

Privacy-Preserving Federated Learning for Distributed Healthcare Analytics Platforms

Rajagopal P^{1*}, Narasimhan S² and Sivakumar P³

¹Department of Computer Science and Engineering, Chaliyar Institute of Quantum Computing, South Korea Email: prakash.rajagopal@chaliyar.edu

ORCID: <https://orcid.org/8691-0886-1800-3650>

²Department of Intelligent Systems, Tamiraparani Center for Cognitive Computing, South Korea Email: srinivas.narasimhan@tamiraparani.edu

ORCID: <https://orcid.org/0968-0928-7550-3020>

³Department of Computational Intelligence, Bhavani Center for Neuromorphic Engineering, Malaysia Email: praveen.sivakumar@bhavani.edu

ORCID: <https://orcid.org/6226-8586-1600-7188>

ABSTRACT

Patient data is uniquely sensitive and, under HIPAA, the GDPR, and comparable regimes, is rarely permitted to leave the institution that collected it, yet the machine learning models most needed in medicine are exactly the ones that benefit from scale and demographic diversity no single hospital can supply alone. Federated learning offers a way to train shared models across many institutions' data without centralizing it, but sharing model updates instead of raw records is not, by itself, a sufficient privacy guarantee: gradients can leak training data, and trained models can leak membership. This article synthesizes the literature on privacy-preserving federated learning for distributed healthcare analytics, tracing the field from federated averaging and its cross-silo healthcare variant through the differential privacy, secure aggregation, and encryption-based mechanisms developed to close the gaps plain federation leaves open, the multi-institutional deployments on real clinical data that have validated the pattern in oncology, infectious disease, and general clinical prediction, and the attack literature that motivates the privacy layer in the first place. Read together, the evidence supports privacy-preserving federated learning as a genuinely validated foundation for distributed healthcare analytics platforms, while leaving standardized cross-study evaluation, quantified privacy-utility reporting, and robustness to heterogeneous or adversarial participants as an open research agenda.

Keywords: federated learning; privacy-preserving machine learning; healthcare analytics; differential privacy; secure aggregation

Citation: Rajagopal P, Narasimhan S and Sivakumar P [2022]. Privacy-Preserving Federated Learning for Distributed Healthcare Analytics Platforms. **DOI:** <https://doi.org/10.5281/zenodo.22843844>

Copyright: © 2022 by the authors. Open access under [CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/).

Article Information: Received: 30 June 2021 Accepted: 8 October 2021 Published: 1 January 2022

Article Class: Research Article

1. Introduction

Healthcare generates an enormous amount of data that, by both law and good sense, is not supposed to leave the walls of the institution that collected it. Electronic health records, radiology archives, pathology slides, and genomic panels sit behind institutional firewalls not because hospitals are secretive about their science but because patient data is uniquely sensitive, re-identifiable in ways that are hard to fully anticipate, and governed by regulatory regimes — HIPAA in the United States, the GDPR in Europe, and comparable frameworks elsewhere — that make centralizing it across institutions legally fraught even when everyone involved would like to collaborate. This creates a genuine tension for machine learning research in medicine: the models that would benefit most from scale, robustness, and demographic diversity are exactly the ones that a single hospital's data cannot supply on its own, because any one institution's patient population is a narrow, geographically and demographically biased slice of the disease it is trying to characterize.

Federated learning (FL) was proposed, in its now-canonical form, as a way to train a shared model across many clients' local data without ever moving that data off-device or off-site [1]. A central coordinator broadcasts model parameters to participating clients; each client trains locally on its own data; only the resulting model updates, not the raw data, are sent back and aggregated. For consumer applications — next-word prediction on a phone keyboard, say — this was originally framed as a bandwidth and latency optimization as much as a privacy one. For healthcare, the framing inverts: bandwidth is rarely the binding constraint, but the inability to centralize patient data is close to an absolute one, and federated learning offers a plausible way to train genuinely multi-institutional models — spanning dozens of hospitals, several countries, and thousands of patients — without any single participant ever having to hand over identifiable records to anyone else [2].

It would be a mistake, though, to treat "federated" as a synonym for "private." Sharing model gradients instead of raw data closes the most obvious leak, but a substantial body of security research has shown that gradients themselves can leak a surprising amount about the data that produced them — in some cases enough to reconstruct training images pixel-for-pixel [3]

— and that a trained model's outputs alone can sometimes reveal whether a specific patient's record was in the training set at all [4]. Healthcare federated learning has therefore developed as a genuinely interdisciplinary effort, combining the distributed-optimization machinery of federated learning with a second layer of formal privacy-preserving techniques — differential privacy, secure aggregation, and homomorphic encryption chief among them — that are designed specifically to close the gaps that "just don't share the raw data" leaves open [5].

This article synthesizes that combined literature as it applies specifically to distributed healthcare analytics platforms. We trace federated learning from its general-purpose optimization foundations through the privacy-preserving mechanisms that make it suitable for clinical deployment, survey the architectures and multi-institutional deployments that have actually been built and evaluated on real patient data, examine the attack surface that motivates the privacy layer in the first place, and close with the open challenges that, in our reading, still separate a promising research pattern from a routinely deployable clinical infrastructure. Throughout, we have tried to hold ourselves to citing only results that the underlying papers actually report, rather than the somewhat rosier version of "federated learning works in healthcare" that circulates informally; where the evidence is qualified, we say so.

It is worth being explicit up front about what kind of evidence this literature actually offers. Some of the strongest results — the EXAM network's federated COVID-19 outcome model [6], the Federated Tumor Segmentation (FeTS) initiative's brain-tumor delineation results [7], and the ten-institution federated brain-tumor segmentation study led by Sheller and colleagues [8] — come from real multi-institutional deployments on real clinical data, not simulations, which is unusual and valuable in a field where most machine learning papers are evaluated on public benchmarks. Others, particularly on the attack and defense side, remain largely demonstrations on standard image-classification benchmarks rather than clinical data specifically. We try to keep that distinction visible rather than blur it.

2. Federated Learning Foundations for Healthcare Data

The algorithmic core of most healthcare federated

learning systems remains Federated Averaging (FedAvg), introduced by McMahan and colleagues: a central server distributes the current global model to a selected subset of clients, each client performs several epochs of local stochastic gradient descent on its own data, and the server averages the returned model updates, weighted by local dataset size, to produce the next global model [1]. The appeal of this scheme for healthcare is almost immediate once stated: a hospital's local training loop never has to expose an individual patient record, only the net effect that its entire local dataset had on the model's parameters after a round of training. Kairouz and colleagues' widely cited survey of federated learning catalogues the broader design space this basic loop sits inside — synchronous versus asynchronous aggregation, cross-device versus cross-silo deployments, and the communication-efficiency and robustness techniques that have accumulated around the core FedAvg idea — and is a useful map for readers who want the general-purpose version of the story before the healthcare-specific one [9].

Healthcare deployments are, almost without exception, what the federated learning literature calls "cross-silo" rather than "cross-device": a federation of perhaps five to fifty hospitals or research consortium members, each with a substantial, well-curated (if still biased) local dataset and reliable connectivity, rather than millions of intermittently connected phones. This distinction matters more than it might initially seem to. Cross-silo federations can typically assume that participants stay in the federation for the whole training run, that each participant's local compute budget is nontrivial (a hospital research infrastructure, not a smartphone), and that the number of participants is small enough that per-round communication and coordination costs which dominate cross-device FL research are rarely the binding constraint. What dominates instead is statistical heterogeneity — different hospitals serve different patient populations, use different imaging protocols and scanner vendors, and code diagnoses with different conventions — which is precisely the condition under which plain FedAvg is known to struggle. Li and colleagues' FedProx algorithm, which adds a proximal term to the local training objective to bound how far a client's local update can drift from the current global model, was proposed explicitly to make federated optimization more robust to this kind of heterogeneity, and remains a standard building block that healthcare

federated learning systems draw on or adapt [10].

Yang and colleagues' widely used taxonomy of federated learning distinguishes horizontal federation, where participants share the same feature space but hold different patients (the situation for most multi-hospital imaging federations discussed below), from vertical federation, where participants hold different features on an overlapping patient population (a hospital and an insurer, say, or a hospital and a genomics lab) [11]. This distinction turns out to matter a great deal for healthcare specifically, because the two settings call for genuinely different aggregation and privacy machinery — horizontal federation fits FedAvg-style parameter averaging naturally, while vertical federation, as we discuss below, generally requires cryptographic techniques operating on partial feature computations rather than simple weight averaging. Bonawitz and colleagues' systems-level account of what it takes to run federated learning at production scale, developed originally for cross-device mobile deployments, is also a useful point of contrast for healthcare's more typical cross-silo setting: client selection, fault tolerance for participants dropping out mid-round, and the orchestration layer needed to coordinate a federation reliably are all real engineering concerns their paper works through in detail, even though the specific scale and reliability assumptions of a keyboard-prediction federation of millions of phones differ substantially from a federation of a few dozen hospital research infrastructures [12].

Xu and colleagues' review of federated learning for healthcare informatics frames the field's foundational technical challenges specifically in clinical terms: statistical heterogeneity arising from differing patient demographics and disease prevalence across sites, system heterogeneity arising from the wide range of computational infrastructure different healthcare institutions maintain, and the specific privacy and regulatory requirements — informed consent, institutional data-use agreements, HIPAA and GDPR compliance — that shape what a federated system is even allowed to do before any algorithmic consideration enters the picture [13]. Their review is useful precisely because it treats these constraints as first-class design requirements rather than implementation detail, which is the correct posture for anyone building a system that will actually touch patient data rather than a benchmark

dataset. Li, Sahu, Talwalkar, and Smith's companion survey of federated learning's core challenges — statistical heterogeneity, systems heterogeneity, privacy, and communication cost — offers the shorter, general-purpose version of the same map, useful as a first orientation before the healthcare-specific literature [14].

3. Privacy-Preserving Mechanisms: Differential Privacy, Secure Aggregation, and Encryption

Federated learning's baseline privacy property — raw data never leaves the client — is necessary but not sufficient for healthcare deployment, and the field has developed three complementary mechanisms to close the remaining gaps: differential privacy, secure aggregation, and homomorphic or other cryptographic encryption of updates. Each addresses a different threat, and healthcare federated learning systems increasingly combine more than one of them rather than treating them as interchangeable alternatives.

Differential privacy provides a formal, mathematically provable guarantee that the presence or absence of any single individual's data in a dataset changes the probability of any particular output by at most a bounded, quantifiable amount. Abadi and colleagues showed how to train deep neural networks under this guarantee in practice, by clipping per-example gradients to a bounded norm and adding calibrated Gaussian noise before the aggregation step, along with a moments-accountant technique for tracking the cumulative privacy cost across many training iterations more tightly than naive composition would allow [15]. Applied to a federated round, differential privacy is typically added at the point where client updates are aggregated: noise is injected into the aggregate (or into each client's update before it leaves the client, in the "local" variant), which bounds how much any single institution's, or in principle any single patient's, contribution can be inferred from the resulting global model. The tradeoff is well documented and not merely theoretical: tighter privacy budgets require more noise, and more noise degrades model accuracy, a tradeoff that becomes more binding as the number of contributing parties shrinks — which is exactly the cross-silo healthcare regime, where a federation of ten or twenty hospitals offers far less noise-averaging headroom than a federation of ten million phones.

Secure aggregation addresses a related but distinct

threat: it ensures that the central coordinator itself never sees any individual client's update in the clear, only the sum (or weighted average) across all participating clients in a round. Bonawitz and colleagues' practical secure aggregation protocol achieves this using pairwise-random masking, in which each pair of clients agrees on a shared random mask that cancels out exactly when all masked updates are summed, so the server can compute the aggregate without ever recovering an individual update, and the protocol tolerates clients dropping out mid-round without exposing the masks of the clients that stayed connected [16]. Their reported experiments quantify what this protection actually costs in communication: for sixteen-bit input values, the protocol adds roughly a 1.73-fold expansion in communication over sending updates in the clear at 2^{10} clients and 2^{20} -dimensional model vectors, rising to roughly 1.98-fold at 2^{14} clients and 2^{24} -dimensional vectors — a real but bounded overhead, and one that is easier to absorb in a cross-silo healthcare federation of a few dozen well-connected institutions than it would be in a cross-device federation of millions of intermittently connected phones. This is precisely the kind of number a healthcare deployment needs before deciding whether to add secure aggregation to a federated pipeline, and it is worth noting explicitly rather than treating "adds some overhead" as sufficient.

Homomorphic and other cryptographic techniques go further still, encrypting model updates such that computation (addition, and in the fully homomorphic case, multiplication as well) can be performed directly on ciphertexts, so that the coordinator aggregates encrypted updates without ever decrypting them individually or in aggregate until a designated party applies the decryption key. Kaissis and colleagues' review of secure, privacy-preserving, and federated machine learning for medical imaging specifically surveys where each of these techniques sits on the tradeoff between formal privacy guarantee, computational overhead, and accuracy loss, and makes a point we think is under-appreciated outside the specialist literature: no single technique among differential privacy, secure aggregation, and homomorphic encryption is sufficient on its own for a clinical-grade privacy guarantee, and the emerging consensus in medical-imaging federated learning is to layer them — secure aggregation to hide individual

updates from the coordinator, differential privacy to bound what the aggregate itself can reveal, and selective encryption for the most sensitive intermediate values — rather than to pick exactly one [5]. Cheng and colleagues' SecureBoost framework illustrates a complementary, non-neural-network approach to the same problem for the gradient-boosted tree models common in structured clinical and insurance data: using additively homomorphic encryption to compute split-finding statistics across vertically partitioned data (different institutions holding different feature columns for overlapping sets of patients, rather than different patients entirely), SecureBoost is explicitly designed to be lossless, meaning the resulting federated model achieves accuracy identical to a non-federated model trained on the pooled data, at the cost of the added encryption computation [17]. That losslessness is a meaningfully different value proposition from the accuracy-for-privacy tradeoffs that differential privacy imposes, and it is one reason vertically partitioned federated learning is attracting particular interest for cross-institutional structured healthcare data, where a hospital, an insurer, and a public health registry may each hold different feature sets on the same patient population.

Split learning offers a further architectural variant worth naming separately because it is specifically motivated by, and evaluated on, health data. Vepakomma and colleagues proposed splitting a neural network at a chosen layer, so that each client runs only the earlier layers locally on raw patient data, sends the resulting intermediate activations (rather than gradients or raw data) to a server that runs the remaining layers, and backpropagates gradients only as far back as the split point [18]. For healthcare specifically, the appeal is that split learning can, in principle, be combined with federated averaging of the client-side layers across institutions while keeping the more computationally expensive later layers centralized, trading off some of federated learning's decentralization for reduced local compute requirements — relevant for institutions with limited on-premises machine learning infrastructure, which describes a meaningful fraction of hospitals outside large academic medical centers.

4. Federated Learning Architectures and Deployments in Clinical Analytics

The clearest evidence that federated learning works for

healthcare analytics, as opposed to merely being architecturally plausible, comes from a small number of real multi-institutional deployments evaluated on genuine clinical data rather than simulated federations. Sheller and colleagues ran a federated learning study across ten institutions for brain tumor segmentation from MRI, and reported that the resulting federated model reached ninety-nine percent of the segmentation quality (measured by Dice similarity coefficient) achieved by an equivalent model trained on centrally pooled data from all ten sites, performing within about one percentage point of Dice score of the centralized baseline across both held-out test sets and leave-one-institution-out evaluation [8]. This is a genuinely important result for the field, not because a one-percent accuracy gap is surprising in isolation, but because it demonstrates that federated training on real, heterogeneous multi-site clinical imaging data, collected under each institution's own protocols and scanners, does not require anything close to the accuracy sacrifice that a skeptical reader might reasonably expect going in.

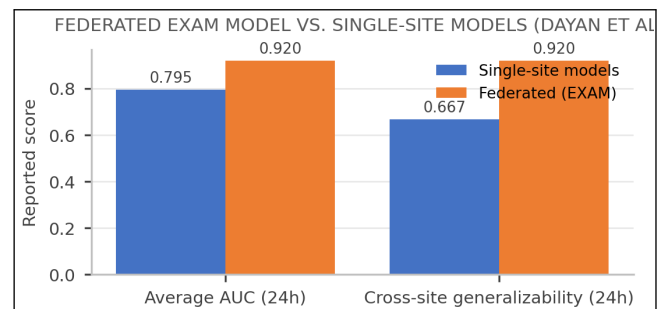


Figure 1: Average AUC and cross-site generalizability of the federated EXAM model versus single-site models for predicting 24-hour oxygen-requirement/ventilation outcomes in hospitalized COVID-19 patients, across twenty participating hospitals -- the federated model improves average AUC by 16% and cross-site generalizability by 38% relative to models trained at any one site alone [6].

The Federated Tumor Segmentation (FeTS) initiative extended this line of work substantially in scale and rigor, coordinating federated training of a brain tumor boundary-detection model across a much larger, more geographically distributed consortium of institutions specifically to address the problem that any single center, even a large academic one, sees too few cases of rare tumor subtypes and genetic markers to train a robust segmentation model on its own data alone [7]. The FeTS consensus model, trained federatively across the consortium, achieved a thirty-three percent improvement in delineation accuracy for the surgically targetable tumor core and a twenty-three percent improvement for the complete tumor extent, relative to

a model trained only on publicly available data — a substantial, clinically meaningful gap that speaks directly to the scale advantage federated learning is meant to provide: rare presentations of disease that no single institution sees often enough to learn well become learnable when the federation as a whole has seen enough of them, without any institution ever exporting a single patient's imaging data.

Dayan and colleagues' EXAM network offers a third, methodologically distinct data point, applying federated learning not to image segmentation but to structured and multimodal clinical prediction: forecasting a COVID-19 patient's oxygen requirement and risk of mechanical ventilation or death at twenty-four and seventy-two hours from emergency department presentation, trained federatively across twenty hospitals on four continents [6]. The federated EXAM model achieved an average area under the curve above 0.92 across participating sites, a sixteen percent improvement in average AUC (from roughly 0.795 to roughly 0.920) over models trained at individual sites alone for the twenty-four-hour prediction task, and an eighteen percent improvement for the seventy-two-hour task. Perhaps the more clinically important number, though, is generalizability rather than raw accuracy: the federated model showed an average thirty-eight percent increase in generalizability across sites relative to single-site models (from roughly 0.667 to roughly 0.920) for the twenty-four-hour outcome and a thirty-four percent increase for the seventy-two-hour outcome, and at the largest independent validation hospital, outside the training federation entirely, the model achieved a sensitivity of 0.950 and specificity of 0.882 for predicting the twenty-four-hour outcome. This generalizability gap — a model trained at one hospital performing markedly worse when deployed at another — is precisely the problem federated, multi-institutional training is meant to solve, and the EXAM results are among the clearest clinical demonstrations to date that it actually does.

Beyond disease-specific studies, Wernat-Herresthal and colleagues' Swarm Learning framework offers a complementary architectural variant worth noting: rather than a single central coordinator aggregating client updates, Swarm Learning distributes the coordination role itself across a blockchain-based peer-to-peer network of participating institutions, so that no single party, including a would-be central server

operator, occupies a privileged position with visibility into the aggregation process [19]. Evaluated across leukemia classification from blood transcriptomes (more than twelve thousand individuals), tuberculosis classification, COVID-19 classification from blood transcriptomes spanning more than sixteen thousand samples across 127 clinical studies, and chest radiograph classification, the authors report that swarm-learned models consistently outperformed models trained at any individual participating node, including under conditions of substantially reduced local training data and non-uniform data distribution across sites — evidence that decentralizing the coordination role itself, not just the training data, is a workable architectural variant for clinical deployment rather than a purely theoretical alternative to server-coordinated FedAvg. Nguyen and colleagues' broader survey of federated learning for smart healthcare situates deployments like these within a wider taxonomy spanning wearable sensor data, electronic health records, and medical imaging, and Antunes and colleagues' systematic review of federated learning for healthcare similarly catalogues the growing body of applied studies while noting, as we do below, that evaluation practice across this literature remains considerably less standardized than the underlying algorithms [20,21].

5. Threats, Attacks, and Defenses in Federated Healthcare Systems

The privacy-preserving mechanisms discussed above exist because federated learning's baseline privacy property — clients never send raw data — turns out to be a meaningfully weaker guarantee than intuition suggests, and demonstrating exactly how it can fail has been its own active research thread. Zhu and colleagues' deep leakage from gradients attack showed that, absent any additional protection, an adversary with access to a client's shared gradient update, and knowledge of the model architecture, can in many cases reconstruct the client's original training input essentially pixel-for-pixel, by initializing a dummy input and label and optimizing them so that the gradient they produce matches the observed gradient as closely as possible, typically converging within a few tens of optimization iterations [3]. For a federated healthcare system, the implication is direct and uncomfortable: a client whose local training batch consists of a handful of patient images, sharing raw, unprotected gradients computed on that batch, may be leaking something close to the images

themselves to any party positioned to intercept or aggregate those gradients — precisely the failure mode that secure aggregation and differential privacy are designed to prevent, and precisely why "we only share gradients, not data" cannot be treated as a sufficient privacy claim on its own.

Table 1: Representative quantitative results from multi-institutional federated learning deployments and the privacy-preserving mechanisms layered onto them, as reported in the cited sources. [8,7,6,16,3,17].

Study / Mechanism	Source	Setting	Reported Effect
10-institution federated brain tumor segmentation	Sheller et al. [8]	Model quality vs. centrally pooled data	Federated model reaches 99% of centralized-data Dice score (within ~1% Dice)
FeTS federated consensus tumor-boundary model	Pati et al. [7]	Delineation accuracy vs. publicly trained model	33% improvement for tumor core, 23% for whole tumor extent
EXAM federated COVID-19 outcome model, 20 hospitals	Dayan et al. [6]	AUC and generalizability vs. single-site models	16-18% higher average AUC; 34-38% higher cross-site generalizability
Practical secure aggregation protocol	Bonawitz et al. [16]	Communication overhead vs. plaintext updates	1.73x expansion at 2^{10} clients / 2^{20} -dim vectors; 1.98x at 2^{14} clients / 2^{24} -dim vectors
Deep leakage from gradients attack	Zhu et al. [3]	Unprotected gradient sharing (no secure aggregation/DP)	Near pixel-accurate input reconstruction typically within tens of iterations
SecureBoost vertical federated gradient boosting	Cheng et al. [17]	Accuracy vs. non-federated pooled-data model	Lossless: accuracy identical to training on pooled data

Membership inference attacks target a different, complementary question: not what a training example looked like, but whether a specific individual's data was used in training at all. Shokri and colleagues showed that an adversary with only query access to a trained model's output probabilities, and no access to its internals, can train a separate "shadow model" to recognize the characteristic confidence patterns a model exhibits on data it was trained on versus data it was not, and use that shadow model to infer, often with meaningfully better than chance accuracy, whether a given record was a training member [4]. Applied to a clinical model, a successful membership inference attack reveals something that is itself sensitive independent of any diagnosis: that a specific named individual was a patient at a specific institution, or was enrolled in a specific

study cohort, which can itself violate patient confidentiality even if no clinical detail beyond membership is exposed. Differential privacy's formal guarantee is specifically constructed to bound exactly this kind of attack, which is one of the clearest lines connecting the abstract privacy-budget mathematics in Abadi and colleagues' work to a concrete, healthcare-relevant threat model rather than an abstract cryptographic concern [15].

The practical response the field has converged on, as Kaissis and colleagues' review argues, is defense in depth rather than reliance on any single mechanism [5]. Secure aggregation closes the specific channel that deep leakage from gradients exploits, by ensuring the coordinator (and any eavesdropper on the coordinator's position) never observes an individual client's update at all, only an aggregate across many clients — which substantially raises the bar for a gradient-inversion attack, since the attacker would need to isolate one client's contribution from a masked sum rather than simply reading it off directly [16]. Differential privacy, layered on top, bounds what even the aggregate itself can reveal about any single contributor, addressing membership inference directly. Neither technique is free: secure aggregation's roughly 1.7-to-2-fold communication expansion and differential privacy's accuracy-versus-privacy-budget tradeoff are real deployment costs, and a healthcare federated learning system has to decide, case by case, how much of each it can afford given its specific threat model, regulatory requirements, and accuracy tolerances — a decision that, in our reading of the literature, is made explicitly and quantitatively far less often than it should be.

6. Open Challenges

Several gaps recur across this literature clearly enough to be worth naming directly rather than leaving implicit. First, evaluation heterogeneity: unlike the general federated learning literature, which has largely converged on a small set of standard benchmarks (CIFAR, Shakespeare, EMNIST) for comparing algorithms, healthcare federated learning studies each use their own institution-specific datasets, cohort definitions, and outcome measures, which makes direct comparison across studies — is EXAM's eighteen percent AUC improvement more or less impressive than FeTS's twenty-three percent segmentation improvement? — largely uninformative beyond each

study's own internal baseline. This is not a criticism of any individual study, since clinical data genuinely cannot be pooled into a shared public benchmark the way image-classification data can, but it does mean the field currently lacks the kind of standardized comparison that has driven rapid progress in general-purpose federated learning research.

Second, the interaction between privacy-preserving mechanisms and the statistical heterogeneity that cross-silo healthcare federations exhibit by construction remains under-characterized quantitatively. Differential privacy's noise-to-signal ratio depends on the effective population size contributing to an aggregate, and a federation of a dozen hospitals with substantially different patient demographics is a considerably less forgiving setting for meaningful privacy budgets than the large, more homogeneous cross-device populations differential-privacy-for-FL was originally evaluated against [15,22]. Very few of the clinical deployment studies surveyed here — including the highest-profile ones, EXAM and FeTS — report having applied formal differential privacy to their federated training at all, which suggests either that the accuracy cost was judged unacceptable for the clinical task at hand, or that the additional protection secure aggregation and institutional data-governance agreements already provide was judged sufficient, or, more troublingly, that the tradeoff simply was not evaluated. We think this is the single largest gap between the privacy-preserving-mechanisms literature and the clinical-deployment literature: the mechanisms exist and are reasonably well understood in isolation, but real deployments have not consistently reported using them together with the rigor the threat model would seem to demand.

Third, regulatory and governance questions remain substantially unresolved rather than merely technical. Federated learning changes what data crosses an institutional boundary, but it does not eliminate the need for institutional review board approval, data-use agreements, or patient consent processes, and different jurisdictions' interpretations of whether a shared model gradient itself constitutes "personal data" under frameworks like the GDPR are not fully settled [13]. A hospital's legal and compliance teams reasonably want to know not just "is this cryptographically secure" but "does this satisfy our specific regulatory obligations,"

and the gap between a formal privacy guarantee and a satisfactory answer to that second question is, in our reading, more of a governance and legal-scholarship problem than a machine learning one at this point, but it is one that gates real deployment as much as any algorithmic advance would.

Fourth, robustness to adversarial and low-quality participants remains a comparatively underexplored concern in the healthcare-specific literature, even though it is well studied in general federated learning. A malicious or simply poorly calibrated participating institution can degrade a federated model for everyone through a "poisoned" or systematically biased local update, and while Byzantine-robust aggregation techniques exist in the broader federated learning literature, few of the clinical deployment studies surveyed here discuss what safeguards, if any, protect the federation against a participating hospital's data being mislabeled at scale, drawn from a miscalibrated instrument, or, in the worst case, deliberately corrupted. Given that clinical federations increasingly involve dozens of institutions with varying levels of data-quality infrastructure, we would flag this as a growing rather than a shrinking concern.

Fifth, and finally, most of the attack literature motivating the defenses discussed above — deep leakage from gradients, membership inference — was demonstrated on standard image-classification benchmarks rather than on clinical imaging or structured health data specifically [3,4]. Whether these attacks transfer with equal or different severity to the specific model architectures, batch sizes, and data modalities used in real clinical federated learning deployments is, to our knowledge, not yet well characterized empirically, and represents a fairly direct and tractable piece of future work that the field has not yet fully undertaken.

7. Conclusion

The evidence reviewed here supports a fairly specific and, we think, defensible claim: federated learning, combined with an appropriately layered set of privacy-preserving mechanisms — secure aggregation to hide individual institutional contributions from the coordinator, differential privacy to bound what the resulting aggregate can reveal, and encryption-based techniques for the most sensitive intermediate computations — is no longer merely an architecturally elegant proposal for distributed healthcare analytics but

a pattern with genuine, multi-institutional clinical validation behind it. Sheller and colleagues' ten-institution brain tumor segmentation study, the FeTS consortium's substantially larger federated tumor-delineation effort, and Dayan and colleagues' twenty-hospital EXAM network for COVID-19 outcome prediction are not simulations on public benchmarks; they are federated models trained on real patient data across real institutional boundaries, evaluated against real clinical outcomes, and in each case they either matched centralized-data performance closely or substantially outperformed the single-site alternative that is the realistic counterfactual for any individual hospital acting alone [8,7,6].

At the same time, the privacy story is more work in progress than finished product. The attacks that motivate the privacy-preserving layer — gradient inversion, membership inference — are real and well demonstrated, even if not yet as thoroughly characterized on clinical data specifically as one would like [3,4], and the defenses that address them, secure aggregation and differential privacy chief among them, carry real, quantifiable costs in communication overhead and model accuracy that the clinical deployment literature has, so far, engaged with less rigorously and less consistently than the security literature that motivates them [16,15,5]. Closing that gap — building federated healthcare analytics platforms that report not just their clinical accuracy but the specific, quantified privacy guarantees they provide and the specific costs those guarantees impose — is, in our view, the most important unfinished task for this field, and one that matters more for healthcare than for almost any other federated learning application domain, precisely because the data being protected is patient data and the stakes of getting the privacy guarantee wrong are correspondingly higher.

It is also worth naming a practical implementation gap that sits underneath all of the above: many of the healthcare federated learning toolkits and reference implementations used across these studies were built by different research groups largely independently, with different assumptions about client dropout handling, checkpoint recovery, and how to reconcile slightly mismatched model architectures across institutions running different software versions. This is unglamorous compared to the cryptography and statistics discussed

elsewhere in this article, but it is frequently the actual source of delay when a new institution tries to join an established clinical federation, and it deserves more attention from the systems side of the field than it has so far received.

None of the open challenges catalogued here strike us as fatal to the pattern; if anything, the consistency with which independent multi-institutional consortia, spanning oncology, infectious disease, and general clinical prediction, have converged on structurally similar federated architectures, and have reported real, clinically meaningful gains from doing so, is itself fairly strong evidence that privacy-preserving federated learning is a sound foundation for distributed healthcare analytics platforms rather than a passing research trend. What the field needs next is less a new algorithmic breakthrough than a more disciplined, more standardized practice of reporting exactly what privacy guarantee a given clinical federation actually provides, so that the considerable, real progress documented here can be trusted and extended rather than taken on faith.

8. References

1. McMahan, H.B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A. (2017) Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 54: 1273-1282. DOI: [10.48550/arXiv.1602.05629](https://doi.org/10.48550/arXiv.1602.05629)
2. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., et al. (2020) The Future of Digital Health with Federated Learning. npj Digital Medicine, 3: 119. DOI: [10.1038/s41746-020-00323-1](https://doi.org/10.1038/s41746-020-00323-1)
3. Zhu, L., Liu, Z., Han, S. (2019) Deep Leakage from Gradients. Advances in Neural Information Processing Systems (NeurIPS) 32. DOI: [10.48550/arXiv.1906.08935](https://doi.org/10.48550/arXiv.1906.08935)
4. Shokri, R., Stronati, M., Song, C., Shmatikov, V. (2017) Membership Inference Attacks Against Machine Learning Models. Proceedings of the 2017 IEEE Symposium on Security and Privacy (S&P), 3-18. DOI: [10.1109/SP.2017.41](https://doi.org/10.1109/SP.2017.41)
5. 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: 305-311. DOI: [10.1038/s42256-020-0186-1](https://doi.org/10.1038/s42256-020-0186-1)
6. Dayan, I., Roth, H.R., Zhong, A., Harouni, A., Gentili, A., Abidin, A.Z., et al. (2021) Federated Learning for Predicting Clinical Outcomes in Patients with COVID-19. Nature Medicine, 27: 1735-1743. DOI: [10.1038/s41591-021-01506-3](https://doi.org/10.1038/s41591-021-01506-3)

7. Pati, S., Baid, U., Edwards, B., Sheller, M., Wang, S.H., Reina, G.A., et al. (2022) Federated Learning Enables Big Data for Rare Cancer Boundary Detection. *Nature Communications*, 13: 7346. DOI: [10.1038/s41467-022-33407-5](https://doi.org/10.1038/s41467-022-33407-5)
8. Sheller, M.J., Edwards, B., Reina, G.A., Martin, J., Pati, S., Kotrotsou, A., et al. (2020) Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data. *Scientific Reports*, 10: 12598. DOI: [10.1038/s41598-020-69250-1](https://doi.org/10.1038/s41598-020-69250-1)
9. Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., et al. (2021) Advances and Open Problems in Federated Learning. *Foundations and Trends in Machine Learning*, 14(1-2): 1-210. DOI: [10.1561/22000000083](https://doi.org/10.1561/22000000083)
10. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V. (2020) Federated Optimization in Heterogeneous Networks. *Proceedings of Machine Learning and Systems (MLSys)* 2: 429-450. DOI: [10.48550/arXiv.1812.06127](https://doi.org/10.48550/arXiv.1812.06127)
11. Yang, Q., Liu, Y., Chen, T., Tong, Y. (2019) Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2): 1-19. DOI: [10.1145/3298981](https://doi.org/10.1145/3298981)
12. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., et al. (2019) Towards Federated Learning at Scale: System Design. *Proceedings of the 2nd SysML Conference*. DOI: [10.48550/arXiv.1902.01046](https://doi.org/10.48550/arXiv.1902.01046)
13. Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J., Wang, F. (2021) Federated Learning for Healthcare Informatics. *Journal of Healthcare Informatics Research*, 5(1): 1-19. DOI: [10.1007/s41666-020-00082-4](https://doi.org/10.1007/s41666-020-00082-4)
14. 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. DOI: [10.1109/MSP.2020.2975749](https://doi.org/10.1109/MSP.2020.2975749)
15. Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L. (2016) Deep Learning with Differential Privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 308-318. DOI: [10.1145/2976749.2978318](https://doi.org/10.1145/2976749.2978318)
16. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., et al. (2017) Practical Secure Aggregation for Privacy-Preserving Machine Learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 1175-1191. DOI: [10.1145/3133956.3133982](https://doi.org/10.1145/3133956.3133982)
17. Cheng, K., Fan, T., Jin, Y., Liu, Y., Chen, T., Papadopoulos, D., Yang, Q. (2021) SecureBoost: A Lossless Federated Learning Framework. *IEEE Intelligent Systems*, 36(6): 87-98. DOI: [10.1109/MIS.2021.3082561](https://doi.org/10.1109/MIS.2021.3082561)
18. Vepakomma, P., Gupta, O., Swedish, T., Raskar, R. (2018) Split Learning for Health: Distributed Deep Learning Without Sharing Raw Patient Data. *arXiv preprint*. DOI: [10.48550/arXiv.1812.00564](https://doi.org/10.48550/arXiv.1812.00564)
19. Warnat-Herresthal, S., Schultze, H., Shastri, K.L., Manamohan, S., Mukherjee, S., Garg, V., et al. (2021) Swarm Learning for Decentralized and Confidential Clinical Machine Learning. *Nature*, 594: 265-270. DOI: [10.1038/s41586-021-03583-3](https://doi.org/10.1038/s41586-021-03583-3)
20. Nguyen, D.C., Pham, Q.V., Pathirana, P.N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., Hwang, W.J. (2022) Federated Learning for Smart Healthcare: A Survey. *ACM Computing Surveys*, 55(3): 1-37. DOI: [10.1145/3501296](https://doi.org/10.1145/3501296)

21. Antunes, R.S., André da Costa, C., Küderle, A., Yari, I.A., Eskofier, B. (2022) Federated Learning for Healthcare: Systematic Review and Architecture Proposal. ACM Transactions on Intelligent Systems and Technology, 13(4): 1-23. DOI: [10.1145/3501813](https://doi.org/10.1145/3501813)
22. Wei, K., Li, J., Ding, M., Ma, C., Yang, H.H., Farokhi, F., Jin, S., Quek, T.Q.S., Poor, H.V. (2020) Federated Learning With Differential Privacy: Algorithms and Performance Analysis. IEEE Transactions on Information Forensics and Security, 15: 3454-3469. DOI: [10.1109/TIFS.2020.2988575](https://doi.org/10.1109/TIFS.2020.2988575)

9. Declarations

Funding

This literature synthesis received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no known competing financial or personal interests that could have influenced this work.

Data Availability Statement

No new data were generated or analysed in preparing this article, as it is a literature synthesis; readers should consult the cited sources directly for their underlying data and code availability statements.

Ethical Approval

Not applicable. This article synthesizes previously published literature and did not involve the collection of primary data from human or animal subjects.