Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Selective State Space Models in Medical Imaging: A Systematic Review of Mamba-Based Architectures and Clinical Applications

Ali Emre Gök

aliemregok@nevsehir.edu.tr

Hacı Bektaş Veli University, Computer Engineering Dept, Nevsehir, Türkiye https://orcid.org/0009-0009-8292-2132

Mustafa Yurdakul

mustafayurdakul@kku.edu.tr

Kırıkkale University, Computer Engineering Dept, Kırıkkale, Türkiye https://orcid.org/0000-0003-0562-4931

Şakir Taşdemir

Selçuk University, Computer Engineering Dept, Konya, Türkiye https://orcid.org/0000-0002-2433-246X

Research Article

Keywords: Mamba, Selective State Space Models, Medical Image Analysis, Medical Image Segmentation, Deep Learning, Vision Mamba, Mamba-UNet, Multimodal Medical Imaging, Biomedical Imaging

Posted Date: June 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10139725/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Selective State Space Models in Medical Imaging: A Systematic Review of Mamba-Based Architectures and Clinical Applications

Ali Emre Gök[1], Mustafa Yurdakul[2], Şakir Taşdemir[3]

1 Hacı Bektaş Veli University, Computer Engineering Dept, Nevsehir, Türkiye, aliemregok@nevsehir.edu.tr
2 Kırıkkale University, Computer Engineering Dept, Kırıkkale, Türkiye, mustafayurdakul@kku.edu.tr
3 Selçuk University, Computer Engineering Dept, Konya, Türkiye, stasdemir@selcuk.edu.tr

Abstract

Although Vision Transformer (ViT) architectures have resolved the global-context problem in medical image analysis, their quadratic computational complexity ($O(N^2)$) creates hardware bottlenecks for high-resolution and 3D volumetric data. Mamba and Selective State-Space Models (Selective SSMs) provide an innovative solution to this problem by processing sequential data with linear computational complexity ($O(N)$). This systematic review was designed to examine the rapid adoption of the Mamba architecture in the field of medical artificial intelligence. In accordance with the PRISMA guidelines, 445 recent studies published between 2024 and 2026 were divided into four principal architectural typologies (Pure Mamba, CNN-Mamba Hybrids, Transformer-Mamba Integrations, and Novel Scanning Mechanisms) and passed through a Qualitative Critical Appraisal filter. The findings show that the most dominant and efficient approach in the literature is the CNN-Mamba hybrid, which combines local feature extraction with global context. Despite high diagnostic accuracy, the principal obstacles delaying clinical integration are the loss of spatial information arising from the flattening of 2D and 3D images into 1D sequences, the excessive memory consumption introduced by multi-directional scanning mechanisms, and the lack of multi-center external validation. Consequently, with the development of native 3D Mamba blocks that preserve spatial topology, the Mamba architecture is anticipated to become the backbone of next-generation foundation models in medical image analysis.

Keywords: Mamba; Selective State Space Models; Medical Image Analysis; Medical Image Segmentation; Deep Learning; Vision Mamba; Mamba-UNet; Multimodal Medical Imaging; Biomedical Imaging

1. Introduction

1.1 Background

Medical image analysis plays a critical role in the processes of disease diagnosis and treatment planning [1]. The automatic segmentation and classification of lesions and anatomical structures in medical images have become a fundamental driving force of intelligent medical imaging technologies [2]. Convolutional Neural Networks (CNNs) [3], [4], [5]–[6], which have long been accepted as the standard in this field, have demonstrated superior performance in local feature extraction owing to their inductive bias, spatial locality, and translation equivariance [7]. However, owing to their limited receptive fields, CNNs remain inadequate for modeling the long-range dependencies of large anatomical structures [8], [9].

To overcome this lack of global context, Vision Transformer (ViT) architectures [10], [11], [12], [13], [14] have been integrated into the medical field. However, ViT models possess a

quadratic computational complexity ($O(N^2)$) with respect to the number of tokens in an image [15]. Given that medical images are inherently high-resolution and 3D volumetric in nature, this quadratic complexity leads to serious hardware and memory bottlenecks. This complicates the integration of such models into clinical environments [16].

In this context, Mamba and Selective State-Space Models (Selective SSMs), introduced by Gu and Dao, have offered a revolutionary alternative in the deep learning literature [17]. Through a hardware-aware optimization design, Mamba can process sequential data with linear computational complexity ($O(N)$) [17], [18]. In this way, the ability of Selective SSMs to model long-range dependencies with linear complexity and the local feature extraction power of CNNs can be successfully combined in hybrid networks [19].

1.2 Research Gap

Recently, Mamba-based architectures have been rapidly integrated into the field of medical imaging. In particular, they have been employed in a variety of tasks such as classification and image synthesis [18], [20], [21]. However, when the literature is examined holistically, it is observed that these innovative studies have predominantly concentrated on medical image segmentation tasks. In terms of data modalities, Mamba models adapt to clinical diversity and are most frequently preferred in Magnetic Resonance (MR), Computed Tomography (CT), dermoscopy, and fundus imaging [22].

Despite this rapid adoption process, current Mamba architectures face several critical limitations arising from the nature of medical data. In particular, standard Mamba scanning algorithms designed for 1D sequences are observed to experience difficulties in preserving the spatial context and modeling the directional features of 2D and 3D medical images [19]. In addition, the scalability problems encountered with the clinically important high-resolution and full 3D volumetric data cause memory bottlenecks on standard hardware [8], [22]. Studies in the literature also report that these models can be fragile against the noise, low-dose acquisitions, sparse sampling, and patient motion frequently encountered in clinical settings [22].

Although the applications of Mamba architectures in medicine are increasing exponentially and new solutions are being produced with hybrid architectures, these efforts generally focus on a specific modality or a specific diagnosis. This situation reveals the lack of a current and comprehensive literature review that holistically synthesizes the current integration strategies, dominant modalities, network designs, and constraints of Mamba-based deep learning methods in medical imaging.

1.3 Aim of the Study

Motivated by the gap in the literature detailed in Section 1.2, the primary aim of this study is to examine the applications of Mamba and Selective SSMs in the field of medical imaging in a systematic and holistic manner. The study aims to evaluate how Mamba is integrated with conventional architectures (CNN and ViT), how it is adapted to high-resolution medical data, and how it performs against operational and methodological constraints.

In line with this overall aim, and in order to fill the existing gaps in the literature, our study seeks to answer the following three principal research questions (RQs):

- **RQ1:** For which clinical tasks and image modalities are Mamba-based architectures most frequently used in medical image analysis, and what structural advantages do they offer compared with conventional methods in this process?
- **RQ2:** Through which structural adaptations and hybrid designs (multi-directional scanning mechanisms or pre-trained model integrations) do Mamba models, which are fundamentally designed for 1D sequences, pass in order to model the spatial context of 2D/3D volumetric medical data?
- **RQ3:** What are the principal methodological bottlenecks encountered in adapting the Mamba architecture to medical images, and which strategies should be followed in the future for clinical integration?

2. Overview of State Space Models and Mamba

2.1 State-Space Models

State-Space Models (SSMs) [23]–[24], which derive their foundation from classical control theory, map an input stimulus $x(t) \in R$, through a hidden state $h(t) \in R^N$, to an output $y(t) \in R$ by means of linear differential equations [17]. This continuous-time mapping is defined as a linear dynamic system using the evolution and projection matrices of the system. The continuous-time differential equations that constitute the backbone of the system are expressed as given in Equations (1) and (2) [17]:

$$h'(t) = Ah(t) + Bx(t) \tag{1}$$

$$y(t) = Ch(t) \tag{2}$$

In Equations (1) and (2), the matrix $A \in R^{NxN}$ governs the hidden-state transitions of the system, the matrix $B \in R^{Nx1}$ determines the influence of the input on the hidden state, and the matrix $C \in R^{1xN}$ defines how the hidden state is projected onto the final output [17]. The standard Mamba architecture is presented in Fig. 1.

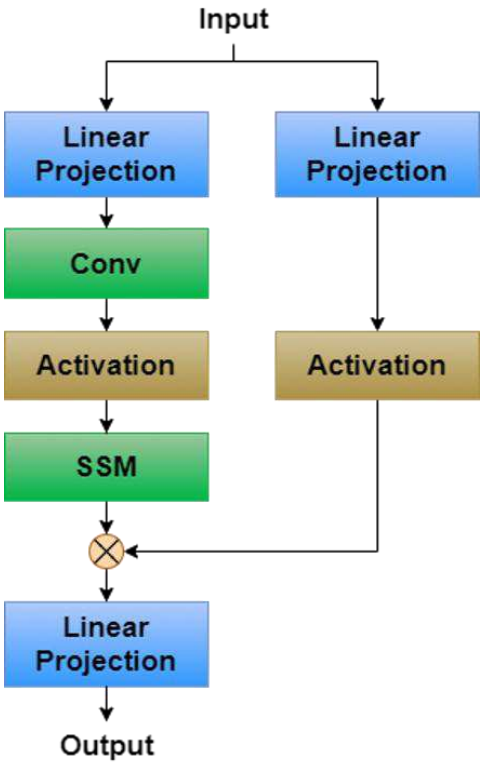


Fig. 1 Architecture of the Mamba block, where the input is split into two linear projection branches, one applying a convolution, an activation and an SSM and the other a gating activation, whose outputs are multiplied and passed through a final linear projection to produce the output.

2.2 Discretization

For continuous-time analytical formulations to operate with the discrete data used in deep learning networks, such as token sequences or image pixels, discretization of the system is required [25]. This transformation process is carried out using a system-specific time-step or scaling parameter $\Delta \in R^{+}$ [17], [25]. In Mamba-based studies in the literature, the Zero-Order Hold (ZOH) rule is widely applied to convert the continuous-time matrices (A, B) into their discrete-time counterparts ($\underline{A}, \underline{B}$). Under the ZOH assumption, the discretized parameters are derived mathematically through the formulas given in Equations (3) and (4) [17]:

$$\underline{A} = exp\,(\Delta A) \tag{3}$$

$$\underline{B} = (\Delta A)^{-1}(exp\,(\Delta A)\ - I) \cdot \Delta B \tag{4}$$

Once the discretization step is complete, the continuous-time system structure is transformed into a linear recurrent model and is operated across the time steps (t) as given in Equations (5) and (6) [19]:

$$h_t = \underline{A} h_{t-1}\ + \underline{B} x_t \tag{5}$$

$$y_t = Ch_t \tag{6}$$

2.3 Selective Scanning Mechanism

Conventional SSMs are Linear Time-Invariant (LTI) systems, in which the matrices A, B, C and Δ are fixed independently of the input. This restricts the capacity of such models for data-driven, content-based reasoning [26]. To eliminate this LTI constraint, the Mamba architecture introduced the Selective Scan Mechanism [27]. In this mechanism, the parameters B, C and Δ are no longer fixed matrices but are instead made data-dependent by being turned into dynamic functions of the input sequence x_t at each time step [17], [27]. In this way, in high-variance data such as medical images, the model can dynamically decide, according to the input content, which features to retain in the hidden state and which noisy information to forget [17]

2.4 Hardware-Aware Optimization

Making the matrices time- and input-dependent through the selective mechanism causes the model to lose the advantage of parallel convolutional computation. This leads to memory bottlenecks on GPU hardware [17], [18]. To resolve this efficiency crisis, Mamba employs a hardware-aware parallel scanning algorithm. Instead of materializing the high-dimensional hidden-state matrices in the slow High-Bandwidth Memory (HBM) of the graphics processor, the algorithm performs the computations directly in the SRAM cache of the GPU [17]. Through this optimization, which minimizes the input/output (I/O) transfer load between the memory hierarchies, Mamba models process long sequences with linear time complexity and thereby manage to avoid the quadratic memory burden of Transformer models [17], [19].

2.5 Mamba Architectures in Image Processing

SSMs and the original Mamba architecture were fundamentally developed to process one-dimensional sequential data such as text. Directly adapting the standard Mamba model to two-dimensional (2D) image data by flattening images into 1D sequences disrupts the natural spatial structures and local correlations inherent to the image. Pixels that are physically adjacent in the 2D plane become distant from one another in the scanning sequence, giving rise to the loss of spatial continuity, topological consistency, and directional dependencies (the directional sensitivity problem). To solve this problem without sacrificing the linear-time complexity advantage of Mamba, the Vision Mamba (ViM) and VMamba architectures [28] were designed.

In the Vision Mamba (ViM) architecture (Fig. 2), there are two fundamental structural innovations to overcome this obstacle. First, position embeddings are added to the image patches in order to ensure the spatial position awareness of the patches. Second, instead of standard unidirectional scanning, a bidirectional SSM approach is used. Thus, image sequences are processed simultaneously in both the forward and backward directions, and the resulting outputs are then combined. In this way, ViM successfully captures the position-aware and data-dependent global visual context in a single forward pass, without requiring the computationally and memory-intensive self-attention mechanisms.

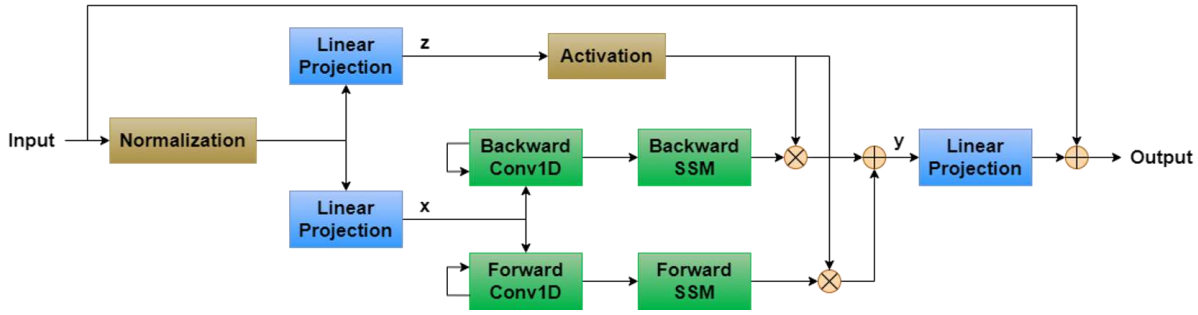


Fig. 2 Architecture of the Vision Mamba block, where the normalized input is processed by parallel backward and forward branches, each comprising a Conv1D and an SSM, that are gated by an activation and summed, then linearly projected and added to the input through a residual connection to produce the output.

On the other hand, in the VMamba architecture (Fig. 3), a 2D-Selective-Scan (SS2D) mechanism is used that contains a Cross-Scan Module (CSM) and essentially consists of three stages, in order to bridge the discrepancy between 1D sequence scanning and the 2D visual plane. In the first stage, called scan expanding, the 2D image features are unfolded along rows and columns and converted into sequences in four different independent directions: top-left to bottom-right, bottom-right to top-left, top-right to bottom-left, and bottom-left to top-right.

These four separate sequences are then processed in parallel by the selective-scan blocks, capturing features from various perspectives. In the final stage, scan merging (cross-merge), these sequences obtained from the independent directions are summed and reshaped back into a single 2D feature map. In this way, a strong global receptive field is constructed in two-dimensional space without compromising computational efficiency, and the spatial relationships are successfully reconstructed.

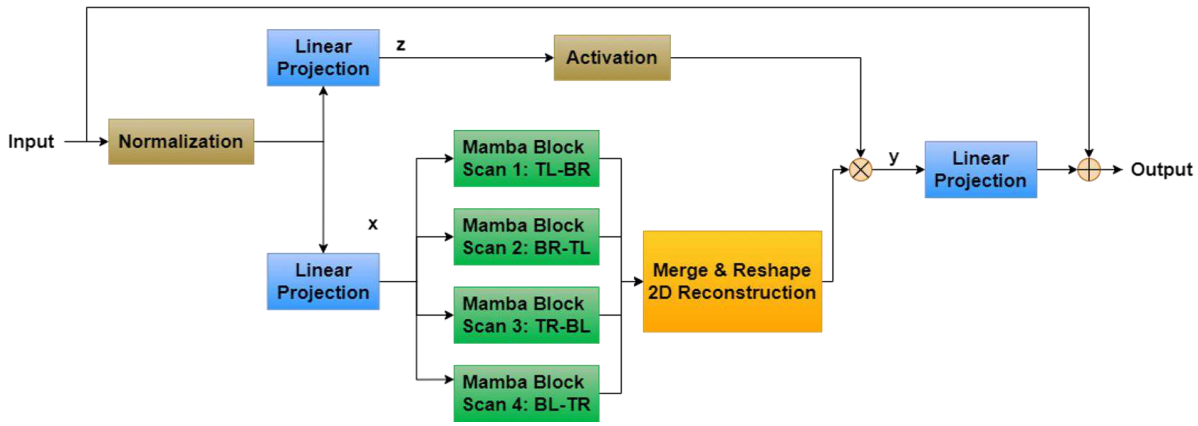


Fig. 3 Architecture of the VMamba block, where the normalized input is scanned in four directions (TL-BR, BR-TL, TR-BL and BL-TR) by parallel Mamba blocks whose outputs are merged and reshaped, gated by an activation, and linearly projected with a residual connection to produce the output.

3. Methods

3.1 Search Strategy

The search strategy of this study is designed to systematically collect research focused on Selective SSMs, particularly Mamba architectures in the medical image analysis domain. To

this end, studies between 2024-2026 were scanned from four databases: Web of Science (SCIE), IEEE Xplore, Scopus and ACM Digital Library.

The search queries were specifically applied to the title, abstract, and keyword fields of the articles. To maximize the retrieval of relevant studies, the following search string was constructed using boolean operators:

("Mamba" OR "MedMamba" OR "U-Mamba" OR "SegMamba" OR "Mamba-UNet" OR "VM-UNet" OR "Swin-UMamba" OR "Vision Mamba" OR "Selective State Space Model*" OR "State Space Model*" OR "VSSM" OR "Vision State Space Model*") AND ("medical image*" OR "biomedical image*" OR "clinical image*" OR "MRI" OR "magnetic resonance" OR "CT" OR "computed tomography" OR "X-ray" OR "Xray" OR "ultrasound" OR "PET" OR "positron emission" OR "histopathology" OR "endoscopy" OR "radiology")

3.2 Inclusion Criteria

Studies were included in this review based on predefined criteria to ensure relevance, quality, and methodological rigor. Specifically, only peer-reviewed journal articles published between 2024 and 2026 and indexed in major databases, including Web of Science (SCIE), IEEE Xplore, Scopus, and the ACM Digital Library, were considered. Eligible studies were required to focus on Mamba-based architectures or Selective SSMs applied to medical or biomedical imaging tasks. Additionally, only English-language articles with full-text availability were included, and studies were required to provide a substantive methodological contribution.

The detailed inclusion criteria are summarized in Table 1.

Table 1. Inclusion Criteria for Study Selection

Criterion	Description
Topic Relevance	Studies focusing on Mamba-based architectures or Selective SSMs in medical or biomedical imaging tasks (e.g., segmentation, classification, detection, reconstruction, fusion, report generation).
Publication Type	Peer-reviewed journal articles only; conference papers, workshops, preprints, book chapters, and theses were excluded.
Time Frame	Publications between 2024 and 2026.
Indexing Databases	Indexed in Web of Science (SCIE), IEEE Xplore, Scopus or ACM Digital Library.
Language	English-language publications only.
Accessibility	Full-text availability required.
Methodological Contribution	Studies must propose or substantially evaluate Mamba-based or SSM-based methods.

3.3 Exclusion Criteria

Studies were excluded from this review if they did not meet the predefined inclusion criteria or failed to provide sufficient methodological or contextual relevance. Specifically, non-peer-reviewed publications such as conference papers, workshop articles, preprints, book chapters, and theses were excluded. Studies published outside the 2024–2026 time frame or not indexed in Web of Science (SCIE), IEEE Xplore, Scopus, or the ACM Digital Library were also omitted. Additionally, articles not written in English or lacking full-text access were excluded. Furthermore, studies presenting architectural incompatibility (e.g., not utilizing Mamba/SSMs), off-field applications (non-medical domains), or out-of-scope tasks (e.g., lack of substantial imaging-focused experimental evaluation) were removed. Works that only mentioned Mamba or Selective SSM concepts without providing meaningful methodological development or experimental evaluation were also excluded.

The detailed exclusion criteria are summarized in Table 2.

Table 2. Exclusion Criteria for Study Selection

Criterion	Description
Off-Field Applications	Studies applying Mamba/SSMs outside the domain of medical or biomedical imaging (e.g., natural images, satellite imagery).
Architectural Incompatibility	Studies not utilizing Mamba-based architectures or SSMs.
Out-of-Scope Tasks	Studies focusing on tasks not relevant to the core objective of the review.
Publication Type	Conference papers, workshops, preprints, book chapters, theses, and other non-peer-reviewed articles.
Time Frame	Studies published before 2024 or after 2026.
Indexing Databases	Studies not indexed in Web of Science (SCIE), IEEE Xplore, Scopus or ACM Digital Library.
Language	Non-English publications.
Accessibility	Studies without accessible full text.

3.4 Study Selection Process

The study selection process was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [29] to ensure transparency, reproducibility, and methodological rigor. The selection procedure consisted of multiple stages, including identification, screening, eligibility assessment, and final inclusion.

In the identification phase, a comprehensive literature search was performed across four major databases: Web of Science (SCIE), IEEE Xplore, Scopus and the ACM Digital Library. Duplicate records retrieved from different databases were identified and removed prior to further analysis. During the screening phase, the titles and abstracts of the retrieved studies were independently reviewed by two researchers to exclude clearly irrelevant publications

based on the predefined inclusion and exclusion criteria. Studies not related to Mamba-based architectures or medical imaging were eliminated at this stage.

In the eligibility phase, the full texts of the remaining articles were carefully assessed to determine their suitability. Studies were excluded if they did not provide sufficient methodological detail, lacked experimental validation, or did not present a meaningful contribution related to Selective SSMs or Mamba-based approaches.

Finally, in the inclusion phase, only studies that fully satisfied all inclusion criteria were selected for qualitative synthesis. Any disagreements during the selection process were resolved through discussion and consensus.

Due to the extensive volume of included literature (n=445), a representative subset of 60 studies, 15 for each architectural typology was purposively selected for detailed qualitative synthesis in the tabular summaries. This subset was strategically chosen to ensure a comprehensive representation of architectural diversity, various imaging modalities, and diverse clinical analysis tasks, rather than selecting based on performance metrics.

3.5 Data Extraction

In this section, data extraction was performed on articles which were selected after the literature review. Two of the authors investigated papers by following a common pattern to ensure the study is conducted in a more effective manner. Firstly, article title and general information were extracted, then method and data sections were focused on. The utilized architectures, methods such as hybridization and datasets were comprehensively analysed. At last, obtained information was compiled into a structured table. The key information gathered from the articles encompasses the following domains:

General Information: Title, authors, publication year and journal.

Clinical Goal and Modality: Data type used in study (MRI, CT, etc.) and task performed.

Architectural Structure: Characteristics of the Mamba architecture used in the study (Mamba variants, pure Mamba, hybrid approaches etc.)

Dataset: Details regarding the datasets.

Performance Metrics: Metrics used for model evaluation.

Limitations: Challenges and limitations encountered in the study.

During this process, in cases where the required data could not be located within the articles, the provided complementary materials were inspected. Consequently, information that could be categorized under common headings were obtained from all the analysed articles. In instances of disagreement among the researchers, deliberations were conducted to reach a mutual consensus.

3.6 Quality Assessment

In this study, research utilizing Mamba models for medical image analysis was examined with a primary focus on the methodological robustness and hardware applicability of novel architectures rather than solely on diagnostic accuracy. In this context, the Qualitative Critical

Appraisal approach was preferred. The articles reviewed were rated according to three fundamental criteria:

1. **Methodological Transparency and Reproducibility:** Whether the code used is open source, the transparency of the architectural improvements, and whether the developed models are compared to current state-of-the-art models.
2. **Dataset Robustness and Validation:** Whether the dataset is private or public, and the implementation of robust validation strategies such as cross-validation.
3. **Computational Feasibility and Clinical Applicability:** The model's parameter load, scanning mechanism used, computational cost, bottlenecks and generalization ability.

Information extracted from reviewed studies was independently evaluated by the authors and each study was rated on a scale of Low, Medium, or High. At this stage, any disagreements were discussed and a consensus was reached. The findings obtained based on these criteria are summarized in Section 4.4 (Summary of Included Studies). In addition, the limitations of the studies are evaluated in Section 5 (Discussion). It should be noted that the quality assessment was strictly used for characterizing the methodological robustness of the included studies, rather than serving as an exclusion criterion.

4. Results

4.1 Study Selection (PRISMA Flow)

The comprehensive literature search yielded a total of 1259 articles across four databases: IEEE (125), Web of Science (447), Scopus (576) and ACM Digital Library (111). The distribution of these studies is given in Fig. 4. With the elimination of duplicated studies, 601 unique studies remained. Afterwards, an initial screening was conducted to eliminate irrelevant articles. During the screening phase, 64 articles were excluded due to out-of-scope tasks or off-field applications. After the screening phase, eligibility assessment was performed and 92 articles were excluded due to architectural incompatibility or falling under different publication types (e.g. reviews). Consequently, 445 studies met all criteria and were included in the systematic review. The complete list of these 445 included studies is provided in Supplementary File S1 These remaining studies were grouped in four distinct categories: Pure Mamba Adaptations, CNN-Mamba Hybrids, Transformer-Mamba Integrations and Novel Scanning Mechanisms. PRISMA flow is given in Fig. 5.

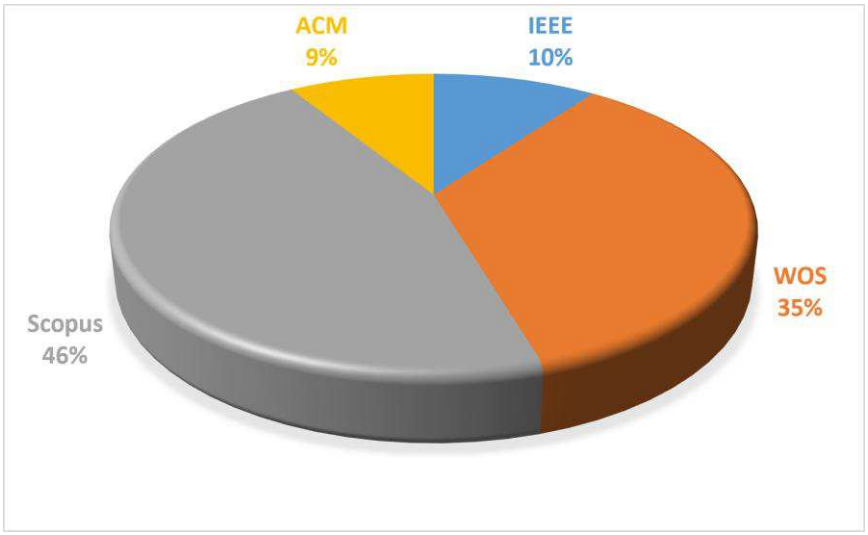


Fig. 4 Distribution of the included studies across the four source databases: Scopus, Web of Science, IEEE Xplore and ACM Digital Library.

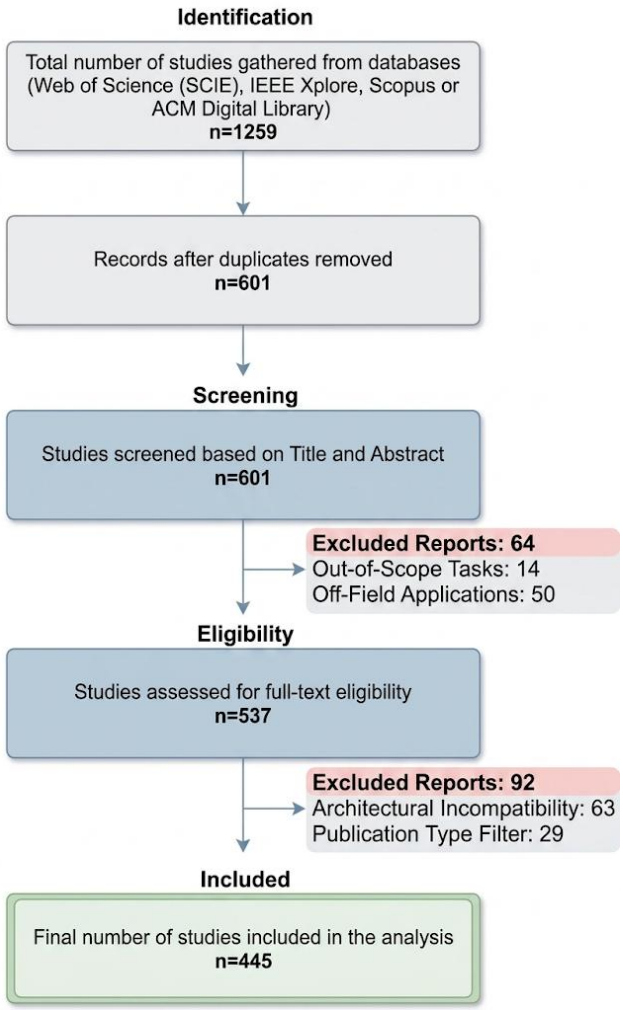


Fig. 5 PRISMA flow diagram of the study selection process, covering the identification, screening, eligibility and inclusion stages that yielded the 445 studies analyzed in this review.

4.2 Descriptive Analysis

The distribution of the studies according to years is presented in Fig. 6. The Mamba architecture developed in 2024, attracted considerable attention immediately after its emergence and was employed in more than 200 medical image analysis studies in 2025. This number demonstrates the Mamba architecture's rapid adoption in the medical field. The main reason for this trend is Mamba's linear computational complexity in contrast to quadratic computational cost of Transformer (ViT) models. Although literature review for this study ended in the second quarter of 2026, it is remarkable that there are almost the same number of studies in comparison to the year 2025. This highlights that the Mamba architecture has become one of the most dynamic research areas in literature.

When examining the clinical focus areas of this rapidly growing literature (Fig. 7), it is observed that multi-modality approaches constitute the largest portion of the studies, while MRI and CT are the next most extensively researched fields. Specifically, 27% (n=121) of the reviewed studies utilized multi-modality data, followed by MRI and CT with 18% (n=82) and 15% (n=68), respectively. The primary reason for the high preference for MRI and CT modalities is their 3D and high-resolution nature. Due to its capability of processing long sequences without bottlenecks, Mamba architecture gains a natural advantage in processing 3D medical data. These three leading fields are followed by general medical imaging studies in which independent clinical datasets are used together. Subsequently, ultrasound and X-ray modalities account for 10% and 5% of the studies, respectively. The fact that multi-modality studies hold the highest proportion among the studies performed demonstrates the flexibility of Mamba architecture across diverse data types.

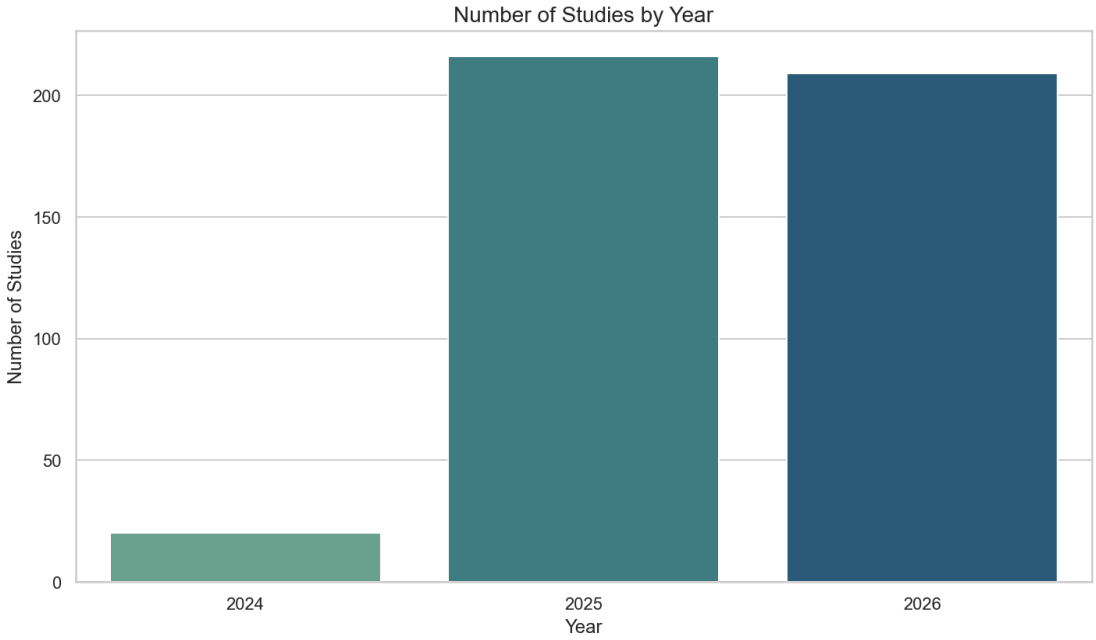


Fig. 6 Number of included studies by publication year across 2024, 2025 and 2026.

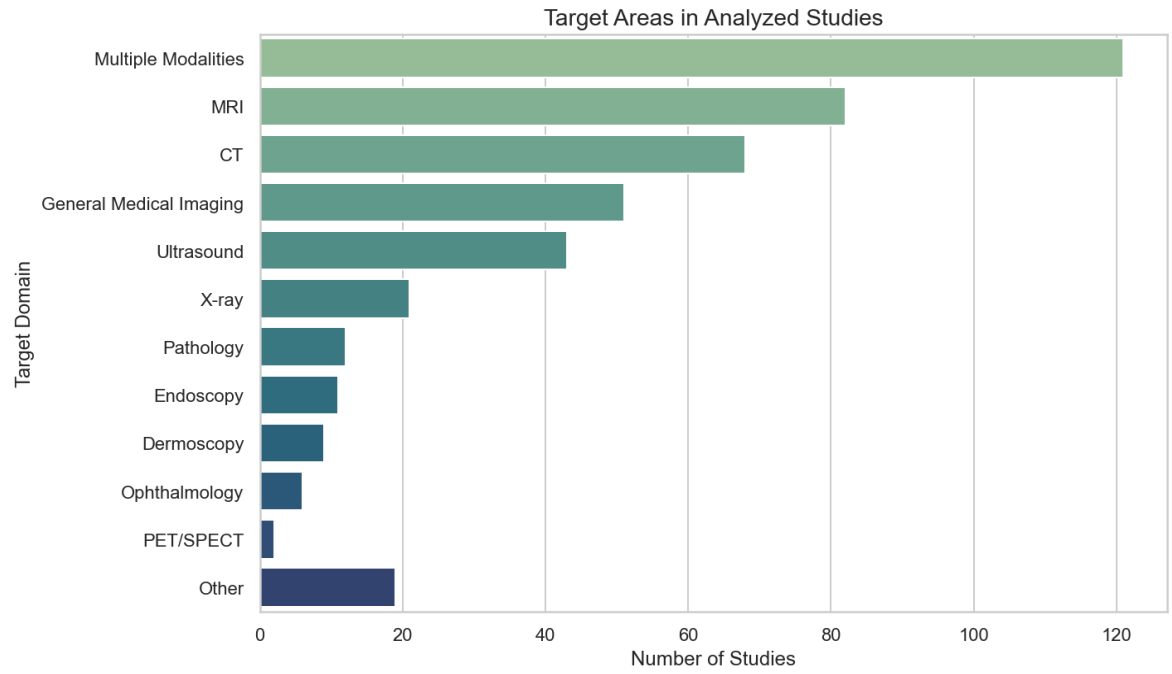


Fig. 7 Distribution of the analyzed studies across the targeted medical imaging domains.

4.3 Thematic Analysis

The reviewed studies are categorized into four main methodological themes. The distribution of these studies is presented in Fig. 8. It is seen that the majority of the studies utilized CNN-Mamba hybridizations, leading the others with a clear margin. This is followed by Pure Mamba Adaptations. Subsequently, these two leading methodologies are followed by Transformer-Mamba Integrations and Novel Scanning Mechanisms which account for an almost equal number of studies.

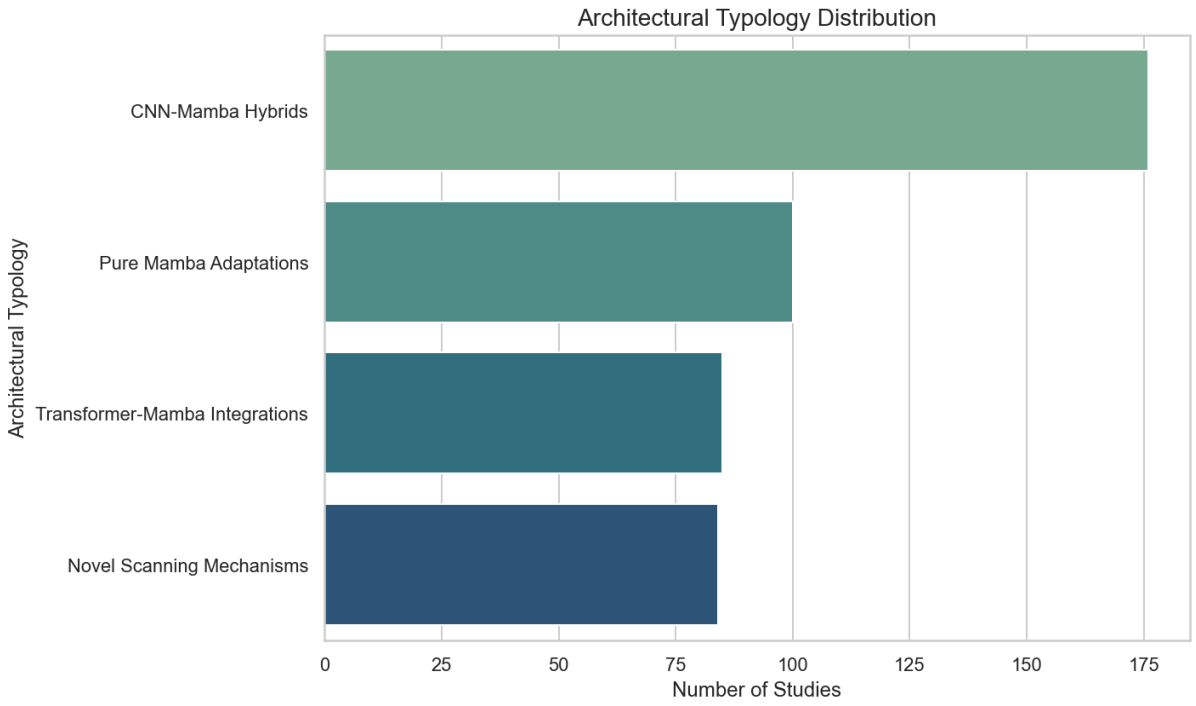


Fig. 8 Distribution of the analyzed studies across the four architectural typologies of Mamba-based models: CNN-Mamba hybrids, pure Mamba adaptations, Transformer-Mamba integrations and novel scanning mechanisms.

4.3.1 CNN-Mamba Hybrids

The number of CNN-Mamba hybrid studies is significantly higher than that of other methodologies. The primary objective of these models, which are employed in various clinical tasks such as classification [22] , segmentation [25], cross-modality synthesis [30], [31] [21] and object detection [32] [33], is to enhance performance by combining the local feature extraction ability of CNNs with the Mamba architecture's capability to capture global context in long sequences. A review of prominent studies in literature reveals that these hybrid approaches are mostly used on complex anatomical structures such as brain [34], breast [7], pathology [22], eye [35], dermoscopy [25], endoscopy [36] and kidney [1]. The main reason for specifically targeting these areas is that the detection of lesions and pathologies requires both micro-level tissue detail (CNN-based) and a broad receptive field capable of grasping the spatial topology of the entire organ (Mamba-based) [19]. Although these hybrid architectures yield high diagnostic accuracy, they introduce certain limitations particularly regarding computation and memory overhead stemming from their dual-branch or complex integration structures [22]. The datasets, architectural details, achieved performance metrics and methodological limitations of the prominent innovative approaches in this category are comprehensively summarized in Table 3.

Table 3. CNN-Mamba Hybrid Studies

Reference	Clinical Task & Dataset	Proposed Architecture	Key Performance Metrics	Methodological Limitations
[37]	Task: Brain Tumor Seg. Modality: MRI Dataset: BraTS 2023, 2020	Hierarchical tri-directional Mamba unit combined with large kernel convolutions to balance local semantics and global awareness.	Significantly surpassed current SOTA methodologies in whole-volume feature dependency and segmentation performance.	Tri-directional scanning alongside large convolutions may substantially inflate parameter counts, hindering deployment on memory-constrained hardware.
[38]	Task: Skin Lesion & Tumor Seg. Modality: Dermoscopy, US Dataset: ISIC17/18, BUSI	Selective Fusion Mamba and Multi-Scale Fusion Mamba mechanisms for dynamic local-global integration.	Outstanding segmentation efficacy across three major datasets with an ultra-lightweight footprint of only 0.038M parameters.	Extreme parameter reduction may severely bottleneck the network's capacity to recognize highly irregular multi-class structures outside the training distribution.
[39]	Task: Breast Cancer Seg. Modality: Ultrasound Dataset: BCDM and BMNP	Brownian Bridge diffusion paradigm integrated with Semantic Correlation Mapping Attention and State Space Fusion Mamba.	SOTA breast cancer segmentation with noticeably fewer inference steps and mitigated cross-domain discrepancies.	Diffusion-based models inherently require significantly more inference time than traditional single-pass feedforward networks.
[40]	Task: Multitask Seg. & Class. Modality: Ultrasound, Derm. Dataset: 4 breast US, 1 skin	Dual-Stream CNN-Mamba extraction module and multi-scale convolutional self-attention preprocessor.	Scored 72.86–78.45% IoU on breast US and 86.57% on skin cancer datasets, outperforming unitask methods.	Complex integration of multiple preprocessors and dual-stream extractions likely demands substantial VRAM, hindering edge deployment.

[41]	Task: Macular Edema Diag. Modality: Fundus Imaging Dataset: IDRiD, Messidor	Integrates graph convolutional networks (GCN) for local spatial correlations with Mamba for long-range global dependencies.	Substantially outperforms existing approaches, ensuring extracted characteristics firmly align with actual pathological lesions.	The mapping matrix for geometric correspondence likely exhibits high sensitivity to variations in retinal topography and severe misalignment.
[42]	Task: Mutation Burden Predict. Modality: Pathology WSI Dataset: Lung cancer cohort	Parallel text-guided attention modules and a VMamba backbone, integrating diagnostic textual features with histopathology.	Achieved a classification AUC of 0.994 and a regression MAPE of 0.25 without requiring whole-exome sequencing.	Heavy reliance on diagnostic textual features indicates that inaccurate or sparse clinical notes could directly compromise mutation burden predictions.
[43]	Task: Deformable Registration Modality: Brain MRI Dataset: Mindboggle, IXI	State-space model embedded within a residual pyramid network encoder to address complex non-rigid brain deformations.	Dice score up to 65.5, outperforming SOTA registration methods while reducing computational complexity.	The model may encounter difficulties with extreme pathological deformations or topological changes not adequately represented in the training datasets.
[44]	Task: Multi-modal Synthesis Modality: MRI, CT, PET Dataset: multi-contrast MRI	Hierarchical multi-scale ConvMamba framework combined with a GLCM loss function and an auxiliary diffusion model.	Decisively outperforms SOTA generative architectures across diverse multi-contrast synthesis benchmarks.	Employing an auxiliary diffusion model within a GAN generates immense computational overhead, significantly hindering real-time synthesis applications.
[45]	Task: Image Denoising/Restor. Modality: Spectral CT Dataset: PCCT and MSCT	Mamba-assisted X-Net architecture guided by latent representation priors for limited-view restoration.	SOTA proficiency in structural restoration and noise mitigation for both photon-counting and multi-source CT.	Utilizing latent representations of full-spectrum prior images requires access to pristine, high-quality spectral data during training, which remains scarce.

[46]	Task: Super-Resolution Modality: Ultrasound Dataset: Five different US datasets	Frequency-spatial domain reconstruction utilizing an Enhanced Transform Combine Module and an Adaptive Top-k Prompt Module.	Exceptional fidelity, achieving an average PSNR 1.31dB higher than SOTA approaches at a 2x scale factor.	The adaptive routing-guided strategy introduces dynamic computational pathways that could lead to inconsistent real-time inference latency.
[47]	Task: Object Detection Modality: Endoscopy Dataset: SUN Polyp, PICCOLO	State Space Model incorporated into the YOLOv10 architecture alongside an Efficient Multi-Scale Attention mechanism.	SOTA diagnostic detection accuracy and robustness on the Etis-Larib dataset for early colorectal cancer screening.	The training data does not fully cover the full range of clinical scenarios, and model pruning optimizations have not yet been implemented for hardware-constrained devices.
[48]	Task: Image Watermarking Modality: X-ray, MRI, CT Dataset: NIH CXR, COVIDx	Integrates Mamba state-space models and CNNs within the Haar wavelet domain for secure patient data embedding.	High-fidelity embedding (PSNR 43.8 dB) improving watermark robustness under complex composite distortion attacks by 10.8%.	Embedding watermarks directly into the wavelet frequency domain runs a persistent risk of subtly altering critical diagnostic textures.
[49]	Task: Active Domain Adaptation Modality: Brain MRI Dataset: BraTS cross-domains	Hybrid Mamba-CNN architecture utilizing a Region-level Uncertainty-Guided Sample Selection strategy.	Approaches fully supervised performance levels across distinct target domains under a strictly limited 5% annotation budget.	Calculating region-level uncertainties to guide active learning sample selection may introduce latent computational bottlenecks.
[50]	Task: EEG-to-fMRI Reconstruct. Modality: EEG & fMRI Dataset: NODDI, Oddball	Multi-directional Time-Frequency Convolutional Attention Encoder with an efficient Vision-Mamba U-Net decoder.	Drastically advanced spatial fidelity, achieving exceptional structural similarity indices of up to 0.788 across public benchmarks.	The model focuses exclusively on static single-volume reconstruction, lacking recurrent mechanisms to map continuous dynamic fMRI time-series.

[51]	Task: CFD Hemodynamics Modality: CFD Simulation Dataset: Computational dataset	Merging Mamba-based state-space modeling with hierarchical point cloud processing (PointNet++) for vascular geometries.	Prediction accuracies comparable to full CFD simulations while improving computational efficiency by several orders of magnitude.	Relying strictly on synthetic CFD simulations for point cloud training inherently assumes rigid boundary conditions, failing to capture true physiological elasticity.

4.3.2 Pure Mamba Adaptations

The studies in this category are research efforts that directly adapt the Mamba and Selective SSM backbone to medical imaging tasks [52], [53]–[54] without using structural hybridization with conventional CNN- or Transformer-based architectures. The primary aim of these approaches is to maximize the use of the linear computational complexity of the Mamba architecture and its ability to model long-sequence dependencies, thereby improving parameter efficiency on high-resolution medical data [17], [55]. Prominent studies demonstrate the applicability of pure Mamba models across a broad clinical spectrum, such as segmentation [56], classification [34] and super-resolution [22]. These architectures are particularly preferred for analyzing high-dimensional data in which conventional Transformers experience memory bottlenecks, such as pathology, functional MR sequences, and high-resolution ultrasound. However, the need for pure Mamba models to compress complex 2D or 3D spatial topologies into 1D sequences (flattening) can lead to local feature attenuation [57].. Examples of studies demonstrating the potential of this architectural approach across multiple modalities and its current limitations are presented in Table 4.

Table 4. Pure Mamba Adaptation Studies

Reference	Clinical Task & Dataset	Proposed Architecture	Key Performance Metrics	Methodological Limitations
[25]	Task: Medical Image Seg. Modality: Dermoscopy, CT/MRI Dataset: ISIC17/18, Synapse	The first pure state-space-based architecture for medical image segmentation utilizing an asymmetrical encoder-decoder structure.	Eliminated quadratic computational complexity, yielding competitive Dice scores of 89.03, 89.71, and 81.08 across diverse datasets.	An asymmetrical structure with fewer convolution layers may reduce the model's capacity to extract extremely fine-grained, localized spatial details.
[58]	Task: Report Gen. & Lesion Det. Modality: X-ray Dataset: IU X-ray, CheXpertPlus	Lightweight Mamba-based vision encoder paired with a large language model decoder and multi-stage contrastive learning.	Accurately localized abnormalities and produced clinically coherent reports, successfully eliminating the high computational overhead typical of ViTs.	Pairing a lightweight encoder with an LLM decoder inherently demands substantial GPU memory for autoregressive text generation, bottlenecking edge deployment.
[59]	Task: CBCT Segmentation Modality: CBCT Dataset: Mixed dentition	Pure U-Mamba Bot and U-Mamba Enc architectures applied for pediatric CBCT segmentation of dental hard tissues and pulp.	Maintained highly consistent volumetric correlations and rapid processing times across varying dental developmental stages.	Segmentation accuracy exhibited noticeable degradation when identifying primary teeth in older children due to root resorption complexities.
[60]	Task: Few-shot Seg. Modality: Medical Imaging Dataset: Not specified	Hierarchical hypercorrelation vision state-space model integrating a multiscale feature matching pipeline.	Efficiently captured long-range dependencies with linear complexity, overcoming severe low-contrast and contextual ambiguity challenges.	Dense computation of target-level and image-level correlations could still demand high VRAM during the training of large-scale multiscale inputs.

[61]	Task: IBD Subtype Ident. Modality: Endoscopy Dataset: WCE, Clinical cohort	Dual-stream architecture combining a multimodal LLM and a hierarchical MambaVision network with cross-attention.	Achieved an exceptional state-of-the-art diagnostic accuracy of 96.49% across a large-scale clinical cohort and public benchmarks.	Integrating an MLLM alongside a hierarchical MambaVision architecture likely imposes severe GPU memory constraints during training.
[62]	Task: Connectivity Modeling Modality: fMRI & EEG Dataset: Public Cohorts	Adapts state-space modeling to Riemannian manifolds with a dual-path design for hierarchical multi-granular aggregation.	Produced highly compact, discriminative representations, demonstrating consistently competitive accuracy and robustness for neurological disease detection.	High-dimensional SPD inputs may still pose computational bottlenecks during progressive dimensionality reduction on large cohort datasets.
[63]	Task: Tumor Segmentation Modality: CT Dataset: LiTS, 3DIRCADb	Integrates anatomy-aware prompts with Mamba-based state-space modeling and structure-aware filtering.	Highly compact model (27.6M parameters) surpassed SOTA CNNs/Transformers, demonstrating exceptional sensitivity to small nodules.	Heavy reliance on predefined prompts requires accurate initial organ localization; anatomical anomalies could disrupt the entire pipeline.
[64]	Task: 2D/3D Registration Modality: DXR & CBCT Dataset: Knee Joint Data	Utilizes a twin-stream Vision Mamba-based encoder and a Multi-Objective Spatial Transformation module.	Achieved a mean target registration error of 3.36 mm and processed single X-ray images within 1.22 seconds.	Registering a continuous high-frame-rate dynamic X-ray video sequence would still demand significant cumulative computational time.
[65]	Task: Disease Detection Modality: Endoscopy Dataset: Not specified	Foundation model leveraging a specialized state-space model and masked autoencoding to learn from unlabeled sequences.	Matched the diagnostic performance of traditional, fully supervised systems using merely 1% of expert-labeled data.	The self-learning mechanism's dependency on unlabeled data quality might propagate biases if the pre-training corpus lacks pathological diversity.

[66]	Task: Deformable Registration Modality: MR-CT Dataset: SR-Reg	Leverages efficient Mamba state space models for robust global feature learning alongside fine-grained extractors.	Demonstrated impressive enhancements in registration efficacy and computational efficiency, vastly outperforming SOTA frameworks.	Deformable registration via 1D modeling may struggle to maintain biologically plausible deformation fields across complex 3D structures without spatial regularization.
[67]	Task: Tumor Segmentation Modality: MRI Dataset: Multicenter MRI	Employs a triple-domains Mamba module fusing channel, spatial, and frequency features to address class imbalance.	Demonstrated superior diagnostic performance and robust adaptability in precisely segmenting both microscopic and massive carcinomas.	Fusing spatial, channel, and frequency domain features simultaneously may result in excessive feature redundancy and prolonged processing times.
[68]	Task: Image Super-Resolution Modality: Pathological Imaging Dataset: Not specified	Integrates a state-space module into a diffusion encoder to achieve authentic optical zoom characteristics.	Enhanced downstream classification accuracy by efficiently reformulating denoising processes to achieve accelerated sampling.	Complex diffusion processes combined with state-space temporal modeling may demand substantial computational resources during whole-slide generation.
[69]	Task: Multimodal Fusion Modality: Multi-modality Dataset: Unnamed Datasets	Utilizes structured state-space Mamba to aggregate embeddings and learn the Gromov-Wasserstein distance for feature alignment.	Produced clearer boundary definitions and vastly superior visual consistency in fused images compared to existing models.	Calculating the Gromov-Wasserstein distance is highly computationally intensive, which may significantly reduce inference speed for high-resolution images.
[70]	Task: Biomarker Prediction Modality: Ultrasound Dataset: Multicenter BC Cohort	Vision-Mamba architecture integrating 2D/strain elastography images with pixel-level radiomic feature maps.	Achieved outstanding AUC values (0.941-0.945) on external cohorts for predicting Ki-67 expression.	Acquiring both B-mode images and specialized elastography maps significantly increases the clinical data acquisition burden.

[71]	Task: 3D Organ Reconstruction Modality: X-ray Dataset: Thoracic benchmarks	Direct 3D multi-organ reconstruction from orthogonal biplanar X-rays via a state-space cross-view fusion module.	Attained impressive Dice scores (up to 93.94%) for lung and heart segmentation without requiring CT priors.	Generating comprehensive 3D anatomy strictly from biplanar X-rays may struggle with pathological deformations that overlap heavily in 2D projections.

4.3.3 Transformer-Mamba Integrations

Although the Mamba architecture offers a linear computational advantage on long sequences, the precise spatial-alignment capabilities provided by the global self-attention mechanisms of ViT-based models are indispensable for some complex medical tasks [8], [22]. For this reason, researchers have developed hybrid networks that combine these two architectures in order to reduce the computational load of Transformers and to enhance the spatial awareness of Mamba [72], [73]. These hybrid networks are generally used in clinical applications such as report generation [74], image enhancement [75] and segmentation [76]. Highly successful results are obtained by combining Swin-Transformer [77], Segment Anything Model (SAM) [78] and diffusion-based encoders [22] with Mamba modules. The fundamental rationale of this integration is to bring together the precise inter-patch token correlations of these structures with the sequence-processing efficiency of Mamba [8], [72]. Nevertheless, the combination of these paradigms generally results in extremely dense architectures with enormous parameter counts [22], [72]. This increases the hardware requirements and complicates the training process [22]. The innovative approaches in which Transformers are integrated with Mamba, and the challenges encountered, are presented in Table 5.

Table 5. Transformer-Mamba Integration Studies

Reference	Clinical Task & Dataset	Proposed Architecture	Key Performance Metrics	Methodological Limitations
[79]	Task: 3D Neuron Seg. Modality: 3D Microscopy Dataset: Three datasets	Dual U-Net architecture utilizing cross-attention modules for diverse fields-of-view and Mamba for adaptive high-resolution fusion.	Achieved the highest F1 scores and ensured continuous neuronal segmentation while maintaining computational efficiency.	Running dual U-Nets in parallel inherently doubles the memory footprint, restricting the maximum processable patch size during training.
[80]	Task: Preoperative Grading Modality: CE-Ultrasound Dataset: Not specified	Utilizes a Vision-Language Model (LLaVA-Med) for knowledge distillation into a Distilled Mamba Transformer Convolutional backbone.	Achieved highly competitive non-invasive clear cell renal cell carcinoma grading accuracy.	Utilizing a massive pre-trained VLM for distillation introduces significant offline computational overhead and requires meticulous prompt engineering.
[81]	Task: Cardiac Segmentation Modality: Echo & MRI Dataset: CAMUS, ACDC	Aggregates multi-scale contexts by combining spline-based Kolmogorov-Arnold Networks (KAN) with Selective Mamba layers and attention gates.	Outperformed U-Net variants, TransUNet, and U-Mamba (95.18% on CAMUS), exhibiting strong zero-shot generalization.	Spline-based KAN layers introduce specialized non-linear modeling that may require highly delicate hyperparameter tuning to avoid overfitting.
[82]	Task: Surgical Phase Rec. Modality: Endoscopic Video Dataset: Cholec80, ESD385	Synthesizes a Mamba-based architecture with Scaled Residual TranMamba blocks via a Hierarchical Sampling Strategy.	Achieved real-time execution speeds alongside a state-of-the-art 87.70% accuracy for endoscopic submucosal dissection.	Modeling extensive temporal sequences via hierarchical sampling risks ignoring instantaneous surgical anomalies or extremely rapid phase transitions.
[83]	Task: Cancer Classification Modality: Histopathology	Employs a Transformer block with Haar wavelet transforms alongside a Cross-Scale Mamba	Synergistic integration of signal processing and sequence	Integrating both multi-scale Transformer attention and Mamba in a single framework may negate the

	Dataset: Chaoyang CRC	Block for hierarchical feature fusion.	modeling achieved state-of-the-art classification performance.	inherent efficiency advantages of pure Mamba.
[84]	Task: Image Synthesis Modality: Medical Images Dataset: Benchmark datasets	Utilizes a MambaTrans architecture to embed radiomics prompts, guiding denoising diffusion implicit models via Monte Carlo sampling.	Significantly outperformed SOTA methods in structural quality, uncertainty quantification, and clinical reliability.	Generating high-resolution images via diffusion combined with complex MambaTrans architectures limits real-time synthesis capabilities.
[85]	Task: Scribble-based Seg. Modality: Medical Imaging Dataset: Two datasets	Integrates CNN, ViT, and VMamba architectures via a cross-supervisory mechanism using pseudo labels.	Successfully surpassed baseline weakly-supervised learning frameworks relying solely on U-Net or Swin-UNet models.	Simultaneous integration of CNN, ViT, and Mamba networks significantly increases VRAM requirements during the training phase.
[86]	Task: Incomplete Modality Seg. Modality: Multimodal MRI Dataset: BraTS 18/20	Uses Multimodal Feature Distillation and a Cross-Modal Fusion module leveraging Mamba-Transformer hybrids for long-range dependencies.	Significantly outperformed current state-of-the-art missing modality methods by minimizing boundary discrepancies.	Accuracy remains inherently restricted when missing sequences contain exclusive, critical pathological signatures not present in available modalities.
[87]	Task: Autonomous Seg. Modality: MRI / CT Dataset: ACDC	Develops a trainable Visual Mamba adapter mechanism paired with a pre-trained Segment Anything Model (SAM) ViT encoder.	Eradicates reliance on manual prompts, achieving a 0.18 Dice improvement and 20.38 mm Hausdorff distance reduction vs. AutoSAM.	Due to reliance on a massive pre-trained ViT foundation model, the architecture remains heavyweight, challenging edge deployment.
[88]	Task: Pathology Grading Modality: Histopathology	Integrates graph attention mechanisms with Mamba's selective state space mechanism	Achieved grading accuracy improvements of up to 2.90%	Combining graph neural networks with Mamba necessitates complex graph construction, increasing

	Dataset: Two datasets	to capture tumor-microenvironment topology.	over the highest baselines with superior interpretability.	computational overhead for gigapixel WSIs.
[89]	Task: Action Triplet Detection Modality: Surgical Video Dataset: CholecT50	Weaves Mamba and Transformer blocks within an enhanced Spatial-Temporal Perception decoder using textual-temporal trails.	Confirmed superior performance in instrument localization and accurate action triplet detection in long-sequence surgical videos.	Dependency on a pre-trained text encoder implies the model's performance may be highly sensitive to the quality of the specific textual prompts.
[90]	Task: Tumor Segmentation Modality: Ultrasound Dataset: Six datasets	Bio-inspired architecture integrating ConvNets (local extraction), self-attention (global focus), and Mamba (linear dependency).	Broadly outperformed existing models in binary accuracy, IoU, and Dice coefficients across six ultrasound cancer datasets.	Tri-integrating ConvNets, self-attention, and Mamba inherently creates a massive footprint, making it unsuitable for portable point-of-care ultrasound.
[74]	Task: Report Generation Modality: X-ray & Pathology Dataset: IU X-Ray, COV-CTR	Integrates Mamba into a Transformer decoder to enhance long sequence dependency capturing for multi-modal clinical narratives.	Mitigated normal/abnormal data bias, resulting in significant improvements in BLEU scores and phrase overlap.	The ability to maintain high coherence over extremely long or complex clinical narratives might still be challenged by rigid vocabulary constraints.
[91]	Task: Super-resolution Modality: MRI Dataset: Clinical datasets	Leverages a conditional Diffusion Transformer framework coupled with a hybrid Mamba decoder (MambaVision Mixer).	Faithfully restores intricate anatomical structures directly from noisy latent inputs, outperforming standard U-Net diffusion approaches.	Relying on a complex MambaVision Mixer within the decoding phase of a diffusion network compounds inferential latency.

[92]	Task: Synthetic CT Gen. Modality: CBCT to CT Dataset: Brain, H&N	Conditional diffusion framework synergizing Group Multi-axis Hadamard Product Attention with a Mamba bottleneck module.	Drastically minimized CBCT artifacts, yielding peak signal-to-noise ratios up to 31.44 dB for synthetic CT generation.	Integrating complex Hadamard attention alongside Mamba mechanisms into diffusion models further compounds inference latency for radiotherapy.

4.3.4 Novel Scanning Mechanisms

The standard Mamba architecture was developed for 1D sequential data processing. However, various enhancements have subsequently been made so that it can be used in image analysis [22]. The 2D and 3D spatial complexity of medical images causes standard scanning methods to disrupt anatomical integrity and to lead to semantic loss [19]. To overcome this limitation, innovative algorithms such as cross-scan [22], fractal [19], bidirectional slicing [9] and graph-based [56] scanning paths have been proposed. Because preserving the continuous boundaries of multi-dimensional anatomical structures is critical, these content-aware scanning methods are widely applied in complex volumetric tasks such as 3D brain tumor segmentation [8], denoising [22] and image reconstruction [9]. Although customized scanning methods reduce directional forgetting by preserving spatial adjacency, they interrupt continuous memory-access patterns and thereby cause latency during the inference stage [19], [22]. Examples of the applications of these innovative scanning algorithms in clinical imaging modalities are presented in Table 6.

Table 6. Novel Scanning Mechanism Studies

Reference	Clinical Task & Dataset	Proposed Architecture	Key Performance Metrics	Methodological Limitations
[2]	Task: 3D Volumetric Seg. Modality: MRI & CT Dataset: AMOS, BraTS	Integrates custom 3D depthwise convolutions and a novel Tri-scan approach to efficiently process high-resolution volumes.	Consistently outperformed leading models (nnUNet, CoTr) with highly competitive accuracy and superior linear computational efficiency.	Applying multi-directional scanning to 3D high-resolution volumes may induce localized latency spikes during real-time surgical applications.
[93]	Task: Artifact Reduction/Recon. Modality: CBCT Dataset: Head and neck data	Unsupervised data adaptation and image restoration pipeline enhanced by a spiral-Mamba module.	Eliminates wedge artifacts and structural distortions, showing highly promising restoration on real limited-angle CBCT.	The multi-stage unsupervised process may struggle to precisely recover fine structures in extreme limited-angle scenarios with insufficient sampling.
[94]	Task: Medical Image Seg. Modality: Ultrasound Dataset: Cross-domain US	Integrates fast Fourier transforms with a deformable windowed selective scan to dynamically adjust scanning paths.	Achieved exceptional Dice scores across four cross-domain ultrasound datasets, substantially outperforming leading SOTA methods.	Dynamically adjusting scanning paths via a learnable offset field may increase susceptibility to severe speckle artifact misinterpretations.
[95]	Task: Polyp Segmentation Modality: Endoscopy Dataset: Kvasir-SEG, CVC	Addresses polyp scale variations through a Hierarchical Scan Mamba implementing global, meso-scale, and local parallel scanning.	Achieved linear complexity and outperformed state-of-the-art benchmarks in mean Dice and IoU metrics.	Running three parallel multi-scale scanning branches simultaneously may introduce memory allocation bottlenecks for high-resolution video.

[96]	Task: Brain Disorder Ident. Modality: fMRI Dataset: ABIDE	Utilizes a Mamba-based multi-scale dynamic graph learning module to extract spatio-temporal features, bypassing sliding windows.	Scored 73.9% accuracy and an AUC of 74.9% for brain disorder identification on the ABIDE multicenter dataset.	Extracting dynamic spatio-temporal graphs from massive 4D fMRI datasets likely imposes steep memory constraints during multi-task optimization.
[97]	Task: Image Denoising Modality: Low-dose CT Dataset: Not specified	Integration of an adaptive gradient-guided scanning strategy to overcome the limitations of fixed scanning paths.	Vastly superior quantitative metrics and visual quality compared to traditional denoising approaches while preserving edges.	The gradient-guided strategy may introduce computational overhead during dynamic path generation, impacting real-time processing speeds.
[98]	Task: Lesion Segmentation Modality: Pathology WSI Dataset: Public/Private	Introduces a bidirectional scanning Mamba encoder synergized with a learnable frequency-domain encoding module for MIL.	Matched fully supervised benchmarks using only sparse image-level labels in a linear-complexity multi-instance learning framework.	Reliance on initial CAMs for pseudo-label generation might propagate noise if self-correction fails to identify false positives.
[99]	Task: Multi-target Seg. Modality: Derm. & Endoscopy Dataset: ISIC, Kvasir-SEG	Multi-directional dynamic modeling featuring Parallel Cross-axis Mamba attention and an Edge Feature Enhancement Model.	Yielded robust Dice similarity coefficients (up to 94.14), demonstrating superior generalization and discontinuity reduction.	Parallel cross-axes mechanisms combined with explicit edge enhancement may increase architectural complexity, slowing edge device inference.
[100]	Task: Tumor Segmentation Modality: PET-CT / MRI Dataset: STS, LMNCLC	Synergizes a diffusion model for missing modality generation with a cross-modal interaction Mamba network.	Significantly surpassed SOTA segmentation methods on three distinct datasets, proving highly resilient to modality incompleteness.	Integrating a score-based diffusion model directly into the segmentation pipeline drastically increases computational cost and inference latency.

[101]	Task: Multimodal Fusion Modality: Visible-IR & Medical Dataset: Unified benchmarks	Implements an interval sampling cross-scan mechanism to convert non-causal images into ordered sequence patches.	Achieves exceptional generation performance and rapid execution speeds for multimodal image fusion tasks.	Interval sampling and bi-stride extraction run the risk of inadvertently skipping isolated, single-pixel anomalies critical for diagnosis.
[102]	Task: Coronary Artery Seg. Modality: X-ray Angiography Dataset: In-house 739 images	Employs a rotation-aligned multi-directional state-space scan alongside a persistent high-resolution bypass.	Achieved a Dice score of 0.8706 and reduced fractal dimension deviation by up to 57% relative to U-Net.	Rotation-aligned scanning and un-downsampled high-resolution bypasses significantly increase memory consumption, restricting batch sizes.
[103]	Task: Fluorosis Diagnosis Modality: Skeletal X-ray Dataset: SFXRay	Dual-branch convolutional state space model featuring a novel multi-window cross-scan mechanism.	Achieved diagnostic accuracies of 83.33% and 66.67% for binary and multi-classification tasks on a newly constructed dataset.	Diagnostic generalizability remains constrained by the limited scale and diversity of the single-source dataset.
[104]	Task: Macular Edema Diag. Modality: Fundus Imaging Dataset: IDRiD, Messidor	Synergizes Dempster-Shafer evidence theory with a multi-scan Mamba module to synthesize light and dark fundus regions.	Delivers unparalleled accuracy and robustness across three independent clinical image cohorts via confidence-aware networks.	Rigidly separating features into light and dark domains may discard valuable diagnostic gradients in mid-tone pathological transitions.
[105]	Task: Retinal Disease Detect. Modality: OCT Dataset: Not specified	Fuses multi-scale convolutions with customized retinal Mamba blocks via a multi-directional selective mechanism.	Yielded formidable overall classification accuracies of 98.98% and 96.21% across two highly variable retinal datasets.	Multi-directional selective scanning likely encounters feature misalignment if OCT images contain severe motion artifacts.

[106]	Task: Noise Estimation Modality: Optical Molecular Dataset: Clinical & in vivo	Integrates proximal gradient descent with deep equilibrium unfolding Mamba networks to decouple complex noise.	State-of-the-art noise suppression performance with highly efficient parameter utilization confirmed via in vivo experiments.	Deep equilibrium models require complex implicit differentiation during training, potentially leading to unstable convergence.

4.4 Summary of Included Studies

When the studies examined in detail in the Thematic Analysis (Section 4.3) are evaluated as a whole, the current state of the Mamba architecture in medical image analysis and the methodological trends become evident. The reduction in parameter counts and processing cost compared with Transformer (ViT) models, and the high precision achieved through the uninterrupted processing of long sequences, particularly in 3D volumetric data, are the two principal contributions offered by the Mamba architecture (see Tables 3-6). The analysis carried out in light of the Quality Assessment criteria applied in Section 3.6 shows that the great majority of the studies in the literature are at the Medium or High level in terms of methodological transparency. In particular, a significant proportion of the Pure Mamba and CNN-Mamba hybrid studies have made a major contribution to reproducibility by sharing their architectural designs and source code as open source. However, when examined in terms of datasets, it was found that the research predominantly relied on standard open-access datasets such as BraTS, Synapse, ISIC, and Kvasir-SEG (see Tables 3-6), and that some studies remained at the Low or Medium level owing to a lack of generalization (domain generalization) to different clinical centers and unseen domains. Although many of the models have overcome hardware limitations, studies tested in real time or on edge devices remain a minority.

A summary of the distribution of the architectural typologies across the prominent clinical imaging modalities and of the study density is presented in Table 7. This table reveals which architectural approach is beginning to become a standard for which clinical data type.

Table 7. Architectural Typologies across Primary Medical Modalities

Architectural Typology	MRI & CT (3D Volumetric / Cross-sectional)	Ultrasound & Echocardiography (Noisy / Dynamic)	Pathology (WSI / Gigapixel)	Endoscopy & Dermoscopy (2D Surface / Video)	Primary Clinical Tasks Evaluated
CNN-Mamba Hybrids	High Focus (Leveraging CNN local feature extraction for diffuse tumors)	Moderate Focus (Addressing speckle noise via hybrid features)	Moderate Focus (Patch-level localized feature extraction)	High Focus (Boundary refinement for polyps and skin lesions)	Segmentation, Detection, Artifact Reduction
Pure Mamba Adaptations	High Focus (Handling massive 3D sequences efficiently)	Low Focus	High Focus (Processing gigapixel sequences without ViT memory bottlenecks)	Moderate Focus	Segmentation, Multi-class Classification, Super-Resolution
Transformer-Mamba Integrations	Moderate Focus (Balancing self-attention with sequence efficiency)	Moderate Focus	Low Focus (Often too computationally heavy for WSIs)	High Focus (Precise spatial alignment for complex boundaries)	Segmentation, Deformable Registration, Image Synthesis
Novel Scanning Mechanisms	High Focus (Multi-directional / Tri-scan adaptations for 3D topologies)	Moderate Focus	Moderate Focus (Graph-driven traversing of patches)	Low Focus (Raster scan is usually sufficient for 2D)	Volume Reconstruction, Multi-target Segmentation, Cross-modal Synthesis

5. Discussion

5.1 The Rise of State-Space Models in Medical Imaging

The fundamental aim in medical image analysis is to capture local lesion features, such as tumor tissue, with high precision while simultaneously preserving the global context of the organ or tissue under examination. Although CNNs, which have been frequently used in studies over the years, have proven their success in local feature extraction in many different studies, they remain inadequate for modeling the long-range dependencies of anatomical structures owing to their limited receptive field. ViT architectures set a new standard by solving this global-context problem through self-attention mechanisms. However, the high-resolution and volumetric 3D structures inherent to medical images, when combined with the sequence-length-dependent quadratic computational complexity ($O(N^2)$) of ViTs, have caused memory and hardware bottlenecks.

The fundamental reason that Mamba and Selective SSMs have gained acceptance in the literature in such a short time is precisely their ability to combine the advantages of these two opposing poles mathematically [17]. By processing sequential data with linear computational complexity ($O(\mathrm{N})$), Mamba has enabled high-resolution 2D and 3D medical data to be analyzed as longer sequences without being subjected to the excessive downsampling found in conventional Transformer models [17], [19]. The findings obtained from the literature review (Table 7), particularly the rapid emergence of Mamba adaptations as dominant in volumetric scans such as MR and CT, show that this is not a methodological coincidence. By combining the global awareness of ViTs with the linear processing efficiency of CNNs, Mamba has formed a natural backbone alternative for medical image analysis that has a low computational cost but high semantic comprehension.

5.2 Methodological Bottlenecks and Architectural Limitations

Although the Mamba architecture solves the quadratic complexity problem of ViTs by theoretically offering a linear processing cost, its practical applications in medical image analysis bring serious hardware bottlenecks [22]. The fundamental reason for this bottleneck is that Selective SSMs, which are inherently designed to process 1D sequential data, require multi-directional scanning mechanisms in order to comprehend 2D and 3D anatomical structures. When the studies in the literature are examined, it is observed that separate state-space matrices must be computed and held in GPU memory for each direction of the image [19]. This particularly increases memory usage in volumetric CT and MR scans. In addition, in the CNN-Mamba hybrid approaches most preferred in the literature, processing local and global features independently in parallel branches and subsequently fusing them increases the parameter count [22]. As a result, despite the high diagnostic accuracy of these hybrid and multi-directional models, their use in their current form on hospital hardware with limited memory, particularly in real-time surgical navigation and ultrasonography applications, is not practical at present [18], [22].

The computational efficiency of Mamba relies on flattening high-resolution 2D and 3D medical volumes into 1D sequences. However, this process leads to spatial semantic loss and to the fragmentation of neighborhood relationships, particularly in tumor tissues with indistinct

boundaries or in complex microscopic vascular networks [9]. Unfolding the image matrix into a one-dimensional vector causes pixels that are adjacent in 2D or 3D space to become distant from one another [9]. The studies examined in the literature (Section 4.3.4) aim to prevent this loss. Although semantic loss is reduced by the various proposed scanning mechanisms, the anatomical continuity and tissue integrity required by the nature of medical images are still not fully preserved in current Mamba architectures.

Another critical bottleneck in the studies examined is the lack of overfitting control and clinical generalizability. The multi-branch hybrid approaches that offer high performance in the literature tend to memorize the datasets owing to their high parameter counts [107]. However, when these networks encounter multi-center variable scanning protocols, noisy clinical ultrasounds (speckle noise), or missing-modality situations, they experience serious performance degradation [107]. For this reason, in order to integrate Mamba-based models into clinical workflows, it is necessary to reduce their dataset dependencies and to increase their robustness against real-world conditions [107].

5.3 Open Challenges and Future Directions

Despite the rapid rise of Mamba and Selective SSMs in medical image analysis, the current state of the literature contains several open challenges and potential research directions on which researchers should focus:

- **Native 3D Selective SSMs:** The majority of current Mamba adaptations force images into 1D sequences (flattening) or multi-directional 2D scans in order to process volumetric medical images [9]. Future studies should focus on the design of memory-efficient Mamba blocks and isotropic scanning algorithms that can directly process 3D volumes without disrupting spatial topology and inter-pixel neighborhood relationships [2], [9] .
- **Medical Foundation Models:** The lack of labeled medical data is one of the greatest obstacles limiting the performance of supervised learning models. The capacity of Mamba to process long sequences at low hardware cost offers an opportunity to build Mamba-based Medical Foundation Models [108], [109] on large, unlabeled clinical datasets using self-supervised learning paradigms [76], [107]. These models will provide an infrastructure that can easily be fine-tuned for different hospital systems and diseases [107].
- **Edge Devices and Real-Time Integration:** Although the hybrid architectures in the literature achieve high accuracy rates, problems are encountered in integrating them into low-capacity devices owing to their parameter density [22]. Mamba networks can be made lighter using model-compression techniques such as pruning, quantization, and knowledge distillation [18], [22]. In this way, edge-computing scenarios such as instant diagnosis on portable ultrasound devices or real-time navigation in surgical robotic systems can be realized [18].

6. Limitations

This study has some methodological limitations that should be taken into account. First, the Mamba architecture is a very new, exponentially growing, and fast-paced research area in the

literature. Because the literature review was concluded in the second quarter of 2026, potentially breakthrough studies published after this date were not included in this review. Second, the clinical tasks, the modalities used (MRI, CT, Ultrasound), and the performance evaluation metrics (Dice, IoU, Accuracy) on which the 445 studies included in the PRISMA flow focused exhibit very high heterogeneity among themselves. Owing to this excessive diversity, a quantitative meta-analysis statistically combining the results of the studies could not be conducted. Instead, a Qualitative Critical Appraisal had to be presented. Finally, the search strategy was limited to full-text articles published only in English. Potentially significant findings in conference proceedings (MICCAI, CVPR), preprints, and unpublished gray literature, which are of great importance in the deep learning literature, were not included in this study.

7. Conclusion

This systematic review comprehensively reveals, across 445 studies, the rapid adoption of the Mamba architecture in medical image analysis and the transformative potential it offers. The review shows that, owing to its linear ($O(N)$) computational complexity, Mamba overcomes the quadratic hardware bottleneck ($O(N^2)$) of ViTs and sets a new standard for high-resolution and 3D volumetric medical data. It was found that the most dominant, efficient, and clinically applicable approach in the literature at present is the CNN-Mamba hybrid, which combines the local feature extraction power of CNNs with the global-context capability of Mamba.

Despite all these innovative steps, several methodological obstacles remain before Mamba-based models can be directly integrated into clinical decision-support systems. Foremost among these obstacles are the spatial information losses arising from the flattening of images into 1D sequences, the heavy parameter loads caused by hybrid networks, and the lack of multi-center external validation.

In the future, native 3D Mamba blocks that preserve spatial topology should be developed, and model-compression techniques aimed at edge-computing hardware should be integrated into the literature. With these methodological improvements, the Mamba architecture is anticipated to become not merely a passing trend in the field of medical artificial intelligence but the principal backbone architecture that will shape next-generation foundation models.

Declarations

Funding

Author Ali Emre Gök is supported by the TÜBİTAK 2211-A National PhD Scholarship Program.

Competing Interests

The authors declare that they have no competing financial or non-financial interests that are relevant to the content of this article.

Ethics Approval and Consent to Participate

Not applicable. This study is a systematic review based exclusively on previously published literature and did not involve any new experiments on human participants or animals performed by the authors.

Data Availability

Data sharing is not applicable to this article, as no new datasets were generated or analysed during the current study. All data discussed in this systematic review are available in the cited published articles, and the complete list of the 445 included studies is provided in the Supplementary Information.

Author Contributions

M.Y. and A.E.G. conceived and designed the study. A.E.G. performed the literature search, study screening, and data extraction. M.Y. and A.E.G. carried out the qualitative critical appraisal and analysis. A.E.G. prepared the figures and tables. M.Y. and A.E.G. wrote the main manuscript text. Ş.T. supervised the study and critically revised the manuscript for important intellectual content. All authors reviewed and approved the final manuscript.

References

[1] Kumar A, Mahendran N (2026) MedScope-LDx: A comprehensive approach for advanced lesion analysis in medical imaging. Biomedical Signal Processing and Control 111:108352

[2] Wang C, Xie Y, Chen Q, Zhou Y, Wu Q (2025) A comprehensive analysis of Mamba for 3D volumetric medical image segmentation. Pattern Recognition 112701

[3] 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

[4] 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 international publishing, Cham

[5] Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25

[6] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention pp 424-432. Springer International Publishing, Cham

[7] Mabrok MA, Zafari Y (2025) X-VMamba: Explainable Vision Mamba. arXiv preprint arXiv:2511.12694

[8] Xu J, Lan Y, Zhang Y, Zhang C, Stirenko S, Li H (2025) CDA-mamba: cross-directional attention mamba for enhanced 3D medical image segmentation. Scientific Reports 15(1):21357

[9] Fan C, Yu H, Huang Y, Wang L, Yang Z, Jia X (2025) Slicemamba with neural architecture search for medical image segmentation. IEEE Journal of Biomedical and Health Informatics

[10] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al (2017) Attention is all you need. Advances in Neural Information Processing Systems 30

[11] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T et al (2020) An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929

[12] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y et al (2021) Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306

[13] Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M (2022) Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision pp 205-218. Springer Nature Switzerland, Cham

[14] Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H (2021) Training data-efficient image transformers & distillation through attention. In: International conference on machine learning pp 10347-10357. PMLR

[15] Ferdous GJ, Sathi KA, Hossain MA, Hoque MM, Dewan MAA (2023) LCDEiT: A Linear Complexity Data-Efficient Image Transformer for MRI Brain Tumor Classification. IEEE Access 11:20337–20350. https://doi.org/10.1109/ACCESS.2023.3244228

[16] Wang A (2025) Vision Transformers (ViTs): A New Era in Computer Vision–A Review. Journal of Computing and Electronic Information Management 18(3):23-30

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

[18] Aalishah R, Navardi M, Mohsenin T (2025) Medmambalite: Hardware-aware mamba for medical image classification. In: 2025 IEEE Biomedical Circuits and Systems Conference (BioCAS) pp 224-228. IEEE

[19] Zhang Z, Ma Q, Zhang T, Chen J, Zheng H, Gao W (2025) Switch-UMamba: Dynamic scanning vision Mamba UNet for medical image segmentation. Medical Image Analysis 103792

[20] Zhou Y, Sun L, Xiong X, Ti G, Yang S (2025) GCNet-Mamba: Leveraging State Space Models and CNN for Medical Image Classification. Expert Systems with Applications 130733

[21] Lin W, Luo Y, Ling J, Li F, Qin J, Yin Z, Yao S (2025) Mamba-Convolutional UNet for multi-modal medical image synthesis. Medical Physics 52(10):e70029

[22] Kurniawan YI, Rachmadi MF, Ramadhan AW, Jatmiko W (2025) Mamba-Based Deep Learning Methods in Medical Image Analysis: A Systematic Literature Review. IEEE Access 13:208801-208831

[23] Gu A, Dao T, Ermon S, Rudra A, Ré C (2020) Hippo: Recurrent memory with optimal polynomial projections. Advances in Neural Information Processing Systems 33:1474-1487

[24] Dao T, Gu A (2024) Transformers are ssms: Generalized models and efficient algorithms through structured state space duality. arXiv preprint arXiv:2405.21060

[25] Ruan J, Li J, Xiang S (2024) Vm-unet: Vision mamba unet for medical image segmentation. ACM Transactions on Multimedia Computing, Communications and Applications

[26] Latibari BS, Nazari N, Brignac D, Sayadi H, Homayoun H, Mahalanobis A (2026) Is Mamba Reliable for Medical Imaging? In: 2026 27th International Symposium on Quality Electronic Design (ISQED) pp 1-6. IEEE

[27] 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, Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/PAIS66004.2025.11126051

[28] Liu Y, Tian Y, Zhao Y, Yu H, Xie L, Wang Y et al (2024) Vmamba: Visual state space model. Advances in Neural Information Processing Systems 37:103031-103063

[29] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372

[30] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S et al (2014) Generative adversarial nets. Advances in Neural Information Processing Systems 27

[31] Ho J, Jain A, Abbeel P (2020) Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems 33:6840-6851

[32] He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision pp 2961-2969

[33] 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):e70139

[34] 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):e70177

[35] Santhosh M, George SV, Surendhan A, Arisuthan C (2025) EM-DR: Efficient-Mamba vision model for early detection and classification of diabetic retinopathy. In: 2025 International Conference on Frontier Technologies and Solutions (ICFTS) pp 1-9. IEEE

[36] Aithal SJ, Bharti V, Singh AK, Gupta BB (2026) EFSVMNet: Enhanced Feature-Selective Vision Mamba Network for Medical Image Segmentation. ACM Transactions on Multimedia Computing, Communications and Applications

[37] Li Q, Ni T, Wang X, Liu H (2026) FocusMamba: A Local–Global Mamba Framework Inspired by Visual Observation for Brain Tumor Segmentation. Applied Sciences 16(7):3571

[38] Li G, Huang Q, Wang W, Liu L (2025) Selective and multi-scale fusion Mamba for medical image segmentation. Expert Systems with Applications 261:125518

[39] Feng F, Liu T, Zheng F, Sun Y, Wang H (2026) Efficient Breast Cancer Segmentation via Brownian Bridge Diffusion with Semantic Fusion Strategy. Pattern Recognition 112993

[40] Hasan M, Hasan MK (2026) Hawk-Net: Medical image segmentation and classification using multi-scale convolutional self-attention-based image processor with DK-CNN-Mamba-xAttention Fusion Network. Biomedical Signal Processing and Control 115:109401

[41] Zhang Y, Zhu H, Qian T, Chen N, Xu X, Huang B (2025) Graph mapping Mamba network for automated macular edema diagnosis from fundus images. Information Fusion 103476

[42] Yu C, Meng X, Li Y, Zhao Z, Zhang Y (2025) TG-Mamba: Leveraging text guidance for predicting tumor mutation burden in lung cancer. Computerized Medical Imaging and Graphics 102626

[43] Chen Y, Wang X, Qian C, Guan X, Dai X, Gao P (2025) MambaBIR: A Residual Pyramid Network for Brain Image Registration with State-Space Model. Expert Systems with Applications 129335

[44] Jin L, Peng Y, Wang J, Jiang Y, Zhang K (2025) Hierarchical multi-scale Mamba generative adversarial network for multi-modal medical image synthesis. Expert Systems with Applications 280:127451

[45] Zhang Y, Wang J, Wang X, Ji X, Chen K, Yang J et al (2025) Spectral-X: Latent prior enhanced spectral CT restoration with mamba-assisted X-net. Computerized Medical Imaging and Graphics 102696

[46] Wang T, Zhou Z, Jin C, Liu T, Shi B, Zhou G et al (2026) USRMamba: Adaptive Routing-Guided State Space Model for Ultrasound Super-Resolution. IEEE Journal of Biomedical and Health Informatics

[47] Chen W, Hou F, Shen Y (2025) CoroYOLO: a novel colorectal cancer detection method based on the Mamba framework. Frontiers in Physics 13:1597378

[48] He Z, Chen Y, Yang Y, Ouyang Z, Luo Y, Chen L (2026) FreqMambaMark: Wavelet-Mamba-Driven Robust Medical Image Watermarking. Knowledge-Based Systems 115344

[49] Zheng H, Wang C, Zhou Y, Cai C, Chen Z (2026) Source-Free Active Domain Adaptation for Brain Tumor Segmentation via Mamba and Region-Level Uncertainty. Brain Sciences 16(3):300

[50] He D, Li S, Jiang B, Yan H (2025) Spec2VolCAMU-Net: a spectrogram-to-volume model for EEG-to-fMRI reconstruction based on Multi-directional Time–Frequency Convolutional Attention Encoder and Vision-Mamba U-Net. Journal of Neural Engineering 22(5):056042

[51] Li M, Zhao K, Zhao J, Chen X, Fang K, Yang W, Zhang X (2025) Real-time generation of renal artery hemodynamic parameters using a point cloud-based deep learning model. Computer Methods in Biomechanics and Biomedical Engineering 1-13

[52] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18(2):203-211

[53] Ma J, Li F, Wang B (2024) U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722

[54] Gong H, Kang L, Wang Y, Wang Y, Wan X, Wu X, Li H (2025) nnmamba: 3d biomedical image segmentation, classification and landmark detection with state space model. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI) pp 1-5. IEEE

[55] Zhu L, Liao B, Zhang Q, Wang X, Liu W, Wang X (2024) Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417

[56] 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 115326

[57] Li Y, Mao Z, Qin F, Peng Y, Zhang G, Xi X et al (2026) A Local-Global Fusion Vision Mamba UNet Framework for medical image segmentation. Engineering Applications of Artificial Intelligence 169:113987

[58] Feng W, Lu Y, Li X, Yung KL, Ip WH (2026) Fog/Edge-Aware State Space Models for Multi-Task Chest X-ray Report Generation and Lesion Detection. IEEE Journal of Biomedical and Health Informatics

[59] Baraka M, Elbadry E, Abourida O, Albaradi A, Wagih Y, Gamal M, Torki M (2026) Multi-Architecture Deep Learning for CBCT Segmentation of Dental Hard Tissues and Pulp in Mixed Dentition. Journal of Dentistry 106344

[60] Cao Z, Du B, Hu D, Zhou W, Gao S, Tan R, Xu J (2026) Few-Shot Medical Image Segmentation With Hierarchical Hypercorrelation Vision State-Space Model. IEEE Transactions on Computational Social Systems

[61] Wang J, Ruan G, Li H, Nian Y, Wei Y, Lei Y (2026) Semba: A semantic–visual fusion method based on MLLM and Mamba for IBD subtype identification. Biomedical Signal Processing and Control 121:110358

[62] Cao Y, Dan T, Yang Y, Wu G (2026) Geo-Mamba: Geometry-informed state-space learning of functional brain organization. Medical Image Analysis 104081

[63] Xia L, Chen HY, Cao YW, Gan CQ, Zhao JZ, Zheng WH et al (2026) Prompt-mamba filtering networks for accurate hepatocellular carcinoma lesion segmentation in abdominal CT. npj Digital Medicine

[64] Wang J, Tang H, Wu Z, Zhang Y, Xi Y, Chen Y et al (2025) Twin-ViMReg: DXR driven synthetic dynamic Standing-CBCTs through Twin Vision Mamba-based 2D/3D registration. Computerized Medical Imaging and Graphics 102648

[65] Wang W, Jin Z, Liu X, Chen X (2025) NaMA-Mamba: Foundation model for generalizable nasal disease detection using masked autoencoder with Mamba on endoscopic images. Computerized Medical Imaging and Graphics 122:102524

[66] Wang Y, Guo T, Yuan W, Shu S, Meng C, Bai X (2025) Mamba-based deformable medical image registration with an annotated brain MR-CT dataset. Computerized Medical Imaging and Graphics 123:102566

[67] Nie X, Chen X, Luo J, Huang S, Lei Q, Wu M et al (2026) M3amba: Multi-focus multi-domain Mamba network for liver tumor segmentation on multicenter MRI. Biomedical Signal Processing and Control 120:110240

[68] Liu Y, Zhang Y, Dong L, Wang T, Wu S, Zhu S, Lei B (2026) Diff-magnifier: Utilizing state space model for diffusion processes in breast tumor pathological image super-resolution and classification. Pattern Recognition 113649

[69] Zhang R, Ding W, Han H, Cao Y, Wang H, Huang Y (2025) StrFilter: Multi-Modal Medical Image Fusion via Structure-Oriented Adaptive Filtering. Information Fusion 103888

[70] Wei W, Xia F, Zhang D, Zhou W, Wang X, Gao Y et al (2026) Pixel-level Radiomics and Deep Learning for Predicting Ki-67 Expression in Breast Cancer Based on Dual-modal Ultrasound Images. Academic Radiology

[71] Pan Z, Zhou H, Ren Q, Hou X, Dai J, Liu X et al (2026) X2Shape: CT-free 3D multi-organ reconstruction with biplanar X-rays. Medical Image Analysis 104085

[72] Lieber O, Lenz B, Bata H, Cohen G, Osin J, Dalmedigos I et al (2024) Jamba: A hybrid transformer-mamba language model. arXiv preprint arXiv:2403.19887

[73] Hatamizadeh A, Kautz J (2025) Mambavision: A hybrid mamba-transformer vision backbone. In: Proceedings of the Computer Vision and Pattern Recognition Conference pp 25261-25270

[74] Yang R, Li F (2026) A hybrid transformer architecture with Mamba for medical image report generation. Multimedia Systems 32(1):53

[75] Xie Z, Miao G, Chang H (2025) MTSR: Mamba-Transformer Super-Resolution Model for Hyperspectral Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

[76] Liu J, Yang H, Zhou HY, Yu L, Liang Y, Yu Y et al (2024) Swin-UMamba†: Adapting Mamba-based vision foundation models for medical image segmentation. IEEE Transactions on Medical Imaging 44(10):3898-3908

[77] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z et al (2021) Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision pp 10012-10022

[78] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L et al (2023) Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision pp 4015-4026

[79] Yan H, Zhang Y, Li Z, Guo J, Zhai H, Liu J et al (2025) Glancing Beyond Patch: Spatial Contextual Cues for 3D Neuron Segmentation. IEEE Transactions on Medical Imaging

[80] Zou M, Liu Y, Shang J, Liu Y, He H, Chen Y et al (2026) VL-Grading: A vision–language distillation framework for preoperative grading of clear cell renal cell carcinoma from multi-phase contrast-enhanced ultrasound. Biomedical Signal Processing and Control 121:110321

[81] Ali H, Mehmood A, Tolno DN, Thierry SSG, Saeed M, Ahmed N (2026) KMS-Net: Kolmogorov–Arnold Based Multi-Scale Attention Network for Cardiac Segmentation. IEEE Access

[82] Zhang X, Zhang Q, Chen J, Zhou C, Wang Y, Zhang Z et al (2026) SPRMamba: surgical phase recognition for endoscopic submucosal dissection with mamba. IEEE Sensors Journal

[83] Chen L, Li C, Meng F, Tai J, Wang K (2026) CRC-Former: frequency-domain adaptive swin-transformer for colorectal cancer histopathology classification. Frontiers in Physiology 17:1792357

[84] Zhao J, Li S (2025) Radiomics-Driven Diffusion Model and Monte Carlo Compression Sampling for Reliable Medical Image Synthesis. IEEE Journal of Biomedical and Health Informatics

[85] Wang Z, Tao T, Ge Y, Chen Z, Chen T, Ye Z, Lei Y (2026) Weak-mamba-unet: Visual mamba makes cnn and vit work better for scribble-based medical image segmentation. IEEE Transactions on Biomedical Engineering

[86] Kang M, Ting FF, Saw SN, Phan RCW, Ge Z, Ting CM (2026) A Multimodal Feature Distillation with Mamba-Transformer Network for Brain Tumor Segmentation with Incomplete Modalities. IEEE Transactions on Artificial Intelligence

[87] Liang P, Shi L, Pu B, Wu R, Chen J, Zhou L et al (2025) Mambasam: A visual mamba-adapted sam framework for medical image segmentation. IEEE Journal of Biomedical and Health Informatics

[88] Qu Y, Zhou X, Huang P, Liu Y, Mercaldo F, Santone A, Feng P (2025) CGAM: An end-to-end causality graph attention Mamba network for esophageal pathology grading. Biomedical Signal Processing and Control 103:107452

[89] Pei J, Zhang J, Qin G, Wang K, Jin Y, Heng PA (2025) Instrument-tissue-guided surgical action triplet detection via textual-temporal trail exploration. IEEE Transactions on Medical Imaging

[90] Atabansi CC, Nie J, Li H, Liu H, Zhou X (2025) BMSNet: Brain-inspired mixed ultrasound image segmentation network in cancer diagnosis. Journal of King Saud University Computer and Information Sciences 37(3):27

[91] Tu X, Li G, Fan Z, Ding X, Liu Y (2025) Taming diffusion transformers for high-fidelity MRI super-resolution. Computers in Biology and Medicine 198:111261

[92] Hu C, Hang X, Li X, Xing F, Zhou H, Cao N (2026) CBCT-based synthetic CT image generation using a boosting conditional diffusion model. Biomedical Signal Processing and Control 112:108409

[93] Liu Y, Wang Y, Zhang Y, Hu D, Liu Q, Zhang Y et al (2026) FDA-Recon: faithful and data-adaptive reconstruction for limited-angle CBCT. IEEE Transactions on Instrumentation and Measurement 75:1-13

[94] Liu Y, Yang Y, Jiang Y, Zhao X, Fan L (2026) Frequency-aware Vision Mamba With Deformable Windowed Selective Scan for Ultrasound Image Segmentation. Ultrasound in Medicine & Biology

[95] Xue J, Nie M, Qiu Y (2026) MambaPolyp: A Hierarchical Mamba-based Network with Similarity Aggregation for Polyp Segmentation. Journal of Mechanics in Medicine and Biology 26(03):2640021

[96] Ma Y, Zhang C, Xiong D, Zhang H, Ying S (2026) Multi-task dynamic graph learning for brain disorder identification with functional MRI. Pattern Recognition 170:111922

[97] An Z, Ma Y, He Z, Jin C, Song Y, Luo T (2026) TGMamba: texture-guided mamba network for low-dose CT denoising with adaptive scanning. Biomedical Signal Processing and Control 121:110195

[98] Cheng H, Dong X, Shen J, Shi G, Liu X, Zhang J et al (2026) FMaMIL: Synergistic spatial-frequency Mamba multi-instance learning for weakly supervised pathology lesion segmentation. Pattern Recognition 113442

[99] Peng R, Shen L, Lu Z, Diao L, Ge F (2025) CVM-fusion: parallel cross-axes mamba fusion for medical image segmentation. Biomedical Engineering/Biomedizinische Technik

[100] Zou Z, Kui X, Lu W, Li Y, Hu Z, Duan J, Zou B (2026) Robust tumor segmentation in incomplete multi-modal imaging via a synergy of diffusion and Mamba models. Biomedical Signal Processing and Control 112:108685

[101] Zhang L, Wang L, Wu Y, Chen M, Zheng D, Cai Y (2025) ISCDFuse: Interval sampling correlation driven visual state space models for multimodal image fusion. Neurocomputing 640:130329

[102] Zhang X, Lu P, Zheng Z, Li W (2026) HR-UMamba++: A High-Resolution Multi-Directional Mamba Framework for Coronary Artery Segmentation in X-Ray Coronary Angiography. Fractal and Fractional 10(1):43

[103] Xu H, Wu Y, Xie R, Xu J, Wu J, Wang R, Tian Y (2025) Convolutional state space model with multi-window cross-scan to advance the automated diagnosis of skeletal fluorosis. Biomedical Signal Processing and Control 103:107439

[104] Zhang Y, Zhu H, Qian T, Hou T, Chen N, Xu X, Huang B (2026) Adaptive Evidential Fusion of Light–Dark Features With Multi-Scan Mamba for Automated Macular Edema Diagnosis. International Journal of Imaging Systems and Technology 36(1):e70265

[105] Zuo Q, Shi Z, Liu B, Ping N, Wang J, Cheng X et al (2024) Multi-resolution visual Mamba with multi-directional selective mechanism for retinal disease detection. Frontiers in Cell and Developmental Biology 12:1484880

[106] Fu L, Li L, Lu B, Guo X, Shi X, Tian J, Hu Z (2025) Deep equilibrium unfolding learning for noise estimation and removal in optical molecular imaging. Computerized Medical Imaging and Graphics 120:102492

[107] Cheng Z, Guo J, Zhang J, Qi L, Zhou L, Shi Y, Gao Y (2025) Mamba-sea: A mamba-based framework with global-to-local sequence augmentation for generalizable medical image segmentation. IEEE Transactions on Medical Imaging

[108] Ma J, He Y, Li F, Han L, You C, Wang B (2024) Segment anything in medical images. Nature Communications 15(1):654

[109] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S et al (2021) Learning transferable visual models from natural language supervision. In: International conference on machine learning pp 8748-8763. PMLR

[110] Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

[111] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 4700-4708

[112] Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning pp 6105-6114. PMLR

[113] Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S (2022) A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp 11976-11986

[114] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K et al (2018) Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999

[115] Gu A, Goel K, Ré C (2021) Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396

[116] Xing Z, Ye T, Yang Y, Cai D, Gai B, Wu XJ et al (2025) Segmamba-v2: Long-range sequential modeling mamba for general 3d medical image segmentation. IEEE Transactions on Medical Imaging