# RESEARCH



# Comparative analysis of deep learning and radiomic signatures for overall survival prediction in recurrent high-grade glioma treated with immunotherapy



Qi Wan<sup>1,2\*</sup>, Clifford Lindsay<sup>3</sup>, Chenxi Zhang<sup>4</sup>, Jisoo Kim<sup>5</sup>, Xin Chen<sup>6</sup>, Jing Li<sup>7</sup>, Raymond Y. Huang<sup>5</sup>, David A. Reardon<sup>8</sup>, Geoffrey S. Young<sup>5\*</sup> and Lei Qin<sup>2\*</sup>

# Abstract

**Background** Radiomic analysis of quantitative features extracted from segmented medical images can be used for predictive modeling of prognosis in brain tumor patients. Manual segmentation of the tumor components is time-consuming and poses significant reproducibility issues. We compare the prediction of overall survival (OS) in recurrent high-grade glioma(HGG) patients undergoing immunotherapy, using deep learning (DL) classification networks along with radiomic signatures derived from manual and convolutional neural networks (CNN) automated segmentation.

**Materials and methods** We retrospectively retrieved 154 cases of recurrent HGG from multiple centers. Tumor segmentation was performed by expert radiologists and a convolutional neural network (CNN). From the segmented tumors, 2553 radiomic features were extracted for each case. A robust feature subset was selected using intraclass correlation coefficient analysis between manual and automated segmentations. The data was divided into a 9:1 ratio and validated through ten-fold cross-validation and tested on a rotating test set. Features selection was done by the Kruskal–Wallis test. The Radiomics-based OS predictions, generated using Support Vector Machine (SVM), were compared between the two segmentation approaches and against OS prediction by the CNN model adapted for classification. Model efficacy was evaluated using the area under the receiver operating characteristic curve (AUC).

**Results** The clinical model AUC for OS prediction was  $0.640 \pm 0.013$  (mean  $\pm 95\%$  confidence interval) in the training set and  $0.610 \pm 0.131$  in the test set. The radiomics prediction of OS based on manual segmentation outperformed automatic segmentation (AUC of  $0.662 \pm 0.122$  vs.  $0.471 \pm 0.086$ , respectively) in the test set. Robust features improved the performance of manual segmentation to AUC of  $0.700 \pm 0.102$ , of automated segmentation to  $0.554 \pm 0.085$ . The CNN prognosis model demonstrated promising results, with an average AUC of  $0.755 \pm 0.071$  for training sets and  $0.700 \pm 0.101$  for the test set.

\*Correspondence: Qi Wan qiwan@gzhmu.edu.cn Geoffrey S. Young gsyoung@bwh.harvard.edu Lei Qin lei\_gin@dfci.harvard.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusion** Manual segmentation-derived radiomic features outperformed automated segmentation-derived features for predicting OS in recurrent high-grade glioma patients undergoing immunotherapy. The end-to-end CNN prognosis model performed similarly to radiomics modeling using manual-segmentation-derived features without the need for segmentation. The potential time-saving must be weighed against the lower interpretability of end-to-end black box modeling.

Keywords High-grade glioma, Convolutional neural networks, Deep learning, Radiomics, Overall survival

#### Introduction

High-grade gliomas (HGG) represent the most common and aggressive infiltrative glioma in adults, characterized by rapid growth and high invasiveness, with glioblastoma (GBM) being the most common type [1]. The prognosis for patients with HGG is universally poor, despite advancements in standard care, including maximal surgical resection, concurrent chemoradiotherapy, and maintenance chemotherapy [2, 3]. Immunotherapies are transforming the care of patients with melanoma and lung cancer. Although these therapies have not shown substantial benefit for HGG patients as a whole, data suggests that a subset of patients may respond [4]. Identifying these subsets of patients who may benefit from immunotherapy remains a critical goal.

Radiomics, the extraction and analysis of quantitative imaging features from brain tumor MRI, allows automated, quantitative, diagnostic and prognostic modeling [5, 6]. Feature extraction for radiomic analysis requires 'tumor segmentation', by annotating a region of interest (ROI) containing the tumor on brain MRI. Manual segmentation by radiologists is the gold standard, but it is both time-consuming and labor-intensive, hindering research use and precluding clinical translation of radiomics [7]. Deep-learning approaches that allow automated or annotation-independent methods could significantly accelerate the research and clinical application of radiomics.

In this study, we report the development and validation of two deep learning approaches, based on a convolutional neural network (CNN), designed to accelerate radiomics analysis of recurrent HGG patients undergoing immunotherapy: a CNN-segmentation model for automated image segmentation, and an end-to-end, CNN-prognosis model that inputs unsegmented MRI images and outputs a survival prediction. We compare the accuracy of patient overall survival (OS) prediction based on radiomic features obtained from the automated segmentation, compared to features extracted by manual segmentation, and compared to the accuracy of the CNN-prognosis model.

#### **Materials and methods**

#### **Study participants**

This retrospective medical records study was approved by our Institutional Review Board with waiver of informed consent, and conducted in compliance with HIPAA regulations. The patient cohort used in this study was part of our previous research [8]. 154 patients with histopathologically confirmed recurrent high-grade glioma who received immunotherapy between April 2014 and February 2019 were consecutively included. All patients received immunotherapy, with the majority undergoing anti-programmed death-ligand 1 (PD-L1) treatment (n = 151). A few patients received alternative agents, including anti-cluster of differentiation 137 (CD137; n = 1) and anti-lymphocyte activation gene 3 (LAG-3; relatlimab, n = 2) instead of anti-PD-L1. Many patients also received concurrent treatments alongside immunotherapy, including: Anti-cytotoxic T-lymphocyte-associated protein 4 (CTLA-4; ipilimumab, n = 1), anti-CD27 (varlilumab, n=5), anti-colony-stimulating factor-1 (CSF-1; cabiralizumab, n = 1), anti-indoleamine 2,3-dioxygenase 1 (IDO-1; epacadostat, n = 4), anti-LAG-3 (n = 1), vaccines (n=3), isocitrate dehydrogenase (IDH) inhibitor (n = 1), bevacizumab (Avastin, n = 51), re-irradiation (n = 22), re-irradiation combined with Avastin (n = 7), reirradiation combined with temozolomide (TMZ) (n=6), gene therapy (Ad-RTS-12 veledimex, n = 2), carboplatin (n=1), and gene-mediated cytotoxic immunotherapy (n = 1).

Exclusion criteria include participants with intracranial conditions (e.g., epidural hematoma) causing brain tissue displacement or swelling(n = 2), which could affect assessment. Incomplete MRI data (e.g., missing T1-weighted contrast-enhanced images, n = 4); unsatisfying automated segmentation (n = 5). OS was defined as the time from immunotherapy initiation to death. For patients alive at the last follow-up, OS was censored at the time of last recorded contact. OS was dichotomized into intervals of less than 1 year, and 1 year or greater [9]. Figure 1 shows the flow chart of our study.

# Image acquisition and preprocessing

The MRI data retrieved included fluid-attenuated inversion recovery T2-weighted (FLAIR), Apparent Diffusion Coefficient (ADC) maps calculated from diffusion-weighted imaging, and post-contrast T1-weighted (T1CE) sequences. These images were sourced from 12 different scanners made by two vendors (GE Medical Systems<sup>®</sup> and Siemens Healthineers<sup>®</sup>), at two institutions (Dana-Farber Cancer Institute and Brigham Women's



Fig. 1 Workflow for our study. Institutional MRI data were processed and segmented using both manual and automated methods to extract radiomic features. A clinical model based on clinical variables, four radiomics models (two based on original radiomics features and the other two based on robust radiomics features), and a deep learning classification model were constructed and evaluated for predictive accuracy

Hospital). To ensure consistency and comparability, all images were co-registered to the T1CE and interpolated to a uniform resolution of  $1 \times 1 \times 1$  mm<sup>3</sup>.

#### **Tumor segmentation**

Manual segmentation was performed independently by two radiologists, with 10 and 7 years of experience respectively, using 3D Slicer<sup>®</sup> software (Harvard Medical School, Boston, MA, USA, https://www.slicer.org). ROIs were delineated separately on T1CE, ADC, and FLAIR sequences to generate a comprehensive whole-tumor ROI, that comprised both the central enhancing and surrounding non-enhancing portions of the tumor.

For automated segmentation, we employed a Seg-ResNet [10] based CNN architecture, optimized with the Adam optimizer and cosine annealing as the scheduler. The model was trained using a hybrid loss function, which combined DiceLoss and CrossEntropy Loss. The CNN was initially trained on the publicly available BraTS 2018 [11–13] (*n*=210) and BraTS 2020 [11–13] (*n*=369) datasets. To enhance the model's performance and ensure it is effective for both newly diagnosed and recurrent HGG, we fine-tuned it using 100 randomly selected cases from our institution: 73 randomly selected from the 154 included recurrent HGG cases, augmented by 27 newly diagnosed HGG cases. The inclusion of both newly diagnosed and recurrent cases was driven by our goal to create a model robust enough for glioma segmentation across different stages of the disease.

The segmentation model used in this study was part of our previous research [8]. Details of the model's architecture, initial training process and segmentation details, including DICE scores for each sequence and the overall subjective score, were reported in our previous paper [8].

#### **Deep learning model**

For deep learning prognosis classification, we truncated the SegResNet, keeping only the encoder arm, while replacing the decoder arm with an integrated classification head composed of a convolutional layer followed by a dense layer with dimensions  $128 \times 1024$ . This modification was designed to facilitate the transition from feature extraction to classification, ensuring that the network effectively interprets and categorizes the extracted features.

#### **Radiomics analysis**

The sample size ratio between the training set and the testing set was 9:1. Within the training set, we performed ten-fold cross-validation, meaning that in each fold, the training set was further divided into a smaller training subset and a validation subset. To minimize bias and ensure a robust evaluation of the models, we used a stratified ten-fold cross-validation with a rotating test set. This means that in each fold, a different subset of the data is used as the test set, while the remaining data is used for training and validation. This process ensures that every data point is used for testing exactly once. (Fig. 2)

We used Pyradiomics to extract a comprehensive set of features (n = 851) from each of the manually-segmented and CNN-segmented whole-tumor ROI, including shape,



Fig. 2 Schematic of data splitting and cross-validation strategy. To minimize bias, a stratified ten-fold cross-validation with a rotating test set was employed

intensity, texture, and wavelet-based features, from each of the ADC, FLAIR, and T1CE. This yielded a total of 2553 ( $851 \times 3$ ) original features for each ROI. For cases with multifocal tumors, each lesion was labeled, and radiomic features were extracted from all lesions combined as a single dataset per patient.

After Z-score normalization, For each pair of features with a PCC>0.90, one feature was randomly removed. Further feature selection was then performed using the Kruskal-Wallis (KW) test. Then the feature set was input into the support vector machine (SVM) classifier, and performance of data from each segmentation type was compared. Next, original features with ICC>0.75 between manual and auto-segmented ROIs were selected as robust features, reducing the total number of included features to 1278 in each model. These robust features also underwent the same normalization and feature selection process and input into the SVM classifier and the performance of data from each segmentation type was compared. The whole process was repeated 10 times for different folds. FeAture Explorer [14] (FAE, V0.5.3, https://github.com/salan668/FAE) was used for all the analysis.

### Statistical analysis

The Kolmogorov-Smirnov test was used to assess the normality of data distribution. For normally distributed data, differences were evaluated with the independent sample t-test; for non-normally distributed data, the Mann-Whitney U test was used for group comparison. Cox regression analysis was performed to determine the association between clinical variables and the endpoint. We incorporated variables with P < 0.05 in the univariable Cox regression analysis into the multivariable analysis. Multivariate logistic regression was used to build a binary prognostic model. Receiver operating characteristic

#### Table 1 Clinical characteristics of patients

Clinical features	Overall survival (%)		Р
	Short survival (<1 year, n=90)	Long survival (≥1 year, n=64)	value
Age	57.78±12.13	53.27±12.61	0.028*
Sex			0.259
Male	56(62.22)	34(53.13)	
Female	34(37.78)	30(46.88)	
IDH-status			0.867
Mutant	12(13.33)	8(12.70)	
Wild type	78(86.67)	55(87.30)	
MGMT promoter methylation			0.056
Unmethylated	46(58.97)	23(38.33)	
Partially methylated	6(7.69)	7(11.67)	
Methylated	26(33.33)	30(50.00)	

\* denotes *p* < 0.05

(ROC) curve analysis assessed the diagnostic performance of clinical, radiomics, and deep learning models. The area under the ROC curve (AUC) and its 95% confidence intervals (CIs) were calculated across ten data folds to ensure robustness. All statistical analyses were performed using Python (version 3.9). Statistical significance was defined as a p-value less than 0.05.

# Results

### Participant characteristics

The demographic and clinical characteristics of patients are shown in Table 1. In our cohort, there were 133 cases of IDH-wildtype GBM and 21 cases of other HGG, including 1 case of diffuse midline glioma (H3K27Mmutant), 18 cases of IDH-mutant Astrocytoma, and 1 case of IDH-mutant, 1p/19q co-deleted Oligodendroglioma. Patients with short survival had an average age of 57.78 years (±12.13), while those with long survival were younger, averaging 53.27 years (±12.61) (P=0.028). There was a notable trend in O<sup>6</sup>-methylguanine-DNA methyltransferase (MGMT) promoter methylation status. In the short survival group, 46 (58.97%) patients had unmethylated, 6 (7.69%) partially methylated, and 26 (33.33%) methylated MGMT promoters. Conversely, in the long survival group, these numbers were 23 (38.33%), 7 (11.67%), and 30 (50.00%) respectively. The trend approached statistical significance (P=0.056). Gender and IDH status did not show a significant correlation with OS (P=0.259 and 0.867 respectively).

#### Univariate and multivariate Cox regression

The univariate Cox regression analysis identified age(p=0.039) and MGMT methylation(p=0.035) are significant OS predictors in HGG patients. The multivariate Cox regression analysis revealed that the hazard ratio (HR) for age was 1.0178 (95% CI: 1.001–1.034, p=0.037). Patients with methylated MGMT had a significantly reduced hazard of death (HR=0.658, 95% CI: 0.444–0.975, p=0.037) compared to those with unmethylated MGMT. However, the partially methylated MGMT

group did not show a statistically significant difference in survival (HR = 0.730, 95% CI: 0.378-1.409, p = 0.348).

#### Performance of clinical models

We included MGMT promoter methylation and age as predictors in our clinical model. The model achieved an AUC of  $0.640 \pm 0.013$  in the training set and  $0.610 \pm 0.131$  in the test set across 10 folds.

## Lesion segmentation

Figure 3 shows the distribution of diameter and volumerelated features of manual and auto segmentation indicating a good consistency between the automated and manual methods. Figure 4 shows some examples of the comparison of the two segmented ROIs. All lesions were detected by CNN and all segmentations were checked by a radiologist with 10 years' experience. Subjective reviews revealed no significant difference between automatic and manual segmentation [8].

#### Performance of radiomics models

Figure 5 illustrates the performance of different radiomics models compared to clinical models and differentiated by segmentation technique. Models using features from



Fig. 3 Distributions of a selection of diameter and volume-related features from both manual and deep learning segmentation methods. There is a considerable overlap between the blue (manual segmentation) and red (deep learning auto-segmentation) histograms suggesting that both segmentation methods yield fairly similar morphological features across the dataset



Fig. 4 Manual and automatic segmentation examples in short overall survival (left) and long overall survival group (right). Top row: FLAIR, contrastenhanced (C+), and ADC images. Middle row: Manual segmentation overlay on the respective images, with enhancing tumor highlighted in semitransparent dark red. Bottom row: Automatic segmentation results using a CNN-based approach, depicted with the same color scheme

manual segmentation achieved a mean AUC in the test set of  $0.662 \pm 0.122$ , compared to models using features from automatic segmentation that achieved a mean AUC of  $0.471 \pm 0.086$  in the test set across 10 folds.

Models using only the 1287 robust features selected based on ICC>0.75 demonstrated improved performance: manual segmentation mean AUC of  $0.700 \pm 0.102$ , CNN-segmentation mean AUC of  $0.554 \pm 0.085$ .

#### Performance of deep learning models

The CNN-prognosis classification model achieved an average AUC of  $0.755 \pm 0.071$  in the training set, and  $0.700 \pm 0.101$  in the test set, across 10 folds.

## Discussion

Accurate prediction of survival may help select HGG patients for trials of novel immunotherapies or targeted molecular therapies in which variability in patient responses poses significant challenges, and can be helpful for personalizing HGG therapy in individual patients. We compare automated CNN-segmentation with manual-segmentation based radiomics OS prediction and clinical OS prediction model in patients with recurrent HGG treated with immunotherapy. Radiomics models using features extracted by expert manual segmentation yielded better performance than equivalent models using features extracted by automated segmentation. The performance of both radiomics models improved after applying robust features. The CNN-prognosis endto-end classification model, combining an encoder-arm trained for tumor segmentation with a classifier trained to predict overall survival performed comparably to the radiomics model.

In this study, we demonstrate some important clinical prognostic factors in recurrent HGG patients treated with immunotherapy. Our findings suggest the significance of MGMT promoter methylation as a positive prognostic marker, aligning with a recent meta-analysis [15]. However, the partially methylated MGMT group did not show a statistically significant difference in survival. Conversely, advanced age correlates with reduced OS, potentially due to age-related immune suppression in the brain [16]. We developed clinical models aimed at predicting OS, which demonstrated moderate efficacy. This indicates that relying solely on clinical variables might offer limited predictive power in forecasting survival outcomes for recurrent HGG patients undergoing immunotherapy.

The application of radiomics and machine learning for imaging-based prognosis prediction in HGG patients undergoing immunotherapy remains under-explored. A previous study indicates that high PD-L1 levels, as predicted by these methods, are associated with a better prognosis [17]. A recent study demonstrated that a



Performance of different models for predicting overall survival

**Fig. 5** Performance of clinical, radiomics and deep learning models for predicting overall survival in recurrent high-grade glioma patients who underwent immunotherapy. Performance is measured by the area under the receiver operating characteristic curve with a 95% confidence interval, comparing training (blue) and test (orange) datasets. (Auto Seg: Automated segmentation, Manual Seg: Manual segmentation)

radiomics-based machine learning model, using first ontreatment MR imaging features, can predict survival rates in patients with glioblastoma undergoing PD-L1 inhibition immunotherapy [18]. Previous studies largely relied on manual segmentation, which is time-consuming and subject to inter-operator variations [19, 20]. We report the development and testing of two distinct CNN-based approaches to automate prognostic analysis: CNN-segmentation and end-to-end CNN-prognosis models, and compare them to OS prediction using manual segmentation-based radiomics and known clinical and molecular features. Our investigation revealed that, although the CNN segmentation was qualitatively similar to manual segmentation, predictive models based on manual segmentation by radiologists outperformed those using automated segmentation. It seems likely that the shape (diameter and volume) features and semi-quantitative analysis of the ROIs were not sufficiently sensitive to capture the finer distinctions that are pivotal for prognostic evaluations. On the one hand, this highlights an ongoing need for expert input, especially in complex scenarios like postoperative cases where automated segmentation is often challenging [21, 22]. It illustrates that more work is needed to fully automate radiomics-based modeling in reccurrent HGG.

Our finding that the selection of robust feature subsets substantially improved radiomics model performance suggests that this strategy deserved further research. In the context of GBM MRI radiomics, studies have investigated the robustness of radiomic features across various image preprocessing methods, registration algorithms, and segmentations by different experts [23-26]. Our results demonstrated that the identification of robust features shared between manual and automated segmentation improved OS prediction. Our findings align with prior research showing that selecting stable features enhances the accuracy of prognostic models compared to those trained with less robust features [27]. This raises the question of whether a subset of radiomics features could be identified that would yield more consistent results across a wide range of different datasets.

Previous studies have introduced deep learning models as a promising tool for predicting the prognosis of glioblastoma [28, 29]. The majority of these studies built their prediction models based on tumor segmentation or cropped image regions that contain only the tumor, neglecting information available in the surrounding tissue [30]. This may omit crucial prognostic data, as highlighted by recent research indicating that sarcopenia metrics derived from temporalis muscle, as analyzed by a DL model, correlate with GBM outcomes [31]. This underscores the potential prognostic value embedded within non-tumoral regions. Our results are consistent with this, demonstrating promising results for recurrent HGG prognosis prediction using a modified CNN-prognosis model that extracts features from the whole brain, circumventing the need for specific tumor delineation.

There are several limitations in our study. First, although the heterogeneity of our image data, comprising 12 scanners, 2 vendors and 2 institutions, should help with generalizability, our sample size is relatively small. To mitigate the small sample size, we employed a 10-fold cross-validation approach with a rotating test set. Validation of these results in larger and more diverse patient cohorts will be essential. Second, in this study, radiomics analysis was performed based on the whole tumor segmentation rather than on separate masks. The primary reason for this choice was that whole tumor segmentation showed the highest similarity between manual and CNN segmentation. To avoid issues related to segmentation quality, we opted for whole tumor analysis. Future research should explore the impact of radiomic features derived from individual sequences on predictive performance. Third, while our radiomics and deep learning models predict prognosis in recurrent HGG patients receiving immunotherapy, the lack of a non-immunotherapy control group limits the attribution of outcomes specifically to immunotherapy. Additionally, concurrent treatments introduce variability in effects. Our findings serve as a preliminary step towards identifying immunotherapy-specific prognostic markers, though further research with control groups and standardized treatments is needed to validate these markers. Fourth, the choice of machine learning methods could impact the results of predictive modeling [32]. In this study, we used SVM, as it is a relatively common and generally well-performing method. However, exploring different machine learning methods for predicting HGG prognosis may be beneficial in future studies. Finally, interpretability remains a major challenge for CNN and remains a major barrier to the clinical translation of these methods.

In conclusion, our study demonstrates that CNN-based brain tumor MRI segmentation has the potential as a tool to facilitate research in HGG patients, but is not yet ready to fully replace human-expert annotation in complex scenarios such as postoperative cases. Identification of robust radiomic feature subsets can enhance radiomics model performance and deserve further investigation. End-to-end, segmentation-independent, CNN prediction of prognosis, demonstrated comparable performance to radiomics models in this small dataset, but at the cost of posing greater problems for interpretability.

#### Abbreviations

AUC	area under the curve
CNN	convolutional neural networks
DL	deep learning
OS	overall survival
SVM	Support Vector Machine
GBM	glioblastoma
MGMT	O6-methylguanine-DNA methyltransferase
IDH	isocitrate dehydrogenase
HGG	high grade glioma

#### Acknowledgements

Not applicable.

#### Author contributions

The study was conceived and designed by LQ and GSY. Material preparation, data collection were performed by JK, RH, DR, QW, XC and JL. Data analysis was performed by QW, CL, CZ. The first draft of the manuscript was written by QW and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

#### Funding

This work was funded by the National Institutes of Health awards R01LM012434, R01LM013891, and R01LM013772.

#### Data availability

The data are not publicly available due to privacy or ethical restrictions.

### Declarations

#### Ethics approval

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by our institutional Ethics Committee.

# **Consent for publication**

**Competing interests** 

The authors declare no competing interests.

#### Author details

Not applicable.

<sup>1</sup>Department of Radiology, the Key Laboratory of Advanced Interdisciplinary Studies Center, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China <sup>2</sup>Department of Imaging, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA <sup>3</sup>Department of Radiology, Division of Biomedical Imaging and Bioengineering, UMass Chan Medical School, Worcester, MA, USA <sup>4</sup>Digital Medical Research Center, School of Basic Medical Sciences, Fudan University, Shanghai, China <sup>5</sup>Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA <sup>6</sup>Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, China <sup>7</sup>Department of Radiology, the Affiliated Cancer Hospital of Zhengzhou University (Henan Cancer Hospital), Zhengzhou, China <sup>8</sup>Center for Neuro-Oncology, Dana-Farber Cancer Institute, Harvard

Medical School, Boston, MA, USA

#### Received: 3 July 2024 / Accepted: 18 December 2024 Published online: 21 January 2025

#### References

- Whitfield BT, Huse JT. Classification of adult-type diffuse gliomas: impact of 1. the World Health Organization 2021 update. Brain Pathol. 2022;32(4):e13062.
- 2. Kotecha R, Odia Y, Khosla AA, Ahluwalia MS. Key clinical principles in the management of Glioblastoma. JCO Oncol Pract. 2023;19(4):180-9.

- Mensali N, Inderberg EM. Emerging biomarkers for Immunotherapy in Glioblastoma. Cancers (Basel). 2022;14(8).
- Chiu F-Y, Yen Y. Imaging biomarkers for clinical applications in neuro-oncology: current status and future perspectives. Biomark Res. 2023;11(1):1–23.
- Lohmann P, Franceschi E, Vollmuth P, Dhermain F, Weller M, Preusser M, et al. Radiomics in neuro-oncological clinical trials. Lancet Digit Health. 2022;4(11):e841–9.
- Aboian M, Bousabarah K, Kazarian E, Zeevi T, Holler W, Merkaj S et al. Clinical implementation of artificial intelligence in neuroradiology with development of a novel workflow-efficient picture archiving and communication systembased automated brain tumor segmentation and radiomic feature extraction. Front NeuroSci. 2022;16.
- Wan Q, Kim J, Lindsay C, Chen X, Li J, lorgulescu JB, et al. Auto-segmentation of adult-type diffuse gliomas: comparison of transfer learning-based convolutional neural network model vs. radiologists. J Imaging Inf Med. 2024;37(4):1401–10.
- de Groot J, Penas-Prado M, Alfaro-Munoz K, Hunter K, Pei BL, O'Brien B et al. Window-of-opportunity clinical trial of pembrolizumab in patients with recurrent glioblastoma reveals predominance of immune-suppressive mac rophages. Neurooncology. 22(4):539–49.
- Myronenko A, editor. 3D MRI brain tumor segmentation using autoencoder regularization. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4; 2019: Springer.
- Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data. 2017;4:170117.
- 12. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging. 2015;34(10):1993–2024.
- Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:181102629. 2018.
- Song Y, Zhang J, Zhang YD, Hou Y, Yan X, Wang Y, et al. FeAture Explorer (FAE): a tool for developing and comparing radiomics models. PLoS ONE. 2020;15(8):e0237587.
- Hu W, Liu H, Li Z, Liu J, Chen L. Impact of molecular and clinical variables on survival outcome with immunotherapy for glioblastoma patients: a systematic review and meta-analysis. CNS Neurosci Ther. 2022;28(10):1476–91.
- Ladomersky E, Zhai L, Lauing KL, Bell A, Xu J, Kocherginsky M, et al. Advanced Age increases Immunosuppression in the brain and decreases immunotherapeutic efficacy in subjects with Glioblastoma. Clin Cancer Res. 2020;26(19):5232–45.
- Sforazzini F, Salome P, Kudak A, Ulrich M, König L, Warta R, et al. PD-L1-R: a MR based surrogate for PD-L1 expression in Glioblastoma multiforme. J Clin Oncol. 2021;39(15suppl):2041.
- George E, Flagg E, Chang K, Bai HX, Aerts HJ, Vallières M, et al. Radiomicsbased machine learning for Outcome Prediction in a Multicenter Phase II

study of programmed death-ligand 1 inhibition immunotherapy for Glioblastoma. AJNR Am J Neuroradiol. 2022;43(5):675–81.

- Visser M, Muller DMJ, van Duijn RJM, Smits M, Verburg N, Hendriks EJ, et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. Neuroimage Clin. 2019;22:101727.
- Bo HK, Solheim O, Jakola AS, Kvistad KA, Reinertsen I, Berntsen EM. Intra-rater variability in low-grade glioma segmentation. J Neurooncol. 2017;131(2):393–402.
- Sorensen PJ, Carlsen JF, Larsen VA, Andersen FL, Ladefoged CN, Nielsen MB et al. Evaluation of the HD-GLIO Deep Learning Algorithm for Brain Tumour Segmentation on postoperative MRI. Diagnostics (Basel). 2023;13(3).
- Abayazeed AH, Abbassy A, Müeller M, Hill M, Qayati M, Mohamed S, et al. NS-HGlio: a generalizable and repeatable HGG segmentation and volumetric measurement AI algorithm for the longitudinal MRI assessment to inform RANO in trials and clinics. Neurooncol Adv. 2023;5(1).vdac184.
- Shiri I, Hajianfar G, Sohrabi A, Abdollahi H, S PS, Geramifar P, et al. Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: test-retest and image registration analyses. Med Phys. 2020;47(9):4265–80.
- Pati S, Verma R, Akbari H, Bilello M, Hill VB, Sako C, et al. Reproducibility analysis of multi-institutional paired expert annotations and radiomic features of the Ivy Glioblastoma Atlas Project (Ivy GAP) dataset. Med Phys. 2020;47(12):6039–52.
- Um H, Tixier F, Bermudez D, Deasy JO, Young RJ, Veeraraghavan H. Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets. Phys Med Biol. 2019;64(16):165011.
- Tixier F, Um H, Young RJ, Veeraraghavan H. Reliability of tumor segmentation in glioblastoma: impact on the robustness of MRI-radiomic features. Med Phys. 2019;46(8):3582–91.
- Verma R, Hill VB, Statsevych V, Bera K, Correa R, Leo P, et al. Stable and discriminatory Radiomic features from the Tumor and its Habitat Associated with Progression-Free Survival in Glioblastoma: a multi-institutional study. AJNR Am J Neuroradiol. 2022;43(8):1115–23.
- Han W, Qin L, Bay C, Chen X, Yu KH, Miskin N, et al. Deep transfer learning and Radiomics Feature Prediction of Survival of patients with high-Grade Gliomas. AJNR Am J Neuroradiol. 2020;41(1):40–8.
- Zhang D, Luan J, Liu B, Yang A, Lv K, Hu P, et al. Comparison of MRI radiomicsbased machine learning survival models in predicting prognosis of glioblastoma multiforme. Front Med (Lausanne). 2023;10:1271687.
- Kaur G, Rana PS, Arora V. State-of-the-art techniques using pre-operative brain MRI scans for survival prediction of Glioblastoma Multiforme patients and future research directions. Clin Transl Imaging. 2022;10(4):355–89.
- Mi E, Mauricaite R, Pakzad-Shahabi L, Chen J, Ho A, Williams M. Deep learningbased quantification of temporalis muscle has prognostic value in patients with glioblastoma. Br J Cancer. 2022;126(2):196–203.
- Wan Q, Zhou J, Xia X, Hu J, Wang P, Peng Y, et al. Diagnostic performance of 2D and 3D T2WI-based radiomics features with machine learning algorithms to distinguish solid solitary pulmonary lesion. Front Oncol. 2021;11:683587.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.