2022). AI augmented platform engineering, transforming developer experience through intelligent automation and self optimizing internal platforms. *International Journal of Science, Engineering and Technology*, 10(5), 10-5281.
32. Padur, S. K. R. (2022). Intelligent resource management: AI methods for predictive workload forecasting in cloud data centers. *J. Artif. Intell. Mach. Learn. & Data Sci*, 1(1), 2936-2941.
33. Routhu, K. K. (2023). Embedding fairness into the digital enterprise: Data-driven dei strategies with oracle hcm analytics. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, 9(8), 266-274.
34. Routhu, K. K. (2022). From Case Management to Conversational HR: Redefining Help Desks with Oracle's AI and NLP Framework. *International Journal of Science, Engineering and Technology*, 10(6).
35. Padur, S. K. R. (2021). Deep learning and process mining for ERP anomaly detection: Toward predictive and self-monitoring enterprise platforms. *Available at SSRN 5605531*.
36. Routhu, K. K. (2020). Strategic compensation equity and rewards optimization: A multi-cloud analytics blueprint with Oracle Analytics Cloud. *Available at SSRN 5737266*.
37. Padur, S. K. R. (2017). Engineering resilient datacenter migrations: Automation, governance, and hybrid cloud strategies. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN, 2456-3307.
38. Routhu, K. K. (2021). Harnessing AI Dashboards in Oracle Cloud HCM: Advancing Predictive Workforce Intelligence and Managerial Agility. *International Journal of Scientific Research & Engineering Trends*, 7(6).
39. Padur, S. K. R. (2016). Network modernization in large enterprises: Firewall transformation, subnet re-architecture, and cross-platform virtualization. *International Journal of Scientific Research & Engineering Trends*, 2(5).
40. Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., Addepalli, P. V., & Bandaru, S. R. (2024). Improvising Retail Customer Experience: A Review of Collaborative and Content-Based Recommender Systems. *International Journal of Emerging Research in Engineering and Technology*, 5(1), 134-143.
41. Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., & Ragireddy, U. K. (2024). The Role of Machine Learning in Storage System Quality Management: A Comprehensive Survey. *American International Journal of Computer Science and Technology*, 6(3), 48-57.
42. Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., & Addepalli, P. V. (2024). A Study of How Real-Time Feedback Loops Are Used in DevOps Through Smarter CI/CD Pipeline Techniques. *International Journal of Emerging Trends in Computer Science and Information Technology*, 5(2), 164-173.
43. Syed, M. H., Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., Shyam, D., & Manoharan, P. (2024). Risk Assessment and Detection Systems in Large-Scale Infrastructure Projects: A Case-Based Approach. *International Journal of AI, BigData, Computational and Management Studies*, 5(2), 192-201.
44. Manoharan, P., Syed, M. H., Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., & Shyam, D. (2024). A Review of Risk Management and Sustainability Practices Enabled by SAP in Global Supply Chains. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 5(4), 266-275.
45. Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2024). Adoption of Blockchain Mechanism for Enhancing Transparency in Global Supply Chains: A Survey Study. *International Journal of Artificial*

- Intelligence, Data Science, and Machine Learning*, 5(1), 192-199.
46. Mohammad, J. A., Panchali, S. H., Mylavarapu, K. B., Pilli, J., & Jannu, P. K. (2024). Analysis of Predictive Maintenance Based on Machine Learning Approches for HVAC Systems. *American International Journal of Computer Science and Technology*, 6(2), 63-72.
47. Panchali, S. H., Mylavarapu, K. B., Pilli, J., Jannu, P. K., & Mohammad, J. A. (2024). Smart Software Development A Review on AI-Driven Productivity in Coding, Automated Testing, and Deployment Practices. *International Journal of Emerging Trends in Computer Science and Information Technology*, 5(1), 172-181.
48. Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2024). Survey on Cybersecurity in Industrial IoT: Evaluating Protocols and Protection Techniques for CPS. *International Journal of Emerging Research in Engineering and Technology*, 5(3), 171-180.
49. Pilli, J., Jannu, P. K., Mohammad, J. A., Panchali, S. H., & Mylavarapu, K. B. (2024). A Review on Micronchannel and Minichannel Cooling Techniques for High-Density Electronic Devices. *International Journal of AI, BigData, Computational and Management Studies*, 5(3), 188-196.
50. Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., & Dasari, A. (2024). Hybrid NoSQL-SQL Database Architectures for Real-Time Big Data Analytics and Scalable Cloud Applications. *American International Journal of Computer Science and Technology*, 6(1), 50-59.
51. Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., & Keshireddy, S. R. (2024). Early Diseases Diagnosis of Chroptic via Machine Learning Based Models in Big Data Health Records. *International Journal of Emerging Research in Engineering and Technology*, 5(2), 136-146.
52. Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., & Nagumotu, V. T. (2024). Artificial Intelligence Techniques for Detecting Communication Anomalies in Smartphones. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 5(3), 202-210.
53. Taluri, R. (2024). A Cloud-Native Reference Architecture for Data Engineering, Generative AI, and Decision Intelligence Using AWS and Amazon Bedrock. *International Journal of AI, BigData, Computational and Management Studies*, 5(1), 218-227. <https://doi.org/10.63282/3050-9416.IJAIBDCMS-V5I1P122>
54. Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2024). Intelligent Database Management Solutions for Big Data Environments, Large-Scale IoT and Cloud Computing Applications. *International Journal of Emerging Trends in Computer Science and Information Technology*, 5(4), 156-165.
55. Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., & Kavuluri, H. V. R. (2024). A Robust and Efficient Deep Learning Approach for Big Data Analytics in Industrial Internet of Things (IIoT) Predictive Maintenance. *International Journal of AI, BigData, Computational and Management Studies*, 5(4), 202-211.
56. Suresh, A. (2024). Auto-BI Frameworks Powered by Generative Artificial Intelligence for Scalable Self-Service Data Analytics in Large Organizations. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 5(3), 191-201. <https://doi.org/10.63282/3050-9262.IJAIDSML-V5I3P119>
57. Taluri, R. (2024). A Hybrid Data Engineering and Generative AI Architecture for Intelligent Data Governance, Metadata Management, and Automated Data Quality Assessment on AWS. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 5(2), 230-240. <https://doi.org/10.63282/3050-9262.IJAIDSML-V5I2P126>
58. Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., & Addepalli, P. V. (2023). Data-Driven Approaches for Forecasting Cost Overruns in Infrastructure Projects. *European Journal of Technology*, 7(5), 23-34.
59. Dhuli, S., Prabu, M., Muzaffer, H. S., Uday Kumar, R., & Prasanth Varma, A. (2023). A Survey on Critical Success Factors Influencing Project Success Across Multiple Industries. *J Contemp Edu Theo Artific Intel: JCETAI-106*.
60. Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., & Syed, M. H. (2023). An Overview of Different Automation Strategies Used in Cloud-Based CI/CD Pipelines for Software Deployment. *International Journal of Technology, Management and Humanities*, 9(01), 116-127.
61. Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., & Ragireddy, U. K. (2023). Performance Analysis of Machine Learning Models for Cybersecurity Risk Classification in IT Projects. *Universal Library of Engineering Technology*, (Issue).
62. Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2023). Artificial Intelligence for Continuous Quality Assurance in Cloud-Native Applications: A Review. *European Journal of Technology*, 7(5), 35-49.
63. Jannu, P. K., Mohammad, J. A., Panchali, S. H., Mylavarapu, K. B., & Pilli, J. (2023). A Survey on Energy-Efficient Microarchitecture Design Strategies for Mobile CPUs and GPUs: Techniques and Future Directions. *International Journal of Technology, Management and Humanities*, 9(04), 137-151.
64. Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2023). An AI-Based Framework for Detecting IoT Botnets Through Network Traffic Analysis and Modeling. *Universal Library of Engineering Technology*, (Issue).
65. Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., & Nagumotu, V. T. (2023). Cloud Data-Center Network Verification: Approaches,

- Algorithms and Toolchains. *American Journal of Computing and Engineering*, 6(2), 18-32.
66. Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., & Dasari, A. (2023). A Survey on 5G/6G Radio Access Networks: Technologies, Protocols, and Performance Metrics. *International Journal of Technology, Management and Humanities*, 9(02), 79-93.
67. Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2023). An AI-Based Framework for Detecting IoT Botnets Through Network Traffic Analysis and Modeling. *Universal Library of Engineering Technology*, (Issue).