GMR
The Original

A COMPARATIVE EXPERIMENTAL ANALYSIS OF DEEP LEARNING MODELS FOR BRAIN TUMOR IDENTIFICATION AND PREDICTION USING MRI IMAGES

Parimal A. Trivedi[1], Ishbir Singh[2], Sheetal Pandya[3]

[1]Assistant professor.,Computer science Indus University
[2]Director, Computer engineering, Indus University
[3]Assistant professor, Computer engineering, Indus University

ABSTRACT:
Segmentation of brain tumors from magnetic resonance imaging (MRI) remains difficult due to significant variations in tumor regions in terms of size, shape, tissue type and location. In this study, a multimodal deep learning framework named DRA-MSCNet was designed and tested, which included denoising-aware preprocessing, residual learning, multiscale feature extraction and attention-based refinement. The BraTS 2023 Pediatric MRI dataset (99 eligible cases) was used, where 69 cases were used for training, 15 cases for validation, and 15 cases for a locked test set. Whole Tumor (WT), Tumor Core (TC) and Enhancing Tumor (ET) were segmented together from four different MRI modalities: T1n, T1c, T2w and T2-FLAIR. The proposed model achieved Dice scores of 0.7636, 0.7233, and 0.4422 for WT, TC, and ET, respectively, with a mean Dice of 0.6430. Structured pruning was then used to reduce the number of computations. The chosen 10% compressed model was able to reduce the parameters from 2,207,137 to 1,968,491 and the model size from 8.476 MB to 7.567 MB with a mean Dice of 0.6376. There was no significant difference in the original and compressed models (adjusted $p = 0.0637$) and Mean HD95 improved from 37.8670 mm to 31.4338 mm. Results show that moderate pruning can decrease complexity of computations without sacrificing most of the segmentation performance.

KEYWORDS: Brain tumor segmentation; Multimodal MRI; DRA-MSCNet; Deep learning; Model compression

INTRODUCTION

Brain tumors come in many forms, vary in size and location, and are a broad range of neurological diseases, so they are complex conditions. The delineation in paediatric cases is especially difficult as the appearance and the subdivision between subregions may vary significantly from one patient to another. Magnetic resonance imaging (MRI) is commonly used for tumour assessment due to its good soft tissue contrast with complementary imaging sequences. Multiparametric MRI (T1, contrast-enhanced T1, T2, and FLAIR) helps better visualize areas of the tumor and its surrounding abnormalities. The clinically relevant regions like Whole Tumor (WT), Tumor Core (TC) and Enhancing Tumor (ET) have been standardized for automated brain tumor segmentation in the BraTS benchmark [1]. Reliable segmentation methods have also been developed and evaluated using expert annotated glioma datasets [2].

In brain tumour segmentation, three-dimensional approaches are especially relevant as they take into account the inter-slice spatial information and maintain volumetric context. Multimodal techniques have indicated that complementary MRI sequences have an advantage for tumor representation. Multimodal approaches for brain tumor segmentation were explored in MM-UNet [4] and self-supervised hybrid fusion approaches showed the benefit of fusion of information from imaging sequences [5]. Tumor boundary localization and discrimination has also been enhanced by pixel-level and feature-level fusion strategies [6]. But segmentation is difficult if large diffuse WT regions, smaller TC regions and highly heterogeneous ET regions are to be identified within a single framework, and yet keep the computation of segmentation manageable.

Due to the high variability in size, shape, intensity and spatial extent of tumour subregions, automated segmentation of paediatric brain tumour is still a difficult task. Generally, the size and delineation of ET and TC are smaller than that of WT, and ET is susceptible to severe class imbalance and possibly smaller size and irregular shape, along with fragmentation. These attributes make it hard for consistent segmentation across all three regions.

Another challenge is the computational complexity of 3D deep learning architectures. More depth networks would be able to learn rich context information, but are likely to be very expensive from memory, parameters, and computational operations. Multi-scale residual architectures are able to deliver better representation of tumors at varying spatial resolutions [7] and attention mechanisms are able to better select features by highlighting relevant spatial and semantic information [8]. But, the use of residual learning and multi-scale processing along with attention might further complicate the architecture.

To reduce the computational requirements, lightweight three dimensional attention networks have been thus proposed to keep steady segmentation performance [9]. Broad receptive field, residual learning, and attention mechanisms have been others combined in dilated multi-scale residual attention models to segment volumetric tumors [10]. However, the challenge of accurate WT, TC and ET segmentation along with computational efficiency is not yet solved. Model

complexity, model generalization, efficiency, and accurate delineation of subregions are recurring issues in multimodal MRI segmentation as found in the recent literature [11]. In the present study, we tackle this problem by designing a denoising-aware pre-processing network, DRA-MSCNet, which integrates denoising-aware pre-processing, residual learning, multi-scale feature extraction, attention-based refinement, and structured pruning.

In this study, we particularly address the problem of 3D segmentation of paediatric brain tumours from denoised four channel multimodal MRI volumes. The proposed DRA-MSCNet segments T1n, T1c, T2w and T2-FLAIR sequences to segment the WT, TC and ET regions. Structured model compression is also assessed to assess the potential of reducing the computation complexity without compromising the segmentation performance.

The experimental analysis only involves one publicly available pediatric MRI data set and static patient-level divisions for training, validation and locked test. Segmentation performance is assessed by overlap and boundary-based measures, as well as computational measures like the number of parameters, model size, memory consumption, MACs, FLOPs, and inference latency. It is possible to compare only the original DRA-MSCNet with the compressed DRA-MSCNet configuration selected. There is no controlled comparison with other additional independent baseline architectures or external dataset validation. The study does not cover clinical deployment, prospective testing or diagnostic decision-making.

The study helps in segmenting brain tumors in 3D with the integration of multimodal MRI information with the residual multi-scale attention-based feature learning which makes it computationally efficient. To retain relevant tumor information across each heterogeneous region (WT, TC, ET), while retaining a compact architecture, a DRA-MSCNet is designed.

Another contribution is the evaluation of structured pruning following model development. The study does not just measure segmentation accuracy, but also considers the tradeoff between predictive performance and computational needs. A comparison between the original and compressed DRA-MSCNet configuration will show if parameters, storage requirement, memory usage and number of computation operations can be reduced without statistical loss of segmentation performance. This is a combined accuracy-efficiency evaluation which is more deployment oriented evaluation of proposed framework.

Research Objectives

The study aims to:

- Develop DRA-MSCNet for three-dimensional segmentation of Whole Tumor, Tumor Core, and Enhancing Tumor from multimodal pediatric MRI.
- Evaluate the segmentation performance of DRA-MSCNet using regional overlap and boundary-based metrics.
- Investigate structured pruning for reducing the computational complexity of DRA-MSCNet.
- Statistically compare the segmentation performance of the original and compressed DRA-MSCNet configurations using the locked test set.

2. LITERATURE REVIEW

Large variations in intra-tumoral heterogeneity should be well represented for brain tumor segmentation. Whole Tumor (WT) typically has a relatively large spatial extent, while Tumor Core (TC) and Enhancing Tumor (ET) can be small, irregular and discontinuous. Thus, one receptive field size could be too small to capture the whole tumor context and too large to capture the fine regional boundaries. To solve this problem, DMRA U-Net introduced three-dimensional dilated multi-scale residual processing with attention, which can be used to enhance the receptive field while preserving the spatial feature resolution [12]. This design trend shows the significance of combining contextual feature extraction with propagation of residual information in volumetric segmentation.

Computational complexity is still one of the outstanding issues as 3D convolutions significantly add to the number of parameters, memory usage and floating-point operations. LATUP-Net presented a lightweight 3D attention U-Net with parallel convolutions, which maintains multi-scale representation, yet with limited architecture overhead [13]. Inception processing and guided attention for multimodal MRI segmentation were then combined with residual learning, demonstrating the importance of selectively focusing attention on tumor-relevant representations with effective gradient propagation [14].

Interactions between multiple views, as well as between multiple scales, have also been investigated for enhancing representation through complementary pieces of information. To better integrate heterogeneous tumor properties, multi-view attention with multi-scale feature interaction was integrated in MVSI-Net [15]. Likewise, MResU-Net fused residual learning and multi-scale representation for multimodal MRI segmentation, enabling the reuse of the hierarchy of features and the representation of the different-scale structures [16]. Such architectures suggest that residual pathways are not exclusive mechanisms to modelling at multiple scales, but instead complementary.

The recent attention has been directed towards the reduction of the computation load of volumetric segmentation. To enhance the computational efficiency while maintaining the learning of contextual features, TDPC-Net added a lightweight multi-scale processing to the three-dimensional attention mechanism. Preetha et al. combined multi-scale attention with EfficientNetB4 encoder which is another approach to enhance the feature discrimination using efficient hierarchical representation [18]. In addition, Chen et al. also used a multiscale attention U-Net to enhance the spatially relevant MRI features at various network scales [19].

The fusion of multimodal features is also crucial since various MRI sequences offer complementary anatomical and pathological information. The DPAFNet used the dual-path attention fusion to fuse the multimodal representations and retain the discriminative information in each pathway [20]. RDAU-Net incorporated the three methods of residual

convolution, DFP and CBAM-based refinement, further highlighting the importance of channel-spatial attention in filtering out irrelevant responses and enhancing tumor-related features [21].
Attention is no longer the only thing that is designed for efficiency. FFLUNet presented feature fusion in the lightweight U-Net structure, which adopted the strategy that less complex architectures can be coupled with appropriate feature aggregation [22]. To represent the tumor information at progressively different semantic and spatial levels, HMSA-Net used hierarchical multi-scale attention for multimodal MRI segmentation. A hybrid multi-resolution U-Net also integrated residual dual attention and deep supervision, and proved the effectiveness of multi-resolution learning in retaining detailed structures and boosting optimization with each layer of the network [24].
These architectural advances need to be evaluated with standardized datasets. The BraTS benchmark set the multimodal MRI-based evaluation of clinically-relevant tumor regions and gave a common framework for the evaluation of WT, TC, and ET segmentation [25]. Expert-annotated glioma MRI collections also facilitated reproducible segmentation research by offering voxel-level tumor annotations and radiomic information for quantitative analysis [26].
So far, the studies show that residual learning, multi-scale feature extraction, multimodal fusion, attention refinement, lightweight decoding and deep supervision can help to enhance brain tumor segmentation from various aspects. Yet, when this happens together, one should consider the important accuracy/efficiency tradeoff. Growing of the receptive-field diversity, the attention capacity and the architectural depth can lead to more parameters, more memory consumption, more multiply-accumulate operations and higher inference cost. On the other hand, too strong architectural simplification can decrease the computational cost at the expense of localization accuracy of small and heterogeneous tumor area, especially for ET.
DRA-MSCNet sits in the middle of this undecidable trade-off. The proposed framework combines the four-channel MRI processing with denoising-aware, residual feature propagation, multi-scale contextual extraction, attention guided refinement, and lightweight 3D decoding. More importantly, the architecture is then structurally pruned and assessed based on the regional segmentation performance and computational-efficiency metrics. This design does not just consider correctness/accuracy in building development, but explicitly the question if the complexity of the model can be reduced without losing the WT, TC and ET segmentation performance. The framework hence aims at a computationally efficient multimodal segmentation approach that can be further optimized towards deployment.

3. METHODOLOGY

3.1 Research Design

For three-dimensional pediatric brain tumor segmentation from multimodal magnetic resonance imaging (MRI), a quantitative controlled experimental design was used for the development and evaluation of DRA-MSCNet.
This workflow involved the acquisition of datasets, quality control, modality-specific pre-processing, denoising, patient-wise data partitioning, model development, structured compression, performance evaluation and paired statistical analysis. For segmentation, Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET) were targeted. DRA-MSCNet incorporated residual feature learning, extraction of multi-scale contextual information, attention-based feature refinement, and a lightweight decoder. Structured pruning was then taken to examine the segmentation performance versus the computational efficiency. The patient-level partition was reproducible with a seed of 42. Out of 99 eligible cases, 69 cases were used for training, 15 for validation and 15 for testing. The test subset was used for final evaluation only, and not in the training or during validation-based model selection. To allow for paired comparison it was evaluated how the original and compressed DRA-MSCNet configurations performed with the same test patients.

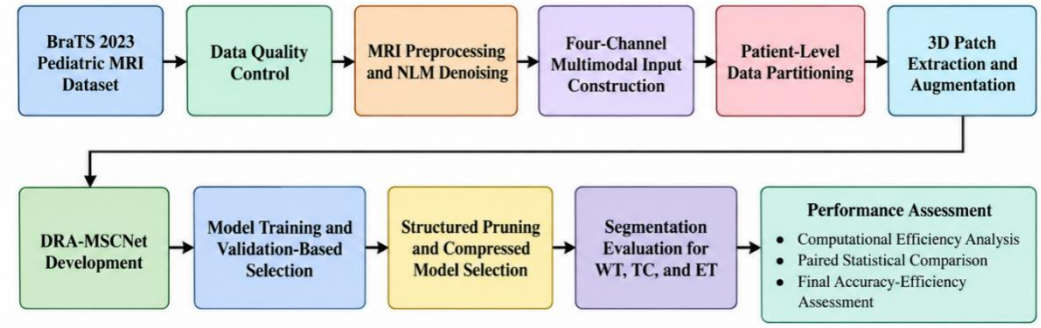

Figure 1. Overall workflow of the proposed DRA-MSCNet framework.

3.2 Data Collection and Preprocessing

The BraTS 2023 Pediatric MRI dataset was downloaded from MedOtter repository on Hugging Face [27]. All cases were provided with four MRI modalities (native T1-weighted MRI (T1n); post-contrast T1-weighted MRI (T2w); T2-weighted MRI (T2w); and T2-FLAIR) as well as voxel-level tumour annotations in NIfTI format. Modality completeness, dimensions of images, affine information, finite voxel values, availability of segmentation-masks and validity of tumor-labels were tested in quality-control.
All NIfTI volumes were processed with NiBabel and transformed into canonical RAS orientation. All the source images were of the same spatial resolution and no resampling was needed. Normalization per modality was performed for each MRI modality. The foreground voxels' intensities were clipped between 0.5th and 99.5th percentiles, while the background voxels were excluded from the intensity estimation. Then, z-score normalization was used as

$$x' = \frac{x - \mu}{\sigma + \epsilon},$$

where xis the voxel intensity, μand σdenote the mean and standard deviation of clipped foreground intensities, and $\epsilon = 10^{-8}$prevents numerical instability.

Slice-wise two-dimensional Non-Local Means denoising was performed to reduce imaging noise while maintaining the structural information and the tumor-boundary information. The four modalities were preprocessed and combined as a four channel volumetric input. The 96 × 96 × 96 voxel patches were used for training. Tumor aware sampling gave a probability of 0.70 to patches with tumor voxels. Data augmentation consisted of three independent random flips (with probability 0.5) along the three spatial axes, and random in-plane rotations of 0°, 90°, 180°, and 270°. These procedures enhanced the tumor representation and the spatial variability in the model optimization procedure.

3.3 Proposed Model: DRA-MSCNet

DRA-MSCNet was designed as a four-level 3D encoder-decoder with residual learning, multi-scale contextual extraction, channel-spatial attention, attention gated skip connections, and lightweight decoding. Four modalities of MRI data were provided as input channels and the segmentation head produced the four classes at the voxel level. The encoder used channel widths of 16, 32, 64, and 128. The first stage kept the 96 × 96 × 96 input resolution, and the subsequent stride-2 convolutions gradually downsampled the feature-map to 48 × 48 × 48, 24 × 24 × 24 and 12 × 12 × 12 voxels. There were two 3 × 3 × 3 convolutions with Instance Normalization and PReLU activation in the residual blocks.

Where necessary, channel alignment was achieved by projection shortcuts with 1 × 1 × 1 convolutions. The used bottleneck was a three parallel 3 × 3 × 3 dilated convolution blocks with dilation rate 1, 2, 4 respectively. They were concatenated and fused via a 1 × 1 × 1 conv, and then residual added and channel-spatial attention applied. This layout increased the effective receptive field without sacrificing the best spatial resolution. Channel attention focused on feature maps and spatial attention on informative volumetric regions. Encoder skip features were added to attention gates prior to a decoder fusion. The decoder comprised three stages of lightweight operations: trilinear interpolation, 1 × 1 × 1 channel projection, attention-gated skip fusion and depthwise-separable 3D convolution. The final segmentation output was produced by a 1 × 1 × 1 convolution. Table 1. The following figure illustrates the Architectural Configuration of DRA-MSCNet.

Table 1. Architectural Configuration of DRA-MSCNet

Stage	Spatial Resolution	Main Operation	Channels
Input	96 × 96 × 96	Four MRI modalities	4
Input block	96 × 96 × 96	Adaptive modality weighting	16
Encoder 1	96 × 96 × 96	Residual 3D block	16
Encoder 2	48 × 48 × 48	Strided residual blocks	32
Encoder 3	24 × 24 × 24	Strided residual blocks	64
Encoder 4	12 × 12 × 12	Deep residual blocks	128
Bottleneck	12 × 12 × 12	Dilated branches d = 1,2,4+ channel-spatial attention	128
Decoder 3	24 × 24 × 24	Attention-gated skip + depthwise-separable convolution	64
Decoder 2	48 × 48 × 48	Attention-gated skip + depthwise-separable convolution	32
Decoder 1	96 × 96 × 96	Attention-gated skip + depthwise-separable convolution	16
Output	96 × 96 × 96	1 × 1 × 1 segmentation head	4

The uncompressed architecture contained 2,207,137 trainable parameters.

Training used a combined Dice and cross-entropy objective. Background was excluded from the Dice component, and both terms contributed equally to the optimization objective:

$$\mathcal{L} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{CE}}.$$

AdamW was used with an initial learning rate of 1×10^{-4}, weight decay of 1×10^{-5}, batch size of 1, and gradient-norm clipping at 12.0. Validation used sliding-window inference with 50% overlap. Model selection was based on validation segmentation performance, and the strongest validation checkpoint was retained for final evaluation.

3.4 DRA-MSCNet Compression

To reduce the complexity of the network without compromising the segmentation performance, DRA-MSCNet was structured using pruning. During the compression analysis using validation, pruning ratios of 10%, 20%, 30%, 40%, and 50% were investigated. This was done by reducing network capacity in the structure, instead of creating a singular weight masking. The basic FP32 configuration was used for the reference. The compression configurations were evaluated based on the segmentation performance and efficiency characteristics from the validation subset. After some pruning, the 10% structurally pruned FP32 model was selected for final evaluation as it offered a good balance between model-size reduction and maintaining segmentation performance. The selected compressed configuration and the original model were tested against the same test cases, to evaluate their segmentation accuracy and computational efficiency under the same evaluation conditions.

3.5 Evaluation Metrics

Segmentation performance was assessed separately for WT, TC, and ET using overlap, voxel-classification, and boundary-based metrics.

Dice similarity was calculated as

$$\text{Dice} = \frac{2 \mid P \cap G \mid}{\mid P \mid + \mid G \mid},$$

where Pand Grepresent predicted and reference masks.

Intersection over Union was defined as

$$\text{IoU} = \frac{\mid P \cap G \mid}{\mid P \cup G \mid}.$$

Precision, sensitivity, and specificity were calculated as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

Agreement at the boundary was assessed based on Hausdorff distance (HD95), with the lower the value the closer the agreement between the predicted and reference boundaries. Patient- and tumor-region-level metrics were calculated. WT, TC and ET values have been averaged to get mean Dice and mean IoU. The number of trainable parameters, the model size, the number of MACs and FLOPs, the peak memory during inference and the inference latency per patient were used to evaluate the computational efficiency.

3.6 Statistical Analysis

Because both model configurations were tested on the same 15 test patients, a paired statistical analysis was used. A patient-level comparison of the mean Dice values for the original and compressed DRA-MSCNet configurations was performed using Wilcoxon signed-rank test. The test was two sided with the significance level set at $p<0.05$. The mean paired difference with 95% CI were reported. Holm adjustment has been used in case multiple comparisons have been considered. Rank-biserial correlation and paired Cohen's d_zwere used to characterize effect magnitude.

3.7 Ethical Considerations

The study did not involve recruitment of participants, intervention or interaction with the participants, or collection of new clinical medical imaging data and was conducted with a publicly available secondary medical imaging data set. MRI volumes were linked to anonymised patient identifiers and no effort made to identify individual participants. The imaging data and segmentation masks were only used for the development of the computational model, for segmentation analysis, for compression assessment and for performance evaluation. Anonymised and aggregated measures were used to report results. Both DRA-MSCNet and DRA-MSCNet (compressed) were thought of as research-oriented computational models and not alternatives to radiological interpretation or clinical decision making.

4. RESULTS

4.1 Dataset and Experimental Characteristics

After quality control, a total of 99 pediatric brain MRI cases were used for the analysis. For each case, 4 MRI modalities - T1n, T1c, T2w and T2-FLAIR - along with voxel-level tumor annotations were provided. The patient level partition was 69 training, 15 validation and 15 test cases. Original MRI volumes were 240 × 240 × 155 voxels, each 1 × 1 × 1 mm in size. Four channel 96 × 96 × 96 voxel patches were used for training.

Table 2. Dataset and Experimental Characteristics

Characteristic	Value
Total patients	99
MRI modalities	T1n, T1c, T2w, T2-FLAIR
Original dimensions	240 × 240 × 155 voxels
Voxel spacing	1 × 1 × 1 mm
Training cases	69
Validation cases	15
Test cases	15
Training patch size	96 × 96 × 96 voxels
Input channels	4
Evaluated regions	WT, TC, ET

4.2 Segmentation Performance

The mean test Dice score of DRA-MSCNet was 0.6430. The Dice scores were 0.7636 in Region-specific Whole Tumor, 0.7233 in Region-specific Tumor Core and 0.4422 in Region-specific Enhancing Tumor. The mean Dice score for the compressed DRA-MSCNet was 0.6376 with region scores of 0.7649, 0.7085, and 0.4393.

The regional performance was very similar between the two configurations. The figure for segmentation agreement was highest for WT and lower for TC and ET. A slightly better WT Dice score was observed in the compressed configuration, and marginally better TC and ET Dice scores were observed in the original model.

The mean HD95 for DRA-MSCNet was 37.8670 mm and 31.4338 mm for the compressed configuration, which represent lower high percentiles of the boundary distance for the compressed configuration.

Table 3. Test-Set Segmentation Performance

Model	WT Dice	TC Dice	ET Dice	Mean Dice	Mean HD95 (mm)
DRA-MSCNet	0.7636	0.7233	0.4422	0.6430	37.8670
Compressed DRA-MSCNet	0.7649	0.7085	0.4393	0.6376	31.4338

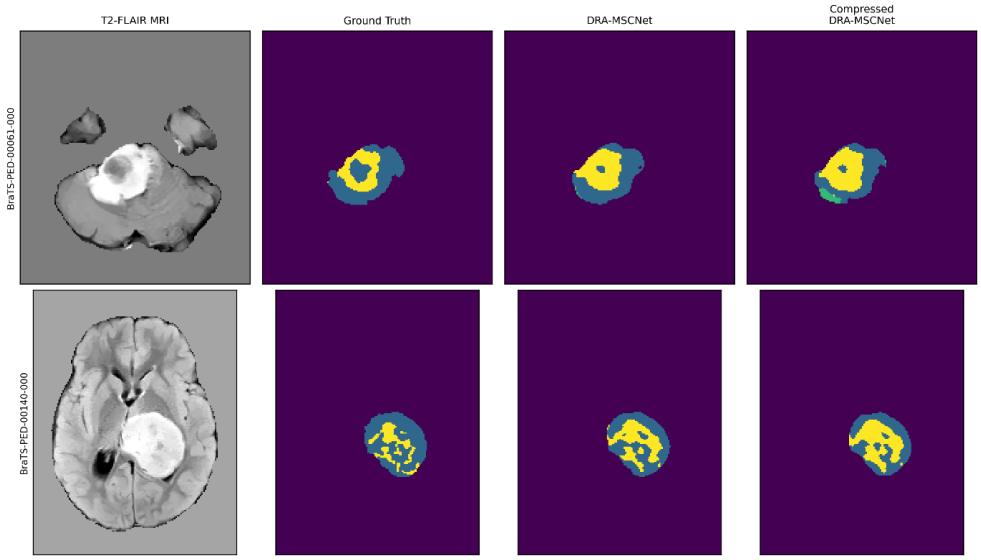

Figure 2. Qualitative comparison of ground-truth annotations and segmentation predictions generated by DRA-MSCNet and compressed DRA-MSCNet across representative MRI slices.

The qualitative examples illustrate the spatial correspondence between reference tumor masks and model-generated segmentation regions and complement the quantitative overlap and boundary measures.

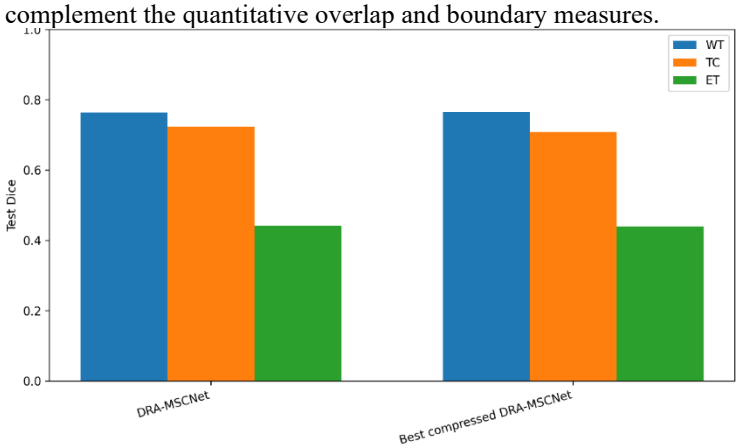

Figure 3. Regional Dice performance of DRA-MSCNet and compressed DRA-MSCNet across WT, TC, and ET.

The regional comparison demonstrates consistent performance patterns across both configurations, with stronger overlap for WT and TC than for ET.

4.3 Computational Efficiency

The original DRA-MSCNet contained 2,207,137 trainable parameters and required 8.476 MB of model storage. The compressed configuration contained 1,968,491 trainable parameters and required 7.567 MB. Structured compression reduced the trainable parameter count by approximately 10.8% and model storage by approximately 10.7%. Peak inference memory decreased from 82.21 MB to 76.73 MB, corresponding to a reduction of approximately 6.7%.

Computational complexity decreased from 36.443 G to 32.416 G MACs and from 72.885 G to 64.833 G FLOPs, representing reductions of approximately 11%. Inference latency remained comparable between configurations at 42,052.48 ms per patient for the original model and 42,116.42 ms for the compressed model.

Table 4. Segmentation Accuracy and Computational Efficiency

Model	Mean Dice	Trainable Parameters	Size (MB)	Latency (ms/patient)	Peak Memory (MB)	MACs (G)	FLOPs (G)
DRA-MSCNet	0.6430	2,207,137	8.476	42,052.48	82.21	36.443	72.885
Compressed DRA-MSCNet	0.6376	1,968,491	7.567	42,116.42	76.73	32.416	64.833

These results demonstrate improved parameter efficiency, reduced storage requirements, lower memory consumption, and lower arithmetic complexity after structured compression.

4.4 Compression Analysis

Validation-based compression analysis identified the 10% structured-pruning configuration as the preferred efficiency-performance configuration. The original FP32 DRA-MSCNet achieved a validation mean Dice of 0.6586 with a model size of 8.476 MB. The selected compressed configuration achieved a validation mean Dice of 0.6364 while reducing model size to 7.564 MB.

Validation-stage latency was 1693.15 ms for the original configuration and 1723.18 ms for the selected compressed model. The corresponding Dice change was −0.0221.

Table 5. Selected Compression Configuration

Configuration	Pruning Ratio	Precision	Validation Mean Dice	Model Size (MB)	Validation Latency (ms)	Dice Change
Original DRA-MSCNet	0%	FP32	0.6586	8.476	1693.15	0.0000
Compressed DRA-MSCNet	10%	FP32	0.6364	7.564	1723.18	−0.0221

The selected configuration provided a practical balance between reduced model complexity and preservation of segmentation performance.

4.5 Statistical Comparison

A paired statistical comparison was conducted using the 15 test cases evaluated by both configurations. The estimated mean Dice difference was 0.00545 with a 95% confidence interval of −0.0115 to 0.0224.

The Wilcoxon signed-rank test yielded $W = 27$, with a raw p-value of 0.0637 and a Holm-adjusted p-value of 0.0637. The difference therefore did not reach statistical significance at the 0.05 level. The rank-biserial effect size was 0.55, while paired Cohen's d_z was 0.178, indicating a limited standardized mean difference between the two model configurations.

Table 6. Paired Statistical Comparison of DRA-MSCNet Configurations

Comparison	Metric	Mean Difference	95% CI	Wilcoxon W	Raw p-value	Holm-Adjusted p-value	Rank-Biserial Effect	Cohen's d_z
DRA-MSCNet vs. compressed DRA-MSCNet	Mean Dice	0.00545	−0.0115 to 0.0224	27	0.0637	0.0637	0.55	0.178

Figure 7 illustrates the patient-level paired mean Dice values for the original and compressed DRA-MSCNet configurations. Most observations remained closely aligned across the two models, with only modest patient-specific variation. This pattern is consistent with the paired statistical analysis, which showed that the overall difference in mean Dice did not reach statistical significance.

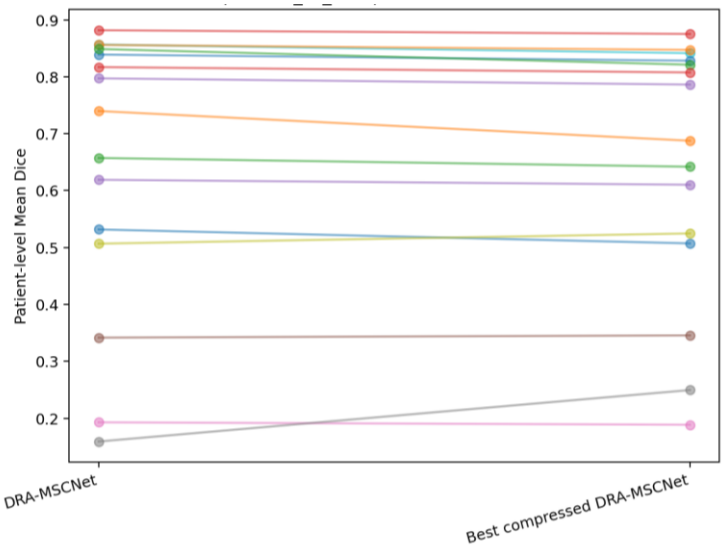

Figure 7. Patient-level paired mean Dice comparison between DRA-MSCNet and compressed DRA-MSCNet across the 15 test patients.

4.6 Overall Findings

DRA-MSCNet demonstrated consistent three-dimensional segmentation of WT, TC, and ET from multimodal pediatric MRI. The original architecture achieved a mean test Dice of 0.6430, while the compressed configuration achieved 0.6376. Regional segmentation performance remained closely comparable between the two models, and the compressed configuration achieved a lower mean HD95.

Structured pruning reduced trainable parameters by approximately 10.8%, model storage by 10.7%, peak memory by 6.7%, and MACs and FLOPs by approximately 11%. Statistical comparison showed that the difference in mean Dice did not reach statistical significance. Overall, the selected compressed DRA-MSCNet achieved meaningful reductions in computational complexity while retaining segmentation performance comparable to the original architecture.

5. DISCUSSION

Consistent 3D segmentation of Whole Tumor (WT), Tumor Core (TC) and Enhancing Tumor (ET) from multimodal pediatric MRI was achieved by DRA-MSCNet. The original model had a mean Dice score of 0.6430, and regional Dice scores of 0.7636 for WT, 0.7233 for TC, and 0.4422 for ET. The performance pattern was related to the various spatial and morphological features of the tumour regions. WT demonstrated the highest overlap, whereas ET demonstrated the lowest level of segmentation agreement, as improving regions tend to be smaller, more disjointed and less certain to be correct [1] and/or [25]. It is worth noting that the performance distribution also corroborates the architectural design of the combination of residual learning, multi-scale contextual extraction and attention-based refinement. Multi-scale processing allows network to learn both global (tumor context) and local (structural) information and the informative representations are kept in the deeper layers of the network (residual pathways). To further focus the network on tumour-relevant responses in the context of the heterogeneity of volumetric data, attention-based refinement is further added.

The mean Dice score of the compressed DRA-MSCNet was nearly the same at 0.6376. WT Dice increased slightly from 0.7636 to 0.7649, while TC and ET reached 0.7085 and 0.4393, respectively. Mean HD95 decreased from 37.8670 mm to 31.4338 mm. This outcome shows high percentile agreement for boundaries at the compressed setting with some variation in regional overlap. Therefore, the Dice and HD95 are combined to give a more holistic measure of the quality of segmentation. This regional performance trend has been seen in earlier brain tumor segmentation works where larger tumor regions are more likely to be represented accurately than the smaller highly heterogeneous regions. The benefit of combining complementary MRI modalities to capture the heterogeneous tumor tissue has been shown for multimodal architectures, like MM-UNet [3]–[5] or hybrid fusion networks. This is in line with the four-channel multimodal approach that is being followed in DRA-MSCNet.

The use of architecture with both residual and multi-scale has also been proposed to tackle the challenge of modeling tumor morphology at varying scales. In order to enhance the contextual representation and feature re-using, multi-scale feature processing is used for GMetaNet and MResU-Net. Selective feature refinement was also applied to three-dimensional tumor segmentation on attention-oriented architectures like LATUP-Net, GAIR-U-Net and RDAU-Net, which further illustrated its usefulness [13] [14] [21]. DRA-MSCNet is inspired by this technical trend, and on top of that, it introduces the residual, multi-scale, and attention mechanisms in a streamlined encoder-decoder structure. Direct numerical comparison with these studies should be interpreted with care as the composition of the datasets, preprocessing, partitioning and model configuration, as well as the evaluation protocol, vary from one study to another. This study involved 99 pediatric MR cases, which were split into 69 training, 15 validation and 15 test cases. Thus, the comparison of the literature is mainly to create the methodological consistency in the architecture positioning, but not to show the superiority on performance.

The compression analysis takes the evaluation beyond the accuracy of segmentation. Structured pruning reduced trainable parameters from 2,207,137 to 1,968,491, model size from 8.476 MB to 7.567 MB, peak memory from 82.21

MB to 76.73 MB, MACs from 36.443 G to 32.416 G, and FLOPs from 72.885 G to 64.833 G. These reductions show that architectural complexity can be lowered, without a significant decrease in segmentation performance. The selected compressed DRA-MSCNet reduced the parameters by around 10.8%, reduced the model storage by around 10.7%, reduced the peak memory consumption by 6.7%, and reduced the MACs and FLOPs by around 11%. These results suggest that structured pruning can enhance the compactness of the model and the computation speed without compromising the segmentation behaviour in the region significantly.

Inference latency were similar for the original configuration (42,052.48 ms per patient) and the compressed configuration (42,116.42 ms per patient). This observation shows that not only do the number of parameters and the number of arithmetic operations reduced are not necessarily proportional, but also the measured execution times are not necessarily reduced in the same proportion. GPU architecture, memory access, implementation of the kernels, transfer of data, software level optimizations, and runtime performance are all impacted. These paired statistical analysis also confirmed that both configurations were alike. The mean Dice difference was 0.00545 with a 95% confidence interval of −0.0115 to 0.0224. The Wilcoxon signed rank test showed a W=27 with a Holm adjusted p=0.0637 showing that the difference was not statistically significant. The pairwise Cohen's d_z=0.178were indicative of a small standardized mean difference. These results confirm the notion that the compressed configuration achieved a similar segmentation performance with fewer computations.

The results show the potential of the combination of the multimodal MRI processing, the residual learning, the multi-scale contextual extraction, the attention refinement and the structured pruning as they can all be combined in one three-dimensional segmentation framework. The architecture is a more compact method in terms of computation without compromising the segmentation results between WT, TC and ET. Such an accuracy-efficiency balance is applicable to future deployment-oriented medical image analysis systems where memory, storage and computational resources play an important part of the design. The study was done on one paediatric MRI dataset and with a specified partition of patient level data. Therefore, wider generalization needs to be done by further data, multi-institutional evaluation. Again ET was found to be the most challenging region to segment, suggesting that there is potential for improving small-region sensitivity and localization of boundaries.

The future research should focus on region-aware and boundary-sensitive optimization methods like focal loss, Tversky-based objective, hard-example mining, targeted tumor sampling, etc. Additional optimizations such as hardware-aware pruning, quantization, mixed precision inference, and knowledge distillation can further enhance the computational efficiency. Further validation in other independent cohorts, multiple experimental repetitions, controlled comparison of DRA-MSCNet with existing segmentation architectures, and validation by expert radiologists would further reinforce the translational evidence of DRA-MSCNet.

6. CONCLUSION

In this study, DRA-MSCNet has been developed and tested for 3D brain tumor segmentation in pediatric multimodal MRI. The framework combined denoising-aware preprocessing, residual feature learning, multi-scale contextual extraction, attention based refinement, and lightweight decoding to segment Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET). The original DRA-MSCNet obtained a mean Dice score of 0.6430, the highest regional agreement being for WT, followed by TC and ET. Structured pruning was then used to increase the computational efficiency. The 10% compressed configuration selected had a mean Dice score of 0.6376, reduced the number of trainable parameters from 2,207,137 to 1,968,491, and reduced the model size from 8.476 MB to 7.567 MB. Peak memory usage, MACs and FLOPs also decreased and mean HD95 decreased from 37.8670 mm to 31.4338 mm, the mean difference between the original and compressed configurations was not statistically significant with a Holm adjusted p-value of 0.0637. Overall, the results show that the DRA-MSCNet framework is effective in multi-modal 3D brain tumor segmentation and can be used to compress structured models. The compression configuration resulted in significant model complexity and resource savings without compromising segmentation performance. Future research is needed to further validate on independent as well as multi-institutional datasets, further optimise the segmentation of small enhancing tumour regions, and explore hardware aware compression, quantization and deployment oriented optimisation to further fortify the framework's practicality.

REFERENCES

1. Z. U. Abidin, R. A. Naqvi, A. Haider, H. S. Kim, D. Jeong, and S. W. Lee, “Recent deep learning-based brain tumor segmentation models using multi-modality magnetic resonance imaging: A prospective survey,” Frontiers in Bioengineering and Biotechnology, vol. 12, Art. no. 1392807, 2024, doi: 10.3389/fbioe.2024.1392807.
2. F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,” Nature Methods, vol. 18, pp. 203–211, 2021, doi: 10.1038/s41592-020-01008-z.
3. L. Zhao, J. Ma, Y. Shao, C. Jia, J. Zhao, and H. Yuan, “MM-UNet: A multimodality brain tumor segmentation network in MRI images,” Frontiers in Oncology, vol. 12, Art. no. 950706, 2022, doi: 10.3389/fonc.2022.950706.
4. F. Fang, Y. Yao, T. Zhou, G. Xie, and J. Lu, “Self-supervised multi-modal hybrid fusion network for brain tumor segmentation,” IEEE Journal of Biomedical and Health Informatics, vol. 26, pp. 5310–5320, 2022, doi: 10.1109/JBHI.2021.3109301.
5. Y. Liu, F. Mu, Y. Shi, J. Cheng, C. Li, and X. Chen, “Brain tumor segmentation in multimodal MRI via pixel-level and feature-level image fusion,” Frontiers in Neuroscience, vol. 16, Art. no. 1000587, 2022, doi: 10.3389/fnins.2022.1000587.

6. A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D medical image segmentation," in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2022, pp. 574–584, doi: 10.1109/WACV51458.2022.00181.
7. J. Liang, C. Yang, M. Zeng, and X. Wang, "TransConver: Transformer and convolution parallel network for developing automatic brain tumor segmentation in MRI images," Quantitative Imaging in Medicine and Surgery, vol. 12, pp. 2397–2415, 2022, doi: 10.21037/qims-21-919.
8. J. Liang, C. Yang, and L. Zeng, "3D PSwinBTS: An efficient transformer-based U-Net using 3D parallel shifted windows for brain tumor segmentation," Digital Signal Processing, vol. 131, Art. no. 103784, 2022, doi: 10.1016/j.dsp.2022.103784.
9. X. Li, Y. Jiang, M. Li, J. Zhang, S. Yin, and H. Luo, "MSFR-Net: Multi-modality and single-modality feature recalibration network for brain tumor segmentation," Medical Physics, vol. 50, pp. 2249–2262, 2023, doi: 10.1002/mp.15933.
10. Q. Hou, Y. Peng, Z. Wang, J. Wang, and J. Jiang, "MFD-Net: Modality fusion diffractive network for segmentation of multimodal brain tumor image," IEEE Journal of Biomedical and Health Informatics, vol. 27, pp. 5958–5969, 2023, doi: 10.1109/JBHI.2023.3318640.
11. Y. Lu, Y. Chang, Z. Zheng, Y. Sun, M. Zhao, B. Yu, et al., "GMetaNet: Multi-scale ghost convolutional neural network with auxiliary MetaFormer decoding path for brain tumor segmentation," Biomedical Signal Processing and Control, vol. 83, Art. no. 104694, 2023, doi: 10.1016/j.bspc.2023.104694.
12. "Dilated multi-scale residual attention (DMRA) U-Net: Three-dimensional (3D) dilated multi-scale residual attention U-Net for brain tumor segmentation," Quantitative Imaging in Medicine and Surgery, 2024, doi: 10.21037/qims-24-779.
13. E. J. Alwadee, X. Sun, Y. Qin, and F. C. Langbein, "LATUP-Net: A lightweight 3D attention U-Net with parallel convolutions for brain tumor segmentation," Computers in Biology and Medicine, vol. 184, Art. no. 109353, 2025, doi: 10.1016/j.compbiomed.2024.109353.
14. E. K. Rutoh, Q. Z. Guang, N. Bahadar, R. Raza, and M. S. Hanif, "GAIR-U-Net: 3D guided attention inception residual U-Net for brain tumor segmentation using multimodal MRI images," Journal of King Saud University - Computer and Information Sciences, vol. 36, Art. no. 102086, 2024, doi: 10.1016/j.jksuci.2024.102086.
15. "MVSI-Net: Multi-view attention and multi-scale feature interaction for brain tumor segmentation," Biomedical Signal Processing and Control, vol. 95, Art. no. 106484, 2024, doi: 10.1016/j.bspc.2024.106484.
16. P. Li, Z. Li, Z. Wang, C. Li, and M. Wang, "MResU-Net: Multi-scale residual U-Net-based brain tumor segmentation from multimodal MRI," Medical & Biological Engineering & Computing, vol. 62, pp. 641–651, 2024, doi: 10.1007/s11517-023-02965-1.
17. "TDPC-Net: Multi-scale lightweight and efficient 3D segmentation network with a 3D attention mechanism for brain tumor segmentation," Biomedical Signal Processing and Control, vol. 99, Art. no. 106911, 2025, doi: 10.1016/j.bspc.2024.106911.
18. R. Preetha, J. P. M. Priyadarsini, and J. S. Nisha, "Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis," Scientific Reports, vol. 15, Art. no. 9914, 2025, doi: 10.1038/s41598-025-94267-9.
19. B. Chen, T. He, W. Wang, Y. Han, J. Zhang, S. Bobek, and S. Sternad Zabukovsek, "MRI brain tumour segmentation using multiscale attention U-Net," Informatica, vol. 35, no. 4, pp. 751–774, 2024, doi: 10.15388/24-INFOR574.
20. Y. Chang, Z. Zheng, Y. Sun, M. Zhao, Y. Lu, and Y. Zhang, "DPAFNet: A residual dual-path attention-fusion convolutional neural network for multimodal brain tumor segmentation," Biomedical Signal Processing and Control, vol. 79, Art. no. 104037, 2023, doi: 10.1016/j.bspc.2022.104037.
21. J. Wang, Z. Yu, Z. Luan, J. Ren, Y. Zhao, and G. Yu, "RDAU-Net: Based on a residual convolutional neural network with DFP and CBAM for brain tumor segmentation," Frontiers in Oncology, vol. 12, Art. no. 805263, 2022, doi: 10.3389/fonc.2022.805263.
22. "FFLUNet: Feature fused lightweight UNet for brain tumor segmentation," Computers in Biology and Medicine, vol. 194, Art. no. 110460, 2025, doi: 10.1016/j.compbiomed.2025.110460.
23. G. Ramasamy, T. Singh, X. Yuan, and G. R. Naik, "HMSA-Net: A hierarchical multi-scale attention network for brain tumor segmentation from multi-modal MRI," IEEE Access, vol. 13, pp. 167672–167683, 2025, doi: 10.1109/ACCESS.2025.3606560.
24. "Brain tumor segmentation using a hybrid multi-resolution U-Net with residual dual attention and deep supervision on MR images," Biomedical Signal Processing and Control, vol. 78, Art. no. 103939, 2022, doi: 10.1016/j.bspc.2022.103939.
25. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993–2024, 2015, doi: 10.1109/TMI.2014.2377694.
26. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," Scientific Data, vol. 4, Art. no. 170117, 2017, doi: 10.1038/sdata.2017.117.
27. MedOtter, "BraTS2023 Pediatric Dataset," Hugging Face Datasets. [Online]. Available: https://huggingface.co/datasets/MedOtter/brats2023-ped-dataset. [Accessed: Sep. 11, 2026].