Neuro-Oncology Advances
6(1), vdae108, 2024 | https://doi.org/10.1093/noajnl/vdae108 | Advance Access date 28 June 2024

Early prognostication of overall survival for pediatric diffuse midline gliomas using MRI radiomics and machine learning: A two-center study

Xinyang Liu, Zhifan Jiang, Holger R. Roth, Syed Muhammad Anwar, Erin R. Bonner, Aria Mahtabfar, Roger J. Packer, Anahita Fathi Kazerooni, Miriam Bornhorst, and Marius George Linguraru

All author affiliations are listed at the end of the article

Corresponding Author: Marius George Linguraru, DPhil, Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital, 111 Michigan Avenue NW, Washington, District of Columbia 20010, USA (mlingura@childrensnational.org).

Abstract

Background. Diffuse midline gliomas (DMG) are aggressive pediatric brain tumors that are diagnosed and monitored through MRI. We developed an automatic pipeline to segment subregions of DMG and select radiomic features that predict patient overall survival (OS).

Methods. We acquired diagnostic and post-radiation therapy (RT) multisequence MRI (T1, T1ce, T2, and T2 FLAIR) and manual segmentations from 2 centers: 53 from 1 center formed the internal cohort and 16 from the other center formed the external cohort. We pretrained a deep learning model on a public adult brain tumor data set (BraTS 2021), and finetuned it to automatically segment tumor core (TC) and whole tumor (WT) volumes. PyRadiomics and sequential feature selection were used for feature extraction and selection based on the segmented volumes. Two machine learning models were trained on our internal cohort to predict patient 12-month survival from diagnosis. One model used only data obtained at diagnosis prior to any therapy (baseline study) and the other used data at both diagnosis and post-RT (post-RT study).

Results. Overall survival prediction accuracy was 77% and 81% for the baseline study, and 85% and 78% for the post-RT study, for internal and external cohorts, respectively. Homogeneous WT intensity in baseline T2 FLAIR and larger post-RT TC/WT volume ratio indicate shorter OS.

Conclusions. Machine learning analysis of MRI radiomics has potential to accurately and noninvasively predict which pediatric patients with DMG will survive less than 12 months from the time of diagnosis to provide patient stratification and guide therapy.

Key Points

- Automatic machine learning approach predicts DMG survival at 77%–85% accuracy.
- Homogeneous whole tumor intensity in baseline T2 FLAIR indicates worse prognosis.
- Larger post-RT tumor core/whole tumor volume ratio indicates worse prognosis.

Diffuse midline gliomas (DMG), including diffuse intrinsic pontine gliomas (DIPG), are aggressive central nervous system pediatric tumors located in the brainstem, thalamus, and spinal cord.[1] As one of the most devastating pediatric cancers, DMG represents about 10%–15% of all pediatric tumors of the central nervous system, with an estimated 300 new cases diagnosed annually in the USA.[2] Most DMGs occur between the ages of 5 and 10 years, with a peak at 7 years.[3] There is no curative therapy for DMG, and radiation therapy (RT) is the standard treatment with only transitory benefits.[4] Despite

Importance of the Study

Studies of pediatric diffuse midline gliomas (DMG) prognostication have relied on manual tumor segmentation from MRI, which is impractical and variable in busy clinics. We present an automatic imaging tool based on machine learning to segment subregions of DMG and select radiomic features that predict overall survival. We trained and evaluated our tool on multisequence, 2-center MRIs acquired at the time of diagnosis and post-radiation therapy. Our methods achieved 77%–85% accuracy for DMG survival prediction. The data-driven study identified that homogeneous whole tumor intensity in baseline T2 FLAIR and larger posttherapy tumor core/whole tumor volume ratio indicate worse prognosis. Our tool can increase the utility of MRI for predicting clinical outcome and stratify patients into risk groups. This automated tool has potential to be incorporated in multi-institutional clinical trials to provide consistent and repeatable tumor evaluation.

numerous clinical trials of new agents and novel therapeutic approaches over the last decades,[5] disease outcomes remain dismal with a median overall survival (OS) of less than 1 year, a 2-year OS rate of less than 10%,[6] and a 5-year OS rate of less than 1%.[7]

Magnetic resonance imaging (MRI) is the standard noninvasive test for DMG diagnosis and monitoring of tumor response to therapy. Although pediatric DMGs have a diverse imaging appearance,[8] MRI features have been used to predict H3K27M mutation status[9] and correlate with patient prognosis.[10–15] The features utilized in these studies were either low-dimensional image features[10,11,13–15] or based on texture analysis.[12] The statistical analyses that most of these studies relied on tend to identify inconsistent and inconclusive imaging biomarkers across different studies and data sets. For example, a study of 357 pediatric DIPGs demonstrated that although many MRI features, such as tumor size, enhancement, necrosis, etc., were strongly associated with survival on univariable analysis, very few were significantly associated with survival on multivariable analysis.[11] These findings suggest that only relying on statistical analysis of conventional MRI findings may not be sufficient to predict OS in DMGs.

Machine learning has shown great potential to predict survival or discriminate between certain groups in studies of other brain tumors such as glioblastoma (GBM) and pediatric low-grade gliomas.[16–19] For DMG, machine learning-based regression models were proposed to correlate with patient prognosis based on extracted MRI radiomic features.[20,21] These studies only focused on imaging data at diagnosis, and the tumors were segmented manually, which is generally believed to be time-consuming and to have high interoperator variability. Other studies demonstrated that semiautomated DMG volume measurements are more accurate, prognostically relevant, and consistent than manual measurements.[14,15] In addition to diagnostic scans, it is also important to consider longitudinal data at posttreatment timepoints.[10]

With new therapeutic strategies currently under investigation for DMG, including epigenetic therapy and immunotherapy,[22] there is a great need for noninvasive prognostic imaging tools that can be universally used to accurately identify which patients are at risk for the most rapid deterioration, and thereby assist clinical trial stratification and therapy planning. Such tools should be automatic, objective, and easy to use in multi-institutional clinical trials. With the vast advancements in deep learning techniques, there has been tremendous success in automatic segmentation of brain tumors from MRI, including adult,[23,24] pediatric brain tumors,[25,26] and our previous work of segmenting DMG.[27,28] These advancements have the potential to enable us to create a fully automatic, image-based radiomic analysis, and DMG prognostic tool.

We hypothesize that MRI radiomic features of DMG could be important biomarkers for OS prediction. The purpose of this work is to develop an automatic pipeline to segment subregions of DMG and select MRI features to predict if patient can survive 12 months from diagnosis. Many studies have reported median OS of DMG patients to be approximately 12 months,[4,11] which is also the median OS of our internal cohort. As a first step of quantitative prognostication, accurate prediction if the patient can survive longer or shorter than the median OS could have positive impact on the clinical management of DMG. The proposed method was trained and validated on an internal cohort from Children's National Hospital (CNH) to investigate the accuracy of OS prediction in (1) a baseline study using MRIs obtained at diagnosis prior to any therapy, and (2) a post-RT study using MRIs obtained at both diagnosis and post-RT (ie, within 3 months since the first RT). The method was further tested on an external DMG data set from Children's Hospital of Philadelphia (CHOP) to assess the reproducibility of our findings.

Materials and Methods

Study Cohort

The design of this study was developed with consideration of the CLEAR checklist,[29] which was submitted as Supplementary Material. For this 2-center retrospective study, institutional review board (IRB) approval was obtained at CNH (IRB Protocols #1339 and #14310). The data set from CHOP was obtained through the Children's Brain Tumor Network (CBTN),[30] a research consortium involving multiple institutions, with patients participating under protocols approved by local IRBs. All patients had classic ("typical") DIPG based on radiological imaging defined as T1-hypointense and T2-hyperintense diffusely infiltrative tumors that arise from the pons and involve at least 50% of the pons by cross-sectional area.[31] Our internal cohort from CNH includes 53 pediatric (1 young

adult) patients diagnosed with DMG between 2005 and 2022 (F = 29, M = 24) at CNH. The median patient age at diagnosis is 6.5 years with a range of 3.2–25.9 years. The median OS is 12 months with a range of 3.3–132 months from diagnosis (1 patient is still alive). 23/53 patients underwent biopsy to identify the molecular characteristics of the tumor (1 wild-type, 3 K27M-H3.1, 19 K27M-H3.3). Data of 45/53 patients in our internal cohort were used in our previous publications[27,28] which focused on developing deep learning models for DMG segmentation. This study included a larger number of patients and focused on predicting patient survival.

The external cohort from CHOP includes 16 pediatric patients diagnosed with DMG between 2005 and 2022 (F = 9, M = 7), made available by CBTN. The median age at diagnosis is 9.4 years with a range of 3.8–18.2 years. The median OS is 9.6 months with a range of 1.3–27.1 months from diagnosis. None of the patients underwent biopsy. The differences in median age and OS between the 2 cohorts can be used to analyze whether the model trained on our internal cohort generalizes effectively to the external cohort. Sample size of the 2 cohorts was determined based on availability of MRI data.

MRI Data

Both institutions used scanners and imaging protocols that varied among patients and timepoints because of retrospective data collection. For each patient, 4 MRI sequences at diagnosis and/or post-RT were collected including T1-weighted (T1), contrast-enhanced T1 (T1ce), T2-weighted (T2), and T2-weighted-Fluid-Attenuated Inversion Recovery (T2 FLAIR). The MRIs were acquired either on 1.5T or 3T magnet, with 2D or 3D acquisition protocols, using scanners from GE Healthcare, Siemens AG, or Toshiba. T1 and T1ce MRIs included T1 SE, T1 FSE, T1 MPRAGE, or T1 SPGR. T2 MRI included T2 SE, T2 FSE, T2 FRFSE, or T2 propeller. T2 FLAIR MRI included those with or without gadolinium (Gd) enhancement. The slice thickness range was 0.5–6 mm and matrix range was (256–512) × (256–512) pixels. All images were collected in the DICOM image format.

Manual segmentation of DMG volumes was used as the ground truth for training the deep learning segmentation model. All segmentations were conducted by trained laboratory personnel (E.R.B. and K.B.) using ITK-SNAP[32] and were reviewed by a neuro-oncologist (M.B.). A representative subset of segmentations was jointly reviewed by a neuroradiologist (Gilbert Vezina) and neuro-oncologist (M.B.) to ensure adequacy of the segmentation technique and accuracy of the data. Because necrosis/cyst is not consistently identifiable for DMG, 2 labels were created: tumor core (TC) and whole tumor (WT). TC included 2 components: the Gd-enhancing tumor appearing as enhancement on T1ce, and the necrotic/cystic core appearing as hypointense on T1ce. WT includes TC and the peritumoral edematous/infiltrated tissue appearing as abnormal hyperintense signal on T2 FLAIR.

Automatic DMG Segmentation

Despite the success of deep learning-based automatic segmentation for adult GBMs, the direct application of these methods on rare pediatric brain tumors remains challenging.[33] While GBMs and DMGs are both high-grade brain gliomas, they have distinctive characteristics. GBMs typically locate in frontal/temporal lobe, whereas DMGs typically locate in the pons. Necrosis is common in GBMs but is rare/unclear in DMGs. Our approach was to transfer knowledge learnt from GBM segmentation to DMG segmentation.

The Brain Tumor Segmentation (BraTS) challenge is an ongoing annual event that has been held since 2012. We obtained imaging data of 1251 GBM patients that was publicly available from BraTS.[34] For each patient, 4 MRI sequences (T1, T1ce, T2, and T2 FLAIR) and manual segmentations of TC and WT subregions of GBM were provided. The winning method of the BraTS 2020 challenge was based on nnU-Net,[24] a popular and robust semantic deep-learning segmentation method. nnU-Net analyzes the training data and automatically configures a matching U-Net[35]-based segmentation pipeline.

Figure 1 shows the model architecture of our transfer learning-based approach using nnU-Net. It includes a pretraining phase of nnU-Net using the GBM data set. Because nnU-Net automatically determines the segmentation pipeline based on the specific data set, we changed this pretraining paradigm to first design the segmentation pipeline based on the DMG data set, and then used the planned pipeline to perform pretraining on the GBM data. The pretrained network weights were then used as initialization to finetune the model using the DMG data set. Preprocessing was performed in an automatic fashion and included N4 bias correction to correct for MRI inhomogeneities,[36] rigid registration to the SRI-24 Atlas for spatial alignment,[37] and skull stripping.[38] The output of the segmentation model was the TC and WT volumes, which were used as input to the radiomic feature extraction step.

Experiments and Evaluation for Tumor Segmentation

Data from 45 CNH patients (with manual segmentation) were used for training and validation of the segmentation model. Because tumor regions varied considerably between images at diagnosis and post-RT of the same patient, the images at the 2 timepoints were considered as separate sets for the purpose of segmentation. This yielded a total of 82 sets from the 45 patients (not all patients had qualified images or manual segmentations for both timepoints). Specifically, 41/82 sets were acquired at diagnosis, 34/82 sets were acquired within 1-month post-RT, and the rest of 7 sets were acquired 2–4 months post-RT.

The 82 DMG sets were randomly divided into 5 folds, and 5-fold cross-validation was performed to obtain the TC and WT volumes. Images at 2 timepoints of the same patient were kept in the same fold. Dice coefficient and volume similarity were used as evaluation metrics to compare the predicted and ground truth segmentations, where the volume similarity is calculated as the ratio between the smaller of the compared volumes and the average of the compared volumes.[39]

After 5-fold cross-validation, we trained a final model with all 82 sets and used it to predict TC and WT volumes for the remaining 8 internal patients (these cases did not have manual segmentations) and 16 external patients

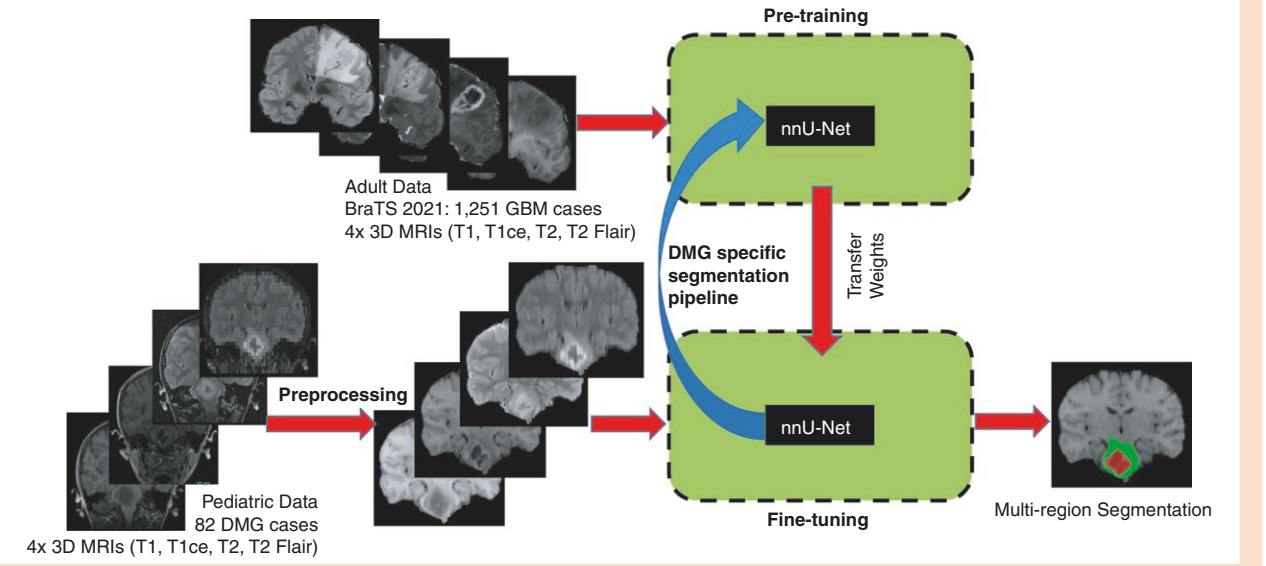

Figure 1. Model architecture of our DMG segmentation method, which employs nnUnet-based pre-training and fine-tuning. The input to pre-training is the adult brain tumor data set from the BraTS 2021 challenge. The input to fine-tuning is the preprocessed DMG data set. Based on the DMG data set, a specific segmentation pipeline is determined and used for pre-training. After pre-training, the obtained weights are used as input for fine-tuning. The output of the model is multi-region segmentation masks.

(14 cases has manual segmentations). For evaluation of automatic segmentation, 5-fold cross-validation of the 82 internal cases and test on the 14 external cases were reported.

Many DMG cases do not have or have very small TC volumes. Thus, comparison between predicted and ground truth in small or absent TC volumes produces extreme metrics (eg, Dice score of 0 or 1). To void bias to small volumes, we cleaned predicted volumes by removing small (ie, <130 mm^3) disconnected regions. Moreover, we let TC/WT denote the ratio between TC volume and WT volume. We did not evaluate TC segmentation performance if $0 < TC/WT < 4\%$ for both predicted and ground truth segmentations. If TC/WT = 0 for both, the metrics were set to be 1. As a result, evaluation of TC segmentation for 7/82 internal cases and 3/14 external cases was omitted. The thresholds of 130 voxels and 4% were determined by a previous study on the pediatric brain tumor data.[40,41]

Radiomic Feature Extraction

Based on automatically segmented DMG volumes, we used the open-source PyRadiomics software[42] with default configuration to extract radiomic features from the original images. These features included 13 volumetric and shape features and 91 gray level features. Please refer to Supplemental Appendix S1 for a complete list of features. The gray level features included: 18 first-order features, 22 gray level co-occurrence matrix (GLCM) features, 16 gray level size zone matrix (GLSZM) features, 16 gray level run length matrix (GLRLM) features, 5 neighboring gray tone difference matrix (NGTDM) features, and 14 gray level dependence matrix (GLDM) features. In addition, we added 2 demographic features (ie, sex and age), and 2 volumetric features of interest (ie, brain volume and relative tumor volume [DMG volume divided by brain volume]). Because gray level features are susceptible to inter-scanner variation due to different acquisition protocol,[43] image gray levels were normalized by removing the mean and scaling to unit variance before the features were calculated.

The baseline study employed 401 features, including sex, age, 35 volumetric and shape features, and 4 sets of 91 gray level features (1 set for each MRI sequence). The volumetric and shape features include brain volume, 14 WT features (ie, 13 from PyRadiomics and relative DMG volume), 10 TC features, and 10 features for the ratio between TC and WT (TC/WT). Because many DMG cases do not have TC volume, 4 features (ie, elongation, flatness, surface area to volume ratio, and sphericity) having measurements of TC in the denominator of their calculation were excluded. The gray level features were calculated based on WT segmentations.

The post-RT study employed 1576 features, including sex, age, 118 volumetric and shape features, and 1456 gray level features. The volumetric and shape features include brain volumes at diagnosis and post-RT, 28 WT features (14 at diagnosis and 14 post-RT), changes of 14 WT features (post-RT values minus values at diagnosis), relative changes of 14 WT features (changes divided by values at diagnosis), 20 TC features (10 at diagnosis and 10 post-RT), changes of 10 TC features, 20 TC/WT features (10 at diagnosis and 10 post-RT), and changes of 10 TC/WT features. We did not include relative changes of TC and TC/WT features because measurements related to TC at diagnosis could be null, which would make the definition of relative change invalid. The gray level features included 4 sets of 91 gray level features at diagnosis, 4 sets of 91 gray level features post-RT, changes of 4 sets of 91 gray level features, and relative changes of 4 sets of 91 gray level features.

Feature Selection

Feature selection was performed on the training data prior to prediction to avoid overfitting. In the first step, feature filtering was performed using the Mann–Whitney *U* test comparing feature values between short OS (<1 year) and long OS (≥1 year). Sixty-nine features with $P < .05$ were selected for the post-RT study. For the baseline study, because there was only 1 feature with $P < .05$, we selected 10% of all features (40 features) with the smallest *P* values. Sequential feature selection was then performed on the filtered features to select the optimal number of discriminative features for each study. Starting from none, the algorithm added 1 feature at an iteration to form a feature subset until the desired number of features (which we capped at 10% of the number of patients to avoid overfitting the model to the training data) was reached. At each iteration, the algorithm went through each feature not currently in the feature subset and chose the feature to add such that the new feature subset achieved the best accuracy in the leave-one-out cross-validation. Specifically, we trained a linear support vector machine (SVM) to classify between short OS and long OS using all subjects in our internal cohort except for 1, which was used for testing. This process was repeated iteratively until all patients were tested. Because of our small data sets, we used leave-one-out cross-validation to maximize the number of training examples, and employed the linear kernel for SVM, which is less prone to overfitting than nonlinear kernels.

Experiments and Evaluation for OS Prediction

Images at diagnosis of 52/53 CNH patients were used for training and validation in the baseline study. There were 26/52 patients with short OS, that is, survival shorter than 12 months from diagnosis. One patient did not have images of all 4 MRI sequences at diagnosis, but the post-RT images were used for training the segmentation model. Images at diagnosis and within 3 months post-RT of 41/52 patients were available and used for training and validation in the post-RT study. There were 22/41 patients with short OS.

After feature selection, the final SVM model was trained with all internal patients with the selected features. Validation of the final model on the internal data set was reported. The final model was used to predict OS based on the same selected features on the external data set. For the baseline study, 16 external patients (9 had short OS) were tested. 9/16 external patients who had post-RT imaging (< 3 months) were tested in the post-RT study. 4/9 patients had short OS.

Results

Segmentation Results

Table 1 shows performance of the automatic DMG segmentation method evaluated on the internal and external data sets. The external evaluation shows performance on out-of-distribution data to reflect generalizability based on scanning and protocol variability and tumor heterogeneity.[44] Metrics of WT segmentation for the external cohort (0.86 mean Dice score and 0.91 mean volume similarity) were similar to those obtained for the internal cohort (0.88 mean Dice score [Mann–Whitney *U* test $P = .10$] and 0.93 mean volume similarity [$P = .13$]). This suggests our method can be successfully generalized for segmenting WT volume of images from different sources. Similarly, metrics of TC segmentation for the external cohort (0.74 mean Dice score and 0.81 mean volume similarity) were similar, although inferior to those obtained for the internal cohort (0.91 mean Dice score [$P = .10$] and 0.94 mean volume similarity [$P = .58$]).

Figure 2 shows qualitative segmentation results on the diagnosis and post-RT images of a DMG patient of the internal cohort. The Dice scores for this case were 0.92 (diagnostic TC), 0.92 (diagnostic WT), 0.97 (post-RT TC), and 0.93 (post-RT WT), which were approximately the median Dice scores for our internal cohort (0.94 for TC and 0.91 for WT in Table 1).

OS Prediction Results

Table 2 shows results of the proposed OS prediction method. Our classification results were reported in 2 ways: emphasizing accuracy, a balanced measure combining sensitivity and specificity, and the other emphasizing sensitivity, potentially more clinically useful for identifying patients with short OS. For accuracy-focused results (first row), the specificity range of 71%–73% is reasonable. For sensitivity-focused results (second and third rows), we achieved a superb sensitivity of 92%–100%. However, when prioritizing sensitivity, specificity tends to be lower, illustrating the typical tradeoff between these metrics. Our model demonstrated better performance on sensitivity than specificity, possibly due to the relatively small size of our training data set. Our results also suggest that adding post-RT data may improve prediction accuracy and sensitivity over the baseline. Despite the small data cohort, the evaluation metrics on our external cohort were generally comparable to those obtained on the internal cohort, indicating overall generalizability of our machine learning predictive model.

Table 1. Mean (median) and Standard Deviation of Dice Coefficient and Volume Similarity Calculated by Comparing Predicted Tumor Core (TC) and Whole Tumor (WT) Volumes and Those Segmented Manually. Results Shown Include Validation on the Internal Cohort (From Children's National Hospital) and Testing on the External Cohort (From Children's Hospital of Philadelphia)

Evaluation data set	TC Dice	WT Dice	TC vol. similarity	WT vol. similarity
Internal cohort	0.91 (0.94) ± 0.12	0.88 (0.91) ± 0.07	0.94 (0.99) ± 0.10	0.93 (0.96) ± 0.07
External cohort	0.74 (0.83) ± 0.32	0.86 (0.89) ± 0.06	0.81 (0.99) ± 0.34	0.91 (0.93) ± 0.07

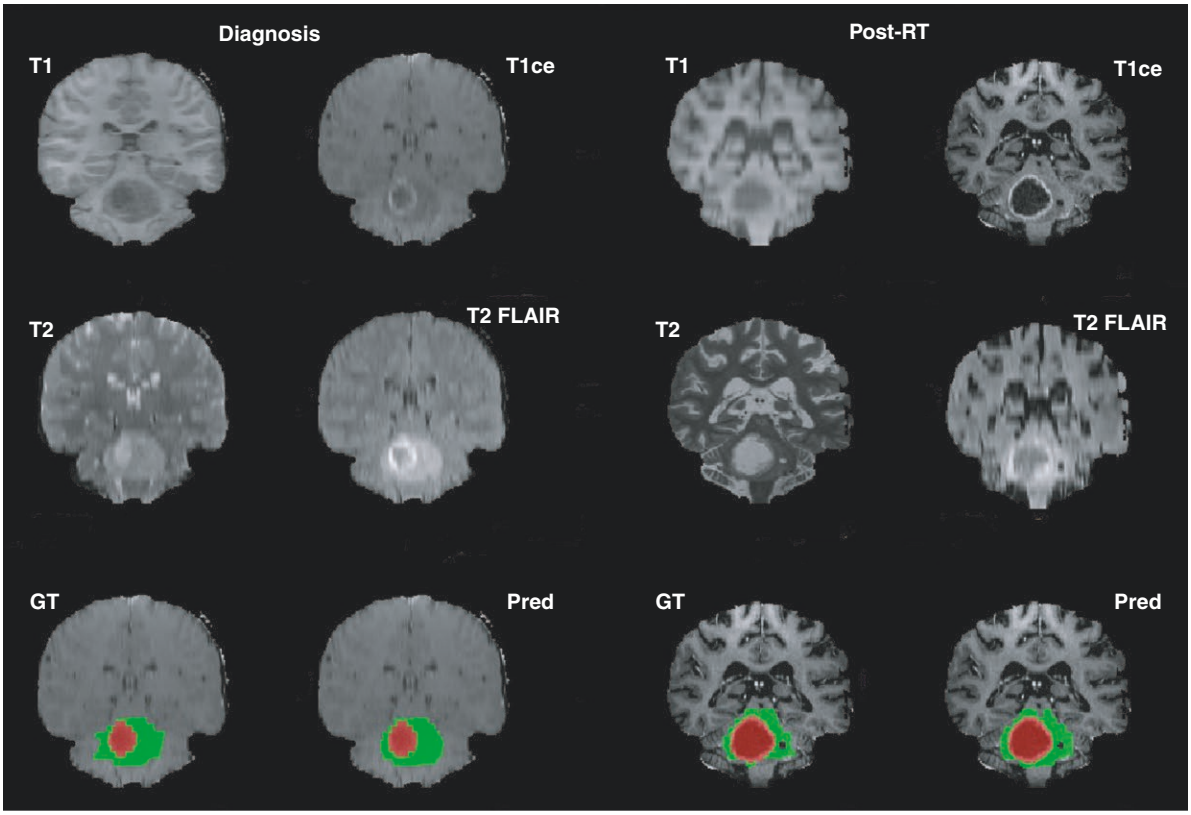

Figure 2. Qualitative segmentation results on the diagnosis and post-RT images of a DMG patient from the internal cohort. The figure shows 4 MRI sequences after preprocessing, the ground truth (GT) segmentation, and the predicted (Pred) segmentation of 2 regions: the tumor core, which is enclosed by the whole tumor.

Table 2. Results of the Proposed OS Prediction Method. Short OS of Less Than 1 Year is Considered Positive. We Present Results at the Operating Points of Maximum Accuracy and Maximum Sensitivity

Study	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
	Internal cohort (52 subjects)			External cohort (16 subjects)		
Baseline max accuracy	77%	81%	73%	81%	89%	71%
Baseline max sensitivity	62%	92%	31%	75%	100%	43%
	Internal cohort (41 subjects)			External cohort (9 subjects)		
Post-RT max accuracy[a]	85%	100%	68%	78%	100%	60%

[a]Also max sensitivity.

The number of selected features for the baseline and post-RT studies was 5 and 4, respectively. We list below the selected features for each study, along with their interpretation for prediction and the *P* values of Mann–Whitney *U* test between short and long OS computed on our internal cohort. The features are listed in the order of their relevance to OS prediction.

The 5 selected features for the baseline study were:

GLCM Information measure of correlation (Imc1) on T2 FLAIR ($P = .118$): quantifies the complexity of the texture. It ranges from –1 to 0 and the higher the value the more complex in texture.

GLSZM High gray level zone emphasis on T1 ($P = .231$): measures the distribution of the higher gray level values.

The median gray level value on T2 FLAIR ($P = .173$)

Skewness on T2 ($P = .061$): measures the asymmetry of the distribution of gray level values about the mean value.

The 10th percentile of gray level value on T2 FLAIR ($P = .217$)

The significant feature for the baseline study was the GLCM Cluster Shade on T2, which is a measure of skewness and uniformity ($P = .009$). However, our feature selection algorithm did not select this feature. This verifies our method selects features that perform best in combination

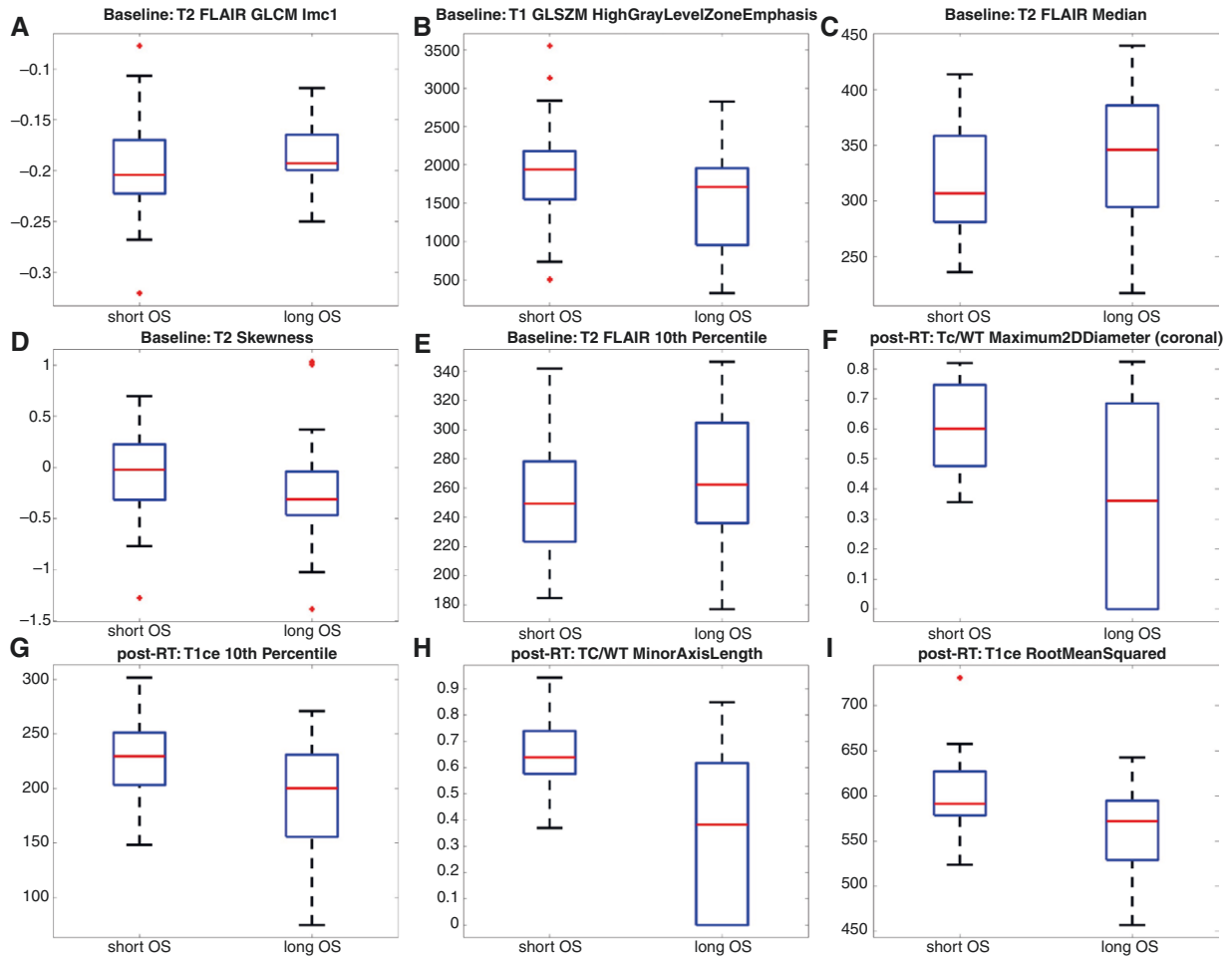

Figure 3. Comparison between short OS and long OS for the selected features of the baseline (A–E) and the post-RT (F–I) studies. Data of both internal and external cohorts were considered. (A) GLCM Imc1on T2 FLAIR; (B) GLSZM high gray level zone emphasis on T1; (C) median gray level on T2 FLAIR; (D) skewness on T2; (E) 10th percentile gray level on T2 FLAIR; (F) the ratio between TC and WT for maximum 2D diameter in the coronal plane; (G) 10th percentile gray level on T1ce; (H) the ratio between TC and WT for minor axis length; and (I): root mean square of gray level on T1ce.

in the machine learning approach, but not necessarily the features with the smallest *P* values.

The 4 selected features for the post-RT study were:

The ratio of maximum 2D diameter (coronal plane) between post-RT TC and post-RT WT ($P = .017$). The maximum 2D diameter is the largest pairwise Euclidean distance between tumor surface mesh vertices on a 2D plane.

The 10th percentile of gray level value on post-RT T1ce ($P = .027$).

The ratio of minor axis length between post-RT TC and post-RT WT (*P*.002). The minor axis length is the second-largest axis length of principal component analysis performed on the volume.

Root mean squared on post-RTT1ce (*P*=.006): is the square-root of the mean of all the squared gray level values. This is a measure of the magnitude of gray level values.

The algorithm selected the 10th percentile of gray level value as an important feature for both studies, in combination with the other chosen features. Figure 3 shows the comparison between short OS and long OS for the selected features of the 2 studies. A visual example of radiomics is shown in Figure 4.

Discussion

The analysis of brain tumors on MRI, and especially of rare pediatric tumors, has been challenged by small data cohorts acquired by different scanners and imaging protocols, and by manual segmentations with interobserver variability. Machine learning models have the potential to extract complex imaging patterns, provide automation, and standardization for the analysis, and support the evaluation of clinical trials—and ultimately of patient therapy—with repeatable and consistent data.

To our best knowledge, this study is the first to report a fully automatic, machine learning-based model to

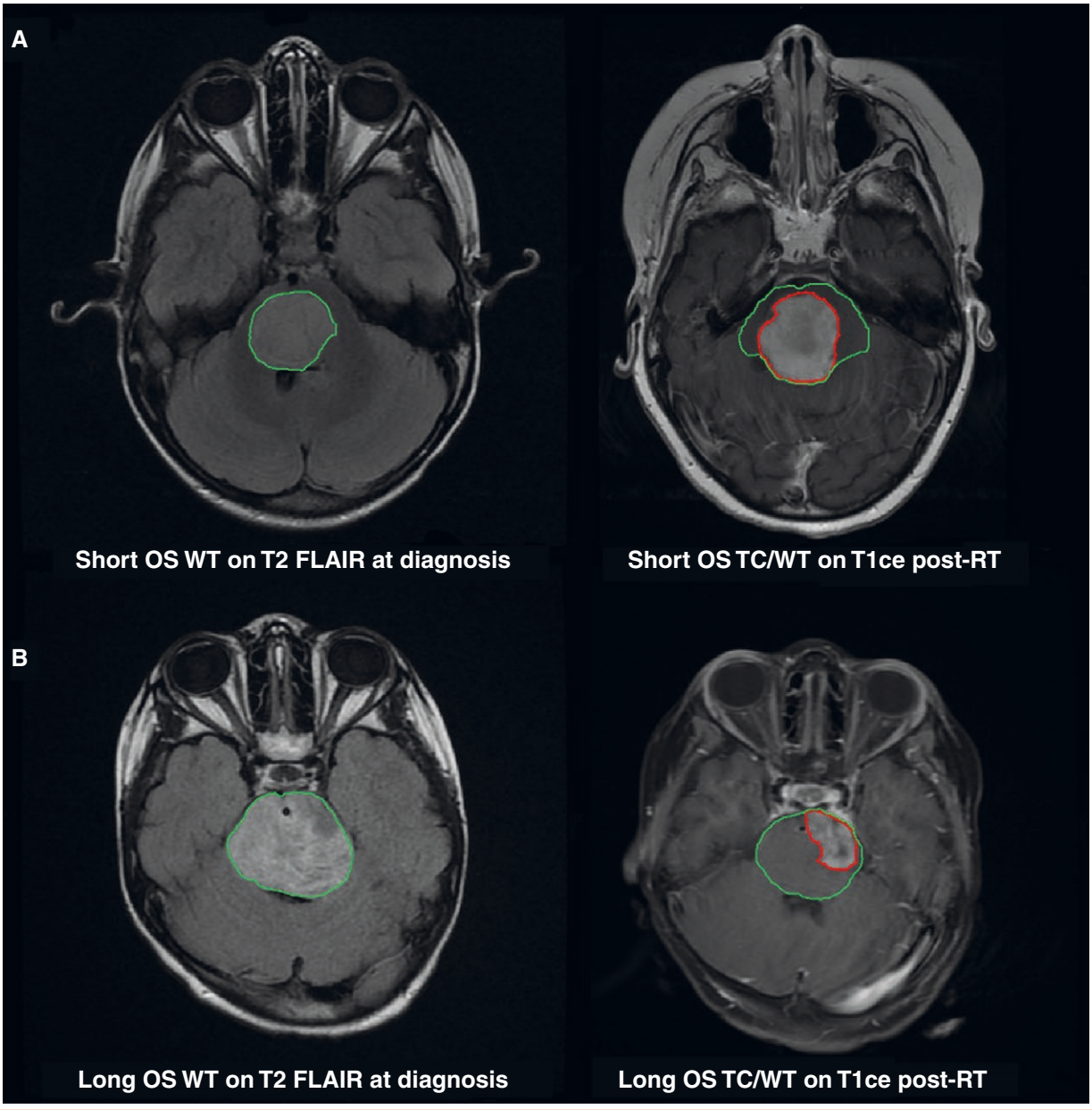

Figure 4. MRI of 2 patients who survived 8 months (A, short OS) and 14 months (B, long OS) from our internal cohort. Manual segmentations were outlined (inner contour: tumor core [TC], outer contour: whole tumor [WT]). Diagnostic T2 FLAIR suggests in WT intensity distribution is more homogeneous and intensity values are relatively lower (in contrast to nearby tissues) for short OS compared with long OS. Post-RT T1ce suggests the TC/WT ratio of short OS is larger than that of long OS. These observations were consistent with our findings in the selected features (Figure 3).

prognosticate DMG survival using MRI features. Our automatic DMG segmentation method generated accurate TC and WT segmentations. The mean Dice scores of 0.91 for TC and 0.88 for WT obtained on the internal cohort were comparable to those reported for adult GBM segmentation using state-of-the-art deep learning models.[45,46] The results on the external cohort were similar for WT, an indication of model generalizability and robustness when applied to independent data with different imaging and patient characteristics. Although results were inferior for TC segmentation for the external cohort (mean Dice = 0.74), they were comparable to the 0.62–0.74 Dice scores reported in a recent study of automatic segmentation of subregions of pediatric brain tumors.[26] Our results are also comparable with results of the winning method (TC Dice = 0.78, WT Dice = 0.82)[40] for the pediatric BraTS challenge 2023.[41]

A recent study based on manual tumor segmentation presented a machine learning-based regression model to correlate MRI radiomic features with DIPG prognosis.[20] The study employed T1ce and T2 MRI acquired at diagnosis, and found that homogeneous tumor pixel intensity or texture, such as the GLCM features, conferred a shorter OS.

A similar pattern was found in our baseline study, where tumors in the short OS group tend to have more homogeneous gray level distribution (ie, smaller value of GLCM Imc1) as shown in Figures 3A and 4.

Although diagnostic features were considered in the post-RT study, all the selected features in the post-RT study were related to post-RT measurements. Tumor volumetric and shape features, which are independent of scanner variation, were selected for the post-RT study, whereas no shape feature was selected for the baseline study. These results suggest post-RT features are more discriminative and potentially more robust compared with diagnostic features. The 2 selected shape features in the post-RT study indicate that larger post-RT TC/WT ratio predicts OS shorter than 12 months (Figure 3F and H). For both baseline and post-RT studies, our method predicted short OS with high sensitivity and specificity for both internal and external cohorts.

Our study is not without limitations. Both of our internal and external cohorts are small data sets, which is a challenge for studies of rare diseases. The findings of this study should be further verified with a larger DMG data set. Given the data size, we used machine learning predictors based on SVM, which perform better on small cohorts and offer feature interpretability. Better DMG segmentation and OS prediction models can be achieved by training on a larger data set, and the fully automatic nature of the proposed method is well suited for such large multi-institutional collaboration. Another potential limitation is the fact that radiomics are susceptible to bias and variation due to inter-scanner factors such as different acquisition protocols. We addressed this limitation by normalizing the distribution of gray level values. Additional feature harmonization methods besides what was performed in our study could be used to remove scanner effects in brain MRI radiomic features.[43,47] One other limitation of our study is that we only classified if the patient could live longer or shorter than 12 months. Adding more clinically relevant timepoints, or directly predicting the patient's lifespan in terms of months, could be natural extensions of this study. Another restriction of our work is that 18/52 cases we used for classification had Gd-enhanced T2 FLAIR, evenly split between 9 with short OS and 9 with long OS. This balanced distribution suggests that the inclusion of Gd-enhanced T2 FLAIR had a limited effect on our results. Additionally, all cases in the external cohort had T2 FLAIR without enhancement. The comparable results across both cohorts further support the minimal impact of Gd enhancement on the study outcomes.

In conclusion, we presented a fully automatic machine learning-based approach to compute radiomic biomarkers of DMGs from multisequence MRI. The approach can accurately and noninvasively predict OS for DMG patients and can be extended to other rare pediatric brain tumors. Our approach offers several advantages over the current standards of evaluation of pediatric brain tumors on MRI. Quantitative image analysis, including volumetrics of tumor components, can support the evaluation of tumor progression and response to treatment. Early prognostication of OS can guide patient risk stratification and clinical decisions. Specifically, the results of this study would allow for a tissue-agnostic stratification of patients where upfront treatment could be modified based on radiomic features to ensure maximum intervention for patients with a predicted survival less than 12 months compared to patients who may have a longer survival. Future studies incorporating radiomic features and genomic findings will allow for a more precise radiogenomic stratification, with the hope to decrease the need for biopsies of the tumor in the future. With automated and standardized analysis, the machine learning tool can also provide data-driven evidence to clinical trials.

Supplementary material

Supplementary material is available online at *Neuro-Oncology Advances* (https://academic.oup.com/noa).

Keywords

diffuse midline glioma | magnetic resonance imaging | machine learning | overall survival | radiomics

Funding

This work was partially supported by the National Cancer Institute [award number 5UH3CA236536-04].

Acknowledgments

The authors would like to acknowledge Dr. Javad Nazarian, PhD, who is the PI of IRB Pro #1339 that was used to identify patients from Children's National Hospital for this study, and to the Children's Brain Tumor Network for making available the data from Children's Hospital of Philadelphia. The authors would like to acknowledge Kristen Bougher, who conducted a subset of manual segmentations, and Dr. Gilbert Vezina, neuro-radiologist, who reviewed a subset of the manual segmentations with the research team to ensure accuracy with measurements. The authors would also like to acknowledge the patients and their families who consented to this research.

Conflict of interest statement

M.B. serves on an External Advisory Board for Alexion. There is no conflict with this manuscript. The remaining authors have no conflicts to declare.

Authorship statement

Conception and design: X.L., Z.J., M.B., and M.G.L. Data acquisition: X.L., E.R.B., A.M., R.J.P., A.F.K., and M.B. Data analysis/interpretation: X.L., Z.J., H.R.R., S.M.A., and M.G.L. Drafting/revising: all. Final approval: all.

Data Availability

Data of external cohort can be requested via Children's Brain Tumor Network (cbtn.org). A preprint version of this manuscript has been deposited on medRxiv (https://doi.org/10.1101/2023.11.01.23297935).

Affiliations

Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital, Washington, District of Columbia, USA (X.L., Z.J., S.M.A., M.G.L.); School of Medicine and Health Sciences, George Washington University, Washington, District of Columbia, USA (S.M.A., M.B., M.G.L.); Center for Genetic Medicine Research, Children's National Hospital, Washington, District of Columbia, USA (E.R.B., M.B.); Brain Tumor Institute, Children's National Hospital, Washington, District of Columbia, USA (R.J.P., M.B.); Center for Cancer and Blood Disorders, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois, USA (M.B.); NVIDIA, Santa Clara, California, USA (H.R.R.); Center for Data-Driven Discovery in Biomedicine (D3b), Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA (A.F.K., A.M.); Department of Neurosurgery, University of Pennsylvania, Philadelphia, Pennsylvania, USA (A.F.K.); Center for AI and Data Science for Integrated Diagnostics (AI2D), Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, Pennsylvania, USA (A.F.K.)

References

1. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *NeuroOncol.* 2021;23(8):1231–1251.
2. Warren KE. Diffuse intrinsic pontine glioma: poised for progress. *Front Oncol.* 2012;2:205.
3. Di Ruscio V, Del Baldo G, Fabozzi F, et al. Pediatric diffuse midline gliomas: an unfinished puzzle. *Diagnostics (Basel).* 2022;12(9):2064.
4. Hoffman LM, DeWire M, Ryall S, et al. Spatial genomic heterogeneity in diffuse intrinsic pontine and midline high-grade glioma: implications for diagnostic biopsy and targeted therapeutics. *Acta Neuropathol Commun.* 2016;4:1.
5. Espirito Santo V, Passos J, Nzwalo H, Nunes S, Salgado D. Remission of pediatric diffuse intrinsic pontine glioma: case report and review of literature. *J Pediatr Neurosci.* 2021;16(1):1–4.
6. Rashed WM, Maher E, Adel M, Saber O, Zaghloul MS. Pediatric diffuse intrinsic pontine glioma: where do we stand? *Cancer Metastasis Rev.* 2019;38(4):759–770.
7. Hayden E, Holliday H, Lehmann R, et al. Therapeutic targets in diffuse midline gliomas—an emerging landscape. *Cancers (Basel).* 2021;13(24):6251.
8. Aboian MS, Solomon DA, Felton E, et al. Imaing characteristics of petriatric diffuse midline gliomas with histone H3 K27M mutation. *Am J Neuroradiol.* 2017;38(4):795–800.
9. Chauhan RS, Kulanthaivelu K, Kathrani N, et al. Prediction of H3K27M mutaiton status of diffuse midline gliomas using MRI features. *J Neuroimaging.* 2021;31(6):1201–1210.
10. Löbel U, Hwang S, Edwards A, et al. Discrepant longitudinal volumetric and metabolic evolution of diffuse intrinsic pontine gliomas during treatment: implications for current response assessment strategies. *Neuroradiology.* 2016;58(10):1027–1034.
11. Leach JL, Roebker J, Schafer A, et al. MR imaging features of diffuse intrinsic pontine glioma and relationship to overall survival: report from the International DIPG Registry. *Neuro Oncol.* 2020;22(11):1647–1657.
12. Szychot E, Youssef A, Ganeshan B, et al. Predicting outcome in childhood diffuse midline gliomas using magnetic resonance imaging based texture analysis. *J Neuroradiol.* 2021;48(4):243–247.
13. Zhu X, Lazow MA, Schafer A, et al. A pilot radiogenomic study of DIPG reveals distinct subgroups with unique clinical trajectories and therapeutic targets. *Acta Neuropathol Commun.* 2021;9(1):14.
14. Gilligan LA, DeWire-Schottmiller MD, Fouladi M, DeBlank P, Leach JL. Tumor response assessment in diffuse intrinsic pontine glioma: comparison of semiautomated volumetric, semiautomated linear, and manual linear tumor measurement strategies. *Am J Neuroradiol.* 2020;41(5):866–873.
15. Lazow MA, Nievelstein MT, Lane A, et al. Volumetric endpoints in diffuse intrinsic pontine glioma: comparison to cross-sectional measures and outcome correlations in the International DIPG/DMG Registry. *Neuro Oncol.* 2022;24(9):1598–1608.
16. Chang K, Zhang B, Guo X, et al. Multimodel imaging patterns predict survival in recurrent globlastoma patients treated with bevacizumab. *Neuro Oncol.* 2016;18(12):1680–1687.
17. Wagner MW, Hainc N, Khalvati F, et al. Radiomics of pediatric low-grade gliomas: toward a pretherapeutic differentiation of BRAF-mutated and BRAF-fused tumors. *Am J Neuroradiol.* 2021;42(4):759–765.
18. Li G, Li L, Li Y, et al. An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. *Brain.* 2022;145(3):1151–1161.
19. Moassefi M, Faghani S, Conte GM, et al. A deep learning model for discriminating true progression from pseudoprogression in glioblastoma patients. *J Neurooncol.* 2022;159(2):447–455.
20. Tam LT, Yeom KW, Wright JN, et al. MRI-based radiomics for prognosis of pediatric diffuse intrinsic pontine glioma: an international study. *Neurooncol Adv.* 2021;3(1):vdab042.
21. Wagner MW, Namdar K, Napoleone M, et al. Radiomic features based on MRI predict progression-free survival in pediatric diffuse midline glioma/diffuse intrinsic pontine glioma. *Can Assoc Radiol J.* 2023;74(1):119–126.
22. Long W, Yi Y, Chen S, et al. Potential new therapies for pediatric diffuse intrinsic pontine glioma. *Front Pharmacol.* 2017;8:495.
23. Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. *Proceedings of International MICCAI Brainlesion Workshop.* 2018;311–320.
24. 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(2):203–211.
25. Madhogarhia R, Kazerooni AF, Arif S, et al. Automated segmentation of pediatric brain tumors based on multi-parametric MRI and deep learning. *Proc SPIE Med Imaging.* 2022;12033:737–745.
26. Kazerooni AF, Arif S, Madhogarhia 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(1):1–12.
27. Liu X, Bonner ER, Jiang Z, et al. From adult to pediatric: deep learning-based automatic segmentation of rare pediatric brain tumors. *Proc SPIE Med Imaging.* 2023;12465.

28. Liu X, Bonner ER, Jiang Z, et al. Automatic segmentation of rare pediatric brain tumors using knowledge transfer from adult data. *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*. 2023;1–4.
29. Kocak B, Baessler B, Bakas S, et al. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. *Insights Imaging*. 2023;14(1):75.
30. 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 (New York, N.Y.)*. 2023;35:100846.
31. Barkovich AJ, Krischer J, Kun LA, et al. Brain stem gliomas: a classification system based on magnetic resonance imaging. *Pediatr Neurosurg*. 1990;16(2):73–83.
32. Paul AY, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116–1128.
33. Drai M, Testud B, Brun G, et al. Borrowing strength from adults: transferability of AI algorithms for paediatric brain and tumour segmentation. *Eur J Radiol*. 2022;151:110291.
34. Baid U, Ghodasara S, Mohan S, et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv:2107.02314*. 2021.
35. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Proc MICCAI*. 2015;9351:234–241.
36. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310–1320.
37. Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. *Hum Brain Mapp*. 2010;31(5):798–819.
38. Thakur S, Doshi J, Pati S, et al. Brain extraction on MRI scans in presence of diffuse glioma: multi-institutional performance evaluation of deep learning methods and robust modality-agnostic training. *Neuroimage*. 2020;220:117081.
39. Cardenes R, Luis-Garcia R, Bach-Cuadra M. A multidimensional segmentation evaluation for medical image data. *Comput Methods Programs Biomed*. 2009;96(2):108–124.
40. Capellan-Martin D, Jiang Z, Parida A, et al. Model ensemble for brain tumor segmentation in magnetic resonance imaging. *International MICCAI Brainlesion Workshop*. In press.
41. 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:2305.17033*. 2023.
42. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–e107.
43. Li Y, Ammari S, Balleyguier C, Lassau N, Chouzenoux E. Impact of preprocessing and harmonization methods on the removal of scanner effects in brain MRI radiomic features. *Cancers (Basel)*. 2021;13(12):3000.
44. Prabhudesai S, Wang NC, Ahluwalia V, et al. Stratification by tumor grade groups in a holistic evaluation of machine learning for brain tumor segmentation. *Front Neurosci*. 2021;15:740353.
45. Isensee F, Jaeger PF, Full PM, et al. nnU-Net for brain tumor segmentation. *International MICCAI Brainlesion Workshop*. 2020;118–132.
46. Aboian M, Bousabarah K, Kazarian E, et al. Clinical implementation of artificial intelligence in neuroradiology with development of a novel workflow-efficient picture archiving and communication system-based automated brain tumor segmentation and radiomic feature extraction. *Front Neurosci*. 2022;16:860208.
47. Stamoulou E, Spanakis C, Manikis GC, et al. Harmonization strategies in multicenter MRI-based radiomics. *J Imaging*. 2022;8(11):303.