



Providing Choice & Value

Generic CT and MRI Contrast Agents



**FRESENIUS
KABI**

CONTACT REP

AJNR








This information is current as
of July 23, 2025.

Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors

Arastoo Vossough, Nastaran Khalili, Ariana M. Familiar,
Deep Gandhi, Karthik Viswanathan, Wenxin Tu, Debanjan
Haldar, Sina Bagheri, Hannah Anderson, Shuvanjan Haldar,
Phillip B. Storm, Adam Resnick, Jeffrey B. Ware, Ali
Nabavizadeh and Anahita Fathi Kazerooni

AJNR Am J Neuroradiol published online 9 May 2024
<http://www.ajnr.org/content/early/2024/05/09/ajnr.A8293>

Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors

 Arastoo Vossough, Nastaran Khalili,  Ariana M. Familiar, Deep Gandhi, Karthik Viswanathan,  Wenxin Tu, Debanjan Haldar,  Sina Bagheri, Hannah Anderson, Shuvanjan Haldar, Phillip B. Storm, Adam Resnick,  Jeffrey B. Ware,  Ali Nabavizadeh, and  Anahita Fathi Kazerooni



ABSTRACT

BACKGROUND AND PURPOSE: Tumor segmentation is essential in surgical and treatment planning and response assessment and monitoring in pediatric brain tumors, the leading cause of cancer-related death among children. However, manual segmentation is time-consuming and has high interoperator variability, underscoring the need for more efficient methods. After training, we compared 2 deep-learning-based 3D segmentation models, DeepMedic and nnU-Net, with pediatric-specific multi-institutional brain tumor data based on multiparametric MR images.

MATERIALS AND METHODS: Multiparametric preoperative MR imaging scans of 339 pediatric patients ($n = 293$ internal and $n = 46$ external cohorts) with a variety of tumor subtypes were preprocessed and manually segmented into 4 tumor subregions, ie, enhancing tumor, nonenhancing tumor, cystic components, and peritumoral edema. After training, performances of the 2 models on internal and external test sets were evaluated with reference to ground truth manual segmentations. Additionally, concordance was assessed by comparing the volume of the subregions as a percentage of the whole tumor between model predictions and ground truth segmentations using the Pearson or Spearman correlation coefficients and the Bland-Altman method.

RESULTS: The mean Dice score for nnU-Net internal test set was 0.9 (SD, 0.07) (median, 0.94) for whole tumor; 0.77 (SD, 0.29) for enhancing tumor; 0.66 (SD, 0.32) for nonenhancing tumor; 0.71 (SD, 0.33) for cystic components, and 0.71 (SD, 0.40) for peritumoral edema, respectively. For DeepMedic, the mean Dice scores were 0.82 (SD, 0.16) for whole tumor; 0.66 (SD, 0.32) for enhancing tumor; 0.48 (SD, 0.27) for nonenhancing tumor; 0.48 (SD, 0.36) for cystic components, and 0.19 (SD, 0.33) for peritumoral edema, respectively. Dice scores were significantly higher for nnU-Net ($P \leq .01$). Correlation coefficients for tumor subregion percentage volumes were higher (0.98 versus 0.91 for enhancing tumor, 0.97 versus 0.75 for nonenhancing tumor, 0.98 versus 0.80 for cystic components, 0.95 versus 0.33 for peritumoral edema in the internal test set). Bland-Altman plots were better for nnU-Net compared with DeepMedic. External validation of the trained nnU-Net model on the multi-institutional Brain Tumor Segmentation Challenge in Pediatrics (BraTS-PEDs) 2023 data set revealed high generalization capability in the segmentation of whole tumor, tumor core (a combination of enhancing tumor, nonenhancing tumor, and cystic components), and enhancing tumor with mean Dice scores of 0.87 (SD, 0.13) (median, 0.91), 0.83 (SD, 0.18) (median, 0.89), and 0.48 (SD, 0.38) (median, 0.58), respectively.

CONCLUSIONS: The pediatric-specific data-trained nnU-Net model is superior to DeepMedic for whole tumor and subregion segmentation of pediatric brain tumors.

ABBREVIATIONS: AI = artificial intelligence; BraTS = Brain Tumor Segmentation Challenge; CBTN = Children's Brain Tumor Network; CC = cystic component; CNN = convolutional neural network; DMG/DIPG = diffuse midline glioma/diffuse intrinsic pontine glioma; ED = edema; ET = enhancing tumor; NET = non-enhancing tumor; TC = tumor core; WT = whole tumor

Pediatric CNS tumors are the second most common childhood cancer and represent the most prevalent solid tumor

and the leading cause of cancer-related mortality in children.^{1,2} These tumors encompass a wide range of histologies and display

Received January 5, 2024; accepted after revision March 1.

From the Center for Data Driven Discovery in Biomedicine (A.V., N.K., A.M.F., D.G., K.V., D.H., S.B., H.A., P.B.S., A.R., A.N., A.F.K.), Department of Radiology (A.V.), and Department of Neurosurgery (P.B.S., A.F.K.), Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; Department of Radiology (A.V., S.B., J.B.W., A.N.), College of Arts and Sciences (W.T.), Center for AI & Data Science for Integrated Diagnostics (A.F.K.), and Center for Biomedical Image Computing and Analytics (A.F.K.), University of Pennsylvania, Philadelphia, Pennsylvania; and School of Engineering (S.H.), Rutgers University, New Brunswick, New Jersey.

This research was supported by grant funding from the diffuse intrinsic pontine glioma/diffuse midline glioma Research Funding Alliance and Pediatric Brain Tumor Foundation.

Please address correspondence to Anahita Fathi Kazerooni, PhD, Center for Data Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA; e-mail: anahitaf@upenn.edu; @anahita_fathi



Indicates article with online supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8293>

SUMMARY

PREVIOUS LITERATURE: The research on automated segmentation of pediatric brain tumors remains sporadic, particularly in the context of multi-institutional and multi-histology data sets, and with validation on independent test sets. Studies employing deep learning for the segmentation of whole lesions in pediatric brain tumors have documented Dice scores from 0.72 to 0.86. Only one study has reported on the segmentation of different tumor subregions, including enhancing tumor, nonenhancing tumor, cystic components, and edema, with Dice scores ranging between 0.35 and 0.74.

KEY FINDINGS: Our automated tumor segmentation method by using nnU-Net demonstrates better performance in segmenting all subregions when compared with existing studies in the literature with high Dice scores of 0.90 for whole tumors, 0.77 for enhancing regions, and a combined score (0.82) for all nonenhancing areas.

KNOWLEDGE ADVANCEMENT: This developed model, which is available for public access, generates reliable results for segmentation of different tumor subregions across various histologies, validated with multi-institutional data.

marked variations in their molecular origins, disease course, and response to therapy, which complicates the clinical decision-making process for their management.³ MR images are imperative to precisely locate, characterize, and monitor the treatment for these tumors. Achieving this precision relies heavily on accurate delineation and characterization of the whole tumor (WT) and tumor subcomponents. Quantitative measures of change in tumors are highly desirable for objective assessment of size and signal intensity, requiring visual and manual measurement of the tumors and tumor components. Accurate, automated tumor segmentation methods can offer rapid determination of tumor volumes with less effort and potentially more consistency than manual segmentation methods.

Manual delineation of brain tumors presents distinct obstacles that require specialized expertise, resources, and time.⁴ Moreover, there are distinct differences in the prevalence, appearance, histology, and behavior of pediatric brain tumors compared with adult brain tumors.^{1,5} The most common pediatric brain tumors include pilocytic astrocytoma, medulloblastoma, and other gliomas. In adults, the most common intracranial tumors are brain metastases and meningiomas, and the most common primary intra-axial brain tumors are glioblastoma.⁶ There is a higher prevalence of circumscribed gliomas in children compared with infiltrating gliomas that are commonly seen in adults.⁷ Necrosis is much more prevalent in adult brain tumors than in children, whereas tumoral cysts are more common in pediatrics.⁷ Contrast enhancement in adult brain tumors is more commonly associated with high-grade tumors, whereas a large proportion of low-grade pediatric tumors demonstrate contrast enhancement.⁸

The response assessment criteria for adult and pediatric brain tumors are also considerably different.⁹⁻¹¹ The differences have implications for brain tumor subregion segmentation. In adult brain tumor segmentations, typically nonenhancing tumor (NET) and edema are combined into a single subregion label, given the difficulty of separating these tissues and of common infiltrative adult brain tumors. Alternatively, NET and tumor necrosis may be combined into 1 label, though they are quite distinct on imaging. As a result of these differences, segmentation models trained on adult brain tumors may not be well-suited for segmentation of pediatric brain tumors, leading to under- or oversegmentation of tumor subregions.¹² Thus, there

is a relatively unmet need for training and validation of more accurate, pediatric-specific, brain tumor segmentation models with tumor subregion delineation.

Advancements in deep learning have broadened the potential applications of artificial intelligence (AI) in the field of medical imaging. In the context of brain tumors, 3D convolutional neural networks (CNNs) have gained widespread use in this field for image segmentation due to their ability to capture spatial features in 3D data, which is particularly relevant for brain MR imaging.¹³ Nevertheless, optimal use of CNNs for pediatric brain tumor segmentation remains inadequate.¹⁴ Distinctive characteristics of the pediatric brain as well as inherent limitations in available data sources present formidable challenges in the development of automated segmentation methods tailored to pediatric cases.¹⁵ Moreover, given the necessity for AI systems to undergo specific training and validation for each distinct application, it is evident that focused research endeavors within the pediatric demographic are essential for optimal application of AI in segmenting pediatric brain tumors.

Recent advancements in CNN structures like ShuffleNet (<https://arxiv.org/abs/1707.01083>), ResNet (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7711146/>), and DenseNet (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8300985/>) have shown promising results in various image analysis tasks, including adult brain tumor segmentation.^{16,17} U-Net, a CNN based on an encoder-decoder architecture, has become popular for medical image segmentation, showing superior performance in small data sets with a limited number of scans.^{18,19} Thus, U-Net and its derivatives are potentially helpful for segmentation of pediatric brain tumors. A few studies have used U-Net models to improve segmentation performance on types of pediatric brain tumors, yielding good agreement between predicted and manual segmentations.^{20,21} Nevertheless, there is a paucity of available highly accurate deep learning models that can be used for pediatric brain tumor segmentation across a wide range of tumor pathologies.

In this article, we investigate 2 innovative, 3D deep learning segmentation architectures that have been successfully applied in adult brain tumor segmentations, namely DeepMedic (<https://deepmedic.org/>) and nnU-Net models (<https://www.nature.com/articles/s41592-020-01008-z>).^{22,23} DeepMedic is a multilayered, multiscale, 3D deep CNN architecture coupled with a 3D fully

Table 1: Characteristics of patients, tumor histology, and MR imaging scanners in internal and external patient cohorts

	Internal Cohort	External Cohort
Total patients	293	46
Median age at imaging (range) (yr)	7.84 (0.24–21.71)	9.42 (0.55–20.73)
Sex		
Male	157 (53.5%)	22 (47.8%)
Female	134 (45.7%)	24 (52.1%)
NA	2 (0.6%)	
Histology		
Low-grade glioma/astrocytoma	152 (51.9%)	21 (45.7%)
High-grade glioma/astrocytoma	23 (7.8%)	3 (6.5%)
Ependymoma	7 (2.4%)	0 (0%)
Medulloblastoma	85 (29%)	17 (36.9%)
Brainstem glioma	15 (5.1%)	2 (4.3%)
Germinoma	2 (0.7%)	0 (0%)
Other	9 (3.1%)	3 (6.5%)
Percentage of tumors containing each subregion		
ET	85.0%	91.3%
NET	95.0%	93.5%
CC	68.3%	84.8%
ED	43.3%	58.7%

Note:—NA indicates not available.

connected conditional random field,²² showing excellent performance in adult brain tumors, notably in the 2017 BraTS Challenge by the Medical Image Computing and Computer Assisted Interventions (MICCAI) organization (<https://arxiv.org/html/2305.17033v6>). nnU-Net, a self-configuring model for biomedical image segmentation, excels in various applications,²³ including The Brain Tumor Segmentation (BraTS) 2020 multiorgan segmentation challenge. nnU-Net extends the original U-Net design with an automated end-to-end pipeline, selecting the optimal configuration for diverse segmentation tasks. Our objective was to harness an extensive collection of multi-institutional ground truth segmentations to train these 2 models on carefully curated, pediatric-specific data and compare the performance for automated pediatric brain tumor subregion segmentation across a wide spectrum of tumor types. We hypothesized that the nnU-NET segmentation model would perform better than DeepMedic.

MATERIALS AND METHODS

Data Description and Patient Cohort

This was a Health Insurance Portability and Accountability Act-compliant, institutional review board (Children's Hospital of Philadelphia)-approved study of previously acquired multi-institutional data from the subjects enrolled onto the Children's Brain Tumor Network (CBTN) consortium (<https://cbtn.org>).²⁴ MR imaging examinations of pediatric patients with histologically confirmed brain tumors from the CBTN consortium were retrospectively collected. Inclusion criteria were the availability of preoperative brain MR imaging comprising 4 conventional MR imaging sequences, ie, precontrast T1WI, T2WI, T2 FLAIR, and gadolinium postcontrast T1WI sequences, all acquired as a part of standard-of-care clinical imaging evaluation for brain tumors. Patients were still included if the only procedure was placement of an external ventricular drain or needle biopsy. Patients were excluded if the images were incomplete or if they were severely

degraded by artifacts. Internal site data were from the Children's Hospital of Philadelphia, and external data were from other consortium members of the CBTN. A total of 339 patients (293 from the internal site, 46 from the external sites) were included in this study. Detailed descriptions of the patients, tumor types, and MR image characteristics are included in Table 1 and the Online Supplemental Data. We also used another independent data set ($n = 92$) from Brain Tumor Segmentation Challenge in Pediatrics (BraTS-PEDs) data set, for the purpose of benchmarking our best performing model (See "Benchmarking Our Model in the BraTS-PEDs Context" for more details). We report a level 5A efficacy for our study.

Details about image preparation, preprocessing, and tumor subregion segmentation can be found in the Online Supplemental Data. Tumors were segmented into 4 subregions,¹² including enhancing tumor (ET), NET, cystic component (CC), and edema (ED). WT segmentation masks were generated by the union of all 4 tumor components (ie, WT = ET + NET + CC + ED). Tumor core (TC) was defined as a combination of ET, NET, and CC.²⁵

Model Training and Validation

We trained and evaluated two 3D convolutional neural networks, DeepMedic and nnU-Net, for automated tumor subregion segmentation on multiparametric MR imaging sequences of 233 subjects from the internal cohort and tested them on withheld sets of 60 internal and 46 external subjects from the CBTN. nnU-Net v1 (<https://github.com/MIC-DKFZ/nnUNet/tree/nnunetv1>) with 5-fold cross-validation was trained with an initial learning rate of 0.0, stochastic gradient descent with Nesterov momentum ($\mu = 0.99$), and number of epochs = 1000×250 minibatches. Because the DeepMedic approach does not inherently include cross-validation, a validation set comprising 20% of the 293 training subjects ($n = 47$) was randomly selected. DeepMedic, Version 0.8.4 (<https://github.com/deepmedic/deepmedic>), was trained from scratch with a learning rate = 0.001, number of epochs = 35, and batch size = 10.

Benchmarking Our Model in the BraTS-PEDs Context

We extended the validation of our best-performing model (based on the results on internal and external data sets) to include the latest benchmarks in automated tumor segmentation, specifically focusing on the multi-institutional data set provided through the BraTS-PEDs 2023 data set.^{25,26} Our analysis involved applying the best model to a cohort of 92 pediatric subjects diagnosed with high-grade gliomas, which include astrocytoma and diffuse midline glioma/diffuse intrinsic pontine glioma (DMG/DIPG).²⁵

The evaluation of the performance of our model was conducted in alignment with the BraTS-PEDs validation criteria,²⁵

Table 2: Results comparing the performance metrics of nnU-Net versus DeepMedic architectures for WT and tumor component segmentations compared with the manual ground truth in terms of Dice score metric

Region	Internal Test Subjects		External Test Subjects	
	Dice Score: Mean (median)		Dice Score: Mean (median)	
	nnU-Net	DeepMedic	nnU-Net	DeepMedic
WT	0.9 (SD,0.07) (0.94)	0.82 (SD, 0.16) (0.88)	0.88 (SD, 0.07) (0.9)	0.78 (SD, 0.18) (0.86)
ET	0.77 (SD, 0.29) (0.86)	0.66 (SD, 0.32) (0.75)	0.75 (SD, 0.26) (0.85)	0.65 (SD, 0.32) (0.8)
NET	0.66 (SD, 0.32) (0.80)	0.48 (SD, 0.27) (0.49)	0.53 (SD, 0.32) (0.64)	0.4 (SD, 0.27) (0.4)
CC	0.71 (SD, 0.33) (0.83)	0.48 (SD, 0.36) (0.55)	0.55 (SD, 0.33) (0.67)	0.37 (SD, 0.33) (0.35)
ED	0.71 (SD, 0.40) (1)	0.19 (SD, 0.33) (0)	0.4 (SD, 0.43) (0.19)	0.21 (SD, 0.32) (0)
NET+CC+ED	0.82 (SD, 0.14) (0.86)	0.67 (SD, 0.2) (0.73)	0.74 (SD, 0.2) (0.8)	0.59 (SD, 0.23) (0.65)
TC	0.91 (SD, 0.07) (0.94)	0.78 (SD, 0.18) (0.83)	0.87 (SD, 0.07) (0.88)	0.76 (SD, 0.18) (0.79)

which focus on the segmentation of ET, TC, and the WT. The TC region encompasses the ET, NET, and CC regions.

Code Availability

All image-processing tools used in this study are freely available for public use (CaPTk, <https://www.cbica.upenn.edu/captk>; ITK-SNAP, <http://www.itksnap.org/>). The pretrained nnUNet tumor segmentation model is publicly available on GitHub (<https://github.com/d3b-center/peds-brain-auto-seg-public>). It is also a software plug-in (“gear”) on the Flywheel platform (<https://flywheel.io/>) and can be found by searching “Pediatric Brain Automated Segmentation” in the Flywheel Gear Exchange library (<https://flywheel.io/gear-exchange/>).

Statistical Methods

The performance of the DeepMedic and nnU-Net models with respect to the expert manual ground truth segmentations was evaluated using several evaluation metrics, including the Dice score (Sørensen-Dice similarity coefficient), sensitivity, and 95% Hausdorff distance. We assessed segmentation of the WT, ET, NET, CC, and ED subregions along with the nonenhancing component/edema, which encompassed the combination of NET, CC, and ED. A paired *t* test was used for comparison of Dice scores between the 2 automated segmentation methods. The comparison between the predicted and expert tumor segmentations was further analyzed by measuring a selection of semantic radiomic features, namely Visually AcceSable Rembrandt Images (VASARI) features (<https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project>), which have been extensively used in the studies on the diagnosis and prognosis of brain tumors.^{27,28} We calculated the proportion of the WT volume, that is ET, NET, CC, or ED, for the nnU-Net and DeepMedic models and compared them with the values obtained using expert segmentations by using Pearson or Spearman correlation coefficients, depending on the data distribution, the Bland-Altman method, and Mann-Whitney *U* test.

RESULTS

The Dice score, sensitivity, and 95% Hausdorff distance metrics of the nnU-Net and DeepMedic deep learning models compared with manual ground truth segmentations are shown in Table 2 and the Online Supplemental Data. The Online Supplemental Data show the breakdown across different histologies. Median values are included in addition to means (SDs) because a discrepancy

in the mere absence or presence of even a small tumor subregion label in one of the model-pair comparisons may tremendously affect the metrics for a particular subject and disproportionately affect the mean (eg, result in a calculated Dice score of 0). Additionally, the distribution of the internal and external test set Dice scores of WT and tumor subregions for the 2 models is shown with violin plots in Fig 1 (The Online Supplemental Data show the breakdown across different histologies). The distribution of Dice scores in WT and across all tumor subregions was more favorable for nnU-Net compared with DeepMedic, with a tighter distribution toward higher Dice scores (Table 2). Both models performed worse in ED segmentation compared with other tumor subregions, but even here the nnU-Net performed much better than DeepMedic ($P < .001$ and $P = .003$ for internal and external sets, respectively), which had rather poor results for ED delineation (Dice score = 0.19 and 0.21 for internal and external test subjects, respectively). Paired *t* test comparison of Dice scores between the 2 automated segmentation methods demonstrated higher nnU-Net Dice scores compared with DeepMedic for ET segmentation in the internal test set mean Dice score: 0.77 (SD, 0.29) (median, 0.86) versus 0.66 (SD, 0.32) (median, 0.75) ($P = .01$), and similarly for WT and all other tumor subregions in both the internal and external test sets ($P < .001$).

Correlation between the volume percentages of tumor subregion segmentation compared with the ground truth for nnU-Net and DeepMedic models are shown in Fig 2 and the Online Supplemental Data. All correlation coefficient *P* values were $< .001$, but again the lowest correlations were seen with DeepMedic ED determination, with $r = 0.48$ and $r = 0.33$ for internal and external test subjects, respectively. All nnU-Net percentage volume correlation coefficients were close to or above $r = 0.9$. Furthermore, the Bland-Altman assessment of agreement between the methods showed tighter 95% intervals for nnU-Net compared with the ground truth as opposed to DeepMedic (Fig 3 and the Online Supplemental Data).

Finally, the results of the Mann-Whitney *U* test comparing tumor subregions as a proportion of WT in nnU-Net and DeepMedic segmentations compared with ground truth segmentation are shown in the Online Supplemental Data. ED segmentation proportions determined by DeepMedic were statistically different from the ground truth for the internal test subjects. For the internal test subjects, the DeepMedic model was significantly different from ground truth in the proportion of ED with respect to the WT (P value = .003).

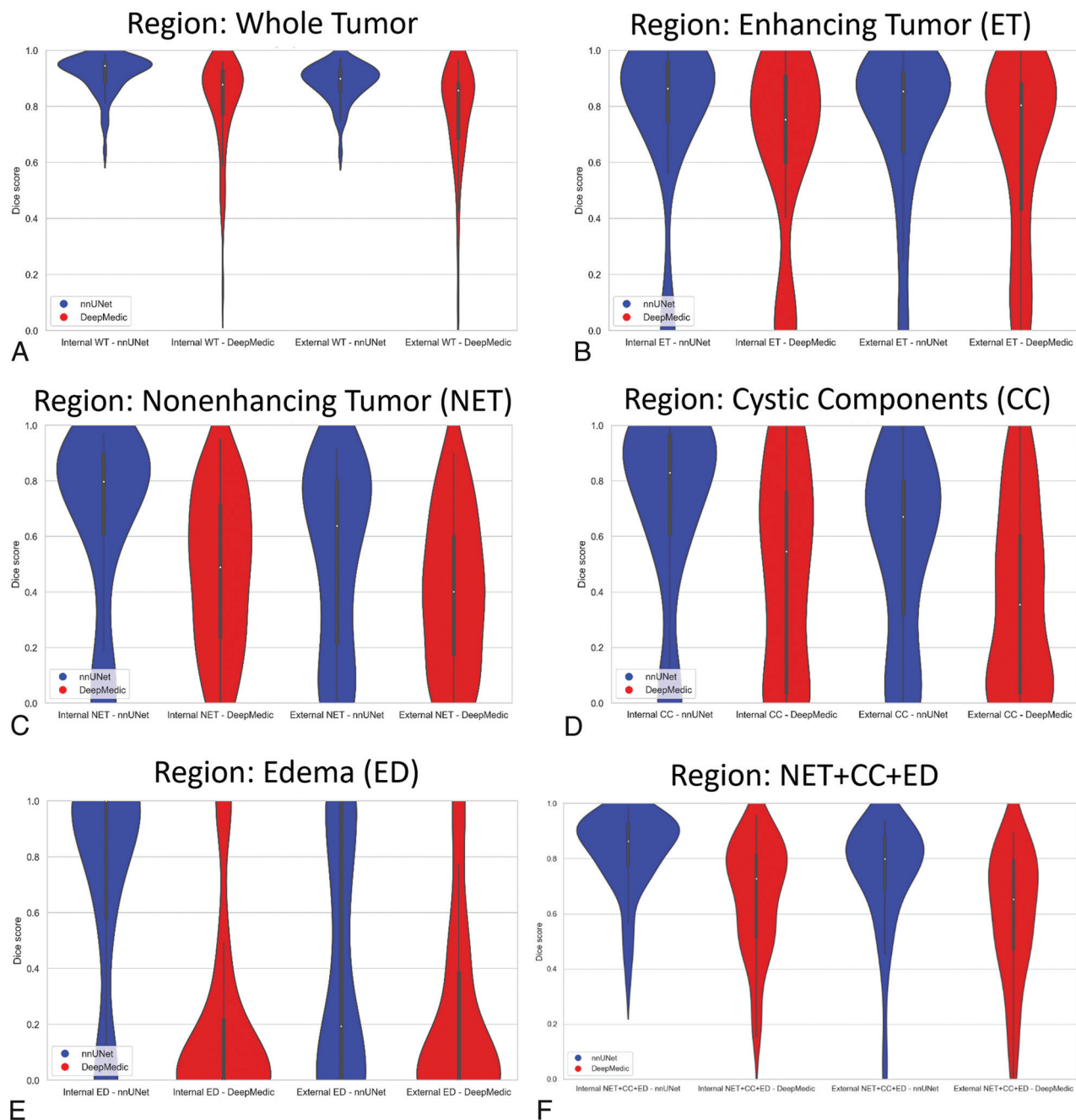


FIG 1. Violin plots demonstrating the distribution of Dice scores for nnU-Net and DeepMedic segmentation compared with ground truth for both internal and external test sets. A, WT. B, ET. C, NET. D, CC. E, ED. F, All NETs (NET + CC + ED).

Sample MR images with results comparing ground truth segmentation with nnU-Net and DeepMedic models are shown in Fig 4 and the Online Supplemental Data.

Because the nnU-Net-based autosegmentation model performed better than the DeepMedic model on our withheld internal and external sets, we applied it to the multi-institutional BraTS-PEDs 2023 data set (Table 3). This model indicates a high degree of generalizability in segmenting both the WT and the tumor core region mean Dice scores of 0.87 (SD, 0.13) (median, 0.91) and 0.83 (SD, 0.18) (median, 0.89), respectively. However, a comparatively lower performance in segmenting the ET region was a mean Dice score of 0.48 (SD, 0.38) (median, 0.58).

DISCUSSION

DeepMedic²² has been effective in pediatric tumor subregion segmentation¹² and has outperformed technicians in brain tumor segmentation on a multi-institutional MR imaging database, as validated by neuroradiologists.²⁹ Its multiscale, parallel-processing approach captures comprehensive contextual information. Using feature maps from its final convolutional layer, DeepMedic efficiently predicts voxel labels in input patches.³⁰ However, its focus on high-level semantic information may limit its performance.³⁰

nnU-Net^{23,31} has also been successfully applied in automated segmentation of brain metastases and specific tumors such as craniopharyngiomas³² or meningiomas.³³ On the basis of the U-Net

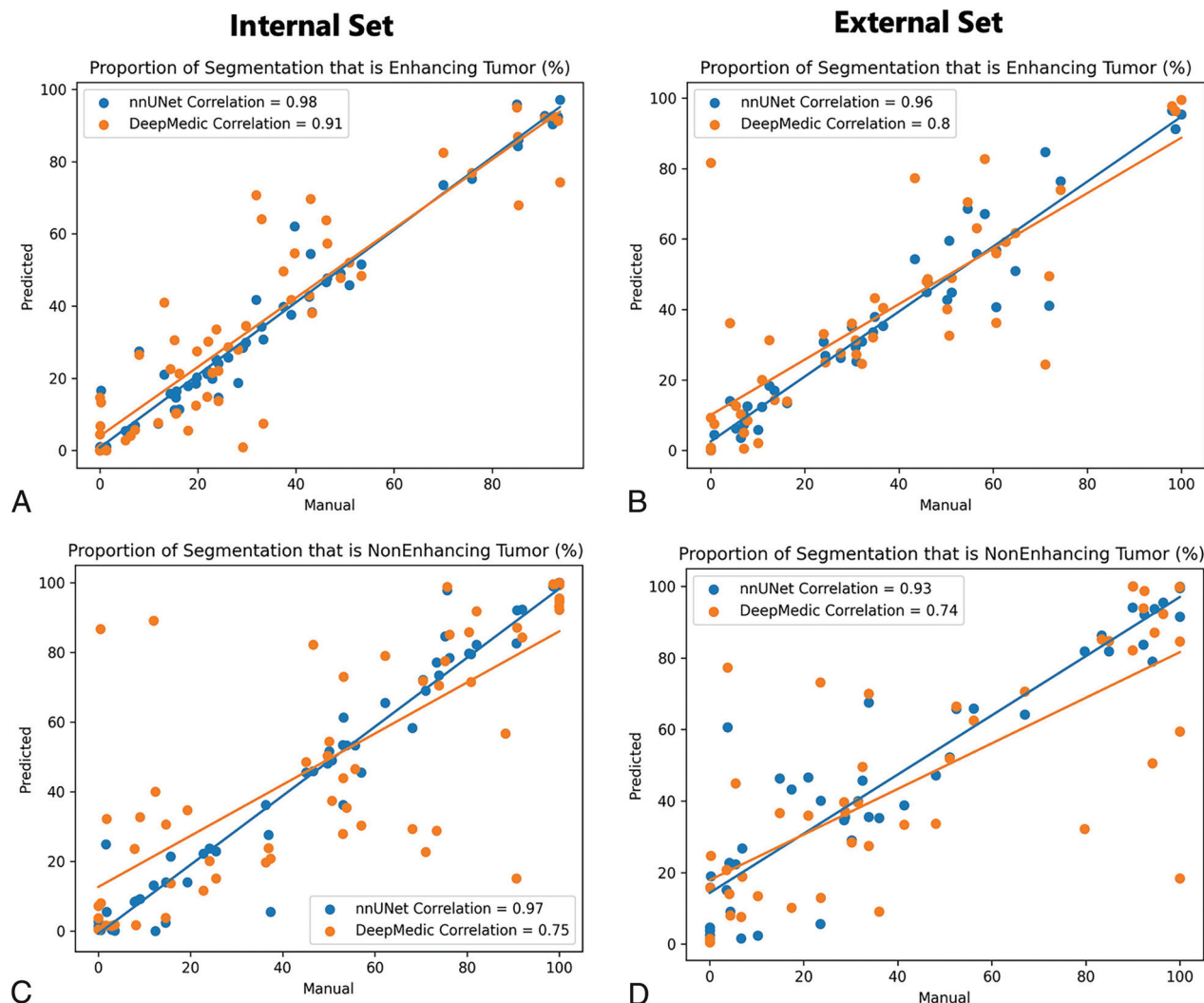


FIG 2. Scatterplots of the correlations between ground truth and automated tumor subregion volume percentages from nnU-Net and DeepMedic for internal and external test sets. A and C, Proportion of tumor that is labeled enhancing, B and D, Proportion of tumor that is labeled NET.

structure, it combines encoding with downsampling and decoding with upsampling. U-Net introduces multilevel information gradually, from encoding to decoding, to optimize prediction accuracy.

Leveraging accurate automated methods can greatly facilitate the rapid delineation of structures and potentially decrease variability. To this end, we leveraged a carefully curated manual segmentation data set of pediatric brain tumors and trained nnU-Net and DeepMedic models. nnU-Net showed high Dice scores, sensitivity, and 95% Hausdorff distance metrics compared with DeepMedic. The distribution scale of Dice scores was also tighter with less variability across subjects for WT and all tumor subregions, resulting in statistically significant higher Dice scores for nnU-Net. Additionally, for a more practical evaluation, volume percentages (proportions) of tumor subregions with respect to the WT were also evaluated using core relational analysis and Bland-Altman plots, which again showed higher correlation and agreement of nnU-Net with the manual ground truth. nnU-Net achieved high Dice scores of 0.90 and 0.77 for WT and ET

regions, respectively, though the Dice scores for NET (0.66), CC (0.71), and ED (0.71) were lower. The Dice score (0.82) on a combination of all the nonenhancing areas (NET + CC + ED) was higher than in these individual subregions. Our model demonstrates better performance in segmenting all subregions compared with existing studies in the literature. Other research using deep learning for the segmentation of the WT region in pediatric brain tumors has documented Dice scores from 0.72 to 0.86 across studies.^{12,20,21,34-36} Only a limited number of studies focus on the segmentation of ET, with Dice scores of 0.68,¹² 0.72,²⁰ and 0.83.³⁵ The highest Dice score of 0.83 in one of these studies³⁵ was achieved through cross-validation instead of evaluation on an independent test set. Only 1 study has reported the segmentation of NET, CC, and ED subregions, with Dice scores ranging between 0.35 and 0.51 and a Dice score of 0.73 for a combination of all nonenhancing regions.¹²

Our results of Dice score values for both models performed worse in delineation of ED, and this outcome was particularly problematic for DeepMedic. Even though a smaller proportion of

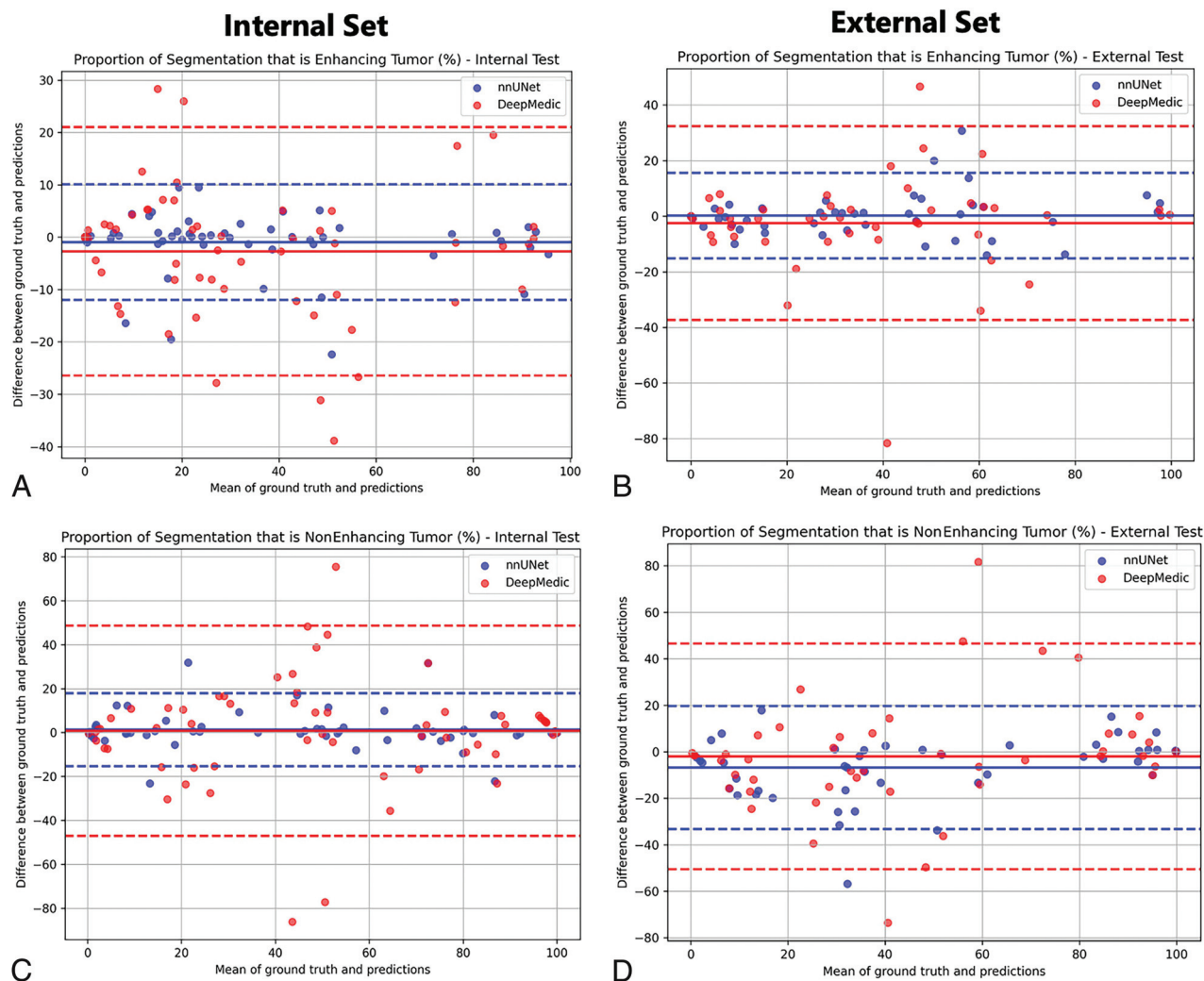


FIG 3. Bland-Altman analysis plots demonstrating the agreement between ground truth and automated tumor subregion volumes/percentages from nnU-Net and DeepMedic for internal and external test sets. *A* and *C*, Proportion of tumor that is labeled enhancing. *B* and *D*, Proportion of tumor that is labeled NET.

pediatric brain tumors are infiltrative compared with adult brain tumors, nevertheless, the ever-present challenge of separating ED from NET persists to some degree. Overall, both models performed slightly worse on the external test cohort than on the internal test set. This result was expected to some degree because the external test set was from multiple institutions with a wider variability of MR imaging acquisition protocols.

This study demonstrates the feasibility of achieving accurate pediatric brain tumor subregion segmentation results on the basis of multiparametric, multiscaner, multihistology, and multi-institutional clinical standard-of-care MR imaging scans. In some cases, manual evaluation and revision may still be required to refine the segmentations, but nevertheless, the time and effort burden will be substantially less using an accurate, automated, segmentation software.

Our study has several advantages. The MR imaging examinations were performed as clinical standard-of-care scans without a predetermined universal research imaging protocol. The imaging was performed using various vendors and models of MR imaging scanners as well as on different field-strength magnets. These

features help with increasing the generalizability of the segmentation model results. The ground truth manual segmentations were also performed in a rigorous fashion through multiple iterations to ensure high-quality ground truth labeling. As evident from the results of our study, segmentation evaluation studies generally benefit from training with larger data sets, and this benefit is true for our study as well. Increasing the sample size may be particularly helpful in improving segmentation performance for external multi-institutional data in which there is wider variability and scan protocols that can contribute to suboptimal performance for some tumor subregions.

The effectiveness of our approach is evidenced by the high Dice scores obtained for segmenting both the WT and tumor core components using the multi-institutional BraTS-PEDs 2023 data set. However, the lower performance in segmenting the ET region with the mean Dice score = 0.48 (SD, 0.38) (median, 0.58) can be attributed to the low prevalence of this subregion in subjects with DMG/DIPG tumors in the BraTS-PEDs data set. Additionally, our original training cohort included a limited number of high-grade gliomas (including DMG/DIPG histology),

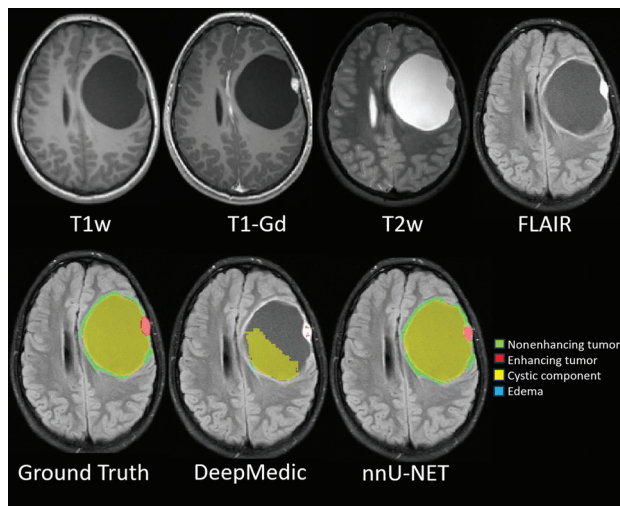


FIG 4. Sample comparison images of ground truth and predicted segmentation from nnU-Net and DeepMedic in a predominantly cystic supratentorial mass with a solid enhancing nodule. While nnU-Net has near-perfect agreement with the ground truth, DeepMedic failed to delineate most of the large cyst, enhancing nodule, and nonenhancing rim. T1w indicates T1-weighted; T1-GD, T1-gadolinium.

Table 3: Results on the performance of our trained autosegmentation model using nnU-Net architecture on multi-institutional BraTS-PEDs 2023 data ($n = 92$)^a

Region/Metric	Dice Score Mean (Median)	Sensitivity Mean (Median)	95% Hausdorff Distance Mean (Median)
WT	0.87 (SD, 0.13) (0.91)	0.82 (SD, 0.16) (0.88)	6.47 (SD, 15.99) (3)
ET	0.48 (SD, 0.38) (0.58)	0.55 (SD, 0.35) (0.71)	11.26 (SD, 20.06) (3.40)
TC	0.83 (SD, 0.18) (0.89)	0.82 (SD, 0.17) (0.87)	7.65 (SD, 16.27) (3.67)

Note:—The metrics are reported for whole tumor, enhancing tumor, and tumor core (ET + CC + NET) in agreement with BraTS-PEDs evaluation method.

which may have contributed to this reduced performance in the ET region. Incorporating the additional BraTS-PEDs 2023 data set into the training data for our future autosegmentation models is a logical step to address this issue. This would likely enhance the generalizability of the model, particularly for the tumors with rare histology in multi-institutional cohorts.

No segmentation performance metric is optimal, exemplified by the limitations of the Dice score, as demonstrated in this study. Specifically, if a model fails to segment a certain label, the Dice score for that label will be equal to 0. This zero score considerably influences the aggregate Dice score for that label across all samples. Furthermore, if the data set contains a limited number of samples with that particular label, the aggregate score may not accurately reflect the overall performance trend of the model. As a result, we used multiple performance metrics and included tumor subregion proportion comparisons as well for a more practical and comprehensive evaluation of the segmentation results.

CONCLUSIONS

We present the results from automated deep learning-based pediatric brain tumor subregion segmentation models from 2 different segmentation models, nnU-Net and DeepMedic. nnU-Net achieved excellent results for WT and ET segmentation and decent results for the nonenhancing components including ET, CC, and peritumoral edema.

Disclosure forms provided by the authors are available with the full text and PDF of this article at www.ajnr.org.

REFERENCES

- Ostrom QT, Gittleman H, Xu J, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009–2013. *Neuro Oncol* 2016;18:v1–v75 [CrossRef Medline](#)
- Udaka YT, Packer RJ. Pediatric brain tumors. *Neurol Clin* 2018;36:533–56 [CrossRef Medline](#)
- Manoharan N, Liu KX, Mueller S, et al. Pediatric low-grade glioma: targeted therapeutics and clinical trials in the molecular era. *Neoplasia* 2023;36:100857 [CrossRef Medline](#)
- Deeley MA, Chen A, Datteri R, et al. Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. *Phys Med Biol* 2011;56:4557–77 [CrossRef Medline](#)
- WHO Classification of Central Nervous System Tumours. Lyons, France: International Agency for Research on Cancer; 2021
- Ostrom QT, Price M, Neff C, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019. *Neuro Oncol* 2022;24:v1–v95 [CrossRef Medline](#)
- Gonçalves FG, Viaene AN, Vossough A. Advanced magnetic resonance imaging in pediatric glioblastomas. *Front Neurol* 2021;12:733323 [CrossRef Medline](#)
- Koob M, Girard N. Cerebral tumors: specific features in children. *Diagn Interv Imaging* 2014;95:965–83 [CrossRef Medline](#)
- Quant EC, Wen PY. Response assessment in neuro-oncology. *Curr Oncol Rep* 2011;13:50–56 [CrossRef Medline](#)
- Wen PY, van den Bent M, Youssef G, et al. RANO 2.0: update to the response assessment in neuro-oncology criteria for high- and low-grade gliomas in adults. *J Clin Oncol* 2023;41:5187–99 [CrossRef Medline](#)
- Warren KE, Poussaint TY, Vezina G, et al. Challenges with defining response to antitumor agents in pediatric neuro-oncology: a report from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. *Pediatr Blood Cancer* 2013;60:1397–401 [CrossRef Medline](#)
- Fathi Kazerooni A, Arif S, Madhogaria R, et al. Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: a multi-institutional study. *Neurooncol Adv* 2023;5:vdad027 [CrossRef Medline](#)
- Rudie JD, Rauschecker AM, Bryan RN, et al. Emerging applications of artificial intelligence in neuro-oncology. *Radiology* 2019;290:607–18 [CrossRef Medline](#)
- Huang J, Shlobin NA, Lam SK, et al. Artificial intelligence applications in pediatric brain tumor imaging: a systematic review. *World Neurosurg* 2022;157:99–105 [CrossRef Medline](#)
- Nabavizadeh A, Barkovich MJ, Mian A, et al. Current state of pediatric neuro-oncology imaging, challenges and future directions. *Neoplasia* 2023;37:100886 [CrossRef Medline](#)
- Kang J, Ullah Z, Gwak J. MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers. *Sensors (Basel)* 2021;21:2222 [CrossRef Medline](#)
- Çinar A, Yildirim M. Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture. *Med Hypotheses* 2020;139:109684 [CrossRef Medline](#)

18. Siddique N, Paheding S, Elkin CP, et al. **U-Net and its variants for medical image segmentation: a review of theory and applications.** *IEEE Access* 2021;9:82031–57 [CrossRef](#)
19. Ummadi V. **U-Net and its variants for Medical Image Segmentation: a short review.** *arXiv* 2022. <https://arxiv.org/abs/2204.08470>. Accessed December 6, 2023
20. Peng J, Kim DD, Patel JB, et al. **Deep learning-based automatic tumor burden assessment of pediatric high-grade gliomas, medulloblastomas, and other leptomeningeal seeding tumors.** *Neuro Oncol* 2022;24:289–99 [CrossRef Medline](#)
21. Vafaeikia P, Wagner MW, Hawkins C, et al. **MRI-based end-to-end pediatric low-grade glioma segmentation and classification.** *Can Assoc Radiol J* 2023;75:153–60 [CrossRef Medline](#)
22. Kamnitsas K, Ferrante E, Parisot S, et al. **Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries.** In: *Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and mTOP 2016, Held in Conjunction with MICCAI 2016*. Athens, Greece. October 17, 2016
23. Isensee F, Jaeger PF, Kohl SA, et al. **nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.** *Nat Methods* 2021;18:203–11 [CrossRef Medline](#)
24. Lilly JV, Rokita JL, Mason JL, et al. **The Children's Brain Tumor Network (CBTN): accelerating research in pediatric central nervous system tumors through collaboration and open science.** *Neoplasia* 2023;35:100846 [CrossRef Medline](#)
25. Kazerooni AF, Khalili N, Liu X, et al. **The Brain Tumor Segmentation (BraTS) Challenge 2023: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs).** *ArXiv* 2023
26. Baid U, Ghodasara S, Mohan S, et al. **The RSNA-ASNR-MICCAI BRATS 2021 benchmark on brain tumor segmentation and radiogenomic classification.** *arXi* 210702314 2021 <https://arxiv.org/abs/2107.02314>. Accessed December 6, 2023
27. You W, Mao Y, Jiao X, et al. **The combination of radiomics features and VASARI standard to predict glioma grade.** *Front Oncol* 2023;13:1083216 [CrossRef Medline](#)
28. Rios Velazquez E, Meier R, Dunn WD Jr, et al. **Fully automatic GBM segmentation in the TCGA-GBM dataset: prognosis and correlation with VASARI features.** *Sci Rep* 2015;5:16822 [CrossRef Medline](#)
29. Mitchell JR, Kamnitsas K, Singleton KW, et al. **Deep neural network to locate and segment brain tumors outperformed the expert technicians who created the training data.** *J Med Imaging (Bellingham)* 2020;7:055501 [CrossRef Medline](#)
30. Chen S, Ding C, Liu M. **Dual-force convolutional neural networks for accurate brain tumor segmentation.** *Pattern Recognition* 2019;88:90–100 [CrossRef](#)
31. Isensee F, Jäger PF, Full PM, et al; nnU-Net for brain tumor segmentation. **Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries:** In: *Proceedings of the 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020*. Springer-Verlag; 2021:118–32
32. Chen C, Zhang T, Teng Y, et al. **Automated segmentation of cranio-pharyngioma on MR images using U-Net-based deep convolutional neural network.** *Eur Radiol* 2023;33:2665–75 [CrossRef Medline](#)
33. Kang H, Witanto JN, Pratama K, et al. **Fully automated MRI segmentation and volumetric measurement of intracranial meningioma using deep learning.** *J Magn Reson Imaging* 2023;57:871–81 [CrossRef Medline](#)
34. Boyd A, Ye Z, Prabhu S, et al. **Expert-level pediatric brain tumor segmentation in a limited data scenario with stepwise transfer learning.** *medRxiv* 2023 Accessed December 6, 2023
35. Liu X, Bonner ER, Jiang Z, et al. **From adult to pediatric: deep learning-based automatic segmentation of rare pediatric brain tumors.** In: *Proceedings of 2023 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2023)*. Springer: 2023;15–19
36. Nalepa J, Adamski S, Kotowski K, et al. **Segmenting pediatric optic pathway gliomas from MRI using deep learning.** *Comput Biol Med* 2022;142:105237 [CrossRef Medline](#)