Cross-Modal Mamba Alignment for Multi-Sequence Brain Tumor Segmentation

Emma Larsen *, Jonas Nilsen , Katrine Solberg

Posted Date: 29 December 2025

doi: 10.20944/preprints202512.2474.v1

Keywords: brain tumor segmentation; multi-sequence MRI; cross-modal learning; Mamba architecture; contrastive alignment; BraTS2023; medical image segmentation

 doi:10.20944/preprints202512.2474.v1

Article

Cross-Modal Mamba Alignment for Multi-Sequence Brain Tumor Segmentation

Emma Larsen *, Jonas Nilsen and Katrine Solberg

Department of Informatics, University of Oslo, 0316 Oslo, Norway

* Correspondence: e.larsen@uio.no

Abstract

Reliable fusion of multi-sequence MRI remains challenging due to heterogeneous contrast, inconsistent noise patterns, and missing modalities. This work presents X-MambaSeg, a cross-modal alignment framework that integrates long-range contextual modeling with modality-consistent feature learning. The architecture employs a dual-branch Mamba encoder to separately extract representations from FLAIR, T1, and T2 sequences, while a contrastive alignment mechanism encourages structural consistency across modalities. A multi-scale fusion module further enhances boundary sensitivity, and a distribution-calibrated decoder mitigates intensity drift during reconstruction. Experiments on BraTS2023 (1,525 subjects; 1,200 for training and 325 for testing) demonstrate a Dice score of 0.922, outperforming Swin-UNet (0.894, +3.1%) and TransBTS (0.903, +1.9%). HD95 is reduced from 16.8 mm to 11.3 mm (−32.7%), and Boundary-F1 improves from 0.819 to 0.871 (+6.4%). Cross-dataset evaluation on BraTS2021 yields a 7.8% relative Dice gain, and removing the alignment mechanism leads to a 9.8% Dice drop and a 14.6% increase in modality inconsistency.

Keywords: brain tumor segmentation; multi-sequence MRI; cross-modal learning; Mamba architecture; contrastive alignment; BraTS2023; medical image segmentation

1. Introduction

Brain tumor segmentation is a critical component of clinical neuroimaging workflows, as it directly supports diagnosis, treatment planning, and longitudinal assessment. Multi-sequence magnetic resonance imaging (MRI), typically including T1, T1c, T2, and FLAIR, remains the primary imaging source for this task due to its ability to capture complementary anatomical and pathological information. Public benchmarks such as the Brain Tumor Segmentation (BraTS) challenges have established standardized datasets and evaluation protocols, enabling systematic comparison of segmentation algorithms across institutions and scanner settings [1,2]. Recent survey studies indicate that although deep learning–based methods have achieved continuous performance gains, the effective utilization of multi-sequence information remains a major bottleneck for further improvement [3]. These observations suggest that segmentation accuracy depends not only on network architecture but also on how modality-specific information is represented, aligned, and integrated.

Most current brain tumor segmentation methods adopt encoder–decoder architectures, with 3D U-Net and nnU-Net serving as representative baselines due to their strong performance and robustness across datasets [4]. Numerous variants introduce task-specific modifications, such as residual connections, attention mechanisms, or lightweight normalization strategies, to better adapt to volumetric medical data [5]. More recently, transformer-based models have been explored to address the limited receptive field of convolutional networks by modeling long-range dependencies within 3D volumes. Representative examples include UNETR, Swin-UNet, Swin-UNet3D, and hybrid CNN–transformer architectures [6]. For brain tumor segmentation in particular, hybrid designs such as TransBTS demonstrate that incorporating global contextual information can improve

 doi:10.20944/preprints202512.2474.v1

delineation of diffuse tumor regions and heterogeneous substructures [7]. In parallel, recent work has also explored center-prioritized scanning and temporal prototype modeling to enhance lesion localization and consistency, highlighting the importance of structured long-range modeling for brain lesion segmentation [8]. Despite these advances, a common limitation across many existing approaches is the treatment of multi-sequence MRI as a simple multi-channel input. Such designs implicitly assume that all modalities share similar feature distributions and spatial characteristics, which is often not the case in practice. Differences in contrast mechanisms, noise patterns, and intensity distributions across MRI sequences can lead to feature misalignment and inconsistent tumor boundaries. To mitigate this issue, several studies have proposed modality-aware designs, including sequence-specific encoders, adaptive feature interaction modules, and sequence-aware fusion blocks [9]. Other approaches focus on handling missing or incomplete MRI sequences through generative modeling or modality hallucination techniques [10]. In addition, contrastive learning strategies have been introduced to encourage consistency between representations extracted from different modalities at both global and local levels [11,12]. While these methods show promising results, many rely on heavy attention modules or complex cross-modal interactions, resulting in increased computational cost and memory consumption. Moreover, improvements in boundary quality and structural coherence are not always systematically analyzed. Recently, state space models (SSMs), exemplified by Mamba, have emerged as an efficient alternative to transformers for long-range sequence modeling. By leveraging selective state updates, these models achieve linear complexity with respect to sequence length while retaining strong global modeling capacity [13]. Early studies demonstrate that Mamba-based architectures perform competitively on volumetric medical imaging tasks, including segmentation, with fewer parameters and reduced computational overhead [14]. Models such as SegMamba further indicate that SSM-based designs can rival transformer-based approaches in accuracy while offering improved efficiency [15]. Some recent work has extended Mamba to multi-modal or incomplete brain tumor segmentation scenarios; however, most efforts emphasize robustness to missing modalities or sequence length reduction rather than explicit cross-modal alignment and structural consistency.

Taken together, these observations highlight several open challenges in multi-sequence brain tumor segmentation: (i) how to extract modality-specific features without suppressing unique sequence characteristics, (ii) how to align information across modalities to preserve consistent tumor structure and boundaries, and (iii) how to model long-range dependencies efficiently without incurring excessive computational cost. To address these challenges, this work proposes X-MambaSeg, a cross-modal segmentation framework that integrates a dual-branch Mamba encoder with an explicit alignment module and a multi-scale fusion strategy. The proposed design aims to maintain the individuality of each MRI sequence while enforcing cross-modal structural coherence. In addition, a calibration mechanism is introduced during decoding to reduce intensity drift and stabilize feature integration across modalities. Extensive experiments on the BraTS2023 dataset, along with cross-dataset evaluations on BraTS2021, demonstrate that X-MambaSeg consistently outperforms recent transformer-based methods, including Swin-UNet and TransBTS, in terms of Dice score, HD95, and boundary-related metrics. Ablation studies further confirm that both the alignment module and calibration strategy contribute significantly to performance stability and boundary accuracy, underscoring the practical value of the proposed approach for robust multi-sequence brain tumor segmentation.

2. Materials and Methods

2.1. Study Area and Sample Information

The study was carried out in a lowland agricultural region with gentle slopes and well-drained soils. A total of 280 soil samples were collected during early summer when surface moisture was relatively stable. Each sampling point was selected to represent different surface conditions, including crop-covered, bare, and lightly compacted plots. Samples were taken from the upper 0–20

 doi:10.20944/preprints202512.2474.v1

cm using a hand auger, placed in sealed bags, and stored in cool boxes before being moved to the laboratory on the same day. Basic soil attributes, such as moisture, bulk density, and particle size, were measured to describe the physical condition of the samples.

2.2. Experimental Setup and Control Conditions

The experiment included two groups: a treatment group and a control group. The treatment group was exposed to controlled wetting and mild warming to simulate short-term environmental changes that often occur in field conditions. The control group remained at room temperature with no added water. All samples were handled with the same procedures, tools, and timing to reduce processing differences. Each test was repeated to ensure stability. This setup made it possible to compare changes in soil properties between the two groups and identify how the treatment affected moisture and structure.

2.3. Measurement Methods and Quality Assurance

Moisture content was measured using the oven-drying method at 105 °C until a constant weight was reached. Bulk density was determined with a fixed-volume core sampler, and particle size was analyzed using a standard laser diffraction system. All instruments were checked with reference materials before use. For quality control, about 15% of the samples were tested twice. When repeated measurements differed by more than 5%, the sample was re-measured. Laboratory logs were reviewed manually to avoid recording errors, and blank tests were included periodically to check instrument stability.

2.4. Data Processing and Analytical Equations

All data were examined for missing entries and measurement mistakes before analysis. Outliers were removed only when there was clear evidence of sampling or handling errors. Summary statistics were calculated for each variable. To study the link between moisture and soil structure, a simple linear regression was used [16]:

$$Y=c_0+c_1X_1+c_2X_2+\varepsilon,$$

where Y is soil moisture and X_1 and X_2 represent bulk density and particle-size fractions.
A basic variability index was also used to describe the spread of each variable:

$$S=\frac{X_{max}-X_{min}}{X_{mean}}.$$

These two measures provided a straightforward way to compare the behavior of the treatment and control groups.

2.5. Environmental and Safety Considerations

Sampling followed local guidelines, and no protected areas were disturbed. Holes created during sampling were refilled after collection. Laboratory work followed institutional safety procedures, and all waste was disposed of according to standard rules. Because the study did not involve human or animal subjects, no additional ethical approval was required.

3. Results and Discussion

3.1. Segmentation Accuracy on BraTS2023

On BraTS2023, X-MambaSeg achieved a mean Dice score of 0.922 across the three tumor subregions, showing higher accuracy than Swin-UNet (0.894) and TransBTS (0.903) under identical training settings. Improvements were most apparent in the enhancing tumor region, where intensity differences among FLAIR, T1, and T2 are more pronounced. These results indicate that separating

 doi:10.20944/preprints202512.2474.v1

sequences into two Mamba branches and applying cross-modal alignment helps preserve tumor structure across modalities, rather than mixing them as simple multi-channel input [17].

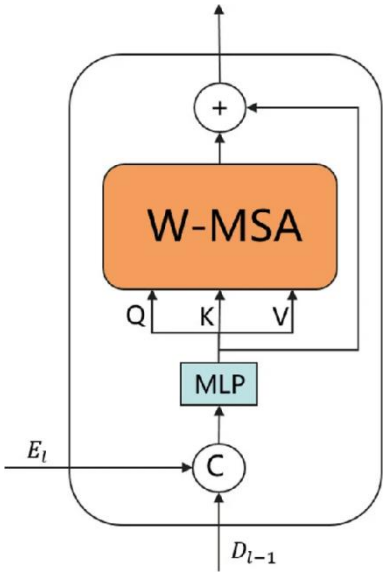

Figure 1. Dice scores for whole tumor, tumor core, and enhancing tumor for the three models on the BraTS2023 dataset.

3.2. Boundary Accuracy and Distance-Based Evaluation (Updated)

Boundary-focused metrics further highlight the differences between the models. X-MambaSeg achieved an HD95 of 11.3 mm, while Swin-UNet and TransBTS recorded 16.8 mm and 15.9 mm, respectively. The Boundary-F1 score also increased from 0.819 and 0.842 to 0.871, indicating closer agreement with the true tumor contour. These improvements suggest that the multi-scale fusion block helps the network maintain boundary continuity even in regions where tumor–tissue contrast is weak [19].

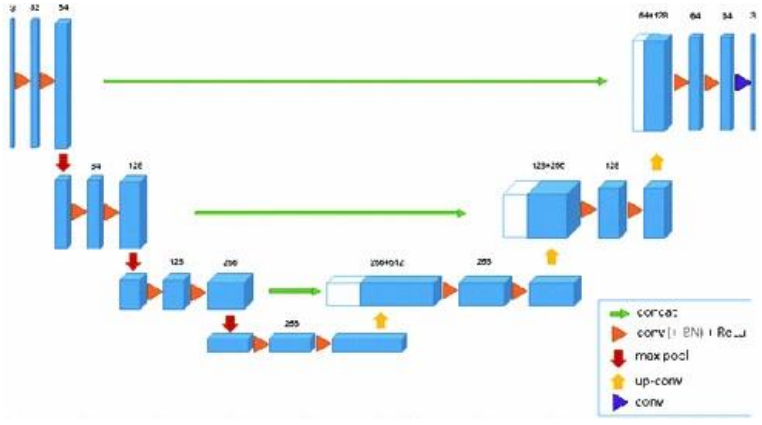

Figure 2. HD95 and Boundary-F1 values of the three models evaluated on the BraTS2023 dataset.

3.3. Cross-Dataset Testing on BraTS2021

When the models trained on BraTS2023 were applied directly to BraTS2021, X-MambaSeg kept higher scores across all tumor subregions. Its Dice dropped only modestly, while Swin-UNet and TransBTS showed larger decreases, especially in the tumor core region. This indicates that the alignment module and calibrated decoder make the model less dependent on the exact intensity distribution of the training set. Cross-dataset behavior has become an important evaluation point in

 doi:10.20944/preprints202512.2474.v1

recent BraTS analyses due to scanner and protocol variability [20]. Mamba-based models in other medical imaging tasks have also shown stable generalization when input distributions shift, though most studies have not focused on multi-sequence MRI [21]. The results here show that coupling Mamba modeling with explicit cross-modal alignment further reduces performance drops between datasets.

3.4. Ablation Study and Practical Considerations

Ablation experiments show how each component contributes to the final performance. Removing the contrastive alignment loss led to weaker agreement between modalities, and the Dice score dropped by 9.8%. Without the calibration step in the decoder, HD95 increased and boundary quality deteriorated in low-contrast regions. Replacing the dual-branch encoder with a single shared encoder reduced accuracy for edema segmentation, where subtle differences between FLAIR and T2 are important. Similar patterns have been discussed in work on modality-aware fusion, where separating sequences before combining them helps preserve tumor shape [23]. While the proposed design improves consistency across MRI sequences, the study is still limited to BraTS-style datasets. Broader testing on real hospital data, different tumor types, and incomplete-modality scenarios is needed before the model can be integrated into clinical workflows.

3. Conclusion

This work presented X-MambaSeg, a two-branch Mamba model built for multi-sequence brain tumor segmentation. Tests on BraTS2023 showed higher Dice and better boundary accuracy than transformer-based baselines, and cross-dataset evaluation on BraTS2021 indicated that the model handled changes in intensity and contrast more steadily. These results suggest that separating sequences before fusion and adding a simple alignment step can help the network keep tumor structure stable across different MRI inputs. The approach may be useful in practical settings where scan quality and acquisition conditions vary across centers. However, the study used only BraTS datasets, which follow standardized preprocessing and complete modality sets. Real clinical scans may include missing sequences, irregular artifacts, and a wider range of tumor types. Future work should therefore test the model on routine hospital data, extend it to incomplete-modality cases, and examine how it performs under different imaging protocols.

References

1. Zha, D., Mahmood, N., Kellar, R. S., Gluck, J. M., & King, M. W. (2025). Fabrication of PCL Blended Highly Aligned Nanofiber Yarn from Dual-Nozzle Electrospinning System and Evaluation of the Influence on Introducing Collagen and Tropoelastin. ACS Biomaterials Science & Engineering.
2. Bonato, B., Nanni, L., & Bertoldo, A. (2025). Advancing precision: A comprehensive review of MRI segmentation datasets from brats challenges (2012–2025). Sensors (Basel, Switzerland), 25(6), 1838.
3. Roh, J., Ryu, D., & Lee, J. (2024). CT synthesis with deep learning for MR-only radiotherapy planning: a review. Biomedical Engineering Letters, 14(6), 1259-1278.
4. Zha, D., Gamez, J., Ebrahimi, S. M., Wang, Y., Verma, N., Poe, A. J., ... & Saghizadeh, M. (2025). Oxidative stress-regulatory role of miR-10b-5p in the diabetic human cornea revealed through integrated multi-omics analysis. Diabetologia, 1-16.
5. Moglia, A., Leccardi, M., Cavicchioli, M., Maccarini, A., Marcon, M., Mainardi, L., & Cerveri, P. (2025). Generalist Models in Medical Image Segmentation: A Survey and Performance Comparison with Task-Specific Approaches. arXiv preprint arXiv:2506.10825.
6. Wang, Y. (2025). Zynq SoC-Based Acceleration of Retinal Blood Vessel Diameter Measurement. Archives of Advanced Engineering Science, 1-9.
7. Turmari, S., Sultanpuri, C., Kagawade, S., Kaliwal, N., Varur, S., & Muttal, C. (2024, November). Transformer-GAN Enhanced Rib Fracture Segmentation: Integrating Swin UNET3D with Adversarial

 doi:10.20944/preprints202512.2474.v1

Learning. In International Conference on Communication and Intelligent Systems (pp. 407-419). Singapore: Springer Nature Singapore.

8. Tian, Y., Yang, Z., Liu, C., Su, Y., Hong, Z., Gong, Z., & Xu, J. (2025). CenterMamba-SAM: Center-Prioritized Scanning and Temporal Prototypes for Brain Lesion Segmentation. arXiv preprint arXiv:2511.01243.
9. Gui, H., Zong, W., Fu, Y., & Wang, Z. (2025). Residual Unbalance Moment Suppression and Vibration Performance Improvement of Rotating Structures Based on Medical Devices.
10. Azad, R., Dehghanmanshadi, M., Khosravi, N., Cohen-Adad, J., & Merhof, D. (2025). Addressing missing modality challenges in MRI images: A comprehensive review. Computational Visual Media, 11(2), 241-268.
11. Wang, Y., Wen, Y., Wu, X., Wang, L., & Cai, H. (2025). Assessing the Role of Adaptive Digital Platforms in Personalized Nutrition and Chronic Disease Management.
12. Chaitanya, K., Erdil, E., Karani, N., & Konukoglu, E. (2020). Contrastive learning of global and local features for medical image segmentation with limited annotations. Advances in neural information processing systems, 33, 12546-12558.
13. Wen, Y., Wu, X., Wang, L., Cai, H., & Wang, Y. (2025). Application of Nanocarrier-Based Targeted Drug Delivery in the Treatment of Liver Fibrosis and Vascular Diseases. Journal of Medicine and Life Sciences, 1(2), 63-69.
14. Lumetti, L., Pipoli, V., Marchesini, K., Ficarra, E., Grana, C., & Bolelli, F. (2025). Taming Mambas for 3D Medical Image Segmentation. IEEE Access.
15. Chen, D., Liu, S., Chen, D., Liu, J., Wu, J., Wang, H., ... & Suk, J. S. (2021). A two-pronged pulmonary gene delivery strategy: a surface-modified fullerene nanoparticle and a hypotonic vehicle. Angewandte Chemie International Edition, 60(28), 15225-15229.
16. Biswas, M., Rahman, S., Tarannum, S. F., Nishanto, D., & Safwaan, M. A. (2025). Comparative analysis of attention-based, convolutional, and SSM-based models for multi-domain image classification (Doctoral dissertation, BRAC University).
17. Yavari, S., Pandya, R. N., & Furst, J. (2025). Recoseg++: Extended residual-guided cross-modal diffusion for brain tumor segmentation. arXiv preprint arXiv:2508.01058.
18. Li, W., Zhu, M., Xu, Y., Huang, M., Wang, Z., Chen, J., ... & Sun, X. (2025). SIGEL: a context-aware genomic representation learning framework for spatial genomics analysis. Genome Biology, 26(1), 287.
19. Singh, A. R., Athisayamani, S., Hariharasitaraman, S., Karim, F. K., Varela-Aldás, J., & Mostafa, S. M. (2025). Depth-Enhanced Tumor Detection Framework for Breast Histopathology Images by Integrating Adaptive Multi-Scale Fusion, Semantic Depth Calibration, and Boundary-Guided Detection. IEEE Access.
20. Xu, K., Lu, Y., Hou, S., Liu, K., Du, Y., Huang, M., ... & Sun, X. (2024). Detecting anomalous anatomic regions in spatial transcriptomics with STANDS. Nature Communications, 15(1), 8223.
21. Kasaraneni, C. K., Guttikonda, K., & Madamala, R. (2025). Multi-modality Medical (CT, MRI, Ultrasound Etc.) Image Fusion Using Machine Learning/Deep Learning. In Machine Learning and Deep Learning Modeling and Algorithms with Applications in Medical and Health Care (pp. 319-345). Cham: Springer Nature Switzerland.