Generative Mamba Models for Data Augmentation in Low-Resource Medical Imaging

Grant Bergman
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
bergman1980@ku.edu

Ishaan Moran
Department of Computer Science, University of North Texas, Denton, TX, USA.
moran1975@unt.edu

Neil D. Ferguson
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
realneil@binghamton.edu

Abstract

Low-resource medical imaging environments face a persistent tension between the need for high-capacity deep learning models and the limited availability of expertly annotated data. Traditional augmentation methods can improve generalization but rarely introduce the semantic diversity required for rare diseases, underrepresented anatomies, and heterogeneous acquisition protocols. Generative models have been proposed as a response to this limitation, yet their deployment in clinical and low-resource settings is often constrained by computational cost, training instability, and governance complexity. This paper examines generative Mamba models as a system-level alternative for data augmentation in medical imaging. Mamba architectures use selective state space sequence modeling to provide efficient long-range dependency modeling, making them attractive for high-resolution image synthesis under constrained computational budgets. The discussion positions data augmentation not as an isolated algorithmic step but as an integrated socio-technical pipeline spanning data curation, generative pretraining, downstream validation, auditing, deployment, and regulatory oversight. The paper analyzes architectural trade-offs between adversarial, variational, diffusion, and state space generative approaches and examines how Mamba-based backbones may reduce memory and inference costs while preserving clinically meaningful variation. It further addresses robustness risks such as mode collapse, shortcut learning, demographic bias amplification, and synthetic data provenance. Infrastructure considerations include federated training, edge deployment, interoperability with clinical imaging systems, and sustainability. The paper argues that generative Mamba models can support low-resource medical imaging, but only when embedded within transparent, auditable, and clinically validated augmentation systems.

Keywords

generative state space models; Mamba; medical imaging; data augmentation; low-resource learning; algorithmic governance.

1. Introduction

Deep learning systems for medical image analysis have produced substantial improvements in standardized benchmark settings, but their clinical translation is frequently constrained by the uneven availability of annotated imaging data [1]. In low-resource environments, the problem is structural rather than purely technical. Hospitals may lack specialized radiologists, rare pathological examples may be geographically dispersed, and privacy regulations may prevent direct aggregation of patient records. Under these conditions, a model trained only on local data often fails to capture the variability needed for reliable decision-making, while a model trained exclusively on external public datasets may not generalize across population differences or equipment-specific imaging characteristics. The result is a persistent gap between the performance reported in controlled research settings and the operational reliability required in clinical practice.

Generative modeling has offered one promising route for reducing this gap. Early generative approaches such as generative adversarial networks [2] and variational autoencoders [3] introduced mechanisms for synthesizing new examples by learning the distribution of observed images. More recently, diffusion probabilistic models [4] have shown that iterative denoising can produce high-fidelity samples with more stable training dynamics than adversarial formulations. However, these models also bring substantial computational demands because they often rely on attention mechanisms that scale poorly with image resolution and sequence length. For institutions with limited computational infrastructure, training or even fine-tuning such models may be impractical. Moreover, the production of synthetic medical images introduces new risks that are not adequately addressed by purely algorithmic evaluations.

Recently, Mamba state space models [5] have emerged as a class of sequence models that can capture long-range dependencies while maintaining linear-time inference and more favorable memory scaling than comparable attention-based architectures. This efficiency makes Mamba attractive for medical imaging tasks in which high resolution and volumetric data create very long input sequences. Although Mamba has received considerable attention for language modeling and sequence analysis, its potential for generative data augmentation in low-resource medical imaging remains underexplored from a systems perspective. This paper therefore adopts an interdisciplinary view that connects model architecture with data governance, fairness, infrastructure, deployment, and regulatory oversight. The central claim is that generative Mamba models can provide meaningful augmentation benefits, but only when integrated into a carefully governed augmentation pipeline that accounts for data provenance, model limitations, and clinical risk.

2. Data Scarcity and Augmentation Systems in Medical Imaging

The performance of deep learning in medical imaging is tightly coupled to the quantity and quality of labeled data. Modern segmentation frameworks such as nnU-Net [6] demonstrate that well-configured convolutional networks can achieve strong performance when sufficient labels are available. Seminal demonstrations in retinal disease [7] and dermatologic classification [8] likewise relied on large curated datasets that are not representative of all clinical contexts. Public collections such as CheXpert [9], MIMIC-CXR [10], and HAM10000 [11] have lowered barriers for research, but they reflect particular institutional populations, acquisition protocols, and annotation practices. Consequently, models trained on these datasets may perform unevenly when transferred to low-resource settings with different equipment, patient demographics, or disease prevalence.

Image augmentation has traditionally been used to compensate for such limitations by expanding the effective training distribution through geometric transforms, intensity adjustments, and noise injection. Reviews of augmentation practice in medical imaging [12] and the broader survey literature on deep learning in medical image analysis [13] identify many useful transformations, but also emphasize their limitations. Geometric and photometric augmentations alter the appearance of existing cases without introducing new anatomical or pathological configurations. They cannot generate rare disease presentations that are absent from the original dataset, nor can they produce novel combinations of clinical features. Learned augmentation policies such as AutoAugment [14] can optimize the composition of transformations, but still operate on source images and do not create fundamentally new disease manifestations.

This limitation matters greatly in low-resource medical imaging. A hospital that possesses only a small number of positive cases for a rare condition cannot realistically expect conventional augmentation to cover the full spectrum of presentations. Generative augmentation is therefore attractive because it can, in principle, synthesize plausible examples that expand the semantic coverage of the training set. However, the introduction of synthetic data into clinical machine learning pipelines raises questions that extend far beyond image quality. The generation process must be traceable, the synthetic cases must be evaluated for realism and clinical coherence, and the downstream model must be monitored for biases that may be inherited or amplified from the generative model. For these reasons, data augmentation must be treated as a system-level intervention rather than a simple preprocessing step.

3. State Space Models and Mamba Architectures for Image Generation

State space models have a long history in signal processing and control theory, but their recent reformulation as deep learning sequence models has opened new possibilities for efficient long-range modeling. The Mamba architecture [5] introduces a selective state space mechanism that allows the model to adapt its state transitions according to the input, thereby retaining contextual sensitivity while avoiding the quadratic complexity of dense attention. Subsequent work on structured state space duality [15] has clarified the relationship between state space models and attention mechanisms, suggesting that many of the representational advantages of attention can be retained in more efficient sequence modeling frameworks. For medical imaging, where images are often represented as long sequences of patches or voxels, this efficiency is particularly relevant.

Mamba-based medical imaging models have begun to appear for segmentation and classification tasks. A segmentation architecture based on long-range sequential modeling has been proposed for three-dimensional medical volumes [16], demonstrating that Mamba can handle spatial dependencies across large volumetric data. Similarly, a cross-fusion Mamba approach for multi-view medical image classification has shown that state space models can integrate complementary views without excessive computational overhead [17]. These efforts are primarily discriminative, but they establish an important architectural foundation for generative Mamba models. By replacing or complementing convolutional and attention layers with selective state space blocks, generative models can process long spatial sequences more efficiently while preserving the ability to model global context.

From a generative perspective, Mamba can serve as the backbone within adversarial, variational, or diffusion formulations. Adversarial models offer fast sampling but may suffer from mode collapse and unstable training. Diffusion models provide high-fidelity samples

and more stable optimization, but their iterative sampling is computationally expensive. A Mamba-based denoiser can reduce the memory footprint of each diffusion step, making high-resolution generation more feasible in low-resource settings. Variational approaches can also benefit from Mamba encoders and decoders that capture long-range dependencies without the cost of full attention. The main architectural trade-off is therefore between sample diversity, sample quality, training stability, and computational efficiency. In low-resource medical imaging, this trade-off cannot be resolved by image quality metrics alone; it must also account for the ability to deploy the model within constrained institutional infrastructure.

4. System Design for Generative Augmentation Pipelines

A generative augmentation pipeline for low-resource medical imaging should be designed as an integrated system that begins with data curation and ends with monitored clinical deployment. The first stage involves assembling a foundation dataset from public sources, institutional archives, and federated partners. Public datasets such as CheXpert [9] and MIMIC-CXR [10] can provide broad pretraining coverage, while local data capture institution-specific variation. The generative Mamba model can be pretrained on the broader dataset and then fine-tuned on local examples, thereby reducing the number of local cases required for adaptation. This transfer strategy is particularly important in low-resource environments because it avoids the need to train large models from scratch.

The second stage is the generation and filtering of synthetic images. Not all generated samples are useful or safe. Some may contain unrealistic anatomical structures, others may exaggerate texture artifacts, and still others may encode shortcuts that confuse downstream classifiers. A robust pipeline therefore requires quality filters that combine quantitative metrics with domain expert review. Automated filters can detect out-of-distribution samples, measure anatomical consistency, and compare generated images against reference distributions. Human review remains essential when generated images are intended for rare disease simulation or when the downstream model will influence diagnostic decisions. The filtered synthetic cases can then be mixed with real cases using controlled ratios to avoid overfitting to generator-specific artifacts.

The third stage is downstream model training and validation. It is important to evaluate the downstream model on real held-out data rather than on synthetic examples alone. Synthetic augmentation may improve performance on rare classes, but it can also introduce subtle distributional shifts that are invisible in aggregate accuracy metrics. Subgroup analysis should be performed across demographic categories, acquisition devices, and disease subtypes. In addition, the augmentation pipeline should be versioned so that each downstream model can be linked to the exact synthetic data generation recipe. This traceability is critical for debugging failures, reproducing results, and meeting regulatory expectations. Without such governance mechanisms, generative augmentation can create a silent form of data contamination that undermines rather than improves clinical reliability.

5. Robustness, Fairness, and Clinical Risk Management

The introduction of generative Mamba models into medical imaging raises important robustness and fairness concerns. Model cards provide a structured way to document intended use, training data, evaluation results, and known limitations [18]. Such documentation is especially important for synthetic data pipelines because the provenance of generated cases affects their clinical validity. Surveys of bias and fairness in machine learning [19] indicate that even well-intentioned data augmentation can amplify existing demographic disparities if

the generative model is trained on unrepresentative source data. For example, if a dermatology generator is trained primarily on images from lighter skin tones, its synthetic outputs may reinforce the same bias when used to augment training data for a diagnostic model. Low-resource settings are particularly vulnerable because local data may be too small to detect these disparities during internal validation.

Algorithmic auditing frameworks [20] offer a starting point for evaluating generative augmentation systems through external or semi-external review. Such audits should examine the demographic distribution of generated images, the stability of downstream model predictions across subgroups, and the potential for synthetic cases to introduce clinically implausible features. Mode collapse is another risk that must be monitored. If the Mamba generator produces a narrow range of plausible-looking images, the downstream model may become overconfident about those specific patterns while failing on genuine clinical variation. Diffusion-based Mamba generators may reduce mode collapse compared with adversarial approaches, but they still require explicit diversity audits. Clinicians should be involved in evaluating whether generated images represent meaningful variation rather than superficial visual novelty.

Clinical risk management must also account for the possibility that synthetic images leak information from training data. Although generative models are often assumed to produce anonymous new samples, membership inference attacks and latent memorization can reveal individual patient features. This is a serious concern in medical imaging because images may contain identifying metadata or unique anatomical markers. Technical safeguards such as differential privacy and data minimization can reduce these risks, but they cannot eliminate them entirely. Hospitals and research institutions must therefore treat generative models as processing sensitive personal data and ensure that synthetic data generation complies with applicable privacy and regulatory requirements. The clinical risk profile of a model trained with synthetic augmentation is different from one trained only on real data, and this difference must be explicitly documented rather than hidden within large-scale performance metrics.

6. Infrastructure, Sustainability, and Deployment Models

The computational benefits of Mamba architectures have direct implications for deployment in low-resource medical imaging. High-performance medical imaging systems often assume access to centralized GPU clusters and stable data engineering teams [21], but many low-resource environments cannot support such infrastructure. Mamba-based generators reduce memory requirements relative to attention-based models, enabling smaller institutions to fine-tune generative models on a single modern GPU rather than requiring large multi-node systems. This efficiency also facilitates federated training across hospitals. In a federated configuration, the generative model can be trained locally on each institution's data without transferring raw images to a central server. The reduced communication and computational footprint of Mamba models makes this approach more feasible than federated training of large attention-based diffusion models.

Sustainability is another important system-level consideration. Training large generative models consumes significant energy and contributes to the carbon footprint of artificial intelligence research. The linear-time properties of Mamba can reduce the energy required for training and inference, but sustainability gains depend on the entire pipeline. If institutions repeatedly retrain generative models to address performance drift, the cumulative energy cost may offset the efficiency of the underlying architecture. Sustainable deployment therefore requires careful capacity planning, model reuse, and evaluation of whether generative

augmentation provides measurable clinical benefit. Smaller specialized models may be more sustainable than continual expansion of model size. Mamba's efficiency can be used to support a strategy of smaller, more frequent local fine-tuning rather than centralized training of ever-larger models.

Deployment models for generative augmentation span a continuum from centralized services to fully local implementations. A centralized generative service can maintain high-quality models and rigorous auditing, but it raises privacy concerns and depends on stable network connectivity. A fully local implementation preserves data sovereignty but may suffer from limited computational resources and less frequent model updates. Hybrid approaches can allow local fine-tuning on top of shared base models, with periodic synchronization of only model updates rather than raw images. Interoperability with clinical imaging systems is essential in all deployment models. Generated images must be stored and transmitted in standard formats, and the augmentation pipeline should integrate with existing picture archiving and communication systems as well as electronic health record workflows. Without such integration, even a technically effective generative Mamba model may remain unused because it does not fit within the daily routines of clinical teams.

7. Policy and Governance Implications

The use of machine learning in health care creates ethical, legal, and policy tensions that extend beyond technical performance [22]. Generative augmentation adds another layer of complexity because synthetic images blur the boundary between real patient data and artificially created information. Regulators may need to determine whether a generative model used to augment training data should be regulated as a medical device, as a development tool, or as part of the broader quality management system. In many jurisdictions, software that influences diagnostic decisions is subject to regulatory oversight, but the precise status of data augmentation components remains ambiguous. Clear guidance is needed on how synthetic data should be labeled, stored, and audited within clinical machine learning pipelines.

Transparency is a central policy requirement. Synthetic images should be accompanied by provenance metadata that records the generative model version, the conditioning variables, the filtering criteria, and the intended use. Model cards [18] and auditing processes [20] can support this transparency, but they require institutional commitment and regulatory incentives. Data protection law also interacts with generative augmentation. Even if generated images are not direct copies of patient data, their close resemblance to real cases may re-identify individuals or reveal sensitive health information. Privacy-preserving techniques should be considered part of the governance baseline, not optional enhancements. In cross-border collaborations, differences in data protection regimes can create legal obstacles that are only partially addressed by technical anonymization.

Finally, policy should address equity in low-resource settings. Generative Mamba models promise to reduce the data advantage held by large academic medical centers, but this promise can be undermined by unequal access to computational resources, pretrained models, and governance expertise. Research funders and health authorities may need to support shared benchmarking platforms, open evaluation protocols, and regional technical assistance programs. Such infrastructure would allow smaller institutions to evaluate whether generative augmentation actually improves local clinical outcomes rather than simply increasing model complexity. The long-term policy goal should be to ensure that generative augmentation supports safe, fair, and sustainable diagnostic services rather than accelerating the deployment

of opaque models. In this respect, Mamba is an enabling technology, not a governance solution. Its efficiency creates new possibilities, but the institutional and regulatory work remains essential.

8. Conclusion

Generative Mamba models offer a promising direction for data augmentation in low-resource medical imaging because they combine expressive long-range modeling with computational efficiency. This paper has argued that the value of such models depends on their integration within broader systems of data governance, clinical validation, infrastructure planning, and regulatory oversight. Mamba architectures can reduce the memory and energy costs of generative modeling, making synthetic augmentation more feasible for hospitals and research groups that lack large GPU clusters. However, efficiency alone does not guarantee clinical utility. Synthetic data must be filtered, audited, and evaluated for robustness across demographic and clinical subgroups. The risk of bias amplification, mode collapse, and privacy leakage must be managed through transparent documentation and independent auditing.

From a systems perspective, generative Mamba models should be viewed as one component of a carefully governed augmentation pipeline. The pipeline must connect public and local data sources, support fine-tuning under constrained resources, generate and filter synthetic images, train downstream models with traceable data mixtures, and monitor performance in real clinical settings. Infrastructure choices such as federated training and edge deployment are as important as model architecture. Policy frameworks must evolve to clarify the regulatory status of synthetic data and to ensure that low-resource institutions can access the benefits without inheriting disproportionate risk. Future work should develop standardized evaluation protocols, open benchmark suites, and governance templates that enable meaningful comparison between Mamba-based synthesis and alternative generative approaches. By addressing these structural dimensions, the medical imaging community can move beyond isolated performance gains and toward augmentation systems that are trustworthy, sustainable, and responsive to the realities of low-resource care.

References

1. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1-48.

2. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Advances in Neural Information Processing Systems 27 (NeurIPS 2014).

3. Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. In Proceedings of the International Conference on Learning Representations (ICLR).

4. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In Advances in Neural Information Processing Systems 33 (NeurIPS 2020).

5. Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752.

6. 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.

7. De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., ... Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine, 24(9), 1342-1350.

8. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

9. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., 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. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 590-597.

10. Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C.-Y., Mark, R. G., & Horng, S. (2019). MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data, 6(1), 317.

11. Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5, 180161.

12. Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology, 65(5), 545-563.

13. 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., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88.

14. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2019). AutoAugment: Learning augmentation strategies from data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 113-123).

15. Dao, T., & Gu, A. (2024). Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In Proceedings of the 41st International Conference on Machine Learning (ICML).

16. Xing, Z., Ye, T., Yang, Y., Liu, G., & Zhu, L. (2024). SegMamba: Long-range sequential modeling Mamba for 3D medical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 578-588). Springer.

17. 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.

18. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT) (pp. 220-229).

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. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT) (pp. 33-44).

21. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.

22. Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in health care: Ethics, law, and policy. JAMA, 319(13), 1359-1360.