

Federated Mamba Learning for Privacy-Preserving Collaborative Medical Image Analysis

Silver Parker

Department of Computer Science, Binghamton University, Binghamton, NY, USA.
parker8543@binghamton.edu

Anders Simmons

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

Abstract

Federated learning has emerged as a promising route to collaborative medical image analysis without raw data exchange, but prevailing architectures often rely on convolutional or attention-based encoders that encounter limitations in modeling long-range dependencies under strict communication and privacy constraints. This paper presents a system-level examination of federated Mamba learning, an approach that integrates selective state space models into distributed medical imaging pipelines. The discussion addresses architectural choices, synchronization trade-offs, privacy accounting, heterogeneous client coordination, governance, clinical deployment, sustainability, robustness, fairness, and policy. Rather than proposing a single algorithm, the paper develops a conceptual and infrastructural framework in which federated Mamba models operate across institutions. The analysis shows that selective state space sequence modeling can reduce some computational bottlenecks of attention while introducing new challenges in gradient compression, calibration, domain shift, interpretability, and auditability. The paper argues that privacy-preserving collaborative analysis depends as much on institutional governance, auditability, and life-cycle carbon accounting as on model architecture. A future-facing agenda is presented for regulated medical data consortia, emphasizing differential privacy, secure aggregation, model cards, and energy-aware training. This synthesis is intended to guide researchers and system designers developing next-generation medical imaging federations.

Keywords

federated learning; selective state space models; medical image analysis; privacy; governance; sustainability.

1. Introduction

Federated learning has emerged as a compelling paradigm for collaborative model development when data cannot be centralized due to privacy, regulatory, or operational constraints [1]. In medical image analysis, multi-institutional collaboration is particularly valuable because rare diseases, scanner variability, and heterogeneous patient populations require broader training distributions than any single hospital can provide [2]. At the same time, patient confidentiality, institutional governance, and data-use agreements often prohibit raw image sharing [3]. Federated algorithms therefore seek to move model updates rather than protected health information. However, the shift from centralized training to decentralized coordination introduces a distinct set of system-level concerns related to communication efficiency, heterogeneity, auditability, and privacy engineering.

Recent advances in sequence modeling have brought selective state space models, often referred to as Mamba, into prominence for their ability to process long sequences with linear-time complexity [4]. Subsequent work has further connected these models to structured state space duality and efficient attention-like mechanisms [5]. In visual recognition, adaptations such as visual state space models have demonstrated competitive performance with lower activation memory in some regimes [6]. Biomedical imaging has also begun to adopt these architectures for tasks requiring long-range spatial dependency, including segmentation [7]. These developments motivate the present paper: if federated medical image analysis could integrate selective state space encoders without assuming central data aggregation, what architectural, governance, and deployment trade-offs would arise?

Traditional medical imaging pipelines frequently rely on convolutional neural networks, exemplified by deep residual architectures, because of their locality and maturity [8]. More recently, vision transformers have shown strong representation capacity but introduce quadratic attention costs that can be burdensome on three-dimensional medical volumes and in edge hospital settings [9]. Federated Mamba learning occupies a middle position: it aims to preserve global receptive fields while reducing some computational costs associated with attention. Yet the system-level implications are not limited to floating-point operations. They extend to gradient compression, differential privacy budgeting, secure aggregation overhead, client drift, and tiered compute across institutions. This paper develops a system-oriented analysis of federated Mamba learning for privacy-preserving collaborative medical image analysis.

2. Background and Privacy Foundations

Privacy-preserving machine learning has developed along three complementary lines: differential privacy, secure multi-party computation, and federated orchestration. Differential privacy provides a rigorous framework for limiting the influence of any single record on model outputs, often through gradient clipping and calibrated noise [10]. Secure aggregation protocols allow a server to compute sums of client updates without inspecting individual contributions [11]. In medical imaging, these techniques have been combined to demonstrate that collaborative learning can operate under realistic institutional constraints while preserving diagnostic performance [12]. Differential privacy in medical image analysis requires careful accounting because imaging datasets are small, high-dimensional, and prone to membership leakage [13]. Foundational treatments of differential privacy clarify how privacy budgets degrade under repeated queries, a particularly acute problem in iterative federated training [14].

The interaction between privacy mechanisms and model architecture is not neutral. Convolutional encoders with bounded receptive fields may require fewer clipping thresholds, while attention-based models can distribute information broadly and complicate sensitivity analysis. Selective state space models add a temporal or spatial recurrence that can accumulate information over long sequences, making per-parameter sensitivity nontrivial. Thus, a federated Mamba system must embed privacy engineering into the architecture selection process. Differential privacy noise may interact with state space dynamics, potentially amplifying or dampening signal depending on how hidden states are normalized. Secure aggregation likewise introduces communication latency that may be more tolerable for state space models with smaller message sizes, but this benefit can be offset if the model update contains large hidden state tensors.

A further consideration is the relationship between record-level privacy and institutional-level inference. In a federation with only a few participating hospitals, even secure aggregation may reveal aggregate patterns that allow one institution to infer another institution’s case mix or imaging protocols. This risk is heightened when selective state space models encode long-range context because a single update may carry more structural information about the underlying patient population. Therefore, privacy engineering for federated Mamba systems must include contribution caps, minimum cohort sizes, and query auditing in addition to differential privacy and secure aggregation. These measures convert a purely statistical notion of privacy into a broader socio-technical safeguard.

3. Federated Mamba Architecture and Structural Trade-offs

A federated Mamba system consists of three main subsystems: the local state space encoder, the task-specific decoder or head, and the coordination plane. The local encoder processes medical images as sequences of patches or slices, using selective state space dynamics to propagate context. The task head produces classification, segmentation, or detection outputs. The coordination plane handles initialization, aggregation, versioning, and participant metadata. Structural trade-offs emerge at each boundary. If the server aggregates only task head gradients, it may reduce leakage but produce weaker global representations. If it aggregates full encoders, representation quality may improve, but the communication and privacy budgets increase. Recent cross-fusion Mamba designs for multi-view medical image classification illustrate that selective state space fusion can integrate heterogeneous views into a single diagnostic decision [15]. However, federated deployment complicates cross-view learning because not all sites possess the same views, and some views may be legally or technically unavailable.

The heterogeneity of medical institutions requires more than standard federated averaging. Client populations differ in disease prevalence, scanner manufacturer, acquisition protocol, and annotation quality. Federated optimization methods such as proximal terms can reduce client drift under heterogeneous data [16]. In a Mamba setting, heterogeneity can manifest in sequence length, patch ordering, and hidden state initialization. A system designer may standardize patch size and sequence length to simplify aggregation, yet this standardization can discard clinically meaningful resolution. Alternatives include multi-scale state space encoders that process low-resolution overviews and high-resolution local patches. These multi-scale designs can balance communication cost and diagnostic granularity, but they also increase the number of hyperparameters and complicate fairness audits.

Another structural trade-off concerns synchronous versus asynchronous coordination. Synchronous rounds provide clean aggregation semantics and simplify privacy accounting. However, they force fast institutions to wait for slow participants, increasing wall-clock time and carbon cost. Asynchronous aggregation allows clients to contribute when ready, but it can introduce stale updates and inconsistent state space parameter versions. Federated Mamba learning may benefit from bounded asynchrony, where state space layer updates are synchronized tightly while lightweight task heads are permitted to update more frequently. This tiered synchronization strategy can reduce idle time, but it requires a governance layer that records exact model versions and update provenance. Without such records, retrospective audits cannot distinguish between model drift, data shift, and privacy intervention.

4. Privacy, Security, and Governance Infrastructure

A robust federated Mamba deployment must distinguish between privacy at the model, record, and institution levels. Differential privacy offers record-level protection but may be insufficient for group-level confidentiality. Secure aggregation protects in-transit updates, while secure enclaves or trusted execution environments protect server-side computations. However, these techniques create different trust boundaries. For example, secure aggregation prevents the server from observing individual vectors but still allows the aggregated state space update to be visible. If only one site contributes rare pathology data, the aggregate may reveal its contribution through distributional anomalies. Thus, privacy infrastructure must include minimum cohort sizes, query auditing, and contribution caps. In addition, the interaction between differential privacy and recurrent state dynamics must be monitored over multiple training rounds because hidden states can accumulate information across patches, slices, or views.

Governance is the operational layer that assigns responsibilities across institutions, data stewards, model stewards, and regulators. A federated Mamba consortium should maintain a data-sharing agreement that covers model ownership, publication rights, and breach notification. Audit logs should record model version, aggregation round, clipping thresholds, noise multipliers, and participating institutions. This provenance is essential for medical device regulation because changes in privacy parameters can alter model behavior. Furthermore, governance must address the right to withdraw a client’s data contribution. With state space models, a single hospital’s updates may be entangled in hidden state matrices over many rounds; naive deletion may require retraining or influence scrubbing. Institutions should negotiate whether withdrawal triggers local model reversion, contribution removal, or full retraining. The architecture should be designed so that withdrawal events are detectable and do not silently degrade the global model.

Security threats in federated medical imaging include gradient inversion, membership inference, and model poisoning. Gradient inversion attempts to reconstruct image patches from shared updates. Differential privacy and secure aggregation reduce but do not eliminate these risks. State space models may expose hidden state gradients that encode longer context windows, increasing the potential information carried by a single update. Therefore, client-side clipping and noise should be applied to both hidden state gradients and final layer gradients. Systems may also use local validation sets to detect anomalous updates without revealing raw data. The governance board should review security incidents through a predefined protocol, balancing transparency with the need not to disclose vulnerabilities before mitigation. These controls transform federated Mamba learning from a purely technical system into a governed socio-technical infrastructure.

5. Deployment, Robustness, and Fairness

Deployment of federated Mamba models in clinical environments requires more than model accuracy. Real-world federated medical studies have validated collaborative learning for clinical outcome prediction across institutions, but they also reveal operational frictions related to data completeness, institutional review, and technical staffing [17]. Chest radiograph benchmarks such as large public models have shown that deep networks can approach expert-level classification in controlled settings [18]; federated models must demonstrate similar reliability across uncurated multi-site data. A federated Mamba system should be evaluated not only on area under the curve but also on calibration stability under site-specific distribution shift. Selective state space models may learn long-range contextual patterns that are clinically meaningful, such as the relationship between a peripheral nodule

and central airway structures, but these patterns can also encode site-specific artifacts. Robustness testing therefore requires audit datasets with scanner, dose, and institution labels.

Fairness is a structural property of the entire federation, not a post hoc metric. If participating sites are predominantly academic medical centers with advanced imaging equipment, the global model may systematically underperform in community hospitals or low-resource settings [19]. Federated Mamba learning can mitigate this by including tiered compute profiles, local adaptation boundaries, and fairness-aware aggregation weights. However, fairness-aware weighting can conflict with privacy preservation because it requires the server to know demographic or institutional characteristics. Governance mechanisms must decide which fairness attributes are disclosed, how they are aggregated, and whether site-level performance is reported. The choice of Mamba hidden state size and patch sequence length also interacts with fairness: aggressive compression may degrade performance on rare findings that require high-resolution context, thereby increasing disparities across patient groups.

Robustness under adversarial and environmental perturbations is especially important in medical imaging. Federated pipelines face risks from label noise, delayed annotations, and malicious clients. Secure aggregation can prevent direct gradient inspection, but it does not by itself eliminate model poisoning. Systems may include update clipping, anomaly detection on client contributions, and differential privacy noise to reduce influence. The energy implications of these defenses are nontrivial. Training large visual models has been shown to carry significant carbon costs, with model architecture, hyperparameter tuning, and retraining cycles all contributing [20]. Federated Mamba models may reduce attention-related activation costs in some settings, but secure aggregation, differential privacy, client heterogeneity, and governance audits can offset these savings. A full life-cycle assessment should include computation at hospitals, aggregation servers, audit nodes, and environmental monitoring infrastructure.

Explainability and clinical trust remain central to adoption. Post hoc visual explanation techniques such as gradient-weighted class activation mapping can help clinicians inspect model decisions [21], but their reliability in federated state space models requires further study. Because Mamba hidden states propagate over long sequences, the spatial attribution of a prediction may be dispersed across distant regions, complicating conventional saliency maps. Federated learning systems originally built for nonclinical applications, such as mobile keyboard prediction, illustrate how frequent small updates and on-device personalization can improve user experience [22], but clinical use imposes stronger traceability and regulatory demands. Healthcare informatics frameworks emphasize that federated learning must align with existing electronic health record workflows, identity management, and audit controls [23]. Ethical analyses similarly stress that algorithmic transparency, informed consent, and public accountability are not optional additions but constitutive requirements [24].

6. Sustainability and Policy Implications

Sustainability in federated medical image analysis is a multi-dimensional concern. It includes energy consumption, hardware longevity, algorithmic efficiency, and institutional capacity. Selective state space models may lower some computational costs compared with attention-based models, but federated training involves repeated communication rounds and multiple local epochs. If each round requires all hospitals to rehearse large volumes of data, the aggregate carbon footprint can exceed centralized training. A federated Mamba system should therefore adopt adaptive round policies, convergence-based stopping, and lightweight local

models where appropriate. Reusing previously trained state space encoders across tasks can amortize environmental costs, but transferability in medical imaging is constrained by domain shift across scanners and populations.

Policy implications extend beyond privacy law. Medical device regulators increasingly expect machine learning systems to be safe, reproducible, and robust across deployment conditions. Federated systems complicate regulatory evaluation because the final model is produced by a distributed process without a stable training dataset. Mamba-based encoders add architectural novelty that may require additional validation evidence, especially for long-range feature propagation in high-resolution imaging. Policies should encourage model registries, versioned training protocols, and third-party auditing. Data protection authorities may accept differential privacy as a safeguard, but the interpretation of sufficient privacy budgets varies across jurisdictions. International federations must therefore harmonize privacy accounting practices and document the residual re-identification risk.

Cross-domain comparison with consumer federated learning is instructive. Mobile keyboard systems have shown that model updates can occur on millions of devices with personalized local adaptation [22], but medical imaging operates under different incentives and failure costs. In consumer settings, occasional model degradation may be tolerated; in clinical decision support, degradation can harm patients. This asymmetry demands more conservative aggregation rules, stricter change management, and deeper human oversight. Federated learning in healthcare also intersects with public health surveillance, where models may be repurposed beyond their original diagnostic scope. Future policy frameworks should separate routine clinical quality improvement from secondary research use, requiring explicit institutional review and public engagement. The architecture should facilitate such governance by exposing metadata about data provenance, update statistics, and performance across sites.

7. Conclusion

Federated Mamba learning represents a promising direction for privacy-preserving collaborative medical image analysis, but its value depends on more than architectural novelty. Selective state space encoders can reduce some attention-related computational costs and capture long-range diagnostic context, yet they introduce new challenges in privacy accounting, gradient leakage, interpretability, and cross-site standardization. This paper has argued that federated medical imaging systems should be treated as governed infrastructures rather than isolated machine learning experiments. The structural boundaries between local encoders, global aggregation, privacy mechanisms, and clinical deployment must be designed jointly. Differential privacy, secure aggregation, audit logging, and fairness monitoring are not peripheral safeguards; they are constitutive elements of a trustworthy federation. Future work should develop reference architectures that combine tiered compute, bounded asynchrony, energy-aware training, and regulatory documentation. Such systems can support multi-institutional medical imaging research while respecting patient confidentiality and the operational realities of heterogeneous healthcare institutions.

References

1. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 1273–1282.

2. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., ... & Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, 10(1), 12598.
3. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ Digital Medicine*, 3, 119.
4. Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.
5. Dao, T., & Gu, A. (2024). Transformers are SSMS: Generalized models and efficient algorithms through structured state space duality. *Proceedings of the 41st International Conference on Machine Learning*, 22041–22090.
6. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., & Liu, Y. (2024). VMamba: Visual state space model. *arXiv preprint arXiv:2401.10166*.
7. Wang, Z., Zheng, J.-Q., Zhang, Y., Cui, G., & Li, L. (Wang, Z., Zheng, J.-Q., Zhang, Y., Cui, G., & Li, L. (2024). Mamba-UNet: Unet-like pure visual mamba for medical image segmentation. *arXiv preprint arXiv:2402.05079*.
8. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
9. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
10. 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*, 308–318.
11. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 1175–1191.
12. 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.
13. 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*, 3–18.
14. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407.
15. Zheng, X., Chen, X., Gong, S., Griffin, X., & Slabaugh, G. (2025, September). Xfmamba: Cross-fusion mamba for multi-view medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 672–682). Cham: Springer Nature Switzerland.

16. 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*, 2, 429–450.
17. Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., ... & Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27(10), 1735–1743.
18. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Lungren, M. P. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
19. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35.
20. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650.
21. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626.
22. Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., ... & McMahan, H. B. (2018). Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*.
23. 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.
24. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.