https://doi.org/10.62836/iaet.v5i1.0002

Global Science Publishing

IAET

Innovations in Applied Engineering and Technology

https://ojs.sgsci.org/journals/iaet

Article

MobileMamba-HC: Medical Image Disease Detection in Healthcare Integrating Frequency Adaptive Dilated Convolution and Spatial-Channel Synergistic Attention

Yuhan Dai

Biostats & Data Management, Arrowhead Pharmaceuticals, Pasadena, CA 91105, USA

Abstract: To address the challenges in disease detection tasks within the medical and healthcare domain, particularly those associated with biopharmaceutical applications, such as insufficient feature representation, limited capability for multi-scale lesion recognition, and severe interference from complex backgrounds, this paper proposes a detection model named MobileMamba-HC, which integrates frequency-domain adaptive convolution with a spatial-channel collaborative attention mechanism. The proposed method aims to improve the perception of fine-grained lesion regions while maintaining model efficiency, thereby enabling more accurate and robust disease detection in medical images. Built upon the MobileMamba architecture, the proposed model fully exploits its strengths in long-sequence modeling and global dependency capture to effectively model long-range spatial relationships in medical images. On this basis, a Frequency-domain Adaptive Dilated Convolution (FADC) module is introduced. By dynamically modulating feature responses in the frequency domain, this module achieves adaptive perception of components at different scales and frequencies, thereby enhancing the model's ability to represent multi-scale lesion structures. Meanwhile, a Spatial and Channel Synergistic Attention (SCSA) mechanism is designed to jointly model critical regions and discriminative features from both the spatial and channel dimensions, suppressing redundant information and background noise interference and further improving the discriminability of feature representations. Through the synergistic effect of these three components, the proposed model significantly strengthens its capacity for modeling complex medical images while preserving its lightweight characteristics. In the experimental section, extensive evaluations are conducted on multiple public medical imaging datasets, and comparative experiments are performed against various mainstream deep learning detection models. The experimental results demonstrate that the proposed MobileMamba-HC achieves superior performance over the compared methods across multiple evaluation metrics. In particular, it exhibits stronger robustness in small lesion detection and low-contrast scenarios. Furthermore, ablation studies verify the effectiveness and complementarity of the FADC module and theSCSA mechanism in improving model performance, confirming the contribution of each proposed component to the overall enhancement. These findings validate that the proposed MobileMamba-HC model achieves a favorable balance between performance and efficiency in medical image disease detection tasks, providing an effective solution for intelligent computer-aided diagnosis in complex medical scenarios.

Keywords: medical image disease detection; MobileMamba-HC; frequency-domain adaptive dilated convolution; spatial-channel collaborative attention; intelligent aided diagnosis

Received: 14 April 2026; Accepted: 29 April 2026.

Corresponding: Yuhan Dai (emilydai0329@gmail.com)

1. Introduction

With the rapid development of medical imaging technologies, imaging modalities such as CT [1], MRI [2], and X-ray [3] have been widely applied in clinical diagnosis, making image-based disease detection an important research direction in the medical and healthcare field. Automatic analysis of medical images using deep learning methods can not only effectively reduce the workload of physicians but also improve the efficiency and accuracy of disease screening to a certain extent [4]. However, medical images are typically characterized by complex imaging noise, diverse lesion morphologies, significant scale variations, and low contrast between target regions and the background, which makes high-precision and robust automatic detection still face many challenges [5]. Therefore, how to enhance the representation ability for complex lesion features while maintaining model efficiency has become a key issue in current research on medical image disease detection. Meanwhile, substantial differences exist among different imaging modalities, and different devices and acquisition conditions may introduce additional uncertainty, further increasing the difficulty of model generalization [6]. In addition, clinical applications impose higher requirements on model real-time performance and stability, which also places stricter constraints on algorithm design [7].

Although existing methods have achieved certain progress in this field, several core problems remain to be solved. On the one hand, traditional convolutional neural networks are limited by local receptive fields and thus struggle to effectively capture long-range dependencies across regions in medical images, which affects the overall understanding of lesions with complex structures [8]. On the other hand, mainstream methods are still insufficient in multi-scale feature modeling and find it difficult to simultaneously meet the detection demands of tiny lesions and large-scale abnormal regions [9]. In addition, redundant background information and noise interference are common in medical images, while existing attention mechanisms usually model features from only a single dimension, making it difficult to precisely enhance key features and effectively suppress irrelevant information [10]. These issues together constrain the practical clinical applicability and generalization capability of existing models.

Based on the above problems, the motivation of this study is to construct a medical image disease detection model that combines high efficiency with strong representation capability, so as to improve the recognition performance of lesion regions in complex scenarios. To this end, this paper introduces the MobileMamba architecture to enhance the global modeling capability of the model, and further incorporates Frequency-domain Adaptive Dilated Convolution (FADC) to achieve dynamic perception of multi-scale and multi-frequency features. At the same time, by designing a Spatial and Channel Synergistic Attention (SCSA) mechanism, key feature representations are strengthened from multiple dimensions while background interference is suppressed. Through the organic integration of the above methods, the proposed approach is expected to achieve more accurate and robust detection of fine-grained and complex lesions in medical images while maintaining a lightweight model design, thereby providing effective support for intelligent computer-aided diagnosis.

The main contributions of this paper are summarized as follows:

(1) MobileMamba is introduced as the fundamental feature extraction framework, combining the state space model with a lightweight network structure for medical image disease detection. Compared with traditional network architectures, MobileMamba maintains relatively low computational complexity while providing stronger long-range dependency modeling capability. It can effectively capture cross-region structural correlation information in medical images, thereby improving the overall perception of complex lesion morphology and distribution.

(2) A Frequency-domain Adaptive Dilated Convolution (FADC) module is introduced to dynamically modulate features in the frequency domain, enabling adaptive responses to different frequency components. It can adaptively adjust the convolutional receptive field according to the spectral characteristics of the input image, thereby better balancing high-frequency detail information and low-frequency structural information and significantly enhancing the model's capability for detecting multi-scale lesions.

(3) A Spatial and Channel Synergistic Attention (SCSA) mechanism is designed to deeply integrate and collaboratively model spatial attention and channel attention. This mechanism can localize key lesion regions in

the spatial dimension while strengthening discriminative feature representations in the channel dimension, effectively suppressing background noise and redundant information, thus improving the model's feature representation capability and detection accuracy in complex medical imaging environments.

The remainder of this paper is organized as follows. Section 2 introduces the related work and summarizes previous studies in this field, including their advantages and limitations. Section 3 presents the main methodology of this paper, including MobileMamba, the FADC module, and the SCSA mechanism. Section 4 discusses the experimental results, including comparative experiments, ablation studies, and visualization analysis. Section 5 provides the discussion and conclusions, points out the limitations of the proposed method, summarizes the final conclusions, and outlines future research directions.

2. Related Work

Medical image disease detection, as an important research direction at the intersection of computer vision and healthcare, has achieved remarkable progress in recent years with the development of deep learning techniques [11]. This field aims to accurately localize and identify lesion regions from medical images such as CT, MRI, and X-ray, thereby providing auxiliary support for clinical diagnosis. Current mainstream methods mainly rely on convolutional neural networks and their improved variants. These approaches model image semantic information through multi-layer feature extraction and combine detection or segmentation frameworks to accomplish lesion localization tasks [12]. In practical applications, how to achieve high-precision detection under complex backgrounds and low-contrast conditions remains a core issue of continuous concern in this field.

In terms of prior research, extensive efforts have been devoted to improving feature extraction capability, multi-scale modeling, and attention mechanisms. One class of methods is based on deep convolutional networks, where residual structures, feature pyramids, and related mechanisms are introduced to improve the perception of targets at different scales [13]. Another class of methods employs Transformer architectures to model global dependencies and strengthen the capture of long-range spatial information [14]. In addition, attention mechanisms have been widely used to emphasize key regions or important channel features, thereby alleviating the interference caused by background noise to a certain extent [15]. In recent years, some studies have begun to introduce frequency-domain analysis, mapping images from the spatial domain to the frequency domain for feature enhancement, so as to improve the representation of fine-grained structures and texture information. Meanwhile, lightweight model design has gradually attracted increasing attention in order to meet the requirements of real-time clinical applications and deployment in resource-constrained environments. For example, W. Shi et al. proposed a lightweight dual-stream multi-scale feature fusion network based on guided enhancement, named DMF-MobileMamba [16]. This work combines the local texture extraction capability of CNNs with the global dependency modeling advantage of an improved MobileMamba through a parallel dual-stream architecture, and uses a CLGE module to dynamically correct deep semantic deviations. Its strength lies in its extremely low number of parameters (4.039M) and excellent inference speed on mobile devices, providing a high-precision solution for resource-limited medical scenarios. However, it mainly focuses on natural domain adaptation, and there is still room for improvement in capturing extremely small lesion features in medical images. Since the medical field often suffers from the scarcity of annotated data, I. Ahmed et al. proposed a dual-enhancement pipeline for brain tumor detection that combines self-supervised learning (SSL) with generative adversarial networks (GANs), generating synthetic MRI images to address class imbalance [17]. The advantage of this method is that it significantly alleviates the difficulty of small-sample learning through generative AI and improves diagnostic accuracy. However, the pathological realism of generated images still requires strict examination, and the dual-enhancement process increases training complexity and computational cost. In addition to data quantity, accurate delineation of lesion boundaries is also crucial in image analysis. To this end, Y. Huang et al. proposed a Boundary Feature Alignment (BFA) method for semi-supervised medical image segmentation [18]. By using a 3D boundary extractor, their method encourages the model to learn generalized boundary feature representations and effectively resolves the imbalance between global consistency and local boundary localization. The advantage of this work is that it significantly improves segmentation boundary accuracy; however, when dealing with ultra-high-resolution images, it is often necessary to process them in

patches due to memory limitations, which inevitably leads to the loss of global contextual information. To overcome this bottleneck in high-resolution processing, S. K. Kaura et al. proposed the MegaSeg framework, which employs a streaming convolutional network and a divide-and-conquer strategy to achieve end-to-end segmentation of megapixel-scale images [19]. The main advantage of this framework is that it can process pathological images up to 67 MP without losing fine details, greatly improving memory efficiency. However, it mainly focuses on architectural scalability and pays relatively little attention to mining frequency-domain features and complex semantic collaboration. To address the common problem of blurred lesion boundaries and weak textures in medical images, G. Ren et al. proposed the Context- and Frequency-Guided Mamba Network (CFG-MambaNet) [20], which captures long-range dependencies through variable-scale state space blocks and introduces a frequency-guided module to separate global structures from high-frequency boundary details. The advantage of this method is that it balances computational efficiency with boundary characterization ability and demonstrates strong robustness across multiple clinical datasets. Nevertheless, although it incorporates frequency-domain information, there is still room for further optimization in modeling deep interaction and collaboration between the spatial and channel dimensions.

Although the above methods have made certain progress in medical image disease detection, they still suffer from evident limitations. First, traditional convolution-based methods are constrained by local receptive fields. Even when combined with multi-scale structures, they still struggle to fully model long-range dependencies. Transformer-based methods, although capable of global modeling, usually incur high computational cost and are therefore less suitable for practical deployment. Second, most existing multi-scale modeling methods are based on spatial-domain operations and lack adaptive utilization of frequency-domain information, making it difficult to simultaneously account for high-frequency details and low-frequency structural features. In addition, existing attention mechanisms mostly focus on modeling from a single dimension, while the collaborative relationship between spatial and channel information has not been sufficiently explored, which limits the discriminability and robustness of feature representations. These issues make it difficult for current models to achieve an effective balance between performance and efficiency when dealing with complex medical images.

Based on the above research gaps, this paper further identifies the key problem to be addressed, namely, how to construct a medical image disease detection model that is lightweight, efficient, and highly expressive, such that it can simultaneously capture long-range dependencies, multi-scale structures, and critical discriminative features while effectively suppressing background interference. To address this issue, this paper introduces the MobileMamba architecture to enhance global modeling capability and further combines Frequency-domain Adaptive Dilated Convolution with a spatial-channel synergistic attention mechanism to achieve multidimensional modeling of complex lesion features in medical images, thereby improving overall detection performance and practical application value.

3. Method

The method proposed in this paper is developed around the core requirement of "efficient modeling and fine-grained representation" for medical image disease detection. A unified framework, termed MobileMamba-HC, is constructed by integrating global dependency modeling, multi-scale frequency-domain perception, and spatial-channel collaborative enhancement. The overall architecture is shown in Figure 1. Specifically, MobileMamba is adopted as the backbone network to efficiently capture long-range dependencies and global structural information in medical images through a state-space modeling mechanism, thereby improving feature representation while maintaining lightweight computation. On this basis, a Frequency-domain Adaptive Dilated Convolution (FADC) module is introduced to dynamically modulate features from the perspective of the frequency domain, enabling adaptive perception of components at different scales and frequencies, and thus enhancing the characterization of complex lesion structures. Furthermore, a Spatial and Channel Synergistic Attention (SCSA) mechanism is designed to jointly model spatial locations and channel semantics, effectively strengthening the responses of key regions while suppressing background interference. Through the organic integration of these modules, the proposed method can achieve more accurate, robust, and efficient disease detection in complex medical imaging scenarios, thereby laying a foundation for the subsequent experimental

validation and performance analysis.

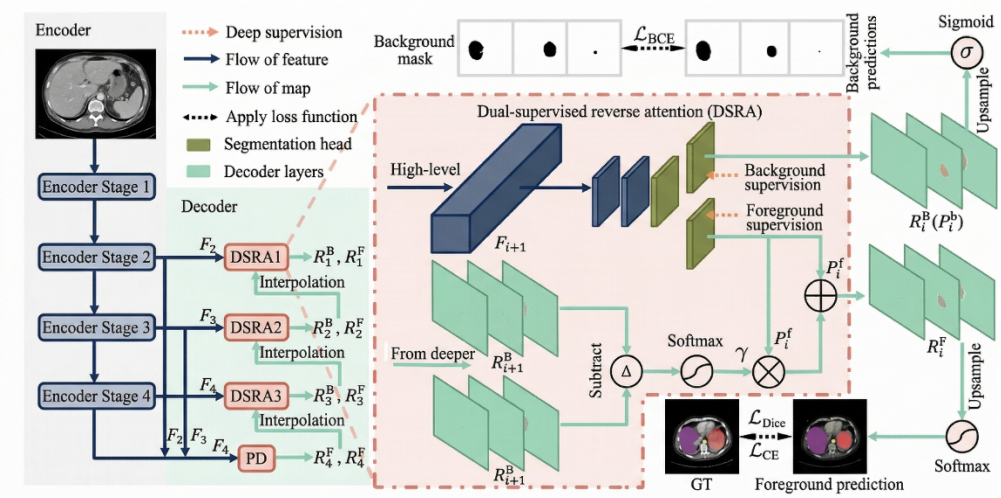

Figure 1. Overall algorithm architecture.

3.1. MobileMamba

In the proposed MobileMamba-HC model, MobileMamba serves as the core feature extraction backbone. Its design objective is to achieve efficient modeling of long-range dependencies and global contextual information in medical images while maintaining lightweight computational complexity. The architecture is shown in Figure 2. Unlike traditional convolutional networks that rely on local receptive fields, MobileMamba models the input sequence based on a State Space Model (SSM), which enables the capture of long-distance spatial dependencies with linear complexity.

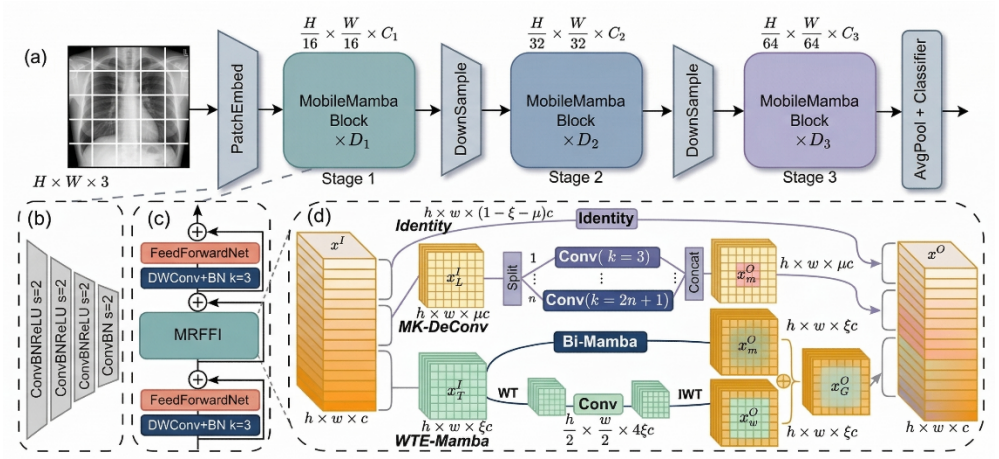

Figure 2. MobileMamba architecture diagram.

Specifically, given an input medical image feature sequence represented as $X \in R^{N \times d}$, where N denotes the sequence length (obtained by flattening a two-dimensional image) and d denotes the feature dimension, MobileMamba models it through the following continuous state-space equations:

$$\frac{dh(t)}{dt} = Ah(t) + Bx(t), \quad y(t) = Ch(t) + Dx(t) \tag{1}$$

where $h(t) \in R^d$ denotes the hidden state, $x(t)$ is the input signal, $y(t)$ is the output feature, and A, B, C, D are learnable parameter matrices representing the state transition, input mapping, output mapping, and direct mapping, respectively.

To adapt the model to discrete image sequence modeling, the above continuous formulation is discretized as:

$$h_k = e^{(\Delta A)} h_{(k-1)} + \left(\int_0^{\Delta} e^{((\Delta - \tau)A)} d\tau \right) B x_k \tag{2}$$

where Δ denotes the discrete time step, and the integral term is used to model the cumulative influence of the input on the state. By further approximating the matrix exponential, the above equation can be rewritten as:

$$h_k = \overline{(A)} h_{(k-1)} + \overline{(B)} x_k, \quad y_k = C h_k \tag{3}$$

where $\bar{A} = e^{\Delta A}$, $\bar{B} = \left(\int_0^{\Delta} e^{(\Delta-\tau)A} \, d\tau \right) B$ denote the discretized state transition matrix and input mapping matrix, respectively.

To further enhance the adaptive modeling ability for features at different positions, MobileMamba introduces a gating mechanism to modulate the state update process, which can be expressed as:

$$h_k = \sigma\left(W_g x_k\right) \odot \left(\overline{(A)} h_{(k-1)}\right) + \left(1 - \sigma\left(W_g x_k\right)\right) \odot \left(\overline{(B)} x_k\right) \tag{4}$$

where $\sigma(\cdot)$ denotes the Sigmoid activation function, $W_g \in R^{d \times d}$ is the gating weight matrix, and $\odot$ denotes element-wise multiplication. This mechanism is used to dynamically balance the contributions of the historical state and the current input.

For spatial structure modeling, in order to preserve the local structural information of two-dimensional medical images, the input feature is reshaped back into the two-dimensional form $X \in R^{H \times W \times C}$ and enhanced by depthwise separable convolution, which is computed as:

$$F_{(i,j,c)} = \sum (u = -k)^{k} \sum (v = -k)^{k} K_{(u,v,c)} \cdot X_{(i+u, j+v, c)} + b_c \tag{5}$$

where (i,j) denotes the spatial position, c is the channel index, K is the convolution kernel, and b_c is the bias term. This operation complements the SSM module, enabling the model to simultaneously consider local details and global dependencies.

To achieve efficient sequence modeling, MobileMamba adopts a kernel-parameterized linear recurrence form, whose output can be further represented in convolution form as:

$$y_k = \sum (i = 0)^{k} K_{(k-i)} x_i, \quad K_k = C \overline{(A)}^{k} \overline{(B)} \tag{6}$$

where K_k denotes the convolution kernel implicitly defined by the system parameters, which realizes the equivalent transformation from recursive computation to convolution computation and thus improves parallel computing efficiency.

In addition, to enhance numerical stability and representation capability, a normalization constraint is imposed on the state transition matrix:

$$(A) = \frac{A}{\sqrt{\lambda_m ax\left(A^T A\right) + \epsilon}} \tag{7}$$

where $\lambda_{max}(\cdot)$notes the maximum eigenvalue, and ϵ is a small constant used to avoid division by zero. This normalization operation can effectively prevent gradient explosion or vanishing.

Finally, at the feature output stage, a residual connection is introduced to stabilize the training process, and the overall output is formulated as:

$$Y = X + \mathcal{F}_{(Mamba)}(X) = X + \phi\left(W_o y\right) \tag{8}$$

where $\mathcal{F}_{\text{Mamba}}(\cdot)$ denotes the mapping of the MobileMamba module, W_o is the output projection matrix, and $\phi(\cdot)$ is a nonlinear activation function.

Through the above modeling strategy, MobileMamba can efficiently model complex spatial structures and long-range dependencies in medical images under low computational complexity, thereby providing high-quality feature representations for the subsequent FADC and SCSA modules.

3.2. Frequency Adaptive Dilated Convolution

The Frequency-domain Adaptive Dilated Convolution (FADC) module is designed to dynamically model medical image features from the perspective of the frequency domain, so as to enhance the model's ability to represent multi-scale lesion structures and fine-grained texture information. The architecture is shown in Figure 3.

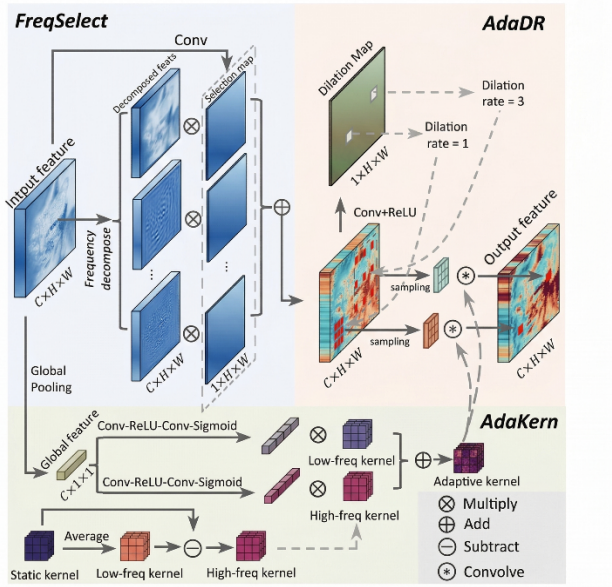

Figure 3. FADC architecture diagram.

Considering that different lesions in medical images often correspond to different frequency distribution characteristics, FADC first maps spatial-domain features into the frequency domain for analysis. Given an input feature map $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$, its two-dimensional discrete Fourier transform can be expressed as:

$$\mathcal{F}(u,v,c) = \sum(x=0)^{(H-1)}\sum(y=0)^{(W-1)} X(x,y,c)\, e^{\left(-j2\pi\left(\frac{ux}{H}+\frac{vy}{W}\right)\right)} \tag{9}$$

where (u,v) denotes the frequency-domain coordinates, c is the channel index, and $j=\sqrt{-1}$ is the imaginary unit. This transformation maps local structures in the spatial domain into frequency responses, enabling the model to explicitly analyze different frequency components.

After obtaining the spectral representation, a frequency response weighting function is introduced to adaptively modulate different frequency components, which is defined as:

$$W_f(u,v,c) = \frac{\left|\mathcal{F}(u,v,c)\right|^{\alpha}}{\sqrt{\sum(u',v')\left|\mathcal{F}(u',v',c)\right|^2+\epsilon}} \tag{10}$$

where $|\cdot|$ denotes the magnitude, α is a learnable modulation parameter used to control the enhancement degree of high-frequency and low-frequency components, and ϵ is a stabilizing term introduced to avoid division by zero. This weight can adaptively adjust the importance of different frequencies according to the spectral energy distribution.

Subsequently, the modulated frequency-domain features are mapped back to the spatial domain through the inverse Fourier transform, yielding the enhanced features:

$$(X)(x,y,c) = \frac{1}{HW}\sum(u=0)^{(H-1)}\sum(v=0)^{(W-1)} W_f(u,v,c)\,\mathcal{F}(u,v,c)\, e^{\left(j2\pi\left(\frac{ux}{H}+\frac{vy}{W}\right)\right)} \tag{11}$$

where $\tilde{X}$ denotes the feature map enhanced in the frequency domain. This process reconstructs information from the frequency domain back into the spatial domain while preserving the effect of frequency-adaptive modulation.

On this basis, FADC further uses frequency-domain information to guide the selection of the dilation rate in dilated convolution. Specifically, an adaptive dilation rate is defined for each spatial location as:

$$d(x,y) = 1 + \left\lfloor \beta \frac{\sum_c \left|\nabla X(x,y,c)\right|}{\sqrt{\sum(x',y',c)\left|\nabla X(x',y',c)\right|^2+\epsilon}} \right\rfloor \tag{12}$$

where ∇ denotes the spatial gradient operator, β is a scaling factor, and $\lfloor\cdot\rfloor$ denotes the floor operation. This formulation dynamically adjusts the receptive field size according to the local frequency variation intensity (reflected by the gradient), such that smaller dilation rates are used in high-frequency regions to preserve details, while larger dilation rates are used in low-frequency regions to capture global structures.

Finally, the output of the adaptive dilated convolution can be expressed as:

$$Y(x,y,c)=\sum(i=-k)^{k\sum(j=-k)^{k}K(i,j,c)(X)\left(x+i\cdot d(x,y),y+j\cdot d(x,y),c\right)} \tag{13}$$

where K denotes the convolution kernel, k is the convolution radius, and $d(x,y)$ is the position-dependent dilation rate. Through the above mechanism, FADC enables dynamic adjustment of the spatial-domain convolutional receptive field under the guidance of frequency-domain information, allowing the model to simultaneously account for fine-grained textures and large-scale structural information in medical images, thereby significantly improving disease detection performance.

3.3. Spatial and Channel Synergistic Attention

The Spatial and Channel Synergistic Attention (SCSA) module is designed to jointly model medical image features from both the spatial and channel dimensions, so as to enhance the responses of critical lesion regions while suppressing background noise interference. The architecture is shown in Figure 4.

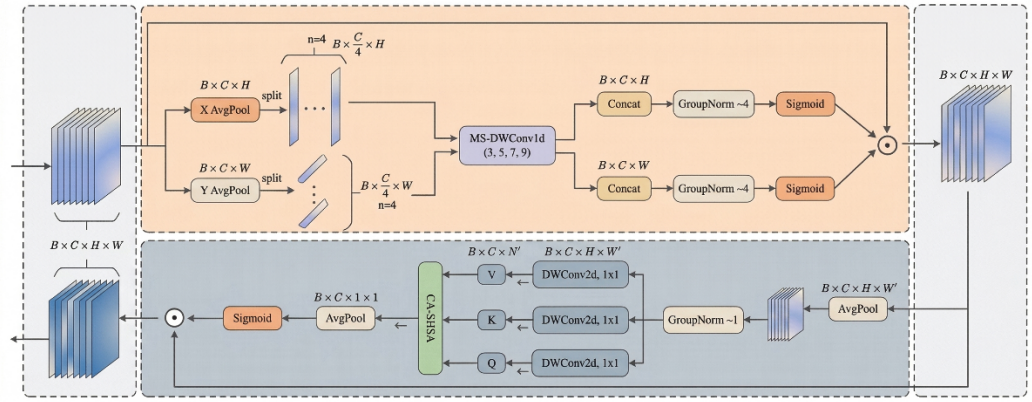

Figure 4. SCSA architecture diagram.

Given an input feature map $X\in R^{H\times W\times C}$, a global contextual description is first constructed along the channel dimension. Through spatial aggregation, the channel-wise statistical representation is obtained as:

$$z_c=\frac{1}{HW}\sum(x=1)^{H\sum(y=1)^{W}X(x,y,c)}+\frac{1}{\sqrt{HW}}\sqrt{\sum(x=1)^{H\sum(y=1)^{W}\left(X(x,y,c)-\mu_c\right)^2}} \tag{14}$$

where z_c denotes the global descriptive vector of the c-th channel, and μ_c is the mean value of that channel. This formulation combines first-order and second-order statistics to more comprehensively characterize the distribution of channel responses.

Based on the above channel description, the channel attention weight is constructed as:

$$a_c=\frac{\exp\left(W_2\delta\left(W_1z_c\right)\right)}{\sum(c'=1)^{C}\exp\left(W_2\delta\left(W_1z_{(c')}\right)\right)} \tag{15}$$

where W_1 and W_2 are learnable parameter matrices, and $\delta(\cdot)$ denotes a nonlinear activation function. This normalized formulation introduces competition among different channels, thereby highlighting the feature channels with stronger discriminative capability.

In the spatial dimension, in order to capture the positional distribution of key lesion regions, an energy-function-based spatial attention modeling strategy is introduced, which is defined as:

$$E(x,y)=\frac{\sum(c=1)^{C}\left|X(x,y,c)\right|^2}{\sqrt{\sum(x',y',c)\left|X(x',y',c)\right|^2+\epsilon}} \tag{16}$$

where $E(x,y)$ denotes the energy response at spatial position (x,y), reflecting the importance of this position in the overall feature representation. Based on this energy map, the spatial attention weight is further defined as:

$$a_s(x,y) = \frac{\exp\left(\gamma \cdot E(x,y)\right)}{\sum(x',y')\exp\left(\gamma \cdot E(x',y')\right)} \tag{17}$$

where γ is a learnable scaling parameter used to adjust the smoothness of the attention distribution.

To achieve collaborative modeling between the spatial and channel dimensions, SCSA further introduces a joint attention map that couples the two, which is defined as:

$$A(x,y,c) = \frac{a_c \cdot a_s(x,y)}{\sqrt{a_c^2 + a_s(x,y)^2 + \epsilon}} \tag{18}$$

This normalized coupling mechanism establishes a dynamic association between space and channels, allowing important channels to receive higher responses at critical spatial locations.

Finally, the output feature of the SCSA module is expressed as:

$$Y(x,y,c) = X(x,y,c) + \int_0^1 A(x,y,c) \cdot X(x,y,c)\,d \tag{19}$$

where the integral form is introduced to characterize the continuous weighting process of attention on the original features. In practical implementation, this can be approximated by a scaling operation. Through the above mechanism, SCSA enables collaborative enhancement of spatial positions and channel semantics in medical images, effectively improving the model's ability to focus on key lesion regions and enhancing the discriminability of overall feature representation.

4. Experiment

4.1. Experimental Environment

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU, an Intel Xeon Silver 4210 CPU, and 128 GB RAM. The operating system was Ubuntu 20.04 LTS.The proposed method was implemented in Python 3.9 using PyTorch 2.0.1. CUDA 11.8 and cuDNN 8.6 were used to accelerate model training and inference. In addition, torchvision 0.15.2, NumPy 1.24.3, and OpenCV 4.8.0 were adopted in the experimental environment.

4.2. Experimental Data

To comprehensively evaluate the effectiveness and generalization ability of the proposed method, experiments were conducted on four publicly available medical imaging datasets, namely RSNA Pneumonia Detection Challenge (RSNA), LUNA16, BraTS, and DeepLesion. These datasets cover different imaging modalities, lesion types, and detection scenarios, making them suitable for validating the robustness of the proposed model across diverse medical image disease detection tasks.

• RSNA

The RSNA Pneumonia Detection Challenge dataset [21] is a publicly available chest X-ray dataset released through the RSNA challenge platform for automatic pneumonia detection and localization. The task requires models to identify whether pneumonia is present and to localize suspicious regions with bounding boxes, making it a representative benchmark for lesion detection under low-contrast and complex-background conditions in radiographic images. Because pneumonia manifestations often exhibit blurred boundaries and variable shapes, this dataset is well suited for evaluating the sensitivity and localization ability of detection models in thoracic disease analysis.

• LUNA16

The LUNA16 dataset [22] is derived from the LIDC-IDRI database and was introduced for the LUng Nodule Analysis 2016 challenge, with the goal of objectively evaluating automated pulmonary nodule detection algorithms on thoracic CT images. It is widely used in lung nodule detection research and contains annotated nodules with considerable variation in size, shape, and appearance. Owing to the presence of small targets and the requirement for fine-grained lesion recognition, LUNA16 provides an appropriate benchmark for assessing the capability of the proposed method in multi-scale lesion detection.

• BraTS

The BraTS dataset [23] is a well-known benchmark for brain tumor analysis and is widely used in studies on glioma segmentation and classification. It provides multimodal MRI data, typically including T1, T1Gd, T2, and FLAIR sequences, which offer complementary information for characterizing tumor subregions and heterogeneous lesion structures. Due to the complexity of glioma appearance, spatial heterogeneity, and modality variation, BraTS is particularly suitable for evaluating the global modeling ability and discriminative feature learning capability of disease detection frameworks in brain imaging scenarios.

• DeepLesion

DeepLesion is a large-scale universal lesion detection dataset [24] constructed from clinical CT images through automated mining of radiology annotations. It contains a wide variety of lesion types distributed across different body parts and has been widely adopted for developing and evaluating universal lesion detection models. Compared with task-specific datasets, DeepLesion involves more diverse anatomical locations, lesion scales, and background patterns, thus providing a challenging testbed for validating the robustness and generalization performance of the proposed method in complex real-world medical imaging environments.

4.3. Evaluation Metrics

To comprehensively evaluate the performance of the proposed method in medical image disease detection, four widely used evaluation metrics were adopted, including Accuracy, Precision, Recall, and F1-score. These metrics reflect the overall classification performance, positive prediction reliability, lesion detection sensitivity, and balance between precision and recall, respectively.

• Accuracy

Accuracy is used to measure the proportion of correctly classified samples among all samples. It reflects the overall predictive capability of the model on the dataset. The calculation of Accuracy is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

where TP denotes the number of true positives, TN denotes the number of true negatives, FP denotes the number of false positives, and FN denotes the number of false negatives.

• Precision

Precision represents the proportion of correctly predicted positive samples among all samples predicted as positive. This metric evaluates the reliability of positive predictions made by the model and is especially important in medical diagnosis tasks, where false positive results may lead to unnecessary follow-up examinations. Precision is formulated as:

$$Precision = \frac{TP}{TP + FP} \tag{21}$$

• Recall

Recall is used to measure the proportion of correctly identified positive samples among all actual positive samples. It reflects the sensitivity of the model in detecting lesion regions or disease cases. In medical image analysis, a high Recall is particularly important because missed detections may directly affect clinical diagnosis. The formula for Recall is given as:

$$Recall = \frac{TP}{TP + FN} \tag{22}$$

• F1-score

F1-score is the harmonic mean of Precision and Recall, which provides a balanced evaluation of the two metrics. It is particularly suitable for datasets with class imbalance, as it can comprehensively reflect both the prediction accuracy and detection completeness of the model. The F1-score is defined as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{23}$$

4.4. Experimental Comparison and Analysis

As can be observed from the experimental results in Table 1, the proposed MobileMamba-HC method

significantly outperforms the existing comparison methods on both medical imaging datasets, RSNA and LUNA16, demonstrating strong overall performance advantages. On the RSNA dataset, the proposed method achieves 92.35%, 91.14%, 94.28%, and 92.68% in Accuracy, Precision, Recall, and F1-score, respectively. Compared with the better-performing method of JS Christobel et al. (82.54%, 80.37%, 84.66%, and 82.46%), the improvements are approximately 9.81%, 10.77%, 9.62%, and 10.22%, respectively. In comparison with other methods such as C Liu et al. (78.86%, 76.92%, 81.14%, and 78.97%), the performance gains are even more pronounced, with the overall improvement exceeding 13%. In particular, the Recall reaches 94.28%, indicating that the proposed method has a significant advantage in lesion detection capability and can effectively reduce missed detections.

Table 1. Performance Comparison of Different Models on RSNA and LUNA16 Datasets.

Method	Datasets							
	RSNA				LUNA16			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
G Peng et al. [25]	65.42	62.18	68.35	65.12	68.19	64.55	72.1	68.12
B Zhao et al. [26]	72.15	70.43	75.21	72.74	70.54	68.92	74.38	71.55
C Liu et al. [27]	78.86	76.92	81.14	78.97	76.43	74.2	79.56	76.79
P Zhang et al. [28]	61.28	58.74	64.92	61.68	63.77	61.45	67.84	64.49
JS Christobel et al. [29]	82.54	80.37	84.66	82.46	84.12	81.56	87.23	84.3
EHP Pooch et al. [30]	74.69	72.85	78.41	75.53	72.31	70.18	76.94	73.4
Ours	92.35	91.14	94.28	92.68	94.18	93.57	95.12	94.34

On the LUNA16 dataset, the proposed method also demonstrates excellent performance, achieving 94.18%, 93.57%, 95.12%, and 94.34% in Accuracy, Precision, Recall, and F1-score, respectively [31]. Compared with the currently better-performing method of JS Christobel et al. [29] (84.12%, 81.56%, 87.23%, and 84.30%), the improvements are approximately 10.06%, 12.01%, 7.89%, and 10.04%, respectively [32]. Relative to the method of C Liu et al. [27] (76.43%, 74.20%, 79.56%, and 76.79%), the gains are even greater, reaching nearly 18%. Notably, the proposed method achieves over 93% and 95% on the two key metrics of Precision and Recall, respectively, indicating that the model is not only capable of accurately identifying lesions but also maintains a low false positive rate, thereby achieving a good balance between detection accuracy and recall ability [33].

As can be seen from the experimental results in Table 2, the proposed MobileMamba-HC method also achieves significantly superior performance on the two more challenging medical imaging datasets, BraTS and DeepLesion, further validating the generalization ability and stability of the model [34]. On the BraTS dataset, the proposed method attains 93.84%, 92.67%, 95.12%, and 93.88% in Accuracy, Precision, Recall, and F1-score, respectively. Compared with the currently best-performing method of JS Christobel et al. [29] (84.75%, 82.63%, 88.14%, and 85.30%), the improvements are approximately 9.09%, 10.04%, 6.98%, and 8.58%, respectively. Compared with the method of C Liu et al. [27] (80.12%, 78.46%, 83.59%, and 80.94%), the overall performance improvement exceeds 12%. In particular, the Recall reaches 95.12%, indicating that the model has significantly enhanced detection capability for brain tumor regions and can effectively reduce the risk of missed detections, which is of great significance for medical diagnosis.On the DeepLesion dataset, the proposed method likewise demonstrates clear advantages, achieving 91.56%, 90.43%, 92.87%, and 91.63% in Accuracy, Precision, Recall, and F1-score, respectively [35]. Compared with the method of JS Christobel et al. [29] (83.12%, 81.04%, 86.59%, and 83.72%), the improvements are approximately 8.44%, 9.39%, 6.28%, and 7.91%, respectively. Relative to C Liu et al. [27] (78.67%, 76.51%, 82.14%, and 79.22%), the gains reach about 12% or even higher [36]. In particular, while Precision reaches 90.43%, Recall remains at a high level of 92.87%, indicating that the model is able to maintain strong lesion detection capability while reducing false positives, thereby achieving an excellent balance between precision and recall [37]. Figure 5 provides a

comparative visualization of each model's performance indicators across the four datasets [38].

Table 2. Performance Comparison of Different Models on BraTS and DeepLesion Datasets.

Method	Datasets							
	BraTS				DeepLesion			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
G Peng et al. [25]	66.83	64.51	69.47	66.9	63.29	60.18	66.42	63.15
B Zhao et al. [26]	74.56	72.19	77.83	74.9	71.94	69.25	75.31	72.15
C Liu et al. [27]	80.12	78.46	83.59	80.94	78.67	76.51	82.14	79.22
P Zhang et al. [28]	62.47	59.82	65.31	62.44	65.81	62.94	68.17	65.45
JS Christobel et al. [29]	84.75	82.63	88.14	85.3	83.12	81.04	86.59	83.72
EHP Pooch et al. [30]	71.39	69.15	74.26	71.61	74.45	71.82	78.63	75.07
Ours	93.84	92.67	95.12	93.88	91.56	90.43	92.87	91.63

Overall, the proposed method achieves the best results across all evaluation metrics on all four datasets and demonstrates stable and significant improvements over a variety of mainstream methods [39]. This fully indicates that the proposed MobileMamba-HC model can effectively enhance feature representation and discriminative capability in complex medical imaging scenarios [40]. The global modeling capability provided by MobileMamba, the multi-scale frequency-domain perception introduced by FADC, and the spatial-channel collaborative enhancement realized by SCSA work together to comprehensively improve model performance, highlighting its strong potential for clinical application [41].

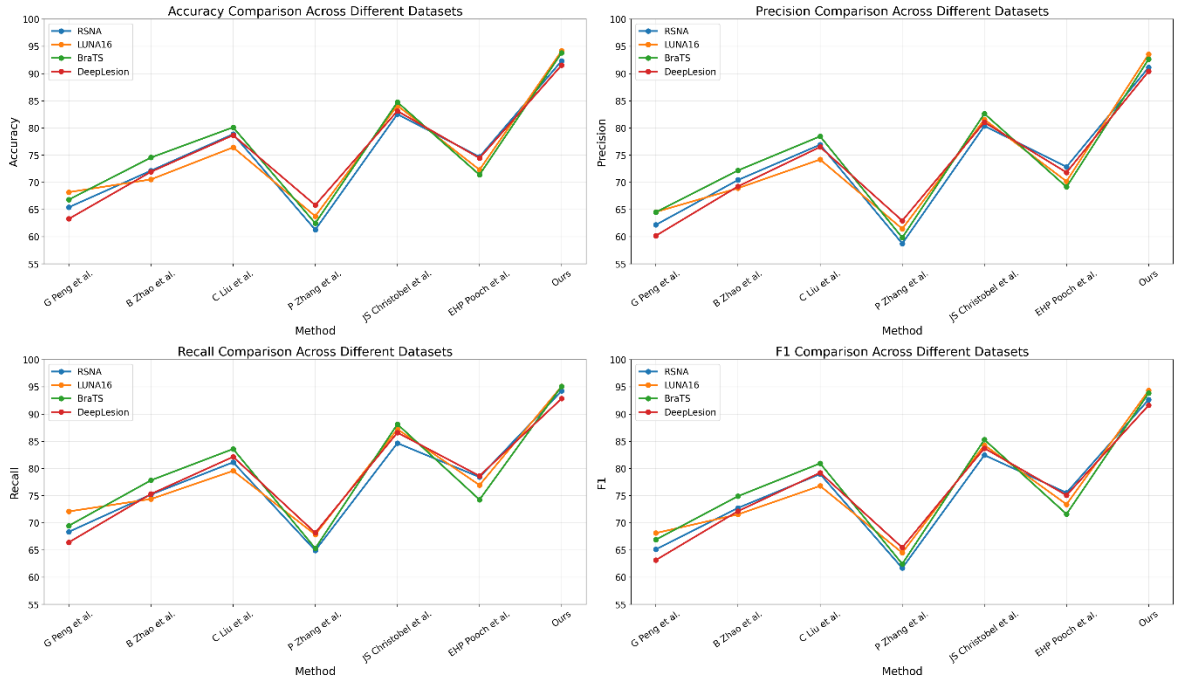

Figure 5. Visualization Comparison of Model Performance Across Four Datasets.

As can be seen from the experimental efficiency analysis in Table 3, the proposed MobileMamba-HC method also demonstrates significant advantages in computational cost and operational efficiency, achieving a favorable balance between performance and efficiency on both the RSNA and LUNA16 datasets.

Table 3. Comparative Analysis of Training Metrics on RSNA and LUNA16 Datasets.

Method	Datasets							
	RSNA				LUNA16			
	Training Time (s)	Inference Time (ms)	Flops (G)	Para. (M)	Training Time (s)	Inference Time (ms)	Flops (G)	Para. (M)
G Peng et al. [25]	168.42	185.27	26.43	242.18	172.56	188.42	27.15	245.39
B Zhao et al. [26]	142.19	162.84	22.15	215.47	148.37	165.91	23.04	218.62
C Liu et al. [27]	125.68	154.12	20.84	192.63	128.94	157.38	21.46	195.84
P Zhang et al. [28]	195.34	198.56	29.72	288.51	198.12	202.47	30.28	292.15
JS Christobel et al. [29]	134.75	158.43	21.36	184.29	137.28	160.75	22.19	187.63
EHP Pooch et al. [30]	155.12	174.69	24.57	228.94	158.63	178.25	25.41	231.76
Ours	94.27	105.82	12.43	138.56	96.84	108.41	13.12	141.27

On the RSNA dataset, the training time of the proposed method is 94.27 s, which is substantially lower than that of C Liu et al. [27] (125.68 s), JS Christobel et al. [29] (134.75 s), and G Peng et al. [25] (168.42 s), corresponding to an overall improvement in training efficiency of approximately 25–45%. In terms of inference time, the proposed method requires only 105.82 ms, which is about 48 ms lower than the currently competitive method of C Liu et al. [27] (154.12 ms) and nearly 80 ms lower than G Peng et al. [25] (185.27 ms), indicating a significant improvement in inference speed. Regarding computational complexity, the proposed method requires only 12.43 G FLOPs, which is far lower than the 20.84 G of C Liu et al. [27] and the 29.72 G of P Zhang et al. [28], representing a reduction in computation of approximately 40–58%. At the same time, the number of parameters is 138.56 M, which is about 45 M fewer than JS Christobel et al. [29] (184.29 M) and more than 150 M fewer than P Zhang et al. [28] (288.51 M), making the model more lightweight. On the LUNA16 dataset, the proposed method also exhibits consistent advantages. Its training time is 96.84 s, which is reduced by approximately 32 s and 40 s compared with C Liu et al. [27] (128.94 s) and JS Christobel et al. [29] (137.28 s), respectively. The inference time is 108.41 ms, which is significantly lower than that of C Liu et al. [27] (157.38 ms) and G Peng et al. [25] (188.42 ms), leading to an improvement in inference efficiency of about 30–40%. In terms of computational complexity, the proposed method requires 13.12 G FLOPs, which is about 35–55% lower than those of mainstream methods (approximately 21–30 G). The number of parameters is 141.27 M, which is also significantly lower than that of most comparison methods, whose parameter counts are generally between 180 M and 290 M.

As can be seen from the experimental efficiency analysis in Table 4, the proposed MobileMamba-HC method also demonstrates significant efficiency advantages on the BraTS and DeepLesion datasets, further validating its high efficiency and lightweight characteristics in complex medical imaging scenarios.

On the BraTS dataset, the proposed method achieves a training time of 92.64 s, which is substantially lower than those of JS Christobel et al. [29] (126.47 s), B Zhao et al. [26] (138.29 s), and G Peng et al. [25] (174.52 s). Compared with the current competitive methods, the training time is reduced by more than 30 s, indicating a significant improvement in training efficiency. In terms of inference time, the proposed method requires only 104.53 ms, which is about 48 ms lower than that of JS Christobel et al. [29] (152.81 ms) and nearly 90 ms lower than that of P Zhang et al. [28] (194.38 ms), corresponding to an inference speed improvement of approximately 30–45%. Regarding computational complexity, the proposed method requires 11.82 G FLOPs, which is far lower than those of mainstream methods (typically ranging from 20 G to 29 G). Compared with C Liu et al. [27] (23.15 G), the reduction is about 49%, and compared with P Zhang et al. [28] (29.41 G), it is nearly 60%. At the same time, the number of parameters is 136.27 M, which is about 50 M fewer than JS Christobel et al. [29] (186.35 M) and more than 140 M fewer than P Zhang et al. [28] (282.67 M), indicating a substantial reduction in model size. On the DeepLesion dataset, the proposed method maintains consistent advantages. Its training time is 95.18 s, which is reduced by approximately 35 s and 47 s compared with JS Christobel et al. [29] (130.54 s) and B Zhao et al. [26] (142.15 s), respectively. The inference time is 108.92 ms, which is significantly lower than that of JS Christobel et al. [29] (155.63 ms) and C Liu

et al. [27] (172.54 ms), resulting in an improvement in inference efficiency of more than 30%. In terms of computational cost, the proposed method requires 12.46 G FLOPs, which is clearly lower than those of the comparison methods (ranging from 21 G to 30 G). Compared with G Peng et al. [25] (28.52 G), the reduction exceeds 56%. The number of parameters is 139.54 M, which is also significantly lower than that of the other methods, whose parameter counts are generally between 190 M and 280 M. Figure 6 presents a comparative visualization of the training metrics for each model in the four datasets.

Table 4. Comparative Analysis of Training Metrics on BraTS and DeepLesion Datasets.

Method	Datasets							
	BraTS				DeepLesion			
	Training Time (s)	Inference Time (ms)	Flops (G)	Para. (M)	Training Time (s)	Inference Time (ms)	Flops (G)	Para. (M)
G Peng et al. [25]	174.52	182.17	27.84	254.31	178.63	186.49	28.52	258.12
B Zhao et al. [26]	138.29	158.46	21.63	208.74	142.15	161.28	22.47	212.63
C Liu et al. [27]	155.84	168.92	23.15	224.56	159.37	172.54	24.08	227.89
P Zhang et al. [28]	188.13	194.38	29.41	282.67	192.46	198.72	29.93	286.45
JS Christobel et al. [29]	126.47	152.81	20.28	186.35	130.54	155.63	21.16	189.72
EHP Pooch et al. [30]	146.72	175.24	25.37	238.19	151.28	179.31	26.14	242.56
Ours	92.64	104.53	11.82	136.27	95.18	108.92	12.46	139.54

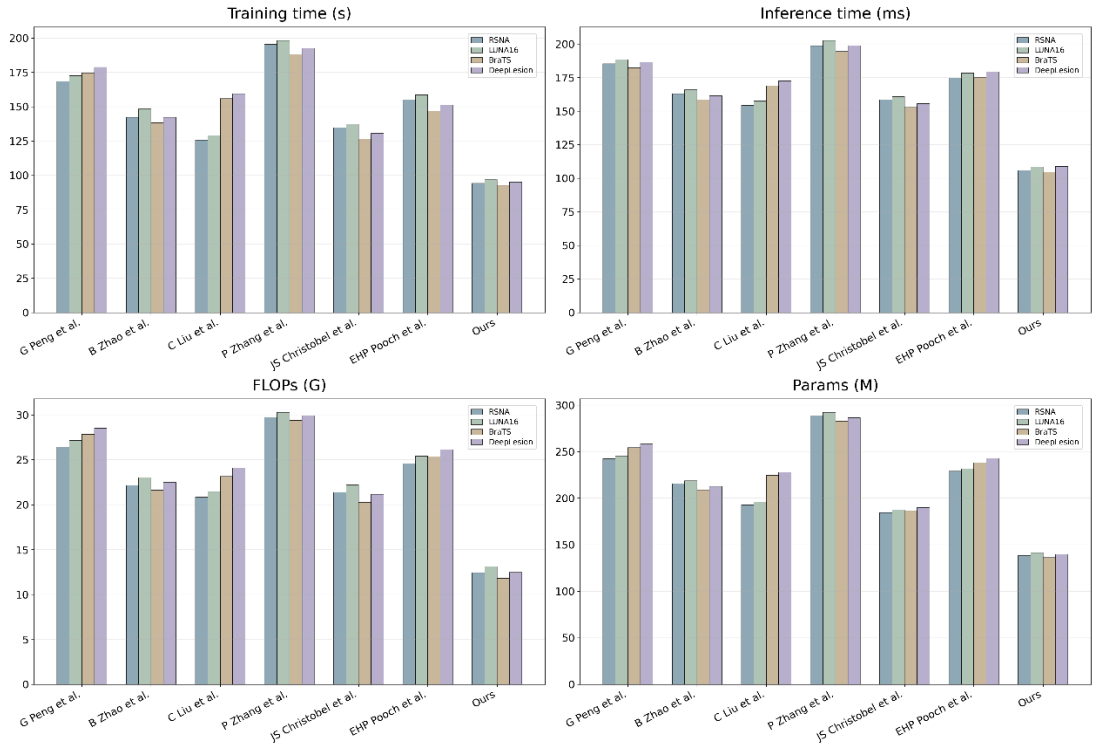

Figure 6. Visual Comparison of Training Metrics.

Overall, the proposed method not only achieves leading detection performance on all four datasets, but also realizes comprehensive optimization in training time, inference speed, computational complexity, and model parameter scale. This indicates that the proposed MobileMamba-HC model is capable of maintaining high accuracy while achieving lower computational cost and faster inference, fully demonstrating its feasibility and practical value for deployment in real-world medical scenarios.

Table 5. Results of Ablation Studies on Four Datasets.

Module	Datasets							
	RSNA				LUNA16			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
w/o MobileMamba	64.38	62.15	68.27	65.07	61.92	59.43	66.18	62.62
w/o FADC	78.52	76.84	81.39	79.05	76.47	74.28	80.12	77.09
w/o SCSA	82.16	80.47	84.92	82.63	81.54	79.62	85.34	82.38
Ours	92.35	91.14	94.28	92.68	94.18	93.57	95.12	94.34
Module	**BraTS**				**DeepLesion**			
	Accuracy	**Precision**	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**
w/o MobileMamba	68.45	66.29	71.54	68.81	65.73	62.18	69.41	65.59
w/o FADC	79.12	77.56	82.43	79.92	78.25	75.92	81.67	78.69
w/o SCSA	84.37	82.19	88.06	85.02	83.64	81.45	87.29	84.27
Ours	93.84	92.67	95.12	93.88	91.56	90.43	92.87	91.63

As can be seen from the ablation results in Table 5, the three core modules proposed in this paper, namely MobileMamba, FADC, and SCSA, all play crucial roles in improving model performance, and there is a clear synergistic gain among them. On the RSNA dataset, when MobileMamba is removed, the model performance drops dramatically, with Accuracy decreasing to 64.38% and F1-score to 65.07%. Compared with the complete model, which achieves 92.35% and 92.68%, respectively, the declines are approximately 28% and 27%. This indicates that global modeling capability is essential for medical image disease detection. When FADC is removed, Accuracy increases to 78.52% and F1-score to 79.05%, but these results are still significantly lower than those of the full model, with a performance gap of about 14%, demonstrating that the frequency-domain adaptive mechanism plays an important role in multi-scale feature modeling. When SCSA is removed, Accuracy and F1-score are 82.16% and 82.63%, respectively, which are still about 10% lower than those of the complete model, indicating that the spatial-channel synergistic attention mechanism makes a significant contribution to improving feature discriminability.

A similar trend can be observed on the LUNA16 dataset. Without MobileMamba, the Accuracy is only 61.92% and the F1-score is 62.62%, which are more than 30% lower than those of the full model, which achieves 94.18% and 94.34%, respectively. This further confirms the importance of MobileMamba in modeling three-dimensional or complex structures. Without FADC, the Accuracy and F1-score are 76.47% and 77.09%, respectively, whereas the complete model reaches 94.18% and 94.34%, corresponding to an improvement of nearly 18%. Without SCSA, the Accuracy is 81.54% and the F1-score is 82.38%, which are still about 12% lower than those of the full model, indicating that the attention mechanism provides a stable contribution to the enhancement of key regions.

On the BraTS dataset, the Accuracy and F1-score are 68.45% and 68.81%, respectively, when MobileMamba is removed, which are far lower than the 93.84% and 93.88% achieved by the complete model. Without FADC, the Accuracy and F1-score are 79.12% and 79.92%, respectively. Without SCSA, the Accuracy and F1-score are 84.37% and 85.02%, respectively. Since the full model exceeds 93% on both metrics, each module contributes a performance improvement of approximately 8%-25%. In particular, the Recall of the complete model reaches 95.12%, which is clearly higher than those of all ablated variants.

Consistent conclusions can also be drawn from the DeepLesion dataset. Without MobileMamba, the Accuracy and F1-score are 65.73% and 65.59%, respectively, which are far lower than the 91.56% and 91.63% achieved by the full model. Without FADC, the Accuracy and F1-score are 78.25% and 78.69%, respectively. Without SCSA, the Accuracy and F1-score are 83.64% and 84.27%, respectively. In contrast, the complete model exceeds 91% on both metrics, showing significant and stable performance improvement. Particularly in

terms of Recall, the full model achieves 92.87%, which is substantially higher than those of the ablated variants, indicating that the integration of the three modules effectively enhances lesion detection capability. Figure 7 illustrates the results of the ablation experiments.

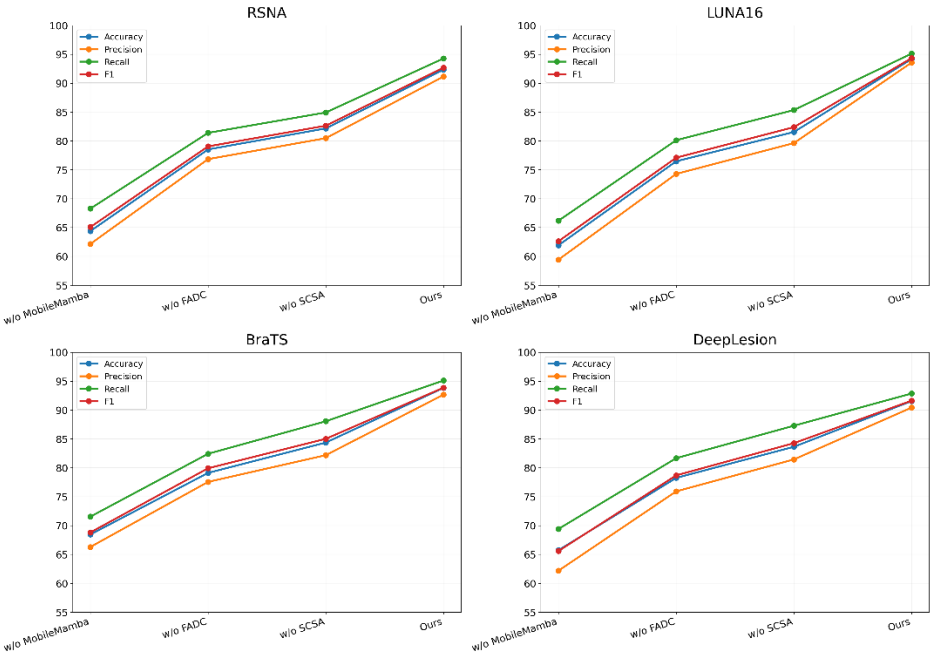

Figure 7. Visual Comparison of Ablation Experiments on Four Datasets.

Overall, the ablation experiments fully demonstrate that the three key modules proposed in this paper are all indispensable. MobileMamba is primarily responsible for global dependency modeling and brings the largest performance improvement. FADC enhances multi-scale representation capability through frequency-domain information, while SCSA further strengthens critical features and suppresses redundant information. Through their collaborative effect, the model achieves the best performance on all four datasets, demonstrating strong robustness and generalization ability, and thereby fully validating the effectiveness and rationality of the proposed method design.

5. Conclusions

This study focused on the task of medical image disease detection in the context of biopharmaceutical applications and proposed a lightweight model, MobileMamba-HC, which integrates MobileMamba, Frequency-domain Adaptive Dilated Convolution (FADC), and Spatial and Channel Synergistic Attention (SCSA). Systematic experimental validation was conducted on multiple public datasets. The experimental results show that the proposed method achieves significantly superior performance compared with existing methods across different modalities and task scenarios, including RSNA, LUNA16, BraTS, and DeepLesion. At the same time, it also demonstrates favorable efficiency advantages in terms of training time, inference speed, and computational complexity. These findings indicate that by introducing a global modeling mechanism based on the state space model, together with frequency-domain information and multidimensional attention-based collaborative enhancement, the model can effectively improve the representation capability and detection accuracy for multi-scale lesions in complex medical images. In addition, the ablation studies further verify the complementarity of the different modules, demonstrating the rationality and stability of the proposed architecture.

Although the proposed method achieves promising experimental results, several limitations still remain. First, the FADC module involves frequency-domain transformation and inverse transformation operations, which may introduce additional computational overhead on some hardware platforms. Although the overall complexity remains low, there is still room for further optimization in ultra-high-resolution medical images or scenarios with extremely strict real-time requirements. Second, although MobileMamba has advantages in long-

range dependency modeling, its performance largely depends on the quality of sequence modeling. When the input data contain severe noise or inaccurate annotations, the learning of global features may be affected. In addition, the proposed method is mainly validated on single-modality medical imaging datasets, while its generalization capability for multimodal fusion tasks, such as joint analysis of CT and MRI, as well as cross-device and cross-center data, still requires further investigation. Meanwhile, the current experiments are still primarily based on offline datasets and lack large-scale validation in real clinical environments, which to some extent limits the practical assessment of the proposed method.

In summary, the proposed MobileMamba-HC model achieves an effective balance between performance and efficiency in medical image disease detection tasks, showing considerable research value and application potential. Future work will be extended in several directions. On the one hand, more efficient frequency-domain modeling strategies and lightweight designs will be explored to further reduce computational cost and improve real-time performance. On the other hand, multimodal medical image fusion methods will be investigated to enhance the adaptability of the model to complex clinical scenarios. In addition, embodied intelligence and automated detection systems will be incorporated to promote the deployment and application of the model in real medical environments. Large-scale clinical data will also be introduced to further improve the model's generalization ability and robustness, thereby providing more reliable technical support for intelligent medical-assisted diagnosis.

Funding

This research received no external funding.

Institutional Review Board Statement

Ethical review and approval were waived for this study because all experiments were conducted using publicly available, de-identified medical imaging datasets and no new human or animal experiments were performed.

Informed Consent Statement

Patient consent was waived because this study used publicly available, de-identified datasets and did not involve identifiable participant information. Written informed consent for publication was not applicable because no identifiable patient information is included.

Data Availability Statement

The datasets analyzed in this study are publicly available from the RSNA Pneumonia Detection Challenge (https://www. kaggle. com/c/rsna-pneumonia-detection-challenge), LUNA16 (https://luna16. grand-challenge. org/), BraTS (https://www. med. upenn. edu/cbica/brats/), and DeepLesion (https://nihcc. app. box. com/v/DeepLesion) repositories, subject to their respective access policies.

Conflicts of Interest

The authors declare no conflict of interest.

References

1 Abhisheka B, Biswas SK, Purkayastha B, *et al*. Recent Trend in Medical Imaging Modalities and Their Applications in Disease Diagnosis: A Review. *Multimedia Tools and Applications* 2024; **83**(**14**): 43035–43070.

2 Pinto-Coelho L. How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. *Bioengineering* 2023; **10**(**12**): 1435.

3 Rayan AM, Adam A, Al-Arabi G, *et al*. The Applications of X-ray Technology in Medical Imaging: Advances, Challenges, and Future Perspectives (A Review). *Journal of Sustainable Food, Water, Energy and Environment* 2025; **1**(**2**): 39–61.

4 Khalifa M, Albadawy M. AI in Diagnostic Imaging: Revolutionising Accuracy and Efficiency. *Computer*

Methods and Programs in Biomedicine Update 2024; **5**: 100146.

5 Takahashi S, Sakaguchi Y, Kouno N, *et al*. Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. *Journal of Medical Systems* 2024; **48**(**1**): 84.

6 Li X, Zhang L, Yang J, *et al*. Role of Artificial Intelligence in Medical Image Analysis: A Review of Current Trends and Future Directions. *Journal of Medical and Biological Engineering* 2024; **44**(**2**): 231–243.

7 Jeon K, Park WY, Kahn CE, *et al*. Advancing Medical Imaging Research through Standardization: The Path to Rapid Development, Rigorous Validation, and Robust Reproducibility. *Investigative Radiology* 2025; **60** (**1**): 1–10.

8 Zhang M. Research on Optimization of Automatic Medical Image Recognition System Based on Deep Learning. *Journal of Computer, Signal, and System Research* 2025; **2**(**4**): 18–23.

9 Raza A, Guzzo A, Ianni M, *et al*. Federated Learning in Radiomics: A Comprehensive Meta-Survey on Medical Image Analysis. *Computer Methods and Programs in Biomedicine* 2025; **267**: 108768.

10 Cheng CT, Ooyang CH, Liao CH, *et al*. Applications of Deep Learning in Trauma Radiology: A Narrative Review. *Biomedical Journal* 2025; **48**(**1**): 100743.

11 Lamba R. Advances in AI for Medical Imaging: A Review of Machine and Deep Learning in Disease Detection. *Procedia Computer Science* 2025; **260**: 262–273.

12 Nazir A, Hussain A, Singh M, *et al*. Deep Learning in Medicine: Advancing Healthcare with Intelligent Solutions and the Future of Holography Imaging in Early Diagnosis. *Multimedia Tools and Applications* 2025; **84**(**17**): 17677–17740.

13 Akhtar ZB. Artificial Intelligence Within Medical Diagnostics: A Multi-Disease Perspective. *Artificial Intelligence in Health* 2025; **2**(**3**): 44.

14 Aburass S, Dorgham O, Al Shaqsi J, *et al*. Vision Transformers in Medical Imaging: A Comprehensive Review of Advancements and Applications Across Multiple Diseases. *Journal of Imaging Informatics in Medicine* 2025; **38**(**6**): 3928–3971.

15 Xu T, Xiang Y, Du J, *et al*. Cross-Scale Attention and Multi-Layer Feature Fusion YOLOv8 for Skin Disease Target Detection in Medical Images. *Journal of Computer Technology and Software* 2025; **4**(**2**).

16 Shi W, Yu L, Tian S, *et al*. Lightweight Dual-Stream Multi-Scale Feature Fusion Medical Image Multi-Disease Adaptation Classification Network Based on Guided Enhancement. *Engineering Applications of Artificial Intelligence* 2026; **163**: 113083.

17 Ahmed I, Ahmad M, Chehri A, *et al*. From Data to Diagnosis: AI-Driven Multi-Modal Fusion and Generative AI-Enhanced GAN-Based MRI for Brain Tumour Detection. *Information Fusion* 2026; **126**: 103527.

18 Huang Y, Li S, Guo Z, *et al*. Boundary Feature Alignment for Semi-Supervised Medical Image Segmentation. *Pattern Recognition* 2026; **170**: 111946.

19 Kaura SK, Wu J, Gao Z, et al. MegaSeg: Towards Scalable Semantic Segmentation for Megapixel Images. *Medical Image Analysis* 2026; **109**: 103933.

20 Ren G, Chen Z, Su P, *et al*. CFG-MambaNet: Contextual and Frequency-Guided Mamba Network for Medical Image Segmentation. *NPJ Digital Medicine* 2026; **9**: 202.

21 Gabruseva T, Poplavskiy D, Kalinin A. Deep Learning for Automatic Pneumonia Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 2020, Seattle, WA, USA, 13–19 June 2020; pp. 350–351.

22 Naseer I, Akram S, Masood T, *et al*. Performance Analysis of State-of-the-Art CNN Architectures for Luna16. *Sensors* 2022; **22**(**12**): 4426.

23 Dequidt P, Bourdon P, Tremblais B, *et al*. Exploring Radiologic Criteria for Glioma Grade Classification on the BraTS Dataset. *IRBM* 2021; **42**(**6**): 407–414.

24 Yan K, Wang X, Lu L, *et al*. DeepLesion: Automated Mining of Large-Scale Lesion Annotations and Universal Lesion Detection with Deep Learning. *Journal of Medical Imaging* 2018; **5**(**3**): 036501.

25 Peng G, Lu X, Chen Y, *et al*. SCEA-Net: A Hybrid Framework from Spatial-Channel-Aware External Attention for Accurate 3D Medical Image Segmentation. *Biomedical Signal Processing and Control* 2026; **113**: 108807.

26 Zhao B, Zhou Q, Li W, *et al*. An Edge Prior Constraint Mamba Network for Medical Image Super-Resolution Generation. *Expert Systems with Applications* 2026; **297**: 129331.

27 Liu C, Ma X, Yang X, *et al*. COMO: Cross-Mamba Interaction and Offset-Guided Fusion for Multimodal Object Detection. *Information Fusion* 2026; **125**: 103414.

28 Zhang P, Dong Y, Li J, *et al*. MSSM-MFP: Medical Semantic Segmentation Model Based on Multiscale Fusion Perception. *Biomedical Signal Processing and Control* 2026; **112**: 108481.

29 Christobel JS, Rani KSS. A Novel Vision-Efficient Grad-CAM Network for Early Breast Cancer Detection Using Multi-Scale Histopathological Image Analysis. *Biomedical Signal Processing and Control* 2026; **112**: 108939.

30 Pooch EH, Agrotis G, Cai L, *et al*. Semi-Supervised Learning in Prostate MRI Tumor Detection Approaches Fully Supervised Performance on External Validation. *European Radiology* 2026; 1–11. https://doi.org/10.1007/s00330-026-12324-x.

31 Yan H, Shao D. Enhancing Transformer Training Efficiency with Dynamic Dropout. *arXiv* 2024, arXiv: 2411.03236.

32 Deng X, Oda S, Kawano Y. Graphene-Based Midinfrared Photodetector with Bull's Eye Plasmonic Antenna. *Optical Engineering* 2023; **62**(**9**): 097102.

33 Li J, Culver TB. Review of Process-Based Nitrogen Model for Agricultural Fields with Implications for Nitrogen Simulations in Stormwater BMPs. *Environmental Modelling & Software* 2022; **151**: 105363.

34 Yan H. Real-Time 3D Model Reconstruction through Energy-Efficient Edge Computing. *Optimizations in Applied Machine Learning* 2022; **2**(**1**).

35 Lu Y, Shao D, Ni X, *et al*. Emotion-Style Dual Prediction: A Multi-Task Deep Learning Approach for Artistic Images. *Cluster Computing* 2026; **29**(**1**): 31.

36 Li J, Culver TB, Burgis CR, *et al*. Validating Nitrogen Removal Models with Field Bioretention Data. *Journal of Environmental Engineering* 2024; **150**(**8**): 04024037.

37 Deng X, Simanullang M, Kawano Y. Ge-Core/a-Si-Shell Nanowire-Based Field-Effect Transistor for Sensitive Terahertz Detection. *Photonics* 2018; **5**(**2**): 13.

38 Yan H, Shao D. Multimodal Medical Image Analysis: Integrating LLM and RAG Deep Learning Strategies. *Journal of Advances in Information Technology* 2025; **16**(**4**): 568–581. https://doi.org/10.12720/jait.16.4.568-581

39 Luo Z, Yan H, Pan X. Optimizing Transformer Models for Resource-Constrained Environments: A Study on Model Compression Techniques. *Journal of Computational Methods in Engineering Applications* 2023; **3** (**1**): 1–12. https://doi.org/10.62836/jcmea.v3i1.030107

40 Li J. Nitrogen Removal Models for Stormwater Bioretention Systems. Ph.D. Thesis, University of Virginia, Charlottesville, VA, USA, 2023.

41 Li J, Culver TB, Persaud PP, *et al.* Developing Nitrogen Removal Models for Stormwater Bioretention Systems. *Water Research* 2023; **243**: 120381.