Hybrid CNN-Mamba Architecture for Automated Detection of Pulmonary Abnormalities in Chest X-Ray Images

Zhudan Du
School of Computing, Clemson University, Clemson, SC, USA.
zhudandu792@clemson.edu

Florian J. Wells
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
florian.wells@colostate.edu

Krish Shukla
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
krish.shukla27@unr.edu

Abstract

Chest X-ray imaging remains one of the most widely used diagnostic modalities for pulmonary disease, yet the rapid and reliable interpretation of these images continues to strain radiology services worldwide. Automated detection systems based on convolutional neural networks have demonstrated strong local feature extraction, but they often struggle to model long-range anatomical and pathological dependencies. Transformer-based architectures address some of these limitations but introduce significant computational and memory burdens. This paper presents a hybrid CNN-Mamba architecture for automated detection of pulmonary abnormalities in chest X-ray images. The proposed system combines local convolutional feature learning with selective state space sequence modeling to capture both fine-grained textural detail and distributed contextual relationships across the lung fields. Rather than focusing exclusively on classification accuracy, this paper provides a system-level analysis of the architectural trade-offs, data governance requirements, training strategies, robustness, fairness, interpretability, deployment constraints, sustainability, and policy implications associated with the hybrid design. The discussion emphasizes that a clinically viable pulmonary abnormality detection system must be evaluated not only as a predictive model but also as a socio-technical infrastructure embedded within existing radiology workflows, regulatory frameworks, and institutional governance structures. The hybrid CNN-Mamba approach offers a promising balance between representational capacity and operational efficiency, but its translation into practice depends on rigorous external validation, transparent uncertainty communication, equitable performance across populations, and alignment with evolving medical device oversight mechanisms.

Keywords

pulmonary abnormality detection; chest X-ray; hybrid deep learning; state space models; Mamba; medical imaging systems.

1. Introduction

Chest radiography is the most frequently performed imaging examination in many health systems and serves as a first-line tool for identifying pneumonia, nodules, effusions, pneumothorax, and other pulmonary abnormalities. The global volume of chest X-ray studies has accelerated demand for computational assistance, especially in under-resourced settings where specialist interpretation may be delayed or unavailable. Large public datasets such as ChestX-ray14 [1], CheXpert [2], and MIMIC-CXR [3] have enabled extensive experimentation with deep learning methods. Early automated systems demonstrated that convolutional networks could achieve radiologist-level pneumonia detection in carefully curated research settings [4]. However, translating these results into dependable clinical infrastructure remains difficult because pulmonary abnormalities often present with localized opacities that are embedded within diffuse lung patterns, requiring simultaneous attention to fine detail and broad anatomical context.

The architectural choices made for automated chest X-ray interpretation are therefore not purely technical. They determine how systems respond to variation in patient positioning, image quality, demographic heterogeneity, and rare or co-occurring conditions. Convolutional neural networks provide robust local feature extraction and translational equivariance, but their receptive fields grow gradually with depth and can miss long-range dependencies. Attention-based models can capture global context but impose quadratic computational complexity that becomes problematic for large radiology datasets. Selective state space models represent an alternative sequence modeling paradigm with linear complexity and strong performance in language and visual tasks. A hybrid architecture that couples convolutional local feature learning with Mamba-based long-range modeling offers a structurally coherent approach to pulmonary abnormality detection.

This paper examines the hybrid CNN-Mamba design from a systems perspective. It addresses architectural structure, data infrastructure, training and evaluation trade-offs, robustness, fairness, interpretability, deployment, sustainability, and policy implications. The central argument is that model architecture is one component of a broader clinical machine learning system. Governance mechanisms, data provenance, human-machine interaction, and institutional accountability shape whether an accurate model becomes a safe and effective diagnostic intervention. The following sections develop this perspective through detailed conceptual analysis, case illustrations, and cross-domain comparisons.

2. Background and Related Work

Convolutional networks have formed the backbone of medical image analysis for more than a decade. U-Net demonstrated that encoder-decoder architectures with skip connections could preserve spatial detail while aggregating high-level semantic information in biomedical segmentation tasks [5]. Deep residual networks further enabled the training of very deep architectures by stabilizing gradient flow and improving generalization across large image classification benchmarks [6]. These convolutional designs remain attractive for chest X-ray analysis because they efficiently learn texture, edge, and local shape features that are important for identifying opacities and structural abnormalities.

More recently, vision transformers introduced attention-based sequence modeling to image analysis and showed that global token interactions can improve representation learning when sufficient data are available [7]. However, self-attention scales quadratically with sequence length, which presents practical challenges for high-resolution medical images and large-scale training pipelines. Selective state space models were proposed to address these limitations by introducing sequence modeling with linear complexity and input-dependent gating [8]. This

formulation allows efficient propagation of information over long sequences while maintaining sensitivity to current inputs. Cross-fusion approaches for multi-view medical image classification have further demonstrated that state space representations can integrate complementary views without the computational burden of full attention [9].

Several visual adaptations of state space models have since emerged. Bidirectional state space encoders have been used to improve visual representation learning across image recognition benchmarks [10]. Visual state space models with cross-scan mechanisms have been proposed to address the spatial locality challenges that arise when applying sequence models to two-dimensional images [11]. Within biomedical image segmentation, Mamba-based U-shaped architectures [12] and pure Mamba UNet variants [13] have shown that state space models can rival or exceed convolutional and transformer-based models for medical imaging tasks. These developments suggest that state space modeling is not merely an alternative to attention but a complementary mechanism for capturing distributed context in medical images.

Despite these advances, chest X-ray abnormality detection presents specific structural challenges that are not fully addressed by generic visual state space models. Pulmonary findings vary in size, density, and location, and clinically important patterns such as interstitial disease often require the integration of subtle bilateral cues. A hybrid architecture that preserves convolutional local feature extraction while leveraging Mamba-based long-range modeling can be better suited to these requirements than either approach alone.

3. System Architecture and Design Rationale

The proposed architecture consists of three main processing stages: local feature extraction, long-range context modeling, and decision fusion. The first stage uses a convolutional stem and a sequence of residual blocks to extract local textural and edge features from the chest radiograph. This stage is designed to capture small opacities, vascular markings, and boundary information that are essential for detecting localized abnormalities. The second stage transforms feature maps into visual sequences and processes them with a bidirectional state space model. This stage enables the system to relate findings across distant regions, such as comparing the left and right lung fields or integrating apical and basal cues. The third stage combines multi-scale features and produces abnormality probabilities.

A central structural trade-off concerns the granularity of the sequence representation. Transforming high-resolution feature maps into very long sequences increases the capacity for fine-grained global reasoning but also increases memory consumption. Conversely, aggressive downsampling reduces computational load but may discard small lesions. In the proposed design, the sequence is constructed from intermediate feature maps rather than raw pixels, allowing the state space branch to reason over semantically meaningful regions while the convolutional branch retains high-resolution detail. This division of labor is conceptually similar to human visual interpretation, in which local inspection and global comparison operate in parallel.

The hybrid structure also influences hardware efficiency. Convolutional operations are highly optimized across accelerators and provide predictable throughput. State space models avoid the quadratic attention matrices that make transformer-based chest X-ray systems costly at scale. Although the recurrent formulation of state space models introduces sequential dependencies, recent implementations use parallel scans that improve training throughput. The result is a system that can be trained on large datasets with moderate infrastructure and deployed in environments with constrained computational resources.

Fusion strategy is another important design consideration. Early fusion combines convolutional and state space features before classification, while late fusion merges predictions after separate branches. Intermediate fusion at multiple resolutions tends to preserve complementary information but requires careful synchronization of feature dimensions and training dynamics. In the proposed system, multi-scale fusion is favored because pulmonary abnormalities operate at different spatial scales, and no single resolution is sufficient for all findings. The architecture should be understood as a flexible scaffold rather than a fixed configuration, allowing institutional adaptation to different acquisition protocols and clinical priorities.

4. Data Infrastructure and Governance

The performance of any pulmonary abnormality detection system depends on the quality, diversity, and governance of training data. Public datasets have accelerated research but they are not neutral research artifacts. ChestX-ray14 was constructed through automated extraction of radiology reports and contains noisy labels that reflect real-world documentation variability [1]. CheXpert introduced label uncertainty handling and expert comparison, but its labels are derived from structured report parsing and remain subject to systematic error [2]. MIMIC-CXR provides paired images and reports from an intensive care population and has enabled multimodal research, yet its patient distribution is shaped by the practices of a single institution [3]. These datasets must be understood as partial representations of clinical reality rather than comprehensive benchmarks.

Data governance for chest X-ray systems must address patient privacy, consent, and secondary use. Chest radiographs contain not only anatomical information but also markers of age, sex, body habitus, and occasionally external devices or artifacts that can contribute to reidentification. Institutional governance structures should require clear data use agreements, access controls, and audit logs. Federated learning infrastructures can allow model training across institutions without centralizing raw images, but they introduce new challenges in coordinating update frequency, ensuring data quality, and detecting distribution shift. Governance is therefore not a single technical solution but an ongoing organizational practice.

Label ontology design also has system-level consequences. A system trained to detect broad categories such as pneumonia or effusion may overlook findings that do not fit predefined labels. Hierarchical label structures, uncertainty-aware supervision, and radiology report embeddings can provide richer training signals. However, these approaches require harmonization across institutions because report style, language, and coding practices vary widely. Without such harmonization, models may learn documentation patterns rather than visual pathology, leading to brittle deployment performance.

5. Training, Optimization, and Evaluation Trade-offs

Training a hybrid CNN-Mamba model for chest X-ray interpretation requires balancing multiple objectives. The convolutional branch benefits from standard image augmentation, while the state space branch requires sequence construction that is invariant to spatial perturbations. Pretraining on large natural image datasets remains a common strategy, but domain adaptation to radiography is nontrivial because chest X-rays are grayscale, high-resolution, and structurally distinct from natural scenes. Self-supervised pretraining on radiographs can improve representation quality, especially when paired with report data or contrastive objectives.

Loss design for multi-label abnormality detection must account for class imbalance and label uncertainty. Common approaches include class weighting, focal objectives, and soft targets derived from report uncertainty. However, these strategies influence model calibration in different ways. A model optimized only for area under the receiver operating characteristic may produce poorly calibrated probabilities, which is problematic for clinical decision support. Evaluation should therefore include discrimination, calibration, and decision-curve analysis, even though such assessments require careful cohort definition and external validation.

External validation is particularly important because chest X-ray acquisition varies by equipment vendor, patient population, and imaging protocol. A model that performs well on research datasets may fail when deployed in a pediatric emergency department or a tuberculosis screening program. Cross-domain comparisons with state space models in language processing and time-series analysis suggest that sequence models adapt well to distribution shift when their inductive biases align with the underlying data. In medical imaging, however, such evidence is still emerging. Case illustrations from multi-site evaluations of pneumonia detection systems demonstrate that performance can degrade substantially when training data are dominated by adult inpatient radiographs and deployment occurs in outpatient or community settings.

6. Robustness, Fairness, and Interpretability

Interpretability is a core requirement for radiology decision support. Gradient-based visualization methods such as Grad-CAM provide coarse localization of image regions that influence predictions [14]. Shapley value explanations offer a complementary perspective by attributing prediction components to input features in a game-theoretic framework [15]. The hybrid CNN-Mamba architecture introduces additional interpretability design options because local convolutional attributions can be separated from long-range contextual influences. Clinically, this separation may help radiologists understand whether a prediction is driven by a localized opacity or by diffuse bilateral patterns. However, visual explanations can also create false confidence, and they should not be treated as definitive evidence of model reasoning.

Fairness in chest X-ray systems is a structural problem that begins with data collection and extends through model evaluation and clinical deployment. A landmark analysis of a widely used health management algorithm demonstrated that unequal access to care can be encoded into predictive models when proxy outcomes reflect systemic disparities [16]. Responsible machine learning frameworks for health care emphasize the need to define fairness in relation to clinical context, specify the groups that matter, and monitor performance across demographic and clinical subgroups [17]. Fairness is not achieved by simply removing protected attributes from inputs; biases can be embedded in image acquisition, report labels, and referral patterns [18].

Robustness to distribution shift is equally important. Chest X-ray models can be sensitive to changes in patient positioning, exposure, portable versus upright imaging, and postprocessing algorithms. Adversarial perturbations are less common in clinical practice than natural variation, but systematic differences across scanners and institutions can produce similar effects. Monitoring systems should track performance drift, alert when input distributions shift, and trigger recalibration or revalidation. The hybrid architecture may offer some robustness advantages because global state space modeling can compensate for missing or

corrupted local regions, but this hypothesis requires empirical evaluation across diverse imaging environments.

The broader evidence base for deep learning in medical imaging underscores both promise and caution. Big data and machine learning approaches have generated enthusiasm for clinical transformation, but high-performance claims in research settings do not automatically translate into improved patient outcomes [19]. High-performance medicine requires iterative integration of human and artificial intelligence, with careful attention to workflow and decision authority [20]. A systematic review and meta-analysis found that deep learning models can match health-care professionals in selected medical imaging tasks, yet the quality and external validity of studies remain highly variable [21]. Key challenges include prospective validation, usability, regulatory clarity, and the avoidance of automation bias [22]. Interpretability research in radiology further shows that explanation methods must be evaluated not only by technical metrics but also by their effect on clinician understanding and patient safety [23].

7. Deployment, Sustainability, and Policy Implications

Deployment of a hybrid CNN-Mamba system requires alignment with existing Picture Archiving and Communication Systems, radiology information systems, and clinical workflows. A model that operates as a standalone application may fail to influence care if it requires radiologists to leave their normal viewing environment or introduces additional cognitive burden. Integration should prioritize workload-aware triggering, prioritization of suspicious studies, and structured reporting templates. The computational efficiency of the hybrid architecture supports edge deployment in low-resource settings, but model compression, quantization, and device-specific optimization must be validated without sacrificing performance for rare abnormalities.

Sustainability has become an important system-level consideration. Transformer-based medical imaging models can consume substantial energy during training and inference, raising concerns about carbon footprint and operational cost in health systems with constrained budgets. The linear complexity of state space models and the efficiency of convolutional operations make the hybrid architecture more sustainable over its lifecycle. However, sustainability assessments should include data storage, network transfer, retraining frequency, and the environmental impact of hardware refresh cycles. A narrow focus on inference latency may overlook the larger footprint of continuous model monitoring and updating.

Policy implications extend beyond technical performance. Regulatory frameworks for medical device software increasingly require evidence of clinical benefit, risk management, postmarket surveillance, and human factors evaluation. A hybrid CNN-Mamba model used for pulmonary abnormality detection would likely be subject to such oversight because erroneous predictions can delay diagnosis or contribute to unnecessary procedures. Liability arrangements remain uncertain when automated systems influence clinical decisions, particularly if radiologists disagree with model outputs. Institutions should establish clear policies for escalation, overrides, and documentation of model-assisted findings.

Global equity is a further policy concern. Pulmonary abnormality detection systems have the potential to expand access to high-quality interpretation in regions with few radiologists, but only if they are trained and validated on representative populations and deployed with appropriate infrastructure. A model developed using high-income tertiary care data may not

generalize to tuberculosis screening or pediatric pneumonia in low-resource settings. System governance should therefore include community engagement, participatory design, and benefit-sharing mechanisms. Without these structures, the deployment of advanced architectures may reinforce existing health disparities rather than alleviate them.

The hybrid CNN-Mamba architecture illustrates how a technical design choice can carry broad system-level implications. Its local-global representational strategy is well matched to the diagnostic challenge of pulmonary abnormality detection, but its clinical value depends on governance, evaluation, and deployment practices that extend far beyond model architecture. Decision support systems are embedded in institutional power structures, professional norms, and regulatory regimes. Treating them as sociotechnical infrastructures requires sustained attention to fairness, transparency, accountability, and environmental responsibility across the entire model lifecycle.

8. Conclusion

This paper has presented a system-level analysis of a hybrid CNN-Mamba architecture for automated detection of pulmonary abnormalities in chest X-ray images. The design combines convolutional local feature extraction with selective state space sequence modeling to address the need for both fine-grained texture analysis and long-range anatomical context. This structural balance offers computational efficiency advantages over attention-based approaches while preserving the localized sensitivity that is critical for identifying small lesions and opacities. However, the paper has argued that architectural innovation alone is insufficient for clinical translation. Data governance, label quality, external validation, interpretability, fairness, robustness, deployment integration, sustainability, and policy alignment are equally constitutive of a dependable diagnostic system. The hybrid CNN-Mamba model should be understood as a flexible socio-technical platform that can be adapted to institutional workflows and population needs. Future work should prioritize prospective clinical evaluation, multi-site robustness studies, participatory governance frameworks, and life-cycle assessments that connect algorithmic design to patient outcomes and health system equity. Only through such integrated inquiry can automated pulmonary abnormality detection fulfill its potential as a safe, effective, and equitable component of modern radiology practice.

References

1. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2097-2106).

2. 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(01), 590-597.

3. 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-JPG, a large publicly available database of labeled chest radiographs. Scientific Data, 6, 317.

4. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225.

5. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241). Springer.

6. 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).

7. 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.

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

9. 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.

10. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., & Wang, X. (2024). Vision Mamba: Efficient visual representation learning with bidirectional state space model. In Proceedings of the 41st International Conference on Machine Learning (pp. 62449-62468). PMLR.

11. 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.

12. Xing, Z., Ye, T., Yang, Y., Liu, G., & Zhu, L. (2024). U-Mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722.

13. 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:2402.02486.

14. 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).

15. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

16. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

17. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337-1340.

18. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866-872.

19. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317-1318.

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

21. Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. The Lancet Digital Health, 1(6), e271-e297.

22. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine, 17, Article 195.

23. Reyes, M., Meier, R., Pereira, S., Silva, C. A., Dahlweid, F.-M., von Tengg-Kobligk, H., Summers, R. M., & Wiest, R. (2020). On the interpretability of artificial intelligence in radiology: Challenges and opportunities. Radiology: Artificial Intelligence, 2(3), e190043.