Mamba-Based Cross-Domain Adaptation for Robust Medical Image Classification

Eero C. Bush
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
eero.bush12@unr.edu

Henri Day
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
henri1992@buffalo.edu

Henrik Park
School of Computing, Clemson University, Clemson, SC, USA.
henrikpark386@clemson.edu

Abstract

Deep learning has substantially improved medical image classification, yet performance frequently degrades when models trained on controlled datasets are applied across different scanners, institutions, patient populations, and acquisition protocols. Transformer-based architectures have increased representational capacity, but their quadratic attention cost and data requirements present significant obstacles for resource-constrained clinical environments. Recent state space models, particularly the Mamba family, offer a promising alternative by combining linear-time sequence modeling with selective state transitions that can capture long-range dependencies in high-dimensional inputs. This paper presents a system-level examination of Mamba-based cross-domain adaptation for robust medical image classification. It analyzes how selective state space mechanisms can be integrated with adversarial, discrepancy-based, and self-supervised domain adaptation principles to improve resilience across heterogeneous imaging environments. Rather than focusing solely on model-level accuracy, the paper addresses architectural trade-offs, data governance, infrastructure design, deployment sustainability, interpretability, fairness, and regulatory policy implications. We argue that robust clinical adoption requires a sociotechnical systems perspective in which adaptation is treated as a property of the full pipeline rather than a standalone algorithmic fix. The paper identifies important research directions including calibration under distribution shift, energy-efficient inference, federated and decentralized adaptation, and alignment with evolving regulatory frameworks. Overall, the work provides a forward-looking framework for integrating Mamba-based sequence modeling with cross-domain adaptation to support more resilient, equitable, and sustainable medical image classification.

Keywords

medical image classification; domain adaptation; state space models; Mamba; robust artificial intelligence; clinical deployment.

1. Introduction

Medical image classification has been transformed by deep learning, moving from handcrafted radiomic features to learned hierarchical representations. Early advances were

dominated by deep convolutional networks, whose translation equivariance and local feature hierarchies proved effective for chest radiographs, pathology slides, and retinal images [1]. Residual learning further stabilized optimization and enabled very deep architectures [2]. More recently, vision transformers introduced global self-attention and demonstrated that long-range spatial dependencies can improve classification performance when sufficient data are available [3]. However, attention mechanisms scale quadratically with sequence length, which is particularly burdensome for high-resolution volumetric medical data and multispectral imaging. State space models, especially the Mamba architecture, have emerged as an alternative that supports linear-time sequence modeling while preserving selective state dynamics [4]. In medical image analysis, transformer-based methods have been extensively surveyed and show growing adoption across segmentation, reconstruction, and classification [5]. A recent cross-fusion Mamba design for multi-view medical image classification further indicates that state space models can aggregate information across complementary views [6]. Despite these advances, robustness across clinical domains remains a major obstacle.

In practice, models trained on one hospital or scanner frequently exhibit performance loss when applied to another institution with different acquisition protocols, patient demographics, or disease prevalence. Domain adaptation addresses this problem by aligning feature distributions, reweighting samples, or learning domain-invariant representations. Adversarial training approaches have been widely adopted to encourage feature extractors to produce representations that confuse a domain discriminator [7], [8]. Theoretical treatments of domain adaptation emphasize the importance of controlling both source risk and divergence between source and target distributions [9]. These ideas have direct relevance for medical image classification because annotated medical data are often scarce, siloed, and heterogeneous. The purpose of this paper is not to advance a single benchmark result, but rather to examine how Mamba-based state space models can be coupled with cross-domain adaptation principles to produce systems that are more robust, deployable, and governable in real clinical settings. The remainder of this paper develops a system-level perspective on architecture, infrastructure, fairness, sustainability, and policy.

2. Background and Related Work

Transfer learning has been a default strategy for medical imaging because large natural image datasets can pretrain low-level visual representations, but evidence suggests that this transfer is not always equally beneficial across medical tasks [10]. Self-supervised learning has recently been explored as a way to build representations from unlabeled medical images, reducing reliance on scarce expert annotations [11]. Contrastive frameworks such as SimCLR [12] and momentum contrast [13] provide strong foundations for learning invariance to augmentations, although their success depends heavily on batch composition, augmentation policy, and dataset diversity. These methods are particularly relevant to cross-domain adaptation because they can learn representations that are less tightly coupled to the superficial characteristics of a single training distribution.

At the infrastructure level, the computational cost of global attention has motivated memory-efficient primitives and hierarchical vision transformers. FlashAttention reduces memory movement through IO-aware exact attention [14], while Swin Transformer introduces shifted window attention to improve scalability [15]. These advances matter for medical imaging because volumetric inputs and large population screening programs create severe resource demands. In parallel, standardized pipelines such as nnU-Net have shown how careful configuration can outperform bespoke architectures for biomedical segmentation [16]. Public

datasets such as CheXpert [17] and MIMIC-CXR [18] have accelerated research by providing large chest radiograph collections with expert-derived labels, but they also reflect specific acquisition environments that may not generalize across sites. These datasets illustrate the domain heterogeneity that persists even when data are publicly available.

3. Mamba-Based Cross-Domain Adaptation Architecture

Selective state space models process inputs through recurrent discretization, but they can be unfolded as linear-time global context aggregation. Visual Mamba variants adapt this mechanism to two-dimensional images by scanning patches along multiple directions [19]. The architecture avoids quadratic attention while retaining global contextual sensitivity. Efficient transformer surveys [20] highlight the broader trade-offs between expressivity and scalability, and these trade-offs are equally relevant to state space models in medical applications. Mamba-based encoders can be designed to process large images without the excessive memory consumption associated with dense attention maps, making them attractive for high-resolution chest radiographs, digital pathology slides, and multimodal imaging.

A Mamba-based cross-domain adaptation system can be organized into three interconnected components: a visual state space encoder, a domain-invariant projection head, and a task classifier. The encoder produces sequence embeddings by scanning image patches in multiple orientations, allowing local texture and long-range anatomical context to interact without dense attention maps. The projection head may incorporate gradient reversal layers to reduce domain discriminability [7], or it may optimize explicit discrepancy measures between source and target feature distributions. The classifier and domain head are trained jointly so that representations remain sensitive to disease-relevant variation while being insensitive to scanner-related nuisance factors. Structural trade-offs arise between selectivity and invariance: strong domain confusion can remove clinically relevant variation, while weak alignment leaves spurious correlations intact [9]. A cross-fusion Mamba approach [6] illustrates how multiple views can be fused within this state space paradigm, although the broader challenge is to ensure that cross-domain robustness does not come at the cost of fine-grained disease localization.

Mamba's selective scan mechanism provides a middle ground between recurrent and attention-like processing. Bidirectional or multidirectional scans are necessary for images because medical findings are not naturally ordered along a single axis. Multiple scanning paths allow the model to aggregate contextual evidence from horizontal, vertical, diagonal, and patch-sequential views. This design introduces a structural tension between memory consumption and representational coverage: each additional scan path increases state dimensions and activations, but can improve robustness to orientation and crop variations. System designers must balance these costs against inference latency and memory availability in hospital picture archiving and communication systems. The exact configuration should therefore be guided by the target deployment environment rather than by accuracy metrics alone.

4. Cross-Domain Robustness and Structural Trade-offs

Robustness in medical image classification is not a monolithic property. It encompasses resilience to scanner variability, patient demographic shifts, acquisition protocol changes, label distribution changes, and temporal drift in clinical practice. Domain adaptation commonly focuses on aligning marginal feature distributions, but this strategy is insufficient when the relationship between images and labels also changes across sites. Some hospitals

may use portable radiographs more frequently for critically ill patients, while others rely on upright posteroanterior images for ambulatory screening. These differences can create strong but misleading associations between image appearance, patient positioning, and disease severity. Mamba-based models that capture global context may unintentionally encode hospital-specific furniture, laterality markers, or tube positioning as predictive features. Adversarial domain adaptation can suppress these features, but excessive suppression may remove legitimate signs of acute illness.

A useful case illustration arises in multi-site chest radiograph classification using datasets such as CheXpert and MIMIC-CXR [17], [18]. These datasets differ in patient mix, annotation procedures, and imaging equipment. A model trained on CheXpert may learn subtle correlations involving portable imaging equipment, while MIMIC-CXR reflects intensive care workflows and a different prevalence of acute findings. Domain adaptation with a Mamba encoder can reduce these gaps by learning scan paths that emphasize anatomical context rather than equipment-related cues. However, the system must be carefully validated across subgroups because aggregate improvement can mask worse performance for rare presentations or underrepresented demographic groups. Robustness should therefore be evaluated with stratified audits, uncertainty quantification, and subgroup-level calibration rather than global accuracy alone.

The integration of self-supervised pretraining with Mamba architectures introduces another structural trade-off. Self-supervised methods such as SimCLR and momentum contrast [12], [13] can learn from large unlabeled image collections, but their augmentation policies are often designed for natural images. Medical images require augmentation strategies that preserve diagnostic content while simulating realistic domain variation, including blur, noise, contrast shift, rotation, and equipment-specific artifacts. Pretraining with overly aggressive transformations can degrade disease sensitivity, whereas conservative transformations may fail to build sufficient invariance. Mamba encoders can be pretrained on unlabeled medical image corpora and then adapted with a domain-confusion head. This combination may reduce dependence on scarce expert labels, but it also increases the complexity of system verification because the pretraining distribution may not be fully documented.

5. Infrastructure and Deployment Considerations

Clinical deployment of Mamba-based classifiers requires integration with existing radiology and pathology informatics infrastructure. Most hospital systems rely on DICOM standards, picture archiving and communication systems, and worklist-based review workflows. A classifier must operate within these systems without introducing unacceptable latency or storage overhead. Mamba's linear-time sequence modeling offers theoretical advantages for large images, but practical throughput depends on implementation maturity, memory bandwidth, batch composition, and hardware support. Memory-efficient primitives such as FlashAttention [14] have shown that IO-aware design can substantially reduce runtime costs for attention-based models, and similar engineering effort is needed for state space models in production settings.

Deployment modes can be broadly divided into on-premises inference, cloud-based inference, and edge inference on local devices. On-premises deployment is often preferred for privacy and regulatory reasons, but it constrains the available GPU memory and energy envelope. Mamba models with large state dimensions may not fit comfortably on older hospital servers. Cloud-based deployment offers elasticity but introduces data transfer risks, dependency on external infrastructure, and variable operational costs. Edge deployment is appealing for

point-of-care ultrasound and low-resource screening, but it requires aggressive quantization, distillation, or architectural simplification. These choices are not purely technical; they reflect institutional risk tolerance, procurement cycles, and the availability of machine learning engineering staff.

Monitoring is another critical infrastructure concern. A deployed classifier should be continuously evaluated for drift in input distribution, label semantics, and clinical workflow. Domain adaptation can be updated when new site data become available, but repeated retraining may destabilize the system if not governed by version control, shadow deployment, and automated rollback procedures. Mamba-based encoders can support lightweight fine-tuning on new domains, but the system should retain source-domain performance through rehearsal buffers, elastic weight consolidation, or adapter modules. In all cases, deployment decisions should be documented with the same rigor as the underlying model training, because unobserved infrastructure failures can create patient safety risks even when the model achieves strong offline metrics.

6. Governance, Fairness, and Policy Implications

Fairness in medical image classification requires more than balanced accuracy across demographic groups. Bias can enter through dataset composition, label noise, acquisition differences, and the selection of adaptation objectives [21]. A Mamba-based domain adaptation system may align source and target hospitals but still fail for smaller subgroups within a site. For example, pediatric patients, patients with rare anatomical variants, or patients with overlapping conditions may be underrepresented in both source and target training data. If domain alignment optimizes aggregate error, it can erase the very variations that are clinically important for these subgroups. Governance processes should therefore mandate subgroup reporting, stratified performance audits, and explicit thresholds for clinically meaningful differences.

Interpretability is a further governance requirement. While post-hoc visual explanations such as gradient-based localization [23] can help radiologists inspect where a model focuses, they do not fully capture the internal state transitions of a Mamba encoder. Intrinsically interpretable models have been advocated in healthcare because they permit clinicians to understand and contest predictions [22]. Mamba models are not inherently interpretable, and their bidirectional scan mechanisms may be difficult to summarize in a simple saliency map. System designers should therefore pair them with uncertainty estimates, counterfactual explanations, or prototype-based interpretations when deployment carries high clinical stakes. Governance bodies increasingly expect documentation of data provenance, intended use, and failure modes before approving machine learning systems for clinical use.

Policy implications extend beyond local hospital governance. Regulatory frameworks increasingly emphasize risk-based classification of artificial intelligence systems, continuous performance monitoring, and human oversight. Mamba-based adaptation systems must be designed with these requirements in mind. This includes maintaining audit trails for training data, adaptation runs, and deployment decisions. It also includes establishing clear procedures for decommissioning a model when drift exceeds acceptable bounds. In federated or decentralized settings, where models are updated across institutions without sharing raw images, governance becomes more complex because adaptation may occur outside a single institutional data controller. Cross-domain adaptation should therefore be accompanied by institutional data-sharing agreements, model cards, and transparent reporting of domain performance.

7. Sustainability and Lifecycle Management

Sustainability in medical artificial intelligence involves computational energy, hardware lifecycle, data storage, and the human labor required for annotation, maintenance, and oversight. Mamba architectures may reduce inference energy compared with large transformer models because they avoid dense quadratic attention. However, this benefit can be offset by the need for multiple scan paths, large hidden state dimensions, and repeated adaptation experiments. A system-level energy budget should account not only for final inference but also for pretraining, hyperparameter sweeps, domain alignment, and periodic retraining. Pretraining strategies based on self-supervision can amortize representation learning across many downstream tasks, but they require substantial computational investment up front [11].

Lifecycle management includes planning for dataset updates, scanner replacements, protocol changes, and evolving disease definitions. A model that performs well on current imaging equipment may degrade when a hospital upgrades from computed radiography to digital radiography or changes reconstruction kernels. Mamba-based adaptation can accommodate these changes through fine-tuning, but frequent fine-tuning can create model sprawl and version proliferation. Institutions should maintain a registry of deployed model versions, the domain distributions they were adapted to, and the evidence supporting their continued use. Sustainable practices also involve sharing adaptation metadata across institutions when privacy constraints permit, because this can reduce redundant computation and improve the collective robustness of the clinical artificial intelligence ecosystem.

Energy efficiency is increasingly relevant to policy and procurement. Hospitals and public health systems may prefer models that can run on existing hardware without requiring specialized accelerators. Mamba-based systems can be competitive in this regard if their scan implementations are optimized and their state dimensions are kept moderate. However, the computational advantage should not be overstated until more standardized benchmarks compare state space models, hierarchical transformers, and convolutional networks under equivalent clinical throughput constraints. Researchers should report runtime, memory use, and energy consumption alongside accuracy to enable meaningful sustainability comparisons.

8. Future Research Directions

Future work should investigate calibration and uncertainty estimation under domain shift. A Mamba-based classifier may produce plausible but overconfident predictions when the target domain differs from the source. Calibration methods, evidential learning, and selective prediction mechanisms can provide safer deployment by identifying cases that require human review. Domain adaptation should be evaluated not only by aggregate accuracy but also by expected calibration error across hospitals and demographic subgroups.

Federated and decentralized domain adaptation represents another important direction. In many settings, medical images cannot be centralized due to privacy, legal, or institutional barriers. Mamba encoders may be well suited to federated fine-tuning because their linear-time computation can operate on local devices, but the aggregation of state space parameters and domain-invariant heads requires careful coordination. Research should examine how to balance local adaptation with global robustness, how to prevent catastrophic forgetting, and how to audit federated models across sites without exposing raw data.

Multimodal adaptation is also promising. Medical classification often benefits from combining images with structured clinical data, laboratory results, and textual reports. Mamba

architectures originally developed for sequence modeling can naturally ingest multiple modalities as sequential tokens, enabling cross-modal interactions that may improve robustness. However, multimodal systems introduce additional governance complexity because missing modalities, heterogeneous formats, and non-image features can create new bias pathways. Future research should address these issues within a unified systems framework rather than treating each modality in isolation.

Finally, stronger links between technical robustness and regulatory evaluation are needed. The clinical community would benefit from open benchmarks that simulate realistic domain shifts, including scanner changes, patient demographic shifts, and temporal drift. These benchmarks should include subgroup metrics, uncertainty measures, and resource constraints so that researchers can compare models on sociotechnical criteria rather than accuracy alone. Mamba-based adaptation should be evaluated in this broader context to ensure that its architectural benefits translate into meaningful clinical resilience.

9. Conclusion

This paper has examined Mamba-based cross-domain adaptation as a systems problem rather than a narrowly algorithmic task. Selective state space models offer a promising way to combine long-range context, linear-time sequence processing, and manageable computational costs for medical image classification. When integrated with domain adaptation principles, these architectures can improve robustness across heterogeneous clinical sites. However, structural trade-offs between invariance and disease sensitivity, between efficiency and representational coverage, and between scalability and interpretability must be managed through deliberate system design. Infrastructure integration, governance, fairness, sustainability, and policy alignment are equally important. Robustness emerges not from a single model component but from the interaction of architecture, data, deployment environment, monitoring, and institutional oversight. Mamba-based adaptation can contribute to more resilient medical imaging systems, but only if researchers and practitioners measure and govern it as part of a complete clinical pipeline.

References

1. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sanchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In International Conference on Learning Representations.

4. Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752. https://doi.org/10.48550/arXiv.2312.00752

5. Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., & Fu, H. (2023). Transformers in medical imaging: A survey. Medical Image Analysis, 88, 102802. https://doi.org/10.1016/j.media.2023.102802

6. 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.

7. Ganin, Y., & Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In Proceedings of the 32nd International Conference on Machine Learning (pp. 1180–1189). PMLR.

8. Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). Adversarial discriminative domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7167–7176). https://doi.org/10.1109/CVPR.2017.761

9. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Vaughan, J. W. (2010). A theory of learning from different domains. Machine Learning, 79(1–2), 151–175. https://doi.org/10.1007/s10994-009-5152-4

10. Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. In Advances in Neural Information Processing Systems (Vol. 32).

11. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., Natarajan, V., & Norouzi, M. (2021). Big self-supervised models advance medical image classification. Nature Biomedical Engineering, 5(12), 1290–1302. https://doi.org/10.1038/s41551-021-00776-1

12. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning (pp. 1597–1607). PMLR.

13. He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9729–9738). https://doi.org/10.1109/CVPR42600.2020.00975

14. Dao, T., Fu, D. Y., Ermon, S., Rudra, A., & Re, C. (2022). FlashAttention: Fast and memory-efficient exact attention with IO-awareness. In Advances in Neural Information Processing Systems (Vol. 35, pp. 16344–16359).

15. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10012–10022). https://doi.org/10.1109/ICCV48922.2021.00986

16. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

17. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R. L., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., & Ng, A. Y. (2019). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 1, pp. 590–597). https://doi.org/10.1609/aaai.v33i01.3301590

18. Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C., Mark, R. G., & Horng, S. (2019). MIMIC-CXR: A large publicly available database of labeled chest radiographs. Scientific Data, 6(1), 317. https://doi.org/10.1038/s41597-019-0322-0

19. 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. https://doi.org/10.48550/arXiv.2401.10166

20. Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2023). Efficient transformers: A survey. ACM Computing Surveys, 55(6), 1–28. https://doi.org/10.1145/3530811

21. 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. https://doi.org/10.1145/3457607

22. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1721–1730). https://doi.org/10.1145/2783258.2788613

23. 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. In Proceedings of the IEEE International Conference on Computer Vision (pp. 618–626). https://doi.org/10.1109/ICCV.2017.74