N6-TreeMamba: Radiology Knowledge-Guided 6-Neighborhood Tree Scanning for 3D MRI-based Stroke Complication Prediction

Chenyang Zhao[1,2⋆], Yaoyao Yu[3⋆], Tianyi Xia[3⋆], Mengxuan Yuan[4], Yunjun Yang[5], and Chuanfei Hu[1]

[1] Faculty of Intelligence Technology, Shanghai Institute of Technology, Shanghai 201418, China
[2] Institute of Robotics and Automatic Information System, Nankai University, Tianjin 30071, China
[3] Department of Radiology, Zhongda Hospital, School of Medicine, Southeast University, Nanjing 210009, Jiangsu, China
[4] Department of Radiology, Jiangdu People's Hospital, Yangzhou University, Yangzhou 225000, Jiangsu, China
[5] Department of Radiology, The First Affiliated Hospital, Wenzhou Medical University, Wenzhou 325035, Zhejiang, China
chuanfei_hu@ieee.org, yyjunjim@163.com

Abstract. Acute ischemic stroke patients are at high risk of stroke-associated pneumonia (SAP) and hemorrhagic transformation (HT), where early prediction from brain MRI can enable timely interventions. However, when modeling long-range dependencies in MRI sequences, most existing CNN- or Transformer-based models either over-emphasize local patterns or incur prohibitive computational cost. Therefore, we propose N6-TreeMamba, a Mamba-based framework that combines a 6-neighborhood tree scanning (N6-TS) strategy to efficiently capture hierarchical long-range interactions. To further align the model with clinical reasoning, we design a radiology knowledge-guided anchoring (RKA) module that leverages white matter hyperintensity (WMH) segmentation to select scanning anchors, guiding the model to attend to clinically relevant regions while maintaining global coverage. Experiments on an in-house stroke MRI dataset demonstrate that N6-TreeMamba consistently improves performance while achieving lower computational complexity. Code is available at: `https://anonymous.4open.science/r/N6Treemamba-925`.

Keywords: Deep Learning · Radiology knowledge · State space model · Stroke Complication Prediction.

1 Introduction

Acute ischemic stroke is frequently complicated by stroke-associated pneumonia (SAP) [1] and hemorrhagic transformation (HT) [7], both of which substantially

⋆ Chenyang Zhao, Yaoyao Yu and Tianyi Xia contributed equally to this work.

worsen functional outcomes and increase mortality. Brain MRI provides rich multi-sequence evidence of tissue injury and microvascular pathology [2], making it an attractive modality for modeling SAP or HT risk. However, accurate MRI-based prediction remains challenging because clinically informative cues may be subtle, spatially diffuse, and vary across sequences, while the target events are influenced by both focal lesions and global brain conditions.

Deep learning approaches [31,8,29] for MRI-based prognosis have been dominated by CNNs and Transformer-style architectures [28,9]. However, CNN-based methods often struggle to model long-range dependencies in 3D MRI sequences, resulting in suboptimal performance. Transformer-based methods, while effective at modeling global context, are often impractical for real-time use due to their high computational complexity. Recently, state space models (SSMs) have emerged as a compelling alternative for long-context modeling [32,10,12], achieving competitive performance with linear-time complexity by replacing quadratic attention with selective sequence scanning. This favorable scaling makes SSMs particularly attractive for 3D medical imaging, where the token length is large.

Nevertheless, the effectiveness of an SSM critically depends on the 1D scan order used to linearize the 3D volume: a naive raster ordering [17] imposes an arbitrary traversal that breaks volumetric locality and fails to reflect the underlying anatomical structure. This observation motivates a tree-based scanning strategy [24], which constructs an input-dependent, locality-preserving token order over the 3D volume rather than relying on a fixed linearization. However, extending such tree-topology scans to 3D MRI sequences is non-trivial, as building input-dependent trees over dense 3D tokens can be expensive, and may still ignore the meaningful radiology knowledge closely associated with stroke complication prediction.

To address these issues, we propose N6-TreeMamba, a radiology knowledge-guided 6-neighborhood scanning framework for 3D MRI-based stroke complication prediction. The naive linearization is replaced with a 6-neighborhood tree scanning (N6-TS) strategy that generates an input-dependent token order, preserving 3D locality while enabling efficient long-context modeling within Mamba. Beyond structural ordering, we introduce a radiology knowledge-guided anchoring (RKA) module driven by white matter hyperintensities (WMH), a well-established biomarker of chronic small-vessel disease and long-term ischemic injury that is associated with SAP and HT risk [20]. Given a pre-aligned binary WMH segmentation at the MRI resolution, our method prioritizes WMH-related regions when selecting scanning anchors and determining traversal emphasis, allowing the model to rapidly attend to clinically meaningful pathology while covering the entire brain MRI sequence. To summarize, the main contributions are as follows:

1. N6-TreeMamba is proposed for 3D MRI-based SAP and HT prediction, ensuring full-sequence coverage while improving the ability to attend to clinically meaningful regions.
2. A 6-neighborhood tree scanning (N6-TS) strategy is proposed to replace naive linearization of 3D MRI sequences, generating sequences that better

preserve 3D neighborhood structure while enabling efficient long-range information propagation.
3. A radiology knowledge-guided anchoring (RKA) module is designed to select scanning anchors and traversal emphasis, encouraging early focus on pathology-related regions.

2 Methodology

2.1 Preliminaries

State space models (SSMs) are widely used to describe linear dynamical systems [23]. Given an input sequence $x(t) \in \mathbb{R}^L$, SSMs produce an output sequence $y(t) \in \mathbb{R}^L$ via a hidden state $h(t) \in \mathbb{R}^N$:

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t), \quad y(t) = \mathbf{C}h(t) + \mathbf{D}x(t), \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the state transition matrix, $\mathbf{B}, \mathbf{C} \in \mathbb{R}^N$ are projection matrices, and $\mathbf{D}$ denotes the skip connection. For deep learning, the continuous system is discretized. Following [5], a timescale $\mathbf{\Delta}$ yields:

$$\overline{\mathbf{A}} = e^{\mathbf{\Delta A}}, \quad \overline{\mathbf{B}} = (\Delta\mathbf{A})^{-1}\big(e^{\mathbf{\Delta A}} - \mathbf{I}\big) \cdot e^{\mathbf{\Delta B}}, \tag{2}$$

and the ZOH-discretized form is $h_t = \overline{\mathbf{A}}h_{t-1} + \overline{\mathbf{B}}x_t, y_t = \mathbf{C}h_t + \mathbf{D}x_t$.

Although vanilla SSMs scale linearly, their sequence-modeling capacity can be limited by time-invariant parameters. Mamba [4] introduces a selective scan (S6) that makes $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{\Delta}$ input-dependent, and combines a gated MLP [11]:

$$\begin{aligned} x'_t &= \sigma(\text{DWConv}(\text{Linear}(x_t))), \quad z_t = \sigma(\text{Linear}(x_t)), \\ \mathbf{B}_t, \mathbf{C}_t, \mathbf{\Delta}_t &= \text{Linear}(x'_t), \quad \overline{\mathbf{A}_t}, \overline{\mathbf{B}_t} = \text{ZOH}(\mathbf{A}, \mathbf{B}_t, \mathbf{\Delta}_t), \\ h'_t &= \overline{\mathbf{A}_t}h_{t-1} + \overline{\mathbf{B}_t}x'_t, \quad y'_t = \mathbf{C}_t h'_t + \mathbf{D}x'_t, \quad y_t = y'_t \otimes z_t, \end{aligned} \tag{3}$$

where σ is SiLU activation and $\otimes$ denotes element-wise multiplication.

2.2 N6-TreeMamba

Overview The overall framework of N6-TreeMamba is shown in Fig. 1, which can be divided into four main components: radiology knowledge-guided anchoring (RKA) module, 6-neighborhood tree scanning (N6-TS) strategy, TreeMamba blocks and prediction head.

Consider a multi-sequence MRI input $\mathbf{M} \in \mathbb{R}^{C\times D\times H\times W}$, with a prognosis label $\mathbf{y} \in \{0,1\}^K$, where C is the number of input sequences, and D, H, W denote depth, height, and width, respectively. First, a 3D convolutional patch embedding Conv3D with kernel (p_d, p_h, p_w) projects the input into non-overlapping visual tokens, and we further prepend a learnable classification token $\mathbf{x}_{cls}$ and add a learnable 3D positional embedding $\mathbf{p}$:

$$\mathbf{X} = \text{CAT}\big(\mathbf{x}_{cls},\, \text{Conv3D}(\mathbf{M})\big) + \mathbf{p}, \tag{4}$$

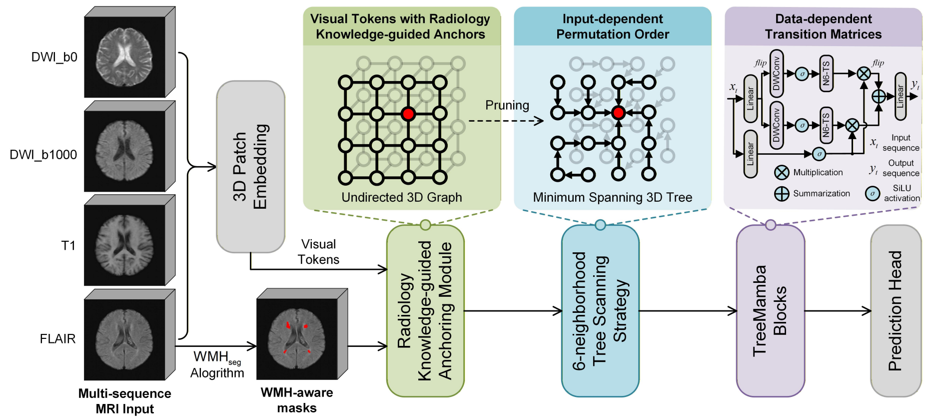

Fig. 1. The overall framework of N6-TreeMamba. Multi-sequence MRI sequences are first tokenized by a 3D patch embedding module to obtain visual tokens. In parallel, WMH-aware masks are introduced by the segmentation algorithm, which are fed into the radiology knowledge-guided anchoring (RKA) module to select clinically meaningful anchors. Based on these anchors, the 6-neighborhood tree scanning strategy (N6-TS) prunes the edges with significant dissimilarity and generates an input-dependent permutation order. The reordered tokens are processed by stacked TreeMamba blocks, and a prediction head outputs the final stroke-complication prognosis.

where $\mathrm{Conv3D}(\mathbf{M}) \in \mathbb{R}^{d \times N}$, d is the token dimension, $N = \frac{D}{p_d} \cdot \frac{H}{p_h} \cdot \frac{W}{p_w}$, $\mathbf{x}_{cls} \in \mathbb{R}^d$, $\mathbf{p} \in \mathbb{R}^{d \times (N+1)}$, and CAT denotes concatenation along the token axis. We then build an undirected 3D graph $\mathcal{G} = (\mathbf{X}, \mathcal{E})$ over tokens using 6-neighborhood connectivity to preserve volumetric locality, where $\mathcal{E}$ denotes the edge set of the graph. The edge weight is calculated by the feature dissimilarity between adjacent vertices.

Then, RKA module segments the WMH regions from the MRI sequence using the $\mathrm{WMH}_{\mathrm{seg}}$ algorithm [13] to obtain WMH-aware masks. We introduce two hyperparameters, K_{wmh} and K_{bg}, which specify the numbers of anchors to be sampled from WMH regions and from non-WMH background regions on the 3D graph, respectively. In practice, $\mathrm{RKA}(\cdot)$ extracts 3D connected components on the pooled WMH-aware masks, samples WMH anchors $\mathcal{A}_{\mathrm{wmh}}$ with farthest-point sampling for spatial coverage, and samples background anchors $\mathcal{A}_{\mathrm{bg}}$ from a narrow ring around WMH.

$$\mathcal{A} = \mathrm{RKA}\Big(\mathrm{Pool}\big(\mathrm{WMH}_{\mathrm{seg}}(\mathbf{M})\big);\ K_{\mathrm{wmh}}, K_{\mathrm{bg}}\Big), \tag{5}$$

$\mathcal{A} = \mathcal{A}_{\mathrm{wmh}} \cup \mathcal{A}_{\mathrm{bg}}$ is a set of anchor indices on the undirected 3D graph. Therefore, N6-TS constructs an input-dependent permutation order based on $\mathcal{A}$, and TreeMamba blocks employ data-dependent transition matrices to obtain the feature representation $\mathbf{F}$:

$$\mathbf{F} = \mathrm{TreeMamba}(\text{N6-TS}(\mathcal{A}, \mathbf{X})), \tag{6}$$

Finally, global pooling and a classification head produce the prognosis prediction:

$$\hat{\mathbf{y}} = \mathrm{Head}\big(\mathrm{Pool}(\mathbf{F})\big), \tag{7}$$

where $\mathrm{Pool}(\cdot)$ is global average pooling, $\mathrm{Head}(\cdot)$ is the prediction head layer.

Radiology Knowledge-guided Anchoring Module RKA injects radiological prior by using WMH as clinically meaningful cues to initialize scanning. Given the input multi-sequence MRI input $\mathbf{M}$, we obtain voxel-level WMH-aware masks $\mathbf{S}$ and align it to the 3D graph via max pooling within each 3D patch block (p_d, p_h, p_w):

$$\mathbf{S} = \mathrm{Pool}\big(\mathrm{WMH}_{\mathrm{seg}}(\mathbf{M})\big) \in \{0,1\}^{\frac{D}{p_d}\times\frac{H}{p_h}\times\frac{W}{p_w}}. \tag{8}$$

We introduce two hyperparameters, K_{wmh} and K_{bg}, which specify the numbers of anchors to be sampled from WMH and non-WMH background regions on the 3D graph, respectively. RKA outputs an anchor set $\mathcal{A} = \mathcal{A}_{\mathrm{wmh}} \cup \mathcal{A}_{\mathrm{bg}}$ following Eq. 5.

For WMH anchors, we extract 3D connected components on $\mathbf{S}$ using 26-connectivity and allocate anchors across components proportional to their sizes. Within each component, we sample anchor positions using farthest-point sampling to encourage spatial coverage of heterogeneous WMH patterns. For background anchors, we preferentially sample from a thin ring around WMH, defined by one-step dilation minus WMH on the undirected 3D graph, and fall back to the global non-WMH area when the ring is empty. These WMH-guided anchors serve as clinically relevant roots for the subsequent N6-TS scan order construction.

6-neighborhood Tree Scanning Strategy Given the undirected graph $\mathcal{G}$ over tokens, we use the Boruvka algorithm [19] to prune the edges with significant dissimilarity and generate a minimum spanning tree $\mathcal{G}_T$. The core of our method lies in the introduction of WMH-guided anchors $\mathcal{A}$ from the RKA module. Instead of a standard traversal, we perform a multi-source Breadth-First Search (BFS) on $\mathcal{G}_T$ starting from $\mathcal{A}$ to generate a WMH-guided scan order.

Following the order, we employ Mamba-style data-dependent transition matrices for state propagation. The aggregated feature h_i for vertex i is defined by the path integrals from all $j \in \Omega$:

$$h_i = \sum_{\forall j \in \Omega} \left(\prod_{k \in N_{ij}} \bar{\mathbf{A}}_k \right) \bar{\mathbf{B}}_j x_j \tag{9}$$

where Ω denotes the index set of all vertices in the tree. N_{ij} denotes the path vertices from j to i, where the distance $\mathrm{dist}(i,j) \le 1$. $\bar{\mathbf{A}}_k$, $\bar{\mathbf{B}}_j$ are the learned state transition and input matrices, respectively. The final feature representation

$\mathbf{F}$ is obtained by fusing the aggregated topological features $\mathbf{H} = \{h_i, i \in \mathcal{A}\}$ with the original input $\mathbf{X}$ through a gated residual connection:

$$\mathbf{F} = \mathbf{C} \odot \text{Norm}(\mathbf{H}) + \mathbf{D} \odot \mathbf{X} \tag{10}$$

where $\mathbf{C}$ and $\mathbf{D}$ are learnable gating parameters that adaptively balance global topological context and local feature details. $\text{Norm}(\cdot)$ is the normalization operation.

3 Experiments

Experimental Setup We evaluated N6-TreeMamba on an in-house dataset containing MRI data from 188 patients, with a total of 752 MRI sequences collected at Jiangdu People's Hospital Affiliated to Yangzhou University. Each sequence has the same spatial resolution of $99 \times 117 \times 95$ with 2mm spacing. Within the collected dataset, stroke-associated pneumonia (SAP) occurs in 41 patients, and hemorrhagic transformation (HT) occurs in 25 patients. The dataset was partitioned into training and test sets using an 8:2 ratio. For the SAP prediction $\mathcal{T}_{\text{SAP}}$ and HT prediction $\mathcal{T}_{\text{HT}}$, 149 patients are assigned to the training set and 39 to the test set.

The experiments are conducted on a workstation with NVIDIA Tesla A100 GPUs. AdamW optimizer [18] with a weight decay of $1e-5$ is adopted to train the network. The total number of epochs is set to 200 and the initial learning rate is $5e-3$, while the learning rate is adjusted by cosine annealing cycles. All MRI sequences are resized to $64 \times 128 \times 128$. We tokenize the input using a 3D patch embedding with patch size/stride $(p_d, p_h, p_w) = (4, 16, 16)$. For RKA, we sample a total of $K_{\text{total}} = 64$ anchors, including $K_{\text{wmh}} = 48$ anchors from WMH regions and $K_{\text{bg}} = 16$ anchors from non-WMH background.

To comprehensively assess the performance of our proposed framework N6-TreeMamba in predicting SAP and HT, we employ metrics including accuracy (**Acc**), and Area Under the Receiver Operating Characteristic Curve (**AUC**). Besides, model parameters (**Params**) and **FLOPs** are utilized to evaluate the computational complexity of the model.

Comparisons with other methods We conduct comparative experiments with state-of-the-art methods under two distinct training paradigms: CNN or Transformer-based models, and Mamba-based models. The compared methods can be further categorized in terms of different visual encoders (Med3D [3], SwinUNETR [6], UniFormer [14], UNET3D [21]), methods designed with different pre-training strategies (BrainMVP [22], SimMIM [25]), and methods designed with different scan strategy (VMamba [16], PlainMamba [27], RSMamba [33], and GrootVL [24]). Since some of the compared models are originally designed for medical image segmentation, we replace their segmentation-specific heads with our prediction head layer for $\mathcal{T}_{\text{SAP}}$ and $\mathcal{T}_{\text{HT}}$. Across all Mamba-based models, we replace the original 2D visual embedding with a 3D embedding layer to accommodate multi-sequence MRI data. In addition, we initialize the root nodes of the tree scan using the attention maps for GrootVL.

Table 1. Comparison of our method with the state-of-the-art methods on the collected dataset. Acc (%)↑ and AUC (%)↑ are utilized to evaluate the performance for $\mathcal{T}_{\mathrm{SAP}}$ and $\mathcal{T}_{\mathrm{HT}}$, Params (M)↓ and FLOPs (G)↓ are utilized to evaluate the computational complexity, where the best results are highlighted in **bold**.

Methods	Year	Backbone	Scan Strategy	Params	FLOPs	$\mathcal{T}_{\mathrm{SAP}}$		$\mathcal{T}_{\mathrm{HT}}$	
						Acc	AUC	Acc	AUC
CNN/Transformer-based models									
Med3D [3]	2019	Res-200	—	130.76	538.42	46.86	41.67	46.88	37.50
UniFormer [14]	2023	UniFormer	—	6.74	**2.46**	73.38	64.74	74.25	46.43
SwinUNETR [6]	2021	Swin-B	—	247.90	1567.38	78.13	60.90	71.25	73.21
SimMIM [25]	2022	Swin-B	—	27.49	364.22	78.76	74.49	71.58	74.46
UNET3D [21]	2015	UNET3D	—	14.14	78.12	36.36	69.23	80.75	80.28
M^3AE [15]	2023	UNET3D	—	18.80	1169.78	75.63	73.21	**81.25**	78.57
BrainMVP [22]	2025	UNET3D	—	22.04	228.61	72.50	76.25	77.73	77.50
Mamba-based models									
VMamba [16]	2024	VMamba	Bidirectional	7.36	27.25	74.57	77.88	70.68	74.81
PlainMamba [27]	2024	VMamba	Continuous Bidirectional	7.38	28.14	78.33	78.90	73.56	77.21
RSMamba [33]	2024	VMamba	Omnidirectional	7.58	51.26	73.26	71.92	72.18	73.82
SegMamba-V2 [26]	2026	VMamba	Tri-orientated	7.41	36.78	75.70	76.82	73.59	76.18
GrootVL [24]	2024	TreeMamba	Tree Scan	7.32	30.03	79.34	80.11	78.45	79.68
N6-TreeMamba	2026	TreeMamba	N6-TS	**6.50**	26.58	**82.57**	**82.61**	76.92	**81.66**

Table 1 reports the comparison on the collected dataset for $\mathcal{T}_{\mathrm{SAP}}$ and $\mathcal{T}_{\mathrm{HT}}$, together with computational complexity. Overall, N6-TreeMamba achieves the best or near-best performance with the lowest model complexity among Mamba-based methods, confirming the effectiveness of RKA module and N6-TS strategy for long-context volumetric modeling. Compared with CNN/Transformer baselines, many 3D models either underperform (Med3D) or require prohibitive FLOPs (SwinUNETR and M^3AE), indicating that simply increasing capacity or using attention-heavy architectures is not ideal for long-context volumetric prognosis.

Among Mamba-based baselines, scan strategy plays a decisive role. The visualizations of different scan strategies is shown in Fig. 2. VMamba provides a strong efficiency baseline with 7.36M Params and 27.25G FLOPs but yields limited accuracy. PlainMamba slightly improves performance with similar complexity, while SegMamba-V2 and RSMamba substantially increase FLOPs without consistent gains, suggesting that adding scan directions alone is inefficient. The tree scan of GrootVL improves the performance of both tasks, and our N6-TreeMamba further advances $\mathcal{T}_{\mathrm{SAP}}$ and $\mathcal{T}_{\mathrm{HT}}$ to the best AUC 82.61% and 81.66%. Such results support that RKA module and N6-TS strategy yield a more spatially coherent and clinically prioritized scan order for selective scanning. Besides, N6-TreeMamba presents the lightest Params and maintaining low FLOPs, which confirms our method achieves a strong accuracy–efficiency trade-off compared to the Mamba-based methods.

Ablation studies Table 2 verifies the effectiveness of both modules. Our method with N6-TS consistently improves performance over the baseline, indicating that

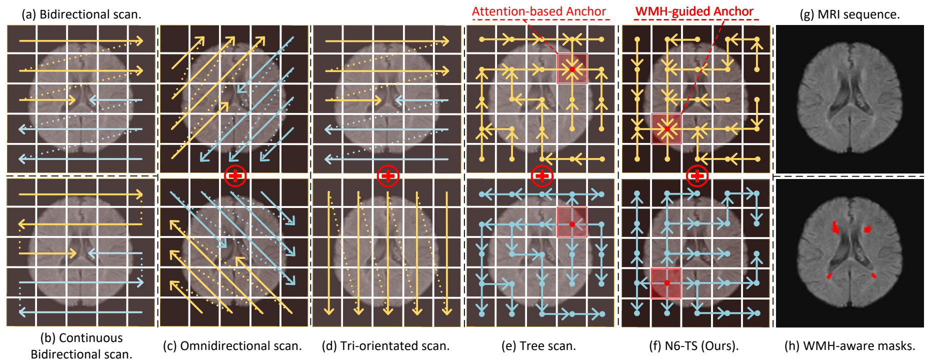

Fig. 2. Different scan strategies used in Mamba blocks. For Tree scan and N6-TS, we select one anchor as the root to visualize the scanning order.

Table 2. Ablation analysis of N6-TS and RKA on the collected dataset.

w/ N6-TS	w/ RKA	$\mathcal{T}_{\mathrm{SAP}}$ Acc	$\mathcal{T}_{\mathrm{SAP}}$ AUC	$\mathcal{T}_{\mathrm{HT}}$ Acc	$\mathcal{T}_{\mathrm{HT}}$ AUC
		79.34	80.11	78.45	79.68
✓		81.50	81.91	79.30	80.02
	✓	81.86	80.77	78.12	81.25
✓	✓	82.57	82.61	76.92	81.66

Table 3. Ablation analysis of different counts of K_{wmh} and K_{bg}.

K_{wmh}	K_{bg}	$\mathcal{T}_{\mathrm{SAP}}$ Acc	$\mathcal{T}_{\mathrm{SAP}}$ AUC	$\mathcal{T}_{\mathrm{HT}}$ Acc	$\mathcal{T}_{\mathrm{HT}}$ AUC
64	0	77.67	81.26	78.12	80.89
32	32	79.60	81.58	80.05	81.34
16	48	75.25	76.47	74.89	78.72
48	16	82.57	82.61	76.92	81.66

a 6-neighborhood scan order better preserves 3D locality for selective scanning. Adding RKA brings further gains, especially in AUC, suggesting that WMH-guided anchors help the model focus on clinically relevant regions. Combining RKA and N6-TS achieves the best AUC on both tasks, demonstrating their complementarity.

Table 3 studies anchor composition. Using only WMH anchors ($K_{\mathrm{wmh}} = 64$ and $K_{\mathrm{bg}} = 0$) is strong, while introducing a small number of background anchors improves robustness. The best setting is $K_{\mathrm{wmh}} = 48$ and $K_{\mathrm{bg}} = 16$. In contrast, overly background-dominant anchors ($K_{\mathrm{wmh}} = 16$ and $K_{\mathrm{bg}} = 48$) significantly degrade AUC, indicating that anchors should remain WMH-dominant with a modest background complement for contextual coverage.

Limitation Despite the promising results, N6-TreeMamba has two main limitations. First, the RKA module relies on automatic WMH segmentation for anchor selection, so segmentation errors may shift the scanning anchors away from clinically stroke-relevant regions. Future work will mitigate the dependence on segmentation quality through more advanced segmentation models [30] and manual refinement of WMH regions. Second, our evaluation is conducted on a single-center in-house cohort with a limited number of positive cases, particularly for HT. As no public 3D MRI dataset for SAP or HT prediction is currently available, we are collecting multi-center data, and an extension of this work will validate N6-TreeMamba across multiple centers with larger cohorts.

4 Conclusion

In this paper, we presented a radiology knowledge-guided TreeMamba framework for 3D MRI-based stroke complication prediction. By introducing a radiology knowledge-guided anchoring (RKA) module, our method leverages WMH priors to select clinically meaningful anchors. Building on these anchors, we propose a 6-neighborhood tree scanning (N6-TS) strategy to construct an input-dependent scan order, enabling efficient and locality-preserving long-context modeling within Mamba blocks. Future works will explore richer clinical priors, adaptive anchor selection, and extending the strategy to other volumetric modalities and downstream tasks.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (grant 82572173 and grant 823B2040), and the Research Personnel Cultivation Programme of Zhongda Hospital, Southeast University (grant CZXM-GSP-RC179).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Assefa, M., Tadesse, A., Adane, A., Yimer, M., Tadesse, M.: Factors associated with stroke associated pneumonia among adult stroke patients admitted to university of gondar hospital, northwest ethiopia. Scientific Reports **12**(1), 12724 (2022)
2. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E., et al.: Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. Medical image analysis **86**, 102789 (2023)
3. Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625 (2019)
4. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
5. Gupta, A., Gu, A., Berant, J.: Diagonal state spaces are as effective as structured state spaces. Advances in Neural Information Processing Systems **35**, 22982–22994 (2022)
6. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
7. Hong, J.M., Kim, D.S., Kim, M.: Hemorrhagic transformation after ischemic stroke: mechanisms and management. Frontiers in Neurology **12**, 703258 (2021)
8. Hu, C., Li, X.: Human-centric context and self-uncertainty-driven multi-modal large language model for training-free vision-based driver state recognition. IEEE Transactions on Intelligent Transportation Systems (2025)
9. Hu, C., Li, X., Pang, J.: Trustworthy driver state perception via contextual interaction-driven evidential vision-language fusion in vehicular cyber-physical systems. IEEE Transactions on Intelligent Transportation Systems **27**(2), 2716–2728 (2026)

10. Hu, C., Xia, T., Cui, Y., Zou, Q., Wang, Y., Xiao, W., Ju, S., Li, X.: Trustworthy multi-phase liver tumor segmentation via evidence-based uncertainty. Engineering Applications of Artificial Intelligence **133**, 108289 (2024)
11. Hua, W., Dai, Z., Liu, H., Le, Q.: Transformer quality in linear time. In: International conference on machine learning. pp. 9099–9117. PMLR (2022)
12. Jiao, R., Zhang, J., Li, C., Hu, L.: Large-kernel spatially parallel feature fusion for monocular 3d perception in autonomous driving. Knowledge-Based Systems **343**, 115998 (2026)
13. Li, J., Santini, T., Huang, Y., Mettenburg, J.M., Ibrahim, T.S., Aizenstein, H.J., Wu, M.: wmh_seg: Transformer based u-net for robust and automatic white matter hyperintensity segmentation across 1.5 t, 3t and 7t. arXiv preprint arXiv:2402.12701 (2024)
14. Li, K., Wang, Y., Zhang, J., Gao, P., Song, G., Liu, Y., Li, H., Qiao, Y.: Uniformer: Unifying convolution and self-attention for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(10), 12581–12600 (2023)
15. Liu, H., Wei, D., Lu, D., Sun, J., Wang, L., Zheng, Y.: M3ae: Multimodal representation learning for brain tumor segmentation with missing modalities. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 1657–1665 (2023)
16. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Liu, Y.: Vmamba: Visual state space model. arXiv preprint arXiv:2401.10166 (2024)
17. Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., Hu, H.: Video swin transformer. In: Proceedings of CVPR. pp. 3202–3211 (2022)
18. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
19. O, B.: O jistém problému minimálním (1926)
20. Pisano, F., Rizzo, G., D'Angelo, L., Bilotta, F.: Predictive role of white matter hyperintensities in poststroke aphasia: A systematic review of clinical evidence. Stroke **57**(1), 89–95 (2026)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
22. Rui, S., Chen, L., Tang, Z., Wang, L., Liu, M., Zhang, S., Wang, X.: Multi-modal vision pre-training for medical image analysis. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5164–5174 (2025)
23. Williams, R.L., Lawrence, D.A., et al.: Linear state-space control systems. John Wiley & Sons (2007)
24. Xiao, Y., Song, L., Huang, S., Wang, J., Song, S., Ge, Y., Li, X., Shan, Y.: Mambatree: Tree topology is all you need in state space model. In: Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C. (eds.) Advances in Neural Information Processing Systems. vol. 37, pp. 75329–75354. Curran Associates, Inc. (2024)
25. Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., Hu, H.: Simmim: A simple framework for masked image modeling. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9653–9663 (2022)
26. Xing, Z., Ye, T., Yang, Y., Cai, D., Gai, B., Wu, X.J., Gao, F., Zhu, L.: Segmamba-v2: Long-range sequential modeling mamba for general 3-d medical image segmentation. IEEE Transactions on Medical Imaging **45**(1), 4–15 (2026)
27. Yang, C., Chen, Z., Espinosa, M., Ericsson, L., Wang, Z., Liu, J., Crowley, E.J.: Plainmamba: Improving non-hierarchical mamba in visual recognition. arXiv preprint arXiv:2403.17695 (2024)

28. Zhang, J., Song, X., Li, Y., Liang, D., Zhang, Z., Cai, J.: Adaptive dual cross-attention network for multispectral object detection in autonomous driving. Expert Systems with Applications **318**, 132012 (2026)
29. Zhang, J., Xiang, M., Hu, Y., Hao, W., Lei, L., Yi, K.: Multivariate feature learning and associative spatial information enhancement for snow object detection in autonomous driving. Engineering Applications of Artificial Intelligence **175**, 114672 (2026)
30. Zhang, Y., Lv, B., Xue, L., Zhang, W., Liu, Y., Fu, Y., Cheng, Y., Qi, Y.: Semisam+: Rethinking semi-supervised medical image segmentation in the era of foundation models. Medical Image Analysis **106**, 103733 (2025)
31. Zhao, C., Hu, C., Cao, Z., Song, Y., Liu, J.: Pipeclip: Defect-conditioned and cross-focus-driven vision-language model for video-based sewer defect inspection. IEEE Transactions on Automation Science and Engineering (2026)
32. Zhao, C., Hu, C., Shao, H., Liu, J.: Pipemamba: State space model for efficient video-based sewer defect classification. IEEE Transactions on Artificial Intelligence **7**(3), 1390–1403 (2026)
33. Zhao, S., Chen, H., Zhang, X., Xiao, P., Bai, L., Ouyang, W.: Rs-mamba for large remote sensing image dense prediction. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–14 (2024)