

Article

Not peer-reviewed version

Selective State-Space Models in Medical Image Processing

Ali Emre Gök^{*}, [Mustafa Yurdakul](#), [Şakir Taşdemir](#)

Posted Date: 14 April 2026

doi: 10.20944/preprints202604.0948.v1

Keywords: medical imaging; selective state space models; mamba architecture; deep learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Selective State-Space Models in Medical Image Processing

Ali Emre Gök¹, Mustafa Yurdakul² and Şakir Taşdemir³

¹ Nevşehir Hacı Bektaş Veli University, Department of Computer Engineering, Turkey

² Kırıkkale University, Department of Computer Engineering, Turkey

³ Selçuk University, Department of Computer Engineering, Turkey

* Correspondence: aliemregok@nevsehir.edu.tr

Abstract

In medical image analysis, modeling local and global features in high-resolution data presents a significant challenge. While the widely used Convolutional Neural Networks (CNNs) struggle to capture long-range dependencies between distant pixels, the high computational cost ($O(N^2)$) of Vision Transformer (ViT) architectures causes bottlenecks in clinical applications. This study investigates the integration of Mamba models which were developed to overcome these limitations and have linear complexity, into medical image analysis, along with recent studies in literature. This fundamentally continuous-time control theory-based architecture dynamically adapts to hardware resolution. The mamba models effectively retain anatomical structures and lesions in memory while filtering out irrelevant noise through their selective mechanism. Moreover, bidirectional scanning (Vision Mamba) and cross-scan (VMamba) methods are used to prevent the loss of spatial information and to overcome the necessity of processing one-dimensional data due to language-based structure of the models. The reviewed literature can be categorized under three main headings: hybrid models, efficient and lightweight designs, and spatial representation studies. Comprehensive analyses of literature indicate that Mamba models deliver significantly higher inference speed and memory efficiency compared to traditional CNN and ViT approaches owing to their hardware-aware design and linear computational efficiency. In conclusion, Mamba architecture has the potential to become a next-generation standard that demonstrates high performance while maintaining global contextual integrity across diverse medical fields such as radiology, ophthalmology, and dermatology.

Keywords: medical imaging; selective state space models; mamba architecture; deep learning

1. Introduction

Modeling local and global features in high resolution medical images presents a significant challenge. Although Convolutional Neural Networks (CNNs), one of the existing methods, successfully extract local details, they struggle to capture long-range dependencies between distant pixels (Jiang et al., 2026; C. Li et al., 2025). On the other hand, since the computational cost of Transformer (ViT) architectures that can capture global dependencies increases quadratically with image size, their use in high-resolution medical imaging is severely limited (Jiang et al., 2026; Y. Li et al., 2026).

In existing studies, approaches such as attention mechanisms or dividing images into small patches are used to overcome these limitations (Jiang et al., 2026; Zhu et al., 2024). Recently, Selective State Space Models (Selective SSM - Mamba) with linear computational cost are developed as an alternative to overcome complexity and prevent local-global information loss (Jiang et al., 2026; C. Li et al., 2025). The Mamba architecture was designed for processing images without bottlenecking and at high speed. In contrast to other deep learning architectures, it has a data-dependent selective state memory (Jiang et al., 2026; Zhu et al., 2024). This mechanism filters out unnecessary spaces and noises

in the images ensuring that only critical anatomical or pathological areas that need to be focused on are kept in memory (Gu & Dao, 2024; C. Li et al., 2025). Recent studies in literature prove that architectural approaches in medical field show a strong trend from CNNs to ViTs and finally to Mamba (Ruan et al., 2024). In this context, hybrid models are developed by integrating convolutional layers, attention modules, and ViTs between Mamba blocks to prevent information loss (C. Li et al., 2025; Y. Li et al., 2026; Zhou et al., 2026).

Furthermore, there is a growing focus on lightweight designs with reduced number of parameters that can operate in mobile platforms or limited resources and various scanning mechanisms that assure preservation of long-range dependencies between distant pixels (Jiang et al., 2026; Y. Li et al., 2026; Zhou et al., 2026).

2. Methodology

2.1. Mamba Architecture

State Space Model (SSM), is a mathematical framework developed to efficiently process sequence data, inspired by continuous dynamic systems (Wang et al., 2025; Yue & Li, 2024). Given the limitations of traditional models and Transformers architectures, development of Mamba, which is a selective SSM, brings out a strong alternative to these limitations. The fundamental operating principle of Mamba relies on a data-dependent selective mechanism that dynamically updates parameters based on input sequences (Gu & Dao, 2024; Ruan et al., 2024; Yue & Li, 2024; Zhu et al., 2024). This mechanism allows the model to filter out noises and retain only crucial information in its memory to ensure high performance on long sequences. In addition, Mamba architecture optimizes GPU memory hierarchy (SRAM and HBM) by using a hardware-aware algorithm. Thus, linear computational complexity is maintained while training and inference processes are speeding up (Gu & Dao, 2024; Hedhoud et al., 2025; Ruan et al., 2024; Yurdusever et al., 2025). Scheme of Mamba (Vision Mamba – bidirectional scan) architecture is given in Figure 1 (Zhu et al., 2024).

When it comes to the mathematical basis, Mamba uses continuous-time ordinary differential equations that map one-dimensional input sequence $(\mathfrak{X}(\mathfrak{T}))$ to output sequence $(\mathfrak{Y}(\mathfrak{T}))$ through a hidden state $(\mathfrak{H}(\mathfrak{T}))$. This relation is demonstrated via Equation (1) (Guo et al., 2025; Yue & Li, 2024).

$$h'(t) = Ah(t) + Bx(t) \quad (1)$$

$$y(t) = Ch(t) \quad (2)$$

In Equations (1) and (2), \mathfrak{A} denotes the state matrix while \mathfrak{B} and \mathfrak{C} denote projection parameters. This continuous system needs to be discretized to integrate it to Deep Learning environment. The Zero-Order Hold (ZOH) method is usually used in this regard (Ruan et al., 2024; Yue & Li, 2024; Yurdusever et al., 2025; Zhu et al., 2024).

$$\bar{A} = \exp(\Delta A) \quad (3)$$

$$\bar{B} = (\Delta A)^{-1} (\exp(\Delta A) - I) \cdot (\Delta B) \quad (4)$$

A timeline parameter (\mathfrak{T}) is added to the system via Equations 3 and 4. These are used to transform continuous \mathfrak{A} and \mathfrak{B} matrices into discrete forms. Finally, the system calculates the next state and output via discrete iteration formulas. The main difference of Mamba from other models is that the variables Δ , \mathfrak{A} and \mathfrak{B} are designed as functions dependent on time and input (Guo et al., 2025; Yue & Li, 2024; Zhu et al., 2024).

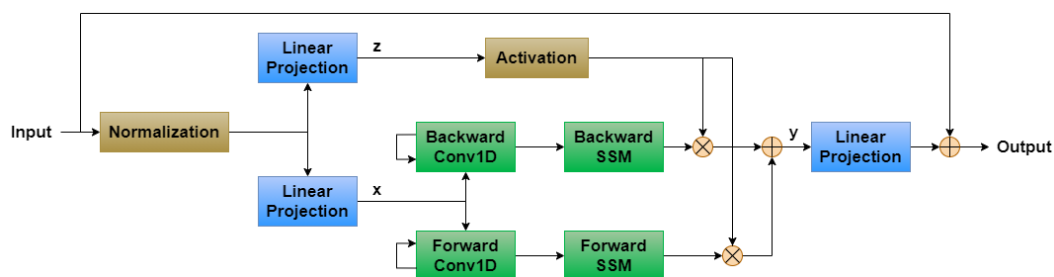


Figure 1. Mamba architecture (bidirectional scan).

2.2. Literature

The literature presented within the scope of this study is created by searching fundamental scientific databases such as Scopus, IEEE, Web of Science and ACM. During the search, the focus was the state-of-the-art studies of Mamba architectures in medical imaging published between 2024 and 2026 (Jiang et al., 2026; Y. Li et al., 2026; Ruan et al., 2024).

3. Findings

Integration of Selective State Space Models (Mamba) into medical imaging area allows researchers to overcome the computational bottleneck in processing high-resolution images and resulted in the establishment of a new generation of clinical applications. The studies inspected illustrated that this innovative architecture is not limited to a specific organ or data type. Mamba architecture is successfully applied in a wide range of fields such as neurological analyses, dermatological examinations, cellular-level pathology diagnoses, and radiological scans. In this section, current Mamba-based studies in literature are analyzed. The studies are divided into two main groups: clinical application areas and architectural approaches.

3.1. Clinical Application Areas

Mamba architecture provides specific and versatile solutions for many sub-branches of medical imaging. Tumor classification using MRI data in neurology and lesion detection using dermatoscopic images are examples of primary research domains. In the lungs, tuberculosis diagnosis and lesion analysis are performed while cell and parasite classification are performed in microscopic examinations. In addition, studies such as multi-organ segmentation and modality synthesis that allow transition between imaging types are noteworthy in literature. The clinical focus, data types used and key achievements of the studies examined are summarized in Table 1.

Table 1. Literature review of clinical studies.

Ref.	Clinical Focus	Imaging Type	Method	Results
(Lai et al., 2025)	Neurology (Brain Tumor)	MRI (Magnetic Resonance Imaging)	Vision Mamba	A significant improvement in multi-class classification accuracy, supported by transfer learning.
(S. Li et al., 2025)	Neurology (3D Tumor Detection)	MRI (BraTS Dataset)	3D MedicalDet-Mamba	High accuracy in 3D localization and object detection processes using the CNN-Mamba hybrid.
(Ruan et al., 2024)	Dermatology (Skin Lesion)	Dermatoscopes (ISIC)	VM-UNet	Competitive performance with the first U-shaped medical segmentation framework based on pure SSM.
(Zhong et al., 2025)	Dermatology and Gastroenterology	Dermatoscopy / Endoscopy	VMAXL-UNet	Modelling of correlations between distant lesions using xLSTM and Mamba integration.
(Hedho ud et al., 2025)	Respiratory Diseases (Tuberculosis)	Chest X-ray	Vision Mamba	80% lower GPU memory consumption compared to ViT models and an accuracy rate of 94.32%.
(Kumar & Mahen)	Respiratory Medicine (Lungs)	CT (Computed Tomography)	MedScope-LDx	A multi-stage analysis for the detection and classification of complex lung lesions.

dran, 2026)				
(Yurdu sever et al., 2025)	Microbiology (Parasite Analysis)	Microscopy	Vision Mamba (Vim-Base)	Thanks to hardware-aware design, 99.85% accuracy on an 8- class microscopic dataset.
(Zhou et al., 2026)	Pathology and Cell Analysis	Microscopy	GCNet-Mamba	State-of-the-art (SOTA) performance in blood cell classification (e.g., BloodMNIST).
(Zhang et al., 2026)	General Surgery (Multi-Organ)	CT / MRI (Synapse, ACDC)	Switch-UMamba	High-precision segmentation of complex tissues using the Dynamic (MoS) scanning strategy.
(Wang et al., 2025)	3D Volumetric Whole-Body Analysis	CT / MRI (AMOS, BraTS)	UlikeMamba	Outperforming Transformer architectures using depth-based convolution in 3D medical data.
(Yue & Li, 2024)	General Image Classification	Multi-modality	MedMamba	Superior performance across a wide range of organ and device data using the SS-Conv-SSM hybrid block.
(Lin et al., 2025)	Cross-Modality Image Synthesis	Conversion from MRI to CT	Mamba-Conv UNet	Successful cross-modal synthesis (e.g., generating CT images from MRI scans) to address data gaps.

3.2. Architectural Approaches

There are some novel approaches in literature to prevent data loss while adapting Mamba's 1D structure to 2D and 3D medical images. These approaches have been evaluated under three main headings: hybrid architectures, advanced scanning mechanisms, and cross-paradigm integrations. Hybrid architectures provide high accuracy with lower parameters by combining CNN's local feature extraction capability with Mamba's global context modeling. Advanced scanning mechanisms use multi-path or dynamic scanning algorithms to analyze undirected structure of medical images while preserving spatial integrity. Finally, in cross-paradigm integration studies, Mamba architecture is combined with other powerful models such as Generative Adversarial Networks (GAN), Diffusion and Kolmogorov-Arnold (KAN). The main purpose of these studies is to improve performance and reduce memory requirements in challenging situations. Methodological innovations of the studies inspected are given in Table 2.

Table 2. Literature review of architectural approaches.

Ref.	Model	Architectural Approach	Fundamental Innovation
(Ruan et al., 2024)	VM-UNet	Pure SSM (Pure Mamba)	The first fully SSM-based asymmetric U-shaped encoder-decoder architecture created without the use of CNNs or Transformers.
(Yue & Li, 2024)	MedMamba	CNN-Mamba Hybrid	An integrated 'SS-Conv-SSM' basic block that uses convolution for local features and SSM for global context.
(S. Li et al., 2025)	3D MedicalDet	CNN-Mamba Hybrid	The 'Locality-Integrated Mamba (LIM) module, which runs Mamba in parallel using multi-core convolutions.

(Zhang et al., 2026)	Switch-UMamba	Dynamic Scanning	The Mixture-of-Scans (MoS) mechanism, which considers scan directions as experts and dynamically selects the most suitable scan path for each data point.
(Jiang et al., 2026)	RMViM-Net	Multi-Path Scanning	'5D Multi-Path Scanning' operating in parallel subspaces and a graph-interactive attention module to enhance spatial modelling.
(Y. Li et al., 2026)	LGFVM-UNet	Local-Global Fusion	A VSS block supported by Dynamic Gating to prevent Mamba's global dominance from overwhelming local features.
(Su et al., 2025)	VMKLA-UNet	SSM + KAN Attention	Combining the VMamba encoder with a decoder featuring a linear attention mechanism based on the Kolmogorov-Arnold Network (KAN).
(Guo et al., 2025)	ABE-Mamba	SSM + GAN Integration	A cross-SS2D scanning block embedded within a Discriminator network for few-shot learning.
(C. Li et al., 2025)	MSM-Diff	SSM + Diffusion Model	Combining the capabilities of denoising diffusion models with 3D Multi-Scale Spatial Mamba (MS-Mamba).

4. Discussion and Conclusions

Although Mamba architecture offers an important advantage by reducing the quadratic computational cost of ViTs to linear level, there are some critical problems in clinical uses. The most fundamental architectural bottleneck is the local information loss during transformation of 2D and 3D medical data to 1D sequences. While researchers attempt to address this issue by using hybrid and multi-path scanning approaches, this can sometimes overshadow Mamba's main promise of reduced number of parameters.

On the other hand, the model's robustness and generalization capabilities are not yet fully optimized for real-world clinical applications. The findings of literature demonstrate that Mamba models may be vulnerable to sensor noises, targeted manipulations, and hardware errors. Furthermore, domain shifts in data distribution between "data obtained from different devices in different hospitals" a common occurrence in clinical practice. It negatively affects the cross-dataset performance of the model.

Future research is expected to progress towards lightweight Mamba models focused on edge computing that are directly optimized for the limited hardware capacities of clinical devices. At the same time, it is anticipated that multi-modality synthesis and few-shot learning strategies will be used in a much more integrated way with Mamba infrastructure to overcome lack of data problems for rare diseases.

In this review, recent methodological studies and clinical applications conducted using Selective State Space Models (Mamba) in medical imaging field are examined. Comprehensive analyses illustrate that Mamba architecture offers high accuracy and memory efficiency in a wide range of fields from neurology to dermatology. It is also seen that structural information loss is significantly reduced due to hybrid approaches and dynamic scanning mechanisms. Despite security and generalization challenges, Mamba architecture has the potential to serve as a cornerstone for future generations of clinical decision support systems thanks to its linear computational advantage and capability of modeling global context.

References

- Gu, A., & Dao, T. (2024). *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*. <http://arxiv.org/abs/2312.00752>
- Guo, B., Huang, W., & Wang, X. (2025). ABE-Mamba: Few-shot medical image segmentation via adversarial bidirectional enhanced Mamba. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2025.129897>
- Hedhoud, Y., Mekhaznia, T., & Amroune, M. (2025). Vision Mamba for efficient Tuberculosis Detection based on Chest X-Rays: A comparative study with CNN and Vision transformers. *PAIS 2025 - Proceeding: 7th International Conference on Pattern Analysis and Intelligent Systems*. <https://doi.org/10.1109/PAIS66004.2025.11126051>
- Jiang, S., Kui, X., Bao, X., Li, Q., Hu, Z., & Zou, B. (2026). RMViM-Net: Residual multi-path vision mamba with graph interaction attention for medical image segmentation. *Knowledge-Based Systems*, 336, 115326. <https://doi.org/10.1016/j.knosys.2026.115326>
- Kumar, A., & Mahendran, N. (2026). MedScope-LDx: A comprehensive approach for advanced lesion analysis in medical imaging. *Biomedical Signal Processing and Control*, 111. <https://doi.org/10.1016/j.bspc.2025.108352>
- Lai, Y., Cao, A., Gao, Y., Shang, J., & Li, Z. (2025). Advancing Efficient Brain Tumor Multi-Class Classification: New Insights From the Vision Mamba Model in Transfer Learning. *International Journal of Imaging Systems and Technology*, 35(5). <https://doi.org/10.1002/ima.70177>
- Li, C., Sun, Q., Zhang, M., & Zhang, J. (2025). A diffusion model based on multi-scale spatial Mamba for medical image segmentation. *Engineering Applications of Artificial Intelligence*, 156. <https://doi.org/10.1016/j.engappai.2025.111028>
- Li, S., Shen, Z., Zhang, Y., Lai, H., Tan, S., & Chen, W. (2025). 3D MedicalDet-Mamba: A Hybrid Mamba-CNN Network for Medical Object Detection and Localization. *International Journal of Imaging Systems and Technology*, 35(4). <https://doi.org/10.1002/ima.70139>
- Li, Y., Mao, Z., Qin, F., Peng, Y., Zhang, G., Xi, X., Ma, X., Yu, H., Zhou, Y., & Zhu, Z. (2026). A Local-Global Fusion Vision Mamba UNet Framework for medical image segmentation. *Engineering Applications of Artificial Intelligence*, 169. <https://doi.org/10.1016/j.engappai.2026.113987>
- Lin, W. L., Luo, Y., Ling, J., Li, F. H., Qin, J., Yin, Z. C., & Yao, S. (2025). Mamba-Convolutional UNet for multi-modal medical image synthesis. *Medical Physics*, 52(10). <https://doi.org/10.1002/mp.70029>
- Ruan, J., Li, J., & Xiang, S. (2024). *VM-UNet: Vision Mamba UNet for Medical Image Segmentation*. <http://arxiv.org/abs/2402.02491>
- Su, C., Luo, X., Li, S., Chen, L., & Wang, J. (2025). VMKLA-UNet: vision Mamba with KAN linear attention UNet. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-97397-2>
- Wang, C., Xie, Y., Chen, Q., Zhou, Y., & Wu, Q. (2025). *A Comprehensive Analysis of Mamba for 3D Volumetric Medical Image Segmentation*. <http://arxiv.org/abs/2503.19308>
- Yue, Y., & Li, Z. (2024). *MedMamba: Vision Mamba for Medical Image Classification*. <http://arxiv.org/abs/2403.03849>
- Yurdusever, K. C., Kablan, E. B., & Ayas, S. (2025). Parasite Classification in Biomedical Imaging Utilizing Vision Mamba. *ISAS 2025 - 9th International Symposium on Innovative Approaches in Smart Technologies, Proceedings*. <https://doi.org/10.1109/ISAS66241.2025.11101907>
- Zhang, Z., Ma, Q., Zhang, T., Chen, J., Zheng, H., & Gao, W. (2026). Switch-UMamba: Dynamic scanning vision Mamba UNet for medical image segmentation. *Medical Image Analysis*, 107. <https://doi.org/10.1016/j.media.2025.103792>
- Zhong, X., Lu, G., & Li, H. (2025). Vision Mamba and xLSTM-UNet for medical image segmentation. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-88967-5>
- Zhou, Y., Sun, L., Xiong, X., Ti, G., & Yang, S. (2026). GCNet-Mamba: Leveraging state space models and CNN for medical image classification. *Expert Systems with Applications*, 303. <https://doi.org/10.1016/j.eswa.2025.130733>
- Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., & Wang, X. (2024). *Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model*. <http://arxiv.org/abs/2401.09417>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.