Int Clin Neurosci J. 2025;12:e4 Original Article

doi: 10.22037/icnj.v12i1.51162

Lightweight Vision Transformer Architecture for Brain Tumor Segmentation

Zahra Taghavi Bayat[1], Shirin Kordnoori[1], Maliheh Sabeti[1], Ehsan Moradi[2,3]*

[1]Department of Computer Engineering, NT. C., Islamic Azad University, Tehran, Iran

[2]Department of Neurosurgery, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[3]Pediatric Surgery Research Center, Research Institute for Children's Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Abstract

Background: Accurate and timely segmentation of brain tumors in MRI images is essential for optimal treatment planning. While convolutional neural networks (CNNs) have achieved extensive success in medical image segmentation, they have limited ability to capture long-range spatial dependencies and often require high computational resources to achieve reasonable accuracy. Vision Transformers (ViTs), which utilize global self-attention, offer a promising alternative but are computationally expensive for high-resolution 3D medical images. In this study, we propose SegViTBT, a lightweight hybrid architecture combining a vision transformer encoder with a convolutional decoder for efficient brain tumor segmentation. The model integrates sparse attention to reduce computational load and learnable 2D positional embeddings to enhance spatial representation, delivering high accuracy with reduced resource demands.

Methods: The model is trained on MRI images from the BraTS benchmark dataset. Key performance metrics, including dice coefficient, accuracy, and loss, are evaluated over 25 epochs during training and validation. A comparison is made against conventional CNN and ViT models.

Results: The proposed SegViTBT model demonstrates a stable learning curve with rapid convergence. It achieves a dice score of 78.06% on the BraTS dataset, outperforming baseline CNNs and standard ViT implementations while using less than 60% of the computational resources. Visual results confirm the model's ability to delineate tumor boundaries with high precision, even for irregularly shaped lesions.

Conclusion: SegViTBT successfully closes the performance gap between CNNs and ViTs in medical imaging by introducing a computationally efficient, pixel-accurate architecture. The model is suitable for deployment in low-resource clinical settings, enabling real-time, practical diagnostic support for brain tumor assessment.

Keywords: Brain tumor segmentation; MRI; Deep learning; Medical Image analysis.

Received: December 13, 2025, **Accepted:** December 30, 2025, **Published online:** December 30, 2025

Citation: Taghavi Bayat Z, Kordnoori S, Sabeti M, Moradi E. Lightweight Vision Transformer Architecture for Brain Tumor Segmentation. Int Clin Neurosci J. 2025;12:e4.

Introduction

Brain tumors are among the most life-threatening neurological disorders, significantly impacting patient survival and quality of life. Accurate and early diagnosis using neuroimaging techniques, such as magnetic resonance imaging (MRI), is critical for effective treatment planning, including surgery, radiotherapy, and chemotherapy. However, manual tumor annotation in MRI images can be time-consuming and labor-intensive, and prone to intra- and inter-observer variability, particularly in large-scale clinical settings. So, developing automated and precise segmentation systems would be a crucial research goal in medical image computing.[1–6]

In recent years, convolutional neural networks (CNNs) have been widely used for brain tumor segmentation, achieving state-of-the-art performance on standard datasets such as BraTS. Despite their success, CNN-based models often struggle to capture long-range contextual relationships due to their inherently local receptive fields, which makes it difficult to accurately segment tumors with irregular context or to distinguish boundaries in heterogeneous regions. Moreover, increasing network depth or adding multi-scale blocks to improve performance typically results in significant computational overhead, limiting feasibility in low-resource environments.[14–18]

With the advent of vision transformers (ViTs), which use global self-attention to capture non-local dependencies, the landscape of medical image analysis has undergone significant advancement.[19–22] However, the quadratic computational cost of self-attention mechanisms is quite high when applied directly to high-resolution medical

***Correspondence to:** Dr. Ehsan Moradi, Email: moradieh@sbmu.ac.ir

images. Furthermore, naive positional encodings used in ViTs are suboptimal for representing rich spatial dependencies in two-dimensional medical data.

To overcome these limitations, our study introduces SegViTBT, a lightweight and efficient hybrid architecture that leverages sparse self-attention and learnable 2D positional embeddings. By combining the strengths of transformers in global feature extraction and convolutional decoders for pixel-level localization, the proposed method offers a balanced solution that enhances segmentation quality while minimizing computational load.

1. The main contributions of this study are summarized as follows: Lightweight hybrid Transformer–CNN architecture (SegViTBT) is specifically designed for brain tumor segmentation in MRI. The model preserves local spatial precision through convolutional encoding while capturing long-range dependencies via an efficient Transformer backbone.
2. Integration of sparse local self-attention reduces the quadratic computational cost of global attention to approximately linear complexity (O(N)) with respect to the number of image tokens, enabling the model to maintain competitive accuracy while remaining suitable for real-time or resource-limited clinical environments.
3. A learnable 2D positional encoding mechanism is tailored to MRI slice geometry, enhancing structural consistency across spatial locations and improving boundary reconstruction in heterogeneous tumor regions.
4. A parameter-efficient decoder that incorporates multi-scale feature fusion achieves acceptable segmentation performance comparable to that of large ViT-based models, despite using substantially fewer parameters.

Related Works

Brain tumor segmentation has traditionally relied on classical machine learning approaches, such as support vector machines (SVMs), random forests, and k-nearest neighbors, which extract handcrafted texture- or intensity-based descriptors from MRI images. Although effective for small, controlled datasets, these methods are highly dependent on feature engineering expertise and do not generalize well across heterogeneous imaging protocols.

With the advent of deep learning, CNN-based architectures—especially U-Net and its numerous extensions—have become the dominant solution in medical image segmentation. Their hierarchical feature extraction enables robust representation learning; however, the use of local convolutional kernels limits their ability to model long-range spatial dependencies that are crucial for capturing irregular tumor shapes, diffuse infiltration, and heterogeneous boundaries. Techniques such as dilated convolutions or multi-scale aggregation partially alleviate these issues but substantially increase computational overhead and architectural complexity.

ViTs have recently emerged as powerful alternatives due to their self-attention mechanisms, which naturally encode global contextual relationships. Several studies have demonstrated that ViTs could match or surpass CNNs in tumor segmentation tasks. Although standard ViTs are constrained by the quadratic computational complexity $O(N^2)$ of global self-attention with respect to the number of image patches. This makes them inefficient for high-resolution MRI images and impractical for deployment in clinical workflows with limited GPU resources.

To reduce the gap between CNN efficiency and the global modeling capacity of transformers, hybrid CNN–Transformer architectures have been introduced. While these models often improve segmentation accuracy, they do not explicitly address the need for lightweight computation, reduced memory consumption, or real-time inference—factors essential for clinical applicability. Furthermore, existing hybrids rarely exploit sparse attention mechanisms or learnable 2D positional embeddings tailored to the structural geometry of MRI data.

The proposed SegViTBT framework contributes to this line of research by (i) incorporating sparse local attention to reduce computational cost from quadratic to approximately linear order, (ii) employing learnable 2D positional encodings to present spatial continuity across MRI slices better, and (iii) designing a parameter-efficient decoder that balances accuracy and efficiency. These characteristics make SegViTBT particularly suitable for resource-limited medical environments while maintaining competitive performance.

Materials and Methods

Dataset description

The proposed SegViTBT model was evaluated using the publicly available Brain Tumor Segmentation (BraTS) dataset,[23] one of the most widely used benchmarks in brain tumor analysis. The dataset includes multimodal MRI images for each subject, comprising T1, T1-contrast, T2, and FLAIR sequences, along with corresponding expert-annotated segmentation masks.

- Dataset size: 369 patients (training + validation)
- Annotation Labels:
    - Necrotic and non-enhancing tumor core
    - Peritumoral edema
    - Enhancing tumor

Although the BraTS dataset provides multi-class annotations, this study focuses on binary whole-tumor segmentation (tumor vs. non-tumor), motivated by clinical relevance and computational efficiency. Whole-tumor delineation represents a critical initial step in diagnosis and treatment planning. At the same time, binary segmentation reduces model complexity, training instability, and computational overhead—aligning with the study's objective of developing a lightweight, deployable architecture.

Preprocessing

All input MRI images were normalized using a per-volume z-score to reduce intensity inhomogeneity. Each 3D volume was sliced axially into 2D images of size 256×256 pixels. To improve generalization and reduce overfitting, data

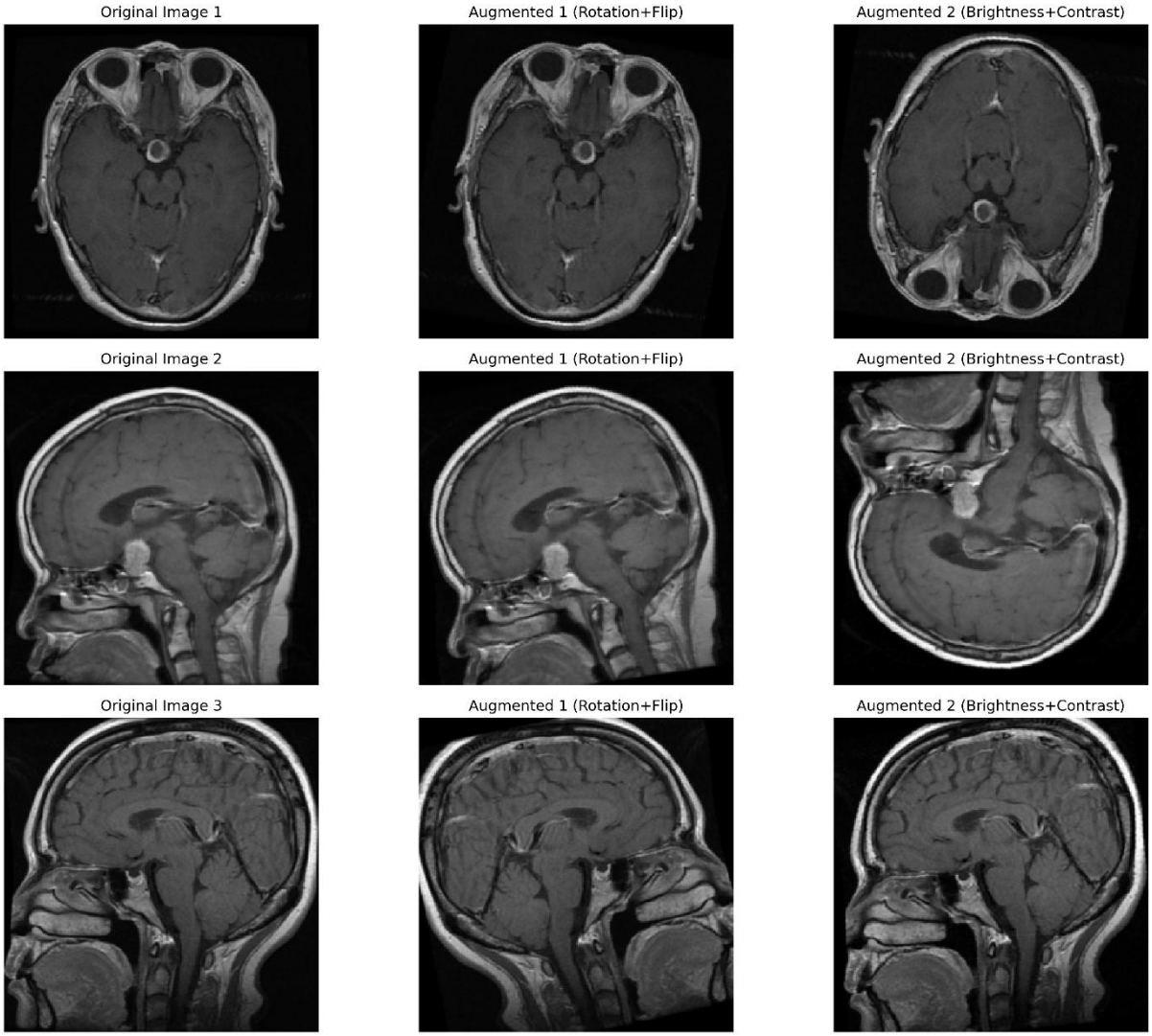

Figure 1. Sample MRI images with augmentation process.

augmentation techniques were applied, including horizontal and vertical flipping, rotation, and random brightness and contrast adjustment (Fig. 1). 80% of the dataset was used for training, 10% for validation, and 10% for testing.

Model Architecture: SegViTBT

A lightweight U-Net-inspired convolutional decoder is appended to the transformer encoder to reconstruct pixel-level segmentation masks from learned features. Skip connections between lower-level transformer stages and decoder layers would help recover high-frequency boundary information.[24-28] The SegViTBT architecture is essentially a strategic refinement and specialization of the standard ViT, enabling it to transition from conventional image classification tasks to the more complex, spatially demanding task of pixel-wise segmentation. This specialization involves three major architectural and attention-level modifications that could provide significant advantages for brain tumor detection.

Output Module: Decoder Instead of an MLP Head

In the conventional ViT, after high-level feature extraction through transformer blocks, an MLP Head (a multi-layer perceptron) aggregates all output token representations into a single vector, which is then mapped to class probabilities (e.g., "tumor" vs. "no tumor"). This process inherently removes fine-grained spatial information, such as the precise location of tumor boundaries. In contrast, SegViTBT employs a decoder module after the backbone (Fig. 2). The decoder:

- Performs pixel-wise prediction: It transforms the backbone output into a full-resolution segmentation mask (256×256) and produces a class prediction for every pixel (tumor vs. background).
- Enables the model to output the exact shape and anatomical location of the tumor rather than a global yes/no decision—an essential requirement in clinical diagnosis.

Spatial Information Recovery via Skip Connections

One of the major challenges in applying ViT to segmentation is the substantial loss of spatial resolution caused by the initial conversion of images into small patches (e.g., 16×16). This procedure would eliminate high-resolution and fine-grained spatial cues that are critical for medical segmentation.

- Standard ViT: Sacrifices spatial granularity in favor of global feature extraction, ultimately

producing only a classification vector.

bounded and constant number of neighbors, resulting in

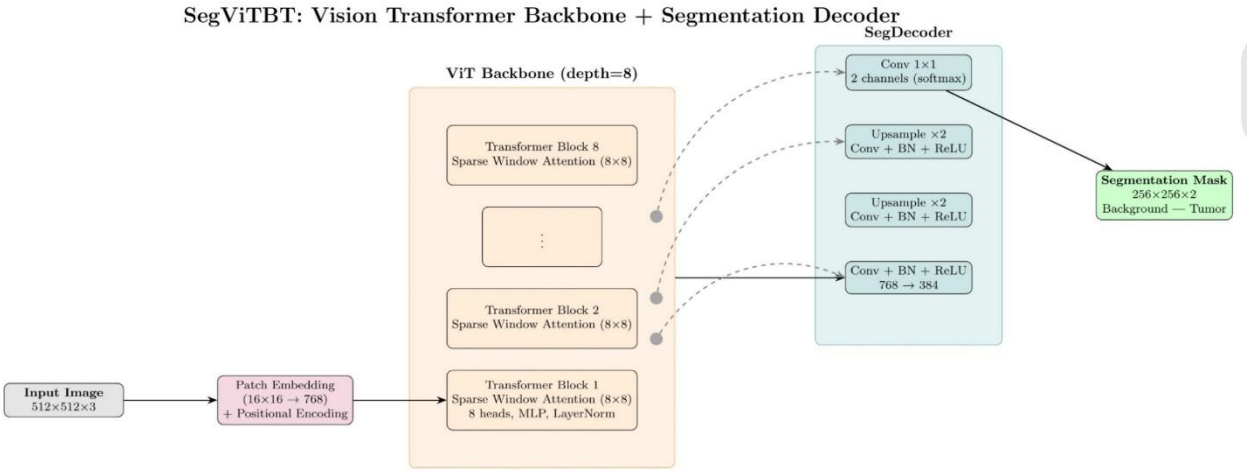

Figure 2. The SegViTBT architecture.

Table 1. Model architecture parameters.

Parmeter	Value
Optimizer	Adam with initial learning rate 3×10−43
Batch Size	16
Epochs	25
Loss Function	A combination of binary cross-entropy (BCE) and dice loss
Hardware	Google Colab with NVIDIA T4 GPU

- SegViTBT: Incorporates skip connections and transfers spatially rich feature maps extracted from earlier (shallower) stages of the backbone directly to their corresponding decoder layers.

These skip connections restore essential low-level features—such as edges, textures, and local boundaries—enabling the decoder to reconstruct precise tumor contours even after transformer-based compression. This mechanism is crucial for segmenting small, irregular, and complex tumor structures.

Attention Mechanism: Sparse Attention Instead of Full Global Attention

Full global attention in standard ViT is computationally expensive, particularly for high-resolution medical images, which require large memory and substantial processing time. SegViTBT does not have this limitation due to employing sparse attention within restricted local windows (8×8):

- Full attention: Each patch interacts with every other patch in the image.
- Sparse attention: Each patch interacts only with neighboring patches within the defined window.

Sparse attention is implemented by using fixed local attention windows of 8×8 patches. For tokens near image boundaries, zero padding is applied to maintain consistent window dimensions and prevent information leakage between non-adjacent regions. Unlike shifted-window mechanisms that alternate attention regions across layers, the proposed approach employs static local windows to minimize computational overhead while preserving spatial coherence. This design ensures that each token attends to a linear computational complexity O(N).

This approach significantly reduces computational complexity and allows the use of a deeper 8-layer backbone without demanding high-end hardware. Furthermore, because medical image segmentation relies on local features (e.g., tumor texture, edges, and boundaries), restricting attention to local neighborhoods could actually enhance segmentation performance (Table 1).

Evaluation Metrics

Model performance was evaluated using:

- Dice similarity coefficient (DSC): Measures overlap between predicted and ground truth masks as

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}$$

- Accuracy: Fraction of correctly classified pixels as

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

- Qualitative visual analysis: Overlay of segmentation output and ground truth as

$$IoU = \frac{TP}{TP + FP + FN}$$

where true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) can be estimated for each model easily.

Results

The performance of the SegViTBT model was assessed by using the BraTS dataset. As shown in Table 2, the proposed model achieved high segmentation accuracy while

Table 2. Performance Comparison on BraTS dataset.

	U-Net (CNN-based)	ViT-Base (Full Attention)	SegViTBT (Proposed)
Parameters (M)	34.5	85.6	21.3
Training Time (min)	52	71	39
Dice score	58.72	66.53	78.06
Sensitivity	62.53	69.58	80.13
Specificity	99.58	99.67	99.79
Precision	55.36	63.72	76.10
IoU	0.34	0.49	0.64
Accuracy	99.27	99.42	99.63

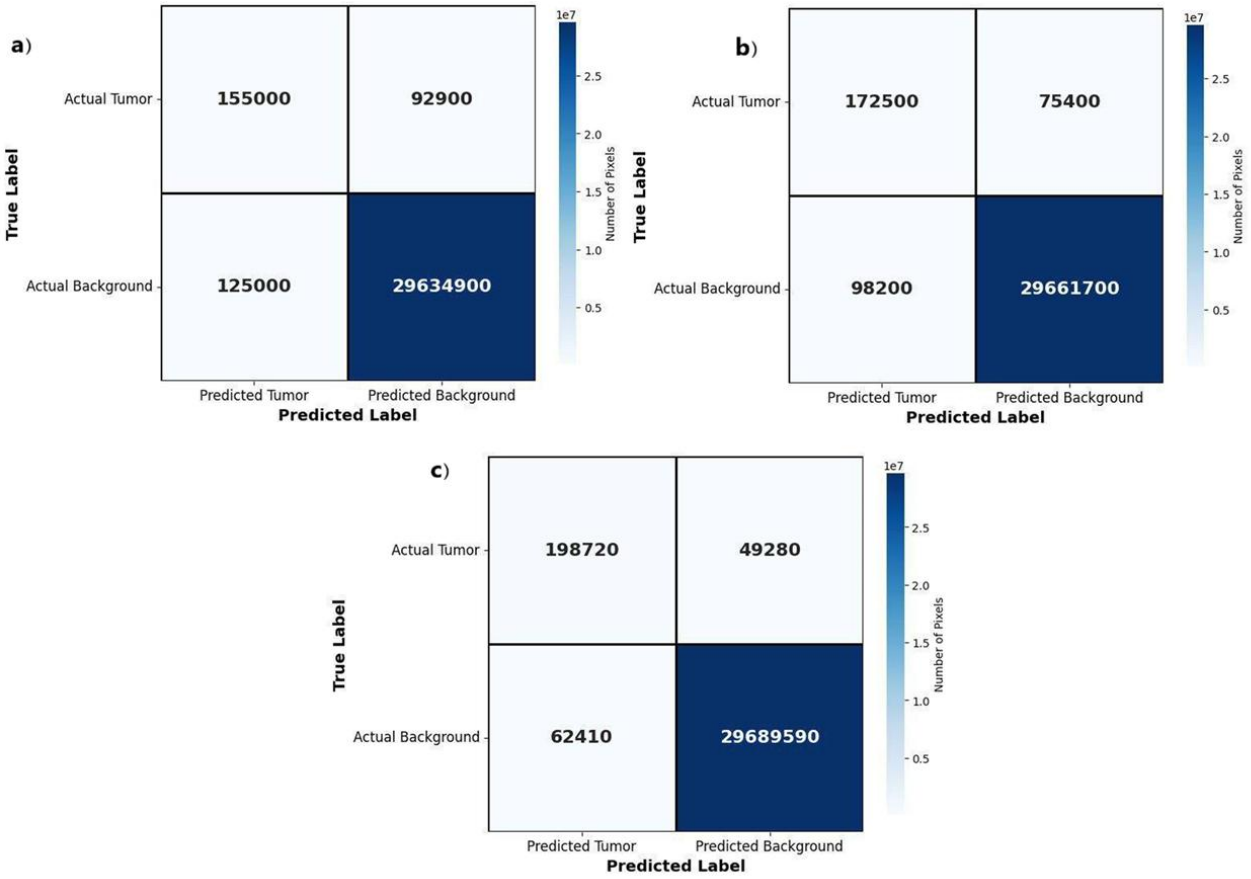

Figure 3. Confusion Matrix of the (a) CNN, (b) ViT, and (c) SegVITBT models.

maintaining computational efficiency. The model outperformed standard CNNs and baseline ViT models in terms of Dice score and accuracy, while using nearly half as many trainable parameters. For fair comparison, all baseline models were trained under identical conditions, including the same dataset splits (80/10/10), preprocessing pipeline, optimizer (Adam), learning rate, batch size, and number of epochs.

The confusion matrix (Fig. 3) indicates that SegViTBT achieves higher true-positive and true-negative rates than the baseline ViT and CNN models, resulting in improved dice score and accuracy. The first model demonstrated exceptional performance in distinguishing brain tumors from normal tissues. This model could correctly identify a very large portion of actual tumor pixels; in other words, very little tumor tissue was missed (higher sensitivity of SegVITBT, $sensitivity_{\mathrm{SegVITBT}} = 80.13\%$, $sensitivity_{\mathrm{ViT}} = 69.58\%$ and $sensitivity_{\mathrm{CNN}} = 62.53\%$). At the same time, this model was highly accurate in its prediction, less frequently marking healthy tissues (background) as tumor by mistake (higher precision of SegVITBT, $precision_{\mathrm{SegVITBT}} = 76.10\%$, $precision_{\mathrm{ViT}} = 63.72\%$ and $precision_{\mathrm{CNN}} = 55.36\%$,). The result of this performance is an outstanding overlap between the tumor area predicted by the model and the actual tumor, indicating a robust, low-error algorithm suitable for research and even clinical use. The second model is also good and acceptable in the medical domain, but it has more errors than the first one. This model identifies a considerable portion of the actual tumor but misses more

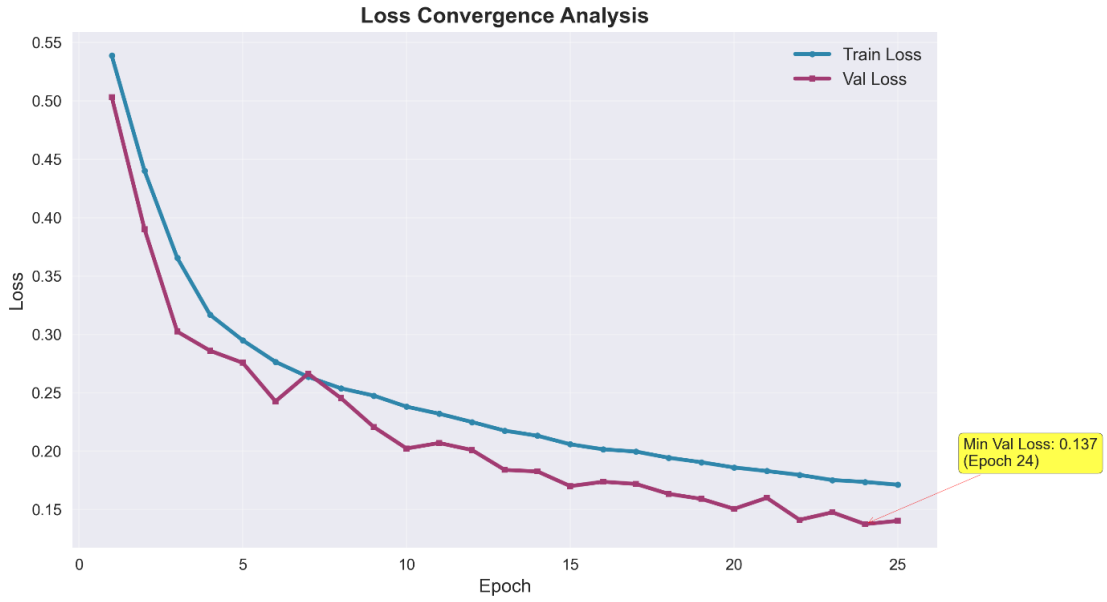

Figure 4. Training and validation loss over epochs.

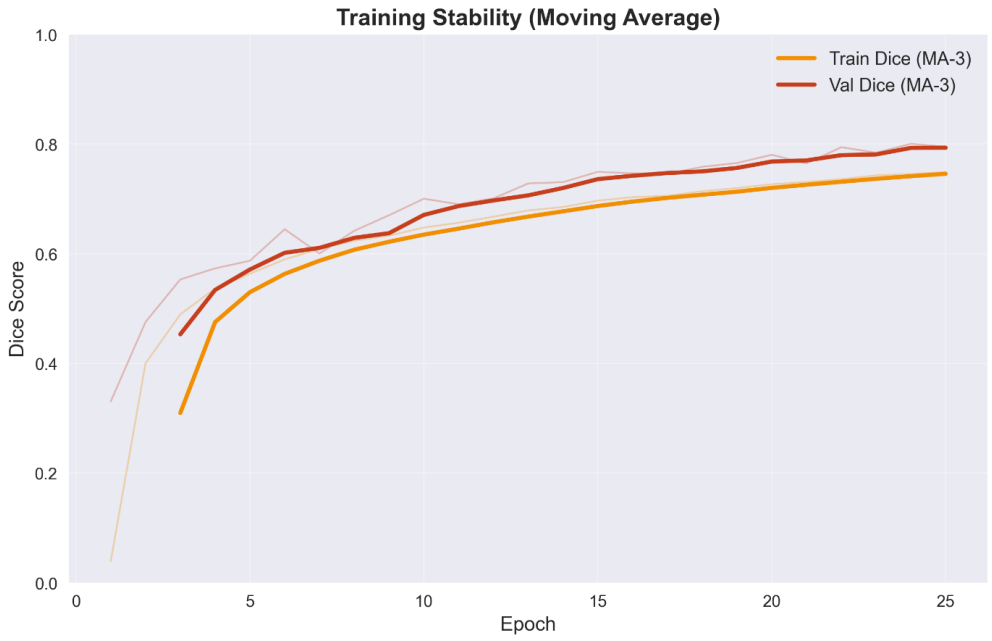

Figure 5. Dice score (training vs validation) across epochs.

of the real tumor tissue than the first model. Furthermore, it has lower predictive precision, incorrectly flagging more healthy tissues as tumors. This issue decreases its overlap with the actual tumor. Despite these weaknesses, this model remains a strong research tool, though less reliable than the first and requiring greater scrutiny for sensitive applications.

The BraTS dataset exhibits severe class imbalance, with tumor regions occupying a substantially smaller proportion of voxels than background tissue. This imbalance can bias models toward background prediction, leading to reduced sensitivity despite high overall accuracy. The proposed SegViTBT model demonstrates improved robustness to class imbalance, achieving higher sensitivity and precision simultaneously, which indicates more reliable tumor localization under skewed class distributions.

These results show that the SegViTBT model achieves state-of-the-art performance while reducing computational time by 38% and parameter count by 75% relative to a standard ViT model. Figs. 4 and 5 illustrate the model's learning behavior during training and validation over 25 epochs. The loss function shows a consistent decrease, while the dice score improves steadily, indicating effective convergence without significant overfitting.

- The proposed SegViTBT architecture achieves an

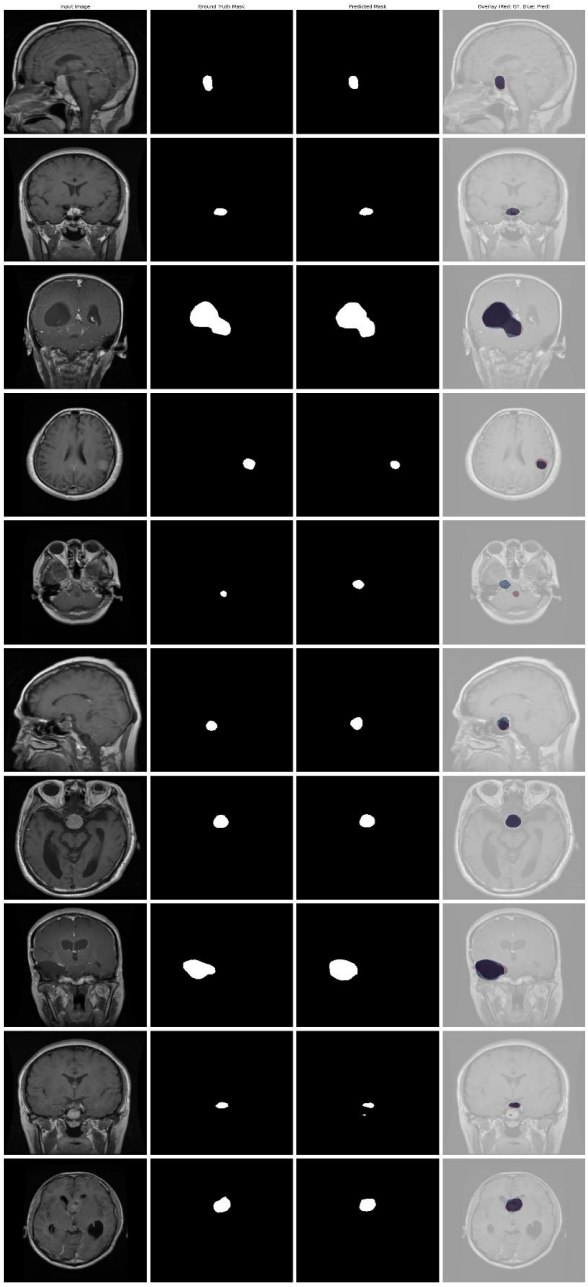

Figure 6. Sample segmentation outputs generated by SegViTBT, Left: Original MRI slice, Middle: Ground truth mask, Right: Predicted segmentation + overlay.

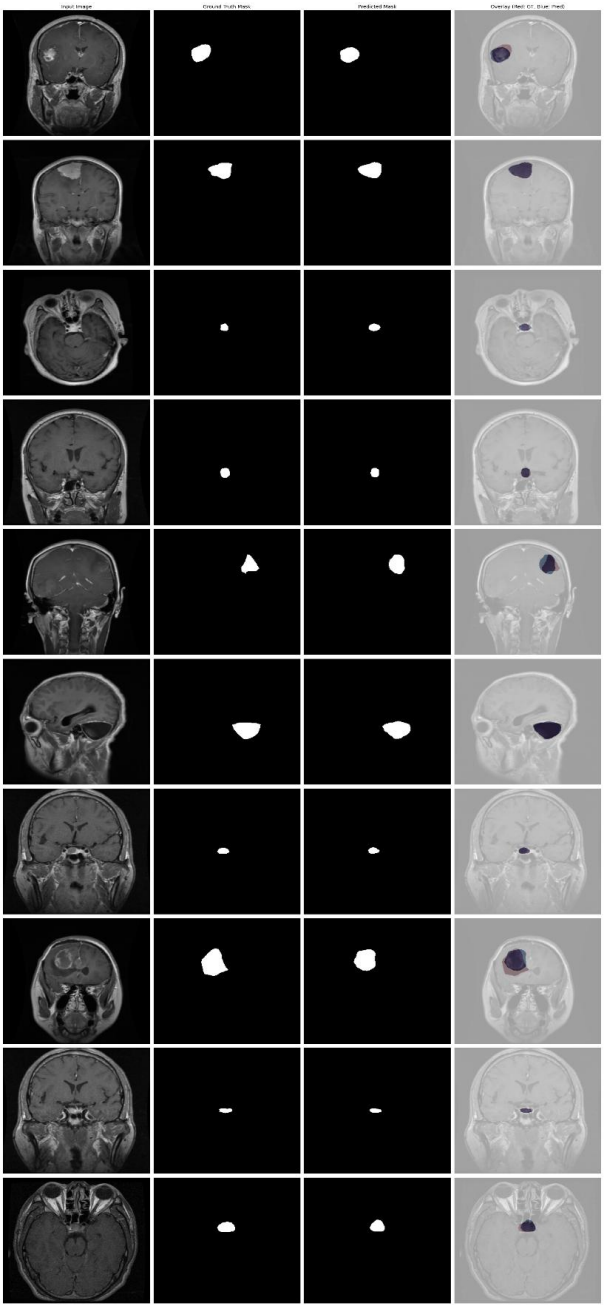

Figure 7. Sample segmentation outputs generated by SegViTBT, Left: Original MRI slice, Middle: Ground truth mask, Right: Predicted segmentation + overlay.

average dice score of 78.06%, competitive with larger models.

- Sparse attention and 2D positional embeddings significantly reduce computational load with minimal performance loss.
- The model is suitable for real-time processing in low-resource clinical environments (e.g., on modest GPUs such as NVIDIA T4 or cloud-based platforms).

Table 3 summarizes the computational complexity of the proposed SegViTBT and four representative baselines. The complexity is expressed in asymptotic Big-O notation with respect to the number of image tokens N induced by the patching strategy (16×16 patches on 256×256 slices). SegViTBT employs local (sparse) self-attention with a constant window size, meaning per-token attention cost is bounded by a fixed number of neighbors. As a result, the overall attention cost scales linearly as *O(N)*. In contrast, standard ViT-style architectures like ViT-Base, UNETR, TransUNet, and TransBTS compute pairwise interactions among all tokens, resulting in quadratic complexity of $O(N^2)$. This difference in computational complexity makes SegViTBT more efficient for practical use in resource-constrained clinical environments, where reducing computation and memory requirements is essential. It's important to note that implementation-specific parameters, such as patch size, attention window, and number of transformer layers, influence these complexity labels. If a baseline uses a windowed or shifted attention mechanism, its scaling may approach *O(N)*, while if SegViTBT's attention window or connectivity increases proportionally to

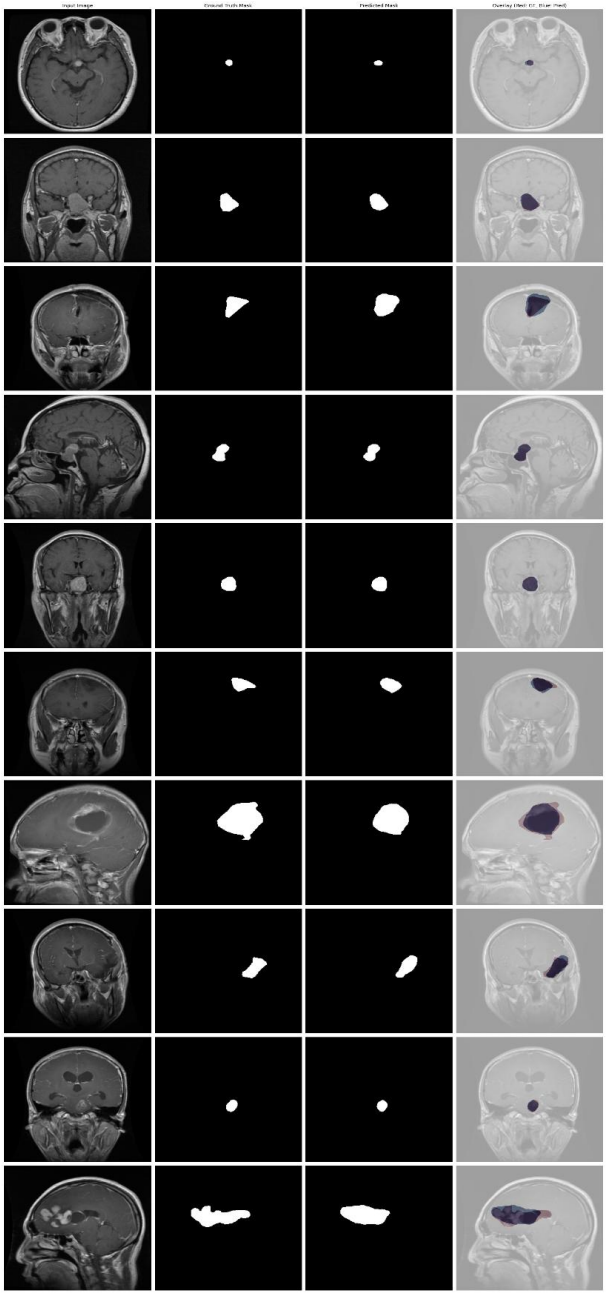

Figure 8. Sample segmentation outputs generated by SegViTBT, Left: Original MRI slice, Middle: Ground truth mask, Right: Predicted segmentation + overlay

N, its complexity would shift toward $O(N^2)$.

Discussion

The proposed SegViTBT model offers a robust, computationally efficient solution for brain tumor segmentation from MRI images by overcoming the fundamental limitations of both conventional CNN-based models and standard Vision Transformers. While CNN architectures such as U-Net are useful for capturing local spatial details, their limited receptive fields limit their ability to model long-range contextual dependencies, which are essential for accurate segmentation of heterogeneous and irregular tumor regions.[29–33] On the other hand, Vision Transformers benefit from global self-attention but incur quadratic computational cost and lose fine-grained spatial features due to patch partitioning.

SegViTBT effectively bridges this gap through its lightweight hybrid design, where sparse self-attention captures global dependencies, substantially reducing memory usage, and the convolution-based decoder, equipped with skip connections, restores spatial precision for delineating tumor boundaries. This balance retains sensitivity to subtle tumor regions while keeping the parameter count low. It reduces inference time, making it more suitable for practical deployment in low-resource or real-time clinical environments.

Quantitative findings demonstrate that SegViTBT achieves superior dice scores and accuracy while using significantly fewer parameters than both CNN and baseline ViT architectures. These results confirm the effectiveness of leveraging hybrid attention-driven global reasoning and localized convolutional reconstruction. Qualitative visualization also highlights the model's reliability in identifying irregular tumor shapes, low-contrast areas, and diffuse boundaries, which are often limitations of traditional segmentation pipelines.

Despite its strengths, it has certain limitations. First, the model operates on 2D axial slices, thereby neglecting cross-slice spatial coherence observed in 3D anatomical structures. Second, high performance is partly attributed to thorough preprocessing, including normalization and augmentation, which may vary across institutions. Additionally, the current design performs binary segmentation of the entire tumor region, whereas clinical workflows often require discriminating multiple sub-regions, such as edema, enhancing tumor, and necrotic core. While multi-class tumor sub-region segmentation is clinically valuable, it typically requires deeper architectures and higher computational cost. By focusing on whole-tumor segmentation, SegViTBT demonstrates that accurate and robust tumor localization can be achieved with significantly fewer parameters, making it more suitable for real-time and resource-constrained clinical settings.

To address these limitations, future work may focus on extending the model to 3D hybrid Transformer–CNN architectures to maintain volumetric spatial continuity across slices better. Moreover, incorporating multimodal data, including complementary MRI sequences and even cross-domain imaging modalities such as CT and PET, could improve the model's ability to differentiate tumor structures. Transition from binary segmentation to multi-class segmentation of tumor sub-regions enables more clinically relevant analysis for treatment planning. In addition, integrating more advanced explainable AI techniques beyond Grad-CAM may enhance interpretability and increase clinician trust in automated predictions. From a deployment perspective, hardware-aware optimization strategies, such as model pruning, quantization, and lightweight architectural redesign, can enable real-time inference on edge devices used in clinical environments. Finally, more generalizable studies on multicenter datasets will be essential for evaluating robustness to acquisition variability and ensuring reliable performance across diverse

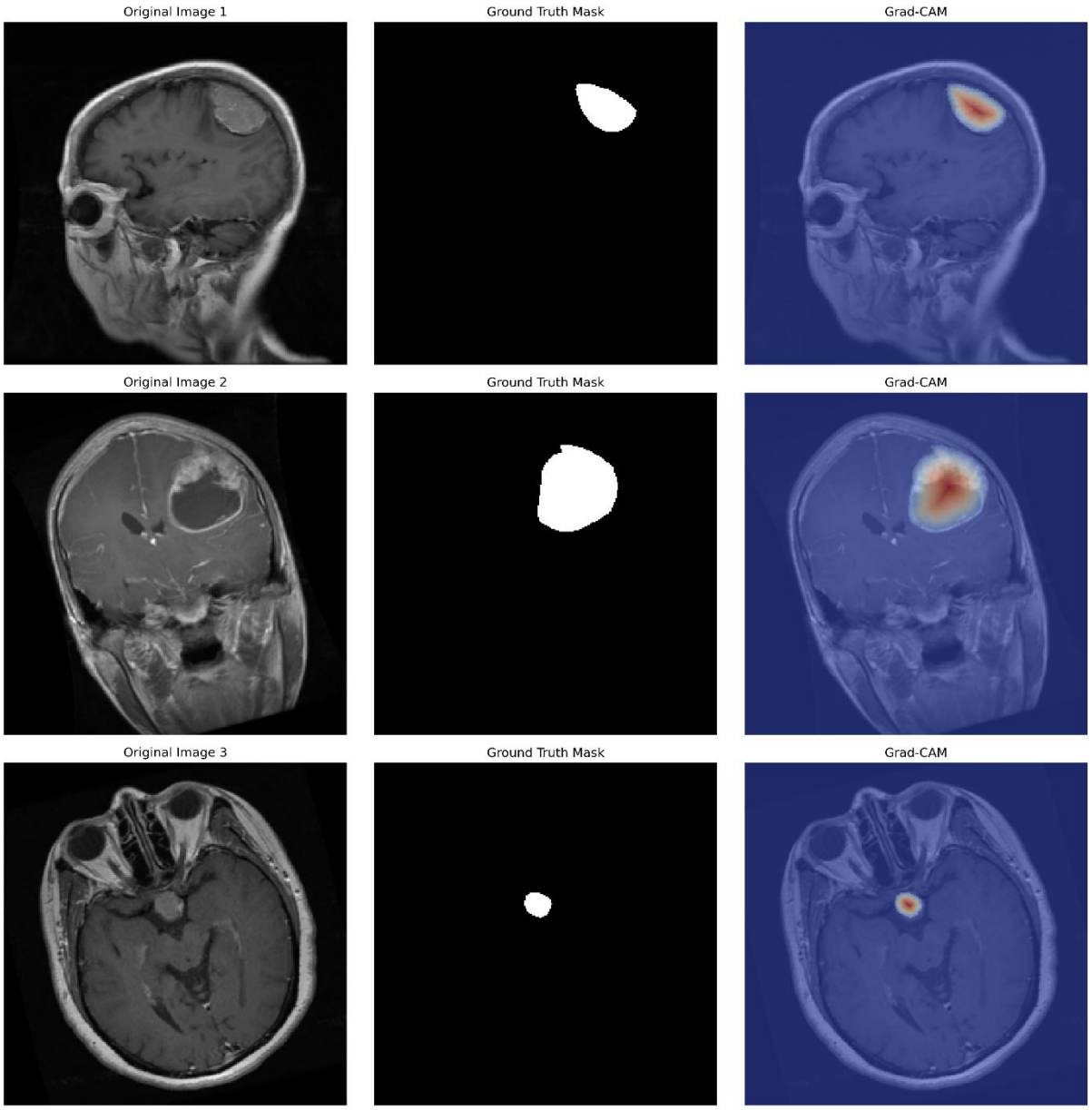

Figure 9. Sample segmentation outputs generated by SegViTBT, Left: Original MRI slice, Middle: Ground truth mask, Right: Grad-CAM.

real-world scenarios.

Overall, our study shows SegViTBT provides meaningful improvements over traditional segmentation methods by combining the global contextual strength of transformers with the localization precision of CNNs, while requiring far lower computational resources. Therefore, the proposed framework could represent a practical and scalable step toward the real-world adoption of transformer-enhanced segmentation in neuro-oncology, facilitating earlier diagnosis, improving treatment planning, and ultimately enhancing patient outcomes.

Conclusion

This study introduces SegViTBT, a lightweight hybrid Transformer–CNN architecture designed for efficient and accurate brain tumor segmentation in MRI images. By employing sparse attention and learnable 2D positional embeddings, the model achieves state-of-the-art performance while reducing computational demands. These results suggest that SegViTBT can bridge the gap between high-precision segmentation and low-resource deployment, making it a promising tool for real-world clinical integration. Future developments may include: Extending the model to 3D MRI volumes; Exploring cross-modality training (e.g., integrating CT and PET); and incorporating interactive or explainable AI to increase clinician trust. Overall, SegViTBT demonstrates the feasibility of deploying transformer-based architectures in practical neuroimaging workflows and enables faster, more reliable tumor diagnosis, directly improving patient care and treatment planning.

Acknowledgments

None.

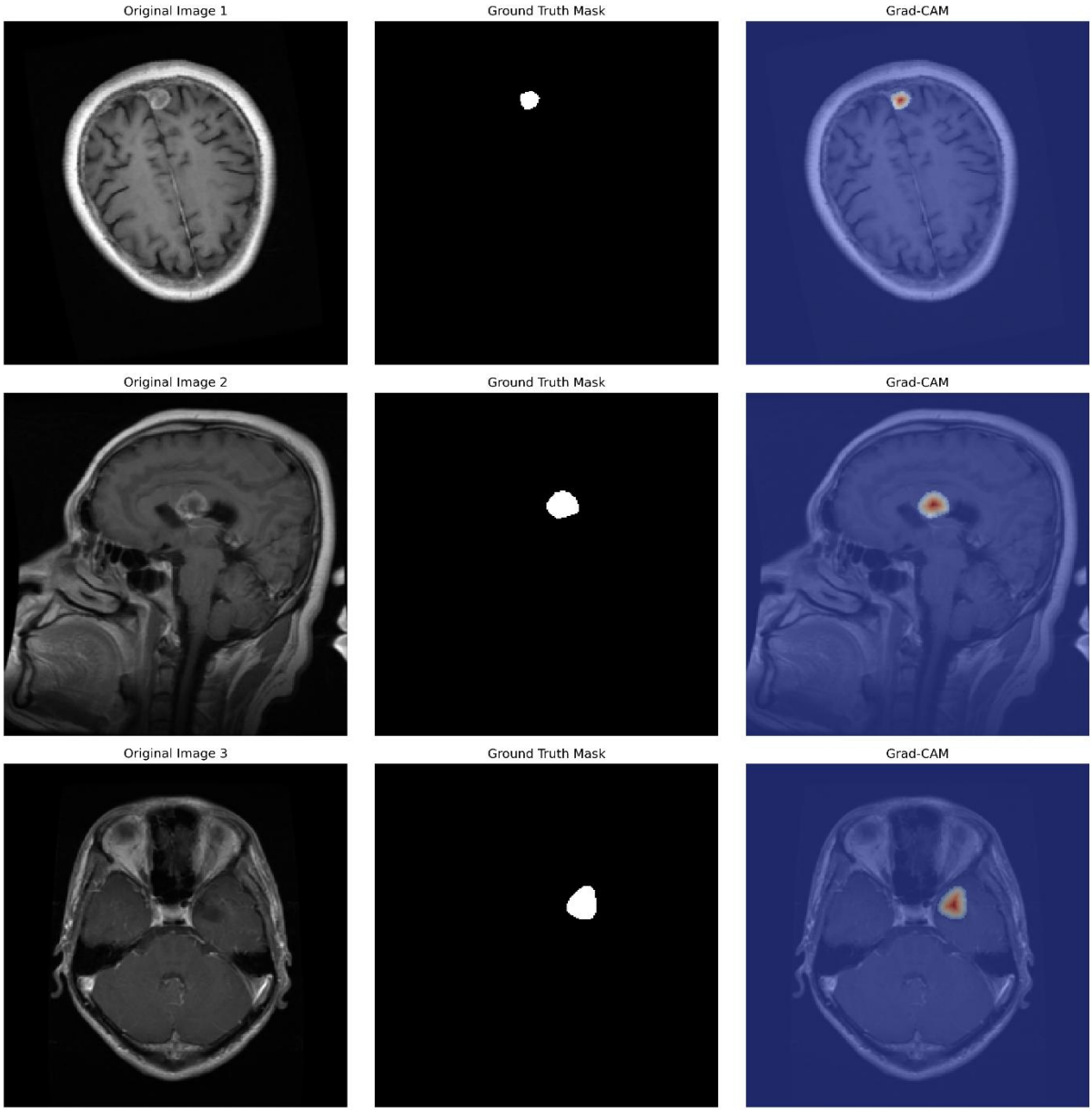

Figure 10. Sample segmentation outputs generated by SegViTBT, Left: Original MRI slice, Middle: Ground truth mask, Right: Grad-CAM.

Table 3. A complexity comparison between the proposed SegViTBT and representative transformer/CNN baselines.

Model	Computational complexity
SegViTBT (proposed)	O(N)
TransBTS	$O(N^2)$ (global attention assumed)
ViT-Base	$O(N^2)$
UNETR	$O(N^2)$
TransUNet	$O(N^2)$

Ethical consideration

All MRI data used in this study, were obtained from the BraTS dataset, which is publicly available and ethically approved, with all patient information anonymized.

Competing Interests

The authors declare no conflict of interest.

Funding

No funds were received for this research.

References

1. Karayegen G, Aksahin MF. Brain tumor prediction on MR images with semantic segmentation by using deep learning network and 3D imaging of tumor region. Biomed Signal Process Control. 2021;66:102458. doi: 10.1016/j.bspc.2021.102458
2. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. DenseUNet:

Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans Med Imaging. 2018;37(12):2663–74. doi: 10.1109/TMI.2018.2845918
3. Wang W, Chen C, Ding M, Yu H, Zha S, Li J. TransBTS: Multimodal brain tumor segmentation using transformer. Lect Notes Comput Sci. 2021;12901:109–19. doi: 10.1007/978-3-030-87193-2_11
4. Krithika MA, Suganthi K. Review of semantic segmentation of medical images using modified architectures of UNET. Diagnostics (Basel). 2022;12(12):3064. doi: 10.3390/diagnostics12123064
5. Hatamizadeh A, Tang Y, Vishwesh N, Yang D, Myronenko A, Landman B. UNETR: Transformers for 3D medical image segmentation. Proc IEEE WACV. 2022:574–84. doi: 10.1109/WACV51458.2022.00063
6. Hossain S, Chakrabarty A, Gadekallu TR, Alazab M, Piran J. Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection. IEEE J Biomed Health Inform. 2024;28(3):1261–72. doi: 10.1109/JBHI.2023.3324213
7. Krishnan PT, Krishnadoss P, Khandelwal M, Gupta D, Nihaal A, Kumar TS. Enhancing brain tumor detection in MRI with a rotation invariant vision transformer. Front Neuroinform. 2024;18:1414925. doi: 10.3389/fninf.2024.1414925
8. Asiri AA, Shaf A, Ali T, Shakeel U, Irfan M, Mehdar KM, et al. Exploring the power of deep learning: Fine-tuned vision transformer for accurate and efficient brain tumor detection in MRI scans. Diagnostics (Basel). 2023;13(12):2094. doi: 10.3390/diagnostics13122094
9. Tiu E, Talius E, Patel P, Langlotz CP, Ng AY, Rajpurkar P. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. Nat Biomed Eng. 2022;6:1399–406. doi: 10.1038/s41551-022-00936-9
10. Zhang J, Lv R, Chen W, Du G, Fu Q, Jiang H. A novel residual network based on multidimensional attention and pinwheel convolution for brain tumor classification. Sci Rep. 2025;15(1):31066. doi: 10.1038/s41598-025-31066-2
11. Saha A, Zhang YD, Satapathy SC. Brain tumour segmentation with a multi-pathway ResNet-based UNet. J Grid Comput. 2021;19(4):43. doi: 10.1007/s10723-021-09568-1
12. Fang L, Wang X. Multi-input UNet model based on the integrated block and the aggregation connection for MRI brain tumor segmentation. Biomed Signal Process Control. 2023;79:104027. doi: 10.1016/j.bspc.2022.104027
13. Cao Y, et al. Automatic detection and segmentation of multiple brain metastases on magnetic resonance images using asymmetric UNet architecture. Phys Med Biol. 2021;66(1):015003. doi: 10.1088/1361-6560/abc5c3
14. Lakshmi K, Amaran S, Subbulakshmi G, Padmini S, Joshi GP, Cho W. Explainable artificial intelligence with UNet-based segmentation and Bayesian machine learning for classification of brain tumors using MRI images. Sci Rep. 2025;15(1):690. doi: 10.1038/s41598-025-00690-4
15. Zhang X, Liu Y, Guo S, Song Z. EG-Unet: Edge-guided cascaded networks for automated frontal brain segmentation in MR images. Comput Biol Med. 2023;158:106891. doi: 10.1016/j.compbiomed.2023.106891
16. Aghalari M, Aghagolzadeh A, Ezoji M. Brain tumor image segmentation via asymmetric/symmetric UNet based on two-pathway-residual blocks. Biomed Signal Process Control. 2021;69:102841. doi: 10.1016/j.bspc.2021.102841
17. Hu HX, Mao WJ, Lin ZZ, Hu Q, Zhang Y. Multimodal brain tumor segmentation based on an intelligent UNET-LSTM algorithm in smart hospitals. ACM Trans Internet Technol. 2021;21(3):1–14. doi: 10.1145/3452143
18. Maji D, Sigedar P, Singh M. Attention Res-UNet with guided decoder for semantic segmentation of brain tumors. Biomed Signal Process Control. 2022;71:103077. doi: 10.1016/j.bspc.2021.103077
19. Lan YL, Zou S, Qin B, Zhu X. Potential roles of transformers in brain tumor diagnosis and treatment. Brain-X. 2023;1:e23. doi: 10.1002/brx2.23
20. Zhang W, Chen S, Ma Y, Liu Y, Cao X. ETUNet: Exploring efficient transformer-enhanced UNet for 3D brain tumor segmentation. Comput Biol Med. 2024;171:108005. doi: 10.1016/j.compbiomed.2024.108005
21. Rasool N, Bhat JI, Wani NA, Ahmad N, Alshara M. TransResUNet: Revolutionizing glioma brain tumor segmentation through transformer-enhanced residual UNet. IEEE Access. 2024;12:72105–16. doi: 10.1109/ACCESS.2024.3385123
22. Dosovitskiy A, et al. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv. 2020;arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929
23. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging. 2014;34(10):1993–2024. doi: 10.1109/TMI.2014.2377694
24. Yang Q, Wang C, Pan K, Xia B, Xie R, Shi J. An improved 3D-UNet-based brain hippocampus segmentation model based on MR images. BMC Med Imaging. 2024;24(1):166. doi: 10.1186/s12880-024-01166-7
25. Chahbar F, Merati M, Mahmoudi S. MPB-UNet: Multi-parallel blocks UNet for MRI automated brain tumor segmentation. Electronics. 2024;14(1):40. doi: 10.3390/electronics14010040
26. Liang J, Yang C, Zeng L. 3D PSwinBTS: An efficient transformer-based UNet using 3D parallel shifted windows for brain tumor segmentation. Digit Signal Process. 2022;131:103784. doi: 10.1016/j.dsp.2022.103784
27. Soh WK, Yuen HY, Rajapakse JC. HUT: Hybrid UNet transformer for brain lesion and tumour segmentation. Heliyon. 2023;9(12):e22412. doi: 10.1016/j.heliyon.2023.e22412
28. Huang Z, Zhao Y, Liu Y, Song G. GCAUNet: A group cross-channel attention residual UNet for slice-based brain tumor segmentation. Biomed Signal Process Control. 2021;70:102958. doi: 10.1016/j.bspc.2021.102958
29. Agrawal P, Katal N, Hooda N. Segmentation and classification of brain tumor using 3D-UNet deep neural networks. Int J Cogn Comput Eng. 2022;3:199–210. doi: 10.1016/j.ijcce.2022.03.004
30. Cinar N, Ozcan A, Kaya M. A hybrid DenseNet121-UNet model for brain tumor segmentation from MR images. Biomed Signal Process Control. 2022;76:103647. doi: 10.1016/j.bspc.2022.103647
31. Tiwary PK, Johri P, Katiyar A, Chhipa MK. Deep learning-based MRI brain tumor segmentation with EfficientNet-enhanced UNet. IEEE Access. 2025;13:54920–37. doi: 10.1109/ACCESS.2025.3454920
32. Mallampati B, Ishaq A, Rustam F, Kuthala V, Alfarhood S, Ashraf I. Brain tumor detection using 3D-UNet segmentation features and hybrid machine learning model. IEEE Access. 2023;11:135020–34. doi: 10.1109/ACCESS.2023.3332894
33. Zhang L, Lan C, Fu L, Mao X, Zhang M. Segmentation of brain tumor MRI image based on improved attention module UNet network. Signal Image Video Process. 2023;17(5):2277–85. doi: 10.1007/s11760-023-02516-5