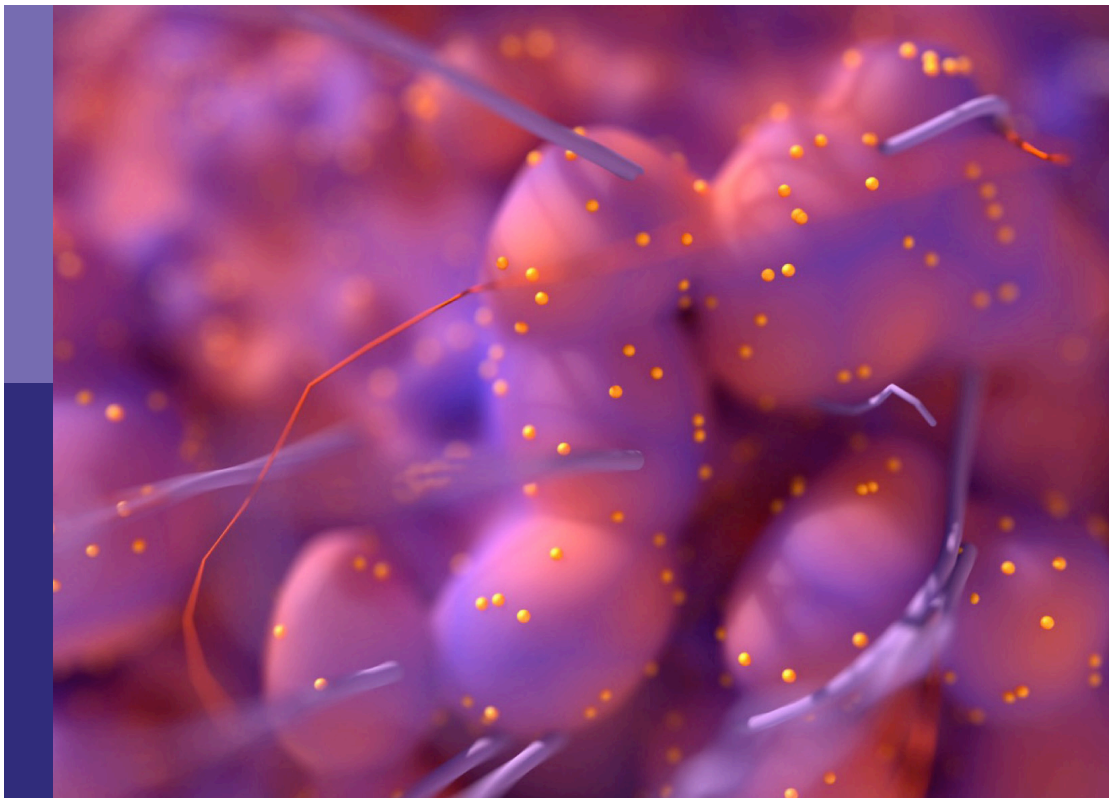
frontiers | Research Topics

Application of emerging technologies in the diagnosis and treatment of patients with brain tumors: new frontiers in imaging for neuro-oncology

Edited by
Domenico Aquino, Valentina Pinzi, Camilla Russo,
Vincenzo Di Nunno and Elena De Martin

Published in
Frontiers in Oncology
Frontiers in Neurology



ISSN 1664-8714
ISBN 978-2-8325-6464-6
DOI 10.3389/978-2-8325-6464-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Application of emerging technologies in the diagnosis and treatment of patients with brain tumors: new frontiers in imaging for neuro-oncology

Topic editors

Domenico Aquino — IRCCS Carlo Besta Neurological Institute Foundation, Italy
Valentina Pinzi — IRCCS Carlo Besta Neurological Institute Foundation, Italy
Camilla Russo — Department of Neuroscience, Santobono-Pausilipon Children's Hospital, Italy
Vincenzo Di Nunno — AUSL Bologna, Italy
Elena De Martin — IRCCS Carlo Besta Neurological Institute Foundation, Italy

Citation

Aquino, D., Pinzi, V., Russo, C., Di Nunno, V., De Martin, E., eds. (2025). *Application of emerging technologies in the diagnosis and treatment of patients with brain tumors: new frontiers in imaging for neuro-oncology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6464-6

Table of contents

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 03 June 2025
DOI 10.3389/fonc.2025.1625256

OPEN ACCESS

EDITED AND REVIEWED BY
Sharon R. Pine,
University of Colorado Anschutz Medical Campus, United States

*CORRESPONDENCE
Elena De Martin
dema.elena02@gmail.com

RECEIVED 08 May 2025
ACCEPTED 19 May 2025
PUBLISHED 03 June 2025

CITATION
Aquino D, De Martin E and Russo C (2025) Editorial: Application of emerging technologies in the diagnosis and treatment of patients with brain tumors: new frontiers in imaging for neuro-oncology. *Front. Oncol.* 15:1625256. doi: 10.3389/fonc.2025.1625256

Editorial: Application of emerging technologies in the diagnosis and treatment of patients with brain tumors: new frontiers in imaging for neuro-oncology

Domenico Aquino[1], Elena De Martin[2*] and Camilla Russo[3]

[1]Neuroradiology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, [2]Health Department, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, [3]Department of Pediatric Neurosciences, Neuroradiology, Santobono-Pausilipon Children's Hospital, Naples, Italy

KEYWORDS

MRI, brain, tumor, radiomic, prognostic

Editorial on the Research Topic

Application of emerging technologies in the diagnosis and treatment of patients with brain tumors: new frontiers in imaging for neuro-oncology

Accurate prognostication across brain tumor types — particularly those with the poorest outcomes, such as high-grade gliomas and especially glioblastoma multiforme (GBM) — remains a central challenge in neuro-oncology. Although advances in surgical techniques, radiotherapy, and systemic therapies have extended survival for some patients, outcomes still vary widely and are difficult to predict based on clinical and molecular data alone. Conventional prognostic models typically incorporate factors such as age, performance status, tumor location, histopathological grade, and molecular alterations (e.g., MGMT promoter methylation, IDH mutation). However, these measures often fail to capture the full extent of biological complexity and spatial heterogeneity within tumors that underlie treatment resistance and progression. Consequently, there is a pressing need for reliable, noninvasive biomarkers that more comprehensively reflect tumor biology, facilitate risk stratification, guide individualized treatment planning, and ultimately improve patient outcomes.

In this context, quantitative neuro-imaging has emerged as a valuable adjunct to traditional assessment, with research concentrating principally on brain tumors with a worse prognosis. Multiparametric MRI — including contrast-enhanced T1-weighted, T2-weighted, FLAIR, diffusion-weighted, and perfusion sequences — provides complementary information on tumor morphology, cellularity, and vascular characteristics. Radiomic analysis further enhances this approach by extracting high-dimensional quantitative features (intensity, shape, texture, and wavelet-based metrics) from defined regions of interest, transforming images into data-rich profiles and underscoring the potential of imaging-derived biomarkers to improve prognostic accuracy and tailor therapy across

diverse brain tumor entities. Notably, Kickingereder et al. used radiomic analysis to identify imaging phenotypes in high grade gliomas associated with patient prognosis; by extracting high-dimensional imaging features, they demonstrated that certain radiomic profiles correlate with overall survival, independent of established clinical variables, with a prognostic accuracy superior when compared with clinical and conventional imaging models (1). Similarly, the integration of quantitative perfusion MRI parameters with genetic profiling and molecular insights has proven to enhance tumor characterization, more accurate prognostication, and allow for more tailored therapeutic strategies. Such synergy paves the way for a more refined and individualized model of care in neuro-oncology, moving beyond traditional one-size-fits-all paradigm and toward a future of truly personalized medicine (2). Thus, there are several studies that illustrate how combining radiomic features with molecular and clinical data yields composite models that capture complementary dimensions of tumor biology, offering refined prognostic insights and potential predictive value for targeted therapies. Clinical applications of diffusion-MRI, which include among others diffusion-weighted imaging (DWI) with apparent diffusion coefficient (ADC) maps, diffusion tensor imaging (DTI) and diffusion kurtosis imaging (DKI), have further illuminated the prognostic significance of cellular density and extracellular matrix composition in gliomas. For instance, Li et al. investigated the prognostic value of DKI in GBM, demonstrating how higher mean kurtosis values (reflecting microstructural tissue complexity) were significantly associated with longer overall survival (3). In parallel, Chen et al. assessed how ADC histogram analysis is able to provide innovative insights into the MGMT and TERT molecular characterization in patients newly diagnosed with GBM, enhancing prognostic stratification (4). These findings support the notion that ADC-derived imaging biomarkers reflect underlying histopathological and molecular traits, enabling noninvasive *in vivo* phenotyping of tumor aggressiveness. Advanced perfusion imaging metrics also have prognostic and predictive implications. Dynamic susceptibility contrast MRI–derived cerebral blood volume maps correlate with microvascular density and angiogenic activity. In the context of antiangiogenic therapy, Kickingereder et al. identified pre-treatment radiomic features that predicted progression-free and overall survival in recurrent GBM patients treated with bevacizumab (5). The integration of perfusion radiomics and treatment response modeling highlights how imaging biomarkers can guide therapeutic decision-making, identifying patients most likely to benefit from targeted agents. Machine learning and deep learning techniques have accelerated the discovery of prognostic imaging signatures by automating tumor segmentation, feature extraction, and pattern recognition. Deep learning has significantly transformed the automated segmentation and classification of intracranial tumors, showing particularly promising performance in the context of gliomas, and offering notable advantages in terms of time efficiency and resource optimization; in particular, deep-learning–based segmentation algorithms improve the reproducibility of region of interest delineation, a critical factor for multicenter studies (6). Lao et al. employed transfer learning to derive deep radiomic features from pre-trained convolutional neural networks, constructing a nomogram that outperformed traditional clinical risk factors and handcrafted radiomic features for survival prediction (7). These deep features capture complex, hierarchical image representations that may elude conventional radiomic pipelines. Radiogenomic approaches integrate imaging phenotypes with genomic and transcriptomic landscapes, providing mechanistic insights and enhancing prognostic models. Radiomics has been applied to various areas of neuro-oncology, with particular success in the differential diagnosis and classification of brain tumors (8). Kickingereder et al. used radiomic subtyping to classify GBM tumors into phenotypic clusters associated with differential survival, revealing that specific subtypes—characterized by heterogeneous texture and angiogenic features—portend poorer outcomes (9). A study by Qi et al. demonstrated that specific imaging features correlated with molecular subtypes of GBM, such as proneural and mesenchymal profiles; this spatial distinction, combined with differences in imaging characteristics, allows for the noninvasive prediction of molecular subtypes using MRI data (10). Integrative models that combine radiomics, genomics, and clinical variables have shown superior prognostic performance compared with single-modality models, underscoring the value of multimodal data fusion in precision neuro-oncology. Despite these advances, several challenges impede the clinical translation of prognostic neuro-imaging biomarkers. Variability in MRI acquisition protocols, scanner manufacturers, and imaging parameters introduces heterogeneity that can compromise feature stability. Initiatives to standardize imaging protocols and develop digital reference objects for quality assurance are critical to ensure reproducibility. Segmentation remains labor-intensive and subject to interobserver variability; robust, validated automated or semi-automated segmentation tools are needed to enable widespread adoption. Retrospective study designs and small sample sizes limit generalizability; large-scale, multicenter, prospective studies with standardized imaging and outcome metrics are required for validation. Finally, regulatory pathways for imaging biomarkers demand clear demonstration of clinical utility and cost-effectiveness. In summary, quantitative neuro-imaging and radiomics have reshaped the landscape of prognostication in brain tumors, offering noninvasive insights into tumor heterogeneity, vascularity, and molecular composition. From handcrafted texture features and ADC histogram metrics to deep-learning–derived signatures and radiogenomic mappings, imaging biomarkers hold promise for risk stratification and treatment personalization. As the field advances toward standardized, validated, and interpretable models, neuro-imaging will become integral to precision management of brain tumors, enabling clinicians to tailor therapeutic strategies to individual tumor biology and ultimately improve patient outcomes. The articles published in this Research Topic exemplify the strong and sustained interest in emerging technologies applied to the initial assessment, differential diagnosis, and biological characterization of brain tumors — particularly gliomas — and reflect the ongoing efforts of the scientific community to advance neuroimaging tools that enhance prognostication and therapeutic outcomes for patients with brain neoplasms.

Author contributions

DA: Conceptualization, Resources, Supervision, Writing – original draft, Writing – review & editing. ED: Conceptualization, Investigation, Supervision, Validation, Writing – original draft, Writing – review & editing. CR: Conceptualization, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 13 May 2024
DOI 10.3389/fonc.2024.1401977

OPEN ACCESS

EDITED BY
Camilla Russo,
Santobono-Pausilipon Children's Hospital, Italy

REVIEWED BY
Ying Zhuge,
National Institutes of Health (NIH), United States
Mario Tortora,
University of Naples Federico II, Italy
Mario Tranfa,
University of Naples Federico II, Italy

*CORRESPONDENCE
Tao Zhu
zhutao5@126.com

†These authors share first authorship

RECEIVED 16 March 2024
ACCEPTED 29 April 2024
PUBLISHED 13 May 2024

CITATION
Tan R, Sui C, Wang C and Zhu T (2024) MRI-based intratumoral and peritumoral radiomics for preoperative prediction of glioma grade: a multicenter study.
Front. Oncol. 14:1401977.
doi: 10.3389/fonc.2024.1401977

MRI-based intratumoral and peritumoral radiomics for preoperative prediction of glioma grade: a multicenter study

Rui Tan[1†], Chunxiao Sui[2†], Chao Wang[3] and Tao Zhu[1*]

[1]Department of Neurosurgery, Tianjin Medical University General Hospital, Tianjin, China, [2]Department of Molecular Imaging and Nuclear Medicine, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China, [3]Department of Neurosurgery, Qilu Hospital of Shandong University Dezhou Hospital (Dezhou People's Hospital), Shandong, China

Background: Accurate preoperative prediction of glioma is crucial for developing individualized treatment decisions and assessing prognosis. In this study, we aimed to establish and evaluate the value of integrated models by incorporating the intratumoral and peritumoral features from conventional MRI and clinical characteristics in the prediction of glioma grade.

Methods: A total of 213 glioma patients from two centers were included in the retrospective analysis, among which, 132 patients were classified as the training cohort and internal validation set, and the remaining 81 patients were zoned as the independent external testing cohort. A total of 7728 features were extracted from MRI sequences and various volumes of interest (VOIs). After feature selection, 30 radiomic models depended on five sets of machine learning classifiers, different MRI sequences, and four different combinations of predictive feature sources, including features from the intratumoral region only, features from the peritumoral edema region only, features from the fusion area including intratumoral and peritumoral edema region (VOI-fusion), and features from the intratumoral region with the addition of features from peritumoral edema region (feature-fusion), were established to select the optimal model. A nomogram based on the clinical parameter and optimal radiomic model was constructed for predicting glioma grade in clinical practice.

Results: The intratumoral radiomic models based on contrast-enhanced T1-weighted and T2-flair sequences outperformed those based on a single MRI sequence. Moreover, the internal validation and independent external test underscored that the XGBoost machine learning classifier, incorporating features extracted from VOI-fusion, showed superior predictive efficiency in differentiating between low-grade gliomas (LGG) and high-grade gliomas (HGG), with an AUC of 0.805 in the external test. The radiomic models of VOI-fusion yielded higher prediction efficiency than those of feature-fusion. Additionally, the developed nomogram presented an optimal predictive efficacy with an AUC of 0.825 in the testing cohort.

Conclusion: This study systematically investigated the effect of intratumoral and peritumoral radiomics to predict glioma grading with conventional MRI. The optimal model was the XGBoost classifier coupled radiomic model based on VOI-fusion. The radiomic models that depended on VOI-fusion outperformed those that depended on feature-fusion, suggesting that peritumoral features should be rationally utilized in radiomic studies.

KEYWORDS

MRI, glioma grade, radiomics, peritumoral features, nomogram

1 Introduction

Glioma is a highly fatal disease that represents the most frequent form of primary cancer in the central nervous system (CNS), accounting for about 80% of all malignant tumors in the brain (1–3). Due to the same standardized treatment that can result in varying prognoses for different patients, it may be necessary to make specialized treatment decisions based on the tumor grade in clinical practice (4–6). Gliomas are categorized into grades I-IV in the Central Nervous System Midstream Classification of the World Health Organization (WHO) of 2021, with grades I-II being low-grade gliomas (LGG) and grades III-IV being high-grade gliomas (HGG) (7, 8). Accurate preoperative grading of gliomas is essential for assessing prognosis and developing individualized treatment plans, such as the extent of surgical resection, and the decision of postoperative chemoradiotherapy (9).

In contemporary clinical practice, gliomas are graded based on surgical or puncture histopathologic investigation (10). This diagnostic method is intrusive and slow, though. Furthermore, tissue biopsies from one area of the tumor might not be indicative of the histology of the entire tumor due to the recognized heterogeneity of gliomas and sampling error (11–13). High-precision noninvasive solutions that can offer preoperative grading information are therefore gaining popularity. Over the past decades, magnetic resonance imaging (MRI) has emerged as a crucial non-invasive diagnostic and assistant therapeutic technique for brain tumors, which is used to aid in differential diagnosis, guide treatment planning, and monitor therapy response (14–17). Nevertheless, competent radiologists may easily spot tumors from MRI sequences with the naked eye, gliomas are difficult to discriminate based on grade because of the variability and diversity of the tumors, which is undoubtedly a great challenge for imaging technology (18–20).

Radiomics, a burgeoning discipline, employing automated data mining algorithms to extract characteristics from medical images in a high-throughput manner, has been demonstrated notable advancements in the realm of medical imaging applications (21–24). These extracted features are then utilized by the machine to train itself and generate the anticipated desired output (25–27). Notably, recent progress has been achieved in the prediction of grading gliomas through the utilization of preoperative MRI scans and various machine learning methods. Gemini et al. evaluated the capacity of the Visually AcceSAble Rembrandt Images (VASARI) scoring system in predicting glioma grades and Isocitrate Dehydrogenase (IDH) status, with a possible application in machine learning (28). You et al. utilized traditional radiomics and the VASARI standard to construct a model determining glioma grade with an Area Under the Curve (AUC) of 0.966 (29). Wang et al. created and assessed a multiparametric MRI-based radiomics nomogram for predicting glioma grading (30). In recent years, deep learning has exhibited excellent performance with broader application prospects and deeper development in clinical applications. Voort et al. developed a single multi-task convolutional neural network that used preoperative MRI scans to predict the molecular subtype and grade of glioma, and the independent dataset evaluated that the approach achieved good performance and generalized well (31). Li et al. compared predictive models established by traditional radiomics and deep learning based on multiparametric MRI for grading gliomas and demonstrated that the latter performed better in most circumstances (32). While deep learning-based models have very respectable efficacy, it is commonly recognized as a black box that lacks satisfactory explanatory power. However, these machine learning approaches have focused mainly on the intratumoral region and neglected the role of the peritumoral environment in glioma grade.

The peritumoral environment holds great potential and may provide insightful information for clinical evaluation of the aggressive biological behavior of the tumor (33–35). Regarding intratumoral and peritumoral radiomic analysis, two main research approaches emerged. The first involved feature-fusion, where features from both intratumoral and peritumoral volumes of interest (VOIs) were separately extracted and then integrated. The second method was VOI-fusion, wherein the tumor region was expanded outward by a specific range to create a new VOI combining intra- and peri-tumoral areas. Radiomic features from this newly generated VOI were then extracted for subsequent analysis. For instance, Li et al. independently delineated the gross-tumor region and the peritumor region which was defined as the parenchyma that fell within a 2-cm distance to the tumor

boundary (35). The radiomic features extracted from the two regions were merged and screened to build the radiomic model. Differently, Shi et al. expanded the originally segmented masks of VOIs by five radial distances outside the tumor at 1 mm intervals, creating five new VOIs (34). The findings indicated that the radiomic signature derived from peritumoral regions, specifically at dilated distances of 1 mm and 3 mm, demonstrated the most effective prediction performance in different MRI sequences, respectively. To date, although numerous studies have explored the radiomics of the peritumoral region, there has been a lack of a definitive study that has determined which among them is more persuasive and authoritative. The growth and infiltration of gliomas lead to the disruption of the blood-brain barrier. Consequently, there is a leakage of water, electrolytes, and proteins from peritumoral blood vessels, resulting in increased water content within the brain parenchyma, which contributes to the formation of peripheral edema (32, 36). Given that the tumor and the surrounding edematous area were closely interconnected, forming the microenvironment crucial for tumor cell growth and infiltration, it would be more reasonable to explore them as an integrated whole.

Drawing on the current state of research, the two methods were performed and compared in our study. We conducted an investigation focusing on the intratumor region and its surrounding peritumoral edema region of preoperative MRI scans to predict the grade of glioma. A total of 30 radiomic models, which depended on four different combinations of predictive features source (intratumoral VOI, peritumoral VOI, VOI-fusion, and feature-fusion), five different machine learning classifiers, and different MRI sequences, were established to select the optimal model, which was used for the construction of the nomogram for accurately predicting glioma grade, thereby assisting in the development of personalized treatment strategies for patients, ensuring they receive optimal benefits.

2 Materials and methods

2.1 Study population

All procedures involving human participants in this study adhered to the ethical guidelines outlined in the 1964 Declaration of Helsinki and its subsequent revisions, as well as other applicable ethical standards. The study was approved by the Ethics Committee of Tianjin Medical University General Hospital and Qilu Hospital of Shandong University Dezhou Hospital. Written informed consent was waived due to the retrospective nature of the study.

This study retrospectively analyzed 213 patients with cerebral gliomas from January 2019 to June 2023, who underwent preoperative MRI followed by surgery. Among them, 132 patients with glioma were from the Tianjin Medical University General Hospital (center 1), which was classified as the training cohort and internal validation set, and the remaining 81 patients were from the Qilu Hospital of Shandong University Dezhou Hospital (center 2), zoned as the independent external testing cohort. The inclusion and exclusion criteria are shown in Figure 1.

2.2 Pathological assessment

Pathologists with more than 10 years of experience graded the postoperative specimens, based on the 2021 WHO classification of

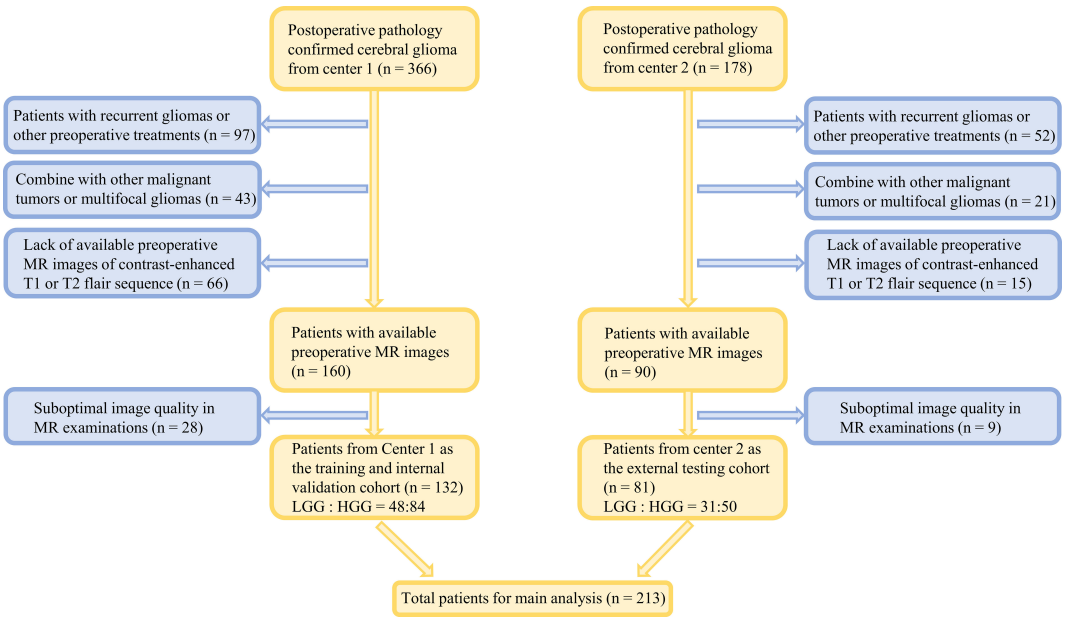


FIGURE 1
Flowchart of the incorporation and expulsion of glioma patients from two centers.

CNS tumors, classifying gliomas into grades I-IV, with grades I-II being LGG and grades III-IV being HGG (19).

2.3 MRI protocol and image preprocessing

Every patient underwent MRI scans within two weeks before the surgery. All MRI studies were performed on the same type of scanners of the two centers, which were acquired using a 3.0 T scanner Discovery MR750 (GE Healthcare) with a 32-channel head coil. MRI acquisition parameters are summarized in Supplementary Table S1. The same MRI protocol was used for training and external testing sets. Patients were told not to move their heads during the scan to reduce the possible impact of head motion. The most useful anatomical multi-contrast MRI sequences included contrast-enhanced T1-weighted images and T2-weighted fluid-attenuated inversion recovery (Flair) images that were analyzed for further study.

For image preprocessing, the preoperative contrast-enhanced T1 images and T2 flair images were spatially aligned by using the rigid registration function of the well-validated Ants software from 3D Slicer (version: 5.2.2) software. Moreover, the quality of the registration was carefully inspected for alignment of the ventricular structures by a radiation oncologist. Then, the spacing of contrast-enhanced T1 and T2 flair images was resampled to 1×1×1 mm^3.

2.4 Image segmentation and feature extraction

Manual segmentation of the contrast-enhanced T1 and T2 flair images of target lesions was performed using 3D Slicer (version: 5.2.2) software by two radiologists who possessed over five years of experience in a blinded manner to the study outcome, to eliminate unstable radiomic features and minimize inter-individual variability. Following clinical studies (8, 33), three VOIs have been delineated, the intratumoral VOI based on the contrast-enhanced T1 images, the peritumoral edema VOI including the intratumor VOI based on the T2 flair images, and the peritumoral edema VOI only. Given the spatial alignment between the contrast-enhanced T1 and T2 flair images, the delineated VOIs were shared across the two sequences.

In this study, the feature extraction was performed utilizing the Pyradiomics module in Python 3.7.0. A comprehensive set of 1288 quantitative radiomic features was extracted individually from each VOI for every sequence, obtaining a total of 7728 features. The available features were categorized as follows: the three-dimensional shape characteristics (n=14), the first-order statistical distribution of voxel intensities (n=252), and the texture features, which comprised gray-level co-occurrence matrix (GLCM) (n=308), gray-level dependence matrix (GLDM) (n=196), gray-level run length matrix (GLRLM) (n=224), gray level size zone matrix (GLSZM) (n=224), and neighborhood gray-tone difference matrix (NGTDM) (n=70).

2.5 Feature selection

Six distinct radiomic models were developed independently by employing extracted conventionally intratumoral radiomic features from two sequences and three kinds of VOIs. Firstly, the TR_{T1} model was a radiomic model based on contrast-enhanced T1-derived radiomic features from the intratumoral VOI. The TR_{T2} model was constructed by utilizing T2 flair-derived radiomic features from the intratumoral VOI. The TR model was established by integrating radiomic features originating from the intratumoral VOI of both sequences. Then, the PR model was based on the combined radiomic features from the peritumoral VOI of both sequences. The $TPR_{VOI\text{-}fusion}$ model was constructed by utilizing the radiomic features from the combining VOI of intratumoral and peritumoral VOIs of both sequences. The $TPR_{feature\text{-}fusion}$ model was established by integrating radiomic features from the intratumoral VOI and peritumoral VOI of both sequences, separately.

Subsequently, Z-score normalization was applied to standardize the intensity range of each radiomic feature across various models, preventing the undue assignment of lower or higher weights to specific features. In the feature selection process, three steps were implemented for the training cohort. The ICC test was conducted between the datasets obtained by the two radiologists. The value exceeding 0.75 was deemed indicative of robust reproducibility and reliability, leading to the exclusion of features with ICC<0.75 from subsequent analysis (37, 38). Furthermore, Pearson's rank correlation coefficient was employed to evaluate the correlation between feature pairs, with one feature randomly excluded from each pair exhibiting a correlation coefficient > 0.9. Lastly, the least absolute shrinkage and selection operator (LASSO) regression, coupled with 10-fold cross-validation, was utilized to identify informative features with non-zero coefficients and calculate the corresponding feature weights.

In addition to radiomic features, machine learning models based on predictive clinical parameters were also constructed, which was referred to as the Clinical model in the study. For the feature selection of clinical features, a two-step procedure was performed. First, univariate analysis was used to identify significant features with a p-value < 0.05. Then, the stepwise multivariate analysis was employed to determine the independent indicator with a p-value < 0.05, which was utilized as the predictive clinical parameters for the prediction of glioma grade.

2.6 Model construction

Then, classifiers including Logistic Regression (LR), Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost), Decision Tree (DT), and Multilayer Perceptron (MLP), employed the features filtrated by Lasso feature screening. In total, 30 machine learning models were formulated by integrating the six distinct radiomic models derived from various sequences and VOIs with the five machine learning classifiers for predicting glioma grade.

To ensure the stability of the prediction models, we randomly divided 30% of the training cohort as an internal validation set, then repeated the steps 100 times and averaged the results as the final prediction result under the model. The performance of the model was then evaluated on the independent external testing cohort, which was not used during the development of the model.

To evaluate the effectiveness of these models, several indicators, such as the AUC, accuracy, sensitivity, specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV), were computed to evaluate the performance of the models. Additionally, to demonstrate the precision and net benefit, both the calibration curve and the decision curve analysis were employed.

Additionally, a nomogram using a logistic regression algorithm involving the optimal radiomic model and significant clinical features was developed to provide a straightforward visual representation in clinical practice. The receiver operating characteristic (ROC) curve, calibration curve, and decision curve analysis were employed correspondingly.

2.7 Statistical analysis

Software version 3.7.0 of Python was used for statistical analysis. The p-value for statistical significance was fixed at 0.05, and all statistical tests were two-sided. For continuous variables, mean ± SD was applied to communicate data that followed a normal distribution; and counts and percentages (n, %) were utilized to convey data for categorical variables. To compare continuous and categorical variables, t-tests and Chi-square were wielded. The prediction result of each model was displayed on an ROC curve, and the prediction performance was evaluated by calculating the AUC, accuracy, sensitivity, specificity, PPV, and NPV. The Delong test was utilized to verify the significance of the AUC from different ROC curves. The Hosmer-Lemeshow test was employed to evaluate the fitting ability of the model. The entire workflow of this study is illustrated in Figure 2.

3 Results

3.1 Patient characteristics

A total of 213 glioma patients fulfilled the requirements for admission from the two centers, 132 patients from center 1 were classified as the training cohort and internal validation set (84HGGs, 48LGGs, mean age 52 years), and the remaining 81 patients from center 2 were zoned as the independent external testing cohort (50HGGs, 31LGGs, mean age 52 years). Table 1 displays the variations in the clinical features of the two groups from different centers. It can be shown that the only factor that significantly distinguished HGG from LGG was age ($p<0.001$). The gender and tumor location were not proposed as potential predictors of glioma grade.

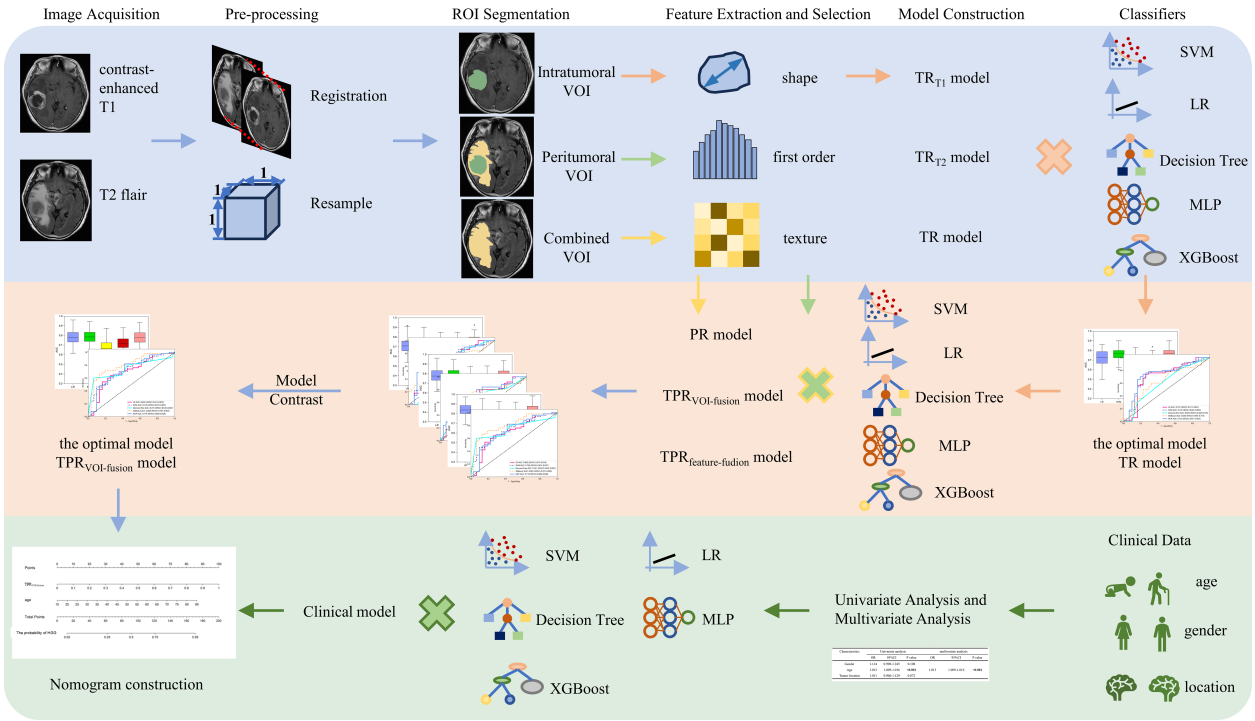


FIGURE 2
The workflow of intratumoral and peritumoral radiomics and clinical analysis for image acquisition, segmentation, feature extraction and selection, model comparison, and nomogram construction in our study.

TABLE 1 Demographic information and clinical characteristics of glioma patients.

Characteristics	Center 1 LGG (n=48)	Center 1 HGG (n=84)	P value	Center 2 LGG (n=31)	Center 2 HGG (n=50)	P value
Gender			0.493			0.21
Male	26 (54.17%)	52 (61.90%)		14 (45.16%)	31 (62.00%)	
Female	22 (45.83%)	32 (38.10%)		17 (54.84%)	19 (38.00%)	
Age	44.81±14.73	56.31±14.24	**<0.001***	44.00±11.57	56.26±13.11	**<0.001***
Tumor location			0.755			0.876
Right brain	25 (52.08%)	40 (47.62%)		16 (51.61%)	28 (56.00%)	
Left brain	23 (47.92%)	44 (52.38%)		15 (48.39%)	22 (44.00%)	

A t-test was used for age. A χ^2 test was used for the rest. *p<0.05
LGG, low-grade glioma; HGG, high-grade glioma.
Significant p values (p< 0.05) are indicated in bold.

3.2 Machine learning models based on intratumoral radiomics

For intratumoral radiomics, a total of 2576 (1288 contrast-enhanced T1-based and 1288 T2 flair-based radiomic features) features were extracted from each tumor region, containing shape, first-order, and texture features. Hence the TR_{T1}, TR_{T2}, and TR models included 1288, 1288, and 2576 radiomic features respectively. The feature statistics of categories and distribution are presented in Supplementary Figure S1. After feature filtration of the ICC test and Pearson's rank correlation coefficient, the LASSO with 10-fold cross-validation was employed to select significant features for building radiomic signatures. At last, 7 features were selected for the TR_{T1} model (Supplementary Figure S2A), 7 features were selected for the TR_{T2} model (Supplementary Figure S2B), and 6 features were selected for the TR model (Supplementary Figure S2C), with weighting coefficient severally. The coefficients and mean standard error (MSE) of the 10-fold validation are also exhibited in Supplementary Figure S2.

Then, 15 machine learning models were constructed based on the various radiomic models above-mentioned and machine learning classifiers (LR, SVM, XGBoost, DT, and MLP) to

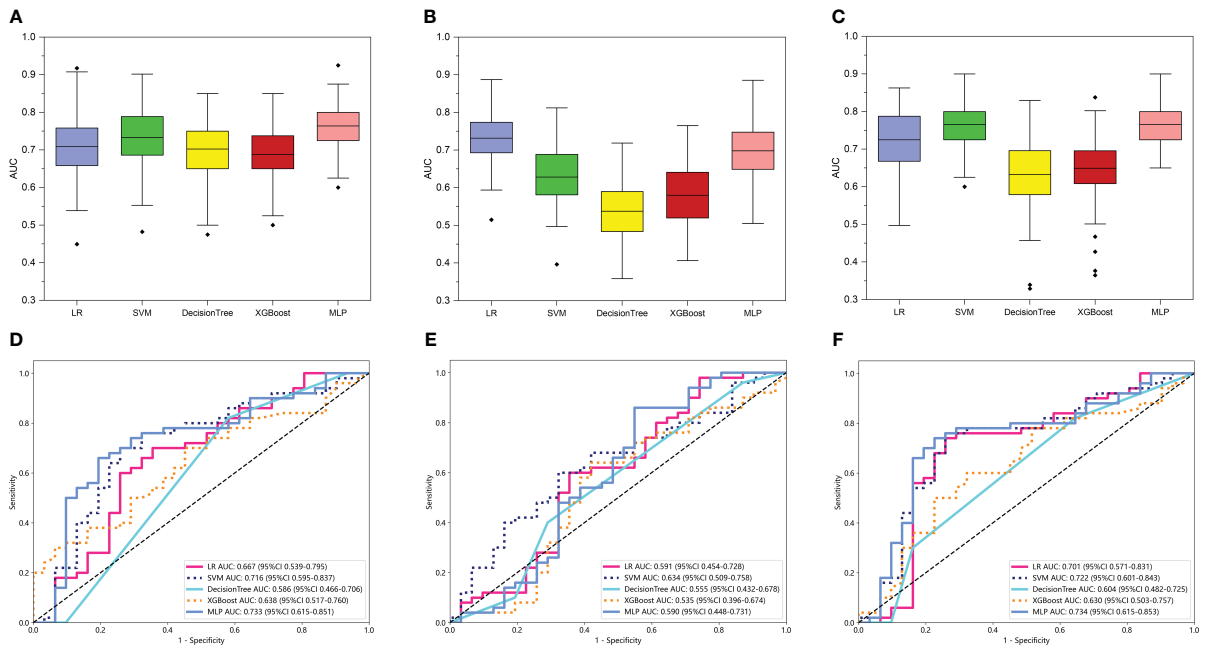


FIGURE 3
The performance of intratumoral radiomic models based on the different MRI sequences. The performance of the internal validation employed 30% of the training cohort randomly, and repeated the steps 100 times of TR_{T1} **(A)**, TR_{T2} **(B)**, and TR **(C)** models respectively. Comparison of ROC curves for the external test of TR_{T1} **(D)**, TR_{T2} **(E)**, and TR **(F)** models of five machine learning models. The models that the mean AUC of the internal validation throughout the 100 repetitions reached more than 0.7, were considered to be relatively stable and efficient. Comprehensively, the MLP-based TR model (AUC = 0.734 in external test) based on the two MRI sequences was considered the optimal model and selected for further analysis.

determine which model was optimal for glioma grade prediction. The results of internal validation with randomized division repeated 100 times are shown in Figures 3A–C for the TR_{T1}, TR_{T2}, and TR models in sequence. Among these five categories of machine learning models, the MLP-based TR_{T1} model, LR-based TR_{T2} model, and MLP-based TR model, which had the highest mean values of 0.76, 0.73, and 0.77, respectively, were considered the most stable models. In the external testing set, the performance of each category classifier on the TR_{T1} and TR model outperformed that of the TR_{T2} model respectively (Figure 3E, Supplementary Table S2). Ultimately, the MLP-based TR model (Figure 3F) with an AUC of 0.734 and MLP-based TR_{T1} model (Figure 3D) with an AUC of 0.733, were considered the optimal model in intratumoral radiomics. Although the Delong test showed a non-significantly statistic (p = 0.966) between these two models, the accuracy, sensitivity, and specificity of the MLP-based TR model were higher than the MLP-based TR_{T1} model. Hence, the TR model based on double-sequence MRI was chosen for further research and evaluation. The decision curves and calibration curves are also depicted in Supplementary Figure S4.

3.3 Machine learning models based on peritumoral radiomics

For peritumoral radiomics, the PR models contained 2576 radiomic features (Supplementary Figure S1). After feature selection, 8 features were finally selected for the PR model (Supplementary Figure S3A). Figure 4A shows the performance of 100-repetition randomized division internal validation. In the external testing cohort, only the AUC of the MLP-based PR model exceeded 0.7, indicating a lower predictive performance compared to the intratumoral radiomics (Figure 4D).

3.4 Machine learning models based on intra- and peri-tumoral fusion radiomics

For intra- and peri-tumoral fusion radiomics, the $TPR_{VOI\text{-}fusion}$ models contained 2576 radiomic features, while the $TPR_{feature\text{-}fusion}$ model included 5152 (Supplementary Figure S1), due to the different combinations of feature sources. After radiomic features dimensionality reduction, the LASSO regression finally selected 6 features for the $TPR_{VOI\text{-}fusion}$ model (Supplementary Figure S3B), and 9 features for the $TPR_{feature\text{-}fusion}$ model (Supplementary Figure S3C). Analogically, the two radiomic models were combined with 5 kinds of classifiers previously mentioned to develop machine-learning models for predicting glioma grade. The 100-repetition randomized division internal validation was conducted to evaluate the model's performance stability. Except for the DT model, the mean AUC of the other four machine learning models reached more than 0.7 in internal validation, which were considered to be relatively stable and efficient predicting models (Figures 4B, C). In the external testing, each categorical classifier on the $TPR_{VOI\text{-}fusion}$ model outperformed that of the $TPR_{feature\text{-}fusion}$ model respectively (Table 2, Figures 4E, F). The XGBoost-based $TPR_{VOI\text{-}fusion}$ model

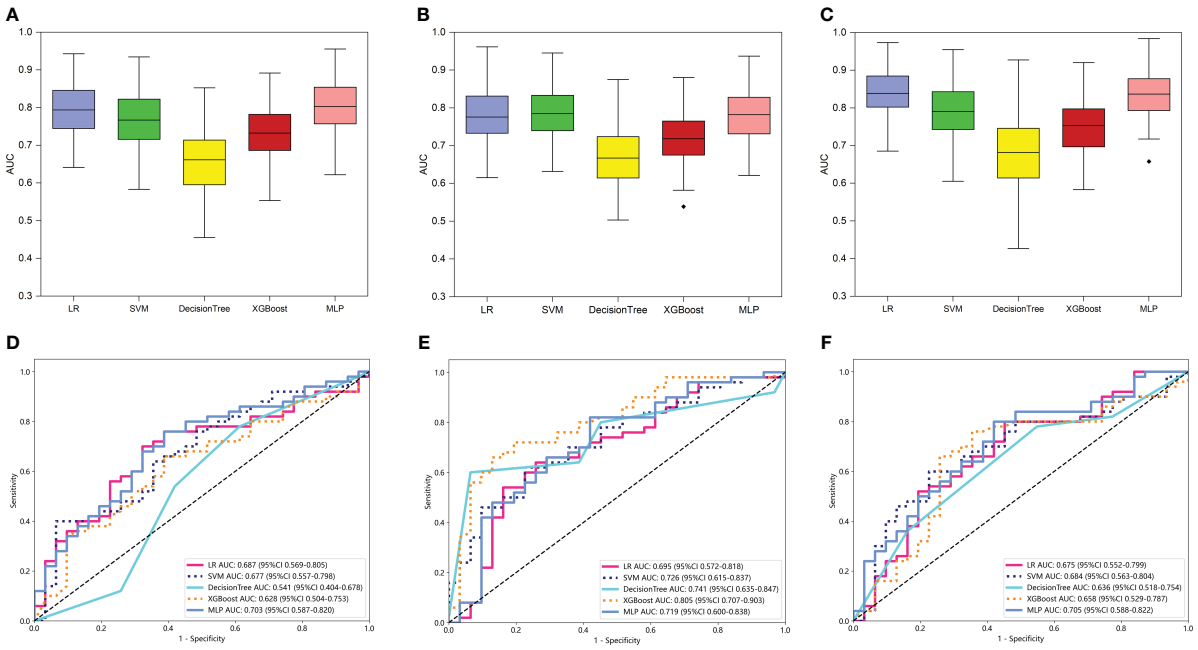


FIGURE 4
The performance of intratumoral and peritumoral radiomic models based on the two MRI sequences. The performance of the internal validation employed 30% of the training cohort randomly, and repeated the steps 100 times of PR **(A)**, $TPR_{VOI\text{-}fusion}$ **(B)**, and $TPR_{feature\text{-}fusion}$ **(C)** models respectively. Comparison of ROC curves for the external test of PR **(D)**, $TPR_{VOI\text{-}fusion}$ **(E)**, and $TPR_{feature\text{-}fusion}$ **(F)** models of five machine learning models. Ultimately, the XGBoost-based $TPR_{VOI\text{-}fusion}$ model (AUC = 0.805 in the external test) was identified as the best model to develop a nomogram.

TABLE 2 Each evaluation index of PR, $TPR_{VOI\text{-}fusion}$, and $TPR_{feature\text{-}fusion}$ models of five machine learning classifiers in the external test.

Model	Classifier	AUC	95%CI	Accuracy	Sensitivity	Specificity	PPV	NPV
PR	LR	0.687	0.5689-0.8053	0.667	0.760	0.516	0.717	0.571
	SVM	0.677	0.5570-0.7978	0.654	0.820	0.387	0.683	0.571
	DT	0.541	0.4037-0.6776	0.556	0.540	0.581	0.675	0.439
	XGBoost	0.628	0.5038-0.7529	0.630	0.680	0.548	0.708	0.515
	MLP	0.703	0.5867-0.8197	0.667	0.840	0.387	0.689	0.600
TPR VOI-fusion	LR	0.695	0.5718-0.8178	0.654	0.820	0.387	0.683	0.571
	SVM	0.726	0.6147-0.8370	0.654	0.840	0.355	0.677	0.579
	DT	0.741	0.6352-0.8474	0.704	0.800	0.548	0.741	0.630
	XGBoost	0.805	0.7069-0.9034	0.691	0.820	0.484	0.719	0.625
	MLP	0.719	0.6002-0.8385	0.667	0.840	0.387	0.689	0.600
TPR feature-fusion	LR	0.675	0.5516-0.7993	0.704	0.800	0.548	0.741	0.630
	SVM	0.684	0.5634-0.8044	0.691	0.800	0.516	0.727	0.615
	DT	0.636	0.5181-0.7536	0.654	0.780	0.452	0.696	0.560
	XGBoost	0.658	0.5293-0.7868	0.667	0.800	0.452	0.702	0.583
	MLP	0.705	0.5880-0.8224	0.667	0.840	0.387	0.689	0.600

AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

with an AUC of 0.805 was considered the optimal model (Figure 4E) in all intratumoral and/or peritumoral radiomic models. Supplementary Figure S5 exhibits the decision curves and calibration curves of corresponding models.

3.5 Nomogram construction

The predictive clinical parameters were chosen based on univariate and multivariate analyses. As shown in Table 3, age was found to be an independent predictor (OR 1.013; 95% CI 1.009-1.016; p<0.001) for glioma grade prediction. Then, the Clinical models were built based on the selected independent predictor and 5 kinds of classifiers.

Subsequently, in exploring the potential utility of the developed MRI-based intratumoral and peritumoral radiomic models for preoperative prediction of glioma grade, a nomogram was constructed by combining the clinical independent predictor with the optimal $TPR_{VOI\text{-}fusion}$ machine learning model (Figure 5A). The individual risk of being predicted as an HGG glioma was derived from the cumulative total points obtained, which allowed for the representation of the prediction model in a more simplified and comprehensive manner. Figure 5B illustrates the superior performance of the nomogram with an AUC of 0.825 (testing cohort), in comparison to both the Clinical model and $TPR_{VOI\text{-}fusion}$ radiomic model. After the Delong test, the nomogram was proved to significantly outperform both of the models (Figure 5C), and the net benefit in the DCA curve of the nomogram was higher than that of the two models at threshold probabilities in the testing cohort (Figure 5D). The Hosmer-Lemeshow test revealed favorable calibration of the nomogram (p = 0.089), suggesting alignment with an ideal fit without significant deviation (Figure 5E).

4 Discussion

In this study, we developed variously dependable models to preoperatively predict grade glioma by using MRI images, which

TABLE 3 Univariate analysis and multivariate analysis of clinical characteristics in all patients.

Characteristics	Univariate analysis			multivariate analysis		
	OR	95%CI	P value	OR	95%CI	P value
Gender	1.114	0.998-1.245	0.108			
Age	1.013	1.009-1.016	**<0.001***	1.013	1.009-1.016	**<0.001***
Tumor location	1.011	0.906-1.129	0.872			

OR, odds ratio; CI, confidence interval.
*p<0.05.
Significant p values (p< 0.05) are indicated in bold.

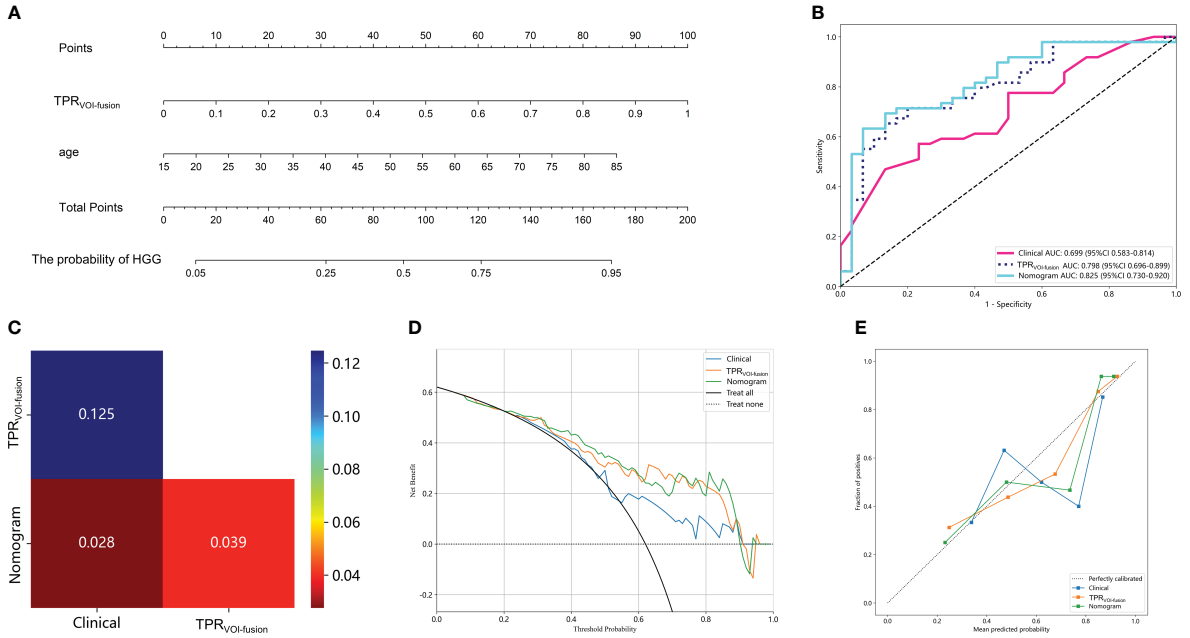


FIGURE 5
Clinical application of the nomogram constructed by radiomic model and clinical parameter in predicting the probability of being HGG for glioma patients **(A)**. The total point was obtained by adding the scores located on the $TPR_{VOI\text{-}fusion}$ and age coordinate axis together, and the vertically corresponding value on the bottom line was the probability of being HGG. **(B)** Comparison of ROC curves of the Clinical model, $TPR_{VOI\text{-}fusion}$ radiomic model, and nomogram. **(C)** The Delong test among the three models. The nomogram with an AUC of 0.825 in the test cohort, significantly outperformed the other two models. The DCA **(D)** and calibration curves **(E)** of the three models. The nomogram had a higher net benefit in predicting glioma grade and represented an ideal fit.

were constructed by radiomic features extracted from the intratumoral region and peritumoral edema region. The XGBoost machine learning classifier, incorporating features extracted from the combination of intratumoral and peritumoral VOIs of MRI, exhibited superior performance in distinguishing between LGG and HGG. The internal validation and independent external test underscored the robustness and generalizability of the model. Additionally, the radiomic models derived from VOI-fusion outperformed those derived from feature-fusion in our study, suggesting more extensive investigations into peritumoral radiomics were necessary to determine a more standardized research method and provide more theoretical support for radiomic studies. Finally, a nomogram based on the optimal machine learning model and clinical parameter was established to detect potential applications for predicting glioma grade in clinical practice.

MRI serves as a routine tool for preliminary diagnosing, treatment planning, and monitoring the treatment response of patients with glioma (39, 40). Recent studies have consistently demonstrated a robust association between radiomic features extracted from multiparametric MRI scans and various applications related to gliomas (2, 41). Kim et al. distinguished between glioblastoma and primary CNS lymphoma using multiparametric MRI sequences that included contrast-enhanced T1-weighted, T2-weighted, and diffusion-weighted imaging (42). Fifteen features were chosen from all imaging modalities, and the model rendered an AUC of 0.979, demonstrating the potent prediction potential of multi-parameter MRI. Nevertheless, the prediction of models based on a single MRI sequence was not performed and compared. In our study, we constructed and compared various radiomic models based on the intratumoral region of the single contrast-enhanced T1 sequence, the single T2 flair sequence, and the combination of two sequences. As expected, the radiomic models derived from dual-sequence MR imaging outperformed those solely based on single contrast-enhanced T1 or single T2 flair sequence. Hence, multicontrast MRI-based radiomics was poised to enhance the predictive capability for glioma grading compared to that of single MRI sequence.

Previous studies have demonstrated that the heterogeneity of gliomas extended beyond the tumor interior to include the peritumoral region, where approximately 90% of gliomas recurred (43, 44). Glioma cells interacted with molecules in the peritumoral area to cause hypoxia, angiogenesis, and tumor infiltration, which would ultimately accelerate the growth of gliomas (45). Consequently, the significant potential existed in the peritumoral environment, which could provide important information for evaluating the aggressive biological behavior of the tumor clinically (46, 47). In studies of other cancers, Ding et al. investigated the effect of peritumoral features for predicting sentinel lymph node metastasis in breast cancer (48). They created peritumoral regions by expanding tumor regions of interest at thicknesses of 2 mm, 4 mm, 6 mm, and 8 mm. By incorporating peritumoral features, the accuracy in the validation set increased from 0.704 to 0.796. Shan et al. developed a prediction model using peritumoral radiomic signatures extracted from a 2 cm peritumoral area, assessing its effectiveness in predicting the early

recurrence of hepatocellular carcinoma post-curative treatment (49). In the validation cohort, ROC curves and decision curves revealed superior prediction efficiency and greater clinical benefits with the peritumoral model. Consequently, the extraction and integration of peritumoral and intratumoral features present a promising avenue. However, existing radiomic-based techniques for grading gliomas focused primarily on the interior of the tumor and less on the peritumoral environment.

Peripheral edema and peritumor in gliomas are two distinct concepts, the peritumor is typically an area within a specified radius around the tumor, whereas the peripheral edema around a glioma is irregular and frequently dispersed along the cerebral gyrus. Previous research has shown that the degree of peritumoral edema increases with the pathological grade and aggressiveness of glioma (14). It was demonstrated that individuals with severe edema (>10 mm) had mean or overall survival rates that were more than 50% lower than those with mild edema (8, 9, 44). Cheng et al. compared the predictive power of the peritumor and peripheral edema region for grading gliomas and found the most predictive features were extracted from the peritumor region within an immediate distance of 1 mm from the tumor core based on MRI scans (8). Notably, the study did not explore a predictive model based on the peripheral edema region in combination with the intratumoral region. Considering that the prognosis of gliomas was strongly correlated with the occurrence of peritumoral edema, we attempted to investigate the radiomic models based on peripheral edema for glioma grade.

Concerning the two main research approaches in intratumoral and peritumoral radiomic analysis, which was simplified as feature-fusion and VOI-fusion, definitive studies remained absent establishing a persuasive conclusion about which approach made more sense and produced more predictive features. Our study filled this gap by conducting a comparative analysis of the two methodologies for the first time. In the internal validation, the AUC of models based on both two methods showed no obvious differences, the difference in AUC was consistent across classifiers. In the external test, each categorical classifier on the VOI-fusion model outperformed that of the feature-fusion model respectively, which indicated that the model constructed by extracting features from the intratumoral and peritumoral regions as a whole yielded higher prediction efficiency. Regarding VOI fusion, the radiomic features, such as shape, first-order, and texture, extracted comprehensive information taking into account both intratumoral and peripheral edema region. Regarding feature fusion, individual peritumoral features exhibited minimal statistical variance in effectively distinguishing HGG from LGG, rendering them prone to elimination during screening. In our study, the constructed model just by feature fusion performed poorly in external tests, indicating unstable performance. This highlighted the need for further studies to unravel the intricacies of intratumoral and peritumoral radiomics.

We assessed 1288 radiomics features for every MRI sequence, and 7728 features in total, which was distinctly more than most recent findings and included all significant variables for radiomic analysis (30, 45). To identify the most optimal model for our dataset, we applied the six previously discussed categories of radiomic models across five classifiers: LR, SVM, XGBoost, DT, and MLP. This meticulous process ensured the exploration of fully optimized models best suited for our data. Meanwhile, previous studies have identified several clinical parameters crucial in distinguishing between LGG and HGG. Wang et al. selected clinical factors including age and sex as well as radiomic signature to develop a nomogram to predict glioma grading (30). It is well-recognized that HGG tends to be diagnosed in the elderly (31, 50). Consistently, we found that only the age parameter was statistically significant in predicting glioma grading, using the uni-multivariate analyses. Despite the lower predictive capacity of clinical model based solely on age feature compared to radiomic models, the nomogram amalgamating clinical parameter and radiomic models surpassed the predictive efficacy of either model in isolation. In terms of the predictive efficiency of various machine learning methods, Voort et al. utilized a deep learning model to predict the grade of glioma, achieving an AUC of 0.81 in the external testing cohort of 240 patients from 13 different institutes (31). Similarly, Li et al. distinguished LGG from HGG by developing deep convolutional neural network models, achieving an AUC of 0.89. In comparison, the nomogram in our study showed great performance with an AUC of 0.825 in the independent external testing cohort, which was equivalent to the state-of-the-art research aforementioned.

For all this, there were limitations in this study. First, our study required a larger sample size from more centers to make the findings more convincing. Second, only two MRI sequences were employed in this study. Some advanced parametric MRI scans, such as DWI and DTI, have shown powerful potential in tumor research, and new scanning techniques should be explored (7, 18, 36). Third, the VOIs in our study were manually annotated, which was time-consuming and laborious, and even prone to inaccurate annotation. Deep learning-based tumor segmentation methods are expected to be employed to improve the accuracy and reliability of image segmentation. Finally, our study lacked molecular subtyping of the samples, which was critical for the prognosis of gliomas, and planned to integrate such information in future studies.

5 Conclusion

In this work, we assessed the function of radiomic models of intratumoral and peripheral edema regions in MRI scans for predicting glioma grade and validated the methodology on an independent external test dataset, which provided a fresh viewpoint on the disease. The nomogram combined clinical parameter and the optimal radiomic model was efficient in glioma grade, and this non-invasive approach was expected to promote clinical research and guide the management of individualized glioma treatment.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Tianjin Medical University General Hospital's Institutional Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because The requirement for written informed consent forms from all patients taking part in the study was waived because of the retrospective investigation.

Author contributions

RT: Data curation, Writing – original draft. CS: Methodology, Writing – original draft. CW: Writing – original draft. TZ: Conceptualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by Scientific Research Program of Tianjin Municipal Science and Technology Bureau (NO. 19ZXDBSY00040).

Acknowledgments

We express our gratitude for the technical support offered by the OnekeyAI platform.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1401977/full#supplementary-material

References

1. Abdel Razek AAK, Alksas A, Shehata M, AbdelKhalek A, Abdel Baky K, El-Baz A, et al. Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging. *Insights Imaging*. (2021) 12:152. doi: 10.1186/s13244-021-01102-6

2. Singh G, Manjila S, Sakla N, True A, Wardeh AH, Beig N, et al. Radiomics and radiogenomics in gliomas: a contemporary update. *Br J Cancer*. *Aug*. (2021) 125:641–57. doi: 10.1038/s41416-021-01387-w

3. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

4. Yi Z, Long L, Zeng Y, Liu Z. Current advances and challenges in radiomics of brain tumors. *Front Oncol*. (2021) 11:732196. doi: 10.3389/fonc.2021.732196

5. Liu D, Chen J, Hu X, Yang K, Liu Y, Hu G, et al. Imaging-genomics in glioblastoma: Combining molecular and imaging signatures. *Front Oncol*. (2021) 11:699265. doi: 10.3389/fonc.2021.699265

6. Lee DH, Park JE, Kim N, Park SY, Kim YH, Cho YH, et al. Tumor habitat analysis by magnetic resonance imaging distinguishes tumor progression from radiation necrosis in brain metastases after stereotactic radiosurgery. *Eur Radiol*. (2022) 32:497–507. doi: 10.1007/s00330-021-08204-1

7. Yang Y, Yan LF, Zhang X, Han Y, Nan HY, Hu YC, et al. Glioma grading on conventional MR images: A deep learning study with transfer learning. *Front Neurosci*. (2018) 12:804. doi: 10.3389/fnins.2018.00804

8. Cheng J, Liu J, Yue H, Bai H, Pan Y, Wang J. Prediction of glioma grade using intratumoral and peritumoral radiomic features from multiparametric MRI images. *IEEE/ACM Trans Comput Biol Bioinform*. (2022) 19:1084–95. doi: 10.1109/TCBB.2020.3033538

9. Li Y, Ammari S, Lawrance L, Quillent A, Assi T, Lassau N, et al. Radiomics-based method for predicting the glioma subtype as defined by tumor grade, IDH mutation, and 1p/19q codeletion. *Cancers (Basel)*. (2022) 14. doi: 10.3390/cancers14071778

10. Hooper GW, Ginat DT. MRI radiomics and potential applications to glioblastoma. *Front Oncol*. (2023) 13:1134109. doi: 10.3389/fonc.2023.1134109

11. Calabrese E, Rudie JD, Rauschecker AM, Villanueva-Meyer JE, Clarke JL, Solomon DA, et al. Combining radiomics and deep convolutional neural network features from preoperative MRI for predicting clinically relevant genetic biomarkers in glioblastoma. *Neurooncol Adv*. (2022) 4:vdac060. doi: 10.1093/noajnl/vdac060

12. Park JE, Kim HS, Kim N, Park SY, Kim YH, Kim JH. Spatiotemporal heterogeneity in multiparametric physiologic MRI is associated with patient outcomes in IDH-wildtype glioblastoma. *Clin Cancer Res*. (2021) 27:237–45. doi: 10.1158/1078-0432.CCR-20-2156

13. Waqar M, Van Houdt PJ, Hessen E, Li KL, Zhu X, Jackson A, et al. Visualising spatial heterogeneity in glioblastoma using imaging habitats. *Front Oncol*. (2022) 12:1037896. doi: 10.3389/fonc.2022.1037896

14. Brancato V, Cerrone M, Lavitrano M, Salvatore M. Cavaliere C. A systematic review of the current status and quality of radiomics for glioma differential diagnosis. *Cancers (Basel)*. (2022) 14. doi: 10.3390/cancers14112731

15. Beig N, Bera K, Prasanna P, Antunes J, Correa R, Singh S, et al. Radiogenomic-based survival risk stratification of tumor habitat on gd-T1w MRI is associated with biological processes in glioblastoma. *Clin Cancer Res*. (2020) 26:1866–76. doi: 10.1158/1078-0432.CCR-19-2556

16. Gillies RJ, Balagurunathan Y. Perfusion MR imaging of breast cancer: Insights using "Habitat imaging". *Radiology*. (2018) 288:36–7. doi: 10.1148/radiol.2018180271

17. Wu J, Cao G, Sun X, Lee J, Rubin DL, Napel S, et al. Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy. *Radiology*. (2018) 288:26–35. doi: 10.1148/radiol.2018172462

18. Kim M, Jung SY, Park JE, Jo Y, Park SY, Nam SJ, et al. Diffusion- and perfusion-weighted MRI radiomics model may predict isocitrate dehydrogenase (IDH) mutation and tumor aggressiveness in diffuse lower grade glioma. *Eur Radiol*. (2020) 30:2142–51. doi: 10.1007/s00330-019-06548-3

19. Zhou H, Xu R, Mei H, Zhang L, Yu Q, Liu R, et al. Application of enhanced T1WI of MRI radiomics in glioma grading. *Int J Clin Pract*. (2022) 2022:3252574. doi: 10.1155/2022/3252574

20. Wu J, Gong G, Cui Y, Li R. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. *J Magn Reson Imaging*. (2016) 44:1107–15. doi: 10.1002/jmri.25279

21. Aoude LG, Wong BZY, Bonazzi VF, Brosda S, Walters SB, Koufariotis LT, et al. Radiomics biomarkers correlate with CD8 expression and predict immune signatures in melanoma patients. *Mol Cancer Res*. (2021) 19:950–6. doi: 10.1158/1541-7786.MCR-20-1038

22. Wang X, Yang W, Zhou Q, Luo H, Chen W, Yeung SJ, et al. The role of (18)F-FDG PET/CT in predicting the pathological response to neoadjuvant PD-1 blockade in combination with chemotherapy for resectable esophageal squamous cell carcinoma. *Eur J Nucl Med Mol Imaging*. (2022) 49:4241–51. doi: 10.1007/s00259-022-05872-z

23. Hannequin P, Decroisette C, Kermanach P, Berardi G, Bourbonne V. FDG PET and CT radiomics in diagnosis and prognosis of non-small-cell lung cancer. *Transl Lung Cancer Res*. (2022) 11:2051–63. doi: 10.21037/tlcr-22-158

24. Napel S, Mu W, Jardim-Perassi BV, Aerts H, Gillies RJ. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. *Cancer*. (2018) 124:4633–49. doi: 10.1002/cncr.31630

25. Wang X, Xie T, Luo J, Zhou Z, Yu X, Guo X. Radiomics predicts the prognosis of patients with locally advanced breast cancer by reflecting the heterogeneity of tumor cells and the tumor microenvironment. *Breast Cancer Res*. (2022) 24. doi: 10.1186/s13058-022-01516-0

26. Cui L, Yu T, Kan Y, Dong Y, Luo Y, Jiang X. Multi-parametric MRI-based peritumoral radiomics on prediction of lymph-vascular space invasion in early-stage cervical cancer. *Diagn Interv Radiol*. (2022) 28:312–21. doi: 10.5152/dir.2022.20657

27. Stoyanova R, Chinea F, Kwon D, Reis IM, Tschudi Y, Parra NA, et al. An automated multiparametric MRI quantitative imaging prostate habitat risk scoring system for defining external beam radiation therapy boost volumes. *Int J Radiat Oncol Biol Phys*. (2018) 102:821–9. doi: 10.1016/j.ijrobp.2018.06.003

28. Gemini L, Tortora M, Giordano P, Prudente ME, Villa A, Vargas O, et al. Vasari scoring system in discerning between different degrees of glioma and IDH status prediction: A possible machine learning application? *J Imaging*. (2023) 9. doi: 10.3390/jimaging9040075

29. You W, Mao Y, Jiao X, Wang D, Liu J, Lei P, et al. The combination of radiomics features and VASARI standard to predict glioma grade. *Front Oncol*. (2023) 13:1083216. doi: 10.3389/fonc.2023.1083216

30. Wang Q, Li Q, Mi R, Ye H, Zhang H, Chen B, et al. Radiomics nomogram building from multiparametric MRI to predict grade in patients with glioma: A cohort study. *J Magn Reson Imaging. Mar*. (2019) 49:825–33. doi: 10.1002/jmri.26265

31. van der Voort SR, Incekara F, Wijnenga MMJ, Kapsas G, Gahrmann R, Schouten JW, et al. Combined molecular subtyping, grading, and segmentation of glioma using multi-task deep learning. *Neuro Oncol*. (2023) 25:279–89. doi: 10.1093/neuonc/noac166

32. Li Y, Wei D, Liu X, Fan X, Wang K, Li S, et al. Molecular subtyping of diffuse gliomas using magnetic resonance imaging: comparison and correlation between radiomics and deep learning. *Eur Radiol*. (2022) 32:747–58. doi: 10.1007/s00330-021-08237-6

33. Tian Y, Hua H, Peng Q, Zhang Z, Wang X, Han J, et al. Preoperative evaluation of gd-EOB-DTPA-enhanced MRI radiomics-based nomogram in small solitary hepatocellular carcinoma (</=3 cm) with microvascular invasion: A two-center study. *J Magn Reson Imaging*. (2022) 56:1459–72. doi: 10.1002/jmri.28157

34. Shi J, Dong Y, Jiang W, Qin F, Wang X, Cui L, et al. MRI-based peritumoral radiomics analysis for preoperative prediction of lymph node metastasis in early-stage cervical cancer: A multi-center study. *Magn Reson Imaging*. (2022) 88:1–8. doi: 10.1016/j.mri.2021.12.008

35. Li N, Wan X, Zhang H, Zhang Z, Guo Y, Hong D. Tumor and peritumor radiomics analysis based on contrast-enhanced CT for predicting early and late recurrence of hepatocellular carcinoma after liver resection. *BMC Cancer*. (2022) 22:664. doi: 10.1186/s12885-022-09743-6

36. Su C, Jiang J, Zhang S, Shi J, Xu K, Shen N, et al. Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour. *Eur Radiol*. (2019) 29:1986–96. doi: 10.1007/s00330-018-5704-8

37. Xie D, Xu F, Zhu W, Pu C, Huang S, Lou K, et al. Delta radiomics model for the prediction of progression-free survival time in advanced non-small-cell lung cancer patients after immunotherapy. *Front Oncol*. (2022) 12:990608. doi: 10.3389/fonc.2022.990608

38. Chen K, Wang J, Li S, Zhou W, Xu W. Predictive value of 18F-FDG PET/CT-based radiomics model for neoadjuvant chemotherapy efficacy in breast cancer: a multi-scanner/center study with external validation. *Eur J Nucl Med Mol Imaging*. (2023) 50:1869–80. doi: 10.1007/s00259-023-06150-2

39. Gu D, Xie Y, Wei J, Li W, Ye Z, Zhu Z, et al. MRI-based radiomics signature: A potential biomarker for identifying glypican 3-positive hepatocellular carcinoma. *J Magn Reson Imaging*. (2020) 52:1679–87. doi: 10.1002/jmri.27199

40. Yan J, Zhang B, Zhang S, Cheng J, Liu X, Wang W, et al. Quantitative MRI-based radiomics for noninvasively predicting molecular subtypes and survival in glioma patients. *NPJ Precis Oncol*. (2021) 5:72. doi: 10.1038/s41698-021-00205-z

41. Xu C, Peng Y, Zhu W, Chen Z, Li J, Tan W, et al. An automated approach for predicting glioma grade and survival of LGG patients using CNN and radiomics. *Front Oncol*. (2022) 12:969907. doi: 10.3389/fonc.2022.969907

42. Kim Y, Cho HH, Kim ST, Park H, Nam D, Kong DS. Radiomics features to distinguish glioblastoma from primary central nervous system lymphoma on multi-parametric MRI. *Neuroradiology*. (2018) 60:1297–305. doi: 10.1007/s00234-018-2091-4

43. Bailo M, Pecco N, Callea M, Scifo P, Gagliardi F, Presotto L, et al. Decoding the heterogeneity of Malignant gliomas by PET and MRI for spatial habitat analysis of hypoxia, perfusion, and diffusion imaging: A preliminary study. *Front Neurosci*. (2022) 16:885291. doi: 10.3389/fnins.2022.885291

44. Shaheen A, Bukhari ST, Nadeem M, Burigat S, Bagci U, Mohy-Ud-Din H. Overall survival prediction of glioma patients with multiregional radiomics. *Front Neurosci*. (2022) 16:911065. doi: 10.3389/fnins.2022.911065

45. Li G, Li L, Li Y, Qian Z, Wu F, He Y, et al. An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. *Brain*. (2022) 145:1151–61. doi: 10.1093/brain/awab340

46. Kazerouni AS, Hormuth DA 2nd, Davis T, Bloom MJ, Mounho S, Rahman G, et al. Quantifying tumor heterogeneity via MRI habitats to characterize microenvironmental alterations in HER2+ Breast cancer. *Cancers (Basel)*. (2022) 14. doi: 10.3390/cancers14071837

47. Algohary A, Shiradkar R, Pahwa S, Purysko A, Verma S, Moses D, et al. Combination of peri-tumoral and intra-tumoral radiomic features on bi-parametric MRI accurately stratifies prostate cancer risk: A multi-site study. *Cancers*. (2020) 12. doi: 10.3390/cancers12082200

48. Ding J, Chen S, Serrano Sosa M, Cattell R, Lei L, Sun J, et al. Optimizing the peritumoral region size in radiomics analysis for sentinel lymph node status prediction in breast cancer. *Acad Radiol*. (2022) 29 Suppl 1:S223–8. doi: 10.1016/j.acra.2020.10.015

49. Shan Q-Y, Hu H-T, Feng S-T, Peng Z-p, Chen S-I, Zhou Q, et al. CT-based peritumoral radiomics signatures to predict early recurrence in hepatocellular carcinoma after curative tumor resection or ablation. *Cancer Imaging*. (2019) 19. doi: 10.1186/s40644-019-0197-5

50. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep*. (2017) 7:10353. doi: 10.1038/s41598-017-10649-8

frontiers | Frontiers in Oncology

TYPE Systematic Review
PUBLISHED 30 July 2024
DOI 10.3389/fonc.2024.1409760

OPEN ACCESS

EDITED BY
Camilla Russo,
Santobono-Pausilipon Children's Hospital, Italy

REVIEWED BY
João Santinha,
Champalimaud Foundation, Portugal
Flavio Giordano,
Hospital Antonio Cardarelli, Italy

*CORRESPONDENCE
Junqiang Lei
Leijq2011@126.com

RECEIVED 31 March 2024
ACCEPTED 08 July 2024
PUBLISHED 30 July 2024

CITATION
Chen X, Lei J, Wang S, Zhang J and Gou L (2024) Diagnostic accuracy of a machine learning-based radiomics approach of MR in predicting IDH mutations in glioma patients: a systematic review and meta-analysis. *Front. Oncol.* 14:1409760. doi: 10.3389/fonc.2024.1409760

Diagnostic accuracy of a machine learning-based radiomics approach of MR in predicting IDH mutations in glioma patients: a systematic review and meta-analysis

Xiaoli Chen[1,2,3], Junqiang Lei[1,2,3]*, Shuaiwen Wang[1,2,3], Jing Zhang[4] and Lubin Gou[1,2,3]

[1]Department of Radiology, The First Hospital of Lanzhou University, Lanzhou, China, [2]Intelligent Imaging Medical Engineering Research Center of Gansu Province, Lanzhou, China, [3]Accurate Image Collaborative Innovation International Science and Technology Cooperation Base of Gansu Province, Lanzhou, China, [4]Department of MRI, Shaanxi Provincial People's Hospital, Xi'an, China

Objectives: To assess the diagnostic accuracy of machine learning (ML)-based radiomics for predicting isocitrate dehydrogenase (IDH) mutations in patients with glioma.

Methods: A systematic search of PubMed, Web of Science, Embase, and the Cochrane Library from inception to 1 September 2023, was conducted to collect all articles investigating the diagnostic performance of ML for the prediction of IDH mutations in gliomas. Two reviewers independently screened all papers for eligibility. Methodological quality and risk of bias were assessed using the METhodological RadiomICs Score and Quality Assessment of Diagnostic Accuracy Studies-2, respectively. The pooled sensitivity, specificity, and 95% confidence intervals were calculated, and the area under the receiver operating characteristic curve (AUC) was obtained.

Results: In total, 14 original articles assessing 1740 patients with gliomas were included. The AUC of ML for predicting IDH mutation was 0.90 (0.87–0.92). The pooled sensitivity, specificity, and diagnostic odds ratio were 0.83 (0.71–0.90), 0.84 (0.74–0.90), and 25 (12,50) respectively. In subgroup analyses, modeling methods, glioma grade, and the combination of magnetic resonance imaging and clinical features affected the diagnostic performance in predicting IDH mutations in gliomas.

Conclusion: ML-based radiomics demonstrated excellent diagnostic performance in predicting IDH mutations in gliomas. Factors influencing the diagnosis included the modeling methods employed, glioma grade, and whether the model incorporated clinical features.

Systematic review registration: https://www.crd.york.ac.uk/PROSPERO/#myprospero, PROSPERO registry (CRD 42023395444).

KEYWORDS

glioma, isocitrate dehydrogenase (IDH), MRI, machine learning, deep learning, radiomics

Introduction

The 2016 World Health Organization (WHO) classification of central nervous system tumors incorporated molecular markers (1). The 2021 WHO classification emphasizes the role of molecular markers in both the classification and grading of gliomas (2). The primary markers for gliomas include isocitrate dehydrogenase (IDH), classified as IDH-mutant, 1p/19q-non-codeleted (IDHmut-Noncodel), and IDH wild-type (IDHwt). Patient outcomes and therapeutic options in glioma vary across subtypes (3, 4). Patients with an IDH-mutated glioma have a better prognosis than those with an IDH wild-type tumor. Recent studies have demonstrated that IDH may be a potential therapeutic target for IDH-mutant gliomas (5). Therefore, preoperative prediction of IDH mutation status is important for prognosis and therapeutic decision-making. Although histopathology is the current diagnostic gold standard, it has some limitations such as sampling errors, complications, and invasiveness. Thus, noninvasive assessment of IDH mutation status is an urgent requirement.

Radiomics can transform images into mineable data for quantitative analysis through high-throughput extraction and analysis, providing support for decision-making (6). Machine learning and deep learning combined with radiomics have excellent potential for preoperative diagnosis, staging, and therapeutic effect evaluation of gliomas (7, 8), as well as for predicting IDH mutation status. A previous systematic review (9) dealing with this subject was published, but it was not quantitative enough to evaluate the predictive performance. In addition, because radiomics research is a complicated process that includes multiple stages, it is critical to evaluate the quality of the method to ensure the reliability and reproducibility of the model before use in clinical work.

The aim of this systematic review and meta-analysis was to evaluate the accuracy of radiomics models in predicting the IDH status of gliomas and to evaluate the methodological quality and risk of bias in radiomics workflows.

Materials and methods

This meta-analysis was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (10) guidelines and registered to the PROSPERO registry (registration number, CRD 42023395444).

Literature search and study selection

The PubMed, EMBASE, Cochrane Library, and Web of Science databases were searched up to 1 September 2023 by two reviewers, C.X.L and Z.J. To identify the relevant articles, only English articles were considered. The following keywords were used to identify relevant studies: ("Glioma" OR "Gliomas") AND ("Isocitrate Dehydrogenase" OR " IDH") AND ("MRI" OR "magnetic resonance imaging") AND ("machine learning" OR "radiomics" OR "deep learning" OR "Artificial Intelligence") The details of search strategies are provided in the Supplementary Materials.

The included articles fulfilled all the following criteria: 1) patients with pathologically confirmed glioma; 2) histopathological examination with the IDH mutation as a reference standard; 3) sufficient data for the reconstruction of 2×2 tables in terms of the diagnostic performance of MR-based radiomics in predicting the IDH of glioma; and 4) original research articles. The exclusion criteria were as follows: 1) each study had at least 10 patients; 2) reviews, case reports, letters, and editorials; 3) studies not focusing on the diagnostic performance of MR-based radiomics in predicting IDH mutations; and 4) insufficient data for the reconstruction of 2×2 table studies with overlapping cohorts. Two authors, C. X.L and Z.J, independently evaluated the eligibility of the included articles, and any disagreements were resolved via discussion with a third author (W.S.W, with 10 years of experience in neuroimaging).

Quality assessment and data extraction

The included articles' methodological quality and the risk of bias at the study level were assessed using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS)-2 (11) and METhodological RadiomICs Score (12), respectively. The QUADAS-2 tool included four parts: (a) patient selection, (b) index test, (c) reference standard, and (d) flow and timing. The METhodological RadiomICs Score (METRICS tool included 30 items within 9 categories for evaluating the quality of the radiomics workflow. Two reviewers, C.X.L and G.L.B, assessed the quality of the articles separately and resolved any disagreements through discussion with a third author (W.S.W).

The following data were extracted from the included articles: 1) study characteristics (authors, year of publication, country of origin, study design (prospective vs. retrospective)); 2) patient and clinical characteristics (number of patients, age, WHO grade, reference standard); 3) technical characteristics of magnetic resonance imaging (MRI) (magnetic field strength (T), scanner, scan sequence) and machine learning details (classifier, method of segmentation, VOI or ROI, and external or internal validation).

Statistical analysis

This meta-analysis was performed using Stata 16 Review Manager 5.3 software and Meta-disc 1.4. Pooled sensitivity, specificity, diagnostic odds ratio (DOR), positive likelihood ratio (PLR), and negative likelihood ratio (NLR) with 95% confidence intervals (CIs) were calculated using bivariate random effects, and a summary receiver operating characteristic (SROC) curve and area under the curve (AUC) were generated to illustrate the diagnostic performance.

The heterogeneity of the included studies was calculated using the Q-test (p value ≤ 0.05) and I^2 statistic (>50%) (13). A Spearman coefficient >0.6 indicated the threshold effect (14). Subgroup analysis was performed to further investigate the potential cause of heterogeneity, and the following four covariates were included: 1) machine learning (ML) vs. deep learning (DL), 2) only radiomics vs. combination of radiomics and clinical information, 3) low-grade glioma (LGG) vs. high-grade glioma (HGG), and 4) external validation vs. internal validation.

Results

Characteristics of included studies

The flowchart of the literature search and selection process is displayed in Figure 1, which yielded 161 studies from PubMed, 279 from Embase, four from the Cochrane Library, and 198 from the Web of Science. After removing 253 duplicate articles, the remaining 389 articles were screened for their title and abstract. The full text of 66 eligible articles was reviewed, and 14 articles (15–28) were included in this meta-analysis.

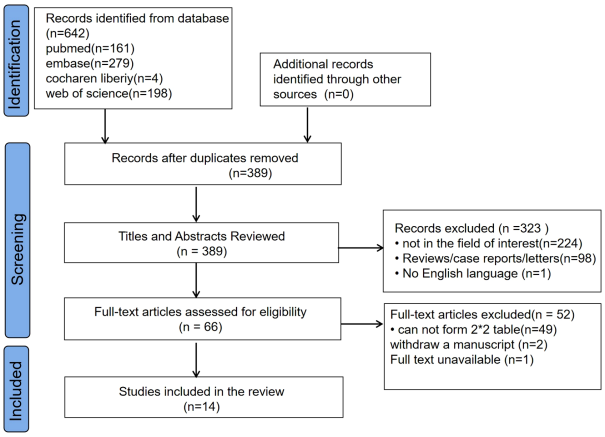


FIGURE 1
Flow chart of study selection.

The characteristics of the included studies are shown in Tables 1 and 2. One study was prospective, and the remaining studies were retrospective. Eight (15–17, 20–22, 26, 27) of the 14 studies used 3-TMRI, four (18, 23, 25, 28) used 1.5-T or 3-TMR, and two (19, 24) used 1.5T. Among these, 14 included studies, 12 (16–23, 25–28) used radiomics combined with ML, while two (15, 24) used DL assessment. The most commonly used ML classifiers were SVM and RFC. In addition, five (15, 16, 23, 24, 26) of the 14 studies employed external validation, three (17, 25, 28) had no validation set, and the remaining six studies (18–21, 25, 27) used internal validation. In terms of glioma grade, three studies (15, 16, 18) were low-grade gliomas, and three studies (19, 20, 25) were high-grade gliomas, whereas the remaining studies included both low- and high-grade gliomas. For the ML analysis, eleven (15–23, 25, 26, 28) studies included only radiomics information, and three (22, 24, 27) used radiomics and clinical information.

Quality assessment

The risk of bias and applicability assessment of the included studies, performed using the QUADAS-2 tool, are shown in Figure 2. In terms of the patient selection, two (17, 28) studies were deemed to have a low risk of bias, six (15, 18, 23, 25–27) exhibited a high risk of bias owing to unclear information regarding the time range and consecution of patients, and six (16, 19–22, 24) were considered to have an unclear risk because of uncertainties in the consecution of patients. Regarding the index test, 13 studies had an unclear risk of bias owing to ambiguity regarding the use of a threshold. All the studies indicated a low risk of bias in the reference standards. Regarding flow and timing, 13 studies had an unclear risk of bias because there was no mention of the time interval between imaging and molecular testing.

The mean METRICS score of the included studies was 60.3% (range, 50%-75%), the quality of six (15–17, 22, 25, 28) studies was moderate (40≤score<60%),and eight studies (17–21, 23, 24, 26) were good (60≤score<80%). The highest score of 75% was obtained

TABLE 1 Basic characteristics and details of the 14 included studies (1).

Study	Country	Study design	No. of patients	Mean age	MRI field intensity	Vendor Scanner	Sequences
Li 2017 (15)	China	Retrospective	151	40.7 ± 10.8	3T	Siemens Trio	T1CE,T2 flair
Yu 2016 (16)	China	Retrospective	110	40.3 ± 11.3 (years)	3T	Siemens Trio	T2 flair
Bisdas 2018 (17)	UK	Prospective	37	63.2 ± 7.6	3T	Siemens Skyra	T1,T1CE T2 flair,DKI
Zhang 2018 (18)	China	Retrospective	103	43.5 ± 12.9	1.5T(37),3T(66)	NA	T1,T2, T1CE T2 flair
Deniz Alis 2019 (19)	Turkey	Retrospective	142	40.87 ± 12.25	1.5T	Siemens Avanto	T1CE,T2 flair,DWI
Niu 2020 (20)	China	Retrospective	182	44 ± 11	3T	GE SIGNA	T1CE
Cao 2020 (21)	China	Retrospective	102	44.6 ± 14.9	3T	GE SIGNA	T1,T2, T1CE T2 flair,DWI
Huang 2021 (22)	China	Retrospective	59	46 ± 15	3T	Siemens MAGNETOM	T1,T2, T1CE T2 flair
Manikis 2021 (23)	Greece	Retrospective	160	58.4 ± 15.9	1.5T,3T	GE, Siemens,Philip	T1,T2, T1CE T2 flair, DCE-MR
Hrapşa 2022 (24)	Romania	Retrospective	21	48.6 ± 15.6	1.5T	GE, Siemens, TOSHIBA	T2, T1CE T2 flair
Kandalgaonkar 2022 (25)	United States	Retrospective	100	52	1.5T,3T	GE,Philip	T2, T1CE
Wang 2022 (26)	China	Retrospective	140	40	3T	Siemens	T1,T2,T1CE,T2 FLAIR, DWI
Wang, J 2022 (27)	China	Retrospective	100	48 ± 13	3T	GE	T1,T2, T1CE T2 flair, DWI,DCE-MR
CaroleH 2020 (28)	UK	Retrospective	333	NA	1.5T,3T	GE, Siemens	DSC-MR

NA, not available.

TABLE 2 Basic characteristics and details of the 14 included studies (2).

Study	WHO Grade	reference standard	Machine learning classifier	Validation	Segmentation	Region/volume of interest
Li 2017 (15)	grade 2	Sanger sequencing	CNN	External validation	Automatic	VOI
Yu 2016 (16)	grade 2	Sanger sequencing	SVM and AdaBoost	External validation	Automatic	VOI
Bisdas 2018 (17)	grade 2,3	Sanger sequencing	SVM	No validation	Manual	VOI
Zhang 2018 (18)	grade 2,3	NA	SVM	Internal validation	Manual	VOI
Deniz Alis 2019 (19)	grade 3,4	Histopathological	RFC	Internal validation	Manual	VOI
Niu 2020 (20)	grade 3,4	Immunohistochemistry	biclassification mode	Internal validation	Manual	ROI
Cao 2020 (21)	grade 2,3	Histopathological	RFC	Internal validation	Manual	VOI
Huang 2021 (22)	grade 2,3	DNAsingle-step assay	Multivariate logistic regression	No validation	Manual	VOI
Manikis 2021 (23)	grade 2,3,4	histologically	SVM,RF,KNN, LR,AdaBoost	External validation	Automatic	VOI
Hrapşa 2022 (24)	grade 2,3,4	MLPA-Multiplex PCR) or immunohistochemistry	CNN	External validation	Automatic	VOI

(Continued)

TABLE 2 Continued

Study	WHO Grade	reference standard	Machine learning classifier	Validation	Segmentation	Region/volume of interest
Kandalgaonkar 2022 (25)	grade 4	Immunohistochemistry, Sanger sequencing	SVM,10-fold cross-validation	No validation	Manual	ROI
Wang 2022 (26)	grade 2,3,4	Histologically	ANN	External validation	Segmentation and manual	ROI
Wang, J 2022 (27)	NA	IDH1 R132H mutation-specific antibody	Liner SVM	Internal validation	Automatic	ROI
CaroleH 2020 (28)	grade 2,3,4	Histopathology	Random-forest algorithm	No validation	Manual	VOI

NA, not available.

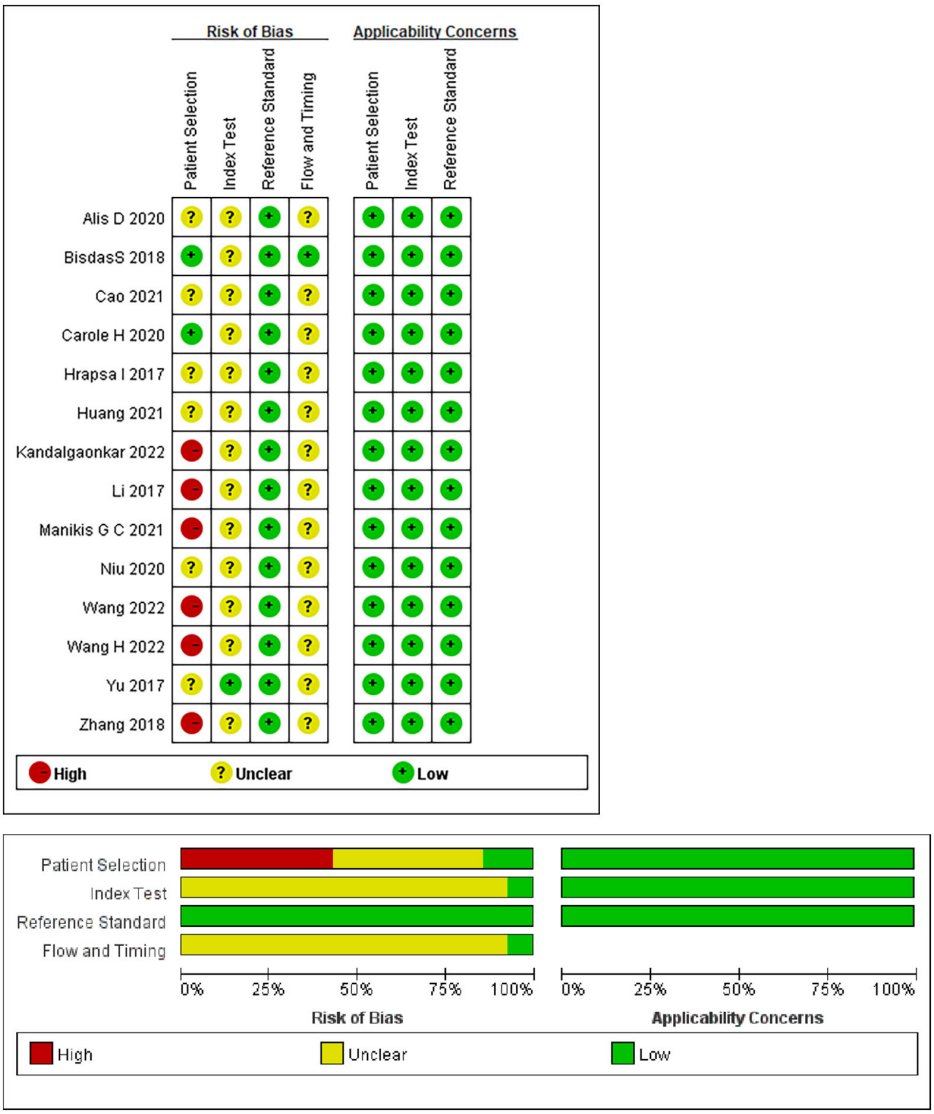


FIGURE 2
Summary of the risk of bias and applicability assessments: the authors' judgements for each domain of each included study were reviewed. The proportion of included studies that indicated low, unclear, or high risk and applicability concerns are shown in green, yellow, and red, respectively.

in one study (26) and the lowest score of 50% was observed in two studies (22, 25), primarily attributed to a lack of a validation cohort. The item of "Model availability" was assigned zero points as none of the included studies addressed it. Only one study (24) publicly shared the code. A detailed description of the METRICS scores is provided in Table 3.

Data analysis

Summaries of the ML models predicting IDH mutations in patients with glioma were analyzed using the random-effects method because of significant statistical heterogeneity (I^2 = 92%). For all 14 studies, the pooled sensitivity, specificity, PLR, NLR, and DOR were 0.83 (0.71,0.90), 0.84 (0.74, 0.90), 5.0 (3.2, 7.8), 0.21 (0.12, 0.35), and 25 (12,50), respectively. The overall pooled AUC was 0.90 (0.87, 0.92), indicating a high diagnostic performance. Forest plots for sensitivity and specificity are illustrated in Figure 3, and the SROC curve is presented in Figure 4.

Cochran's Q test showed significant heterogeneity (Q=25.320, p=0.00) across the studies, with a Higgins's I^2 statistic of 79% for sensitivity and 74.1% for specificity. The Spearman correlation coefficient between the sensitivity and false-positive rate was 0.525 (p=0.054), which indicated no threshold effect among the included studies.

Subgroup analysis

Subgroup analysis was performed by comparing studies with different variables. Supplementary Table 4 shows the results of the subgroup analysis. Studies using DL had a higher specificity and a lower sensitivity (0.91 [0.76, 0.98], 0.77 [0.55, 0.92]) than those using ML (0.78 [0.72, 0.83], 0.86 [0.82, 0.90]). Compared with the studies that only used radiomics features, studies combining the use of radiomics and clinical information showed higher sensitivity and lower specificity (0.89 [0.72, 0.98] vs 0.73 [0.68, 0.78], 0.79 [0.66, 0.88] vs. 0.83 [0.79, 0.86]). In addition, the sensitivity of diagnosing LGG was higher (0.93 [0.85, 0.98]) than that of diagnosing HGG (0.71 [0.58, 0.83]), but the specificity of diagnosing LGG was lower than that of diagnosing HGG (0.71 [0.48, 0.89] vs. 0.91 [0.85, 0.95]). Studies that performed external validation showed lower sensitivity and specificity than those that used internal validation (0.78 [0.70, 0.85] vs. 0.81 [0.74, 0.87]; 0.83 [0.72,0.91] vs. 0.87 [0.82,0.91], respectively).

Discussion

This systematic review and meta-analysis evaluated the diagnostic performance of radiomics in predicting IDH mutations. The pooled sensitivity, specificity, and AUC were 83% (95% CI, 0.71–0.90), 84% (95% CI, 0.74–0.90), and 0.90 (95% CI, 0.87–0.92), respectively. This indicates that radiomics combined with ML and DL could be an effective and accurate diagnostic tool for predicting IDH mutations in gliomas.

TABLE 3 METRICS of the included studies.

Categories	No.	Wang, J.2022 (27)	Wang, H.2022 (26)	Kandalgaonkar, P.2022 (25)	Yu, J.2017 (16)	Hrapsa, I 2022 (24)	Niu L 2020 (20)	Cao, M. 2021 (21)	Bisdas, S.2018 (17)	Li, Z 2017 (15)	Zhang Xi, 2018 (18)	Alis, D.2020 (19)	Huang, W. Y.2021 (22)	Manikis, G. C.2021 (23)	Carole H. Sudre.2020 (28)
Study Design	#1	0.0368	0.0368	0.0368	0.0368	0.0368	0.0368	0.0368	0	0.0368	0.0368	0.0368	0.0368	0.0368	0.0368
	#2	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735	0.0735
	#3	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919	0.0919
Imaging Data	#4	0	0.0438	0	0	0	0	0	0	0	0	0	0	0.0438	0.0438
	#5	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292
	#6	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438	0.0438
	#7	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292	0.0292
Segmentation	#8	0.0337	0.0337	0.0337	0.0337	0	0.0337	0.0337	0.0337	0.0337	0.0337	0.0337	0.0337	0.0337	0.0337
	#9	0	0	0	0.0225	0.0225	0	0	0	0.0225	0	0	0	0.0225	0
	#10	0	0.0112	0	0.0112	0.0112	0	0	0	0.0112	0	0.0112	0	0	0.0112

(*Continued*)

TABLE 3 Continued

Categories	No.	Wang, J.2022 (27)	Wang, H.2022 (26)	Kandalgaonkar, P.2022 (25)	Yu, J.2017 (16)	Hrapsa, I 2022 (24)	Niu L 2020 (20)	Cao, M. 2021 (21)	Bisdas, S.2018 (17)	Li, Z 2017 (15)	Zhang Xi, 2018 (18)	Alis, D.2020 (19)	Huang, W. Y.2021 (22)	Manikis, G. C.2021 (23)	Carole H. Sudre.2020 (28)
Image Processing and Feature Extraction	#11	0	0.0622	0	0	0.0622	0	0	0.0622	0	0	0.0622	0	0.0622	0.0622
	#12	0.0311	0.0311	0.0311	0	0.0311	0.0311	0.0311	0.0311	0	0.0311	0.0311	0.0311	0.0311	0.0311
	#13	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415
Feature Processing	#14	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0
	#15	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	#16	0.03	0	0	0	0	0.03	0.03	0	0	0.03	0	0	0	0
	#17	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Preparation for Modeling	#18	0.0599	0.0599	0	0	0	0.0599	0.0599	0	0	0.0599	0.0599	0	0	0
	#19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Metrics and Comparison	#20	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352
	#21	0.0234	0.0234	0	0	0.0234	0.0234	0	0	0	0.0234	0.0234	0.0234	0	0
	#22	0.0176	0	0	0	0	0	0	0	0	0	0	0	0	0
	#23	0	0	0.0117	0.0117	0	0.0117	0	0	0	0	0	0.0117	0	0
	#24	0	0	0	0	0	0	0.0293	0	0	0	0	0	0	0
	#25	0.0176	0	0	0	0	0	0	0	0	0	0	0	0	0
Testing	#26	0.0375	0	0	0	0	0.0375	0.0375	0.0375	0	0.0375	0.0375	0	0	0
	#27	0	0.0749	0	0.0749	0.0749	0	0	0	0.0749	0	0	0	0.0749	0
Open Science	#28	0	0.0075	0.0075	0	0	0	0	0	0	0	0	0	0	0
	#29	0	0	0	0	0.0075	0	0	0	0	0	0	0	0	0
	#30	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	total	65%	75%	50%	56%	63%	63%	62%	53%	54%	62%	66%	50%	67%	58%

#1 Adherence to radiomics and/or machine learning-specific checklists or guidelines; #2 Eligibility criteria that describe a representative study population ; #3 High-quality reference standard with a clear definition ; #4 Multi-center ; #5 Clinical translatability of the imaging data source for radiomics analysis ; #6 Imaging protocol with acquisition parameters ; #7 The interval between imaging used and reference standard ; #8 Transparent description of segmentation methodology ; #9 Formal evaluation of fully automated segmentation[2]; #10 Test set segmentation masks produced by a single reader or automated tool ; #11 Appropriate use of image preprocessing techniques with transparent description ; #12 Use of standardized feature extraction software[3]; #13 Transparent reporting of feature extraction parameters, otherwise providing a default configuration statement ; #14 Removal of non-robust features[4]; #15 Removal of redundant features[4]; #16 Appropriateness of dimensionality compared to data size[4]; #17 Robustness assessment of end-to-end deep learning pipelines[5]; #18 Proper data partitioning process ; #19 Handling of confounding factors ; #20 Use of appropriate performance evaluation metrics for task ; #21 Consideration of uncertainty ; #22 Calibration assessment ; #23 Use of uni-parametric imaging or proof of its inferiority ; #24 Comparison with a non-radiomic approach or proof of added clinical value ; #25 Comparison with simple or classical statistical models ; #26 Internal testing ; #27 External testing ; #28 Data availability ; #29 Code availability ; #30 Model availability.

10.3389/fonc.2024.1409760

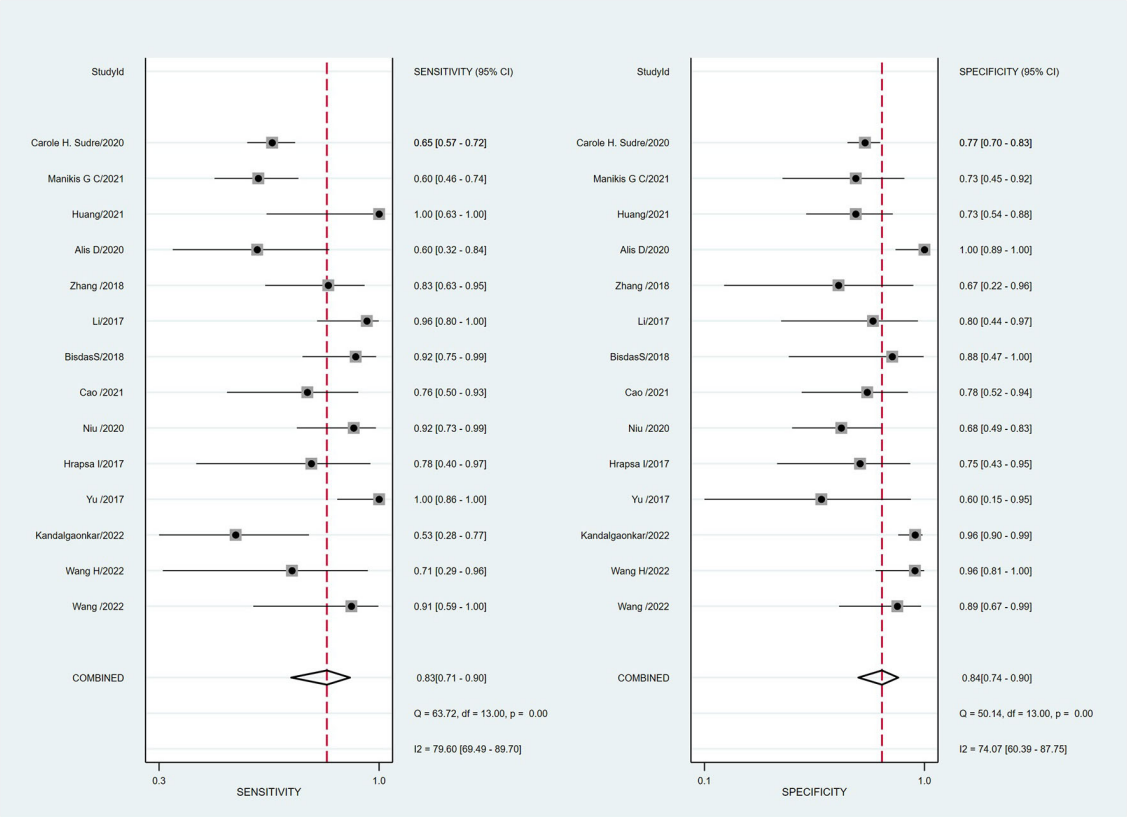


FIGURE 3
Coupled forest plots of the pooled sensitivity and specificity for the diagnostic performance of machine learning-based radiomics for the prediction of IDH mutation glioma.

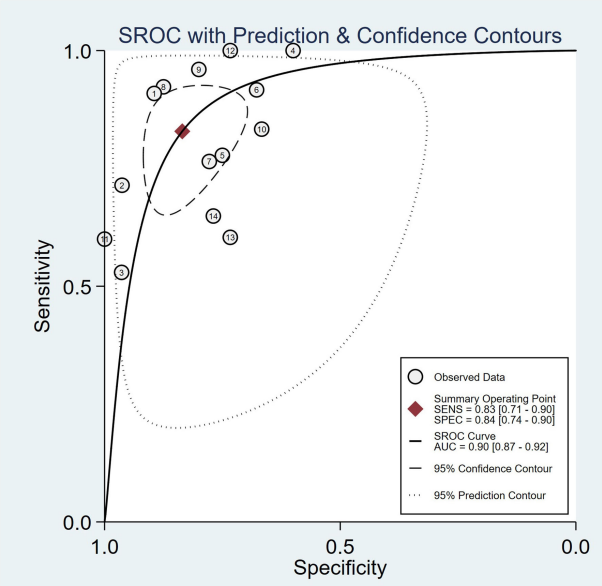


FIGURE 4
Hierarchical summary receiver operating characteristic (SROC) curve of the diagnostic performance of machine learning-based radiomics for the prediction of IDH mutation glioma.

Obviously, heterogeneity was noted in the specificity (I^2=79.6%) and sensitivity (I^2=74.1%), Thus, we performed subgroup analysis to explore the source of the heterogeneity which included the modeling methods (ML vs. DL), glioma grade, whether the model incorporated clinical features, and validation methods (external and internal validation).The results of the present meta-analysis showed that studies using ML had a better diagnostic performance than those using DL. This could be attributed to the small sample sizes of the included studies. DL is capable of training multi-layer deep neural networks, which show significant potential in handling very large datasets with thousands or even millions of instances; however, in scenarios where the size of the dataset is small, DL tends to exhibit lower performance compared to ML. Similar findings have been previously reported for ML in other studies (29, 30). However, only two studies included in our study used DL; thus, future work should incorporate a greater number of studies with sufficient datasets to explore its true diagnostic capabilities. A previous study (31) demonstrated that the combined model of magnetic resonance (MR) and clinical features with ML exhibits better diagnostic performance than that using only MR features. Clinical features such as age, sex, and exposure to ionizing radiation were closely related to the pathological process of glioma (32, 33). For example, age is a risk factor for the development of high-grade glioma; young patients are more likely to suffer from IDH1-mutant glioma, and their postoperative survival and clinical prognosis may

be more optimistic (20). Our findings are consistent with the previous study; therefore, we recommend the combined use of MR and clinical features with ML in future radiomics studies to verify their true diagnostic capabilities in predicting IDH mutation status in gliomas. The diagnostic performance in predicting the IDH mutation of HGG was better than that of LGG in the present study, which is consistent with that of a previous meta-analysis (31); however, it is essential to note that more studies are required to validate this conclusion, given the limited number of included studies. Additionally, we found that studies using external verification models had a diagnostic performance similar to that of studies using internal verification models, demonstrating the stability of the model. Internal validation tends to overestimate the diagnostic value owing to the model's lack of generalizability (34); thus, external validation prediction models are required to reliably estimate the diagnostic capabilities of other datasets.

METRICS is a new quality assessment tool which includes 30 items within 9 categories to evaluate the key steps in the radiomics research workflow. It was developed by a large group of international experts in the field recently and is easy to use, specifically aimed at improving the methodological quality of radiomics research. The METRICS score of the included studies ranged from 50% to 75% and the mean score was 60.3%. The quality of 6 studies was moderate (40≤score<60%) and 8 studies were good (60≤score<80%). For the items with the highest weights, such as high-quality reference standards with a clear definition and eligibility criteria that describe a representative study population, all the included studies received a full score. Only one study (24) publicly shared the code and two studies (25, 26) publicly shared the data, however, these two items which belong to the "open science" category had the lowest weight. Although METRICS is a valuable tool for evaluating the quality of radiomics research, it is not without limitations. Further revision of METRICS may enhance its comprehensiveness in assessing the quality of radiomics studies.

QUADAS-2 quality assessment revealed other issues in the included studies that can be avoided in future investigations. For example, the majority of the studies did not mention the consecution of patients and the time interval between imaging and molecular testing, which led to a high or unclear bias risk. In 13 studies, it remained unclear whether thresholds were pre-specified or not, potentially resulting in an overestimation of the diagnostic performance of the models.

This study had several limitations. First, most of the included studies had a retrospective design, and only one had a prospective design; thus, selection bias was inevitable. Therefore, prospective multicenter studies with larger scales are required to validate our findings. Second, the sample size of the included studies was not large enough for training and validation, which limited the statistical power of the study and may affect the generalizability of the results. Third, significant heterogeneity was observed, which is observed in other meta-analyses of diagnostic accuracy using ML based on radiomics. Finally, the mean METRICS score of the included studies was 60.3%, indicating moderate overall quality. Therefore, further high-quality radiomics studies are required to verify our results. Despite these limitations, our review provided new insights into the accuracy of ML-based radiomics models for predicting IDH mutations in gliomas.

In conclusion, ML-based radiomics demonstrated excellent diagnostic performance for predicting IDH mutations in gliomas. Nevertheless, owing to the limitations in the quality and quantity of the included studies, caution should be exercised when applying the results, and more standardized and prospective studies are warranted to improve the application and reliability of radiomics.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

XC: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft. JL: Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing. SW: Investigation, Methodology, Writing – original draft. JZ: Data curation, Methodology, Software, Writing – original draft. LG: Methodology, Software, Validation, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Youth Science and Technology Fund Program of Science and Technology Department of Gansu Province (21JR11RA071).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1409760/full#supplementary-material

References

1. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1

2. Kurokawa R, Kurokawa M, Baba A, Ota Y, Pinarbasi E, Camelo-Piragua S, et al. Major changes in 2021 world health organization classification of central nervous system tumors. *Radiographics*. (2022) 42:1474–93. doi: 10.1148/rg.210236

3. Cancer Genome Atlas Research Network, Brat DJ, Verhaak RG, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med*. (2015) 372:2481–98. doi: 10.1056/NEJMoa1402121

4. Kawaguchi T, Sonoda Y, Shibahara I, Saito I, Kanamori I, Kumabe I, et al. Impact of gross total resection in patients with WHO grade III glioma harboring the IDH 1/2 mutation without the 1p/19q co-deletion. *J Neurooncol*. (2016) 129:505–14. doi: 10.1007/s11060-016-2201-2

5. Pellegatta S, Valletta L, Corbetta C, Patanè M, Zucca I, Sirtori FR, et al. Effective immunotargeting of the IDH1 mutation R132H in a murine model of intracranial glioma. *Acta Neuropathol Commun*. (2015) 3:4. doi: 10.1186/s40478-014-0180-0

6. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

7. Buda M, AlBadawy EA, Saha A, et al. Deep radiogenomics of lower-grade gliomas: convolutional neural networks predict tumor genomic subtypes using MR images. *Radiol Artif Intell*. (2020) 2:e180050. doi: 10.1148/ryai.2019180050

8. Matsui Y, Maruyama T, Nitta M, Saito T, Tsuzuki S, Tamura M, et al. Prediction of lower-grade glioma molecular subtypes using deep learning. *J Neurooncol*. (2020) 146:321–7. doi: 10.1007/s11060-019-03376-9

9. Bhandari AP, Liong R, Koppen J, Murthy SV, Lasocki A. Noninvasive determination of IDH and 1p19q status of lower-grade gliomas using MRI radiomics: A systematic review. *AJNR Am J Neuroradiol*. (2021) 42:94–101. doi: 10.3174/ajnr.A6875

10. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The prisma-dta statement. *Jama*. (2018) 319:388–96. doi: 10.1001/jama.2017.19163

11. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. (2011) 155:529–36. doi: 10.3760/cma.j.issn.0254-6450.2018.04.028

12. METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. *Insights into Imaging*. (2024) 15. doi: 10.1186/s13244-023-01572-w

13. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions, Version 6.2 (2021). Available online at: https://training.cochrane.org/handbook/current/chapter-10#section-10-10-2 (Accessed 17 October 2021).

14. Deville WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol*. (2002) 2:9. doi: 10.1186/1471-2288-2-9

15. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep*. (2017) 7:5467. doi: 10.1038/s41598-017-05848-2

16. Yu J, Shi Z, Lian Y, Li Z, Liu T, Gao Y, et al. Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. *Eur Radiol*. (2017) 27:3509–22. doi: 10.1007/s00330-016-4653-3

17. Bisdas S, Shen H, Thust S, Katsaros V, Stranjalis G, Boskos C, et al. Texture analysis- and support vector machine-assisted diffusional kurtosis imaging may allow in *vivo* gliomas grading and IDH-mutation status prediction: a preliminary study. *Sci Rep*. (2018) 8:6108. doi: 10.1038/s41598-018-24438-4

18. Zhang X, Tian Q, Wang L, Liu Y, Li B, Liang Z, et al. Radiomics strategy for molecular subtype stratification of lower-grade glioma: detecting IDH and TP53 mutations based on multimodal MRI. *J Magn Reson Imag*. (2018) 48:916–26. doi: 10.1002/jmri.25960

19. Alis D, Bagcilar O, Senli YD, Yergin M, Isler C, Kocer N, et al. Machine learning-based quantitative texture analysis of conventional MRI combined with ADC maps for assessment of IDH1 mutation in high-grade gliomas. *Jpn J Radiol*. (2020) 38:135–43. doi: 10.1007/s11604-019-00902-7

20. Niu L, Feng WH, Duan CF, Liu YC, Liu JH, Liu XJ, et al. The value of enhanced MR radiomics in estimating the IDH1 genotype in high-grade gliomas. *BioMed Res Int*. (2020) 2020:1–6. doi: 10.1155/2020/4630218

21. Cao M, Suo S, Zhang X, Wang X, Xu J, Yang W, et al. Qualitative and quantitative MRI analysis in IDH1 genotype prediction of lower-grade gliomas: A machine learning approach. *BioMed Res Int*. (2021) 2021:1235314. doi: 10.1155/2021/1235314

22. Huang WY, Wen LH, Wu G, Hu MZ, Zhang CC, Chen F, et al. Comparison of radiomics analyses based on different magnetic resonance imaging sequences in grading and molecular genomic typing of glioma. *J Comput Assist Tomogr*. (2021) 45:110–20. doi: 10.1097/RCT.0000000000001114

23. Manikis GC, Ioannidis GS, Siakallis L, Nikiforaki K, Iv M, Vozlic D, et al. Multicenter DSC-MRI-based radiomics predict IDH mutation in gliomas. *Cancers (Basel)*. (2021) 13:3965. doi: 10.3390/cancers13163965

24. Hrapşa I, Florian IA, Şuşman S, Farcaş M, Beni L, Florian IS. External validation of a convolutional neural network for IDH mutation prediction. *Medicina (Kaunas)*. (2022) 58:526. doi: 10.3390/medicina58040526

25. Kandalgaonkar P, Sahu A, Saju AC, Joshi A, Mahajan A, Thakur M, et al. Predicting IDH subtype of grade 4 astrocytoma and glioblastoma from tumor radiomic patterns extracted from multiparametric magnetic resonance images using a machine learning approach. *Front Oncol*. (2022) 12:879376. doi: 10.3389/fonc.2022.879376

26. Wang H, Zhang S, Xing X, Yue Q, Feng W, Chen S, et al. Radiomic study on preoperative multi-modal magnetic resonance images identifies IDH-mutant TERT promoter-mutant gliomas. *Cancer Med*. (2023) 12:2524–37. doi: 10.1002/cam4.5097

27. Wang J, Hu Y, Zhou X, Bao S, Chen Y, Ge M, et al. A radiomics model based on DCE-MRI and DWI may improve the prediction of estimating IDH1 mutation and angiogenesis in gliomas. *Eur J Radiol*. (2022) 147:110141. doi: 10.1016/j.ejrad.2021.110141

28. Sudre CH, Panovska-Griffiths J, Sanverdi E, Brandner S, Katsaros VK, Stranjalis G, et al. Machine learning assisted DSC-MRI radiomics as a tool for glioma classification by grade and mutation status. *BMC Med Inf Decision Making*. (2020) 20. doi: 10.1186/s12911-020-01163-5

29. Cuocolo R, Cipullo MB, Stanzione A, Romeo V, Green R, Cantoni V, et al. Machine learning for the identification of clinically significant prostate cancer on MRI: a meta-analysis. *Eur Radiol*. (2020) 30(12):6877–87. doi: 10.1007/s00330-020-07027-w

30. van Kempen EJ, Post M, Mannil M, Witkam RL, Laan MT, Patel A, et al. Performance of machine learning algorithms for glioma segmentation of brain MRI: a systematic literature review and meta-analysis. *Eur Radiol*. (2021) 31(12):9631–58. doi: 10.1007/s00330-021-08035-0

31. Zhao J, Huang Y, Song Y, Xie D, Hu M, Qiu H, et al. Diagnostic accuracy and potential covariates for machine learning to identify IDH mutations in glioma patients: evidence from a meta-analysis. *Eur Radiol*. (2020) 30(8):4664–74. doi: 10.1007/s00330-020-06717-9

32. Shibahara I, Sonoda Y, Shoji T, Kanamori M, Saito R, Inoue T, et al. Malignant clinical features of anaplastic gliomas without IDH mutation. *Neuro Oncol* (2015) 17 (1):136–44. doi: 10.1093/neuonc/nou112

33. Halligan S, Menu Y, Mallett S. Why did European adiology reject my radiomic biomarker paper? How to correctly evaluate imaging biomarkers in a clinical setting. *Eur Radiol*. (2021) 31:9361–8. doi: 10.1007/s00330-021-07971-1

34. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J Radiol*. (2019) 20:405–10. doi: 10.3348/kjr.2019.0025

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 04 September 2024
DOI 10.3389/fonc.2024.1435204

OPEN ACCESS

EDITED BY
Domenico Aquino,
IRCCS Carlo Besta Neurological Institute Foundation, Italy

REVIEWED BY
Alfonso Mastropietro,
National Research Council (CNR), Italy
Andrea Bianconi,
University Hospital of the City of Health and Science of Turin, Italy

*CORRESPONDENCE
Yang Gao
1390903990@qq.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 19 May 2024
ACCEPTED 19 August 2024
PUBLISHED 04 September 2024

CITATION
Zhang C, Wang P, He J, Wu Q, Xie S, Li B, Hao X, Wang S, Zhang H, Hao Z, Gao W, Liu Y, Guo J, Hu M and Gao Y (2024) Super-resolution reconstruction improves multishell diffusion: using radiomics to predict adult-type diffuse glioma IDH and grade. *Front. Oncol.* 14:1435204. doi: 10.3389/fonc.2024.1435204

Super-resolution reconstruction improves multishell diffusion: using radiomics to predict adult-type diffuse glioma IDH and grade

Chi Zhang[1†], Peng Wang[1†], Jinlong He[1], Qiong Wu[1], Shenghui Xie[1], Bo Li[1], Xiangcheng Hao[1], Shaoyu Wang[2], Huapeng Zhang[2], Zhiyue Hao[1], Weilin Gao[1], Yanhao Liu[1], Jiahui Guo[1], Mingxue Hu[1] and Yang Gao[1*]

[1]Department of Radiology, Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China, [2]MR Research Collaboration, Siemens Healthineers, Shanghai, China

Objectives: Multishell diffusion scanning is limited by low spatial resolution. We sought to improve the resolution of multishell diffusion images through deep learning-based super-resolution reconstruction (SR) and subsequently develop and validate a prediction model for adult-type diffuse glioma, isocitrate dehydrogenase status and grade 2/3 tumors.

Materials and methods: A simple diffusion model (DTI) and three advanced diffusion models (DKI, MAP, and NODDI) were constructed based on multishell diffusion scanning. Migration was performed with a generative adversarial network based on deep residual channel attention networks, after which images with 2x and 4x resolution improvements were generated. Radiomic features were used as inputs, and diagnostic models were subsequently constructed via multiple pipelines.

Results: This prospective study included 90 instances (median age, 54.5 years; 39 men) diagnosed with adult-type diffuse glioma. Images with both 2x- and 4x-improved resolution were visually superior to the original images, and the 2x-improved images allowed better predictions than did the 4x-improved images (P<.001). A comparison of the areas under the curve among the multiple pipeline-constructed models revealed that the advanced diffusion models did not have greater diagnostic performance than the simple diffusion model (P>.05). The NODDI model constructed with 2x-improved images had the best performance in predicting isocitrate dehydrogenase status (AUC_validation=0.877; Brier score=0.132). The MAP model constructed with the original images performed best in classifying grade 2 and grade 3 tumors (AUC_validation=0.806; Brier score=0.168).

Conclusion: SR improves the resolution of multishell diffusion images and has different advantages in achieving different goals and creating different target diffusion models.

KEYWORDS

glioma, diffusion magnetic resonance imaging, deep learning, radiomics, diagnosis

1 Introduction

Following the identification and clarification of brain tumor pathogenesis and diagnostic and therapeutic processes, the World Health Organization released criteria for classifying tumors of the central nervous system in 2021 (1), which has further increased the clinical value of the convergence of tumor grade and molecular genotype. According to this classification system, most individuals diagnosed with adult-type diffuse glioma (approximately 80%) have poorer outcomes than patients with other types of glioma (2). The formulation of clinical decisions and prognosis prediction are dependent on the mutation status of isocitrate dehydrogenase (IDH), while an increase in tumor grade implies, to a certain extent that patients will require frequent radiotherapy and chemotherapy treatments, accompanied by a greater possibility of recurrence. MRI, the gold standard for preoperatively diagnosing glioma (3), may have potential as part of the development of an accurate method of predicting tumor pathology through imaging alone; such an approach could potentially optimize surgical decisions and improve clinical treatment strategies.

Previous studies have shown that, compared with conventional imaging, diffusion imaging can be used to capture additional brain tissue microstructural alterations and pathological changes caused by nuclear heterogeneity (4–6). Theoretically, advanced diffusion models, such as those based on neurite orientation dispersion and density imaging (NODDI) (7), itself based on the three-compartment model, and the mean apparent propagator (MAP) (8), which does not rely on *a priori* assumptions, should allow clinicians to better characterize the complexity and nonuniformity of the tissue microenvironment (9) and further improve the diffusion description of brain tissue over simple Gaussian diffusion models (such as diffusor tensor imaging [DTI]-based models). However, the findings of some clinical studies do not support these theoretical advantages (10, 11), reporting that these techniques may be limited by the spatial resolution of the acquired diffusion images (12). A high spatial resolution mitigates the partial volume effect, the phenomenon by which signal mixing occurs at the interfaces between different tissues. This enhancement facilitates more precise identification of boundaries between lesion areas and normal tissues, thereby improving diagnostic accuracy and reliability.

Super-resolution reconstruction (SR) is a technology through which the physical limitations of imaging systems can be overcome by generating high-resolution maps from one or more corresponding low-resolution images (13). SR methods are currently used in a variety of computer vision applications ranging from security and surveillance imaging (14) to object recognition (15). SR systems have also shown good applicability in the medical field. For example, they have been employed in the development of high angular resolution diffusion imaging brain templates from low angular resolution diffusion data from a single subject (16). Unlike Varentsova et al., Iglesias et al. (17) used SynthSR (a type of convolutional neural network) to synthesize higher spatial resolution images from portable low-field-strength MR images; notably, the high morphological correlation of different regions of interest in the brain demonstrated that SR was able to suitably improve the enhancement in the original image. In another study, a generative adversarial network (GAN)-based network architecture was used for quantitative analysis after migration (18), and the results suggested that the use of SR improved the diagnostic efficacy of radiomic models (which provide biological transformations of multiple feature matrices with more varied attempts for identifying imaging markers). Some scholars have also overcome the issue of inaccurate automatic glioma segmentation due to missing sequences or poor image quality through the combined application of U-Net and transfer learning (19). The use of SR is likely to increase the potential clinical applicability of multishell diffusion images through resolution improvement and facilitate the exploration of imaging markers for adult-type diffuse gliomas (Supplementary Data Sheet 1).

In this study, we attempted to use a GAN-based SR technique to improve the resolution of multishell diffusion images. We used a GAN for the following reasons: 1) Compared with other deep learning models, GANs perform adversarial training between the generator and discriminator to generate high-resolution images with richer details and more realistic textures than the original-resolution images. 2) The combination of content loss and adversarial loss enables GANs to consider both pixel-level accuracy and visual realism in performing super-resolution tasks. 3) The adversarial training mechanism of GANs easily adapts to highly complex image distributions and reduces the risk of overfitting. Two tasks (i.e., predicting the IDH status of adult-type diffuse glioma and predicting whether gliomas would be classified as grade 2 or 3) were subsequently performed with the constructed models to determine whether SR could be beneficial to clinical processes and to determine the practical applications of the diffusion models.

2 Materials and methods

We conducted a prospective study in accordance with the Declaration of Helsinki. The study was approved by our ethics committee (KY2023064), and all instances signed an informed consent form before enrollment.

We used the CLEAR checklist in the conduction of this study (20); it is included in this submission as Supplementary Data.

2.1 Instances and clinical data

In this study, we prospectively recruited consecutive instances who visited our institution between June 2018 and June 2023. Data from ninety instances (median age, 54.5 years; range, 21-77 years; 39 [43%] males) with a pathological diagnosis of adult-type diffuse glioma (in accordance with the 2021 WHO Classification of Tumors of the Central Nervous System) were used for image reconstruction and model building. Instances who were treated by methods including radiotherapy, chemotherapy, or concurrent radiotherapy and chemotherapy prior to pathology sampling and who had poor image quality were excluded. Previous studies (10) analyzed cohorts with overlapping data, in contrast to the present study, in which the scan-to-pathology time (median time of 4.5 days in this study) and the inclusion of new instances (n=13) were controlled and SR was employed more frequently.

2.2 Magnetic resonance scanning

All study instances underwent preoperative conventional MRI and diffusion imaging with a 3T scanner (MAGNETOM Skyra; Siemens Healthcare, Erlangen, Germany) equipped with a 32-channel head/neck coil.

The conventional MRI (cMRI) sequences included axial T1-weighted, axial T2-weighted, axial T2-weighted fluid-attenuated inversion-recovery (FLAIR), and 3D contrast-enhanced T1-weighted imaging, the last of which was performed after intravenous administration of 0.1 mmol/kg gadobutrol (Gadovist, Bayer AG, Berlin, Germany). The diffusion imaging sequences included axial diffusion-weighted imaging (DWI) and diffusion spectrum magnetic resonance imaging. The diffusion spectrum imaging scheme included the acquisition of a total of 128 diffusion samples, consisting of 16 b-values (200, 350, 400, 550, 750, 950, 1150, 1500, 1700, 1850, 1900, 2050, 2250, 2450, 2650 and 3000 s/mm2). The in-plane resolution was 2.65 mm. Detailed information on the parameters is provided in Supplementary Table S1.

2.3 Preprocessing

The diffusion parameters were calculated with Neuro-Diffusion Lab (NeuDiLab, Chengdu ZhongYing Medical Technology Co., Ltd., Chengdu, China), software developed in-house with Python based on the open-source tool DIPY (Diffusion Imaging in Python, https://dipy.org). The software is equipped with FSL-based brain extraction, eddy current and head motion correction, and smoothing functions and allowed acquisition of the final quantitative parametric maps for 25 features from the 4 diffusion models (listed below) and B0 maps. Three advanced models (diffusional kurtosis imaging (DKI), NODDI and MAP-MRI) and 1 simple model (DTI)) were constructed. The cMRI data were bias corrected with the N4ITK MRT bias correction module in 3D-Slicer. Coregistration of the cMRI data and diffusion parameter maps was performed with ANTs in 3D-Slicer. Default parameters were used for bias field correction and image registration.

2.4 Image SR and effect assessment

We performed SR on the original images with a pretrained GAN (https://github.com/OnekeyAI-Platform/onekey). The core of the migration model is a deep residual channel attention network (21) (https://github.com/XPixelGroup/BasicSR/blob/master/README_CN.md and https://github.com/yulunzhang/RCAN). Before training the GAN model, the input images underwent preprocessing to remove noise and artifacts, followed by normalization of the intensity values. The depth of the network is increased by the residual, which consists of several residual clusters with long jump connections. Each residual cluster contains several channel residual blocks with short jump connections. Each channel attention residual block consists of a simple residual block and a channel attention mechanism. Moreover, residual information allows rich low-frequency information to be bypassed via multiple jump connections so that the main network focuses on learning high-frequency information. The channel attention mechanism performs adaptive rescaling of channel features by accounting for the interdependence between channels to change the influence weights of different channel features on the reconstructed image. Finally, diffusion parameter maps with 2-fold and 4-fold spatial resolution enhancements are obtained. The flowchart of this study is shown in Figure 1.

The effect of SR was analyzed with visual characterization and image loss quantification. A physician (JL.H.) with 13 years of experience in neuroradiology and blinded to the clinical and pathological details of the instance but otherwise aware of the tumor diagnosis evaluated the SR-enhanced images by eye. Forced-choice pairwise comparisons were used to evaluate all samples, in which the sharpness, contrast, and both noise and artifacts were compared among the original, 2x resolution, and 4x resolution images. The images were quantitatively evaluated in three steps. First, we downsampled the original images 2x and 4x. Gaussian noise was subsequently introduced into the images. Finally, the image resolution was increased with a GAN-based transfer model, after which the loss between the original image and the newly generated images was calculated. We measured both structural similarity (SSIM) and the normalized root-mean-square error (NRMSE) for the whole brain and the tumor level (i.e., the solid tumor and peritumor edema regions).

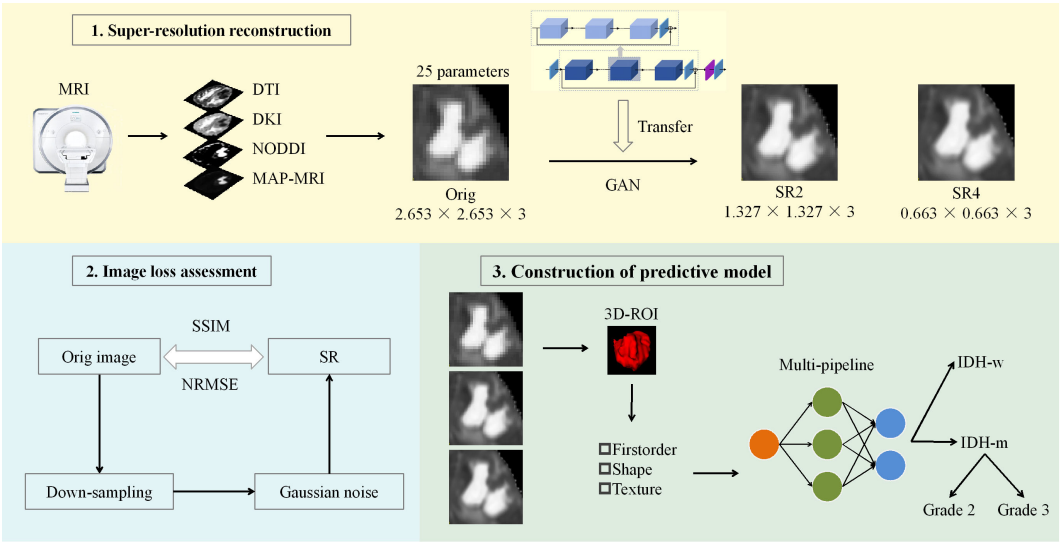


FIGURE 1
Workflow of the study. A GAN-based migration model was used for SR 25 parameter maps of the 4 diffusion models at two magnifications. After the reconstruction effect was evaluated, the actual clinical significance of the SR was judged via machine learning models constructed through multiple pipelines. ROC curve, decision curve and calibration curve analyses were used to evaluate and compare the performance of the models. DTI, diffusion tensor imaging; DKI, diffusion kurtosis imaging; NODDI, neurite orientation dispersion and density imaging; MAP, mean apparent propagation diffusion; GAN, generative adversarial network; SR2, super-resolution reconstruction 2x; SR4, super-resolution reconstruction 4x; SSIM, structural similarity; NRMSE, normalized root-mean-square error; ROI, region of interest; IDH-w, isocitrate dehydrogenase wild-type; IDH-m, isocitrate dehydrogenase mutant.

2.5 Region of interest segmentation and feature extraction

Two physicians (P.W. and C.Z., with 4 and 3 years of neuroimaging experience, respectively) independently delineated regions of interest (ROIs) from the original images via a semiautomatic process in 3D-Slicer. Both radiologists were aware of the tumor diagnosis of the instance but were blinded to the clinical and pathological details. The ROIs were outlined on the B0 map while referring to the cMRI images. The solid tumor and peritumoral edema regions at all levels were selected as the ROIs; these regions are typically depicted radiologically as areas surrounded by T2-FLAIR abnormalities/high signals. For multicentric lesions, all regions were included in the ROIs. The ROIs for the 2x and 4x resolution images were obtained by upsampling the ROIs of the original images with the corresponding multiples, and then all the samples were examined to ensure that the extracted features corresponded to the same region as that in the original image. The Dice coefficient between the two physicians was 0.804; then the ROIs delineated by the more experienced physician was used for further feature extraction after review by another physician (Y.G.) with 28 years of experience in neuroradiology.

FeAture Explorer (FAE v0.5.8, https://github.com/salan668/FAE) (22) was used to extract radiomic parameters from the 3D-ROIs. Feature extraction was performed via PyRadiomics (version 3.0). In total, 107 features were extracted from the original images (Supplementary Table S2), including first-order features [n=18], shape-based features [n=14], gray level co-occurrence matrix features [n=24], gray level dependence matrix features [n=14], gray level run-length matrix features [n=16], gray level size-zone matrix features [n=16], and neighboring gray-tone difference matrix features [n=5]. The features were discretized by fixing the bin count (16 gray levels), and the remaining parameters were assigned the default configurations. Twenty-five sets of diffusion parameters were used for feature extraction; thus, 2675 features were extracted from the images at each resolution, for a total of 8025 features.

2.6 Model construction

Data were included in the training and internal test sets according to the instances' enrollment time (23). Multiple pipeline combinations were then considered during model development, including 1 data balancing method (random upsampling), 3 feature normalization methods (mean, min–max and Z score normalization), 2 data dimensionality reduction methods (principal component analysis and Pearson correlation coefficients (cutoff = 0.85)), 4 feature selection methods (analysis of variance, recursive feature elimination, Kruskal-Wallis, and Relief), and 10 classifier methods (linear [logistic regression, logistic regression via least absolute shrinkage and selection operator, linear discriminant analysis, and support vector machine] and nonlinear [autoencoder, decision tree, random forest, AdaBoost, Gaussian process, and naïve Bayes]) (Scikit-Learn (version 0.24.2)) for a total of 240 basic pipelines. For pipelining, data balancing was used only for the training set, and feature scaling was independently applied to the validation and test sets. Ultimately, the number of features included in the model was restricted according to a rule of

thumb; specifically, the maximum number of features was the number of instances in the training set divided by 10). The selection of the number of model features and hyperparameter tuning were performed via leave-one-out cross-validation. A time-independent internal test set that was not involved in the model selection was then used for model testing. Note that the division of the dataset was fixed, even for different pipelines.

For each imaging resolution for each task, 5 diagnostic models were selected from among the multiple pipelines, including 4 single-modal prediction models based on a single diffusion technique (DTI, DKI, MAP, and NODDI) and 1 fusion prediction model incorporating all the diffusion techniques; in this way, a total of 30 models (2 tasks × 3 image resolutions × 5 models) were selected for assessment.

2.7 Statistical analysis

Model performance was evaluated with receiver operating characteristic curve analysis. The output of the prediction models corresponds to the predicted probability of the task (ranging from 0 to 1). The results were then converted into binary predictions, where the threshold depends on the maximum Youden index. Model performance was compared in the cross-validation set through the integrated discrimination improvement metric; small sample sizes and an overreliance on the choice of p value and cutoff points limited the use of the DeLong test and net reclassification index (24). Calibration curves and Brier scores were used to assess deviations between the model results in the training set and the actual results. Decision curve analysis was used to assess the net clinical benefits of the different models under different threshold probabilities. The sample size calculations are presented in Supplementary Data Sheet 1.

Quantitative data are expressed as the means ± standard deviations. Student's t test was used to compare instance ages, and the $\chi 2$ test, Fisher's exact test or Mann-Whitney U test was used to compare categorical variables between groups. Comparisons between multiple groups were performed with one-way ANOVA, and the Bonferroni correction was used for p value adjustment for multiple comparisons. All the statistical analyses were two-sided, and P<.05 was used to indicate statistical significance. All the statistical analyses were performed with SPSS (version 24.0), R (version 4.3.1), and Python (version 3.9.18).

3 Results

3.1 Instance characteristics

In Task 1, seventy-two instances diagnosed between June 2018 and November 2021 were assigned to the training set; 28 had IDH-mutant glioma, and 44 had IDH-wild type glioma. Eighteen instances diagnosed between December 2021 and June 2023 were assigned to the internal test set, which included 6 with IDH-mutant glioma and 12 with IDH-wild type glioma (Table 1; Supplementary Table S4). Note that owing to sample size limitations, the dataset for Task 2 did not include the internal test set; rather, the data for all Grade 2 (n=12) and 3 (n=21) instances were used as the training set (Supplementary Table S4).

TABLE 1 Instance characteristics.

Variable	Full set
Total number of instances	90
2021 WHO Integrated Diagnosis (CNS WHO grade)	
Astrocytoma, IDH-mutant (2)	7
Astrocytoma, IDH-mutant (3)	7
Astrocytoma, IDH-mutant (4)	1
Oligodendroglioma, IDH-mutant and 1p/19q-codeleted (2)	5
Oligodendroglioma, IDH-mutant and 1p/19q-codeleted (3)	14
Glioblastoma, IDH-wild type (4)	56
Age (years)	
Median	54.5
Range	21-77
Sex	
Male	39
Female	51

Nine morphologic features (necrosis, cysticity, calcification, hemorrhage, tumor enhancement pattern, location, side, solid tumor border clarity, and edema) (10) were extracted from the imaging reports to be used as baseline features for within-set description and comparison. For the IDH predictions, only age (P=.001), necrosis (<.001), tumor location (<.001), and mode of enhancement (<.001) within the training set differed between the groups. For classifying grade 2 and grade 3 tumors, only tumor laterality (P=.032) within the training set differed between the groups. Because only a small number of baseline and morphological characteristics differed between the groups, no additional confounders were considered.

3.2 Image visualization and loss assessment

SR4 had the highest sharpness and contrast, followed by SR2 and the original image. SR4 also had the most amount of noise and highest number of artifacts; however, the patterns of noise and artifacts were not identical among the three sets of images.

The effects of image reconstruction were determined for both positive and negative SSIMs and NRMSEs. At both the whole-brain and the tumor region levels, the image loss in the 4 diffusion models constructed from the 2x-resolution images was less than that of the models constructed from the 4x-resolution images (P<.001) (Figure 2; Supplementary Figure S1). At the whole-brain level, SR2 corresponded to mean SSIM and NRMSE values of 0.827

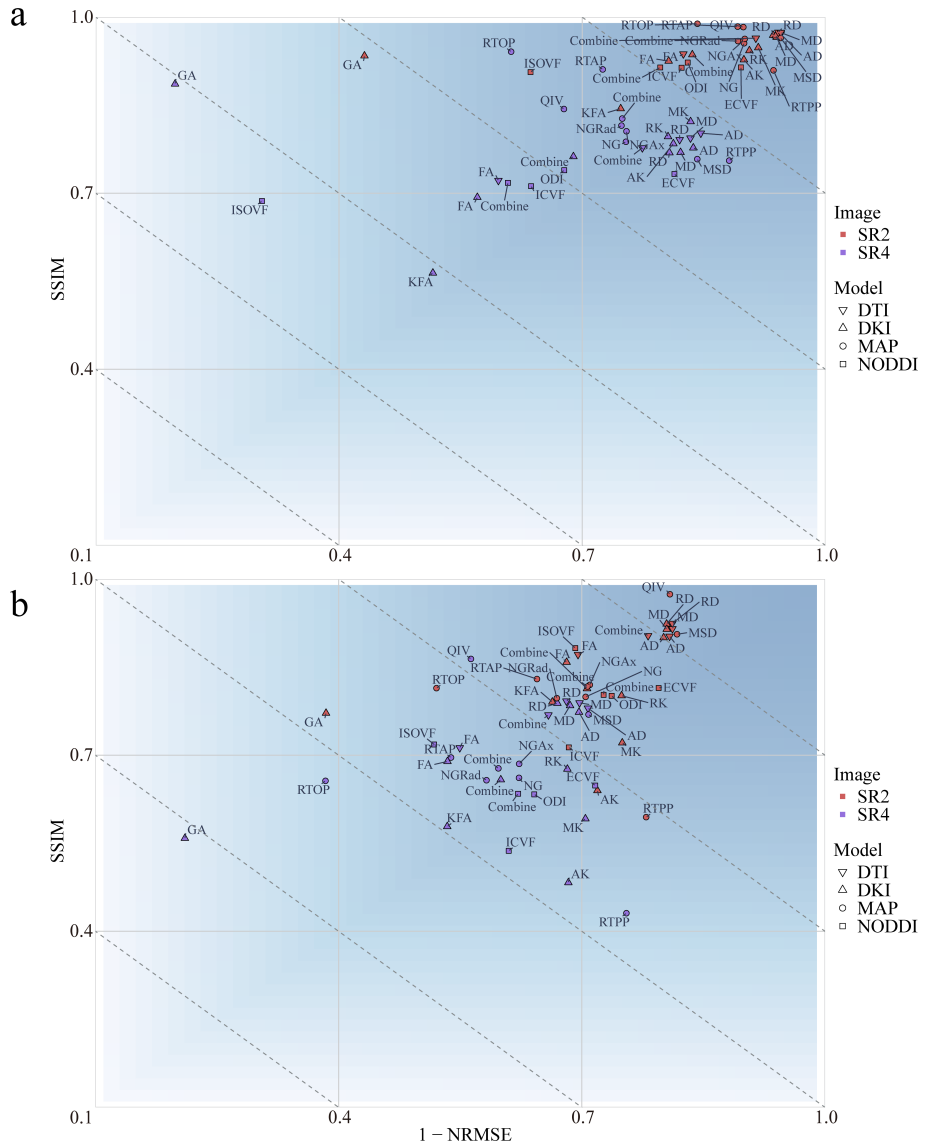


FIGURE 2
Mean loss for the multiple diffusion parameters corresponding to SR at different magnifications at the whole-brain **(A)** and tumor levels **(B)**. SSIM, structural similarity; NRMSE, normalized root-mean-square error; SR2, super-resolution reconstruction 2x; SR4, super-resolution reconstruction 4x; DTI, diffusion tensor imaging; DKI, diffusion kurtosis imaging; MAP, mean apparent propagation diffusion; NODDI, neurite orientation dispersion and density imaging.

and 0.278, respectively, whereas those of SR4 were 0.678 and 0.388, respectively. At the tumor region level, the mean SSIM and NRMSE for SR2 were 0.946 and 0.137, respectively, and those for SR4 were 0.778 and 0.291, respectively. In addition, the 4 diffusion models based on SR2 and the DKI and MAP models based on SR4 had greater image loss at the whole-brain level than at the tumor region (P<.001). The mean SSIM and NRMSE at the whole-brain level were 0.753 and 0.333, respectively, and those at the tumor region level were 0.862 and 0.214, respectively.

At the whole-brain level, the DTI achieved a smaller loss than did the other 3 diffusion models (P<.001). At the tumor region level, the loss with the NODDI was greater than that with the remaining 3 diffusion models (P<.001). At both the whole-brain and tumor-region levels, the fractional anisotropy (FA) loss was greater than that of the other 3 parameters in the DTI model (P<.001). No differences in parameter losses were observed for the other 3 diffusion models (P>.05).

3.3 Features correlated within different resolutions

According to Pearson correlation analysis, the correlation coefficient between the original images and the SR2 images ranged from -0.984 to 0.999 (Supplementary Figure S2), whereas that between the original image and SR4 ranged from -0.578 to 0.585. Radiomic features extracted from the original images were more highly correlated with the radiomic features extracted from the SR2 images than from those extracted from the SR4 images.

3.4 Comparisons of multiple pipelines and model selection

We considered a subgroup and a selected model to have high or low diagnostic efficacy when they exhibited the same trend within all the sets. The results showed that when predicting IDH, the SR4-based DTI prediction model had the lowest diagnostic performance across all sets (Figure 3; Supplementary Figure S3); when classifying Grade 2 and Grade 3 tumor classification, the performance did not seem to differ among the SR image sets.

For each of the 15 predictive models, the areas under the ROC curve (AUCs) in the cross-validation set ranged from 0.593-0.877 and 0.607-0.861 separately for the two tasks. We then further selected the models with the highest diagnostic performance (P>.05) from the 30 models by comparing the integrated discrimination improvement metric in the cross-validation sets and identified three predictive IDH status models and eight models for classifying grade 2 and grade 3 tumors (Table 2; Figure 4). Models whose outputs were closer to the true results were then further selected according to the Brier scores (Supplementary Table S5). Finally, we found that the SR2-based NODDI model best predicted IDH status, while the original image-based MAP model performed best in classifying grade 2 and grade 3 tumors.

The NODDI model of SR2 was constructed via logistic regression and consisted of six features chosen after principal component analysis (Supplementary Figure S4), including principal components 13, 11, 5, 31, 24, and 75, the first and last of which had the highest and lowest mean absolute feature contribution values, respectively (Supplementary Figure S5). In addition, we determined the local SHapley Additive exPlanation (SHAP) values for individual samples. The AUCs of the model in the training and validation sets were 0.903 (0.832-0.975) and 0.877 (0.789-0.966), respectively (Table 3). In the internal test set, with pathological confirmation as the reference standard, 4 (66.7%) of 6

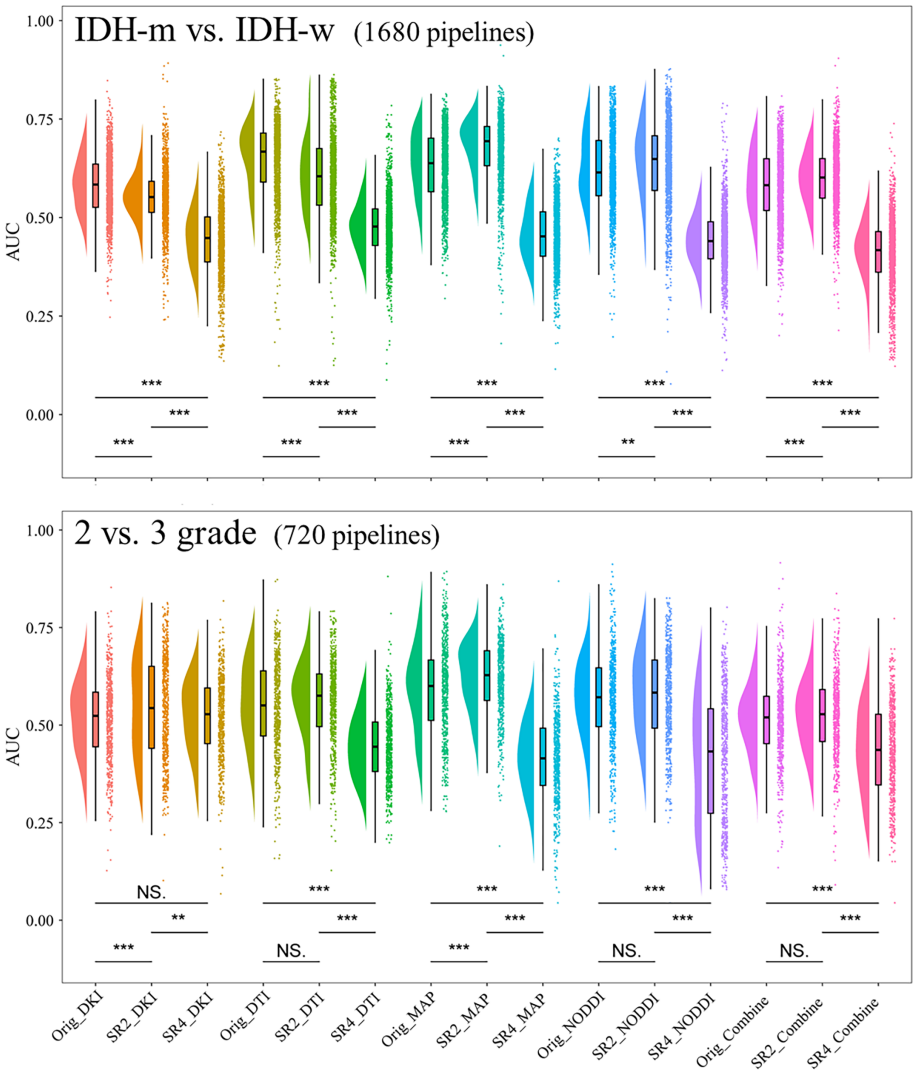


FIGURE 3
Comparison of the AUCs in the cross-validation sets of models constructed through different pipelines. **P <.01, ***P <.001. AUC, area under the curve; SR2, super-resolution reconstruction 2x; SR4, super-resolution reconstruction 4x; DTI, diffusion tensor imaging; DKI, diffusion kurtosis imaging; MAP, mean apparent propagation diffusion; NODDI, neurite orientation dispersion and density imaging. NS, No significance.

TABLE 2 Selected features for model construction.

Task	Feature origin (N)[a]	Pipeline[c] (normalization/ dimension reduction/feature selector/classifier)	Feature name
IDH-m vs. IDH-w	SR2-NODDI[b] (N = 6)	Mean/PCA/KW/LR	PCA_feature_5
			PCA_feature_11
			PCA_feature_13
			PCA_feature_24
			PCA_feature_31
			PCA_feature_75
IDH-m vs. IDH-w	Orig-MAP[b] (N = 6)	Minmax/PCC/Rel/NB	MSD_gldm_LargeDependenceLowGrayLevelEmphasis
			QIV_glrlm_LongRunHighGrayLevelEmphasis
			MSD_firstorder_Kurtosis
			QIV_gldm_SmallDependenceLowGrayLevelEmphasis
			RTOP_firstorder_Variance
			NG_glszm_LargeAreaLowGrayLevelEmphasis
IDH-m vs. IDH-w	Orig-NODDI[b] (N = 4)	Z score/PCC/RFE/SVM	ICVF_firstorder_Energy
			ODI_firstorder_Skewness
			ODI_glcm_Correlation
			ODI_glszm_SmallAreaHighGrayLevelEmphasis
Grade 2 vs. grade 3	Orig-MAP[b] (N = 2)	Minmax/PCC/Rel/NB	NG_gldm_LargeDependenceLowGrayLevelEmphasis
			QIV_firstorder_Kurtosis
Grade 2 vs. grade 3	Orig-DKI[b] (N = 1)	Minmax/PCC/RFE/LR	MK_glszm_LargeAreaHighGrayLevelEmphasis
Grade 2 vs. grade 3	Orig-DTI[b] (N = 2)	Minmax/PCA/ANOVA/LR	PCA_feature_1
			PCA_feature_40
Grade 2 vs. grade 3	SR2-Combine[b] (N = 1)	Minmax/PCC/RFE/LDA	MK_glrlm_LongRunHighGrayLevelEmphasis
Grade 2 vs. grade 3	SR2-MAP[b] (N = 2)	Mean/PCA/ANOVA/LR-Lasso	PCA_feature_2
			PCA_feature_41
Grade 2 vs. grade 3	SR2-NODDI[b] (N = 2)	Zscore/PCA/KW/AE	PCA_feature_1
			PCA_feature_34
Grade 2 vs. grade 3	SR4-MAP[b] (N = 2)	Minmax/PCA/KW/LDA	PCA_feature_6
			PCA_feature_36
Grade 2 vs. grade 3	SR4-NODDI[b] (N = 2)	Zscore/PCA/RFE/LR	PCA_feature_18
			PCA_feature_36

[a]The total number of features in the group.
[b]We included 11 different modeling approaches for identifying IDH type and predicting grade 2 and 3 glioma.
[c]Pipeline for processing valid data features for modeling.

instances were correctly predicted as having the IDH-mutant type, and 6 (50%) of 12 instances were correctly predicted as having the IDH-wild type (Supplementary Figure S6). Overall, the model achieved a favorable AUC of 0.819 (0.576-1) in the internal test set. When the threshold probability was greater than 20%, the SR2-based NODDI provided greater net clinical benefits than did the other 2 models in predicting IDH status compared with the case where no predictive model was used (Figure 4).

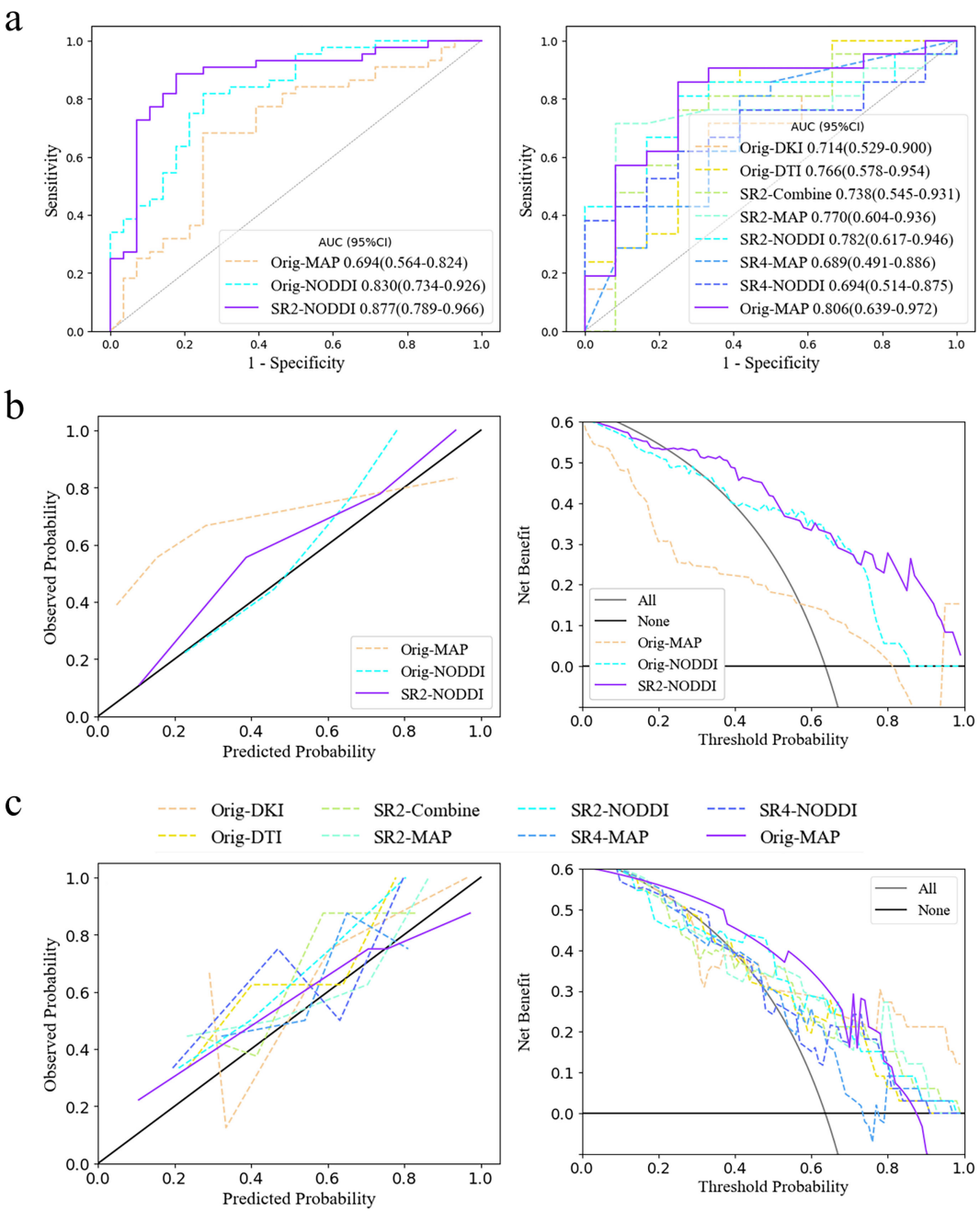


FIGURE 4
Model selection and clinical benefits. **(A)** ROC curves for the two tasks in the cross-validation set. The third and eight models had the highest diagnostic efficacy according to the integrated discrimination improvement index. **(B, C)** Calibration curve and decision curve analysis results in the training set in predicting IDH status and classifying tumors as grade 2/3. The models corresponding to the solid purple lines had the lowest Brier scores, 0.132 and 0.168, and their predictive ability was subsequently visualized via calibration curves (full Brier scores are provided in Supplementary Table S5). Additionally, these two models had the greatest net clinical benefits according to the decision curve analysis. AUC, area under the curve; SR2, super-resolution reconstruction 2x; SR4, super-resolution reconstruction 4x; DTI, diffusion tensor imaging; DKI, diffusion kurtosis imaging; MAP, mean apparent propagation diffusion; NODDI, neurite orientation dispersion and density imaging.

The original image-based MAP model was constructed with the naïve Bayes algorithm and consisted of 1 texture feature (non-Gaussianity), which had the highest mean absolute feature contribution value (0.35), and 1 first-order feature (q-space inverse variance). The AUCs of the model were 0.814 (0.642-0.985) and 0.806 (0.639-0.972) in the training and validation sets, respectively.

4 Discussion

In this study, we used an SR technique (GAN-based model migration) to improve the resolution of 25 quantitative parameter maps with 4 diffusion models at different magnifications and then constructed a multiparametric radiomic model to predict adult-type diffuse glioma IDH status and classify tumors into grades 2 and 3.

TABLE 3 Optimal model performance for the two tasks.

Task	Training set	Cross-validation set	Internal test set
IDH-m vs. IDH-w			
AUC*	0.903 (0.832–0.975)	0.877 (0.789–0.966)	0.819 (0.576–1)
Sensitivity	0.932 (41/44)	0.886 (39/44)	0.5 (6/12)
Specificity	0.786 (22/28)	0.821 (23/28)	0.667 (4/6)
PPV	0.872 (41/47)	0.886 (39/44)	0.75 (6/8)
NPV	0.88 (22/25)	0.821 (23/28)	0.4 (4/10)
ACC	0.875 (63/72)	0.861 (62/72)	0.556 (40/72)
Grade 2 vs. grade 3			
AUC*	0.814 (0.642–0.985)	0.806 (0.639–0.972)	–
Sensitivity	0.81 (17/21)	0.857 (18/21)	–
Specificity	0.75 (9/12)	0.75 (9/12)	–
PPV	0.85 (17/20)	0.857 (18/21)	–
NPV	0.692 (9/13)	0.75 (9/12)	–
ACC	0.788 (26/33)	0.818 (27/33)	–

Data in parentheses are the numerator/denominator of participants included for each parameter, unless otherwise indicated. The values correspond to the optimal threshold according to the maximum Youden index.
*Data are the means (95% CI).

Our study demonstrated that the diffusion MRI radiomic models noninvasively predicted IDH mutation status and tumor grade. The NODDI model based on SR2 images had the highest diagnostic efficacy (validation_AUC = 0.877) and stability (Brier value = 0.132) in predicting IDH mutation status; furthermore, the use of higher-resolution reconstructed images resulted in greater loss and a decrease in diagnostic performance. In addition, comparisons of the diffusion models indicated that advanced diffusion techniques were not always advantageous over DTI.

SR methods can be broadly divided into two categories: interpolation-based and learning-based techniques. Interpolation-based super-resolution reconstruction relies primarily on mathematical interpolation algorithms, such as nearest neighbor interpolation, bilinear interpolation, and bicubic interpolation. These algorithms estimate the values of newly added pixels by calculating the mathematical relationships among existing pixels. However, the primary limitation of this kind of approach is that it cannot introduce new high-frequency information, leading to a certain degree of image smoothing and, consequently, image blurriness, particularly around image edges and in regions with complex textures. Furthermore, the enhancement in spatial resolution is relatively limited, making it difficult to meet the demands of high-precision applications. In contrast, learning-based techniques employ deep learning models to learn the mapping between low-resolution and high-resolution images (19). Compared with interpolation-based methods, learning-based techniques offer superior reconstruction quality, stronger generalization capabilities, and the ability to support arbitrary magnification factors. The 3D super-resolution reconstruction technology presented in this study has achieved reliable results in enhancing the spatial resolution of CT images, which has improved clinical predictions for atherosclerotic plaques (25) and lung tumors (26).

Traditional morphologic visual evaluation, radiomics, and deep learning approaches for assessing tumor heterogeneity through imaging involve qualitative and quantitative analyses within limited regions of the image, regardless of the image source (e.g., radiological, pathological, and so on). Images with high quality actually make this task easier, whereas those with low quality can lead to 'incorrect' outcomes, especially when the lesion area is small and noise is present. Many medical image studies (16–18) have demonstrated the clinical benefits of deep learning-based SR both qualitatively and quantitatively. However, as the resolution increases, the number of image artifacts and the amount of noise also increases. At this point, effective estimation and integration of blur models may be more important than the utilization of image priors (27), i.e., models that use high- and low-resolution images. Images generated with default fixed blur models may be too unclear or contain oversharpened artifacts. In our study, higher magnification resulted in more artifacts and greater noise, which we believe is one of the main reasons for the decrease in diagnostic model performance. Various methods, such as mean filtering, have been developed for removing noise from different sources during (28) and after image generation. We did not apply additional denoising in our study because the tumor region image losses remained low, and excessive image smoothing could reduce the ability to visualize tumor heterogeneity in the images. Additionally, the residual channel attention network used in the present study is effective at both denoising and enhancing the resolution of medical images (29).

In terms of diffusion model theory, non-Gaussian diffusion models can better describe pathophysiological states in the brain than Gaussian models can. This means that the results derived from DKI, NODDI, and MAP data should be better than those derived from DTI data. However, the theoretical specificity of diffusion models leads to different practical strengths for each model. For example, NODDI better captures microstructural changes resulting from white matter diseases by quantifying changes in neurite direction (30), whereas non-Gaussianity provides a more detailed assessment of diffusion characteristics than fiber bundle imaging does (9).

In this study, we did not obtain sufficient evidence to support the hypothesis that advanced diffusion models offer greater clinical benefits than simple diffusion models. Studies by Gao et al. (11), Guo et al. (31), and Wang et al. (10) produced results similar to our results. However, some scholars (32–34) believe that the MAP and NODDI approaches can be used to better predict glioma heterogeneity or differentiate gliomas from metastatic tumors. Studies that drew the latter conclusion, however, used simple inferential statistics based on histogram averages of tumor entities or peritumoral edema without involving the extraction of additional features or the establishment of comprehensive models, and these limited attempts may not facilitate accurate conclusions. Notably, other factors, such as variations in the selection of regions of

interest, may have contributed to the differences in the results (35). FA has been included in various prediction models in previous studies; however, in the present study, DTI exhibited high losses when FA achieved the best SR effect, which may be one of the reasons why the clinical benefits of DTI were not significantly improved.

The results of this study indicate that the combined application of super-resolution reconstruction and magnetic resonance diffusion imaging may provide a more comprehensive understanding of the pathological features of gliomas, including cell density, invasiveness, and vascular distribution, thereby providing strong support for an accurate diagnosis. Additionally, clearer preoperative visualization translates to more precise preoperative planning, especially with the current high reliance on intraoperative fiber tracking navigation. However, issues such as technical complexity and generalizability, data processing and storage, standardization and normalization, and patient safety and privacy protection limit the implementation of this technology in clinical practice. The establishment of proprietary deep learning models on the basis of specific data and theoretical improvements in diffusion models are necessary.

This study has several limitations. First, the prospective nature of the study limited the number of subjects that could be analyzed, and diagnostic model building was relatively limited due to possible overfitting issues with the integrated model. Instead, clinical features and imaging morphological features were used for assessing the effects of potential confounding factors. Further large-scale, multicenter studies are needed to validate the discriminative performance of these diffusion-based indicators, reducing potential prospective biases and issues related to insufficient statistical power. Second, the study lacked true high-resolution multishell diffusion images for evaluating the reconstruction effect of GAN-based super-resolution; instead, a downsampling method followed by the addition of noise was adopted. Although this is a commonly used analysis method, additional data augmentation or optimization of the model structure can improve the model's generalization ability and reconstruction quality. Third, the number of diffusion models analyzed in the study was limited, and certain commonly used or higher-order diffusion models, such as the diffusion-weighted model and the continuous-time random walk model, were not included. Future studies should incorporate these common models when assessing the effects of SR. Fourth, the test-retest procedure for diffusion MRI was not attempted in the same instances, and the ROIs created by a small number of physicians may not be representative of all situations. However, in practice, retest procedures are limited in terms of time, cost, and instance tolerance.

In conclusion, we used GAN-based SR to improve the resolution of four diffusion models (DTI, DKI, NODDI, and MAP), allowing better visualization on multishell diffusion images and the possibility of quantitatively predicting IDH status and tumor grades 2 and 3 in adult-type diffuse glioma patients. Future work should include proprietary GAN model training and applications for specific diffusion models to further determine whether parameters fitted by multishell diffusion models can serve as imaging markers for adult-type diffuse glioma or other types of tumor.

Data availability statement

The datasets presented in this article are not readily available because data sharing that may reveal personally identifiable information about persons shall be done so carefully and in accordance with applicable laws and regulatory agencies. Requests to access the datasets should be directed to Yang Gao, 1390903990@qq.com.

Ethics statement

The studies involving humans were approved by Ethics Committee of Affiliated Hospital of Inner Mongolia Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

CZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. PW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review & editing. JH: Formal analysis, Methodology, Visualization, Writing – review & editing. QW: Formal analysis, Methodology, Supervision, Writing – review & editing. SX: Formal analysis, Methodology, Writing – review & editing. BL: Data curation, Investigation, Writing – review & editing. XH: Data curation, Investigation, Writing – review & editing. SW: Methodology, Software, Writing – review & editing. HZ: Resources, Writing – review & editing. ZH: Formal analysis, Investigation, Writing – review & editing. WG: Formal analysis, Investigation, Writing – review & editing. YL: Formal analysis, Investigation, Writing – review & editing. JG: Formal analysis, Investigation, Writing – review & editing. MH: Formal analysis, Investigation, Writing – review & editing. YG: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Science and Technology Planning Project of Inner Mongolia Autonomous Region (2019GG047); the research project of Inner Mongolia Medical University Affiliated Hospital, Inner Mongolia Autonomous Region Clinical Medicine Research Center of Nervous System Diseases, Hohhot Religion High-quality

Developmental and Advantageous Key Clinical Project of Neurological System Disease (2023NYFY LHYB008) and Natural Science Foundation of Inner Mongolia Autonomous Region (2024MS08015).

Acknowledgments

The authors gratefully acknowledge the essential contributions of the research staff of Affiliated Hospital of Inner Mongolia Medical University. Meanwhile, we are grateful to OnekeyAI and its developers for their invaluable assistance in this scientific research endeavor.

Conflict of interest

Authors SW and HZ were employed by the company Siemens Healthineers.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1435204/full#supplementary-material

References

1. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol.* (2021) 23:1231–51. doi: 10.1093/neuonc/noab106

2. Schaff LR, Mellinghoff IK. Glioblastoma and other primary brain Malignancies in adults: A review. *Jama.* (2023) 329:574–87. doi: 10.1001/jama.2023.0023

3. Weller M, van den Bent M, Preusser M, Le Rhun E, Tonn JC, Minniti G, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol.* (2021) 18:170–86. doi: 10.1038/s41571-020-00447-z

4. Smits M. MRI biomarkers in neuro-oncology. *Nat Rev Neurol.* (2021) 17:486–500. doi: 10.1038/s41582-021-00510-y

5. de Godoy LL, Chawla S, Brem S, Mohan S. Taming glioblastoma in "Real time": integrating multimodal advanced neuroimaging/AI tools towards creating a robust and therapy agnostic model for response assessment in neuro-oncology. *Clin Cancer Res.* (2023) 29:2588–92. doi: 10.1158/1078-0432.Ccr-23-0009

6. Zhang H, Liu K, Ba R, Zhang Z, Zhang Y, Chen Y, et al. Histological and molecular classifications of pediatric glioma with time-dependent diffusion MRI-based microstructural mapping. *Neuro Oncol.* (2023) 25:1146–56. doi: 10.1093/neuonc/noad003

7. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: practical in *vivo* neurite orientation dispersion and density imaging of the human brain. *Neuroimage.* (2012) 61:1000–16. doi: 10.1016/j.neuroimage.2012.03.072

8. Özarslan E, Koay CG, Shepherd TM, Komlosh ME, İrfanoğlu MO, Pierpaoli C, et al. Mean apparent propagator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure. *Neuroimage.* (2013) 78:16–32. doi: 10.1016/j.neuroimage.2013.04.016

9. Ning L, Laun F, Gur Y, DiBella EV, Deslauriers-Gauthier S, Megherbi T, et al. Sparse Reconstruction Challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use? *Med Image Anal.* (2015) 26:316–31. doi: 10.1016/j.media.2015.10.012

10. Wang P, Xie S, Wu Q, Weng L, Hao Z, Yuan P, et al. Model incorporating multiple diffusion MRI features: development and validation of a radiomics-based model to predict adult-type diffuse gliomas grade. *Eur Radiol.* (2023) 33:8809–20. doi: 10.1007/s00330-023-09861-0

11. Gao A, Zhang H, Yan X, Wang S, Chen Q, Gao E, et al. Whole-tumor histogram analysis of multiple diffusion metrics for glioma genotyping. *Radiology.* (2022) 302:652–61. doi: 10.1148/radiol.210820

12. Hirschler L, Sollmann N, Schmitz-Abecassis B, Pinto J, Arzanforoosh F, Barkhof F, et al. Advanced MR techniques for preoperative glioma characterization: part 1. *J Magn Reson Imaging.* (2023) 57:1655–75. doi: 10.1002/jmri.28662

13. Gendy G, He G, Sabor N. Lightweight image super-resolution based on deep learning: State-of-the-art and future directions. *Inf Fusion.* (2023) 94:284–310. doi: 10.1016/j.inffus.2023.01.024

14. Zou WW, Yuen PC. Very low resolution face recognition problem. *IEEE Trans Image Process.* (2012) 21:327–40. doi: 10.1109/tip.2011.2162423

15. Sajjadi MSM, Schölkopf B, Hirsch M. (2017). EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy. (2017) pp. 4501–10. doi: 10.1109/ICCV.2017.481

16. Varentsova A, Zhang S, Arfanakis K. Development of a high angular resolution diffusion imaging human brain template. *Neuroimage.* (2014) 91:177–86. doi: 10.1016/j.neuroimage.2014.01.009

17. Iglesias JE, Schleicher R, Laguna S, Billot B, Schaefer P, McKaig B, et al. Quantitative brain morphometry of portable low-field-strength MRI using super-resolution machine learning. *Radiology.* (2023) 306:e220522. doi: 10.1148/radiol.220522

18. Hou M, Zhou L, Sun J. Deep-learning-based 3D super-resolution MRI radiomics model: superior predictive performance in preoperative T-staging of rectal cancer. *Eur Radiol.* (2023) 33:1–10. doi: 10.1007/s00330-022-08952-8

19. Bianconi A, Rossi LF, Bonada M, Zeppa P, Nico E, De Marco R, et al. Deep learning-based algorithm for postoperative glioblastoma MRI segmentation: a promising new tool for tumor burden assessment. *Brain Inform.* (2023) 10:26. doi: 10.1186/s40708-023-00207-6

20. Kocak B, Baessler B, Bakas S, Cuocolo R, Fedorov A, Maier-Hein L, 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) 14:75. doi: 10.1186/s13244-023-01415-8

21. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. *2018 European Conference on Computer Vision*, Munich, Germany. (2018) pp. 294–310. doi: 10.1007/978-3-030-01234-2_18

22. 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:e0237587. doi: 10.1371/journal.pone.0237587

23. Halligan S, Menu Y, Mallett S. Why did European Radiology reject my radiomic biomarker paper? How to correctly evaluate imaging biomarkers in a clinical setting. *Eur Radiol.* (2021) 31:9361–8. doi: 10.1007/s00330-021-07971-1

24. Kerr KF. Net reclassification index statistics do not help assess new risk models. *Radiology.* (2023) 306:e222343. doi: 10.1148/radiol.222343

25. Wang L, Guo T, Wang L, Yang W, Wang J, Nie J, et al. Improving radiomic modeling for the identification of symptomatic carotid atherosclerotic plaques using deep learning-based 3D super-resolution CT angiography. *Heliyon.* (2024) 10:e29331. doi: 10.1016/j.heliyon.2024.e29331

26. Xing X, Li L, Sun M, Yang J, Zhu X, Peng F, et al. Deep-learning-based 3D super-resolution CT radiomics model: Predict the possibility of the micropapillary/solid component of lung adenocarcinoma. *Heliyon.* (2024) 10:e34163. doi: 10.1016/j.heliyon.2024.e34163

27. Efrat N, Glasner D, Apartsin A, Nadler B, Levin A. (2013). Accurate Blur Models vs. Image Priors in Single Image Super-resolution. *2013 IEEE International Conference*

on Computer Vision, Sydney, NSW, Australia. (2013) pp. 2832–2839. doi: 10.1109/ICCV.2013.352

28. HashemizadehKolowri SK, Chen RR, Adluru G, Dean DC, Wilde EA, Alexander AL, et al. Simultaneous multi-slice image reconstruction using regularized image domain split slice-GRAPPA for diffusion MRI. *Med Image Anal*. (2021) 70:102000. doi: 10.1016/j.media.2021.102000

29. Chen J, Sasaki H, Lai H, Su Y, Liu J, Wu Y, et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nat Methods*. (2021) 18:678–87. doi: 10.1038/s41592-021-01155-x

30. Lawrence KE, Nabulsi L, Santhalingam V, Abaryan Z, Villalon-Reina JE, Nir TM, et al. Age and sex effects on advanced white matter microstructure measures in 15,628 older adults: A UK biobank study. *Brain Imaging Behav*. (2021) 15:2813–23. doi: 10.1007/s11682-021-00548-y

31. Guo H, Liu J, Hu J, Zhang H, Zhao W, Gao M, et al. Diagnostic performance of gliomas grading and IDH status decoding A comparison between 3D amide proton transfer APT and four diffusion-weighted MRI models. *J Magn Reson Imaging*. (2022) 56:1834–44. doi: 10.1002/jmri.28211

32. Sun Y, Su C, Deng K, Hu X, Xue Y, Jiang R. Mean apparent propagator-MRI in evaluation of glioma grade, cellular proliferation, and IDH-1 gene mutation status. *Eur Radiol*. (2022) 32:3744–54. doi: 10.1007/s00330-021-08522-4

33. Zeng S, Ma H, Xie D, Huang Y, Yang J, Lin F, et al. Tumor multiregional mean apparent propagator (MAP) features in evaluating gliomas-A comparative study with diffusion kurtosis imaging (DKI). *J Magn Reson Imaging*. (2023). doi: 10.1002/jmri.29202

34. Mao J, Zeng W, Zhang Q, Yang Z, Yan X, Zhang H, et al. Differentiation between high-grade gliomas and solitary brain metastases: a comparison of five diffusion-weighted MRI models. *BMC Med Imaging*. (2020) 20:124. doi: 10.1186/s12880-020-00524-w

35. Chu JP, Song YK, Tian YS, Qiu HS, Huang XH, Wang YL, et al. Diffusion kurtosis imaging in evaluating gliomas: different region of interest selection methods on time efficiency, measurement repeatability, and diagnostic ability. *Eur Radiol*. (2021) 31:729–39. doi: 10.1007/s00330-020-07204-x

frontiers | Frontiers in Neurology

TYPE Case Report
PUBLISHED 11 October 2024
DOI 10.3389/fneur.2024.1471668

OPEN ACCESS

EDITED BY
Vincenzo Di Nunno,
AUSL Bologna, Italy

REVIEWED BY
Reza Assadsangabi,
University of Southern California,
United States
Hernan Chaves,
Fundación Para la Lucha Contra las Enfermedades Neurológicas de la Infancia (FLENI), Argentina

*CORRESPONDENCE
Ryota Amano
✉ d10sm003@yahoo.co.jp

RECEIVED 28 July 2024
ACCEPTED 23 September 2024
PUBLISHED 11 October 2024

CITATION
Amano R, Sunouchi A, Yokota Y and Mochizuki K (2024) Case report: An autopsy report of patient with metastatic brain tumor and carcinomatous meningitis mimicking paraneoplastic neurological syndrome. *Front. Neurol.* 15:1471668. doi: 10.3389/fneur.2024.1471668

Case report: An autopsy report of patient with metastatic brain tumor and carcinomatous meningitis mimicking paraneoplastic neurological syndrome

Ryota Amano[1]*, Azusa Sunouchi[1], Yuka Yokota[2] and Kunio Mochizuki[2]

[1]Department of Neurology, Fujiyoshida Municipal Medical Center, Yamanashi, Japan, [2]Department of Pathology, University of Yamanashi, Yamanashi, Japan

Differential diagnosis of metastatic brain tumor, carcinomatous meningitis, and paraneoplastic neurological syndrome (PNS) can be challenging in atypical cases. When examining patient with increased T2 fluid-attenuated inversion recovery (FLAIR) hyperintensities in the temporal polar white matter, autoimmune encephalitis, including PNS, should be considered. Herein, we report the case of an 85-year-old man with carcinomatous meningitis due to lung large cell carcinoma. He showed disturbance of consciousness, abnormal behavior, incomprehensible speech, and apathy, which suggested brain dysfunction. Magnetic resonance imaging revealed high intensities on the whole cerebellum on a diffusion-weighted image and bilateral T2 FLAIR hyperintensities in the temporal polar white matter. Cerebrospinal fluid analysis and cytology showed elevated total protein levels, pleocytosis, and atypical cells with nuclear enlargement, hyperchromasia, and irregular shape. Autopsy revealed lung large cell carcinoma and its brain metastasis. Tumor cells were disseminated to the central nervous system along the subarachnoid space. Furthermore, plenty of carcinoma cells and peritumoral enlarged perivascular space were observed in the temporal poles. To our knowledge, this is the first report of bilateral T2 FLAIR hyperintensities in the temporal polar white matter caused by carcinomatous meningitis with pathological confirmation. In patient with carcinomatous meningitis, abnormal T2 FLAIR hyperintensities may not be derived from ischemia or tumor invasion to parenchyma.

KEYWORDS

T2 FLAIR hyperintensity on bilateral temporal polar white matter, paraneoplastic neurological syndrome, lung large cell carcinoma, metastatic brain tumor, carcinomatous meningitis

1 Introduction

Metastatic brain tumors and carcinomatous meningitis are two forms of secondary brain involvement that occur when cancer cells spread from a primary tumor located elsewhere in the body to the brain and its surrounding structures. Patients with metastatic brain tumors or carcinomatous meningitis generally have a poor prognosis despite

the treatment (1). When examining patients with malignant tumors who present with neurologic symptoms, neurologists should consider paraneoplastic neurological syndrome (PNS) as a differential diagnosis. PNS is an immune-mediated neurological disorder caused by antibodies against intracellular, neuronal surface, or synaptic proteins expressed by cancer cells. Patients with PNS exhibit various neurological symptoms and frequent abnormal intensities on magnetic resonance imaging (MRI) in regions of central nervous systems (2, 3). Although detecting anti-neuronal antibodies can aid in diagnosing PNS, their sensitivities and specificities are not necessarily high. Therefore, the possibility of PNS cannot be ruled out, even if anti-neuronal antibodies are not detected (4).

Generally, carcinomatous meningitis presents with abnormal enhancement of the meninges or cranial nerve on T1-weighted gadolinium imaging, as well as hyperintensities on T2 fluid-attenuated inversion recovery (FLAIR), and hydrocephalus can be observed in carcinomatous meningitis (5). Detecting T2 FLAIR hyperintensities in cerebral white matter by carcinomatous meningitis was exceptional, and, in that case, metastatic brain tumors and its surrounding edema should be considered. The representative disease that shows T2 FLAIR hyperintensities in cerebral white matter in cancer patients is PNS, except for metastatic brain tumors. Differential diagnosis of metastatic tumor, carcinomatous meningitis, and PNS is not difficult in typical cases; however, it is challenging in atypical cases.

Herein, we reported a case of metastatic brain tumor and carcinomatous meningitis in a patient with lung large cell carcinoma. MRI revealed high intensities on the whole cerebellum on diffusion-weighted imaging (DWI) and hyperintensities in the bilateral temporal polar white matters on T2 FLAIR. At autopsy, lung large cell carcinoma and its metastasis were found. Tumor cells had disseminated throughout the central nervous system along the subarachnoid space. However, no inflammatory cell infiltration into the brain parenchyma was observed in the temporal poles, suggesting a tumor immunity to cancer cells rather than autoimmune encephalitis, including PNS. This is the first report demonstrating bilateral T2 FLAIR hyperintensities in the temporal polar white matters caused by carcinomatous meningitis with pathological confirmation.

2 Case description

An 85-year-old man presented to our hospital with an altered consciousness and an abnormal shadow on the right hilar region. He exhibited abnormal behavior, incomprehensible speech, and apathy, for at least 7 days before admission, which suggested brain dysfunction. Upon admission, neurological examination confirmed altered consciousness [Glasgow Coma Scale of 9 (E4V1M4)], with nuchal rigidity and slightly decreased deep tendon reflexes in the lower limbs.

Initial laboratory examinations revealed slightly high serum levels of C-reactive protein (CRP: 0.91 mg/dL) and indicated dehydration and renal failure (blood urea nitrogen: 56.6 mg/dL, creatinine: 1.54 mg/dL). Tumor-specific laboratory tests revealed high serum levels of neuron-specific enolase (NSE: 59.6 ng/mL).

Cerebrospinal fluid (CSF) analysis revealed pleocytosis with mononuclear cell predominance (40 leucocytes/μL with 80% of mononuclear leucocytes) and increased total protein levels (TP; 244 mg/dL) with decreased glucose levels (glucose; 24 mg/dL). CSF cytology showed atypical cells with nuclear enlargement, hyperchromasia, and irregular shape, which suggested carcinomatous meningitis.

Chest radiography and computed tomography (CT) revealed a right hilar tumorous lesion (Figures 1A, B). Brain MRI revealed T2 FLAIR hyperintensities in the bilateral temporal polar white matter and left-predominant edematous cerebellar lesions (Figures 1E–G). DWI showed high intensities on the whole cerebellum but no apparent abnormalities in bilateral temporal poles (Figures 1H–J). No tumorous lesions were apparent, and contrast-enhanced MRI could not be performed because of renal failure.

According to the MRI findings, we considered the possibility of PNS; however, paraneoplastic screening by a fixed tissue-based assay using rat hippocampus and cerebellum (Euroimmune) showed no significance. Moreover, EUROLINE PNS12 Ag (Euroimmune) only showed a weak positive between 6 and 10, suggesting a low titer of anti-recoverin antibodies, an autoantibody typically associated with autoimmune retinopathy.

He received two courses of high-dose methylprednisolone over 2 weeks (1,000 mg/day × 3 days intravenously as one course per week); however, increased pleural effusion and cryptogenic acute subdural hematoma were observed 10 days after admission (Figures 1C, D). Despite the palliative care, he died 15 days after admission.

3 Postmortem pathological findings

An autopsy was performed 23 h and 40 min after his death. A tumor measuring 6.0 cm diameter was identified in the pulmonary hilum of the right upper lobe of the lung (Figure 2A). Histological examination revealed cells with a high nucleocytoplasmic ratio and large polygonal cells, proliferating in a solid or a trabecular pattern, with a mixture of cells with well-defined nucleoli (Figure 2B). Extensive necrosis was observed within the tumor area. Immunohistochemical analysis of tumor cells revealed the following results: AE1/AE3(+), TTF-1(−), p40(−), chromogranin A(−), synaptophysin(−), CD56(−), INSM1(−). Mitotic counts reached up to 15 mitoses/high-power field (HPF), with an overall Ki67 labeling index of ∼5% and up to 40% for hot spots. Based on these findings, we diagnosed the case as large cell carcinoma (null immunophenotype) (6). Additionally, lymph node metastasis and bilateral adrenal metastasis were present.

The brain weighed 1,270 g and showed a subdural hematoma (Figure 2C). There was a 6 mm diameter nodule in the left cerebellum and a 4 mm diameter nodule in the right globus pallidus, which were histologically identified as brain metastases of lung cancer (Figures 2D, E). Additionally, tumor cells were found infiltrating the subarachnoid space, indicating meningeal dissemination. The meningeal dissemination extended over the surface of the entire central nervous system, including the cerebrum, brainstem, cerebellum, pituitary gland, and spinal cord.

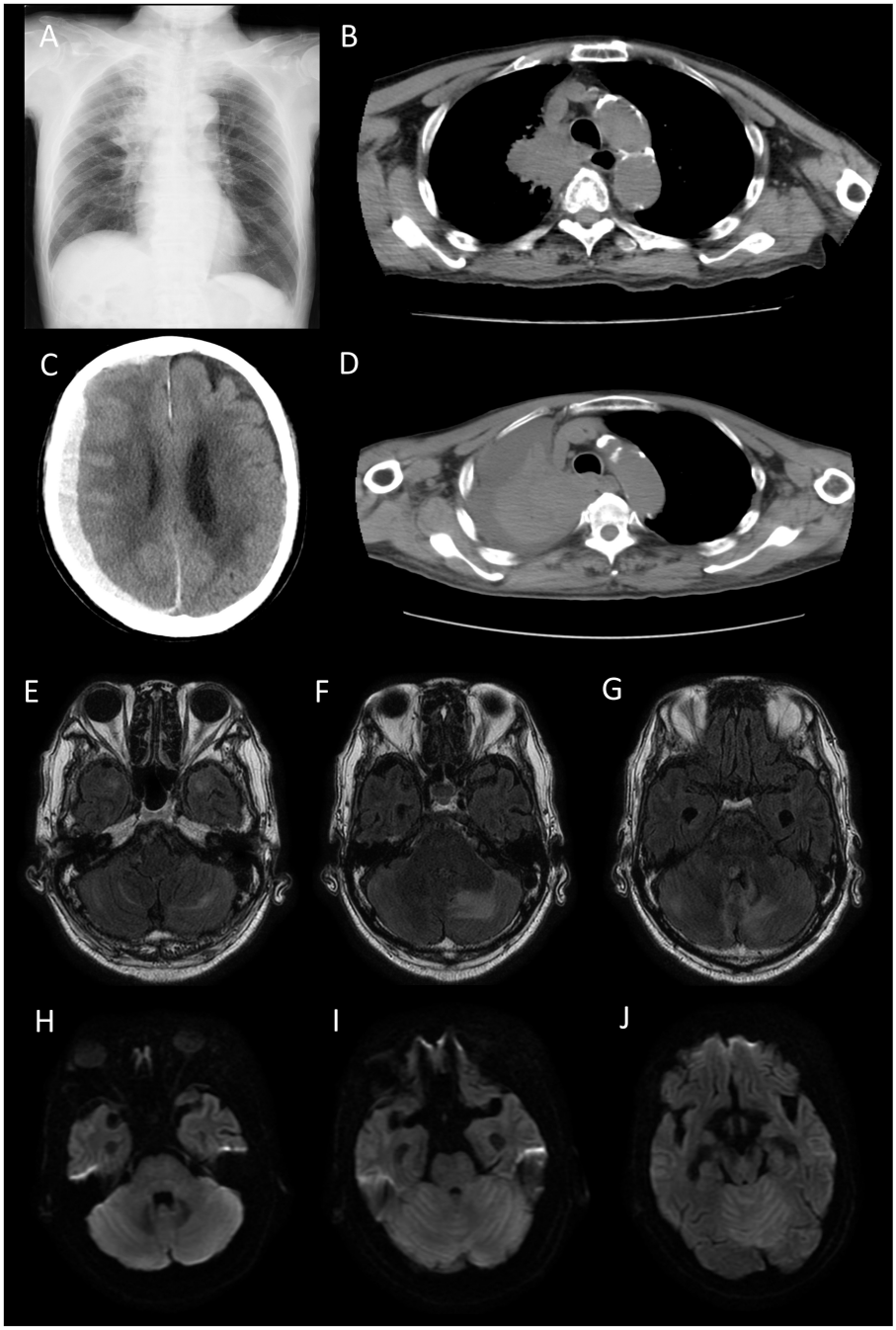


FIGURE 1
Chest radiograph **(A)** and computed tomography (CT) scan **(B)** on admission. A right hilar tumorous lesion was observed. Brain CT **(C)** and chest CT **(D)** on 10 days after admission. Cryptogenic acute subdural hematoma and increased pleural effusion were apparent. Fluid-attenuated inversion recovery (FLAIR) **(E–G)** and Diffusion-weighted image (DWI) **(H–J)** on admission. FLAIR showed T2 hyperintensities in the bilateral temporal polar white matter and light-predominant edematous cerebellar lesions. DWI showed high intensities in the whole cerebellum but no apparent abnormalities in the bilateral temporal poles. No tumorous lesions were apparent.

The cause of the subdural hematoma was considered to be vascular rupture on the brain surface due to tumor cell infiltration.

In the cerebellum, the tumor extended into the perivascular space (PVS) (Figure 3A) and invaded arteries (Figures 3A', A"). CD3-immunopositive lymphocytes were limited to the tumoral area (Figure 3A"'). Moreover, the tumor extended the brain parenchyma through the PVS (Figure 3B) and invaded arteries (Figures 3B', B"). CD3-immunopositive lymphocytes were limited to the tumoral area (Figure 3B"').

In the temporal pole, the tumor infiltrated along the subarachnoid space and PVS (Figures 4A, C). The PVS was edematous and enlarged (Figure 4B). The tumor also invaded arteries (Figures 4C', C"). CD3-immunopositive lymphocytes were limited to the tumoral area, with no inflammatory cell infiltration

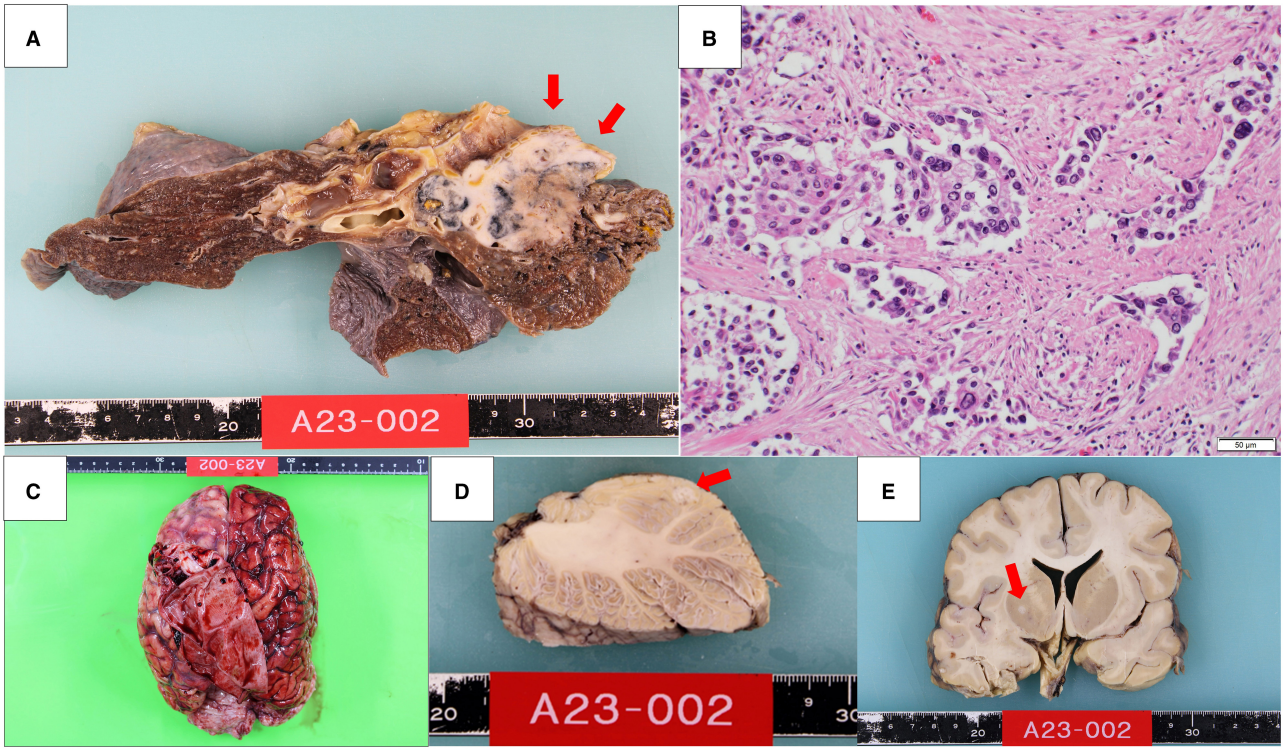


FIGURE 2
A 6.0 cm diameter tumor was found in the pulmonary hilum of the right upper lobe of the lung **(A)**. Histologically, cells with a high nucleocytoplasmic ratio and large polygonal cells proliferated in a solid or a trabecular pattern **(B)**. Macroscopic features of the brain **(C)**. There was a 6 mm diameter nodule in the left cerebellum **(D)** and a 4 mm diameter nodule in the right globus pallidus **(E)**.

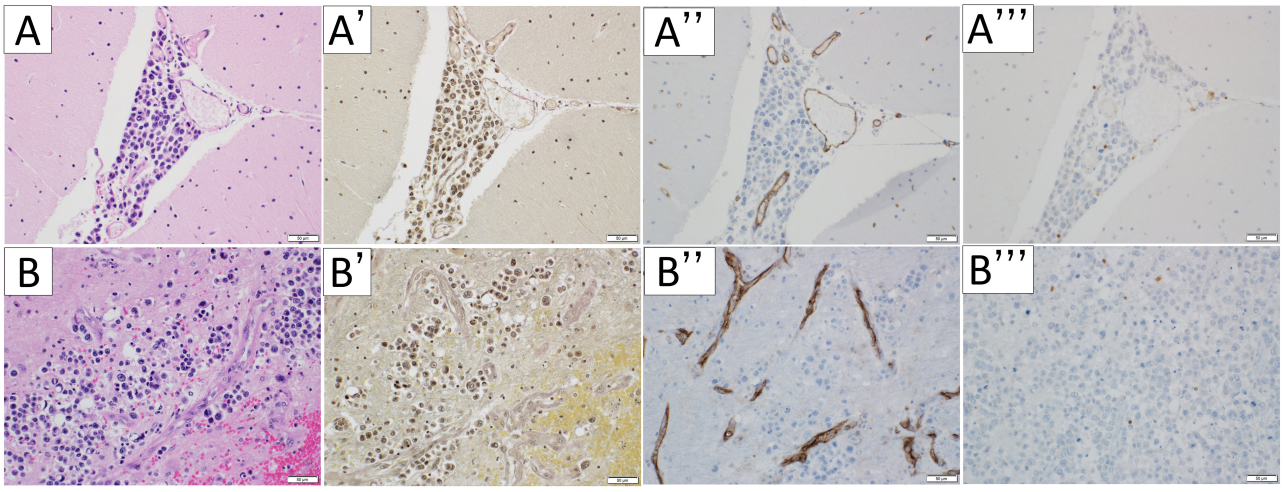


FIGURE 3
Microscopic appearance of the cerebellum. Hematoxylin-eosin staining **(A, B)**, Elastica van Gieson stain **(A', B')**, immunohistochemical staining for CD31 (**A", B"**), and CD3 **(A"', B"')** were performed respectively. In the cerebellum, the tumor extended into the perivascular space **(A)**. The vessels the tumor invades are arteries **(A', A")**. CD3-immunopositive lymphocytes were limited to the tumoral area **(A"', B"')**. The tumor extended the brain parenchyma through the perivascular space **(B)** and invalid arteries **(B', B")**. The scale bar length is 50 μm.

into the brain parenchyma, suggesting tumor immunity to cancer cells rather than autoimmune encephalitis including PNS (Figures 4B"', C"') (7). There was no evidence of neuronal cell loss or gliosis. The infiltrating tumor cells were larger than other sites, and the PVS was edematous and enlarged around the tumor lesion. However, there were no apparent differences between the PVS area of the temporal pole and those of the other regions, except for peritumoral PVS.

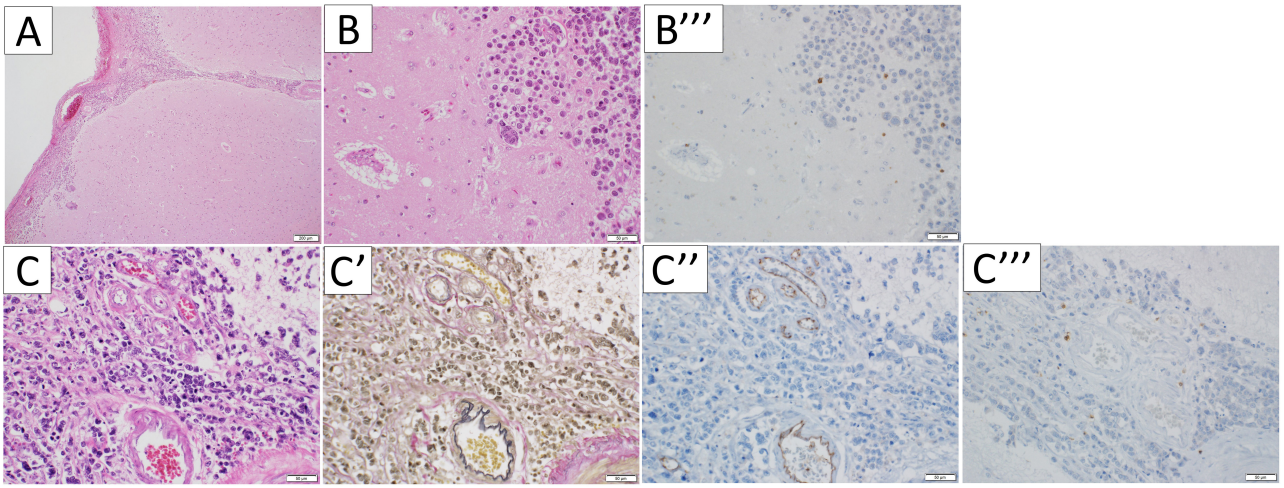


FIGURE 4
Microscopic appearance of the temporal pole. Hematoxylin-eosin staining **(A, B, C)**, Elastica van Gieson stain **(C′)**, and immunohistochemical staining for CD31 **(C″)** and CD3 **(B‴, C‴)**, were performed respectively. In the temporal pole, the tumor infiltrated along the subarachnoid and perivascular spaces **(A, C)**, and the perivascular spaces were edematous and enlarged **(B)**. The vessels the tumor invades are arteries **(C′, C″)**. CD3-immunopositive lymphocytes were limited to the tumoral area **(B‴, C‴)**.

4 Discussion

In this case, the presence of carcinomatous meningitis was confirmed through CSF cytology, which revealed malignant cells, and paraneoplastic screening using a fixed tissue-based assay in rat hippocampus and cerebellum (Euroimmune) showed no considerable changes. Moreover, EUROLINE PNS12 Ag (Euroimmune) did not detect any antibody that could explain the patient's symptoms. Furthermore, postmortem pathological findings did not support autoimmune mechanisms. Despite this, we considered the possibility of co-existing PNS before patient's death, because T2 FLAIR hyperintensities spread to anatomically distant regions (bilateral temporal polar white matter and cerebellum). This pattern is atypical for carcinomatous meningitis. Contrast enhanced T1 weighted image is useful for differentiating brain metastasis, carcinomatous meningitis, and PNS; however, some patients with carcinomatous meningitis have been reported to show brain stem or cerebral cortical or white matter T2 FLAIR hyperintensities without gadolinium enhancement (8, 9). Therefore, it was difficult to rule out overlapping conditions, even if our patient had received a contrast-enhanced T1 weighted image.

Although T2 FLAIR hyperintensities in temporal polar white matter are especially suggested to be diagnostic for cerebral autosomal dominant arteriopathy or myotonic dystrophy type 1 (10), dilated PVS, which mimic cystic neoplasm on brain MRI, are also known to cause T2 FLAIR hyperintensities in temporal polar white matter (11). In conditions such as cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) and dilated PVS, T2 FLAIR hyperintensities are believed to result from the inhibition of fluid drainage (11, 12). While these can occur through various mechanisms, their occurrence in other conditions, apart from autoimmune encephalitis and PNS, is rarely reported. Moreover, although T2 FLAIR hyperintensities in the cerebral white matter which are caused by carcinomatous meningitis without pathological confirmation are sometimes observed in daily clinical practice, to our knowledge, they have rarely been reported.

The possible factors that contribute to developing edema (T2 FLAIR hyperintensity) in bilateral temporal polar white matter are as follows: physical blood-brain barrier disruption by tumor infiltration, secretion of cytokines that increase the permeability of blood vessels, inhibition of interstitial and CSF drainage system (glymphatic system), and inflammatory response against tumor cells. In this case, the number of malignant cells was higher in the temporal poles than in other sites. Additionally, PVS, which are potential spaces where layers of connective tissue surround the blood vessels and facilitate the passage of interstitial and CSF of the brain acting as a drainage system, known as the glymphatic system (13, 14), were enlarged around the metastatic regions. Thus, the inhibition of the glymphatic system and the inflammatory response against tumor cells may have contributed to the development of edema.

Ayzenberg et al. reported a case of carcinomatous meningitis with diffuse cortical and white matter hyperintensities on DWI and T2 FLAIR imaging (8). They also reported histopathological findings and considered that perivascular and intravascular infiltration by tumor cells resulted in ischemic cortical and white matter stroke; however, in our case, DWI did not show abnormalities in the temporal polar white matter, suggesting that the mechanism behind T2 FLAIR hyperintensities in the temporal polar white matter may not be due to ischemic stroke. It is difficult to explain the mechanisms of edema developing in the bilateral temporal polar white matter since multiple factors may be involved however, we suspect that the disruption of glymphatic system by tumor cell is at least one contributing factor.

In conclusion, we reported a case of metastatic brain tumor and carcinomatous meningitis mimicking PNS in a patient with lung large cell carcinoma. Although brain MRI, especially in the

temporal polar white matter, suggested PNS rather than metastatic brain tumor and carcinomatous meningitis, pathological findings revealed the presence of tumor cells and enlarged PVS, with no evidence of autoimmune mechanisms. This is the first report of bilateral T2 FLAIR hyperintensities in the temporal polar white matter caused by carcinomatous meningitis. Physicians should consider carcinomatous meningitis as a differential diagnosis in patients with cancer who show bilateral T2 FLAIR hyperintensities in the temporal polar white matter.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

RA: Data curation, Writing – original draft, Writing – review & editing. AS: Data curation, Writing – review & editing. YY: Data curation, Writing – original draft, Writing – review & editing. KM: Data curation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank Editage (http://www.editage.com) for the English language editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Pellerino A, Garbossa D, Rudà R, Soffietti R. The role of the neurologist in the diagnosis and treatment of brain metastases and carcinomatous meningitis. *Rev Neurol.* (2023) 179:464–74. doi: 10.1016/j.neurol.2023.03.005

2. Devine MF, Kothapalli N, Elkhooly M, Dubey D. Paraneoplastic neurological syndromes: clinical presentations and management. *Ther Adv Neurol Disord.* (2021) 14:1756286420985323. doi: 10.1177/1756286420985323

3. Liang C, Chu E, Kuoy E, Soun JE. Autoimmune-mediated encephalitis and mimics: a neuroimaging review. *J Neuroimaging.* (2023) 33:19–34. doi: 10.1111/jon.13060

4. Amano R, Kim YJ, Yoshida T, Hara M, Nakajima H, Ohtsuka T, et al. Case report: reversible brain atrophy with low titer anti-amphiphysin antibodies related to gastric adenocarcinoma. *Front Neurol.* (2023) 14:1211814. doi: 10.3389/fneur.2023.1211814

5. Sakai M, Kashiwagi N, Nakanishi K, Maeda N, Nakaya Y, Tanaka J, et al. Nonbrain metastases seen on magnetic resonance imaging during metastatic brain tumor screening. *Jpn J Radiol.* (2023) 41:367–81. doi: 10.1007/s11604-022-01362-2

6. Rossi G, Leighl NB, Lu S, Nicholson AG, Smit EF. Large cell carcinoma of lung. In: Lokuhetty D, White VA, Cree IA, editors. *WHO Classification of Thoracic Tumours*. 5th ed. Lyon: International Agency for Research on Cancer (2021). p. 97–9.

7. Kerner C, Kotobelli K, Appleby BS, Cohen ML, Abboud H. Pathological findings in autoimmune encephalitis autopsy specimens from cases of suspected prion disease. *J Neurol.* (2024) 271:446–56. doi: 10.1007/s00415-023-12003-7

8. Ayzenberg I, Börnke C, Tönnes C, Ziebarth W, Lavrov A, Lukas C. Extensive cortical involvement in leptomeningeal carcinomatosis. *J Clin Neurosci.* (2012) 19:1723–5. doi: 10.1016/j.jocn.2011.12.029

9. Crombe A, Alberti N, Durieux M, Frulio N, Dousset V, Tourdias T. Exceptional symmetric anterior brainstem involvement in leptomeningeal carcinomatosis. *J Neuroradiol.* (2014) 41:279–81. doi: 10.1016/j.neurad.2014.05.003

10. Kim H, Lim YM, Oh YJ, Lee EJ, Kim KK. Comparison of brain magnetic resonance imaging between myotonic dystrophy type 1 and cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *PLoS ONE.* (2018) 13:e0208620. doi: 10.1371/journal.pone.0208620

11. Rawal S, Croul SE, Willinsky RA, Tymianski M, Krings T. Subcortical cystic lesions within the anterior superior temporal gyrus: a newly recognized characteristic location for dilated perivascular spaces. *Am J Neuroradiol.* (2014) 35:317–22. doi: 10.3174/ajnr.A3669

12. Yamamoto Y, Ihara M, Tham C, Low RW, Slade JY, Moss T, et al. Neuropathological correlates of temporal pole white matter hyperintensities in CADASIL. *Stroke.* (2009) 40:2004–11. doi: 10.1161/strokeaha.108.528299

13. Wardlaw JM, Benveniste H, Nedergaard M, Zlokovic BV, Mestre H, Lee H, et al. Perivascular spaces in the brain: anatomy, physiology and pathology. *Nat Rev Neurol.* (2020) 16:137–53. doi: 10.1038/s41582-020-0312-z

14. Hablitz LM, Nedergaard M. The glymphatic system: a novel component of fundamental neurobiology. *J Neurosci.* (2021) 41:7698–711. doi: 10.1523/JNEUROSCI.0619-21.2021

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 14 November 2024
DOI 10.3389/fonc.2024.1449235

OPEN ACCESS

EDITED BY
Domenico Aquino,
IRCCS Carlo Besta Neurological Institute Foundation, Italy

REVIEWED BY
Fulvia Palesi,
University of Pavia, Italy
Riccardo Pascuzzo,
IRCCS Carlo Besta Neurological Institute Foundation, Italy

*CORRESPONDENCE
Rosellina Russo
rosellina.russo@policlinicogemelli.it

†These authors have contributed equally to this work

‡These authors share senior authorship

RECEIVED 14 June 2024
ACCEPTED 16 October 2024
PUBLISHED 14 November 2024

CITATION
Pignotti F, Ius T, Russo R, Bagatto D, Beghella Bartoli F, Boccia E, Boldrini L, Chiesa S, Ciardi C, Cusumano D, Giordano C, La Rocca G, Mazzarella C, Mazzucchi E, Olivi A, Skrap M, Tran HE, Varcasia G, Gaudino S and Sabatino G (2024) Development and validation of a MRI-radiomics-based machine learning approach in High Grade Glioma to detect early recurrence. *Front. Oncol.* 14:1449235. doi: 10.3389/fonc.2024.1449235

Development and validation of a MRI-radiomics-based machine learning approach in High Grade Glioma to detect early recurrence

Fabrizio Pignotti[1,2†], Tamara Ius[3†], Rosellina Russo[4*‡], Daniele Bagatto[5], Francesco Beghella Bartoli[6], Edda Boccia[6], Luca Boldrini[6], Silvia Chiesa[6], Chiara Ciardi[5], Davide Cusumano[7], Carolina Giordano[4], Giuseppe La Rocca[8], Ciro Mazzarella[6], Edoardo Mazzucchi[1,2], Alessandro Olivi[2,8], Miran Skrap[3], Houng Elena Tran[6], Giuseppe Varcasia[4], Simona Gaudino[2,4‡] and Giovanni Sabatino[1,2,8‡]

[1]Department of Neurosurgery, Mater Olbia Hospital, Olbia, Italy, [2]Institute of Neurosurgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Catholic University, Rome, Italy, [3]Neurosurgery Unit, Head-Neck and NeuroScience Department, University Hospital of Udine, Udine, Italy, [4]Advanced Radiodiagnostics Centre, Unità Operativa Semplice Dipartimentale (UOSD) Neuroradiology, Fondazione Policlinico Universitario Agostino Gemelli IRCSS, Rome, Italy, [5]Department of Neuroradiology, Azienda Sanitaria Universitaria Friuli Centrale (ASUFC) "Santa Maria Della Misericordia", Udine, Italy, [6]Department of Radiology, Radiation Oncology and Hematology, Fondazione Policlinico Universitario Agostino Gemelli IRCSS, Rome, Italy, [7]Medical Physics Unit, Mater Olbia Hospital, Olbia, Italy, [8]Institute of Neurosurgery, Fondazione Policlinico Universitario Agostino Gemelli IRCSS, Rome, Italy

Purpose: Patients diagnosed with High Grade Gliomas (HGG) generally tend to have a relatively negative prognosis with a high risk of early tumor recurrence (TR) after post-operative radio-chemotherapy. The assessment of the pre-operative risk of early versus delayed TR can be crucial to develop a personalized surgical approach. The purpose of this article is to predict TR using MRI radiomic analysis.

Methods: Data were retrospectively collected from a database. A total of 248 patients were included based on the availability of 6-month TR results: 188 were used to train the model, the others to externally validate it. After manual segmentation of the tumor, Radiomic features were extracted and different machine learning models were implemented considering a combination of T1 and T2 weighted MR sequences. Receiver Operating Characteristic (ROC) curve was calculated with relative model performance metrics (accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)) at the best threshold based on the Youden Index.

Results: Models performance were evaluated based on test set results. The best model resulted to be the XGBoost, with an area under ROC curve of 0.72 (95% CI: 0.56 - 0.87). At the best threshold, the model exhibits 0.75 (95% CI: 0.63 - 0.75) as accuracy, 0.62 (95% CI: 0.38 - 0.83) as sensitivity 0.80 (95% CI: 0.66 - 0.89 as specificity, 0.53 (95% CI: 0.31 - 0.73) as PPV, 0.88 (95% CI: 0.72 - 0.94) as NPV.

Conclusion: MRI radiomic analysis represents a powerful tool to predict late HGG recurrence, which can be useful to plan personalized surgical treatments and to offer pertinent patient pre-operative counseling.

KEYWORDS

machine learning, recurrence, high grade glioma, radiomics, prognosis

Introduction

Despite recent advances in surgical technology and genetic discoveries, patients diagnosed with High Grade Gliomas (HGG), which is considered as grade 4, generally have a dismal prognosis with high risk of early tumor recurrence (TR) after post-operative radio-chemotherapy (1–4).

Compelling evidence, based on objective tumoral volume analysis, supports the role of the extent of resection (EOR) in HGG patients as the first step of patients management (5, 6).

Recent technological innovations have increased the safety of surgical resection, while expanding surgical options and indications for HGG surgical treatment (7, 8): several techniques can be currently used during the surgical procedure, such as intraoperative ultrasound (iUS), cortical mapping, sodium fluorescein (9) and 5-ALA fluorescence, with the aim of fostering higher rates of total resection and so increasing the survival chance (10–19).

Surgical treatment, however, can be rarely considered as radical, due to the infiltrating tumor nature, its multifocal presentation, and ill-defined tumor margins.

Although the Stupp protocol was introduced as post-operative standard treatment more than 15 years ago, alternative recent approaches have not been developed so far, and the 5-year survival has not significantly changed in these last decades (20).

Unfortunately, the infiltrative growth, the rapid proliferative rate of malignant cells and the appearance of treatment-resistant cell clones shortly after initial therapy tend to recur within 2 cm of resection margins, independently by the initial EOR exhibited by the patient (21). In this challenging setting, assessment of pre-operative risk of early versus delayed TR assumes a crucial role to develop a personalized surgical approach (with respect to surgery versus biopsy).

A presurgical identification of HGG patients with high risk of recurrence after 6 months from surgery may have several advantages (22): first of all, a more aggressive surgical resection may be pursued in patients with low risk of TR, planning the use of all intra-operative tools and strategies that allow a maximal safe resection. Furthermore, deep genetic sequences may be considered to assist clinicians during postsurgical decision-making involving patients with high risk of TR.

In addition, detection of early TR risk should encourage efforts to better understand the role of early intensified bridging therapies for HGG between surgery and postsurgical treatments (23).

Radiomics is a field of medical imaging that focuses on the extraction and advanced analysis of large amounts of quantitative features from medical images called "radiomic features". The goal of radiomics is to convert medical images into high-dimensional data that can be analyzed to uncover underlying patterns related to disease characteristics, prognosis, and treatment response (24).

The aim of this study is to develop a Machine Learning (ML) model based on radiomics features extracted from MRI images able to stratify the risk of TR (within 6 months) in newly diagnosed HGG and support clinicians in the decision-making process.

Materials and methods

Patient population and image data

This retrospective study was focused on patients affected by high grade glioma, which is defined as grade 4, enrolled in two Institutions: Fondazione Policlinico Universitario Agostino Gemelli IRCCS in Rome (FPG) and Santa Maria della Misericordia in Udine (SMM).

The experimental protocol adopted in this study was approved by the Ethical Committee of Policlinico Universitario Agostino Gemelli IRCCS.

The informed consent was obtained for all the patients included in the study.

Patients from FPG were treated from January 2016 to December 2019, patients from SMM from November 2014 to June 2020. Tumor grading was defined considering the 2021 WHO staging (25).

Inclusion criteria were: age >/= 18 years; no previous surgery; no preoperative chemo- or radiotherapy; at least 6 months of follow up; objective evaluation of preoperative tumor volume on MRI images in DICOM format based on post-contrast T1-weighted MRI sequences and T2-weighted MRI sequences; objective estimation of EOR on post-contrast T1-weighted MRI sequences; revision of histopathological specimens by using the new 2021 WHO Classification of Tumors of the Central Nervous System; MGMT promoter methylation and IDH1/IDH2 mutation status assessment.

All the patients were subjected to the following therapeutic approach: one month before the surgical procedure, all the patients were scanned with an MR acquisition, which was carried out using the GE and Philips scanners 1.5 T in Rome and the Siemens scanner 1.5 T in Udine.

Patients with the following criteria were excluded from the study: diagnostic images blurred or some of required sequences missing.

As for surgery, all the patients underwent the following procedure: the surgical technique was carried out with an intraoperative protocol that involved the use of contrast enhancement ultrasounds (CEUS) and the fluorescence of 5-aminolevulinic acid (5-ALA). Neuro-navigation system was used in all cases, while the intra-operative neurophysiological monitoring was performed in all cases of proximity of the tumor to the cortico-spinal tract. Awake surgery was selected for patients that presented HGG in the dominant hemisphere, close to the inferior front-occipital fasciculus (IFOF) or the superior longitudinal fasciculus (SLF). In a subgroup of selected patient (young age, tumor not close to ventricular system) Carmustine Wafers (CWs) were implanted after surgical tumor removal and intra-operative pathological confirmation of HGG. CWs were not utilized when tumor removal required the creation of a large opening of ventricle and/or the basal cistern.

All methods were performed in accordance with the national guidelines and regulations and in accordance with the Declaration of Helsinki.

Image acquisition

All examinations were performed using a 1.5 T MR scanner (Siemens Aera in Udine and GE—Optima mr450 and Philips-Ingenia in Rome) with an eight-channel head coil. All study protocols included axial T2-weighted TSE/FSE images with a slice thickness between 4 and 5 mm and post-contrast volumetric T1-weighted images MPRAGE/FSPGR/WATS with a slice thickness between 1 and 1.2 mm. Imaging parameters are described in Table 1.

Image pre-processing and radiomic features extraction

Presurgical MRI performed on 1.5T scanners in FPG and SMM of HGG were analyzed by three neuroradiologists who assessed image quality, excluding patients with images degraded by artifacts or who did not present axial 3D T1 weighted post contrast and axial 2D T2 weighted. Manual segmentation of the tumoral areas was performed by one Neuroradiologist with the software "3D Slicer image computing platform" with ROIs drawn separately on T1 and T2W images as follows: on axial 3D T1 weighted contrast images post-contrast, the ROI on the "enhancing" component of the tumor was delineated, while on axial 2D T2 ROIs were outlined on the

TABLE 1 Technical details of MR sequences.

SIEMENS AERA 1,5 T		
Sequence	T1-MPRAGE	T2-TSE
Echo time	2,74 ms	95 ms
NEX	1	2
Repetition time	2200 ms	2380 ms
No. of sections	256	23
Receiver bandwidth	190 Hz/Px	163 Hz/Px
Echo train length	–	28
FOV	250 mm	230mm
Section thickness	1 mm	5 mm
Section spacing	0	1,3 mm
Matrix size	256x232	208x320
Phase direction	RL	RL
PHILIPS Ingenia 1,5T		
Sequence	T1 WATS	T2 TSE
Echo time	6.2ms	100 ms
NEX	1	3
Repetition time	13 ms	3651 ms
No. of sections	155	36
WFS	216.6 Hz/Px	212.3 Hz/Px
TFE factor	195	23
FOV	250mm	240mm
Section thickness	1 mm	4 mm
Section spacing	0	0.4 mm
Matrix size	252x200	513x331
Phase direction	RL	RL
GE Optima 1,5 T		
Sequence	T1 FSPGR	T2 FSE
Echo time	2.2ms	130 ms
NEX	1	2
Repetition time	7.7 ms	3867 ms
No. of sections	150	34
Receiver bandwidth	22.73 Hz/Px	31.25 Hz/Px
Echo train length	–	21
FOV	250mm	240mm
Section thickness	1.2 mm	4 mm
Section spacing	0	0.4 mm
Matrix size	288x288	356x288
Phase direction	RL	RL

solid component of the tumor and the infiltrative one, excluding the frankly edematous areas (with higher signal in T2) (Figure 1). Before any analysis, the program anonymizes any DICOM.

Image pre-processing and radiomic analyses were performed using Python 3.7.

Image pre-processing was performed via N4 bias field correction to correct low frequency intensity non uniformity and subject-specific Z-score normalization to normalize MR intensity scales and allow inter-patients comparison.

For each patient, a total of 100 radiomic features (26) were extracted from the specific ROI defined in each pre-processed MR sequence. Features were extracted in their 3D version using PyRadiomics (27).

Extracted features belonged to three families: morphology-based (14), first order (18) and second order (68). For the second order features we used a bin width discretization strategy with a bin width value of 25. Morphology-based features capture geometric characteristics of the ROI. First order features describe statistical properties of the grey level histogram, offering global metrics for the distribution of the grey levels within the ROI. Second order features provide localized measures of grey level distribution within the ROI.

Radiomic features selection and modeling

The training set included the patients enrolled in Rome, while the test set included the patients enrolled in Udine. We decided to follow this approach to develop a prediction model using data from one hospital and test its generalizability on data from a different hospital, for external testing to pursue a TRIPOD 3 approach (36).

Each radiomic feature was normalized using the z-score in the training set, and the z-score parameters of the training set were used to normalize the features in the test set.

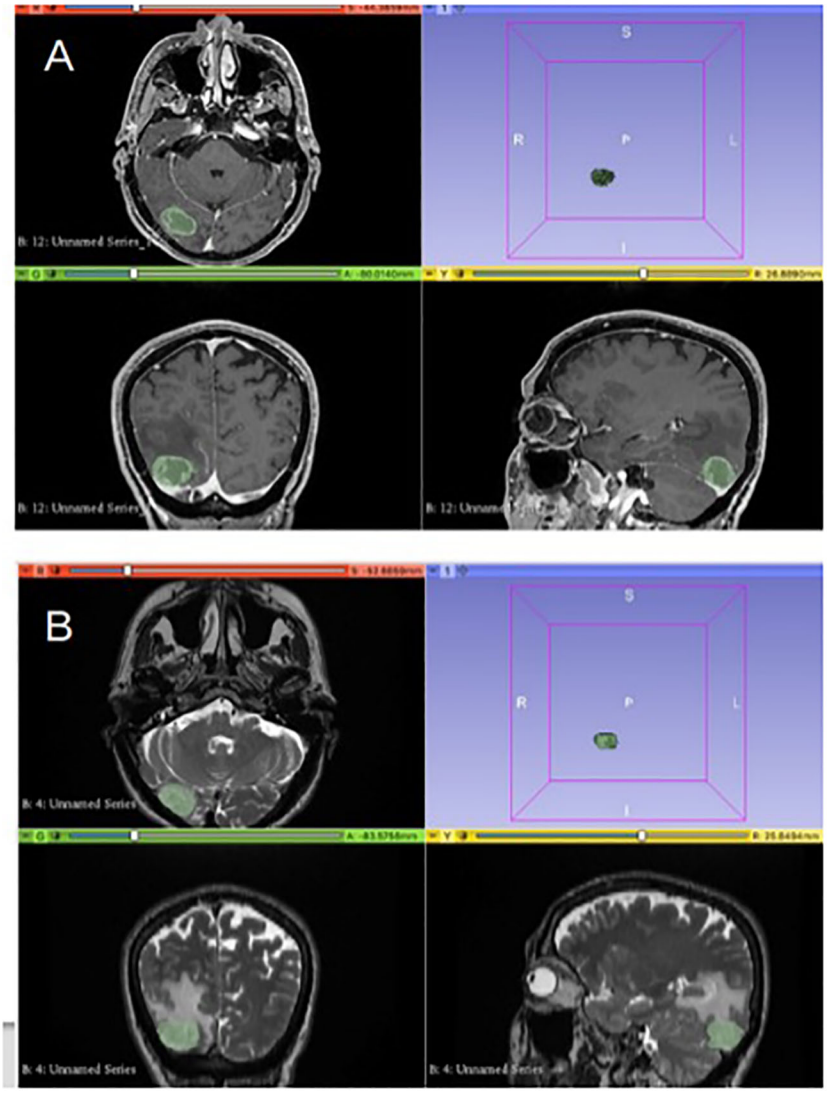


FIGURE 1
Segmentation of the tumoral areas: in image **(A)**, delineation of the ROI on 3D T1 W post contrast images on the "enhancing" component of the tumor. In image **(B)**, the ROI on axial 2D T2w images is delineated for the solid and infiltrative components of the tumor.

After features extraction, feature selection and radiomics modeling was performed on the training set only. The reproducibility of radiomic features was evaluated with respect to the MRI scanner manufacturer using the t-test. Features that resulted not reproducible (p-value < 0.05) were excluded from the following analysis.

Further feature selection methods were applied to prevent overfitting, reduce linear correlations among variables and obtain a minimal and stable set of radiomic features. These methods included the univariate analysis with the Wilcoxon-Mann-Whitney statistical test (WMW, significance level of 0.05) and the computation of the Pearson correlation coefficient (PCC) with the exclusion of features that were highly correlated with any other feature (PCC > 0.9).

Different machine learning models, namely regularized logistic regression, random forest, XGBoost and Support Vector Machine (SVM) were trained on the selected radiomic features extracted from axial 3D T1w and axial 2D T2w.

The predictive models aimed to address a binary classification problem for the prediction of the 6-months progression free survival: class 0 denoted patients without 6-months TR, class 1 denoted patients with 6-months TR.

For each machine learning model, hyper-parameters tuning was performed with a grid search strategy using 3-fold cross-validation (28).

The developed models were externally evaluated on the testing set in terms of discriminative ability and predictive performance. In particular, the area under the curve (AUC) of the receiver operating characteristics (ROC) curve was calculated, and the model performance metrics based on the classification matrix were computed at the best cut-off threshold, identified by maximizing the Youden Index calculated on the training set. The metrics investigated were accurancy, sensitvity, specificity, positive and negative predictive values (PPV and NPV). 95% confidence intervals (CI) of AUC and classification metrics were computed according to bootstrap (29) and Jeffreys (30) methods, respectively.

Results

Patients population

When performing radiomic analysis and modeling, the initial dataset consisting in 273 patients affected by HGG of first diagnosis and undergone respective surgery followed by Stupp protocol, was shrunk to 248 patients on the basis of the availability of the 6-month TR outcome.

Data were split into train (approx. 75%) and test (approx. 25%) sets resulting in 188 patients from Rome and 60 patients from Udine, respectively.

Demographic and clinicopathological characteristics of the included patients are reported in Table 2.

In the training set 109 patients were classified as class 0 and 79 patients as class 1. In the test set, 44 patients belonged to class 0 and 16 to class 1 (Figure 2).

Radiomic features selection

Features selection was applied to features extracted from T1w and T2w sequences.

As for T1w sequence, the t-test evaluating features stability with respect to the manufacturer shrank the number of features from 100 to 58, further reduced by the WMW test and PCC which led to four stable relevant features, reported in Table 3 with their corresponding WMW test p-values.

As for T2w sequence, 85 features out of 100 showed stability with respect to the manufacturer but only one of them was significant at univariate analysis, as reported in Table 3.

Thus, Table 3 included all the features used for the radiomics modeling.

Boxplots of the selected features showing stability with respect to the outcome are shown in Figure 3, with corresponding p-values resulting from the WMW test.

Modeling

A combined T1w&T2w (T1wT2w) modeling method was implemented, consisting in grouping the T1wT2w relevant features in a unique input dataset (see Table 3).

The cross-correlation matrix of the T1wT2w significant features is reported in Figure 4.

Table 4 reports the discriminative and predictive performance metrics for all the implemented ML models trained on T1w and T2w for both training and validation sets.

Based on the metrics evaluated on the test set, the best model resulted to be the XGBoost, with a test set AUC of 0.72 and 95% CI of 0.56 - 0.87, and an accuracy of 0.75 with a 95% CI of 0.63 - 0.85. This model obtained high values for the specificity equal to 0.80 (95% CI: 0.66 - 0.89), and NPV equal to 0.85 (95% CI: 0.72 – 0.94).

The other models showed the problem of model over-fitting to the training data (i.e. Random Forest) or poor model performance for the training and test data (i.e. regularized logistic regression and SVM).

Discussion

Machine learning approach

Today MRI has a primary role in diagnosis, planning and monitoring of HGG patients: clinicians typically use brain MRI to evaluate radiological HGG features such as size, location, edema and enhancement characteristics. MRI features are today not sufficient to predict the risk of recurrence in HGG (31, 32):for this reason, there is a huge need to assess additional imaging biomarkers via computational methods (33, 34).

In the last years, a radiology-based approach focusing on prognosis prediction has gained an important burst fostered by the fast development of advanced computational tools able to manage a significant amount of MRI and clinical data.

TABLE 2 Baseline characteristics of the study population.

Parameters	Trainig set (Rome)	Test set (Udine)	Total
Initial patients	206	67	273
No FU	18	7	25
Patients included	**188**	**60**	**248**
Mean age (years)	62.84 ± 4.86	62.52 ± 5.32	62.76 ± 4.98
Sex			
Male	119 (63.3%)	36 (60%)	155 (62.5%)
Female	69 (36.7%)	24 (40%)	93 (37.5%)
Side			
Left	77 (41%)	28 (46.7%)	105 (42.3%)
Right	111 (59%)	32 (53.3%)	143 (57.7%)
Tumor site			
Precentral	63 (33.5%)	18 (30%)	81 (32.7%)
Postcentral	50 (26.6%)	16 (26.7%)	66 (26.6%)
Temporal/insular	75 (39.8%)	26 (43.3%)	101 (40.7%)
5-ALA			
yes	126 (67%)	47 (78.3%)	173 (69.8%)
no	62 (32%)	13 (21.7%)	75 (31.2%)
Biological features			
MGMT methylation (yes vs no)	124 vs 64 (66% vs 34%)	40 vs 20 (66.7% vs 33.3%)	164 vs 84 (66.1% vs 33.9%)
IDH 1/2 mutation (yes vs no)	8 vs 180 (4.3% vs 95.7%)	3 vs 57 (5% vs 95%)	11 vs 237 (4.4% vs 95.6%)
Ki-67 (mean)	24.25 (3-90)	41,2 (5-75)	29 (4-80)
OS			
Alive (yes vs no)	53 vs 135 (28.2% vs 71.8%)	6 vs 54 (10% vs 90%)	59 vs 189 (23.8% vs 76.2%)
Average of FU times (months)	13.4 (0-35)	17.1 (0-29)	14.3 (0-31)
PFS			
Class 0 (no 6-months TR)	**109** (58%)	**44** (73.3%)	153 (61.7%)
Class 1 (6-months TR)	**79** (42%)	**16** (26.7%)	95 (38.3%)

Characteristics of the study population are described using means ± s.d. (standard deviation) or median and range for continuous variables, number of cases with relative percentages reported in parentheses for categorical variables. Bold values are the corresponding value to Figure 2.
(FU, follow-up; 5-ALA, 5-aminolevulinic acid; MGMT, O-6-Methylguanine-DNA Methyltransferase; IDH, Isocitrate dehydrogenase 1; OS, overall survival; PFS, progression free survival; TR, tumor recurrence).

Radiomics has recently emerged as a powerful data-driven approach that can offer insights into clinically relevant questions related to diagnosis, prediction, prognosis, as well as treatment response assessment (35).

In this investigation radiomic analysis and modeling were performed with the aim to select the most significant and robust MRI features able to predict which patients affected by HGG at first diagnosis would develop progression within 6 months or later. Prediction performance resulting from training was also externally tested following a TRIPOD 3 study approach, according to the original TRIPOD guidelines (36). Clustering was not accounted for in this study limiting model predictive performance and generalizability (48, 49). Future work may be conducted to evaluate the presence of potential cluster effects.

The 6-months TR time-point was chosen as cut-off value to discriminate patient with early TR from those with later TR for two main clinical reasons: it matches the standard timing of post-operative radiological assessment after conventional postsurgical treatments and it makes homogenous the study populations before the non-standardized rescue therapies at tumor recurrence.

Thus, detecting patients with low risk of TR after 6 months from surgery encourage neurosurgeons to extent tumor resection with the aim of exploiting the most modern intraoperative tools and strategies that allow a maximal safe resection.

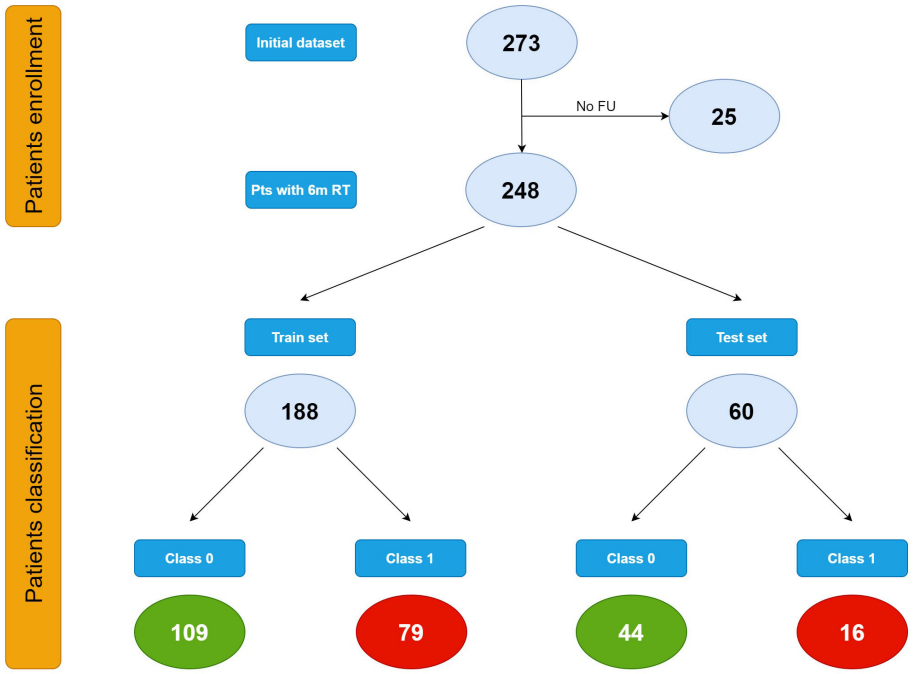


FIGURE 2
Patients classification.

We found that, among all features' families, morphological features show most of the predictive power and that smaller features values decrease the probability to observe 6-month tumor recurrence in patients.

The best model resulted to be the XGBoost (eXtreme Gradient Boosting), an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements ML algorithms under the Gradient Boosting framework, providing distributed gradient-boosted decision tree (GBDT). Boosting is an ensemble learning method that combines a set of weak base estimators into a strong learner to minimize training errors. For these reasons, XGBoost is considered to be one of the leading ML libraries for regression, classification and ranking problems.

Modeling results are reported in Table 3: the developed XGBoost showed a high value of specificity, which reflects the model ability in identifying patients with lower risk to experience tumor recurrence within 6 months and that might undergo a more aggressive surgical resection.

High values of specificity were observed in training (84%) and test set (80%). Features selection applied to this study highlighted that morphological features hold most of the predictive power in discriminating patients with positive and negative outcome. Although it is hard to find a direct biological interpretation of these findings, it is worth mentioning that radiomic features are not relevant if considered only "per se". It is actually important to relate them to the model context and take linear and non-linear interactions between variables into account. In this light, a possible explanation can be given assuming that changes in morphological features might reflect tissues structural alterations (e.g. shape, volume etc…) and be more related to tumor developments. On the other hand, it is also important to ensure that selected features describing the model are actually stable, non-redundant and independent from noise or other non-relevant variables.

TABLE 3 Relevant features resulting from features selection performed on T1w and T2w sequences and corresponding p-values and used for radiomics modeling.

MR sequence	Feature	WMW test p-value
T1w	original_shape_MajorAxisLength	0.02
T1w	original_shape_Maximum2DDiameterColumn	0.03
T1w	original_shape_Maximum2DDiameterSlice	0.04
T1w	original_firstorder_TotalEnergy	0.04
T2w	Original_first_order_kurtosis	0.03

These features belong to morphology-based ("shape") and first order families.

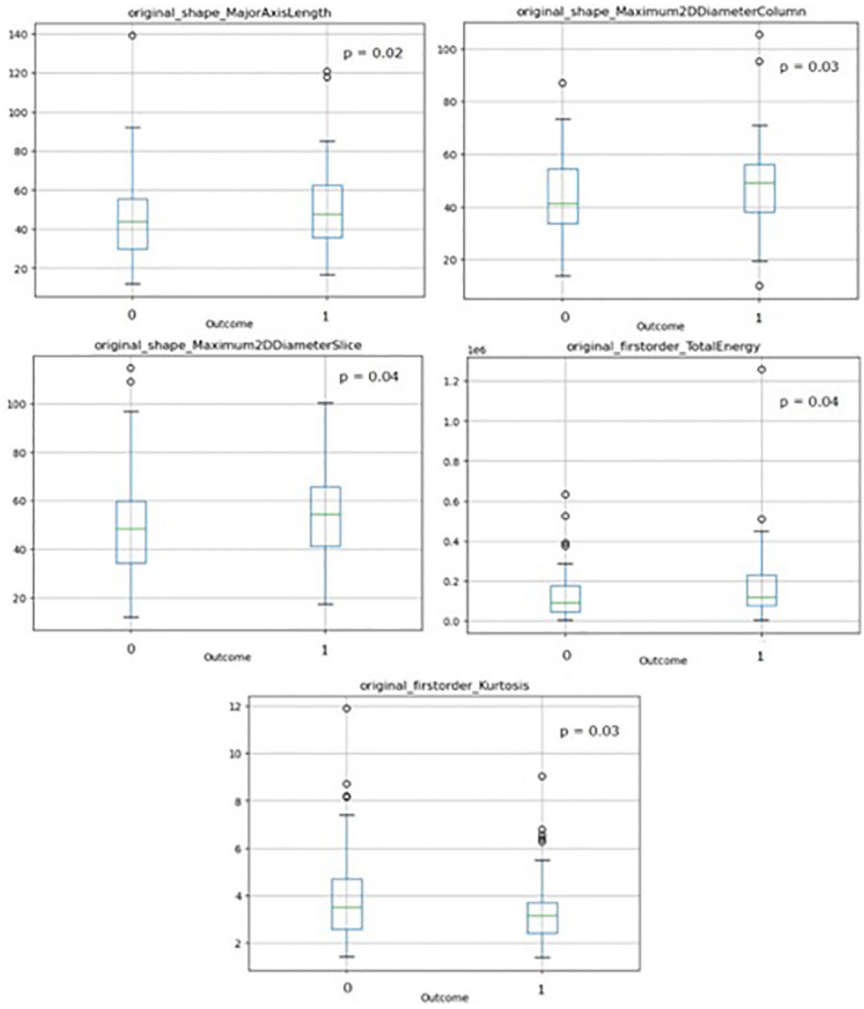


FIGURE 3
Boxplots of the selected radiomics features used for radiomics modeling showing stability with respect to the outcome with corresponding p-values resulting from the WMW test. ('1'=patients with 6-months TR, '0'=patients without 6-months TR).

Pre-operative estimation of HGG biological behavior could help clinicians in detecting cases that could benefit from a maximal safe resection (e.g. HGG patients with pre-operative estimated of low risk of early TR).

On the other hand, pre-operative estimation of high risk of early TR (especially in elderly patients) could drive the choice to biopsy rather than surgery (37).

In addition, for patients with higher risk of TR, multiple tissue samplings should be extracted in order to investigate *ad hoc* target therapies related to lesions' high spatial heterogeneity (38).

Our results are aligned with the study performed by Li et al. (39) with AUC of 0.70 in the training set for the prediction of disease progression at 6 months that used radiomics features extracted from multiple MRI sequences (T2 and FLAIR).

Other studies using smaller sample sizes developed MRI-based radiomics models for the prediction of the progression-free survival in patients with glioblastoma obtaining similar results. Choi et al. (40) obtained an integrated time-dependent AUC of 0.62, while Bathla et al. (41) achieved a C-index up to 0.64.

This investigation presents several limitations. First of all, this was a retrospective study including only two centers: further studies with more heterogeneous datasets would be an interesting point for future development.

Radiomic features reproducibility can be strongly affected by image acquisition parameters and scan protocols, which can vary widely across and within institutions. We took this heterogeneity into account and in order to evaluate features robustness among different centers, we performed a t-test analysis with respect to MRI scanners manufacturers. This allowed to exclude unstable features. In this light, to minimize data variability, harmonization methods might be introduced, as proposed by several research groups which focused on different modified ComBat algorithms (42–46). Although being very powerful, radiomic analysis offers no insights toward biological interpretation of the achieved findings and this study also shows this limitation: many efforts have been made in these recent years to reintroduce biological meaning into radiomics. However, some recent studies also suggest that biological correlation with radiomic features is not mandatory (47).

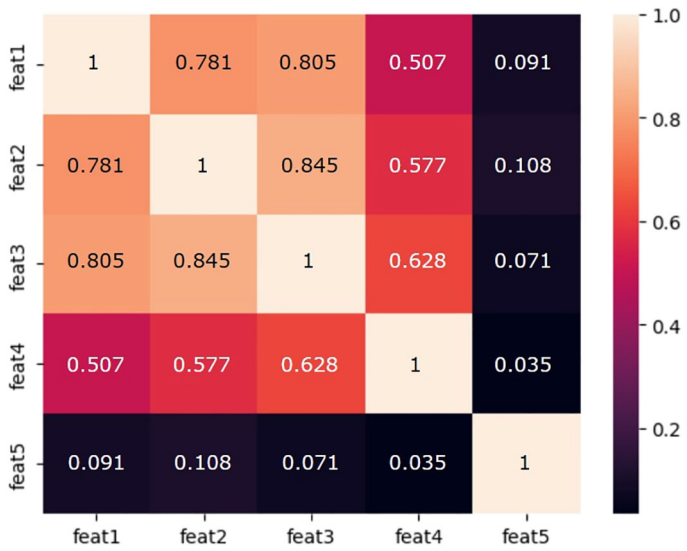


FIGURE 4
Cross-correlation matrix of the T1wT2w significant features used for radiomics modeling. feat1: original_shape_MajorAxisLength; feat2: original_shape_Maximum2DDiameterColumn; feat3: original_shape_Maximum2DDiameterSlice; feat4: original_firstorder_TotalEnergy; feat5: Original_first_order_kurtosis.

In addition, considering the integration of features and clinical variables in a clinical-radiomic model could help improving models' performance in predicting 6-months TR.

Lastly, we did not assess the overall survival rate for this cohort. Data regarding selection criteria adopted at TR to plan the salvage treatment were not available. Each patient underwent an individualized management at TR, so we have not developed standardized protocols for treatments at TR. With our sample size, we did not have the statistical power to tease out the survival rate for patients undergoing different treatments. A future multi-center prospective study of HGG recurrence will be necessary to properly assess survival rate according to the salvage treatment adopted at TR.

The high incidence of early TR should encourage efforts to better understand the role of early intensified bridging therapies for HGG between surgery and postsurgical treatments.

In conclusion, the methodology adopted in this investigation is extremely time-consuming and makes it unsuitable for clinical daily implementation. The next step in this field, beyond increasing accuracy and simplifying the workflow, will be the development of an open source, easily scalable and efficient artificial intelligence algorithm requiring simple or null external intervention from physicians.

TABLE 4 Predictive performance metrics for ML models trained on T1w and T2w for training and validation sets.

		AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
XGBoost	Training 95% CI	0.68 (0.60- 0.75)	0.66 (0.60 - 0.73)	0.42 (0.31 - 0.53)	0.84 (0.77 - 0.90)	0.66 (0.52- 0.78)	0.67 (0.59 - 0.74)
	Test 95% CI	0.72 (0.56 - 0.87)	0.75 (0.63 - 0.85)	0.62 (0.38 - 0.83)	0.80 (0.66 - 0.89)	0.53 (0.31 - 0.73)	0.85 0.72 - 0.94)
Regularized Logistic Regression	Training 95% CI	0.63 (0.55 - 0.70)	0.60 (0.54 - 0.67)	0.57 (0.46 - 0.67)	0.63 (0.54 - 0.72)	0.53 (0.42 - 0.63)	0.67 (0.58 - 0.76)
	Test 95% CI	0.52 (0.36 - 0.68)	0.41 (0.30 - 0.54)	0.56 (0.33 - 0.78)	0.36 (0.23 - 0.51)	0.24 (0.13 - 0.40)	0.69 (0.49 - 0.85)
Random Forest	Training 95% CI	0.99 (0.99 - 1.0)	0.97 (0.95 - 0.99)	0.98 (0.94 - 1.0)	0.97 (0.93 - 0.99)	0.96 (0.90 - 0.99)	0.99 (0.96 - 1.0)
	Test 95% CI	0.51 (0.33 - 0.70)	0.38 (0.27 - 0.51)	0.68 (0.44 - 0.87)	0.27 (0.16 - 0.42)	0.25 (0.14 - 0.40)	0.70 (0.47 - 0.88)
SVM	Training 95% CI	0.54 (0.46 - 0.63)	0.60 (0.54 - 0.67)	0.21 (0.14 - 0.32)	0.90 (0.82 - 0.94)	0.59 (0.41 - 0.75)	0.61 (0.53 - 0.68)
	Test 95% CI	0.61 (0.45 - 0.76)	0.68 (0.56 - 0.79)	0.25 (0.09 - 0.49)	0.84 (0.71 - 0.93)	0.36 (0.14 - 0.65)	0.75 (0.62 - 0.86)

Conclusions

A thorough and reliable ML-model based on combined T1w&T2w sequences to detect the lower risk of TR in newly diagnosed HGG was trained and validated on external cohort.

Our results confirm the potential role of pre-operative MRI analysis beyond the classical anatomical and morphological parameters. MRI radiomic analysis represents a powerful tool to predict early HGG recurrence, to plan personalized surgical treatment and to offer patients pre-operative counseling. In the future, a prospective multicenter study with a larger sample size is needed in order to validate our results, to optimize prediction models for clinical practice, and to overcome the intrinsic limitations of retrospective studies met so far.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ethical Committee of Policlinico Universitario Agostino Gemelli IRCCS. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FP: Writing – original draft, Writing – review & editing. TI: Writing – original draft, Writing – review & editing. RR: Writing – original draft, Writing – review & editing. DB: Writing – original draft, Writing – review & editing. FB: Writing – original draft, Writing – review & editing. EB: Writing – original draft, Writing – review & editing. LB: Writing – original draft, Writing – review & editing. SC: Writing – original draft, Writing – review & editing. CC: Writing – original draft, Writing – review & editing. DC: Writing – original draft, Writing – review & editing. CG: Writing – original draft, Writing – review & editing. GL: Writing – original draft, Writing – review & editing. CM: Writing – original draft, Writing – review & editing. EM: Writing – original draft, Writing – review & editing. AO: Writing – original draft, Writing – review & editing. MS: Writing – original draft, Writing – review & editing. HT: Writing – original draft, Writing – review & editing. GV: Writing – original draft, Writing – review & editing. SG: Writing – original draft, Writing – review & editing. GS: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Ius T, Pignotti F, Della Pepa GM, La Rocca G, Somma T, Isola M, et al. A novel comprehensive clinical stratification model to refine prognosis of glioblastoma patients undergoing surgical resection. *Cancers (Basel)*. (2020) 12. doi: 10.3390/cancers12020386

2. Lemee JM, Clavreul A, Menei P. Intratumoral heterogeneity in glioblastoma: don't forget the peritumoral brain zone. *Neuro Oncol*. (2015) 17:1322–32. doi: 10.1093/neuonc/nov119

3. Ohgaki H. Epidemiology of brain tumors. In: Verma M, editor. *Cancer Epidemiology*. Humana Press, Totowa, NJ (2009). p. 323–42.

4. Weller M, van den Bent M, Preusser M, Le Rhun E, Tonn JC, Minniti G, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol*. (2021) 18:170–86. doi: 10.1038/s41571-020-00447-z

5. Karschnia P, Vogelbaum MA, van den Bent M, Cahill DP, Bello L, Narita Y, et al. Evidence-based recommendations on categories for extent of resection in diffuse glioma. *Eur J Cancer*. (2021) 149:23–33. doi: 10.1016/j.ejca.2021.03.002

6. Wang L, Liang B, Li YI, Liu X, Huang J, Li YM. What is the advance of extent of resection in glioblastoma surgical treatment-a systematic review. *Chin neurosurgical J*. (2019) 5:2. doi: 10.1186/s41016-018-0150-7

7. Delgado-Lopez PD, Corrales-Garcia EM. Survival in glioblastoma: a review on the impact of treatment modalities. *Clin Trans Oncol*. (2016) 18:1062–71. doi: 10.1007/s12094-016-1497-x

8. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J Neurosurg*. (2001) 95:190–8. doi: 10.3171/jns.2001.95.2.0190

9. Nikova AS, Vlotinou P, Karelis L, Karanikas M, Birbilis TA. Gross total resection with fluorescence could lead to improved overall survival rates: a systematic review and meta-analysis. *Br J Neurosurg*. (2022) 36:316–22. doi: 10.1080/02688697.2021.1950637

10. Diez Valle R, Tejada Solis S, Idoate Gastearena MA, Garcia de Eulate R, Dominguez Echavarri P, Aristu Mendiroz J. Surgery guided by 5-aminolevulinic fluorescence in glioblastoma: volumetric analysis of extent of resection in single-center experience. *J Neurooncol*. (2011) 102:105–13. doi: 10.1007/s11060-010-0296-4

11. Eatz TA, Eichberg DG, Lu VM, Di L, Komotar RJ, Ivan ME. Intraoperative 5-ALA fluorescence-guided resection of high-grade glioma leads to greater extent of resection with better outcomes: a systematic review. *J Neurooncol*. (2022) 156:233–56. doi: 10.1007/s11060-021-03901-9

12. Eljamel S. 5-ALA fluorescence image guided resection of glioblastoma multiforme: A meta-analysis of the literature. *Int J Mol Sci*. (2015) 16:10443–56. doi: 10.3390/ijms160510443

13. Gandhi S, Tayebi Meybodi A, Belykh E, Cavallo C, Zhao X, Syed MP, et al. Survival outcomes among patients with high-grade glioma treated with 5-aminolevulinic acid-guided surgery: A systematic review and meta-analysis. *Front Oncol*. (2019) 9:620. doi: 10.3389/fonc.2019.00620

14. La Rocca G, Della Pepa GM, Menna G, Altieri R, Ius T, Rapisarda A, et al. State of the art of fluorescence guided techniques in neurosurgery. *J neurosurgical Sci*. (2019) 63:619–24. doi: 10.23736/S0390-5616.19.04854-9

15. Mazurek M, Kulesza B, Stoma F, Osuchowski J, Mandziuk S, Rola R. Characteristics of fluorescent intraoperative dyes helpful in gross total resection of high-grade gliomas-A systematic review. *Diagnostics*. (2020) 10. doi: 10.3390/diagnostics10121100

16. Pichlmeier U, Bink A, Schackert G, Stummer W, Group ALAGS. Resection and survival in glioblastoma multiforme: an RTOG recursive partitioning analysis of ALA study patients. *Neuro Oncol*. (2008) 10:1025–34. doi: 10.1215/15228517-2008-052

17. Stummer W, Reulen HJ, Meinel T, Pichlmeier U, Schumacher W, Tonn JC, et al. Extent of resection and survival in glioblastoma multiforme: identification of and adjustment for bias. *Neurosurgery*. (2008) 62:564–576; discussion 564-576. doi: 10.1227/01.neu.0000317304.31579.17

18. Della Pepa GM, Ius T, Menna G, La Rocca G, Battistella C, Rapisarda A, et al. Dark corridors" in 5-ALA resection of high-grade gliomas: combining fluorescence-guided surgery and contrast-enhanced ultrasonography to better explore the surgical field. *J neurosurgical Sci*. (2019) 63:688–96. doi: 10.23736/S0390-5616.19.04862-8

19. La Rocca G, Sabatino G, Menna G, Altieri R, Ius T, Marchese E, et al. 5-aminolevulinic acid false positives in cerebral neuro-oncology: not all that is fluorescent is tumor. A case-based update and literature review. *World Neurosurg*. (2020) 137:187–93. doi: 10.1016/j.wneu.2020.01.238

20. Poon MTC, Sudlow CLM, Figueroa JD, Brennan PM. Longer-term (>/= 2 years) survival in patients with glioblastoma in population-based studies pre- and post-2005: a systematic review and meta-analysis. *Sci Rep*. (2020) 10:11622. doi: 10.1038/s41598-020-68011-4

21. Delgado-Lopez PD, Rinones-Mena E, Corrales-Garcia EM. Treatment-related changes in glioblastoma: a review on the controversies in response assessment criteria and the concepts of true progression, pseudoprogression, pseudoresponse and radionecrosis. *Clin Trans Oncol*. (2018) 20:939–53. doi: 10.1007/s12094-017-1816-x

22. Zhou Q, Ke X, Xue C, Li S, Huang X, Zhang B, et al. A nomogram for predicting early recurrence in patients with high-grade gliomas. *World Neurosurg*. (2022) 164: e619–28. doi: 10.1016/j.wneu.2022.05.039

23. Ius T, Sabatino G, Panciani PP, Fontanella MM, Ruda R, Castellano A, et al. Surgical management of Glioma Grade 4: technical update from the neuro-oncology section of the Italian Society of Neurosurgery (SINch(R)): a systematic review. *J Neurooncol*. (2023) 162:267–93. doi: 10.1007/s11060-023-04274-x

24. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

25. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol*. (2021) 23:1231–51. doi: 10.1093/neuonc/noab106

26. Z A. EP-1677: multicentre initiative for standardisation of image biomarkers. *Radiotherapy Oncol*. (2017) 123:S914–915. doi: 10.3252/pso.eu.ESTRO36.2017

27. Available online at: https://www.radiomics.io/pyradiomics.html. (accessed September 15, 2022).

28. B D. Cross-validation. In: *Encyclopedia of Bioinformatics and Computational Biology*. Amsterdam. Netherlands: Elsevier Science (2019). p. 542–5.

29. DiCiccio TJE B. Bootstrap confidence intervals. *Statist Sci*. (1996) 11:189–228. doi: 10.1214/ss/1032280214

30. Brown LDC T, Dasgupta A. Interval estimation for a binomial proportion. *Stat Sci*. (2001) 16:101–33. doi: 10.1214/ss/1009213286

31. Barboriak DP, Zhang Z, Desai P, Snyder BS, Safriel Y, McKinstry RC, et al. Interreader variability of dynamic contrast-enhanced MRI of recurrent glioblastoma: the multicenter ACRIN 6677/RTOG 0625 study. *Radiology*. (2019) 290:467–76. doi: 10.1148/radiol.2019181296

32. Lundemann M, Munck Af Rosenschold P, Muhic A, et al. Feasibility of multi-parametric PET and MRI for prediction of tumour recurrence in patients with glioblastoma. *Eur J Nucl Med Mol Imaging*. (2019) 46:603–13. doi: 10.1007/s00259-018-4180-3

33. Kim JY, Park JE, Jo Y, Shim WH, Nam SJ, Kim JH, et al. Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. *Neuro Oncol*. (2019) 21:404–14. doi: 10.1093/neuonc/noy133

34. Tien AC, Li J, Bao X, Derogatis A, Kim S, Mehta S, et al. A phase 0 trial of ribociclib in recurrent glioblastoma patients incorporating a tumor pharmacodynamic- and pharmacokinetic-guided expansion cohort. *Clin Cancer Res*. (2019) 25:5777–86. doi: 10.1158/1078-0432.CCR-19-0133

35. Beig N, Bera K, Tiwari P. Introduction to radiomics and radiogenomics in neuro-oncology: implications and challenges. *Neuro-oncology Adv*. (2020) 2:iv3–iv14. doi: 10.1093/noajnl/vdaa148

36. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj*. (2015) 350:g7594. doi: 10.1161/CIRCULATIONAHA.114.014508

37. Ius T, Somma T, Altieri R, Angileri FF, Barbagallo GM, Cappabianca P, et al. Is age an additional factor in the treatment of elderly patients with glioblastoma? A new stratification model: an Italian Multicenter Study. *Neurosurgical Focus*. (2020) 49:E13. doi: 10.3171/2020.7.FOCUS20420

38. Menna G, Manini I, Cesselli D, Dai G, Duan L, Luo Y, et al. Immunoregulatory effects of glioma-associated stem cells on the glioblastoma peritumoral microenvironment: a differential PD-L1 expression from core to periphery? *Neurosurgical Focus*. (2022) 52:E4. doi: 10.3171/2021.11.FOCUS21589

39. Li Z, Chen L, Song Y, et al. Predictive value of magnetic resonance imaging radiomics-based machine learning for disease progression in patients with high-grade glioma. *Quantitative Imaging Med Surg*. (2023) 13:224–36. doi: 10.21037/qims-22-459

40. Choi Y, Nam Y, Jang J, Shin NY, Lee YS, Ahn KJ, et al. Radiomics may increase the prognostic value for survival in glioblastoma patients when combined with conventional clinical and genetic prognostic models. *Eur Radiol*. (2021) 31:2084–93. doi: 10.1007/s00330-020-07335-1

41. Bathla G, Soni N, Ward C, Pillenahalli Maheshwarappa R, Agarwal A, Priya S. Clinical and magnetic resonance imaging radiomics-based survival prediction in glioblastoma using multiparametric magnetic resonance imaging. *J Comput assisted tomography*. (2023) 47:919–23. doi: 10.1097/RCT.0000000000001493

42. Da-Ano R, Masson I, Lucia F, Dore M, Robin P, Alfieri J, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci Rep*. (2020) 10:10248. doi: 10.1038/s41598-020-66110-w

43. Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*. (2018) 167:104–20. doi: 10.1016/j.neuroimage.2017.11.024

44. Fortin JP, Parker D, Tunc B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage*. (2017) 161:149–70. doi: 10.1016/j.neuroimage.2017.08.047

45. Mahon RN, Ghita M, Hugo GD, Weiss E. ComBat harmonization for radiomic features in independent phantom and lung cancer patient computed tomography datasets. *Phys Med Biol*. (2020) 65:015010. doi: 10.1088/1361-6560/ab6177

46. Masson I, Da-Ano R, Lucia F, Dore M, Castelli J, Goislard de Monsabert C, et al. Statistical harmonization can improve the development of a multicenter CT-based radiomic model predictive of nonresponse to induction chemotherapy in laryngeal cancers. *Med Phys*. (2021) 48:4099–109. doi: 10.1002/mp.14948

47. Fournier L, Costaridou L, Bidaut L, Michoux N, Lecouvet FE, de Geus-Oei LF, et al. Incorporating radiomics into clinical trials: expert consensus endorsed by the European Society of Radiology on considerations for data-driven compared to biologically driven quantitative biomarkers. *Eur Radiol*. (2021) 31:6001–12. doi: 10.1007/s00330-020-07598-8

48. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD +AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. (2024) 16:385. doi: 10.1136/bmj-2023-078378

49. Debray TPA, Collins GS, Riley RD, Snell KIE, Van Calster B, Reitsma JB, et al. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. *BMJ*. (2023) 7:380. doi: 10.1136/bmj-2022-071018

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 December 2024
DOI 10.3389/fonc.2024.1449982

OPEN ACCESS

EDITED BY
Domenico Aquino,
IRCCS Carlo Besta Neurological Institute Foundation, Italy

REVIEWED BY
Shihong Li,
Fudan University, China
Flavio Giordano,
Hospital Antonio Cardarelli, Italy

*CORRESPONDENCE
Alberto Negro
Alberto.Negro@hotmail.it

RECEIVED 16 June 2024
ACCEPTED 18 November 2024
PUBLISHED 23 December 2024

CITATION
Negro A, Gemini L, Tortora M, Pace G, Iaccarino R, Marchese M, Elefante A, Tortora F, D'Agostino V and members of ODM Multidisciplinary Neuro-Oncology Group (2024) VASARI 2.0: a new updated MRI VASARI lexicon to predict grading and *IDH* status in brain glioma. *Front. Oncol.* 14:1449982. doi: 10.3389/fonc.2024.1449982

VASARI 2.0: a new updated MRI VASARI lexicon to predict grading and *IDH* status in brain glioma

Alberto Negro[1]*, Laura Gemini[1], Mario Tortora[2], Gianvito Pace[1], Raffaele Iaccarino[1], Mario Marchese[3], Andrea Elefante[2], Fabio Tortora[2], Vincenzo D'Agostino[1] and members of ODM Multidisciplinary Neuro-Oncology Group

[1]NeuroRadiology Unit, Ospedale del Mare, Azienda Sanitaria Locale Napoli 1 Centro (ASL NA1 Centro), Naples, Italy, [2]Department of Advanced Biomedical Sciences, University of Naples Federico II, Naples, Italy, [3]Department of Health Medicine and Science "Vincenzo Tiberio", University of Molise, Campobasso, Italy

Introduction: Precision medicine refers to managing brain tumors according to each patient's unique characteristics when it was realized that patients with the same type of tumor differ greatly in terms of survival, responsiveness to treatment, and toxicity of medication. Precision diagnostics can now be advanced through the establishment of imaging biomarkers, which necessitates quantitative image acquisition and processing. The VASARI (Visually AcceSAble Rembrandt Images) manual annotation methodology is an ideal and suitable way to determine the accurate association between genotype and imaging phenotype. Our work proposes an updated version of the VASARI score that is derived by changing the evaluation ranges of its components in an effort to increase the diagnostic accuracy of the VASARI manual annotation system and to find neuroimaging biomarkers in neuro-oncology with increasing reliability.

Materials and methods: We gathered the histological grade and molecular status of 126 patients with glioma (Men/Women = 75/51; mean age: 55.30) by a retrospective analysis. Two residents and three neuroradiologists blindedly examined each patient using all 25 VASARI characteristics, after having appropriately modified the reference ranges in order to implement an innovative VASARI lexicon (VASARI 2.0). It was determined how well the observers agreed. A box plot and a bar plot were used in a statistical analysis to assess the distribution of the observations. After that, we ran a Wald test and univariate and multivariate logistic regressions. To find cutoff values that are predictive of a diagnosis, we also computed the odds ratios, confidence intervals, and evaluation matrices using receiver operating characteristic curves for each variable. Finally, we performed a Pearson correlation test to evaluate whether the variable grades and *IDH* were correlated.

Results: An excellent Intraclass Correlation Coefficient (ICC) estimate was obtained. In this study, five features were part of the predictive model for determining glioma grade: F4, enhancement quality [area under the curve (AUC): 0.87]; F5, tumor-enhancing proportion (AUC: 0.70); F6, tumor–non-enhancing proportion (AUC: 0.89); F7, necrosis proportion (AUC: 0.79); and

F17, diffusion characteristics (AUC: 0.75). Furthermore, six features were found to predict *IDH* mutation status: F4, enhancement quality (AUC: 0.904); F5, tumor-enhancing proportion (AUC: 0.73); F6, tumor–non-enhancing proportion (AUC: 0.91); F7, necrosis proportion (AUC: 0.84); F14, proportion of edema (AUC: 0.75); and diffusion characteristics F17 (AUC: 0.79). VASARI 2.0 models showed good performances according to the AUC values, which are also compared with traditional VASARI scores.

Discussion and conclusion: Glioma grade and isocitrate dehydrogenase (*IDH*) status can be predicted using specific magnetic resonance imaging (MRI) features, which have significant prognostic consequences. The accuracy of texture-derived metrics from preoperative MRI gliomas and machine learning analysis for predicting grade, *IDH* status, and their correlation can be enhanced by the suggested new and updated VASARI manual annotation system. To help with therapy selection and enhance patient care, we intend to create prediction models that incorporate these MRI findings with additional clinical data.

KEYWORDS

VASARI, MRI, glioma, *IDH* status, grade tumor, neuroradiology

1 Introduction

The cerebral glioma, a sizable and heterogeneous family of brain tumors with various features, is the most prevalent primary malignant brain tumor that exhibits variable treatment response and patient prognosis (1). Given the considerable variations in the care of these many glioma subtypes, an accurate diagnosis is essential. Glioblastoma and oligodendroglioma, for instance, respond very differently to treatment. Furthermore, tumors belonging to the same histologic subtype could exhibit distinct behaviors in other patient cohorts. Previously, the phenotypic characteristics of the cells were used to identify the tumor grade, but, today, this seems too simplistic (2), and it is not possible to identify these distinctions between the different glioma subtypes based on the purely histology-driven older classification system. Part of the challenge was that many gliomas can contain mixed cell types, which result in high inter-observer variability of diagnosis among neuropathologists (3).

Thus, a paradigm change in the diagnosis and categorization of gliomas has resulted from new discoveries on their genetic composition.

The updated glioma classification system incorporates molecular markers into tumor subgrouping, which has been shown to better correlate with tumor biology and behavior as well as patient prognosis than the previous purely histology-based classification system (4). The isocitrate dehydrogenase (*IDH*) gene changes and the co-deletion of chromosomal arms 1 and 19 (1p/19q) are the two main alterations taken into account when differentiating tumors in the World Health Organization (WHO) classification (5). It could be comparable to biomarkers that influence the prognosis and biological behavior of a patient. For instance, it has been demonstrated that *IDH* gene family mutations offer higher overall survival in high-grade gliomas than their *IDH*–wild-type counterparts, regardless of the histological grade (6–8). Additionally, the degree of cellular differentiation and the molecular state have an impact on the course of treatment. For instance, low-grade gliomas are often not treated with adjuvant radiotherapy and/or chemotherapy. Clinicians found that patients' responses to treatment, the severity of side effects, and even prognosis could differ even when they shared the same tumor. This implies that therapeutic care tailored to the needs of particular people or "precision medicine" may be the direction of the future (9–11). The method most frequently used today for identifying glioma mutations is immunohistochemical analysis after biopsy or surgical resection (12). By categorizing radiological gliomas in a non-invasive way with relevant prognostic consequences, clinical therapeutic planning may be recommended (13). Magnetic resonance imaging (MRI) is used as a gold standard for radiological examination of gliomas. Because there are no objective measurements that can be extensively duplicated and validated, determining the tumor grade accurately is far from simple (14).

Whereas advanced MRI techniques (e.g., diffusion Magnetic Resonance Imaging (dMRI), Perfusion Magnetic Resonance Imaging (pMRI), and Magnetic Resonance Spectroscopy (MRS)) are more specific to biophysical, cellular, and microstructural processes, conventional MRI methods (e.g., T1-weighted and T2-weighed sequences) give macrostructural anatomical evidence. Unlike standard MRI techniques, which only yield relative image contrasts, these advanced techniques have the potential to be (semi) quantitative. For the purpose of acquiring imaging biomarkers, sensitivity, specificity, and quantification are crucial (15). While a number of recent investigations have concentrated on applying sophisticated MRI methods (such as perfusion, spectroscopy, and

machine learning approaches) for radiogenomic purposes (16–18), standard MRI sequences continue to be the gold standard for the investigation and characterization of brain tumors.

An optimal and adequate method to identify the right correlation between imaging phenotype and genotype, based on the evaluation of specific radiological characteristics, mainly conventional MRI features, and, at the same time, to standardize the assessment of gliomas is represented by the VASARI manual annotation system.

A collection of standardized descriptors called VASARI (Visually AcceSAble Rembrandt Images) MRI characteristics is used to describe brain tumors on contrast-enhanced MRI imaging. These characteristics aid in the diagnosis, grading, and prognostication of gliomas by offering qualitative and quantitative information regarding the visual appearance and properties of the tumor (19). The location, shape, enhancement quality, necrosis proportion, edema proportion, and other geometric parameters of the tumor are all included in the VASARI features (20).

Since its development in 2016, VASARI score has undergone a development from the number of the features, now 25, to the field of application in neuro-oncology.

Certain specific MRI features [enhancement quality (F4), tumor-enhancing proportion (F5), tumor–non-enhancing proportion (F6), and necrosis proportion (F7)] have been shown in our previously published study (21) to be predictive of the grade and *IDH* status of gliomas, with significant prognostic implications.

Inter-observer agreement and multicenter collaborations are made possible by the reliability and consistency in the interpretation of MRI scans made possible by the standardization of the VASARI features (22, 23). The communication between radiologists, oncologists, and other medical professionals involved in the treatment of patients with glioma is improved when VASARI elements are used in structured reporting systems (24). To increase the precision of glioma grading, prognosis prediction, and treatment response assessment, they have been used in machine learning algorithms and radiomics studies (24, 25). Predictive models to inform therapy choices and patient care can be created by merging VASARI variables with additional imaging features and clinical data (26, 27).

Such a model's potential resides in its capacity to evaluate tumor features objectively. In actuality, although VASARI started out as a visual assessment scale, a numerical estimate of the features under consideration can be obtained by using "regions of interest" (ROIs) to calculate the areas of the various tumor components. However, we think that such a system is not useful in terms of outpoint prediction and is too sophisticated and challenging for doctors to utilize, especially in light of the outcomes documented in the literature and our earlier study.

In this regard, we propose a new VASARI glioma score, which we refer to as VASARI 2.0. This system evaluates only those tumor features that can be objectively described by ROI (manual segmentation) and that predict the outpoint (*IDH* status and grade) with area under the curve (AUC) > 0.8. For this purpose, we modify the evaluation intervals/ranges as explained in the following (Table 1), all in order to provide the scientific community with a system suitable with clinical practice.

2 Materials and methods

2.1 Ethics statements

The Institutional Review Board accepted the study because the surgery was routinely carried out and was not considered experimental. Each patient completed and signed a proper written informed consent. There was no indication of a conflict of interest from the writers. No funding was given to support this study.

TABLE 1 Modified ranges between VASARI and VASARI 2.0.

	F4 Enhancement quality	*F5 Enhanced area*	*F6 No-enh area*	*F7 Necrosis area*	*F14 Edema area*	*F17 Diffusion quality*
VASARI 2.0	1. Absent	1. Absent or <5%	1. Absent or <5%	1. Absent or <5%	1. Absent or <5%	1. Augmented
	2. Minimal	2. 6–25	2. 6–25	2. 6–25	2. 6–25	2. Reduced
	3. Avid	3. 26–50	3. 26–50	3. 26–50	3. 26–50	3. Mixed
		4. 51–75	4. 51–75	4. 51–75	4. 51–75	
		5. 76–95	5. 76–95	5. 76–95	5. 76–95	
		6. >95%	6. >95%	6. >95%	6. >95%	
VASARI	1. Absent	3. <5%	3. <5%	2. No	2. No	2. Augmented
	2. Minimal	4. 6%–33%	4. 6%–33%	3. <5%	3. <5%	3. Reduced
	3. Avid	5. 34%–67%	5. 34%–67%	4. 6%–33%	4. 6%–33%	4. Mixed
		6. 68%–95%	6. 68%–95%	5. 34%–67%	5. 34%–67%	
			7. 96%–99%			
			8. 100%			

2.2 Patient population

Retrospective analysis of patients who underwent MRI for pre-surgical glioma evaluation between 2018 and 2021 has been done on the database at our institution. Additionally, pathology reports were gathered to determine the glioma grade. According to the following criteria, a number of patients were disqualified from the study: (a) poor acquisition quality imaging; (b) no intravenous contrast; (c) medications prior to the MR examination, such as steroid medication that may alter edema and contrast enhancement; and (d) lack of a pathology report. In the end, 126 individuals with glioma were included. The study group included 51 women and 75 men, ages 14 to 84 (further information in Table 2).

2.3 Image acquisition

MRI at 1.5 T (Magnetom Amyra; Siemens Medical Systems, Erlangen, Germany) was used for the imaging. In addition to T2-weighted images with dark fluid on the axial planes, the MR technique also includes T1-weighted images taken before and after the administration of gadolinium-based contrast media. In addition to this, we also carried out Diffusion Weighted Imaging (DWI) and Susceptibility Weighted Imaging (SWI) on the axial plane, as well as T1-w and T2-w sequences on additional planes. These were the precise imaging parameters: (1) axial T1-weighted MR: repetition time of 250 ms, echo time of 2.46 ms, slice thickness of 5 mm, matrix dimensions of 320 × 256, and field of view of 220 × 220 mm^2; (2) axial T2-weighted MR: repetition time of 6000 ms, echo time of 93 ms, slice thickness of 5 mm, matrix dimensions of 320 × 288, and field of view of 198 × 220 mm; and (3) axial T2WI dark-fluid MR: repetition time of 8000 ms, echo time of 97 ms, slice thickness of 5 mm, matrix dimensions of 320 × 224, and field of view of 181 × 220 mm.

2.4 Magnetic resonance imaging assessment and analysis

The VASARI lexicon can be easily understood by following the specific guide downloadable from the public website of The Cancer

TABLE 2 Demographic data about our study population.

Demographic data		Glioma grade				
		1 (n = 3)	2 (n = 21)	3 (n = 18)	4 (n = 84)	Total (126)
Age (yr.)	*< 50*	*3*	*10*	*8*	*18*	*39*
	> 50	*0*	*11*	*10*	*66*	*87*
Sex	*Male*	*2*	*8*	*11*	*54*	*75*
	Female	*1*	*13*	*7*	*30*	*51*
Location	*Frontal*	*0*	*12*	*9*	*28*	*49*
	temporal	*0*	*7*	*4*	*17*	*28*
	Insular	*2*	*2*	*1*	*6*	*11*
	Parietal	*0*	*0*	*1*	*22*	*23*
	Occipital	*0*	*0*	*2*	*2*	*4*
	Brain steam	*1*	*0*	*1*	*5*	*7*
	Other (cerebellum)	*0*	*0*	*0*	*4*	*4*
Side	*Right*	*0*	*11*	*5*	*47*	*63*
	Left	*2*	*0*	*2*	*5*	*9*
	Central/Bilateral	*1*	*10*	*11*	*32*	*54*
Eloquent area	*No*	*2*	*15*	*13*	*45*	*75*
	Motor speech	*1*	*2*	*1*	*7*	*11*
	Receptive speech	*0*	*4*	*2*	*16*	*22*
	Motor area	*0*	*0*	*1*	*15*	*16*
	Visual area	*0*	*0*	*1*	*1*	*2*
IDH status	*Positive*	*2*	*13*	*3*	*4*	*22*
	Negative	*1*	*8*	*15*	*80*	*104*

Imaging Archive, in the specific section "Supporting Documentation and Metadata" (https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project).

We considered that the entire lesion was made up of necrotic tissue, edema, enhancing area, and non-enhancing area in accordance with the VASARI approach. Furthermore, we extracted the score system and morphological features. Therefore, an enhancing area was defined as any region of the tumor that shows a discernible increase in signals on the post-contrast T1-weighted pictures in comparison to those in the pre-contrast. Any region displaying T2-weighted hyperintensity (less than the intensity of the Cerebral Spinal Fluid (CSF) fluid) and corresponding T1-weighted hypointensity, as well as a mass effect and architectural distortion, such as blurring of the gray-white interface, was deemed to be a non-enhancing area. An irregular border, a high signal on T2-weighted and proton density imaging, and either no enhancement at all or a significantly decreased enhancement are characteristics of a necrotic section of the tumor. By calculating the ratio of the total lesion area to the necrosis area (internal to it), a quantitative evaluation of the necrosis was produced. On the T2-, T1-, and SWI T2*–weighted sequences, a bleeding was detected and assessed in connection with the existence of hemoglobin breakdown products. Based on an apparent diffusion coefficient (ADC) map, the diffusion characteristics are classified as mostly facilitated or restricted in the enhancing or non-contrast–enhanced tumor (nCET) region of the tumor. They are described as mixed when there is a roughly equal amount of both limited and assisted diffusion.

Three neuroradiologists, two residents and one senior, independently evaluated the imaging characteristics.

2.5 Statistical analysis

The aims of the statistical analysis were as follows: 1) analyze the statistically significance of each variable with respect to the prediction of the variable levels; 2) analyze the statistically significance of each variable with respect to the prediction of the variable *IDH*; and 3) analyze the relationship between *IDH* and GRADE.

To consider only relevant columns in the dataset, a sub-dataset was created, only with the following columns: GRADE, F4, F5, F6, F7, F14, F15, and F24, and another column was added to the dataset named "levels." This variable is binary, with level = 0 denoting a grade of 1 or 2 (low grade) and level = 1 denoting a grade of 3 or 4 (high grade). The variable "*IDH* mutate" has been turned into a dummy binary variable too (*IDH* 0 = non-mutate = neg; *IDH* 1 = mutate = pos).

To see whether the variable GRADE and *IDH* are correlated, we built a contingency table where, on one side, there are the levels of GRADE and, on the other, the *IDH*, negative or positive. Then, we also did the Pearson correlation test and built a correlation plot.

As is often the case in real datasets, the VASARI dataset that we analyzed is highly imbalanced (80% vs. 20%) and is of high grade, so we divided the dataset into train (70%) and test (30%), and, then, we balanced the training data in such a way that we obtained 2,000 observation and perfectly balanced classes (p = 0.5). Then, we proceeded to the analysis of the statistically significant variables in the prediction of levels. As the first step in this part of the analysis, we conducted a multiple logistic regression using the dichotomic variable levels as response variable and each variable as covariate on the balanced training set. We also conducted a Wald test on the categorical variables to confirm statistical significance.

We computed the odds ratios (OR) and confidence intervals for each variable and evaluated the matrices with ROC curves.

After the balancing of the classing, we finally conducted a multiple logistic regression using all the variables in the dataset to have an idea of which variable is statistically significant and to see how does our classifier performs having the whole set of information in the prediction of *IDH*. We then applied this model to the test set to evaluate the predictive performance in the analysis of the relationship between *IDH* and GRADE.

We compared the results with the same value obtained in our previous studies using the gold standard for VASARI score and traditional statistics method; in particular, we focused on the AUC value to compare diagnostic accuracy.

The analysis has been done on the software R, using the package ROSE for the balancing purpose.

Applying more or less complex machine learning methods to this type of data is very risky because the more unbalanced the classes, the greater the risk of having results biased by lack of observation in one class.

More specifically, the classification problem's confusion matrix indicates how well our model classifies the target classes, and it is from this confusion matrix that we derive the model's accuracy, which is determined by dividing the total number of predictions made by the model correctly by the total number of predictions. Thus, in cases where a class has few observations, it may be categorized as the most popular class, potentially yielding a high accuracy score. For example, one of the most often used parametric techniques for binary classification is logistic regression, which is heavily biased in cases when the classes are not balanced because it underestimates the conditional probabilities of the rare class. To solve these problems, many methods have been proposed in the literature, such as oversampling, undersampling, SMOTE (Synthetic Minority Oversampling Technique), and ROSE (Random Over-Sampling Examples). In this paper, we chose to use the most recent ROSE technique. It is a bootstrap-based method that helps with binary classification when there are uncommon classes present. By creating synthetic examples from a conditional density estimate of the two classes, it can handle both continuous and categorical data. We selected this approach because of the strong theoretical underpinnings of ROSE. It also draws synthetic examples from an estimate of the conditional density underlying the data, thus providing confidence that the distribution of the data into the classes has not changed because the balancement has been performed.

3 Results

In previous studies, it has been demonstrated that some of that specific MRI features [enhancement quality (F4), tumor-enhancing

proportion (F5), tumor–non-enhancing proportion (F6), and necrosis proportion (F7)] can be used to predict the grade and *IDH* status of gliomas, with important prognostic implications. The standardization and improvement of these data can be used for programming machine learning software (28).

3.1 Part 1: preliminary analysis and relationship between GRADE/*levels* and *IDH*

We obtained a clear relation between a negative *IDH* and high GRADE. In fact, 75.2% of observations have a negative *IDH* and a high grade (= 3 or 4) with a p-value of 2.2e−16 < 0.05, so the coefficient is statistically significant. We obtained a correlation coefficient of −0.661, meaning that these two variables are significantly negatively correlated. The resulting confidence interval is [−0.750, −0.550], respectively, at 2.5% and 97.5%. These results are shown in Figure 1.

3.2 Part 2: analysis of the statistically significant variables in the prediction of *levels/grade*

We applied this model to the test set to evaluate the predictive performances, and we obtain accuracy = 0.811, sensitivity = 0.781, and specificity = 1, with AUC = 0.906. We proceeded with the analysis of the statistical significance of each variable in the prediction of *levels*. We obtained that variable F4 is significant with a p-value = 2.7e−10 < 0.05 with no significant differences between levels 1 and 3. For every one unit increase in F4 = 2, the odds of being in the level = 1 (high grade) increases by a factor of 4.699.

We calculated the predicted probabilities to be in the high-level grades (vs. low level) at each level of F4: F4 = 3 has 100% probability to be in the high level (grades 3 and 4), whereas the probability for the other two levels is much lower (5% and 20%). This model has accuracy = 0.811, sensitivity = 0.781, and specificity = 1.000, with AUC = 0.869, which is superior to that of the first study (0.73).

Variable F5 is statistically significant with a p-value < 2e−16 < 0.05 with significant differences between levels 2 and 6 of the F5 variable and cutoff level F5 = 4.

For every one unit increase in F5 = 3, the odds of being in the level = 1 (high grade) increases by a factor of 0.49.

F5 = 5 and F6 = 6 have a probability to be in the high level (grades 3 and 4) of 73% and 63%, whereas the probability for the other three levels is much lower (19%, 73%, and 63%).

This model has accuracy = 0.672, sensitivity = 0.908, and specificity = 0.438, with AUC = 0.712.

Variable F6 is statistically significant with a p-value = 0.01409 < 0.05.

For every one unit increase in F6 = 2, the odds of being in the level = 1 (high grade) increases by a factor of 1. We also obtained that there is a significant difference between the F6 = 3 and F6 = 4 (p-value = 0.0049 < 0.05). We calculated the predicted probabilities to be in the high-level grades (vs. low level) at each level of F6: F6 = 1, F6 = 2, and F6 = 3 have a probability to be in the high level (grades 3 and 4) of 100%, whereas the probability for the other levels is much lower (24.1%, 15.9%, and 0%, respectively). We can also clearly identify a cutoff level for this variable. This model has accuracy = 0.757, sensitivity = 0.719, and specificity = 1.000, with AUC = 0.8969. Also, this result is superior to that of our first study.

Variable F7 is statistically significant with a p-value = 2e−16 > 0.05, but none of the levels of F7 is statistically significant. For every one unit increase in F7 = 2, the odds of being in the level = 1 (high grade) increases by a factor of 3.35; whereas for every one unit increase in F7 = 4, the odds of being in the level = 1 (high grade) increases by a factor of 3.22e+09.

Moreover, we calculated the predicted probabilities to be in the high-level grades (vs. low level) at each level of F7: F7 = 4, F7 = 5, and F7 = 6 have the probability to be in the high level (grades 3 and 4) of 100%; whereas when F7 = 1 and F7 = 2, the probability to be in the high level is 8.9% and 24.6%, respectively. This model has accuracy = 0.837, sensitivity = 0.906, and specificity 0.400, with AUC = 0.8 (vs. an AUC of 0.738).

Variable F17 is statistically significant with a p-value = 3.634e −05 < 0.05. Moreover, there is a significant difference between levels F17 = 1 and F17 = 2 with a p-value of 0.017 < 0.05 and also between F17 = 1 and F17 = 3 with a p-value of 0.0017 < 0.05. We can then identify a cufoff in this case.

For every one unit increase in F17 = 2, the odds of being in the level = 1 (high grade) increases by a factor of 7.71; whereas for every one unit increase in F17 = 3, the odds of being in the level = 1 (high grade) increases by a factor of 18.48.

We can calculate the predicted probabilities to be in the high-level grades (vs. low level) at each level of F17: F17 = 3 has the probability to be in the high level (grades 3 and 4) of 74%; whereas when F17 = 1 and F17 = 2, the probability to be in the high level is 13.4% and 24.3%, respectively.

This model has accuracy = 0.840, sensitivity = 0.921, and specificity 0.500, with AUC = 0.8.

Variables F14, F15, F18, F19, and F20 were not statistically significant with a p-value < 0.05. These results are shown in Figure 2.

3.3 Part 3: analysis of the statistically significant variables in the prediction of *IDH* mutation

The pre-processing part is identical to the one described in part 2: We divided the dataset into train (70%) and test (30%), and, then, we balanced the training data in such a way that we obtained 2,000 observation and perfectly balanced classes (p = 0.5). Then, we conducted a multiple logistic regression and obtained that all the variables are statistically significant. We then applied this model to the test set to evaluate the predictive performances, and we obtained accuracy = 0.88, sensitivity = 0.6, and specificity = 0.95, with AUC = 0.93. We proceed with the analysis of the significance of each variable in the prediction of *IDH*. We obtained that variable F4 is significant with a p-value = 1.9e−05 < 0.05 with a significant

Relationship between GRADE/*levels* and *IDH*

	GRADE				
	1	2	3	4	Sum
IDH-	1.6	3.2	10.4	64.8	80.0
IDH+	0.8	13.6	4.0	1.6	20.0
Sum	2.4	16.8	14.4	66.4	100.0

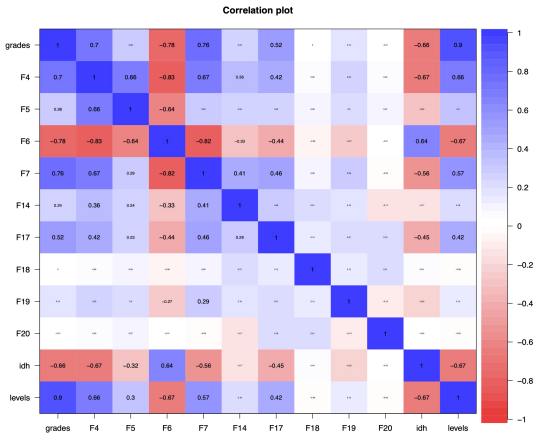


FIGURE 1
In our analysis, we obtained a clear relationship between negative *IDH* and high GRADE. In fact, 75.2% of the observations have negative *IDH* and high grade (= 3 or 4) with a p-value of 2.2e−16 < 0.05, so the coefficient is statistically significant. We obtained a correlation coefficient of −0.661, which means that these two variables are significantly negatively correlated. The resulting confidence interval is [−0.750, −0.550] at 2.5% and 97.5%, respectively. Statistical analysis by contingency table, which is then confirmed by Pearson's test, obtained a statistically significant correlation between *IDH* mutation and glioma grade: In 75% of the observations, we found *IDH* WT and high grade.

difference between F4 = 2 and F4 = 3 (p-value 8.9e−05 < 0.05). The results of the OR have a key role in interpreting the role of the variable. In particular, the OR column means that, for every one unit increase in F4 = 2, the odds of having a positive *IDH* increases by a factor of 0.14; whereas for every one unit increase in F4 = 3, the odds of having a positive *IDH* increases by a factor of 0.0. Moreover, we can calculate the predicted probabilities of having a positive *IDH* (vs. negative) at each level of F4: F4 = 3 has 1.2% probability of having a positive *IDH*, whereas the probability for the other two levels is higher (76.9% and 45.2%). This model has accuracy = 0.90, sensitivity = 0.9, and specificity 0.90, with AUC = 0.904, which is superior to that of our first study (0.73).

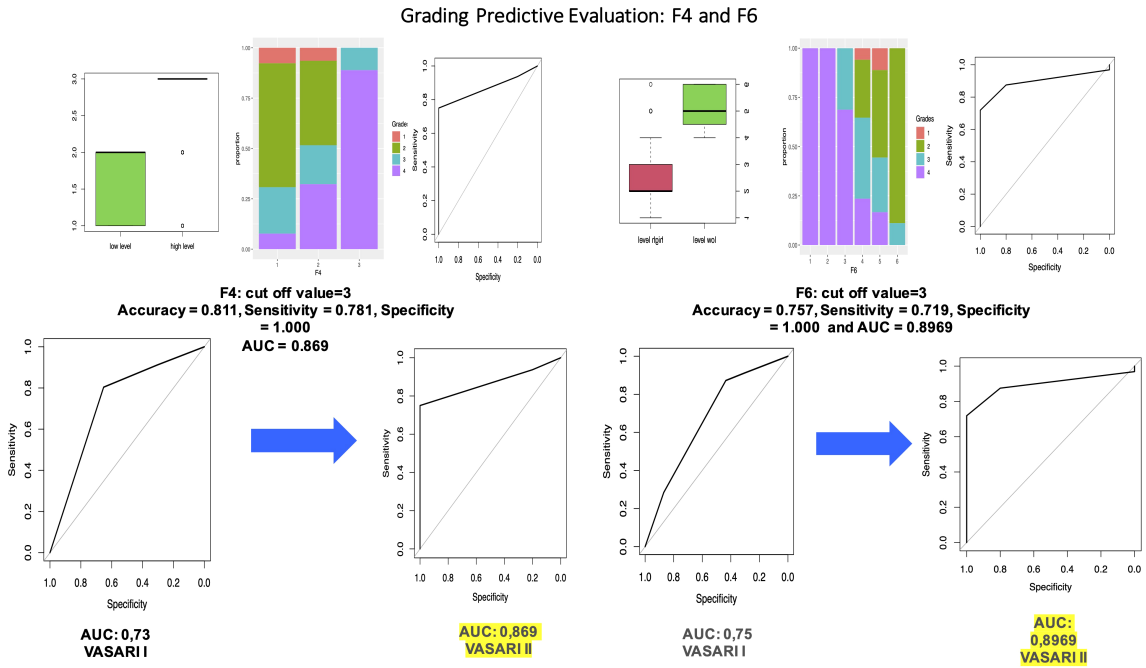


FIGURE 2
Through logistic regression using the dichotomous variable (high or low grade) as the response variable and F4 as the covariate, we obtained that variable F4 is significant with a p-value = 0.000157 < 0.05. We also compared the performance of VASARI 2.0 with that of traditional VASARI in predicting grade.

10.3389/fonc.2024.1449982

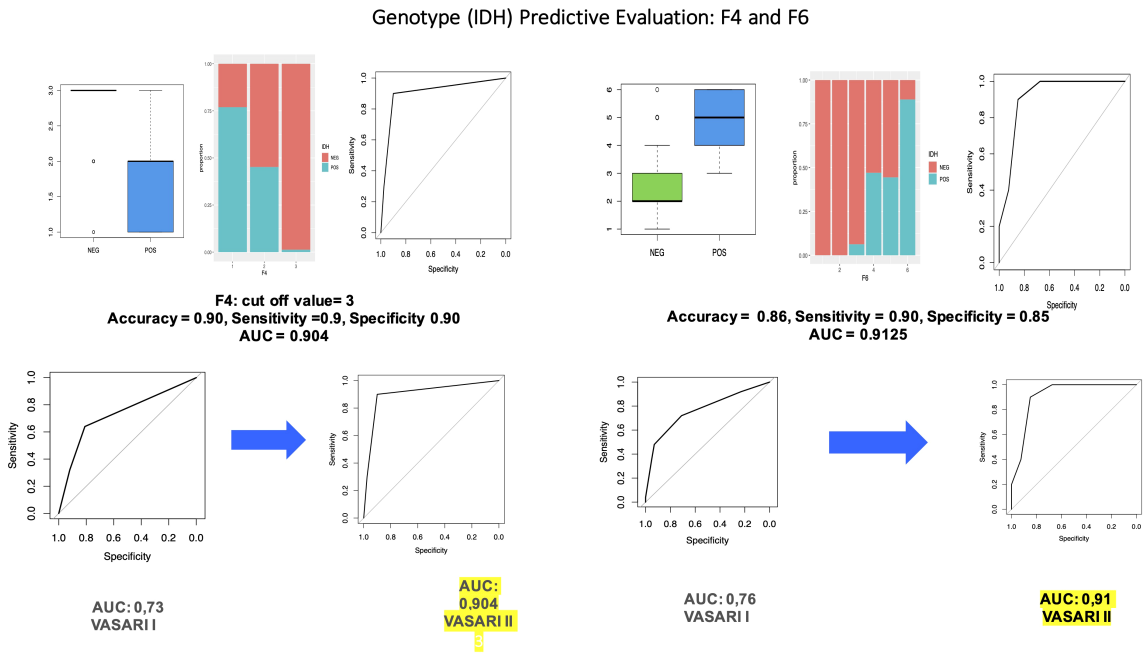


FIGURE 3
Through univariate logistic regression to evaluate the influence of variables in predicting/*IDH* levels, we obtained that variable F6 is statistically significant with a p-value = 0.0032 < 0.05. We also performed a Wald test on variable F6, which concluded that this variable is statistically significant (p-value = 0.01409 < 0.05). We also compared the performance of VASARI 2.0 with that of traditional VASARI in predicting *IDH* status.

Variable F5 is statistically significant with a p-value = 0.04386 < 0.05 with a significant difference also between F5 = 3 and F5 = 4 with a p-value = 3.6e−11 < 0.05 and between F5 = 3 and F5 = 5 with a p-value = 0.0077 < 0.05. Hence, we can conclude that the cutoff level for variable F5 is F5 = 3. The OR column showed that, for every one unit increase in F5 = 2, the odds of having a positive *IDH* increases by a factor of 4.978539e−08; whereas for every one unit increase in F5 = 7, the odds of having a positive *IDH* increases by a factor of 5.524675e−16. These values are so low because F5 = 1 is taken as baseline, where only positive values are observed.

Moreover, we calculated the predicted probabilities of having a positive *IDH* (vs. negative) at each level of F5: F5 = 1, F5 = 2, and F5 = 3 have a high probability to have a positive *IDH* of 100%, 67%, and 81%, respectively, whereas the probability for the other three levels is much lower (53.3%, 14%, and ~0%). This modes has accuracy = 0.808, sensitivity = 0.48, and specificity 0.890, with AUC = 0.73.

Variable F6 is statistically significant with a p-value = 5.162e−05< 0.05 with a significant difference between levels F3 and F4 with a p-value of 2e−16 < 0.05 and between F5 and F6 with a p-value of 1.1e−14 < 0.05. With these results, we may identify two different cutoffs.

The OR column showed that, for every one unit increase in F6 = 2, the odds of having a positive *IDH* increases by a factor of 1; whereas for every one unit increase in F6 = 4, the odds of having a positive *IDH* increases by a of 6.704338e+08.

Moreover, we calculated the predicted probabilities to be in the high-level grades (vs. low level) at each level of F6: F6 = 4, F6 = 5, and F6 = 6 have the highest probability to have a positive *IDH*: 68.1%, 75.5%, and 97.4%, respectively, whereas the probability for the other levels is 0%. This model has accuracy = 0.86, sensitivity = 0.90, and specificity = 0.85, with AUC = 0.9125 (vs. AUC = 0.7648).

We obtained that variable F7 is statistically significant with a p-value = 2.31e−06 < 0.05 and a significant difference between F7 = 3 and F4 = 4 (p-value = 0.00085 < 0.05). Another significant difference is between F7 = 2 and F7 = 4. We can then conclude that we have a cutoff for values higher than F7 = 3. The OR column showed that, for every one unit increase in F7 = 2, the odds having a positive *IDH* increases by a factor of 2.083333e−01. Moreover, we can calculate the predicted probabilities to have a positive *IDH* at each level of F7: F7 = 1, F7 = 2, and F7 = 3 have a probability of having a positive *IDH* of 61.5%, 25%, and 21.2%, respectively, whereas F7 = 4, F7 = 5, and F7 = 6 have a probability of having a positive *IDH* mush lower: 0%. This model has accuracy = 0.740, sensitivity = 1.000, and specificity = 0.675, with AUC = 0.835 (vs. AUC = 0.789).

Variable F14 is barely statistically significant with a p-value = 0.0648 > 0.05 with no significant difference between the levels in F14. The OR column showed that, for every one unit increase in F14 = 2, the odds having a positive *IDH* increases by a factor of 0.25; whereas for every one unit increase in F14 = 3, the odds of being in the odds having a positive *IDH* increases by a factor of 0.625. Moreover, we can calculate the predicted probabilities to have a positive *IDH* at each level of F14: F14 = 3 has the highest probability of having a positive *IDH* with the 38.5%; whereas F14 = 2, F14 = 4, and F14 = 5 have a probability of having a positive *IDH* much lower: 20%, 8.8%, and 12.5%, respectively. This model has accuracy = 0.780, sensitivity = 0.600, and specificity 0.825, with AUC = 0.75.

Variable F17 is statistically significant with a p-value = 9.372e−06 < 0.05 and a significant difference between the levels in F17 = 1 and

F17 = 2 with a p-value of 0.00049 < 0.05. The OR column showed that, for every one unit increase in F17 = 2, the odds having a positive *IDH* increases by a factor of 0.11; whereas for every one unit increase in F17 = 3, the odds of being in the odds having a positive *IDH* increases by a factor of 0.05. Moreover, we can calculate the predicted probabilities to have a positive *IDH* at each level of F17: F17 = 3 has the lowest probability of having a positive *IDH* with the 8.7%; whereas F17 = 1 has a higher probability of having positive *IDH*: 65%. This modes has accuracy = 0.860, sensitivity = 0.400, and specificity 0.975, with AUC = 0.7925.

Other features, qualitative or non-statistically significant, were not taken into account. These results are shown in Figure 3.

4 Discussion

In our study, we confirm that there is a positive statistical evidence between some VASARI features and *IDH* and glioma grade (Figure 1). The most significant variables in the prediction of IDH are F4 (AUC: 0.904), F5 (AUC: 0.73), F6 (AUC: 0.91), F7 (AUC: 0.84), F14 (AUC: 0.75), and F17 (AUC: 0.79).

The most significant variables in the prediction of *levels* are F4 (AUC: 0.87), F5 (AUC: 0.7), F6 (AUC: 0.89), F7 (AUC: 0.79), and F17 (AUC: 0.75).

The statistical significance for all the features is increased using VASARI 2.0 compared to result obtained with traditional VASARI ranges, demonstrating how the new proposed VASARI lexicon promotes an increase in the sensitivity, specificity, and AUC of these features to increase the statistical significance with AUC > 0.8 and to make this system more suitable for clinical practice. We can distinguish low-grade gliomas from high-grade gliomas using this model. Numerous studies have assessed the application of the VASARI Lexicon in the categorization of cerebral gliomas (29–34). Due to a much more extensive destruction of the blood–brain barrier in high-grade gliomas compared to that in low-grade gliomas, our study, for instance, produced similar results to those in the work done by Su et al., in which the enhancement quality and the proportion enhancing were significantly higher in high-grade gliomas (33). High-grade gliomas exhibit much higher cell growth proliferation. In the context of tumor tissue, new, irregular, and aberrant vessels form very rapidly, without an adequate blood–brain barrier, a process known as neoangiogenesis. Newly formed vessels are much more permeable than normal vessels. Therefore, in an MRI study after intra venous of contrast agent, the solid tumor component appears much brighter or "enhanced" in the resulting images, given that the contrast agent can cross the blood–brain barrier very easily, escaping from the blood vessels into the surrounding tumor tissue. Furthermore, areas of necrosis can very often be found in high-grade gliomas, given that they have a very high rate of proliferation and cell growth, which is associated with an inadequate blood supply. These necrotic phenomena, therefore, contribute to the breakdown of the blood–brain barrier, allowing a greater leakage of contrast agents into the tumor tissue and promoting greater further enhancement (31, 32). In high-grade gliomas, there are often infiltrative lesional margins in the surrounding healthy tissue. In MRI, all this appears as an area of pathological alteration of the signal intensity, with hyperintensity in the Fluid Attenuated Inversion Recovery (FLAIR)-weighted sequences, much larger and more extensive than the pathological area visible on the T1-weighted MRI sequences. Our study supports this theory by showing that the infiltrative T1/FLAIR ratio has a high predictive value for glioma grade (OR = 41.99, p = 0.001). The fast and uncontrolled phenomenon of neoangiogenesis is indirectly indicated by the presence of hemorrhagic components and necrosis in the context of high-grade gliomas. The newly created blood vessels are easily injured and do not have an uneven morphology, which results in bleeding. In our study, the presence of necrosis was a strong predictive factor with an OR of 13.57 (p < 0.001) for high-grade glioma, a result that is consistent with previous research findings (30–34).

In our study, we found that edema proportion was a significant factor. These findings, in line with other recent studies, highlight that mass effect is an important predictor of astrocytoma grade. Tumor perilesional cerebral edema correlates with the WHO pathological grading as recently demonstrated (35). Measuring areas of non-enhancing tumor have been highlighted by The Response Assessment in Neuro-Oncology Working Group (36). Astrocytomas represent very heterogeneous neoplasms. The component characterized by contrast enhancement does not always contain anaplastic parts, unlike the component without contrast enhancement, which often contains both anaplastic parts and low-grade parts. The evaluation of the nCET component, therefore, is very important both in the diagnosis and in the follow-up to better assess the therapeutic monitoring of astrocytoma. However, understanding the real extent of a high-grade astrocytoma by evaluating the nCET proportion is very difficult given the extremely heterogeneous nature of the tumor and the extreme difficulty in delineating its peripheral margins. In reality, all this represents a false dichotomy, given that, in infiltrating gliomas, there is an "infiltrative edema," consisting of tumor cells and edema in the background of the brain. Furthermore, even with the use of sophisticated techniques like T2 mapping, diffusion tensor imaging, and perfusion imaging, it remains difficult to differentiate pure vasogenic edema from infiltrative edema. In this work, using multivariate analysis based on VASARI, we showed that nCET percentage was a predictive factor of grade 4 astrocytoma. The new classification of tumors of the central nervous system published in 2021 highlights the importance of evaluating the molecular status, particularly, and first of all, the *IDH* mutation. Usually, one of the two *IDH* genes, *IDH1* and *IDH2*, is affected by the mutation. The mutation most frequently found in gliomas is that affecting the *IDH*1 gene. Typically, a specific mutation (R123H) occurs that causes a single–amino acid change in the active enzyme site. The mutated *IDH* enzyme promotes the conversion of alpha ketoglutarate into 2-hydroxyglutarate, an oncological metabolite that induces cancer formation (6). The presence of the *IDH* mutation is associated with a significantly better prognosis compared to *IDH*–wild-type gliomas, and its identification is, therefore, very important for the classification of cerebral gliomas and for clinical therapeutic management. In our study, six VASARI features were found to predict *IDH* mutation status: F4, enhancement quality (AUC: 0.904); F5, tumor-enhancing proportion (AUC: 0.73); F6, tumor–non-enhancing proportion (AUC: 0.91); F7, necrosis proportion

(AUC: 0.84); F14, proportion of edema (AUC: 0.75); and F17, diffusion characteristics (AUC: 0.79). Our study confirms the results of our recently published study for VASARI characteristics and demonstrates an increase in their diagnostic accuracy, especially regarding F4 and F6 with an AUC value greater than 85%, in line with the ESR statement on the validation of imaging biomarkers (European Society of Radiology, 2020). Olar et al. identified a significant correlation between proportion of enhancing tumor and *IDH* mutation. In the study, the researchers studied the role that the *IDH* mutation may have in the grading and mitotic index in grade II–III diffuse astrocytomas, and their results demonstrated that the *IDH* mutation determines the effect of mitotic index on patient outcome (37, 38). *IDH*–wild-type gliomas usually have a higher contrast enhancement than *IDH*-mutant gliomas (39–43), and *IDH* wild-type gliomas have undefined margins (44). Weighted FLAIR sequences are very useful for evaluating areas with pathological signal intensities. In the case of *IDH*–wild-type gliomas, the areas with pathological signal hyperintensity in the weighted FLAIR sequences, which extend beyond the margins of enhancement, usually represent the infiltrative edema component, characterized by the presence of infiltration of tumor cells in the peripheral tissue (31). Using multivariate analysis, we discovered that, in our cohort, the percentage of necrosis accurately predicts the status of *IDH* mutation. In our investigation, necrosis accounted for less than 25% of the total tumor volume in *IDH*-mutant cohorts and more than 50% in *IDH*-wild phenotypes. These results support the conclusions of multiple investigations. *IDH*-mutants were frequently linked to a cutoff necrosis of less than 33% of the tumor, according to Park et al. (45). Excessive tumor necrosis in *IDH*–wild-type gliomas is determined by increased hypoxia, which is brought on by intravascular thrombosis and the coagulation pathway activation (45–47). Furthermore, our results highlight two other VASARI features that can be used in the prediction of *IDH* status, which were not highlighted in our recently published work, namely, F14 (proportion of edema) and F17 (diffusion characteristics). In *IDH*-mutant gliomas, absent edema or edema with an extension smaller than that of the solid tumor component was found; whereas in *IDH*–wild-type gliomas, the extension of the edematous alteration was greater than or equal to the volume of the tumor. Similar results were documented in the studies by Lasocki et al. (40) and Patel et al. (48). Lasocki et al. found a cutoff value of 33% to distinguish *IDH*-mutant gliomas from *IDH*–wild-type gliomas. Furthermore, the presence of cysts was documented more frequently in *IDH*-mutant gliomas, in line with other published studies (49). In several other studies, *IDH*-mutant gliomas had higher average ADC values than *IDH*–wild-type gliomas, underlining that their edematous component is usually less infiltrative and destructive (50–54). Nowadays, the main attraction of the scientific interest is radiomic and machine learning that has been applied to tumor grading and diagnosis, tumor segmentation, non-invasive genomic biomarker identification, detection of progression, and patient survival prediction. It has been suggested that machine learning models are capable of more accurate prognosis prediction than histopathological categorization. These considerations could be the starting point for subsequent studies. The standardization and improvement of these data can be used for programming machine learning software (55). In the field of brain tumor, interest in machine learning methods is increasing, especially in diagnosis and pre-surgery planning though un-invasive histopathological categorization. Radiomics may be able to determine a tumor's response to treatment, make an accurate diagnosis, and forecast a prognosis (56). In addition to offering extra prognostic data, radiomic analysis' ability to non-invasively differentiate between various glioma molecular subtypes would aid in the choice of targeted chemotherapy for patients with numerous genetic mutations and possibly high-grade tumor types (57–59). Additionally, it would improve surgery, which is necessary to maintain median survival (60). Therefore, treatment responses, progression-free survival, and overall survival can all be more accurately predicted with the application of radiomic risk models (61, 62). It is possible to evaluate the effectiveness of anti-angiogenic therapy without endangering the patient by non-invasively acquiring the radio-genomic profile of a tumor (63). Our study has some limitations that need to be clarified and discussed: a single retrospective center study and a small sample size. This would not allow a validation of the new VASARI 2.0 method proposed by use on a large scale quickly. However, the number of cases is in line with for the type of analysis described. Future studies with multi-center data or larger cohorts needed for a full validation of the new VASARI lexicon that we proposes, in order to eliminate the risk of data bias, which could affect the generalizability of the study results. However, there are also strong points for the use of the proposed VASARI 2.0 lexicon in daily clinical radiological practice. It does not require specific software to automate the scoring process. There are methodological innovations in the evaluation of the MRI VASARI features, thanks to adequate changes in the reference intervals as reported in Table 1. Familiarization with this new lexicon is easy. It can be easily used in daily clinical practice also because it can be of valid help in capturing the most salient aspects to be described in the report. It can represent a valuable tool for producing a structured and standardized report with the aim of offering with simplicity and clarity all the salient information needed by the neuro-oncology core group (oncologist, radiotherapist, neurosurgeon, and neuoradiologist). It would be desirable to conduct a large-scale multi-center study to then draft a new VASARI lexicon guide based on the validation results.

5 Conclusion

The evaluation of gliomas with modified ranges/score of VASARI 2.0 allows the prediction of the outpoint (*IDH* status and grade) with AUC > 0.8, higher than that of traditional VASARI. Thus, the new score could be used in pre-surgical evaluation of gliomas in a method both suitable with clinical practice and that can also be the starting point for subsequent studies of radiomics and machine learning.

Data availability statement

All relevant data is contained within the article. Additional data are available upon request to interested researchers.

Ethics statement

The studies involving humans were approved by Institutional Review Board - ASL NA1 Centro. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by-product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/ next of kin in accordance with the national legislation and institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

AN: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. LG: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. MT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. GP: Data curation, Methodology, Writing – review & editing. RI: Data curation, Methodology, Visualization, Writing – review & editing. MM: Data curation, Writing – review & editing. AE: Supervision, Writing – review & editing. FT: Supervision, Validation, Writing – review & editing. VD'A: Supervision, Validation, Visualization, Writing – review & editing. Members of ODM Multidisciplinary Neuro-Oncology Group: Data curation, Formal analysis, Resources, Software.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

Authors want to recognize all members of ODM Multidisciplinary Neuro-Oncology Group as effective participants to the study. Ospedale del Mare, ASL NA1 Centro, Naples, Italy: Oncology Unit: Pasqualina Giordano and Bruno Daniele; Radiotherapy Unit: Maria Francesca Giugliano, Marcella Gaetano, and Cesare Guida; Nuclear Medicine Unit: Eugenio Di Giorgio and Marco Spadafora; Pathological Anatomy Unit: Alfredo D'Avino and Antonio D'Antonio; Neurosurgery Unit: Carmela Chiaramonte, Federico Frio, Francesco De Falco, Francesco Colaleo, and Giuseppe Catapano.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Omuro A, DeAngelis LM. Glioblastoma and other Malignant gliomas: A clinical review. *JAMA*. (2013) 310:1842–50.

2. Parsons DW, Jones S, Zhang X, Lin JC-H, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. (2008) 321:1807–12.

3. DeWeerdt S. The genomics of brain cancer. *Nature*. (2018) 561:S54–5.

4. Bai J, Varghese J, Jain R. Adult glioma WHO classification update, genomics, and imaging: what the radiologists need to know. *Top Magn Reson Imaging*. (2020) 29:71–82.

5. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro Oncol*. (2021) 23:1231–51.

6. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. *IDH*1 and *IDH*2 mutations in gliomas. *N Engl J Med*. (2009) 360:765–73.

7. Hartmann C, Meyer J, Balss J, Capper D, Mueller W, Christians A, et al. Type and frequency of *IDH*1 and *IDH*2 mutations are related to astrocytic and oligodendroglial differentiation and age: A study of 1010 diffuse gliomas. *Acta Neuropathol*. (2009) 118:469–74.

8. Houillier C, Wang X, Kaloshi G, Mokhtari K, Guillevin R, Laffaire J, et al. *IDH*1 or *IDH*2 mutations predict longer survival and response to temozolomide in low-grade gliomas. *Neurology*. (2010) 75:1560–6.

9. König IR, Fuchs O, Hansen G, Von Mutius E, Kopp MV. What is precision medicine? *Eur Respir J*. (2017) 50:1700391.

10. Yang K, Wu Z, Zhang H, Zhang N, Wu W, Wang Z, et al. Glioma targeted therapy: Insight into future of molecular approaches. *Mol Cancer*. (2022) 21:39.

11. Leibetseder A, Preusser M, Berghoff AS. New approaches with precision medicine in adult brain tumors. *Cancers*. (2022) 14:712.

12. Dunbar E, Yachnis AT. Glioma diagnosis: Immunohistochemistry and beyond. *Adv Anat. Pathol*. (2010) 17:187–201.

13. Nicholson JG, Fine HA. Diffuse glioma heterogeneity and its therapeutic implications. *Cancer Discovery*. (2021) 11:575–90.

14. Weller M, van den Bent M, Preusser M, Le Rhun E, Tonn JC, Minniti G, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol*. (2021) 18:170–86.

15. Smits M. MRI biomarkers in neuro-oncology. *Nat Rev Neurol*. (2021) 17:486–500.

16. Habib A, Jovanovich N, Hoppe M, Ak M, Mamindla PR, Colen R, et al. MRI-based radiomics and radiogenomics in the management of low-grade gliomas: evaluating the evidence for a paradigm shift. *J Clin Med*. (2021) 1:1411.

17. Bhandari A, Sharma C, Ibrahim M, Riggs M, Jones R, Lasocki A. The role of 2-hydroxyglutarate magnetic resonance spectroscopy for the determination of isocitrate dehydrogenase status in lower grade gliomas versus glioblastoma: A systematic review and meta-analysis of diagnostic test accuracy. *Neuroradiology*. (2021) 63:1823–30.

18. Lu J, Li X, Li H. Perfusion parameters derived from MRI for preoperative prediction of *IDH* mutation and MGMT promoter methylation status in glioblastomas. *Magn Reson Imag*. (2021) 83:189–95.

19. You W, Mao Y, Jiao X, Wang D, Liu J, Lei P, et al. The combination of radiomics features and VASARI standard to predict glioma grade. *Front Oncol*. (2023) 13:1083216. doi: 10.3389/fonc.2023.1083216

20. The Cancer Imaging Archive. VASARI Research Project, in: *The Cancer Genome Atlas* (2020). Available online at: https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project (Accessed 4 Jun 2022).

21. Gemini L, Tortora M, Giordano P, Prudente ME, Villa A, Vargas O, et al. Vasari scoring system in discerning between different degrees of glioma and *IDH* status prediction: A possible machine learning application? *J Imaging*. (2023) 9:75.

22. Sampaio L, Linhares P, Fonseca J. Detailed magnetic resonance imaging features of a case series of primary gliosarcoma. *Neuroradiol J*. (2017) 30:546–53.

23. Setyawan NH, Choridah L, Nugroho HA, Malueka RG, Dwianingsih EK. Beyond invasive biopsies: using VASARI MRI features to predict grade and molecular parameters in gliomas. *Cancer Imaging*. (2024) 24:3.

24. Nam YK, Park JE, Park SY, Lee M, Kim M, Nam SJ, et al. Reproducible imaging-based prediction of molecular subtype and risk stratification of gliomas across different experience levels using a structured reporting system. *Eur Radiol*. (2021) 31(10):7374–85. doi: 10.1007/s00330-021-08015-4

25. Peeken JC, Goldberg T, Pyka T, Bernhofer M, Wiestler B, Kessel KA, et al. Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. *Cancer Med*. (2019) 8(1):128–36.

26. Rajput SR, Ravals MS. A review on end-to-end methods for brain tumor segmentation and overall survival prediction. *Azerbaijan J High Perform Comput*.. (2020) 3(1):119–38.

27. Kocher M, Ruge MI, Galldiks N, Lohmann P. Applications of radiomics and machine learning for radiotherapy of malignant brain tumors. *Strahlenther Onkol*.. (2020) 196(10):856–67.

28. Shen G, Wang R, Gao B, Zhang Z, Wu G, Pope W. The MRI features and prognosis of gliomas associated with *IDH*1 mutation: A single center study in southwest china. *Front Oncol*. (2020) 10:852. doi: 10.3389/fonc.2020.00852

29. Li X, Zhu Y, Kang H, et al. Glioma grading by microvascular permeability parameters derived from dynamic contrast-enhanced MRI and intratu moral susceptibility signal on susceptibility weighted imaging. *Cancer Imaging*.. (2015) 15(1):4.

30. Hsu CCT, Watkins TW, Kwan GNC, Haacke EM. Susceptibility-weighted imaging of glioma: Update on current imaging status and future directions. *J Neuroimaging*. (2016) 26(4):383–90.

31. Smits M, van den Bent MJ. Imaging correlates of adult glioma geno types. *Radiology*. (2017) 284(2):316–31.

32. Jain R, Griffith B, Alotaibi F, et al. Glioma angiogenesis and perfusion imaging: Understanding the relationship between tumor blood volume and leakiness with increasing glioma grade. *Am J Neuroradiol*. (2015) 36(11):2030–5.

33. Su CQ, Lu SS, Han QY, Zhou MD, Hong XN. Intergrating conventional MRI, texture analysis of dynamic contrast-enhanced MRI, and susceptibility weighted imaging for glioma grading. *Acta Radiol*. (2019) 60(6):777–87.

34. Yu J, Wang M, Song J, Huang D, Hong X. Potential utility of visually AcceS able rembrandt images assessment in brain astrocytoma grading. *J Comput Assist Tomogr*.. (2016) 40(2):301–6.

35. Rong JD, Ding F, Li J, Zhu GY, He Z. Correlation of influencing factors with the extent of peritumoral brain glioma edema. *Zhonghua Yi Xue Za Zhi*. (2013) 93:169–71.

36. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, De Monte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection and survival. *JNeurosurg*. (2001) 95:190–8.

37. Olar A, Wani KM, Alfaro-Munoz KD, Heathcock LE, van Thuijl HF, Gilbert MR, et al. *IDH* mutation status and role of WHO grade and mitotic index in overall survival in grade II–III diffuse gliomas. *Acta Neuropathol*.. (2015) 129(4):585–96.

38. Zhang B, Chang K, Ramkissoon S, Tanguturi S, Bi WL, Reardon DA, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro Oncol*. (2017) 19(1):109–17.

39. Nakae S, Murayama K, Sasaki H, et al. Prediction of genetic subgroups in adult supra tentorial gliomas by pre- and intraopera tive parameters. *J Neurooncol*. (2017) 131:403–12.

40. Lasocki A, Tsui A, Gaillard F, Tacey M, Drummond K, Stuckey S. Reliability of noncontrast-enhancing tumor as a biomarker of *IDH*1 mutation status in glioblastoma. *J Clin Neurosci*. (2017) 39:170–5.

41. Wang K, Wang Y, Fan X, Wang J, Li G, Ma J, et al. Radiological features combined with *IDH*1 status for predicting the survival outcome of glioblastoma patients. *Neuro Oncol*. (2016) 18:589–97.

42. Qi S, Yu L, Li H, Ou Y, Qiu X, Ding Y, et al. Isocitrate dehydrogenase mutation is associated with tumor location and magnetic resonance imaging characteristics in astrocytic neoplasms. *Oncol Lett*. (2014) 7:1895–902.

43. Carrillo JA, Lai A, Nghiemphu PL, Kim HJ, Phillips HS, Kharbanda S, et al. Relationship between tumor enhancement, edema, *IDH*1 mutational status, MGMT promoter methylation, and survival in glioblastoma. *AJNR Am J Neuroradiol*. (2012) 33:1349–55.

44. Delfanti RL, Piccioni DE, Handwerker J, et al. Imaging correlates for the 2016 update on WHO classification of grade II/III gliomas: implications for *IDH*, 1p/19q and ATRX status. *J Neurooncol*. (2017) 135:601–60.

45. Yamashita K, Hiwatashi A, Togao O, Kikuchi K, Hatae R, Yoshimoto K, et al. MR imaging-based analysis of glioblastoma multiforme: Estimation of *IDH*1 mutation status. *Am J Neuroradiol*. (2016) 37:58–65.

46. Kickingereder P, Sahm F, Radbruch A, Wick W, Heiland S, Deimling AV, et al. *IDH* mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma. *Sci Rep*. (2015) 5:16238.

47. Park YW, Han K, Ahn SS, Bae S, Choi YS, Chang JH, et al. Prediction of *IDH*1-mutation and 1p/19q-codeletion status using preoperative MR imaging phenotypes in lower grade gliomas. *Am J Neuroradiol*. (2018) 39:37–42.

48. Patel SH, Poisson LM, Brat DJ, Zhou Y, Cooper L, Snuderl M, et al. T2–FLAIR mismatch, an imaging biomarker for *IDH* and 1p/19q status in lower-grade gliomas: A TCGA/TCIA project. *Clin Cancer Res*. (2017) 23:6078–86.

49. Villanueva-Meyer JE, Wood MD, Choi B, Mabray MC, Butowski NA, Tihan T, et al. MRI features and *IDH* mutational status in grade II diffuse gliomas: Impact on diagnosis and prognosis. *AJR Am J Roentgenol*.. (2018) 210:621.

50. Xing Z, Yang X, She D, Lin Y, Zhang Y, Cao D. Noninvasive assessment of *IDH* mutational status in world health organization grade II and III astrocytomas using DWI and DSC-PWI combined with conventional MR imaging. *AJNR Am J Neuroradiol*. (2017) 38:1134–44.

51. Leu K, Ott GA, Lai A, Nghiemphu PL, Pope WB, Yong WH, et al. Perfusion and diffusion MRI signatures in histologic and genetic subtypes of WHO grade II–III diffuse gliomas. *J Neurooncol*. (2017) 134:177–88.

52. Xiong J, Tan W, Wen J, Pan J, Wang Y, Zhang J, et al. Combination of diffusion tensor imaging and conventional MRI correlates with isocitrate dehydrogenase1/2 mutations but not 1p/19q genotyping in oligodendroglial tumours. *Eur Radiol*. (2016) 26:1705–15.

53. Wasserman JK, Nicholas G, Yaworski R, Wasserman AM, Woulfe JM, Jansen GH, et al. Radiological and pathological features associated with *IDH*1-R132H mutation status and early mortality in newly diagnosed anaplastic astrocytic tumours. *PloS One*. (2015) 10:e0123890.

54. Lee S, Choi SH, Ryoo I, et al. Evaluation of the micro-environmental heterogeneity in high-grade gliomas with *IDH*1/2 gene mutation using histogram analysis of diffusion-weighted imaging and dynamic-susceptibility contrast perfusion imaging. *J Neurooncol*. (2015) 121:141–50.

55. Merkaj S, Bahar RC, Zeevi T, Lin M, Ikuta I, Bousabarah K, et al. Machine learning tools for image-based glioma grading and the quality of their reporting: Challenges and opportunities. *Cancers*. (2022) 14:2623.

56. Rapisarda A, Melillo G. Overcoming disappointing results with antiangiogenic therapy by targeting hypoxia. *Nat Rev Clin Oncol*. (2012) 9:378–90.

57. Beig N, Patel J, Prasanna P, Hill V, Gupta A, Correa R, et al. Radiogenomic analysis of hypoxia pathway is predictive of overall survival in glioblastoma. *Sci Rep*. (2018) 8:7.

58. Kickingereder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer H-P, et al. Radiomic profiling of glioblastoma: Identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology*. (2016) 280:880–9.

59. McGarry SD, Hurrell SL, Kaczmarowski AL, Cochran EJ, Connelly J, Rand SD, et al. Magnetic resonance imaging-based radiomic profiles predict patient prognosis in newly diagnosed glioblastoma before therapy. *Tomography*. (2016) 2:223–8.

60. Liu X, Li Y, Qian Z, Sun Z, Xu K, Wang K, et al. A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas. *NeuroImage Clin*. (2018) 20:1070–7.

61. Zhou H, Vallières M, Bai HX, Su C, Tang H, Oldridge D, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol*. (2017) 19:862–70.

62. Kickingereder P, Götz M, Muschelli J, Wick A, Neuberger U, Shinohara RT, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clin Cancer Res*. (2016) 22:5765–71.

63. Grossmann P, Narayan V, Chang K, Rahman R, Abrey L, Reardon DA, et al. Quantitative imaging biomarkers for risk stratification of patients with recurrent glioblastoma treated with bevacizumab. *Neuro Oncol*. (2017) 19:1688–97.

frontiers | Frontiers in Oncology

TYPE Mini Review
PUBLISHED 30 January 2025
DOI 10.3389/fonc.2025.1497195

OPEN ACCESS

EDITED BY
Domenico Aquino,
IRCCS Carlo Besta Neurological Institute Foundation, Italy

REVIEWED BY
Chuanwei Wang,
Shandong University, China
Ayse Gul Eker,
Kocaeli University, Türkiye
Guillermo Iglesias,
Universidad Politécnica de Madrid, Spain
Widi Hastomo,
ITB Ahmad Dahlan, Indonesia

*CORRESPONDENCE
Aurélien Corroyer-Dulmont
a.corroyer-dulmont@baclesse.unicancer.fr

RECEIVED 16 September 2024
ACCEPTED 07 January 2025
PUBLISHED 30 January 2025

CITATION
Moreau NN, Valable S, Jaudet C, Dessoude L, Thomas L, Hérault R, Modzelewski R, Stefan D, Thariat J, Lechervy A and Corroyer-Dulmont A (2025) Early characterization and prediction of glioblastoma and brain metastasis treatment efficacy using medical imaging-based radiomics and artificial intelligence algorithms. *Front. Oncol.* 15:1497195. doi: 10.3389/fonc.2025.1497195

Early characterization and prediction of glioblastoma and brain metastasis treatment efficacy using medical imaging-based radiomics and artificial intelligence algorithms

Noémie N. Moreau[1,2], Samuel Valable[2], Cyril Jaudet[1], Loïse Dessoude[3], Leleu Thomas[3], Romain Hérault[4], Romain Modzelewski[5,6], Dinu Stefan[3], Juliette Thariat[3,7], Alexis Lechervy[4] and Aurélien Corroyer-Dulmont[1,2]*

[1]Medical Physics Department, Centre François Baclesse, Caen, France, [2]Université de Caen Normandie, CNRS, Normandie Université, ISTCT UMR6030, GIP CYCERON, Caen, France, [3]Radiation Oncology Department, Centre François Baclesse, Caen, France, [4]UMR GREYC, Normandie Univ, UNICAEN, ENSICAEN, CNRS, Caen, France, [5]LITIS - EA4108-Quantif, University of Rouen, Rouen, France, [6]Nuclear Medicine Department, Henri Becquerel Center, Rouen, France, [7]ENSICAEN, CNRS/IN2P3, LPC UMR6534, Caen, France

Among brain tumors, glioblastoma (GBM) is the most common and the most aggressive type, and brain metastases (BMs) occur in 20%–40% of cancer patients. Even with intensive treatment involving radiotherapy and surgery, which frequently leads to cognitive decline due to doses on healthy brain tissue, the median survival is 15 months for GBM and about 6 to 9 months for BM. Despite these treatments, GBM patients respond heterogeneously as do patients with BM. Following standard of care, some patients will respond and have an overall survival of more than 30 months and others will not respond and will die within a few months. Differentiating non-responders from responders as early as possible in order to tailor treatment in a personalized medicine fashion to optimize tumor control and preserve healthy brain tissue is the most pressing unmet therapeutic challenge. Innovative computer solutions recently emerged and could provide help to this challenge. This review will focus on 52 published research studies between 2013 and 2024 on (1) the early characterization of treatment efficacy with biomarker imaging and radiomic-based solutions, (2) predictive solutions with radiomic and artificial intelligence-based solutions, (3) interest in other biomarkers, and (4) the importance of the prediction of new treatment modalities' efficacy.

KEYWORDS

Glioblastoma (GBM), machine learning (ML), brain tumors, artificial intelligence, treatment efficacy, medical imaging, radiotherapy

Introduction

Brain tumors are highly heterogeneous neoplasms not only from a histological point of view but also from an intratumor temporal and spatial perspective.

Despite treatments including surgery, chemotherapy and radiotherapy patients with brain tumors respond heterogeneously. The same treatment will result in different treatment outcomes. Treatment efficacy is currently evaluated using anatomical MRI several months after treatment initiation. Differentiating non-responders from responders as early as possible in order to tailor treatment in a personalized medicine fashion to optimize tumor control is the most pressing unmet therapeutic challenge.

In this review, we will provide an overview of current research on treatment response assessment for a very aggressive and brain tumor called glioblastoma (GBM) and for a frequent brain tumor: brain metastasis (BM). To provide a clear structure and taxonomy of the reviewed literature, we have categorized the studies into the following sections:

Introduction:

- Overview of brain cancer and the therapeutic challenge of early characterization and prediction of treatment response.

Early characterization of brain cancer treatment efficacy:

- Review of studies using functional imaging biomarkers with MRI, PET, and CT with intensity thresholding for early detection in the days after treatment initiation.

Prediction of treatment response:

- Brief introduction to radiomics and its potential in medical imaging and treatment response assessment.
- Studies utilizing radiomics for extracting quantitative features from clinical routine MRI as input for predicting treatment response in brain cancer patients before its initiation.
- Brief introduction to AI and its potential in medical imaging and treatment response assessment.
- Research on various machine learning (ML) algorithms [e.g., support vector machines (SVMs), random forests, and neural networks] and studies using deep learning (DL) techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers to predict treatment outcome before its initiation.

Challenges and future directions for assessment of new treatment efficacy.

Abbreviations: AI, artificial intelligence; AUC, area under the ROC curve; BM, brain metastasis; CNN, convolutional neural network; DL, deep learning; GBM, glioblastoma; ML, machine learning; SVM, support vector machine; UNETR, UNEt Transformer.

Following the structure of the research and the taxonomy above, in this review, we will firstly focus on early characterization, which involves evaluation shortly after treatment initiation and mainly relies on imaging biomarkers/readouts. We will then focus on the ability to predict treatment efficacy before its initiation using radiomics and new innovative approaches using artificial intelligence (AI) (Figure 1). AI aims to mimic human intelligence through algorithms executed in a computer environment. AI algorithms are increasingly being studied in the field of medical imaging, whether for image processing, diagnosis, or the prediction of patient prognosis (1). One of the benefits of AI is its ability to handle large datasets and extract relevant information that is difficult to obtain through human intelligence. For those reasons, more important focus was made on AI solutions.

Article selection methodology

Databases: We conducted a comprehensive search using multiple databases, including PubMed, Web of Science, and Google Scholar. The search terms used were "Artificial Intelligence," "radiomics," "brain cancer," "treatment response," "glioblastoma," "brain metastases," "prediction," "machine learning," "deep learning," and related synonyms.

Time frame: We considered articles published from January 2012 to the present to ensure the inclusion of the most recent advancements.

Selection process: Articles were selected based on their novelty, number of patients, unique or multicenter approaches, and relevance to the therapeutic challenge.

Study design: We included original research articles and review papers; no case studies were included.

Quality: Only peer-reviewed articles from reputable journals and conferences were considered to ensure the reliability and validity of the findings.

Data extraction: Relevant data, including study design, novelty of the AI techniques used, outcomes, and limitations, were extracted from each selected article.

Current management of brain tumors

GBM is the most common and most aggressive primary brain tumor. Despite treatments including surgical resection, radiotherapy, and chemotherapy, the overall survival remains low (survival median of 15 months) with a high rate of tumor recurrence (2). While GBM has an incidence of 3.22 per 100,000 (3), BMs affect 20% to 40% of cancer patients (4) and represent the most common primary tumor with an incidence 3 to 10 times higher than primary brain tumors (5). BMs occur more frequently in patients with melanoma, lung cancer, or breast cancer (70%, 40%, and 20%, respectively (6)). As for GBM, despite aggressive treatment with radiotherapy and surgery that often led to cognitive decline due to healthy brain tissue dose toxicity, the

10.3389/fonc.2025.1497195

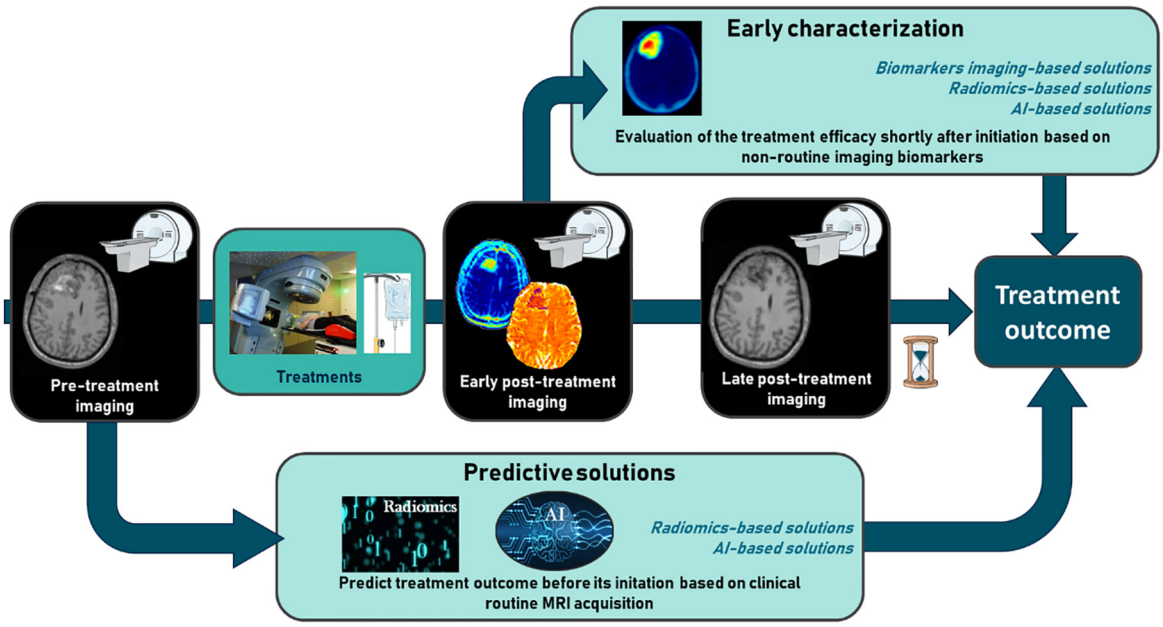


FIGURE 1
The challenge of early characterization in predicting therapeutic efficacy in glioblastoma and brain metastases.

survival median for patients with BM is very short and is about 6 to 9 months from the diagnosis of BM (7).

Therapeutic challenges

Patients with GBM (as well as patients with BM) present heterogeneous treatment responses (8). For the standard treatment (corresponding to surgery plus Stupp regimen), some GBM patients (a minority) are responders and present overall survival higher than 30 months and others are non-responders and die in few months (9). The pressing unmet therapeutic need is to be able to discriminate as soon as possible the non-responder patients from the responders to adapt treatment in a personalized medicine manner to optimize tumor control as well as healthy brain tissue preservation.

The process of evaluating therapeutic response is similar for GBM and BM. The assessment is mainly based on response evaluation criteria in solid tumors (RECIST) (10) and response assessment in neuro-oncology (RANO-BM) (11) criteria, which evaluate the evolution of lesion size on anatomical MRI, at different times after the treatment.

However, the issue is that assessment of the efficacy or non-efficacy of therapies, using conventional anatomical MRI, is only possible approximately 2 months after the beginning of treatment (12). Indeed, there is too much pseudoprogression or inflammatory response before and anatomical MRI is only able to reach the morphological aspect of the tumor. Focusing on other imaging biomarkers that are more specific to tumor biology could help shorten this wasted time, allowing for earlier assessment of treatment efficacy (13).

Subsections relevant for the subject

Early characterization of treatment efficacy

Biomarker imaging-based solutions

As shown in Table 1A, several publications have explored which imaging biomarkers might be more effective than anatomical MRI in predicting early therapeutic response (chemotherapy combined with anti-angiogenic therapy) and overall survival in patients with GBM and recurrent GBM at the clinical and preclinical level. Li and colleagues (14) have shown, on patients, that [^{18}F]-AlF-NOTA-PRGD2 PET/CT ([^{18}F]-RGD PET/CT) and dynamic contrast-enhanced MRI (DCE-MRI) can assess response to treatment, demonstrating that a greater decrease in SUV mean predicts better progression-free survival. Magnetic resonance spectroscopy (MRS) can predict early treatment efficacy. Talati et al. (15) performed a longitudinal MRI/MRS to study whether changes in N-acetylaspartate (NAA)/Choline (Cho) and Lactate (Lac)/NAA from different times after treatment can predict early therapy failures. Changes noted in metabolic levels of NAA/Cho and Lac/NAA were able to predict treatment failure as early as 1 day after anti-angiogenic treatment. This is in accordance with the review made by Qi and colleagues (16), who showed the different modalities and biomarkers that enable early characterization of therapeutic efficacy. At the preclinical level, Corroyer-Dulmont et al. have shown that [^{18}F]-fluoro-thymidine ([^{18}F]-FLT PET) (marker of cell proliferation), compared with other PET {[^{18}F]-fluorodeoxyglucose ([^{18}F]-FDG PET)} or MRI biomarkers, can characterize treatment efficacy from 3 days after treatment initiation, at a time when anatomical MRI shows no differences (17). Predicting treatment efficacy in recurrent GBM is also an important therapeutic challenge.

One clinical study and one preclinical study have shown the importance of using [^{18}F]-FLT PET to predict progression-free survival and overall survival in recurrent GBM (18, 19).

However, early characterization has a limitation. Even if it is effective, the patient has already undergone treatments (radiotherapy and chemotherapy) and may be exposed to their side effects (20).

The recent development of innovative computer techniques such as radiomics or more recently AI could lead to predict treatment effectiveness before its initiation. This will lead to a more personalized medicine where non-responder patients will gain precious months without undergoing an unnecessary costly treatment that could potentially lead to adverse effects (21).

Predictive solutions

Radiomic-based solutions

The term "radiomics" first appeared in literature in 2012 through an article published by Lambin et al. (22). This approach, focused on medical imaging data, aims to extract a large set of features from an image for a better characterization of tumor. Radiomic protocols require the following six steps: image acquisition, image reconstruction and pre-processing, segmentation, resampling, features extraction, and features selection and model-based feature construction (23). Because of these various steps, the use of radiomics aims to be potentially predictive compared to imaging biomarker analyses based on basic features such as mean or peak intensity. Image characteristics are subjected to a more in-depth analysis, making the features more relevant for prediction, and consequently, the results are more effective. Radiomics models are capable of predicting therapeutic response or overall survival (23).

In that context, the use of radiomics to develop models capable of predicting treatment response prior to brain tumor treatment initiation has been explored in several studies.

One of these studies (24) investigated the extraction of radiomic features from post-treatment MRI in patients with BM to predict local tumor control with an estimation of the tumor volume percentage compared to pre-treatment and overall survival with 256 and 237 patients, respectively. Three models were constructed through the training of SVMs using a Gaussian kernel and Bayesian optimization for hyperparameter tuning: (i) clinical features (age, gender overall survival, numbers of tumors, local tumor control, and median dose), (ii) radiomic features, and (iii) combined clinical and radiomic features. For both prediction objectives, the model combining clinical and radiomic features achieved very interesting performances with an area under the receiver operating characteristic curve (AUC) of 0.95 for local tumor control and 0.82 for overall survival.

Furthermore, a clinical study (25) was conducted to predict survival stratification of 125 patients with GBM. Radiomic features were extracted from MRI images. Among the three tested ML models, the SVM model demonstrated the best performance, with an AUC of 0.92.

Table 1B (24–28) summarizes several studies on the prediction of treatment response based on radiomics obtained from pre-treatment imaging. In all studies, the AUC is between 0.62 and 0.95. All these studies highlighting combining radiomic features with clinical features enhance prediction performance. However, radiomics has some limitations for routine clinical application. Most published studies have a relatively small patient cohort especially for GBM. However, to develop effective models, a sufficiently large training and test set is mandatory (29). Because of its complexity, radiomics presents the challenge of low interpretability of the features and models used, raising caution among physicians regarding the use of radiomics models in clinical settings (30). Beyond these points, the main limitation of radiomics remains the low stability and inter-hospital portability of the models (29). To resolve to this challenge, initiatives like the "Imaging Biomarker Standardization Initiative (IBSI)" (31) have been developed to complement radiomic features extraction; however, the robustness of these predictive models remains an issue before their adoption as a standard of care as shown by Peerlings and colleagues (32) for diffusion MRI or CT (33) or even for Test–Retest in PET imaging (34).

Therefore, it is timely to explore more innovative current developments in AI that may enable predictive characterization of treatment efficacy. DL is known to be able to extract more complex and a larger number of features in medical imaging than radiomics, which could lead to better performance (35).

Artificial intelligence-based solutions

Several studies have evaluated the use of AI algorithms to assess the therapeutic efficacy of GBM and BM. A clinical study (36) involving 124 patients with BM developed a CNN-based architecture to extract features from each MRI slice to predict the outcome of local control/failure in BM treated with stereotactic radiation therapy. A CNN is a type of DL neural network specifically designed to process structured data arrays, such as images. They integrated an InceptionResentV2 CNN architecture and a transformer (to consider spatial dependences between MRI slices during modeling). Depending on the mechanism of integration of information from each MRI slice, the AUC ranged from 0.72 to 0.86. The best performance was obtained with the combination of DL features obtained from anatomical MRI with clinical variables (tumor size, age, gender, tumor location, histology, total dose, previous WBRT, and number of BM).

In a study including 30 patients (15 with low-grade glioma and 15 with GBM), Vollmuth et al. (37) demonstrated that AI using artificial neural network (ANN) for brain and then tumor segmentations has the potential to provide a more reproducible and standardized assessment of treatment response on MRI compared to manual two-dimensional measurements of tumor burden using RANO criteria. Time to progression (TTP) was initially evaluated according to RANO criteria based on MRI and then revaluated by incorporating additional information from AI-enhanced MRI sequences that describe longitudinal changes in tumor volume. The inter-observer concordance correlation coefficient (CCC) for TTP measurements was 0.77 using the RANO criteria alone. With the addition of AI, the CCC increased to 0.91. This improvement was most observed in patients with low-grade gliomas (0.70 without AI vs. 0.90 with AI). Because of the less aggressive nature of these tumors, reliable assessment of TTP can be more difficult.

In a previous study, Luckett et al. (38) show a good performance with an accuracy of 90.6% in classifying survival (<1 year, 1–2 years, and >2 years), employing a deep feedforward CNN comprising three

10.3389/fonc.2025.1497195

TABLE 1 Biomarker imaging-based solutions for the early characterization of treatment efficacy (A), radiomic-based solutions for treatment efficacy prediction (B), and AI-based solutions for treatment efficacy prediction (C).

Table 1A Biomarker imaging-based solutions for the early characterization of treatment efficacy.								
Studies	Cohorts (*n*)	Tumor type	Treatment	Imaging modality	Imaging schedule	Outcome prediction	Results	Reference
Clinical	20 patients	GBM	Anti-angiogenic (Bevacizumab) plus conventional radiotherapy and chemotherapy (Temozolomide) (CRT) Adjuvant chemotherapy (Temozolomide) plus anti-angiogenic (Bevacizumab)	^{18}F-RGD PET/CT DCE-MRI	Before CRT Before anti-angiogenic Seven weeks after anti-angiogenic	Treatment efficacy	Prediction of response to treatment after 3 weeks	(14)
Clinical	33 patients	Recurrent GBM	Anti-angiogenic (Bevacizumab) monotherapy or combination therapy	MRI/MRS (NAA/Cho and Lac/NAA)	1 day and 2, 4, 8, and 16 weeks after treatment	Treatment efficacy	Prediction of treatment failure to therapy 1 day after treatment	(15)
Preclinical	25 rats and 29 rats	GBM (U87 and U251: human cell line)	Chemotherapy (Temozolomide), anti-angiogenic (Bevacizumab), or both	Anatomical MRI Diffusion MRI CBV MRI [^{18}F]-[FLT] PET [^{18}F]-FDG PET	5, 10, or 12 days after treatment	Treatment efficacy	[^{18}F]-FLT was more predictive: 3 days after initiation treatment	(17)
Preclinical	49 rats	Recurrent GBM (Human U251 cell line)	Chemotherapy (Temozolomide), anti-angiogenic (Bevacizumab), or both	Anatomical MRI Diffusion MRI CBV MRI [^{18}F]-[FLT] PET [^{18}F]-FDG PET	3, 10, and 17 days after treatment	Treatment efficacy	[^{18}F]-FLT was more predictive: 3 days after the end of treatment	(18)
Clinical	30 patients	Recurrent malignant glioma	Chemotherapy (Temozolomide) and anti-angiogenic (Bevacizumab)	Anatomical MRI [^{18}F]-FLT PET	MRI: 6 weeks after treatments PET: 1 to 5 days and at 2 and 6 weeks after treatments	Treatment efficacy	[^{18}F]-FLT can be used to determine the treatment efficacy 2 weeks after treatments	(19)

CBV, cerebral blood volume; Cho, choline; CRT, radiotherapy and temozolomide; CT, computed tomography; DCE, dynamic contrast-enhanced; [^{18}F]-FDG, [^{18}F]-fluorodeoxyglucose; [^{18}F]-FLT, [^{18}F]-fluoro-thymidine; GBM, glioblastoma; Lac, lactate; MRI, magnetic resonance imaging; MRS, magnetic resonance spectroscopy; NAA, N-acetylaspartate; PET, positron emission tomography; [^{18}F]-RGD, [^{18}F]-AlF-NOTA-PRGD2.

TABLE 1B Radiomic-based solutions for treatment efficacy prediction.

Studies	Cohorts (*n*)	Tumor type	Treatment(s)	Imaging modality	Features numbers	Models	Outcome prediction	Results	Reference
Clinical	237 patients	BM	Gamma knife radiosurgery (GKRS)	MRI	Clinical: 5 Radiomic: 4	SVM	Overall survival	Radiomics and clinical features combination (AUC = 0.82, Acc = 0.80, Sens = 0.77, Spe = 0.81)	(24)
Clinical	256 patients	BM	GKRS	MRI	Clinical: 5 Radiomics: 5	SVM	Local tumor control	Radiomics and clinical features combination (AUC = 0.95, Acc = 0.89, Sens = 0.87, Spe = 0.91)	(24)
Clinical	125 patients	GBM	Radiotherapy and concomitant chemotherapy (Temozolomide)	MRI	Clinical: 6 Radiomics: 21	RF, SVM, LR	Survival stratification	Radiomics and clinical features combination (AUC = 0.92)	(25)
Clinical	76 patients	GBM	Chemoradiotherapy	MRI	Clinical: 2 Radiomics: 6	Naïve Bayes	Distinction in early true progression	Radiomics and clinical features combination	(26)

(Continued)

10.3389/fonc.2025.1497195

TABLE 1B Continued

Studies	Cohorts (*n*)	Tumor type	Treatment(s)	Imaging modality	Features numbers	Models	Outcome prediction	Results	Reference
							between pseudoprogression	(AUC = 0.80, Acc = 0.737, Sens = 0.78, Spe = 0.67)	
Clinical	337 patients	BM	SRS	MRI	Clinical: 4 Radiomics: 223	GNB, kNN, RF, AB, SVM, MLP	Treatment response	Best classifier: SVM Radiomics and clinical features combination (AUC = 0.95)	(27)
Clinical	87 patients	BM	Stereotactic radiosurgery (SRS)	MRI	Clinical: 3 Radiomics: 9	RF	Local tumor control	Radiomics and clinical features combination (AUC = 0.79)	(28)

AB, adaptive boosting; Acc, accuracy; AUC, area under the ROC curve; BM, brain metastasis; GBM, glioblastoma; GKRS, gamma knife radiosurgery; GNB, Gaussian naïve Bayesian; kNN, k-nearest neighbors; LR, logistic regression; MLP, multilayer perceptron; MRI, magnetic resonance imaging; RF, random forest; Sens, sensitivity; Spe, specificity; SRS, stereotactic radiosurgery; SVM, support vector machine.

TABLE 1C AI-based solutions for treatment efficacy prediction.

Studies	Cohorts (*n*)	Tumor localization	Treatment	Imaging modality	Models	Outcome prediction	Results	Reference
Clinical	124 patients	BM	Stereotactic radiation therapy (SRT)	MRI	MLP/Clinical features CNN + Seq2Seq/ Transformers/LSTM CNN + Seq2Seq/ Transformers/LSTM + clinical features	Local tumor control	CNN + LSTM + clinical features (AUC = 0.86, Acc = 0.83, Sens = 0.77, Spe = 0.87)	(36)
Clinical	30 patients	Gliomas (15 GBM)	/	MRI	HD-GLIO-XNAT (https://github.com/NeuroAI-HD/HD-GLIO-XNAT)	Evaluate whether AI-assisted decision support provides a more reproducible and standardized assessment of response to treatment compared to manual measurements using RANO criteria	Lower-grade gliomas (CCP = 0.77 for RANO and 0.91 with AI)	(37)
Clinical	133 patients	GBM	/	MRI	ANN with clinical features	Survival classification	Cross validation: Acc = 0.91	(38)
Preclinical	28 mice	GL261	Chemotherapy (Temozolomide)	MRI/MRS	1D-CNN, LR, SVM, RF, XGBoost	Therapy response assessment	1D-CNN (Acc = 0.9975, Sens = 0.99, Spe = 0.99)	(39)

Acc, accuracy; AI, artificial intelligence; AUC, area under the ROC curve; BM, brain metastasis; CCP, concordance correlation coefficients; CNN, convolutional neural network; GBM, glioblastoma; LR, logistic regression; LSTM, long short-term memory; MLP, multilayer perceptron; MRI, magnetic resonance imaging; MRS, magnetic resonance spectroscopy; RANO, response assessment in neuro-oncology; RF, random forest; Sens, sensitivity; Spe, specificity; SRS, stereotactic radiosurgery; SVM, support vector machine; XGBoost, extreme gradient boosting.

hidden layers with eight neurons in each layer to predict patient survival in a cohort of 133 individuals. Ortega-Martorell and colleagues also showed a good performance of one-dimension CNN in a preclinical study to track therapy response in GBM (39). The 1D-CNN performed better than different ML models, showing the superiority of DL methods.

Our review of the literature reveals that the CNN exhibits superior performance. Although the architecture is not novel, it is particularly suited to medical imaging and currently offers the most effective means of predicting treatment efficacy (40).

Table 1C (36–39) summarizes several studies on the prediction of treatment response based on AI algorithms from pre-treatment MRI. As in the radiomics-based studies, the best performance is achieved by combining imaging data with clinical information.

Many studies applying AI in this field are based on relatively small data cohorts (less than 100 for GBM). However, a large data cohort is essential for optimal training of AI models (41). Centralizing a large amount of data in a single center can be challenging, and the performances of models are not always transferable between centers. Federated learning (42) addresses this issue by enabling learning from distributed data without transferring it between sites. Federated learning is a DL paradigm in which a model is trained across multiple decentralized devices or servers located in various medical centers, each holding local data samples, without the need to exchange the raw data. The only parameters shared among the different hospitals are the model parameters, not the raw medical data.

In addition, AI methodology is constantly evolving and new architectures appear every year. The models we have presented in this review give an overview of what is being done today, but new architectures such as diffusion models or full transformers should be more present in the years to come. One example is the UNEt TRansformer (UNETR) (43), which adapts the CNN encoder/decoder models proposed by UNET to transformer architectures in order to process sequential representations of the input volume more efficiently. Transformers are a type of AI model designed to efficiently process sequential data, such as text. Functional imaging such as proliferation index or other indicators is more relevant for assessing therapeutic efficacy (17). To our knowledge, no study involving AI models uses functional imaging biomarkers for predicting GBM efficacy as all the articles reported in this review used clinical routine anatomical MRI. However, in other cancers with radiomic models, Knuth and colleagues as well as Zhang and colleagues support the add value of function biomarkers in comparison to anatomical MRI in rectal (44) and breast cancers, respectively (45).

Opting for more functional imaging biomarkers instead of anatomical MRI could potentially improve AI performance in predicting treatment efficacy.

It is important to note that current studies were based on 2016 WHO classification rather than the 2021 one. To the best of our knowledge, no study has yet evaluated the potential of AI models to predict treatment outcomes of GBM according to the WHO 2021 classification. These models may not fully reflect current standards and advancements in the field, potentially leading to biases in predictions. However, current performances of the AI models to predict treatment outcome are still valid if they do not take into account the grade of the tumor, for example, if the input data only take the pre-treatment MRI. If the model is capable of predicting the treatment outcome of a brain lesion on an MRI, it should still be able to do so regardless of whether the brain lesion is designated as a GBM or a grade 4 astrocytoma Therefore, it is essential to incorporate recent classifications to ensure that AI models are aligned with best clinical practices and provide reliable and relevant recommendations.

AI models to distinguish pseudo-progression to recurrence

For patients with GBM treated in accordance with the established standard protocol, the prevalence of pseudoprogression is estimated to range between 20% and 30%. This phenomenon typically manifests within 1 to 12 weeks following the conclusion of treatment and is distinguished by an increase in tumor volume and the emergence of new lesions discernible on magnetic resonance imaging (MRI) (46).

This represents a significant challenge in clinical routine, as it complicates the assessment of treatment response and may impact therapeutic decision-making. Distinguishing between pseudoprogression and tumor recurrence is essential for optimal patient management, but this differentiation requires a significant amount of imaging. The acquisition of earlier information on potential pseudoprogression could enable treatment to be adapted more rapidly. Several studies have shown that radiomics and AI could be pertinent tools to predict pseudoprogression. Sun et al. (47) evaluated the diagnostic performance of ML models using a radiomic model based on contrast-enhanced T1-weighted MRI to differentiate pseudoprogression from true progression after standard treatment for 77 patients. The classifier demonstrated limited results with a sensitivity of 78.36% and a specificity of 61.33%. Another study (48), based on 78 patients with GBM, developed a CNN combined with an LSTM to differentiate anatomical MRI pseudoprogression from progression. The AUC results of the three trained models ranged from 0.52 to 0.83. The model that demonstrated the highest performance was the one that combined both MRI data and clinical features including age at the time of surgery, gender, methylation status of the 06-methylguanine-DNA-methyltransferase (MGMT) promoter, mutational status of the isocitrate dehydrogenase (IDH) gene, the total dose and number of fractions of radiotherapy, and other factors. Moassefi and colleagues (49) developed a DL model to distinguish pseudoprogression from true progression for 124 patients, using only clinical routine MRI. The model achieved a mean accuracy of 76.4%, a mean AUC of 0.76, a mean sensitivity of 88.72%, and a mean specificity of 62.05%.

An article using nuclear medicine imaging shows that radiomics based on FET-PET was able to differentiate tumor progression from pseudoprogression (50). Kebir et al. used FET-PET images in 14 patients and applied an unsupervised clustering algorithm for the diagnosis of pseudoprogression, achieving a diagnostic accuracy of 75%.

These studies demonstrate that it is possible to predict pseudoprogression at a relatively early stage, which could potentially optimize patient management. However, it is important to note that (1) performances of the models are limited with a specificity and a sensitivity of approximately 0.7 to 0.8 and (2) none of these studies have explored the prediction of pseudoprogression using pre-treatment imaging, highlighting a significant area for future research.

Interest in other biomarkers

This review focuses on the relevance of imaging biomarkers and the use of radiomics and AI based on MRI before and after treatment. However, molecular biomarkers can also be used to characterize therapeutic efficacy and overall survival. One such molecular biomarker is the methylation status of MGMT (51). The 1p/19q codeletion and loss of chromosome 10 are also predictive of therapeutic response (52). Although these biomarkers are used in routine clinical practice, the cost of testing, limited resources, and analysis time may be limiting factors for some patients (53). In contrast, MRI and RT DOSE are performed for each patient.

In addition, a biopsy is only performed on a part of the tumor. Since GBMs are recognized as highly heterogeneous tumors, molecular or protein expression will not be representative of the entire tumor, introducing a variability in the evaluation of therapeutic response (54). Therefore, imaging biomarkers appear to be the most suitable for routine clinical application.

New treatment modalities

Predicting the efficacy of treatments is of great interest for responder patients. However, for non-responder patients, the use of new treatment modalities, such as proton therapy and carbon ion therapy, is essential. It is important to conduct studies in these areas to assess the appropriateness of using one treatment over another, based on expected therapeutic efficacy. These studies are of crucial importance for the integration of these new treatments, which still need to be validated, especially through clinical trials (55). In this context, an AI tool that predicts treatment efficacy before initiation would be of significant interest.

Conclusion

The practical applications of AI and radiomics in the management of brain cancer are significant. These technologies enable earlier diagnosis, facilitating rapid and personalized treatment plans. For patients, this translates into better clinical outcomes and improved quality of life, particularly through the rapid identification of cases of non-response to treatment, paving the way to more appropriate therapeutic alternatives. As far as healthcare systems are concerned, AI and radiomics offer the possibility of optimizing the use of resources and reducing the financial impact of costly and ineffective treatments.

However, a number of challenges remain. These include the time and effort required to train healthcare professionals in the use of these technologies, as well as the management of administrative and regulatory obstacles.

The review highlights the pressing need for early and accurate characterization of treatment efficacy in GBM and BMs, given their aggressive nature and the heterogeneous responses to standard treatments. Current methods, relying on anatomical MRI, often fail to provide timely assessments due to pseudoprogression, leading to delayed treatment adjustments and potential cognitive decline from radiotherapy.

Early characterization of treatment efficacy

Imaging biomarkers, such as PET/CT, DCE-MRI, and MRS, have shown promise in predicting treatment response and overall survival earlier than conventional MRI. However, these methods still require patients to undergo initial treatments, exposing them to potential side effects.

Predictive solutions

Radiomics and AI offer innovative approaches to predict treatment efficacy before initiation. Studies combining radiomic features with clinical data have achieved high AUC values, indicating strong predictive performance. However, radiomics faces challenges such as low interpretability and limited inter-hospital portability, which initiatives like the IBSI aim to address.

Our review shows that AI, particularly DL techniques like CNNs, has demonstrated superior performance in predicting treatment outcomes. Combining AI-extracted features from MRI with clinical variables has yielded impressive results, with AUC values ranging from 0.72 to 0.99. Federated learning presents a solution to the challenge of data centralization, allowing models to be trained across multiple decentralized sites without exchanging raw data.

Challenges and future directions

Despite the promising results, several challenges remain. Most studies are based on small patient cohorts, which limits the generalizability of the findings. Additionally, the use of functional imaging biomarkers, which may provide more relevant information than anatomical MRI, has not been extensively explored in AI models for brain efficacy prediction. The integration of radiomics and DL in neuro-oncology has led to significant advancements in the management of gliomas, particularly by exploiting complex imaging features to predict molecular and clinical profiles. However, significant challenges remain, including the harmonization of multimodal data. Future research should focus on developing federated learning frameworks and enhancing model interpretability (56).

Pseudoprogression and new treatment modalities

Distinguishing pseudoprogression from true progression is crucial for optimal patient management. Radiomics and AI have shown potential in this area; however, the performance of these models is limited, and predicting pseudoprogression using pre-treatment imaging remains an inadequately explored area.

AI and the radiomics model have some limitations that have to be pointed out:

(a) Bias in training data or learning algorithms: Biases in training data represent a major challenge for training AI models. If the dataset used is not representative of the overall population, model performance is likely to degrade, particularly for more diverse patient groups. To limit these biases and better explain model behaviors, a data quality process is essential. This helps to identify and address potential gaps in the distribution of the data used.

(b) AI reliability in a clinical situation, especially with patient populations that are part of more heterogeneous groups: The reliability of AI systems in the clinical setting is a fundamental issue, especially when it comes to treating heterogeneous patient populations. For example, brain tumors such as GBM and BM present great heterogeneity both between tumors and within the same tumor. This diversity can limit the ability of AI models to generalize effectively. To address this, it is essential to rigorously

validate these models and continuously adapt them using updated data. In addition, the study of model explainability is essential to understand the decisions made by models.

Moreover, patients included in clinical trials are not representative of the general population of patients in clinical practice because the selection criteria are strict. Consequently, the results of most clinical trials do not allow the same conclusions to be drawn in a different population or context (57).

(c) The challenge of integrating new technology into day-to-day clinical practice: Integrating AI technologies into everyday clinical practice involves a number of challenges. Firstly, sufficiently powerful IT infrastructures are needed to run these models. Secondly, medical staff need to be trained in their use, which can come up against a certain resistance to change. In these cases, the explicability of the models plays a key role in instilling confidence and facilitating their adoption. In addition, it is crucial to develop user-friendly interfaces, integrating these models into practical tools for medical staff. Finally, regulatory and ethical aspects, such as data confidentiality and patient safety, must be considered to ensure the safe and responsible deployment of technologies in the clinical environment.

Articles cited in this review evaluate the performance of the AI models with specificity/sensitivity approaches and not with concrete data from clinical routine experiments or case studies on brain tumor treatment efficacy. A study has developed an AI model for diagnosing breast cancer and determined whether it could be useful to radiologists (58). The study showed that AI had better results than radiologists (91% vs. 59%). The integration of AI into clinical practice is raising new challenges while offering considerable opportunities. It is helping to improve the accuracy of diagnoses, optimize administrative tasks, and personalize treatment plans. Moreover, AI allows healthcare staff to spend more time with patients, enhancing the quality of care and the human relationship (59). For example, the authors showed that a BM segmentation system based on DL can be optimally applied to improve the efficiency of BM delineation in clinical practice (60). Another study has developed DL models for the purpose of proposing an alternative solution for patient-specific quality assurance that would make treatment machines more available to patients and thus enable more patients to be treated (61).

In summary, while significant progress has been made in early characterization and prediction of treatment efficacy in GBM and BM using imaging biomarkers, radiomics, and AI, further research is needed to address current limitations and explore new avenues. Integrating functional imaging biomarkers, updating AI models to reflect recent architecture, and investigating new treatment modalities are key areas for future development.

Author contributions

AC-D: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. NM: Writing – original draft, Writing – review & editing. SV: Writing – original draft, Writing – review & editing. CJ: Writing – original draft, Writing – review & editing. LD: Writing – original draft, Writing – review & editing. LT: Conceptualization, Writing – review & editing. RH: Writing – original draft, Writing – review & editing. RM: Writing – original draft, Writing – review & editing. DS: Writing – original draft, Writing – review & editing. JT: Writing – original draft, Writing – review & editing. AL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by the Région Normandie thought “Booster IA” grant. NM was supported by the Région Normandie.

Acknowledgments

We would like to thank Gilles Girault for his help with bibliographic research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

References

1. Bradshaw TJ, Huemann Z, Hu J, Rahmim A. A guide to cross-validation for artificial intelligence in medical imaging. *Radiol: Artif Intell*. (2023) 5:e220232. doi: 10.1148/ryai.220232

2. Tan AC, Ashley DM, López GY, Malinzak M, Friedman HS, Khasraw M. Management of glioblastoma: State of the art and future directions. *CA A Cancer J Clin*. (2020) 70:299–312. doi: 10.3322/caac.21613

3. Aftab K, Aamir FB, Mallick S, Mubarak F, Pope WB, Mikkelsen T, et al. Radiomics for precision medicine in glioblastoma. *J Neurooncol*. (2022) 156:217–31. doi: 10.1007/s11060-021-03933-1

4. Sas-Korczynska B, Rucinska M. WBRT for brain metastases from non-small cell lung cancer: for whom and when?—Contemporary point of view. *J Thorac Dis*. (2021) 13:3246–57. doi: 10.21037/jtd-2019-rbmlc-06

5. Vogelbaum MA, Brown PD, Messersmith H, Brastianos PK, Burri S, Cahill D, et al. Treatment for brain metastases: ASCO-SNO-ASTRO guideline. *JCO*. (2022) 40:492–516. doi: 10.1200/JCO.21.02314

6. Sacks P, Rahman M. Epidemiology of brain metastases. *Neurosurg Clinics North America*. (2020) 31:481–8. doi: 10.1016/j.nec.2020.06.001

7. Brown PD, Jaeckle K, Ballman KV, Farace E, Cerhan JH, Anderson SK, et al. Effect of radiosurgery alone vs radiosurgery with whole brain radiation therapy on cognitive function in patients with 1 to 3 brain metastases: A randomized clinical trial. *JAMA*. (2016) 316:401. doi: 10.1001/jama.2016.9839

8. Skaga E, Kulesskiy E, Fayzullin A, Sandberg CJ, Potdar S, Kyttälä A, et al. Intertumoral heterogeneity in patient-specific drug sensitivities in treatment-naïve glioblastoma. *BMC Cancer*. (2019) 19:628. doi: 10.1186/s12885-019-5861-4

9. Bjorland LS, Fluge O, Gilje B, Mahesparan R, Farbu E. Treatment approach and survival from glioblastoma: results from a population-based retrospective cohort study from Western Norway. *BMJ Open*. (2021) 11:e043208. doi: 10.1136/bmjopen-2020-043208

10. Wen PY, Weller M, Lee EQ, Alexander BM, Barnholtz-Sloan JS, Barthel FP, et al. Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro-Oncology*. (2020) 22:1073–113. doi: 10.1093/neuonc/noaa106

11. Lin NU, Lee EQ, Aoyama H, Barani IJ, Barboriak DP, Baumert BG, et al. Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol*. (2015) 16:e270–8. doi: 10.1016/S1470-2045(15)70057-4

12. Van Dijken BRJ, Van Laar PJ, Holtman GA, van der Hoorn A. Diagnostic accuracy of magnetic resonance imaging techniques for treatment response evaluation in patients with high-grade glioma, a systematic review and meta-analysis. *Eur Radiol*. (2017) 27:4129–44. doi: 10.1007/s00330-017-4789-9

13. Guglielmo P, Quartuccio N, Rossetti V, Celli M, Alongi P, Boero M, et al. 18F] fluorothymidine positron emission tomography imaging in primary brain tumours: A systematic review. *CMIR*. (2022) 18:363–71. doi: 10.2174/1573405617666210917123012

14. Li L, Liu N, Zhang H, Tao R, Zhao S, Chen Z, et al. Potential 18F-RGD PET/CT and DCE-MRI imaging-based biomarkers for postoperative survival prediction among patients with newly diagnosed glioblastoma treated with bevacizumab and chemoradiotherapy. *Front Oncol*. (2022) 12:848266. doi: 10.3389/fonc.2022.848266

15. Talati P, El-Abtah M, Kim D, Dietrich J, Fu M, Wenke M, et al. MR spectroscopic imaging predicts early response to anti-angiogenic therapy in recurrent glioblastoma. *Neuro-Oncol Adv*. (2021) 3:vdab060. doi: 10.1093/noajnl/vdab060

16. Qi D, Li J, Quarles CC, Fonkem E, Wu E. Assessment and prediction of glioblastoma therapy response: challenges and opportunities. *Brain*. (2023) 146:1281–98. doi: 10.1093/brain/awac450

17. Corroyer-Dulmont A, Pérès EA, Petit E, Guillamo JS, Varoqueaux N, Roussel S, et al. Detection of glioblastoma response to temozolomide combined with bevacizumab based on μMRI and μPET imaging reveals [18F]-fluoro-l-thymidine as an early and robust predictive marker for treatment efficacy. *Neuro-Oncology*. (2013) 15:41–56. doi: 10.1093/neuonc/nos260

18. Corroyer-Dulmont A, Pérès EA, Gérault AN, Savina A, Bouquet F, Divoux D, et al. Multimodal imaging based on MRI and PET reveals [18F]FLT PET as a specific and early indicator of treatment efficacy in a preclinical model of recurrent glioblastoma. *Eur J Nucl Med Mol Imaging*. (2016) 43:682–94. doi: 10.1007/s00259-015-3225-0

19. Schwarzenberg J, Czernin J, Cloughesy TF, Ellingson BM, Pope WB, Geist C, et al. 3′-deoxy-3′- ^{18} F-fluorothymidine PET and MRI for early survival predictions in patients with recurrent Malignant glioma treated with bevacizumab. *J Nucl Med*. (2012) 53:29–36. doi: 10.2967/jnumed.111.092387

20. Alemany M, Velasco R, Simó M, Bruna J. Late effects of cancer treatment: consequences for long-term brain cancer survivors. *Neuro-Oncol Pract*. (2021) 8:18–30. doi: 10.1093/nop/npaa039

21. Lehrer EJ, Jones BM, Dickstein DR, Green S, Germano IM, Palmer JD, et al. The cognitive effects of radiotherapy for brain metastases. *Front Oncol*. (2022) 12:893264. doi: 10.3389/fonc.2022.893264

22. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Stiphout RGPMV, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

23. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *Radiol Med*. (2021) 126:1296–311. doi: 10.1007/s11547-021-01389-x

24. Liao C-Y, Lee CC, Yang HC, Chen CJ, Chung WY, Wu HM, et al. Enhancement of radiosurgical treatment outcome prediction using MRI radiomics in patients with non-small cell lung cancer brain metastases. *Cancers*. (2021) 13:4030. doi: 10.3390/cancers13164030

25. Jia X, Zhai Y, Song D, Wang Y, Wei S, Yang F, et al. A multiparametric MRI-based radiomics nomogram for preoperative prediction of survival stratification in glioblastoma patients with standard treatment. *Front Oncol*. (2022) 12:758622. doi: 10.3389/fonc.2022.758622

26. Mouraviev A, Detsky J, Sahgal A, Ruschin M, Lee YK, Karam I, et al. Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery. *Neuro-Oncology*. (2020) 22:797–805. doi: 10.1093/neuonc/noaa007

27. Patel M, Zhan J, Natarajan K, Flintham R, Davies N, Sanghera P, et al. Machine learning-based radiomic evaluation of treatment response prediction in glioblastoma. *Clin Radiol*. (2021) 76:628.e17–27. doi: 10.1016/j.crad.2021.03.019

28. Du P, Liu X, Shen L, Wu X, Chen J, Chen L, et al. Prediction of treatment response in patients with brain metastasis receiving stereotactic radiosurgery based on pre-treatment multimodal MRI radiomics and clinical risk factors: A machine learning model. *Front Oncol*. (2023) 13:1114194. doi: 10.3389/fonc.2023.1114194

29. Hatt M, Rest CCL, Antonorsi N, Tixier F, Tankyevych O, Jaouen V, et al. Radiomics in PET/CT: current status and future AI-based evolutions. *Semin Nucl Med*. (2021) 51:126–33. doi: 10.1053/j.semnuclmed.2020.09.002

30. Zhang Y-P, Zhang XY, Cheng YT, Li B, Teng XZ, Zhang J, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. *Military Med Res*. (2023) 10:22. doi: 10.1186/s40779-023-00458-8

31. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2020) 295:328–38. doi: 10.1148/radiol.2020191145

32. Peerlings J, Woodruff HC, Winfield JM, Ibrahim A, Beers BEV, Heerschap A, et al. Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial. *Sci Rep*. (2019) 9:4800. doi: 10.1038/s41598-019-41344-5

33. Larue RTHM, Van De Voorde L, Van Timmeren JE, Leijenaar RTH, Berbée M, Sosef MN, et al. 4DCT imaging to assess radiomics feature stability: An investigation for thoracic cancers. *Radiother Oncol*. (2017) 125:147–53:2017. doi: 10.1016/j.radonc.2017.07.023

34. Leijenaar RTH, Carvalho S, Velazquez ER, Elmpt WJCV, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability. *Acta Oncol*. (2013) 52:1391–7. doi: 10.3109/0284186X.2013.812798

35. Wagner MW, Namdar K, Biswas A, Monah S, Khalvati F, Ertl-Wagner BB. Radiomics, machine learning, and artificial intelligence—what the neuroradiologist needs to know. *Neuroradiology*. (2021) 63:1957–67. doi: 10.1007/s00234-021-02813-9

36. Jalalifar SA, Soliman H, Sahgal A, Sadeghi-Naini A. Predicting the outcome of radiotherapy in brain metastasis by integrating the clinical and MRI-based deep learning features. *Med Phys*. (2022) 49:7167–78. doi: 10.1002/mp.15814

37. Vollmuth P, Foltyn M, Huang RY, Galldiks N, Petersen J, Isensee F, et al. Artificial intelligence (AI)-based decision support improves reproducibility of tumor response assessment in neuro-oncology: An international multi-reader study. *Neuro-Oncology*. (2023) 25:533–43. doi: 10.1093/neuonc/noac189

38. Luckett PH, Olufawo M, Lamichhane B, Park KY, Dierker D, Verastegui GT, et al. Predicting survival in glioblastoma with multimodal neuroimaging and machine learning. *J Neurooncol*. (2023) 164:309–20. doi: 10.1007/s11060-023-04439-8

39. Ortega-Martorell S, Olier I, Hernandez O, Restrepo-Galvis PD, Bellfield RAA, Candiota AP. Tracking therapy response in glioblastoma using 1D convolutional neural networks. *Cancers*. (2023) 15:4002. doi: 10.3390/cancers15154002

40. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. *Evol Intel*. (2022) 15:1–22. doi: 10.1007/s12065-020-00540-3

41. Sebastian AM, Peter D. Artificial intelligence in cancer research: trends, challenges and future directions. *Life*. (2022) 12:1991. doi: 10.3390/life12121991

42. Pati S, Baid U, Edwards B, Sheller M, Wang SH, Reina GA, et al. Federated learning enables big data for rare cancer boundary detection. *Nat Commun*. (2022) 13:7346. doi: 10.1038/s41467-022-33407-5

43. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, et al. UNETR: transformers for 3D medical image segmentation. *arXiv: arXiv:2103.10504*. (2021). http://arxiv.org/abs/2103.10504.

44. Knuth F, Tohidinezhad F, Winter RM, Bakke KM, Negård A, Holmedal SH, et al. Quantitative MRI-based radiomics analysis identifies blood flow feature associated to overall survival for rectal cancer patients. *Sci Rep*. (2024) 14:258. doi: 10.1038/s41598-023-50966-9

45. Zhang Q, Peng Y, Liu W, Bai J, Zheng J, Yang X, et al. Radiomics based on multimodal MRI for the differential diagnosis of benign and Malignant breast lesions. *Magnet Resonance Imaging*. (2020) 52:596–607. doi: 10.1002/jmri.27098

46. Young JS, Al-Adli N, Scotford K, Cha S, Berger MS. Pseudoprogression versus true progression in glioblastoma: what neurosurgeons need to know. *J Neurosurg*. (2023) 139:748–59. doi: 10.3171/2022.12.JNS222173

47. Sun Y-Z, Yan LF, Han Y, Nan HY, Xiao G, Tian Q, et al. Differentiation of pseudoprogression from true progressionin glioblastoma patients after standard treatment: A machine learning strategy combinedwith radiomics features from T1-weighted contrast-enhanced imaging. *BMC Med Imaging*. (2021) 21:17. doi: 10.1186/s12880-020-00545-5

48. Jang B-S, Jeon SH, Kim IH, Kim IA. Prediction of Pseudoprogression versus Progression using Machine Learning Algorithm in Glioblastoma. *Sci Rep*. (2018) 8:12516. doi: 10.1038/s41598-018-31007-2

49. Moassefi M, Faghani S, Conte GM, Kowalchuk RO, Vahdati S, Crompton DJ, et al. A deep learning model for discriminating true progression from pseudoprogression in glioblastoma patients. *J Neurooncol*. (2022) 159:447–55. doi: 10.1007/s11060-022-04080-x

50. Kocher M, Ruge MI, Galldiks N, Lohmann P. Applications of radiomics and machine learning for radiotherapy of Malignant brain tumors. *Strahlenther Onkol*. (2020) 196:856–67. doi: 10.1007/s00066-020-01626-8

51. Mansouri A, Hachem LD, Mansouri S, Nassiri F, Laperriere NJ, Xia D, et al. MGMT promoter methylation status testing to guide therapy for glioblastoma: refining the approach based on emerging evidence and current challenges. *Neuro-Oncology*. (2019) 21:167–78. doi: 10.1093/neuonc/noy132

52. Śledzińska P, Bebyn MG, Furtak J, Kowalewski J, Lewandowska MA. Prognostic and predictive biomarkers in gliomas. *IJMS*. (2021) 22:10373. doi: 10.3390/ijms221910373

53. Perrier A, Hainaut P, Lamy PJ, Guenoun A, Nguyen DP, Guerber F, et al. Utilisation clinique et évolution des biomarqueurs circulants à l'ère de l'oncologie personnalisée : des marqueurs protéiques aux scores clinicobiologiques. *Bull du Cancer*. (2022) 109:151–69. doi: 10.1016/j.bulcan.2021.11.010

54. Ronvaux L, Riva M, Coosemans A, Herzog M, Rommelaere G, Donis N, et al. Liquid biopsy in glioblastoma. *Cancers*. (2022) 14:3394. doi: 10.3390/cancers14143394

55. Thariat J, Habrand JL, Lesueur P, Chaikh A, Kammerer E, Lecomte D, et al. Apports de la protonthérapie à la radiothérapie d'aujourd'hui, pourquoi, comment? *Bull du Cancer*. (2018) 105:315–26. doi: 10.1016/j.bulcan.2017.12.004

56. Khalighi S, Reddy K, Midya A, Pandav KB, Madabhushi A, Abedalthagafi M. Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. *NPJ Precis Oncol*. (2024) 8:80. doi: 10.1038/s41698-024-00575-0

57. Monti S, Grosso V, Todoerti M, Caporali R. Randomized controlled trials and real-world data: differences and similarities to untangle literature data. *Rheumatol (Oxford)*. (2018) 57:vii54–8. doi: 10.1093/rheumatology/key109

58. Kim H-E, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health*. (2020) 2:e138–48. doi: 10.1016/S2589-7500(20)30003-0

59. Karalis VD. The integration of artificial intelligence into clinical practice. *Appl Biosci*. (2024) 3:14–44. doi: 10.3390/applbiosci3010002

60. Luo X, Yang Y, Yin S, Li H, Shao Y, Zheng D, et al. Automated segmentation of brain metastases with deep learning: A multi-center, randomized crossover, multi-reader evaluation study. *Neuro Oncol*. (2024) 26:2140–51. doi: 10.1093/neuonc/noae113

61. Boutry C, Moreau NN, Jaudet C, Lechippey L, Corroyer-Dulmont A. Machine learning and deep learning prediction of patient specific quality assurance in breast IMRT radiotherapy plans using Halcyon specific complexity indices. *Radiother Oncol*. (2024) 200:110483. doi: 10.1016/j.radonc.2024.110483

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 04 February 2025
DOI 10.3389/fonc.2025.1449911

OPEN ACCESS

EDITED BY
Vincenzo Di Nunno,
AUSL Bologna, Italy

REVIEWED BY
Palash Ghosal,
Sikkim Manipal University, India
Zhifan Gao,
Sun Yat-sen University, China
Shiping Ou,
Zhejiang University, China

*CORRESPONDENCE
Ruipeng Li
drlrp2022@163.com
Chen Song
mcgrady0712@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 16 June 2024
ACCEPTED 13 January 2025
PUBLISHED 04 February 2025

CITATION
Li R, Liao Y, Huang Y, Ma X, Zhao G, Wang Y and Song C (2025) DeepGlioSeg: advanced glioma MRI data segmentation with integrated local-global representation architecture. *Front. Oncol.* 15:1449911. doi: 10.3389/fonc.2025.1449911

DeepGlioSeg: advanced glioma MRI data segmentation with integrated local-global representation architecture

Ruipeng Li[1*†], Yuehui Liao[2†], Yueqi Huang[3†], Xiaofei Ma[2], Guohua Zhao[4], Yanbin Wang[1] and Chen Song[1*]

[1]Department of Urology, Hangzhou Third People's Hospital, Hangzhou, China, [2]College of Medical Technology, Zhejiang Chinese Medical University, Hangzhou, China, [3]Department of Psychiatry, Hangzhou Seventh People's Hospital, Hangzhou, China, [4]Department of Magnetic Resonance Imaging, First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Introduction: Glioma segmentation is vital for diagnostic decision-making, monitoring disease progression, and surgical planning. However, this task is hindered by substantial heterogeneity within gliomas and imbalanced region distributions, posing challenges to existing segmentation methods.

Methods: To address these challenges, we propose the DeepGlioSeg network, a U-shaped architecture with skip connections for continuous contextual feature integration. The model includes two primary components. First, a CTPC (CNN-Transformer Parallel Combination) module leverages parallel branches of CNN and Transformer networks to fuse local and global features of glioma images, enhancing feature representation. Second, the model computes a region-based probability by comparing the number of pixels in tumor and background regions and assigns greater weight to regions with lower probabilities, thereby focusing on the tumor segment. Test-time augmentation (TTA) and volume-constrained (VC) post-processing are subsequently applied to refine the final segmentation outputs.

Results: Extensive experiments were conducted on three publicly available glioma MRI datasets and one privately owned clinical dataset. The quantitative and qualitative findings consistently show that DeepGlioSeg achieves superior segmentation performance over other state-of-the-art methods.

Discussion: By integrating CNN- and Transformer-based features in parallel and adaptively emphasizing underrepresented tumor regions, DeepGlioSeg effectively addresses the challenges associated with glioma heterogeneity and imbalanced region distributions. The final pipeline, augmented with TTA and VC post-processing, demonstrates robust segmentation capabilities. The source code for this work is publicly available at https://github.com/smallboy-code/Brain-tumor-segmentation.

KEYWORDS

automated segmentation, glioma, CTPC, convolutional neural network, magnetic resonance imaging

1 Introduction

Brain tumors, also known as intracranial tumors in medical terminology, are abnormal masses of tissue characterized by uncontrolled cell growth and proliferation. According to the National Brain Tumor Society (1), gliomas account for approximately one-third of all brain tumors. Gliomas predominantly originate from glial cells, which surround and support the neurons in the cerebral cortex. These glial cells include ependymal cells, oligodendrocytes, and astrocytes. Gliomas put pressure on the brain or spinal cord, causing symptoms such as headaches, changes in personality, and weakness in the arms, etc. (2). They can disrupt brain function and pose a significant threat to an individual's life. The exact cause of gliomas remains unclear, and they can develop in all age groups, with a higher incidence observed in adults. Early detection and diagnosis of gliomas are critical to the effectiveness of treatment. Therefore, it is important to identify and diagnose gliomas in a timely manner to improve therapeutic outcomes.

In recent years, advances in medical imaging techniques such as positron emission tomography (PET), computed tomography (CT), and magnetic resonance imaging (MRI) have become increasingly important in the detection and diagnosis of disease. These different imaging modalities have the ability to identify distinct tumor regions within soft tissue (3). Typically, gliomas can be identified using a variety of MRI modalities, including T1-weighted (T1), T1-weighted with contrast enhancement (T1-CE), T2-weighted (T2), and T2-weighted fluid-attenuated inversion recovery (FLAIR). Each of these imaging modalities offers unique perspectives and insights into the properties of the tumor, resulting in different representations of the tumor on the images, as shown in Figure 1. After acquiring multimodal volumetric data of gliomas, a meticulous pixel-by-pixel segmentation process is applied to each individual slice until the entire 3D brain volume is accurately delineated into informative areas, establishing the ground truth (GT). The resulting segmentation output is then central to subsequent stages, including diagnosis, treatment planning, surgical strategies, and ongoing monitoring of tumor dynamics and changes.

In the search for valuable insights into brain tumors, radiologists have traditionally relied on manual segmentation of MRI images, using their expertise in anatomy and physiology (4). However, manual pixel-level segmentation of brain tumors by radiologists is a labor-intensive process (5). Radiologists face significant challenges during manual segmentation due to factors such as indistinct boundaries of gliomas, which includes peritumoral edema, necrotic cores, and tumor core enhancement. As a result, manual segmentation efforts by radiologists typically yield Dice scores in the range of 74% to 85% (6). Furthermore, manual segmentation is time-consuming, with radiologists spending 3–5 hours annotating an MRI scan for a single patient (7). Therefore, fully automated glioma segmentation methods are of paramount clinical importance and practical value (8).

Glioma segmentation faces significant challenges characterized by high heterogeneity and regional imbalances. First, high heterogeneity is evident in the wide variety of tumor shapes, structures, and locations. As shown in Figure 1, gliomas exhibit considerable inter-patient variability in structural characteristics, geometric configurations, and spatial distributions. This inherent variability poses a significant impediment to the accuracy of glioma segmentation. Consequently, an optimal model must effectively capture both local features (such as texture and edges) and global features (including shape, location, and structure) of gliomas. However, most existing convolutional neural networks (CNNs) focus primarily on extracting features at the local level, falling short of achieving a comprehensive representation.

Second, regional imbalance arises from the large size differences between the brain tumor, the background, and various tumor subregions. In the case of the BraTS2020 dataset, pixels within the tumor region represent only 1.1% of the total pixels. This tiny fraction of the tumor region may inadvertently cause the model to prioritize the background region, hindering accurate characterization of tumor features. Moreover, the proportions of each tumor subregion within the total tumor are significantly different (58%, 19.8%, and 22.2% for the whole tumor (WT), enhanced tumor (ET), and tumor core (TC), respectively). This unbalanced distribution among subregions

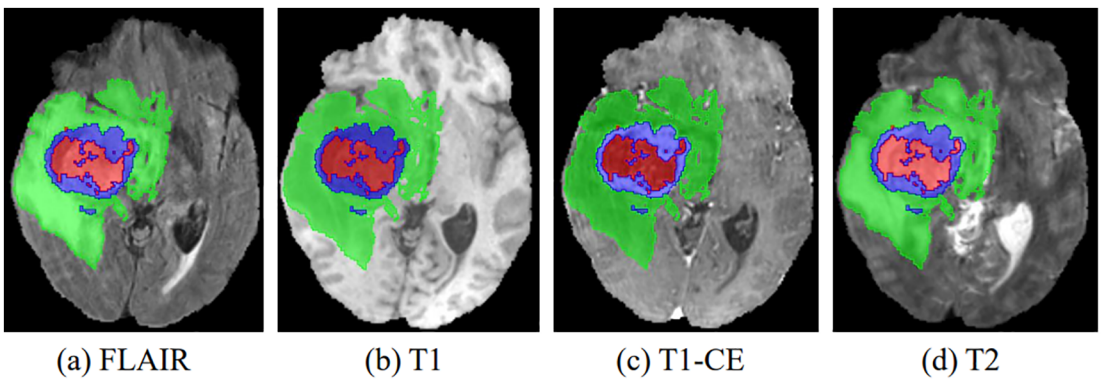


FIGURE 1
The four multimodal MR images with glioma tumors are: **(A)** FLAIR, **(B)** T1, **(C)** T1-CE, and **(D)** T2. Different colors are used to differentiate tumor subregions: red for necrotic and non-enhanced tumor (NCR/NET), green for peritumoral edema (ED), and blue for enhanced tumor (ET).

presents a substantial challenge for the model in classifying categories within these smaller proportions.

Many existing methods incorporate global information to address the challenges mentioned above. Typically, these methods use atrous convolution to expand the receptive field. However, in scenarios involving data types with smaller regions, such as brain tumors, atrous convolution may miss pixels, making it less suitable. A limited number of methods have used self-attention mechanisms to establish long-range dependencies. For example, Chen et al. (9) introduced a Parallel Self-Attention (PSSA) mechanism that transforms self-attention into a standard convolution operation on an appropriately transformed feature. This innovation effectively unifies self-attention and convolution. However, this approach diffuses local features into global features through layer stacking, which may dampen the performance of the method. Notably, the Transformer architecture excels at capturing global representations and requires fewer computational resources compared to traditional self-attention mechanisms (10). For example, Zhang et al. (11) proposed the parallel branched TransFuse network, which combines both Transformer and CNN architectures. This network includes a BiFusion module, consisting of spatial attention and channel attention, to facilitate feature fusion between the two branches. However, a limitation of this approach is the lack of fusion between the Transformer and CNN branches during the down sampling process, as these branches remain independent.

In the present study, we develop a unique DeepGlioSeg framework that enables glioma segmentation in multimodal MRI data. This network adopts a U-shaped architecture with skip connections, strategically used to support the continuous exploitation of contextual information. The DeepGlioSeg network introduces a central CTPC (CNN-Transformer Parallel Combination) module as its core component, comprising parallel branches for both CNN and Transformer networks. This innovative module facilitates the fusion of local and global features within glioma images through the collaborative interaction of these two branches. As a result, it effectively captures both the global and local features of gliomas, mitigating the challenges posed by the high heterogeneity in tumor shape, structure, and location. As shown in Figure 2, this module consistently outperforms convolution-based feature maps in accurately capturing the intricate shape and structure of gliomas. In addition, the DeepGlioSeg network employs a weighted loss approach to address the issue of region imbalance. It extends the generalized Dice loss to account for multiple regions and adjusts the contribution of each region with weighted values. Specifically, larger loss weights are assigned to categories associated with smaller regions, thereby increasing the focus on the tumor region.

We summarize our contributions as follows:

1. Our method introduces a CTPC module, which includes both a Transformer branch and a parallel CNN branch. This module facilitates the fusion of local and global features within glioma images through the interactive cooperation of these two branches, enhancing contextual relationships.
2. Our method assigns specific weights to each region based on the volume ratio of the region relative to the background. This weighting mechanism increases attention to the tumor region.
3. To evaluate the robustness of our algorithm, we curated a private brain tumor dataset consisting of data from 232 patients. Extensive experiments were performed on this private dataset, as well as on three publicly available datasets. The results consistently demonstrate the effectiveness of our proposed approach.

2 Related works

2.1 Brain tumor segmentation methods

Previous research on brain tumor segmentation in MR images can be categorized into (1) machine learning-based segmentation methods and (2) CNN-based segmentation methods. Machine learning-based methods have been adapted for brain tumor segmentation tasks, such as support vector machines (SVM) (12)

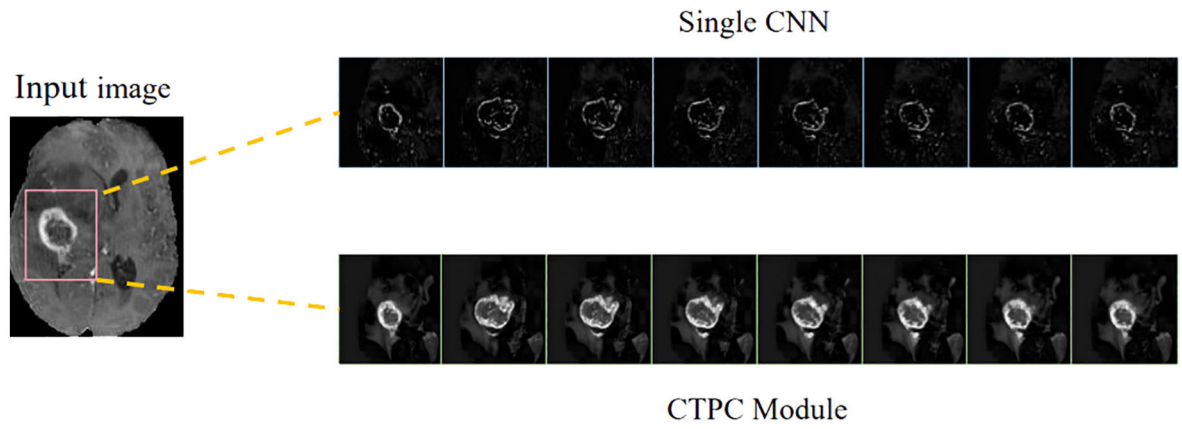


FIGURE 2
Comparison of feature maps between the CTPC module and CNN.

and random forests (13). For example, Bauer and colleagues (14) employed SVM classification methods in conjunction with hierarchical conditional random field regularization to improve the segmentation of brain tumor images.

CNNs have been extensively applied to brain tumor segmentation tasks, yielding remarkable results. For instance, spatial attention gates and channel attention gates were introduced into the U-Net network architecture by Xu et al. (15). Additionally, Xu et al. (15) developed a new FCN with a feature reuse module and a feature integration module (F2-FCN), enabling the extraction of more valuable features by reusing features from different layers. Shen and Gao (16) introduced a network with an encoding path that operates independently across channels and a decoding path focusing on feature fusion. This network utilizes self-supervised training and presents a novel approach to domain adaptation on the feature map, mitigating the risk of losing important restoration information within channels.

Cascaded methods have also emerged as a key research focus in brain tumor segmentation, achieving notable advancements through various strategies. For instance, Le Folgoc et al. (17) introduced lifted auto-context forests, a multi-level decision tree structure that optimizes segmentation via auto-context mechanisms. Wang et al. (18) proposed a cascaded anisotropic convolutional neural network, enhancing tumor edge and structure segmentation with anisotropic convolutional kernels. Lachinov et al. (19) iteratively refined segmentation results using a cascaded 3D U-Net variant, demonstrating its efficacy on the BraTS2018 dataset. Weninger et al. (20) designed a two-step approach with a 3D U-Net for tumor localization, followed by another for detailed segmentation into core, enhanced, and peritumoral edema regions. Finally, Ghosal et al. (21) developed a deep adaptive convolutional network with an adaptive learning mechanism that dynamically adjusts parameters, addressing the complexity of multimodal MRI.

2.2 Segmentation combined with CNN and transformer

The application of Transformer architecture to image segmentation has recently gained prominence, particularly in its integration with CNNs, a fusion that has yielded remarkable results in the field of medical image segmentation (9, 22–24). For example, Cao et al. (22) constructed a Transformer-based U-type skip connection encoder-decoder architecture called Swin-Unet. It is the first pure Transformer segmentation network and successfully demonstrates the applicability of transformers in the visual data domain. Building on Swin-Unet, more and more methods have begun to explore the fusion of Transformer and CNN. For instance, Hatamizadeh et al. (23) presented the architecture of UNet Transformer (UNETR), which uses a pure Transformer as the backbone for learning features in the encoding part, while only CNN is used in the decoding part.

Furthermore, not limited to Transformer, there has been increasing exploration of the application of self-attention mechanisms. For example, Chen et al. (9) theoretically derived a global self-attention approximation scheme that approximates self-attention by performing convolution operations on transformed features. Building on this, some approaches have developed multi-module structures that combine convolution and self-attention to integrate both local and non-local interactions. For instance, Petit et al. (24) presented the U-transformer model, which combines the U-type image segmentation structure with the self-attention and cross-attention mechanisms of the Transformer.

Recent advancements in hybrid CNN-Transformer architectures have significantly improved glioma segmentation by enhancing boundary precision and integrating local and global features. Gai et al. (25) proposed RMTF-Net, which combines ResBlock and mixed transformer features with overlapping patch embedding and a Global Feature Integration (GFI) module to improve decoding quality. Zhu et al. (26) developed a multi-branch hybrid Transformer that combines the Swin Transformer for semantic extraction and a CNN for boundary detection, incorporating a Sobel-based edge attention block to enhance tumor boundary preservation. Hu et al. (27) introduced ERTN, a dual-encoder model with a rank-attention mechanism to prioritize key queries, balancing performance and efficiency. These studies showcase diverse strategies for leveraging CNN-Transformer hybrids to address segmentation challenges, particularly in cases with complex tumor boundaries.

2.3 Category imbalance

A common problem in pixel-level semantic segmentation is class imbalance. This issue tends to reduce accuracy in regions belonging to the minority class (28, 29). For example, Hossain et al. (30) suggested that an effective way to address class imbalance is to adjust the loss function. They propose the bifocal loss function (DFL) to correct the problem of vanishing gradients in focal loss (FL). They introduce a regularization term to impose constraints on the negative class labels, which increases the loss for classes that are difficult to classify. Bressan et al. (31) used pixel-level weights in the training phase to dynamically adjust the importance of individual pixels, either increasing or decreasing their weight as needed. In other words, the contribution of each pixel in the loss function is weighted, which increases the importance of minority class pixels. Pan et al. (32) also faced the challenge of unbalanced foreground and background voxels when performing coronary segmentation. They use the concept of focal loss to optimize the network and achieve good results. To address the significant class imbalance problem observed in brain tumors, we follow the approach of the GDL loss function and assign more weight to small class regions, minimizing the model's focus on background regions.

3 Methodology

The proposed DeepGlioSeg framework consists of two phases: (1) the training phase, which includes data preprocessing, loss calculation, and parameter updating, and (2) the inference phase, which includes data preprocessing, learned model import, and postprocessing. A diagram summarizing this framework is shown in Figure 3.

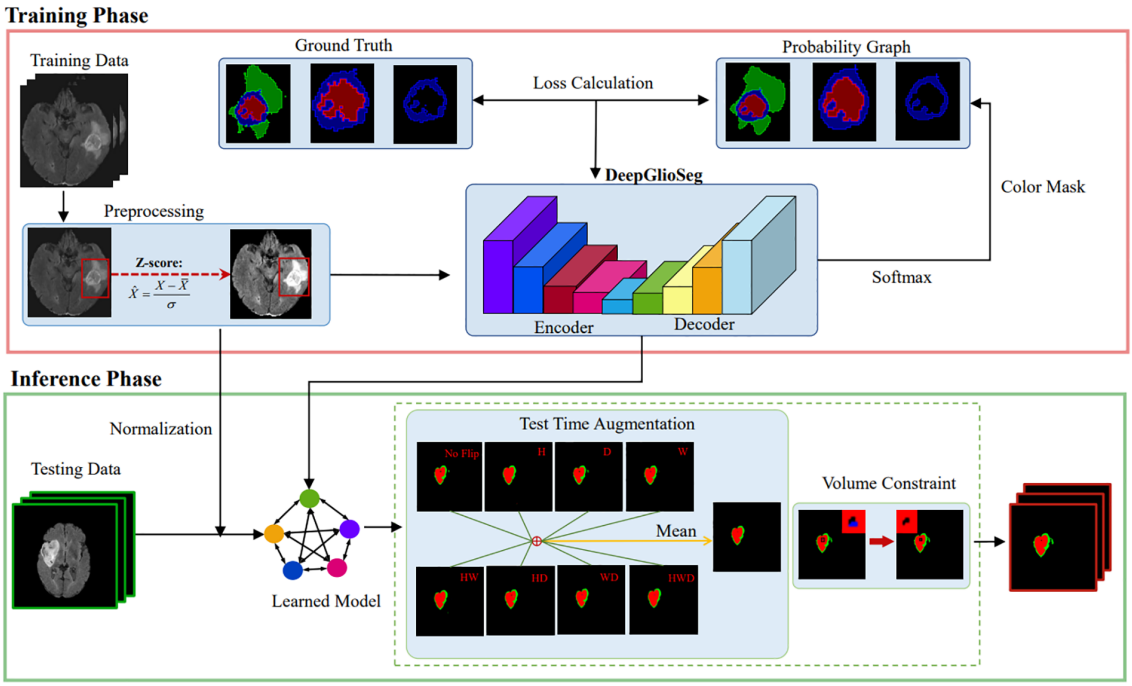


FIGURE 3
The diagram of the proposed DeepGlioSeg for automated glioma segmentation in multimodal MRI images.

3.1 Preprocessing

We used the BraTS and ZZH datasets, where each brain MRI scan includes FLAIR, T1, T1-CE, and T2 modalities, each with distinct eigenvalue distributions due to contrast differences. These variations pose challenges such as slower convergence and overfitting due to inconsistent intensity scales across modalities. To address these issues, we normalized voxel values within brain regions by subtracting the mean and dividing by the standard deviation. This standardization facilitated effective learning and mitigated convergence issues. Voxel values in non-brain regions were set to zero to eliminate interference from irrelevant background data. The normalization formula is as follows:

$$X' = \frac{X - \mu_B}{\sigma_B} \quad (1)$$

where X' represents the processed image, X symbolizes the original voxel value in the brain region, μ_B signifies the average intensity value of the brain region, and σ_B indicates the standard deviation of the brain region. This approach ensures that the model can focus on meaningful information while reducing variability caused by background noise. To further enhance robustness and generalization, we applied data augmentation techniques such as random rotations, flips, and elastic deformations. These augmentations prevent overfitting by exposing the model to diverse variations, improving its performance on unseen data in real-world clinical settings.

Four sets of modal sequences, each with a size of 240×240×155, were merged to obtain 4-channel 3D image data with a size of 240×240×155×4. Each training example has a corresponding label with a size of 240×240×155. The labels consist of four categories: background (label: 0), necrotic and non-enhanced tumor (label: 1), peritumoral edema (label: 2), and GD-enhanced tumors (label: 4). Finally, based on hardware and computational considerations, a training patch with a size of 128×128×128 was extracted from the training case.

3.2 DeepGlioSeg network architecture

The general design of DeepGlioSeg is shown in Figure 4A, which features a symmetric encoder-decoder architecture with skip connections. The basic concept revolves around the alternating stacking of CTPC modules and down sampling layers, combining local features with global representations at different resolution levels. Importantly, the CTPC module maintains consistent feature map sizes, while deconvolution gradually restores resolution. Throughout the network, all convolutional layers are complemented by batch normalization layers and ReLU activation functions. To mitigate overfitting, an initial convolutional layer with dropout functionality is included at the beginning of the model. Additionally, eight successive convolutional layers are implemented at the base of DeepGlioSeg to enhance feature extraction.

To manage the computational demands within the Transformer branch, the Feature Fusion Pathway (FFP) employs different down sampling steps corresponding to various resolution levels while maintaining a patch embedding size of 4096. It is important to note that the feature map input to the Transformer remains constant at 16×16×16.

10.3389/fonc.2025.1449911

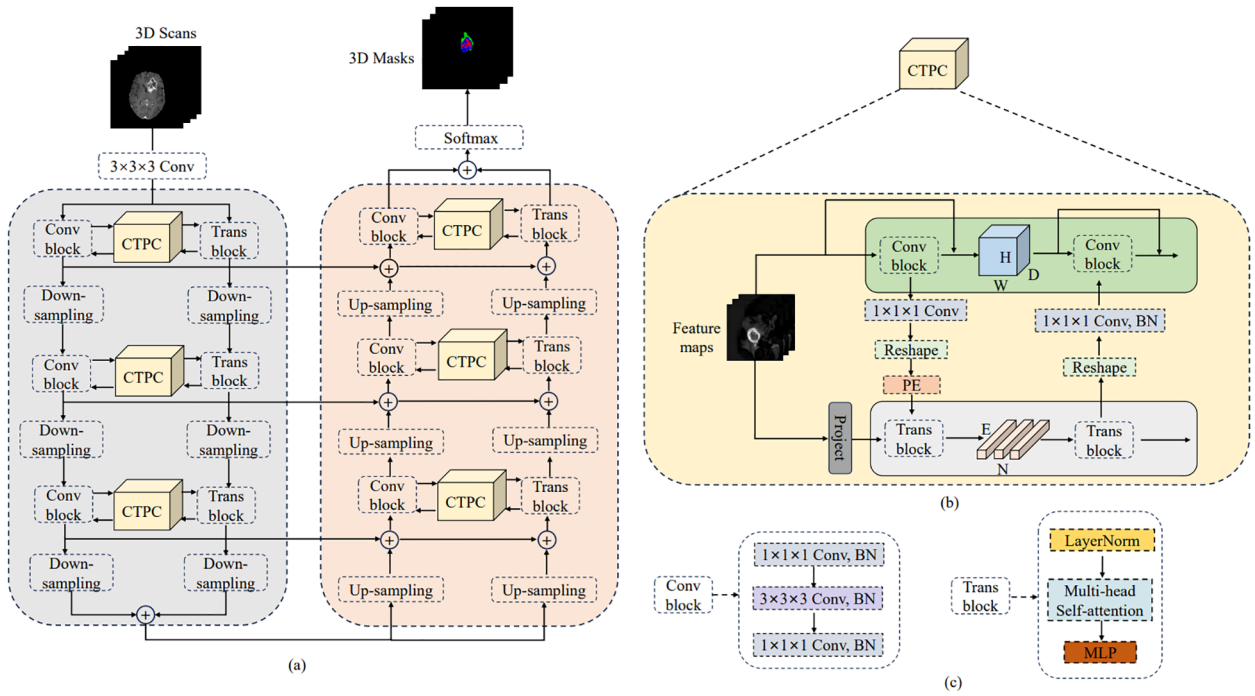


FIGURE 4
Overview of the proposed DeepGlioSeg network architecture: **(A)** The architecture of DeepGlioSeg. **(B)** The specific implementation steps of the CTPC module. **(C)** The detailed composition of the Conv block and Trans block.

3.3 CTPC module structure

In deep learning, CNNs collect local features at different resolutions by applying convolutional operations, effectively preserving local details as feature maps. Vision Transformers, on the other hand, are specifically designed to aggregate global representations by iteratively processing compressed patch embeddings through a series of self-attention modules. The CTPC module, as shown in Figure 4B, consists of three essential elements: the CNN module, the Transformer branch, and the FFP. These components are integrated to facilitate feature fusion between the two branches, effectively enhancing the feature extraction capabilities of the network.

3.3.1 CNN branch

As shown in Figure 4C, the CNN branch consists of two iterative convolution modules. Each module contains a sequence of a 1×1×1 downward convolution layer, a 3×3×3 spatial convolution layer, a 1×1×1 upward convolution layer, and a residual link connecting the module's input and output. While the Vision Transformer encodes image patches into word vectors, potentially leading to a loss of local detail, the CNN branch operates differently. In a CNN, the convolutional kernel glides over the neighborhood map, enabling it to extract continuous local features. This feature allows for the preservation of intricate and detailed local features to a significant extent. As a result, the CNN branch serves as a continuous supplier of local detail to the Transformer branch.

3.3.2 Transformer branch

The Transformer branch includes a multi-head self-attention module and a multi-layer perceptron (MLP) block, as shown in Figure 4C. Layer normalization is applied before both the multi-head self-attention module and the MLP block. Additionally, two residual connections are incorporated at corresponding positions. To balance computational efficiency and feature map resolution, the CNN branch output is down sampled to a 16×16×16 patch embedding.

3.3.3 Feature fusion path

The FFP functions to connect and align the shape disparity between the feature map in the CNN pathway and the patch embedding in the Transformer pathway. It actively promotes the continuous integration of local features with global representations through interactive mechanisms. Notably, the shape of the feature stream differs between the CNN and Transformer pathways. Specifically, the CNN feature map has a shape of $C \times H \times W \times D$, where C, H, W, and D denote the channel, height, width, and depth, respectively. In contrast, the patch embedding takes the form $E \times C$, where E is the embedding size and C is the number of image patches. Prior to inputting the feature map into the Transformer branch, channel alignment of the feature map and patch embedding is achieved by a 1×1×1 convolution. The volume dimensions are then compressed to 16×16×16 using the down sampling module, with different steps chosen for different resolution levels. Finally, the patch embedding is obtained via a reshape operation.

3.3.4 Position embedding

To capture essential positional information crucial for the segmentation task, we introduced learnable positional embeddings that are merged with the patch embedding by direct addition. When transitioning from the Transformer branch back to the CNN branch, it is necessary to upsample the patch embedding to restore it to the original shape of the CNN feature map. A 1×1×1 convolutional layer is then applied to harmonize the channel dimensions. Finally, the resulting output is combined with the feature map. Throughout this process, batch normalization is used to regulate the features.

3.4 *Loss function*

There is a significant data imbalance between tumor and non-tumor tissue for the purpose of identifying and delineating brain tumors and their subregions. Sudre et al. (33) noted that as the degree of data imbalance increases, the loss function based on overlap measurement is less susceptible to fluctuations compared to weighted cross-entropy. Therefore, the Dice coefficient was utilized to focus on different tumor subregions. The formula for the Dice coefficient is given by:

$$L_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g + \epsilon} \tag{2}$$

In this formula, g_i represents the ground truth label for pixel i in category c, and p_i denotes the predicted probability of pixel i belonging to category l. The term N represents the total number of pixels in the image, and ϵ is a small constant added to avoid division by zero, ensuring the stability of the loss function.

For multi-class segmentation tasks, a weight w_l is typically introduced based on the frequency of each category l. According to the statistical analysis of the proportion of each category, the weights were set to 0.1, 1, 2, and 2 for the background, WT, TC, and ET, respectively. The Multi-class Generalized Dice Loss (Multi-GDL) was then used as the model's loss function, which can be written as:

$$L_{\text{GDL}} = 1 - \frac{2\sum_{l=1}^{L} w_l \sum_{i=1}^{N} p_i^{(l)} g_i^{(l)} + \epsilon}{\sum_{l=1}^{L} w_l \left(\sum_{i=1}^{N} p_i^{(l)} + \sum_{i=1}^{N} g_i^{(l)}\right) + \epsilon} \tag{3}$$

In the above formula, L represents the total number of classes, and $p_i^{(l)}$ and $g_i^{(l)}$ denote the predicted probability and ground truth label for pixel i in class l, respectively. The weight w_l ensures that the contribution of each class is appropriately adjusted based on its frequency, addressing the issue of data imbalance.

3.5 Postprocessing

In the inference phase, the original image was sliced from left to right and from top to bottom into eight inference blocks of size 128×128×128 and post-processed with test-time augmentation (TTA) and volume-constraint (VC). For each inference block, seven different flips ((*x*), (*y*), (*z*), (*x*, *y*), (*x*, *z*), (*y*, *z*), (*x*, *y*, *z*)) were performed, as shown in Figure 5A. The flipped data were then fed into the model, and the corresponding inference results were obtained. The rotation angles of the different inference results were restored, and the average value was taken as the final output.

For VC, if the reference segmentation for the ET is missing, the BraTS evaluation assigns a reward of 0 for false positive predictions, and the Dice score is 1. Therefore, in this study, if the ET volume predicted by the model was less than the threshold of 500, the ET region was reclassified as necrotic and non-enhanced tumor tissue. As shown in Figure 5B, the ET region is replaced with necrotic and non-enhanced tumors after volume restriction.

4 Experimental setting

4.1 Dataset

Three public benchmark datasets and one private dataset were used to evaluate the effectiveness of the proposed DeepGlioSeg. The Brain Tumor Segmentation Challenge provided the BraTS2019, BraTS2020, and BraTS2021 datasets used in this study (6, 34, 35). The BraTS2019 dataset consists of 335 training cases and 125 validation cases. The BraTS2020 dataset contains 369 training examples and 125 validation examples. The BraTS2021 dataset includes 1251 training samples and 219 validation samples.

The ZZH dataset was collected by the first affiliated hospital of Zhengzhou University with institutional review board approval (reference number: 2019-KY-231). It consists of 232 patient records in the same format as the BraTS datasets. Each sample was manually labeled by two radiologists at the first affiliated hospital of Zhengzhou University. The dataset was split into training, validation, and test sets in a 7:1:2 ratio. The training set was used to train the model, the validation set was used to guide hyperparameter tuning and early stopping, and the test set was used to evaluate generalization. This approach prevents data leakage and ensures an unbiased performance evaluation.

4.2 Evaluation metrics

The Dice similarity coefficient (Dice), Sensitivity (Sen), and Hausdorff distance (Haus95) are used to assess the segmentation performance of the model. Considering the glioma's anatomical features and structure, the model's performance in segmenting the following three tumor sub-regions is evaluated: WT (necrotic and non-enhanced tumor, peritumoral edema, and enhanced tumor), TC (necrotic and non-enhanced tumor, enhanced tumor), and ET (enhanced tumor).

4.3 Experimental details

The optimization method used is Adam with a learning rate of 0.0002, and training is performed with a batch size of 4. DeepGlioSeg is trained for approximately 1000 iterations. A minimum loss value threshold is set, and the average loss value of each epoch is calculated during training. Training is stopped when

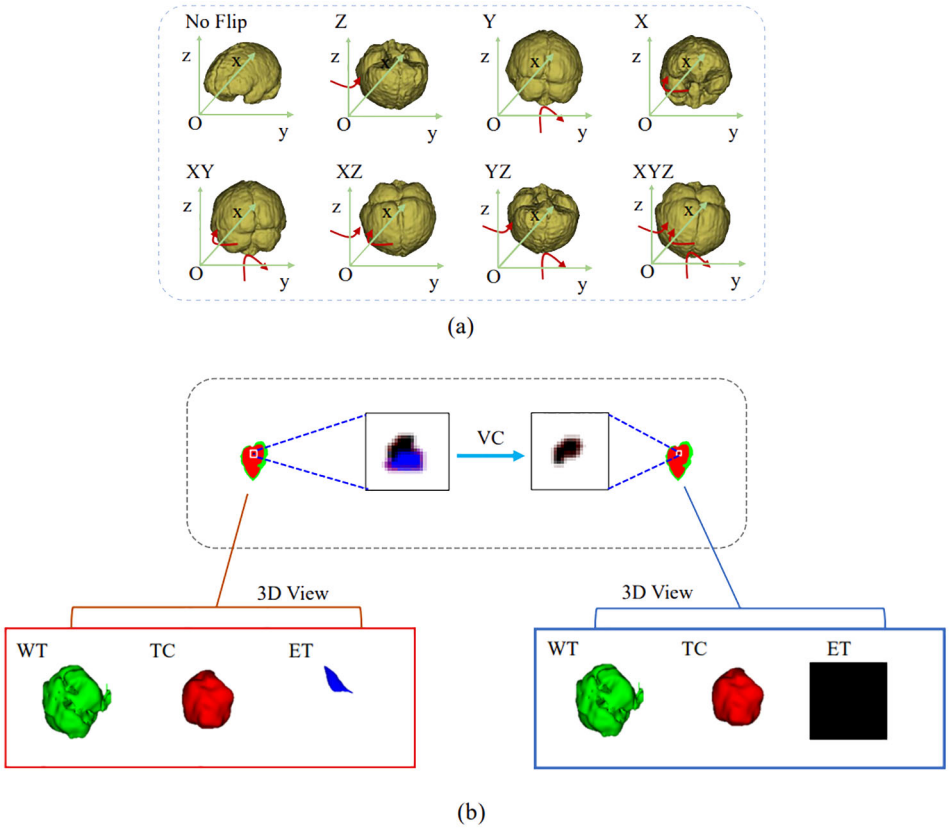


FIGURE 5
Results obtained by the postprocessing: **(A)** TTA; **(B)** VC. TTA was performed using 7 different flips: (x), (y), (z), (x, y), (x, z), (y, z), (x, y, z). VC replaces the ET predicted by the model if its volume is below the 500-voxel threshold.

the loss value drops below the set threshold. The size of the image input to the model is 128×128×128×4.

After each downsampling layer, the size of the feature map is halved, and the number of channels is doubled. The number of initial convolution kernels is 16. The loss weights for the four regions—ET, WT, TC, and background—are set to [2, 1, 2, 0.1]. The following data augmentation techniques are applied: (1) random cropping of the data from 240×240×155 to 128×128×128; (2) random mirroring and rotation in the axial, coronal, and sagittal planes with a probability of 0.5; and (3) random intensity shifts in the range [-0.1, 0.1] and scaling factors in the range [0.8, 1.2]. The network is trained using Multi-GDL, and L2 normalization is applied to regularize the model, with the weight decay rate set to 1e-5.

5 Results

5.1 Ablation study

5.1.1 CTPC module configuration

We conducted experiments to identify the optimal configuration of the CTPC module for better segmentation performance. Table 1 summarizes the impact of different CTPC configurations on model performance. The baseline model employs a standard encoder-decoder architecture with separate CNN and Transformer branches, which independently extract local and global features. Although effective individually, the lack of integration between these branches limits the model's ability to combine local and global information, reducing segmentation accuracy.

DeepGlioSeg addresses this by incorporating the CTPC module, which enables the simultaneous fusion of local and global features. Unlike the baseline, where features are processed separately, the CTPC module integrates the outputs from both branches, fusing local and global features into a unified representation. This enhanced feature fusion improves segmentation accuracy, particularly for complex and heterogeneous tumor regions. The fully embedded CTPC model achieved a 4.5% improvement in Dice score (84.3% vs. 79.8%) over the baseline, demonstrating the effectiveness of this integration.

The CTPC module addresses challenges in feature alignment and compatibility between CNN and Transformer outputs. By using 1×1×1 convolutions for channel alignment, it ensures CNN features match the Transformer input dimensions, preserving local detail while facilitating global feature integration. Downsampling reduces the volume dimensions to 16×16×16, balancing computational efficiency with feature richness for global context. The final reshaping generates patch embeddings that facilitate effective local-global interaction, making the CTPC module highly effective for capturing complex patterns, crucial for tumor segmentation tasks.

TABLE 1 Qualitative comparison of results on the BraTS2020 dataset, including the model architecture without the CTPC module (Baseline), encoding path configuration (EPC), decoding path configuration (DPC), and encoding-decoding path configuration (EDPC).

	Dice				Sen				Haus95			
	ET	WT	TC	Mean	ET	WT	TC	Mean	ET	WT	TC	Mean
Baseline	0.770	0.895	0.728	0.798	0.783	0.899	0.695	0.792	41.7	6.25	26.1	24.7
EPC	0.753	0.892	0.854	0.835	0.775	0.906	0.770	0.817	32.8	6.80	12.9	17.5
DPC	0.753	0.895	0.858	0.835	0.781	0.914	0.799	0.831	36.5	6.88	10.1	17.8
EDPC	0.768	0.897	0.865	0.843	0.785	0.911	0.812	0.836	27.1	5.92	9.94	14.3

Red denotes the best results, and blue means the second best.

5.1.2 Learnable position embedding

In the work of Dosovitskiy (36), a learnable embedding was incorporated into the embedded patch sequence and complemented with position embeddings to preserve critical positional information. Similarly, for glioma segmentation, we introduced a learnable position embedding to encode crucial positional information for the task. Within the CTPC module, the CNN and Transformer branches enable the fusion of feature streams through a shared pathway. Before passing the CNN feature stream into the Transformer branch, we used standard one-dimensional learnable position embeddings to encode position information. The embeddings were then added to the feature map via summation. As shown in Table 2, the introduction of learnable position embeddings improved the average Dice score by 1% (84.3% vs. 83.3%).

5.1.3 Strided convolution

As shown in Figure 3, downsampling the feature map from the CNN branch is necessary to achieve spatial dimension alignment. Peng et al. (37) used average pooling in the feature coupling unit for this purpose. However, pooling can filter out valuable information during downsampling. To mitigate this, we chose strided convolution as the downsampling module. Strided convolution enables multiple downsampling steps while facilitating further feature extraction by adjusting the step size. The network uses four resolution levels (128, 64, 32, 16) from top to bottom, with downsampling modules having step sizes of 8, 4, 2, and 1, respectively. To ensure computational consistency, we maintained the patch embedding size in the Transformer branch at 4096. As shown in Table 2, using strided convolution as the downsampling module within the FFP improved the average Dice score by 0.7% (84.3% vs. 83.6%).

5.1.4 Postprocessing

During inference, we used a dual post-processing approach involving TTA and VC. We evaluated the impact of these strategies on segmentation performance through comparative experiments, summarized in Table 3. The combined use of both strategies led to a 3.5% improvement in the average Dice score (84.3% vs. 80.8%). Importantly, these strategies improved performance without introducing additional computational complexity. We calculated a p-value for this metric, which was less than 0.05, supporting this improvement.

We systematically tested voxel count (VC) thresholds ranging from 100 to 1000 voxels to optimize the Dice score across tumor subregions, focusing on improving segmentation quality. As shown in Figure 6, a 500-voxel threshold achieved the best balance between false positives and true positives. At lower thresholds (e.g.,<500 voxels), over-segmentation occurred, leading to excessive false positives, particularly in the ET region, where small noise regions were incorrectly classified as tumor. Conversely, higher thresholds (>500 voxels) risked under-segmentation, excluding small but clinically significant tumor regions, reducing sensitivity and potentially missing subtle pathological features. The 500-voxel threshold effectively mitigated these issues, ensuring more robust and accurate segmentation across all tumor subregions.

To further justify this choice, we conducted a sensitivity analysis to evaluate the impact of different VC thresholds on segmentation performance. The results, summarized in Table 4, indicate that a threshold of 500 voxels consistently yielded the highest average Dice score while maintaining a favorable balance between precision and sensitivity across the four datasets. For the BraTS2019 dataset, the average Dice score reaches a maximum of 0.834 at a 500-voxel threshold, compared to 0.816 and 0.811 at thresholds of 100 and

TABLE 2 Ablation study of the CTPC architecture on the BraTS2020 dataset, testing the impact of different components. .

	Dice				Sen				Haus95			
	ET	WT	TC	Mean	ET	WT	TC	Mean	ET	Mean	TC	Mean
Pool	0.760	0.889	0.850	0.833	0.782	0.912	0.804	0.833	28.8	8.89	9.13	15.6
Scov	0.761	0.892	0.856	0.836	0.781	0.907	0.808	0.832	27.3	6.56	9.21	14.4
Sconv+PE	0.768	0.897	0.865	0.843	0.785	0.911	0.812	0.836	27.1	5.92	9.94	14.3

Red denotes the best results, and blue means the second best.

TABLE 3 Effect of post-processing on segmentation performance on the BraTS2020 dataset, evaluating strategies such as no post-processing (None), only TTA, and a combination of TTA and VC.

	Dice				Sen				Haus95			
	ET	WT	TC	Mean	ET	WT	TC	Mean	ET	Mean	TC	Mean
None	0.694	0.880	0.850	0.808	0.695	0.905	0.806	0.802	45.8	9.22	11.9	22.3
TTA	0.728	0.897	0.865	0.830	0.726	0.911	0.812	0.816	39.3	5.92	9.94	18.4
TTA+VC	0.768	0.897	0.865	0.843	0.785	0.911	0.812	0.836	27.1	5.92	9.94	14.3

Red denotes the best results, and blue means the second best.

1000, respectively. Sensitivity and precision also achieve their highest values of 0.838 and 0.842 at the 500-voxel threshold. In the BraTS2020 dataset, the average Dice score reaches a maximum of 0.843 when the threshold is 500, with sensitivity and precision also reaching their maximum values of 0.836 and 0.855, respectively. For the BraTS2021 dataset, when the threshold is 500, all evaluation metrics show excellent performance, with the average Dice score at 0.865, sensitivity at 0.855, and precision at 0.846. For the ZZH dataset, although all metrics are relatively low across all thresholds, at the 500-voxel threshold, the Dice coefficient, sensitivity, and precision are 0.616, 0.639, and 0.653, respectively, showing a relative advantage compared to the performance under other thresholds. Overall, setting the threshold at 500 for volume constraints generally yields better segmentation results.

5.1.5 Loss function

Brain tumor segmentation faces significant category imbalances, both between tumor and non-tumor tissue and among different tumor subregions. To address this, we assigned class weights based on category frequencies. Incorporating class weights into the GDL function increased the average Dice score by 5.2% (84.3% vs. 79.1%), as shown in Table 5, demonstrating its effectiveness in handling class imbalance.

Glioma MRI datasets inherently exhibit imbalances among tumor regions, with the TC and ET being significantly smaller compared to the WT. To address this imbalance and emphasize clinically critical regions, we assigned higher weights to the ET and TC during training. Specifically, the model was configured with weights of 0.1 for the background, 1 for the WT, and 2 for the ET and TC, as shown in Table 6. This weighting strategy improved the Dice scores for the smaller regions by encouraging the model to prioritize them over the disproportionately large background and WT regions. We observed that increasing the weights for the ET and TC significantly enhanced their segmentation accuracy, ensuring better representation of these clinically significant areas. Simultaneously, reducing the background weight to 0.1 prevented the model from overfitting to irrelevant regions, which often dominate the data due to their larger size. Conversely, assigning higher weights to the background degraded the segmentation

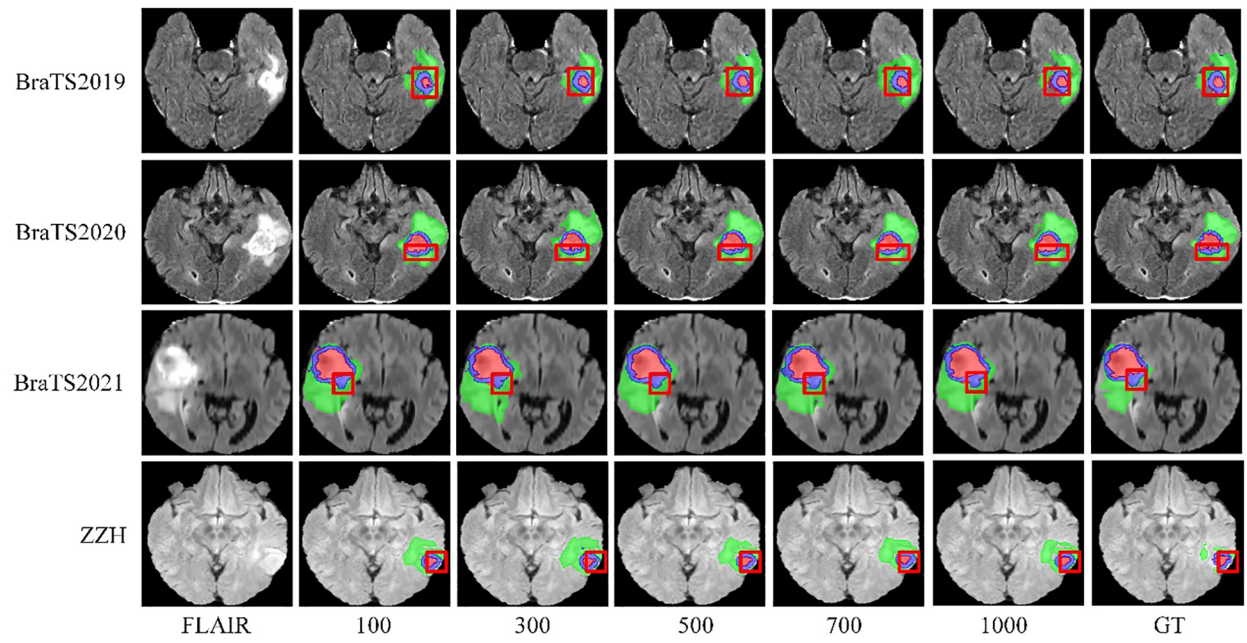


FIGURE 6
Visualization of results from different threshold selections in the VC post-processing technique on both the BraTS and ZZH datasets.

TABLE 4 Comparison of segmentation performance with different thresholds across all the datasets.

Datasets	Threshold	Dice	Sensitivity	Precision
BraTS2019	100	0.816	0.833	0.835
	300	0.831	0.837	0.838
	500	0.834	0.838	0.842
	700	0.827	0.821	0.841
	1000	0.811	0.829	0.828
BraTS2020	100	0.821	0.825	0.828
	300	0.833	0.842	0.834
	500	0.843	0.836	0.855
	700	0.834	0.811	0.863
	1000	0.815	0.823	0.831
BraTS2021	100	0.827	0.824	0.822
	300	0.83	0.827	0.824
	500	0.865	0.855	0.846
	700	0.835	0.833	0.831
	1000	0.833	0.83	0.828
ZZH	100	0.574	0.603	0.621
	300	0.581	0.605	0.623
	500	0.616	0.639	0.653
	700	0.592	0.614	0.628
	1000	0.584	0.611	0.637

performance on the smaller regions, as the model became biased toward identifying the dominant background area. The selected weight configuration effectively strikes a balance by focusing on critical tumor subregions while minimizing distractions from the background, resulting in segmentation that is both accurate and clinically relevant.

5.1.6 CNN branch and transformer branch

The CTPC module consists of two primary components: the CNN and Transformer branches. To better understand their contributions, we conducted ablation studies, with results summarized in Table 7. Removing the CNN branches caused a significant drop in segmentation performance, highlighting their critical role in the CTPC framework.

In contrast, introducing the Transformer branch significantly improved performance at a relatively low parameter cost. This demonstrates the Transformer's high efficiency and underscores its strength within the model.

5.1.7 Different optimization strategy

We conducted experiments to evaluate the impact of different optimization strategies on model performance. Specifically, we tested SGD, Adam, and Adagrad on the BraTS2020 dataset to assess their influence on segmentation metrics. As summarized in Table 8, the SGD optimizer struggled with the ET metric (0.751), indicating difficulty in segmenting complex structures, though it performed slightly better on WT and TC metrics, averaging 0.811. In contrast, the Adam optimizer delivered the best overall performance, excelling in WT (0.897) and TC (0.865), demonstrating its ability to handle intricate segmentation tasks. Adagrad's results were intermediate, performing well in WT (0.875) and TC (0.838) but falling short of Adam. These results highlight the need for effective optimizers like Adam to complement robust model architectures.

5.1.8 Cross-dataset model testing

Table 9 summarizes our evaluation of the model's generalization, trained on BraTS2021 and tested on BraTS2019, BraTS2020, and ZZH datasets. On BraTS2019, the model achieved Dice scores of 0.645 (ET), 0.775 (WT), and 0.735 (TC), averaging 0.718. Performance improved slightly on BraTS2020, with scores of 0.651 (ET), 0.782 (WT), and 0.732 (TC), averaging 0.721, indicating good adaptation to consistent imaging protocols.

In contrast, testing on the more heterogeneous ZZH clinical dataset resulted in lower Dice scores: 0.411 (ET), 0.705 (WT), and 0.431 (TC), averaging 0.515. This performance drop highlights the challenges of domain shifts and non-standardized imaging. These findings show the model's robustness on standardized datasets but underline the need for domain adaptation to handle clinical variability.

5.2 Results

The comparison of the qualitative results for the BraTS2019, BraTS2020, BraTS2021, and ZZH datasets is displayed in Table 10. The segmentation outcomes for the three subregions on the BraTS datasets are similar, with WT achieving the highest accuracy and exhibiting fewer outliers. However, the annotation quality of the ZZH dataset for two subregions, ET and TC, could be improved.

TABLE 5 Comparison of segmentation performance with different loss functions on the BraTS2020 dataset.

	Dice				Sen				Haus95			
	ET	WT	TC	Mean	ET	WT	TC	Mean	ET	Mean	TC	Mean
DL	0.753	0.877	0.745	0.791	0.753	0.891	0.722	0.789	35.9	14.7	16.5	22.4
GDL	0.760	0.886	0.852	0.832	0.775	0.905	0.788	0.823	30.1	7.92	15.4	17.8
Multi-GDL	0.768	0.897	0.865	0.843	0.785	0.911	0.812	0.836	27.1	5.92	9.94	14.3

Red denotes the best results, and blue means the second best.

TABLE 6 Comparison of segmentation performance with different weights using the GDL on all the datasets used.

Datasets	Weights	Dice	Sensitivity	Precision
BraTS2019	0.1, 1, 1, 1	0.802	0.812	0.817
	0.1, 1, 2, 2	0.834	0.838	0.842
	1, 1, 1, 1	0.795	0.801	0.812
	1, 1, 2, 2	0.799	0.806	0.814
BraTS2020	0.1, 1, 1, 1	0.831	0.822	0.838
	0.1, 1, 2, 2	0.843	0.836	0.855
	1, 1, 1, 1	0.805	0.793	0.812
	1, 1, 2, 2	0.812	0.806	0.825
BraTS2021	0.1, 1, 1, 1	0.844	0.835	0.833
	0.1, 1, 2, 2	0.865	0.855	0.846
	1, 1, 1, 1	0.833	0.827	0.822
	1, 1, 2, 2	0.841	0.835	0.828
ZZH	0.1, 1, 1, 1	0.596	0.616	0.635
	0.1, 1, 2, 2	0.616	0.639	0.653
	1, 1, 1, 1	0.574	0.583	0.591
	1, 1, 2, 2	0.586	0.615	0.633

To further demonstrate the efficacy of the proposed DeepGlioSeg, nine advanced image segmentation algorithms in the field of medical image segmentation were reproduced, including 3D U-Net (38), 3D V-Net (39), Attention U-Net (40), nnU-Net (41), nnFormer (42), Segtran (43), SwinUNETR (44), TransBTS (45), and UNETR (23).

For the BraTS2019 dataset, the mean Dice scores for each method across the ET, WT, and TC regions, as well as the overall mean Dice score, are presented in Table 11. Notably, our proposed method demonstrates superior performance, achieving the highest Dice scores across all regions and the overall mean Dice score. Specifically, it achieves Dice scores of 0.761 for ET, 0.887 for WT, 0.854 for TC, and an impressive overall mean Dice score of 0.834. Conversely, other contemporary approaches exhibit varying levels of segmentation accuracy, with Dice scores ranging from 0.656 to 0.750 for ET, 0.831 to 0.879 for WT, 0.781 to 0.835 for TC, and 0.756 to 0.821 for the overall mean Dice score. These results underscore the significant improvement achieved by our proposed method over existing approaches, emphasizing its potential to advance the field of brain tumor segmentation.

TABLE 7 Comparison of segmentation performance of the CNN branch and Transformer branch on the BraTS2020 dataset.

CNN branch	Trans branch	Mean Dice	Params
✗	✓	0.654	5.09M
✓	✗	0.798	5.47M
✓	✓	0.843	6.62M

Red denotes the best results, and blue means the second best.

For the BraTS2020 dataset, the results are summarized in Table 12. Our proposed method stands out as the best-performing approach, achieving the highest mean Dice scores: 0.897 for WT, 0.865 for TC, 0.768 for ET and 0.843 for the overall mean Dice score. It is important to highlight that SwinUNETR is the closest competitor to our proposed method, achieving remarkable mean Dice scores of 0.754 for ET, 0.883 for WT, 0.837 for TC, and 0.823 for the mean Dice score. Comparatively, while several other methods also show competitive performance, our proposed method consistently outperforms them across all regions, demonstrating its effectiveness in accurately segmenting brain tumors on the BraTS2020 dataset.

For the BraTS2021 dataset, Table 13 provides a comprehensive overview of the segmentation performance of several methods. Our proposed method demonstrates exceptional segmentation accuracy, achieving the highest Dice scores across all regions. It stands out with Dice scores of 0.808 for ET, 0.910 for WT, 0.878 for TC, and an impressive mean Dice score of 0.865. This performance underscores the ability of our proposed method to accurately delineate brain tumor regions, indicating potential clinical relevance. SwinUNETR and Segtran also exhibit strong segmentation performance, achieving mean Dice scores of 0.854 and 0.845, respectively. Their robust performance reflects the effectiveness of their segmentation strategies and highlights their potential clinical utility. Our comprehensive evaluation on the BraTS2021 validation dataset confirms the exceptional performance of our proposed approach, outperforming all other methods across different segmentation regions.

To demonstrate the excellence of the proposed DeepGlioSeg in segmenting clinical datasets of suboptimal quality, we present the results of our experiments on the ZZH dataset, comparing the segmentation performance of several advanced technologies for brain tumor segmentation. Table 5 provides a comprehensive evaluation of the segmentation effectiveness of nine advanced image segmentation technologies. From Table 14, we summarize the following key points: (1) The ZZH dataset presents unique challenges for brain tumor segmentation. The Dice scores of all methods are significantly lower compared to previous datasets, indicating the presence of complex tumor phenotypes and irregular shapes in this dataset. (2) Among the evaluated methods, our proposed method consistently achieves the highest Dice scores, demonstrating its effectiveness in addressing the challenges posed by the ZZH dataset. Specifically, it achieves an average Dice score of 0.616, indicating relatively strong segmentation performance even in this challenging context. (3) While our proposed method stands out, there is variability in the performance of other methods. SwinUNETR also shows competitive effectiveness, with an average Dice score of 0.613.

The lower segmentation performance on the ZZH dataset compared to the BraTS datasets arises from differences in data quality, diversity, imaging protocols, and real-world complexities. First, BraTS benefits from high-quality, standardized annotations by multiple radiologists, ensuring consistent labels. In contrast, ZZH annotations reflect varying expertise and subjective judgments, introducing inconsistencies, particularly for smaller regions like ET. Second, the BraTS datasets are diverse, encompassing varied patient demographics, tumor grades, and imaging conditions, enabling better generalization. By

TABLE 8 Quantitative results comparing performance across different optimization strategy.

Strategy	Dice			
	ET	WT	TC	Mean
SGD	0.751	0.859	0.821	0.811
Adam	0.768	0.897	0.865	0.843
Adagrad	0.758	0.875	0.838	0.823

TABLE 9 Performance of cross-dataset model testing.

Cross Dataset	Dice			
	ET	WT	TC	Mean
BraTS2021->BraTS2019	0.645	0.775	0.735	0.718
BraTS2021->BraTS2020	0.651	0.782	0.732	0.721
BraTS2021->ZZH	0.411	0.705	0.431	0.515

comparison, ZZH's limited and less diverse sample restricts feature learning and introduces potential demographic biases. Third, standardized imaging protocols in BraTS ensure consistent data characteristics, whereas ZZH exhibits variability in scanner models, field strengths, and acquisition parameters, affecting tumor visibility and segmentation accuracy. Lastly, ZZH reflects real clinical challenges, such as artifacts, motion blur, and non-standardized conditions, which are less prevalent in BraTS.

Figures 7 and 8 provide visual comparisons of segmentation results from various methods on the BraTS2019, BraTS2020, and BraTS2021 datasets, supported by quantitative metrics such as HD95 and boundary overlap to objectively assess boundary quality. The HD95 metric, which evaluates worst-case boundary deviations, highlights the precision of DeepGlioSeg compared to state-of-the-art models such as SwinUNETR and Segtran. For instance, DeepGlioSeg achieves an HD95 of 5.09 for the ET on BraTS2021, outperforming SwinUNETR (6.45) and Segtran (6.78), indicating superior boundary alignment. Additionally, boundary overlap metrics such as Dice scores reinforce these findings. DeepGlioSeg achieves a Dice score of 0.92 for the WT, outperforming models that struggle with under-segmentation or over-segmentation in intricate regions. This performance is attributed to DeepGlioSeg's hybrid CNN-Transformer architecture, which effectively integrates local detail extraction and global context modeling, enabling precise tumor boundary delineation even in challenging cases. These quantitative results align with the visual comparisons, demonstrating DeepGlioSeg's capability to handle complex boundary variations.

However, challenges persist in tumors with high heterogeneity, where significant variability in appearance affects segmentation consistency. DeepGlioSeg demonstrates superior performance in handling such heterogeneity, achieving consistent results across various datasets, such as ZZH, BraTS2019, BraTS2020, and BraTS2021. Its success lies in the flexible CTPC architecture, which integrates CNNs for capturing localized details and Transformers for modeling global context. This dual-pathway approach enables the seamless fusion of local and global features, ensuring accurate delineation of complex tumor boundaries. As highlighted in Figure 9, DeepGlioSeg produces sharper and more precise tumor segmentations than nnU-Net and SwinUNETR, particularly in real-world datasets (e.g., ZZH), which exhibit greater variability than more standardized datasets such as BraTS.

DeepGlioSeg's design directly overcomes the limitations of existing models. nnU-Net, while efficient in capturing local features, struggles to generalize across datasets due to its lack of global context modeling. SwinUNETR incorporates Transformers for global feature representation but lacks the balanced local-global integration of DeepGlioSeg, limiting its ability to segment tumors with complex boundaries in heterogeneous datasets. In contrast, the CTPC module's efficient fusion of local and global features allows DeepGlioSeg to excel in identifying subtle tumor variations. This capability is crucial for accurate segmentation in real-world clinical settings.

6 Discussion

Gliomas present significant challenges for segmentation due to their complex heterogeneity, including variability in shape, structure, and location. Accurate tumor boundary delineation is essential, requiring models capable of capturing both local and global features. However, many existing methods struggle to achieve this integration effectively. To address this challenge, we propose the DeepGlioSeg framework, which integrates a CTPC module with parallel CNN and Transformer branches. The CNN branch captures fine-grained local details, while the Transformer branch models long-range dependencies. The combination of these two pathways ensures a robust fusion of local and global features, which enhances the model's ability to represent complex tumor

TABLE 10 Qualitative comparison of results on the BraTS2019, BraTS2020, BraTS2021, and ZZH datasets.

Dataset	Dice				Sen				Haus95			
	ET	WT	TC	Mean	ET	WT	TC	Mean	ET	WT	TC	Mean
BraTS2019	0.761	0.887	0.854	0.834	0.785	0.905	0.809	0.836	33.2	7.03	7.09	15.8
BraTS2020	0.768	0.897	0.865	0.843	0.758	0.911	0.812	0.836	27.1	5.92	9.94	14.3
BraTS2021	0.808	0.91	0.878	0.865	0.836	0.925	0.843	0.868	22.4	5.06	10.7	12.7
ZZH	0.491	0.811	0.546	0.616	0.621	0.819	0.65	0.697	41.3	7.74	7.63	18.9

10.3389/fonc.2025.1449911

TABLE 11 Quantitative results comparing the performance of our method with other state-of-the-art segmentation methods on the BraTS2019 dataset.

Method	Dice			
	ET	WT	TC	Mean
3DUnet	0.721	0.864	0.832	0.805
3DVnet	0.712	0.861	0.824	0.799
Atten-Unet	0.738	0.848	0.800	0.795
nnU-Net	0.741	0.868	0.834	0.814
nnformer	0.656	0.831	0.781	0.756
Segtran	0.725	0.858	0.831	0.804
SwinUNETR	0.750	0.879	0.835	0.821
TransBTS	0.741	0.862	0.834	0.812
UNETR	0.736	0.859	0.812	0.802
Proposed	0.761	0.887	0.854	0.834

Red denotes the best results, and blue means the second best.

TABLE 13 Quantitative results comparing the performance of our method with other state-of-the-art segmentation methods on the BraTS2021 dataset.

Method	Dice			
	ET	WT	TC	Mean
3DUnet	0.735	0.874	0.843	0.817
3DVnet	0.728	0.870	0.836	0.811
Atten-Unet	0.716	0.865	0.821	0.800
nnU-Net	0.764	0.887	0.852	0.834
nnformer	0.711	0.861	0.812	0.794
Segtran	0.772	0.896	0.869	0.845
SwinUNETR	0.782	0.905	0.875	0.854
TransBTS	0.761	0.885	0.850	0.832
UNETR	0.756	0.883	0.842	0.827
Proposed	0.808	0.910	0.878	0.865

Red denotes the best results, and blue means the second best.

characteristics—such as varying shapes and structural heterogeneity—resulting in improved segmentation accuracy.

In addition to effective feature fusion, DeepGlioSeg tackles the issue of class imbalance—a common problem in glioma segmentation, where certain tumor regions, such as the ET, are underrepresented. To address this, we employ a weighted loss function that extends the generalized Dice loss. By assigning higher weights to clinically significant but underrepresented regions (like ET), the model can prioritize these areas during training, ensuring accurate segmentation of both larger regions (e.g., the WT) and smaller, more challenging regions crucial for treatment planning. This approach helps mitigate the bias toward larger regions, which is often seen in conventional models.

Table 15 compares the computational complexity of the proposed method with state-of-the-art models across four datasets: BraTS2019, BraTS2020, BraTS2021, and ZZH. Key metrics include the number of parameters (Params in M), computational cost (FLOPs in G), and statistical significance (p-value). Transformer-based models, such as TransBTS and SwinUNETR, exhibit higher parameter counts (15.56M–30.62M) and computational costs (163.73G–254.45G), while lighter models like 3DUNet and Atten-Unet maintain smaller parameter sizes but incur higher FLOPs. The proposed method strikes an optimal balance with 6.92M parameters and 156.79G FLOPs, significantly reducing computational demands while maintaining competitive performance. Lower p-values across datasets confirm the statistical

TABLE 12 Quantitative results comparing the performance of our method with other state-of-the-art segmentation methods on the BraTS2020 dataset.

Method	Dice			
	ET	WT	TC	Mean
3DUnet	0.746	0.870	0.848	0.821
3DVnet	0.734	0.865	0.844	0.814
Atten-Unet	0.742	0.852	0.802	0.818
nnU-Net	0.742	0.871	0.842	0.818
nnformer	0.659	0.833	0.785	0.759
Segtran	0.728	0.875	0.849	0.817
SwinUNETR	0.754	0.883	0.837	0.823
TransBTS	0.744	0.867	0.839	0.816
UNETR	0.734	0.864	0.817	0.813
Proposed	0.768	0.897	0.865	0.843

Red denotes the best results, and blue means the second best.

TABLE 14 Quantitative results comparing the performance of our method with other state-of-the-art segmentation methods on the ZZH dataset.

Method	Dice			
	ET	WT	TC	Mean
3DUnet	0.473	0.745	0.495	0.571
3DVnet	0.446	0.718	0.506	0.556
Atten-Unet	0.430	0.731	0.461	0.540
nnU-Net	0.486	0.808	0.531	0.608
nnformer	0.419	0.701	0.488	0.536
Segtran	0.452	0.804	0.573	0.609
SwinUNETR	0.506	0.803	0.531	0.613
TransBTS	0.434	0.768	0.491	0.564
UNETR	0.441	0.735	0.482	0.552
Proposed	0.491	0.811	0.546	0.616

Red denotes the best results, and blue means the second best.

10.3389/fonc.2025.1449911

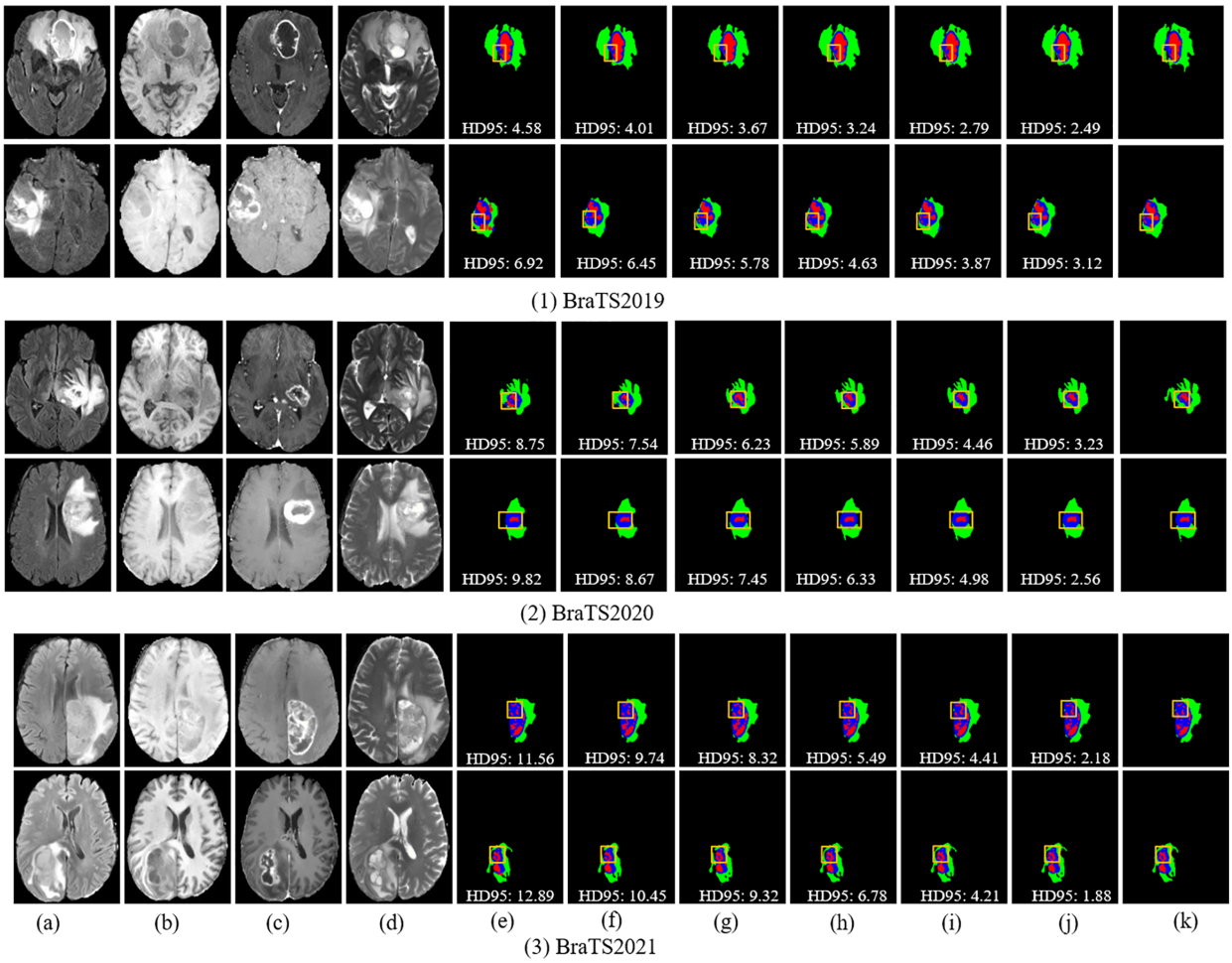


FIGURE 7
Visual comparison of segmentation results with different models on the BraTS2019, BraTS2020, and BraTS2021 datasets. **(A)** FLAIR. **(B)** T1. **(C)** T1-CE. **(D)** T2. **(E)** Attention-Unet. **(F)** nnU-Net. **(G)** Segtran. **(H)** SwinUNETR. **(I)** TransBTS. **(J)** Ours. **(K)** GT.

significance of the proposed method's improvements, emphasizing its efficiency and robustness compared to other models.

Beyond glioma segmentation, the CTPC module is highly adaptable and could be applied to other medical imaging tasks. Its ability to integrate local and global features makes it well-suited for segmenting tumors with irregular boundaries, such as lung or liver tumors. Additionally, the framework supports multi-modal imaging (e.g., PET-CT, MRI-CT fusion), allowing the model to combine complementary information for more accurate segmentation. In the future, we plan to enhance the model further by incorporating dynamic attention-based feature selection and task-specific fusion strategies, broadening its clinical applicability.

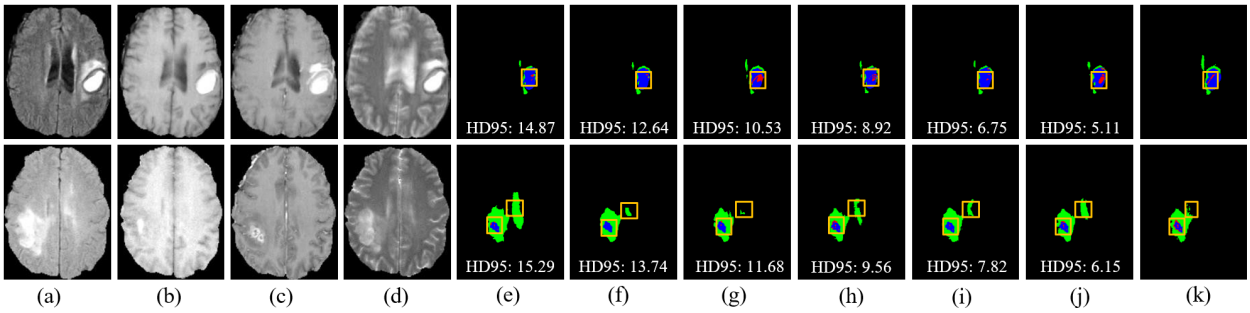


FIGURE 8
Visual comparison of segmentation results with different models on the ZZH dataset. **(A)** FLAIR. **(B)** T1. **(C)** T1-CE. **(D)** T2. **(E)** Attention-Unet. **(F)** nnU-Net. **(G)** Segtran. **(H)** SwinUNETR. **(I)** TransBTS. **(J)** Ours. **(K)** GT.

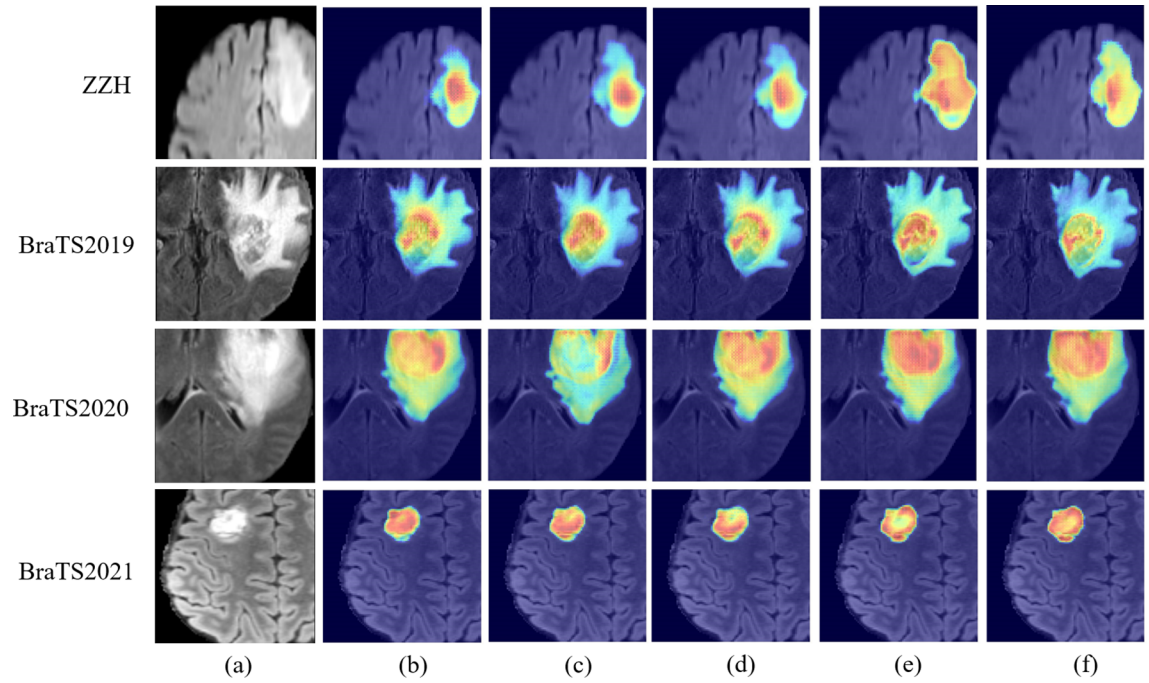


FIGURE 9
Visual comparison of feature maps with different models on the ZZH, BraTS2019, BraTS2020, and BraTS2021 datasets. **(A)** Raw image. **(B)** nnU-Net. **(C)** Segtran. **(D)** SwinUNETR. **(E)** Ours. **(F)** TransBTS.

Despite these strengths, the model is not without limitations. The complexity and large parameter count increase the risk of overfitting, especially when trained on smaller datasets. To mitigate this, regularization techniques such as dropout, data augmentation, and early stopping could improve the model's robustness. Another challenge is the variability in MRI data from different scanners or imaging protocols, which can hinder generalization. Future work will focus on data harmonization methods, such as domain adaptation and intensity normalization, to reduce these challenges and improve generalization across diverse datasets.

Looking ahead, there are several opportunities for further refinement of DeepGlioSeg. Incorporating multi-scale feature extraction, attention mechanisms, and multi-task learning can enhance its ability to handle a broader range of clinical tasks. Additionally, transfer learning from pre-trained models and the inclusion of contextual priors could reduce dependency on large labeled datasets, improving the model's adaptability to various imaging modalities and expanding its clinical utility.

7 Conclusion

In this study, we present DeepGlioSeg, a novel framework developed to address the challenging task of automating brain tumor segmentation. Our proposed method incorporates the CTPC module into an encoder-decoder network architecture, enabling it to capture critical local features of gliomas, such as texture and edges. To tackle the challenge of category imbalance, we introduce the Multi-GDL loss function, which adjusts category weights to rebalance loss contributions, resulting in more accurate identification of tumor

TABLE 15 Comparison of computational complexity between our method and other state-of-the-art methods on all the datasets used.

Datasets	Method	Params (M)	FLOPS (G)	*p*-value
BraTS2019	3DUnet	5.42	275.53	2e-6
	3DVnet	4.76	157.81	4e-5
	Atten-Unet	2.47	164.06	2e-4
	nnU-Net	9.4	84.03	0.0012
	nnformer	14.91	172.04	0.0048
	Segtran	29.19	254.45	0.0035
	SwinUNETR	15.56	206.76	0.0072
	TransBTS	30.62	163.73	0.0223
	UNETR	15.56	206.76	0.0185
	Proposed	6.92	156.79	*
BraTS2020	3DUnet	5.42	275.53	1e-7
	3DVnet	4.76	157.81	3e-4
	Atten-Unet	2.47	164.06	1e-4
	nnU-Net	9.4	84.03	0.0015
	nnformer	14.91	172.04	0.0053
	Segtran	29.19	254.45	0.0041
	SwinUNETR	15.56	206.76	0.0073
	TransBTS	30.62	163.73	0.0244

(Continued)

TABLE 15 Continued

Datasets	Method	Params (M)	FLOPS (G)	*p*-value
	UNETR	15.56	206.76	0.0157
	Proposed	6.92	156.79	*
BraTS2021	3DUnet	5.42	275.53	1e-7
	3DVnet	4.76	157.81	3e-6
	Atten-Unet	2.47	164.06	1e-5
	nnU-Net	9.4	84.03	0.0009
	nnformer	14.91	172.04	0.0052
	Segtran	29.19	254.45	0.0032
	SwinUNETR	15.56	206.76	0.0068
	TransBTS	30.62	163.73	0.0277
	UNETR	15.56	206.76	0.0136
	Proposed	6.92	156.79	*
ZZH	3DUnet	5.42	275.53	1e-8
	3DVnet	4.76	157.81	5e-4
	Atten-Unet	2.47	164.06	6e-4
	nnU-Net	9.4	84.03	0.0018
	nnformer	14.91	172.04	0.0046
	Segtran	29.19	254.45	0.0037
	SwinUNETR	15.56	206.76	0.0076
	TransBTS	30.62	163.73	0.0218
	UNETR	15.56	206.76	0.0165
	Proposed	6.92	156.79	*

The p-value is computed for paired t-tests between our method and other methods. A p-value less than 0.05 indicates the statistical significance of the paired t-tests.
The symbol * indicates that the p-values for the other methods were calculated using a paired samples t-test, with our method serving as the benchmark reference.

structures. To further enhance glioma segmentation during inference, we employ a combination of TTA and VC as post-processing strategies. These improvements highlight the effectiveness of the CTPC module, Multi-GDL loss function, and post-processing strategies. Future enhancements for segmenting complex regions like ET and TC may involve refining the CTPC module with dynamic attention-based feature fusion to adaptively focus on intricate boundaries. Additionally, multi-scale learning and adaptive weighted loss functions could further improve segmentation by capturing multi-resolution features and prioritizing critical regions.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The datasets utilized in this study, including BraTS2019, BraTS2020, and BraTS2021, are publicly accessible at the following URLs: BraTS2019: https://www.med.upenn.edu/cbica/brats-2019/ BraTS2020: https://www.med.upenn.edu/cbica/brats2020 BraTS2021: https://www.med.upenn.edu/cbica/brats2021 However, the private dataset referred to as ZZH, which is also discussed in this paper, is currently not publicly available. The raw and processed data required to replicate these study results are part of an ongoing research project and cannot be shared at this time. For inquiries regarding access to the ZZH dataset, please contact RL at drlrp2022@163.com.

Ethics statement

The studies involving humans were approved by the Institutional Review Board (IRB) of both Hangzhou Third People's Hospital and the First Affiliated Hospital of Zhengzhou University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

RL: Conceptualization, Formal analysis, Investigation, Resources, Validation, Writing – original draft, Writing – review & editing. YL: Formal analysis, Methodology, Software, Visualization, Writing – review & editing. YH: Conceptualization, Investigation, Resources, Writing – original draft. XM: Formal analysis, Methodology, Software, Writing – original draft. GZ: Data curation, Resources, Validation, Writing – original draft. YW: Conceptualization, Investigation, Resources, Writing – original draft. CS: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is funded in part by the Hangzhou Agriculture and Social Development Public Welfare Scientific Research Guidance Projects (No. 20241029Y036), and the Zhejiang Basic Public Welfare Science and Technology Fund Project (No. LGF19H050001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Logeswari T, Karnan M. An improved implementation of brain tumor detection using segmentation based on soft computing. *J Cancer Res Exp Oncol*. (2010) 2:006–14. doi: 10.5897/jcreo2010.0185

2. Schneider T, Mawrin C, Scherlach C, Skalej M, Firsching R. Gliomas in adults. *Deutsches Ärzteblatt Int*. (2010) 107:799. doi: 10.3238/arztebl.2010.0799

3. Akram M, Usman A. (2011). Computer aided system for brain tumor detection and segmentation, in: *International Conference on Computer Networks and Information Technology*. (Abbottabad, Pakistan: IEEE), pp. 299–302.

4. Işın A, Direkoğlu C, Şah M. Review of MRI-based brain tumor image segmentation using deep learning methods. *Proc Comput Sci*. (2016) 102:317–24. doi: 10.1016/j.procs.2016.09.407

5. Bauer S, Wiest R, Nolte LP, Reyes M. A survey of MRI-based medical image analysis for brain tumor studies. *Phys Med Biol*. (2013) 58:R97. doi: 10.1088/0031-9155/58/13/R97

6. Menze B, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging*. (2014) 34:1993–2024. doi: 10.1109/TMI.2014.2377694

7. Kaus M, Warfield S, Nabavi A, Black PM, Jolesz FA, Kikinis R. Automated segmentation of MR images of brain tumors. *Radiology*. (2001) 218:586–91. doi: 10.1148/radiology.218.2.r01fe44586

8. Gordillo N, Montseny E, Sobrevilla P. State of the art survey on MRI brain tumor segmentation. *Magnetic Resonance Imaging*. (2013) 31:1426–38. doi: 10.1016/j.mri.2013.05.002

9. Chen X, Wang H, Ni B. X-volution: On the unification of convolution and self-attention. *arXiv preprint arXiv:2106.02253*. (2021) 6:1–13. doi: 10.48550/arXiv.2106.02253

10. Vaswani A. Attention is all you need. *Adv Neural Inf Process Systems*. (2017) 30:1–11. doi: 10.48550/arXiv.1706.03762

11. Zhang Y, Liu H, Hu Q. Transfuse: Fusing transformers and cnns for medical image segmentation. *Med image computing Comput assisted intervention–MICCAI 2021*. (2021) 12901:14–24. doi: 10.48550/arXiv.2102.08005

12. Ruan S, Lebonvallet S, Merabet A, Constans JM. (2007). Tumor segmentation from a multispectral MRI images by using support vector machine classification, in: *2007 4th IEEE International Symposium on Biomedical Imaging*. (Arlington, VA, USA: IEEE), pp. 1236–9.

13. Ma C, Luo G, Wang K. Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images. *IEEE Trans Med Imaging*. (2018) 37:1943–54. doi: 10.1109/TMI.2018.2805821

14. Bauer S, Nolte LP, Reyes M. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. *Med Image Computing Computer-Assisted Intervention–MICCAI 2011*. (2011) 6893:354–61. doi: 10.1007/978-3-642-23626-6_44

15. Xu X, Zhao W, Zhao J. Brain tumor segmentation using attention-based network in 3D MRI images. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (2020). (Shenzhen, China: Springer), p. 3–13.

16. Shen Y, Gao M. Brain tumor segmentation on MRI with missing modalities. *Inf Process Med Imaging*. (2019) 11492:417–28. doi: 10.1007/978-3-030-20351-1_32

17. Le Folgoc L, Nori AV, Ancha S, Criminisi A. Lifted auto-context forests for brain tumor segmentation. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injurie* (2016). (Athens, Greece: Springer), p. 171–83.

18. Wang G, Li W, Ourselin S, Vercauteren T. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injurie* (2018). (Granada, Spain: Springer), p. 178–90.

19. Lachinov D, Vasiliev E, Turlapov V. (2018). Glioma segmentation with cascaded UNet, in: *International MICCAI Brainlesion Workshop*. (Granada, Spain: Springer), pp. 189–98.

20. Weninger L, Rippel O, Koppers S, Merhof D. Segmentation of brain tumors and patient survival prediction: Methods for the brats 2018 challenge. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. (Shenzhen, China: Springer). (2019) 11384, 3–12. doi: 10.1007/978-3-030-11726-9_1

21. Ghosal P, Reddy S, Sai C, Pandey V, Chakraborty J, Nandi D. (2019). A deep adaptive convolutional network for brain tumor segmentation from multimodal MR images, in: *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. (Kochi, India: IEEE), pp. 1065–70.

22. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. (2022). Swin-unet: Unet-like pure transformer for medical image segmentation, in: *European Conference on Computer Vision*. (Tel Aviv, Israel: Springer), pp. 205–18.

23. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, et al. (2022). Unetr: Transformers for 3d medical image segmentation, in: *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. (Waikoloa, HI, USA: IEEE), pp. 574–84.

24. Petit O, Thome N, Rambour C, Themyr L, Collins T, Soler L. U-net transformer: Self and cross attention for medical image segmentation. *Mach Learn Med Imaging*. (2021) 12966:267–76. doi: 10.48550/arXiv.2103.06104

25. Gai D, Zhang JQ, Xiao YS, Min WD, Zhong YF, Zhong YL, et al. RMTF-Net: Residual mix transformer fusion net for 2D brain tumor segmentation. *Brain Sci*. (2022) 12:1145. doi: 10.3390/brainsci12091145

26. Zhu ZQ, He XY, Qi GQ, Li YY, Cong BS, Liu Y. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI. *Inf Fusion*. (2023) 91:376–87. doi: 10.1016/j.inffus.2022.10.022

27. Hu ZY, Li L, Sui A, Wu GQ, Wang YY, Yu JH. An efficient R-transformer network with dual encoders for brain glioma segmentation in MR images. *Biomed Signal Process Control*. (2023) 79:104034. doi: 10.1016/j.bspc.2022.104034

28. Li Z, Kamnitsas K, Glocker B. Analyzing overfitting under class imbalance in neural networks for image segmentation. *IEEE Trans Med Imaging*. (2020) 40:1065–77. doi: 10.1109/TMI.42

29. Zheng H, Qin Y, Gu Y, Xie FF, Yang J, Sun JY, et al. Alleviating class-wise gradient imbalance for pulmonary airway segmentation. *IEEE Trans Med Imaging*. (2021) 40:2452–62. doi: 10.1109/TMI.2021.3078828

30. Hossain M, Betts J, Paplinski A. Dual focal loss to address class imbalance in semantic segmentation. *Neurocomputing*. (2021) 462:69–87. doi: 10.1016/j.neucom.2021.07.055

31. Bressan PO, Junior J, Martins JAC, Gonçalves DN, Freitas DM, Osco LP, et al. Semantic segmentation with labeling uncertainty and class imbalance. *arXiv preprint arXiv:2102.04566*. (2021) 2:1–15. doi: 10.21203/rs.3.rs-409625/v1

32. Pan LS, Li CW, Su SF, Tay SY, Tran QV, Chan WP. Coronary artery segmentation under class imbalance using a U-Net based architecture on computed tomography angiography images. *Sci Rep*. (2021) 11:14493. doi: 10.1038/s41598-021-93889-z

33. Sudre C, Li W, Vercauteren T, Ourselin S, Cardoso M. (2017). Generalized dice overlap as a deep learning loss function for highly unbalanced segmentations, in: *International Workshop on Deep Learning in Medical Image Analysis*. (Québec City, QC, Canada: Springer), pp. 240–8.

34. 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. *Scientific Data*. (2017) 4:1–13. doi: 10.1038/sdata.2017.117

35. 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:1811.02629*. (2018) 11:1–49. doi: 10.48550/arXiv.1811.02629

36. Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. (2020) 10:1–22. doi: 10.48550/arXiv.2010.11929

37. Peng Z, Guo Z, Huang W, Wang Y, Xie L, Jiao J, et al. (2021). Conformer: Local features coupling global representations for visual recognition, in: *International Conference on Computer Vision*. (Montreal, QC, Canada: IEEE), pp. 367–76.

38. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. (Athens, Greece: Springer), pp. 424–32.

39. Milletari F, Navab N, Ahmadi SA. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: *2016 Fourth International Conference on 3D Vision (3DV)*. (Stanford, CA, USA: IEEE), pp. 565–71.

40. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. (2018) 4:1–10. doi: 10.48550/arXiv.1804.03999

41. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z

42. Zhou H, Guo J, Zhang Y, Yu L, Wang L, Yu Y. nnformer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*. (2021) 9:1–10. doi: 10.48550/arXiv.2109.03201

43. Li S, Sui X, Luo X, Xu X, Liu Y, Goh R. Medical image segmentation using squeeze-and-expansion transformers. *arXiv preprint arXiv:2105.09511*. (2021) 5:1–9. doi: 10.24963/ijcai.2021

44. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. (2021). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images, in: *International MICCAI Brainlesion Workshop*. (Strasbourg, France: Springer), pp. 272–84.

45. Wang W, Chen C, Ding M, Yu H, Zha S, Li J. (2021). Transbts: Multimodal brain tumor segmentation using transformer, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. (Strasbourg, France: Springer), pp. 109–19.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 16 April 2025
DOI 10.3389/fonc.2025.1536122

OPEN ACCESS

EDITED BY
Vincenzo Di Nunno,
AUSL Bologna, Italy

REVIEWED BY
Andrea Bianconi,
University of Genoa, Genova, Italy
Giovanni Dima,
University of Pisa, Italy

*CORRESPONDENCE
Xin Zhang
neuro_zx@163.com

RECEIVED 28 November 2024
ACCEPTED 25 March 2025
PUBLISHED 16 April 2025

CITATION
Zhou J, Hou Z, Guan X, Zhu Z, Wang H, Wang C, Luo W, Tian C, Yang H, Ye M, Chen S, Zhang X and Zhang B (2025) The diagnostic value of advanced tracer kinetic models in evaluating high grade gliomas recurrence and treatment response using dynamic contrast-enhanced MRI.
Front. Oncol. 15:1536122.
doi: 10.3389/fonc.2025.1536122

The diagnostic value of advanced tracer kinetic models in evaluating high grade gliomas recurrence and treatment response using dynamic contrast-enhanced MRI

Jianan Zhou[1,2,3], Zujun Hou[4], Xiuqi Guan[5], Zhengyang Zhu[2,3], Han Wang[6], Cong Wang[7], Wei Luo[5], Chuanshuai Tian[1,2,3], Huiquan Yang[2,3], Meiping Ye[2,3], Sixuan Chen[2,3], Xin Zhang[1,2,3]* and Bing Zhang[1,2,3]

[1]Department of Radiology, Nanjing Drum Tower Hospital Clinical College of Nanjing Medical University, Nanjing, China, [2]Institute of Medical Imaging and Artificial Intelligence, Nanjing University, Nanjing, China, [3]Medical Imaging Center, Department of Radiology, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China, [4]The Second Affiliated Hospital and Yuying Children's Hospital, Wenzhou Medical University, Wenzhou, China, [5]FISCA Healthcare Co., Ltd., Nanjing, China, [6]Nanjing Center for Applied Mathematics, Nanjing, China, [7]School of Electronics and Information Engineering, Suzhou Vocational University, Suzhou, China

Background: The purpose of this study was to investigate the diagnostic value of advanced tracer kinetic models (TKMs) in differentiating HGGs recurrence and treatment response.

Methods: A total of 52 HGGs were included. DCE images were analyzed using the following TKMs: distributed parameter (DP), tissue homogeneity (TH), Brix's two-compartment (Brix) and extended-Tofts model (ETM), yielding the following parameters: cerebral blood flow (CBF), mean transit time (MTT), plasma volume (V_p), extravascular volume (V_e), vascular permeability (PS) and first-pass extraction ratio (E) in advanced TKMs (DP, TH and Brix); K^{trans}, V_e, V_p and K_{ep} in ETM. Two delineation methods were conducted (routine scans and parameter heat maps). The differences between two MRI scanners were compared. Mann–Whitney U test was used to assess the difference of parameter values. Diagnostic performance was assessed using the method of the receiver operating characteristic (ROC) curves, with the areas under the ROC curves (AUC) to determine the discriminating power of DCE parameters between recurrent tumor group and treatment response group . $P<0.05$ indicates statistical significance.

Results: The difference on the normalized kinetic parameter value (with respect to contralateral normal-appearing white matter) between two MRI scanners was statistically insignificant ($P>0.05$). MTT and V_p of advanced TKMs were higher in recurrent than in treatment response group ($P<0.05$). For ROI delineated on parameter heat maps, MTT(DP) attained the best performance with AUC 0.88, followed by MTT(TH) and V_p (DP, Brix) with AUCs around 0.80 (0.81, 0.80, 0.79 respectively). The best performance in ETM was V_p (AUC = 0.73).

Conclusion: MTT (DP, TH), and V_p (DP, Brix) could be potential quantitative imaging biomarkers in distinguishing recurrence and treatment response in HGGs.

KEYWORDS

high grade glioma, dynamic contrast-enhanced, tracer kinetic model, treatment response, recurrence

1 Introduction

High-grade gliomas (HGGs) are the most common primary brain malignancies, and the first line of care consists of surgical resection, radiation therapy (RT), and chemotherapy (CTX) (1). The extent of resection has been validated as a prognostic marker (2). After maximal safe resection, the standard therapy (Stupp protocol) remains RT with concurrent temozolomide (TMZ) 75 mg/m2/day for 6 weeks and maintenance TMZ (150–200 mg/m2/day × 5 days for 6 cycles) (3). In spite of the survival benefit associated with adjuvant radiation and chemotherapy, the majority of HGGs patients relapse after initial therapy.

Contrast-enhanced MRI is the gold standard imaging method in detecting HGGs and defining their extension, and is recommended as the standard method for evaluating treatment response in the Response Assessment in Neuro-Oncology (RANO) 2.0 criteria (4, 5). However, using conventional MRI alone, sensitivity and specificity could be limited in distinguishing tumor recurrence from radiation-induced brain injury (RIBI), including pseudoprogression (PsP) and radiation necrosis (RN), collectively known as treatment response, which may both present with enlarging contrast-enhancing lesions or expanding edema. Although advanced imaging techniques have been investigated to improve diagnostic accuracy, the temporal overlap of imaging features between PsP and recurrence (both predominantly occurring 3–6 months post-treatment) complicates definitive diagnosis based on single-timepoint imaging assessments, which relies on multiple follow-ups imaging evaluations, thereby prolonging the diagnostic timeline (6). The incidence of RN could be up to 24% (7), and the incidence of PsP could be up to 32.3% in HGGs patients treated with standard regimen (8–13), which is related to the radiation dose and the volume of brain tissue irradiated (14). The distinction between recurrent tumor and treatment response has important implications for further treatment.

Advanced MRI techniques have been developed to aid in differentiating PsP from true recurrence (15) and a promising representative is dynamic contrast-enhanced (DCE) MRI, which quantitatively measures tissue microcirculation through analyzing the time-intensity curve using tracer kinetic models (TKMs). A variety of TKMs, such as conventional TKMs (e.g., Tofts model and extended-Tofts model [ETM]) and advanced TKMs (e.g., Brix's conventional two-compartment model [Brix], tissue homogeneity model [TH] and distributed parameter [DP] model), have been proposed and investigated in evaluating glioma diagnosis and treatment response, as detailed in a recent review paper (16). A key difference between conventional and advanced TKMs lies in the characterization of tracer molecular transport type in tissue microenvironment. Two types of transport are accounted for in advanced TKMs, namely the transport due to blood flow within the intravascular space and the exchange through vessel wall between the intravascular space and the extravascular space, which is separately modelled as blood flow and vessel permeability. In contrast, only one type of transport is modelled in conventional TKMs. Tofts model is the only single-compartment model, which assumes that the volume of extravascular extracellular space (EES) is much larger than that of intravascular plasma space (IVPS), hence the compartment of IVPS is neglected in the Tofts model. In the above review paper, inconsistent findings in different studies were highlighted and appraised, and advantages of advanced TKMs over conventional TKMs were discussed, but need to be validated in more studies.

In this study, we attempted to investigate the diagnostic value of advanced TKMs and identify potential quantitative imaging biomarkers in differentiating HGG recurrence from treatment response.

2 Materials and methods

2.1 Participants enrollment

This retrospective study was approved by the institutional review board and performed in accordance with the Declaration of Helsinki. Patients in this study were enrolled between December 2022 to May 2024. The requirement for informed consent was waived. The inclusion criteria were as follows: (1) Pathologically diagnosed as HGGs (WHO grade 3 and 4) according to the 2021 World Health Organization (WHO) criteria, with no prior tumor-related treatment before surgery, and receiving synchronous radio-chemotherapy within 72 hours post-operation; (2) Baseline MRI performed about 4 weeks (21-35 days) after the first radiotherapy;

(3) Follow-up scans including T_1WI, T_2WI, T_2-FLAIR, T_1CE (T_1 contrast-enhanced), and DCE; (4) Development of new enhancing lesions during regular follow-up after the initiation of radio-chemotherapy; (5) Pathological confirmation of recurrence or treatment response via reoperation, or clinical diagnosis of recurrence or treatment response during regular follow-up according to the RANO 2.0 criteria (an increase of ≥25% in the product of the two perpendicular diameters of the maximum cross-section of enhancing lesions compared to baseline, or the emergence of new enhancing lesions outside the radiation target area indicating progression). The exclusion criteria included: (1) Patients who did not receive radio-chemotherapy post-surgery; (2) Cases where enhancement of lesions was not significant or poorly defined during follow-up; (3) Poor image quality due to significant patient motion, resulting in failed DCE data processing. The process for patient inclusion is illustrated in Figure 1.

2.2 MRI examinations

The data acquisition machines used in this study included the Philips Ingenia CX 3.0 T MRI scanner and the United Imaging uMR790 3.0 T MRI scanner (equipped with a 32-channel head coil). The contrast agent used was Gadobutrol injection (Oniyin, GE Healthcare), administered at an injection rate of 3.5 mL/s and a dosage of 0.2 mL/kg. The scanning sequences included T_1WI, T_2WI, T_2-FLAIR, T_1CE, and DCE. The DCE sequence included three precontrast sequences with: TR/TE (3.47 ms /1.9 ms), FOV (240 ×220 mm2), slice thickness (5 mm), number of phases (5), flip angles (5°,10°,15°), and the postcontrast dynamic sequence with the same scanning parameters except for number of phases (90) and flip angle (13°). The temporal resolution was 4 seconds, with total duration 6 mins. Specific scanning parameters were shown in Supplementary Table SA1.

2.3 Image analysis

DCE images were analyzed by a commercially available software for DCE data analytics (MItalytics, FISCA Healthcare, Singapore), using the following TKMs: DP, TH, Brix and ETM, yielding the following parameters: cerebral blood flow (CBF), mean transit time (MTT), plasma volume (V_p), extravascular volume (V_e), vascular permeability (PS) and first-pass extraction ratio (E) in advanced TKMs (DP, TH and Brix); K^{trans}, V_e, V_p and K_{ep} in ETM. All of the DCE analysis models were available with the software. Images were registered when evident movement was observed among the dynamic scans. The software used the method of variable flip angle to compute tissue T_1 value and estimated tracer concentration by the difference in longitudinal relaxation rates between postcontrast and precontrast ($r_1 C = \frac{1}{T_1^c} - \frac{1}{T_1^0}$, where r_1 denotes the longitudinal relaxivity and assumes 4.0 $s^{-1}mM^{-1}$ for the contrast agent used in this study (17, 18). Regions of interest (ROI) delineation was performed independently by two neuroradiologists (ZZ and JZ with 9 and 7 years of experience in neuro-radiography). Two types of delineation methods were conducted, one with reference to the routine clinical scans (based on enhanced lesion and areas of necrotic, cystic and hemorrhages were avoided) and the other with account of parameter heat maps (based on the region of highest signal, no less than 15 voxels). Figure 2 showed an example with two types of ROIs drawn. The observers were blinded to pathohistological results. After manual delineation of all datasets, every case was read by both observers to ensure high-quality measurements. Different opinions were resolved by consensus, with a third observer when necessary. ROIs for contralateral normal-appearing white matter were also delineated. Due to limitation in spatial resolution, substantial partial volume corruption could be arisen in the carotid. Hence, a surrogate for the artery input function (AIF), namely the concentration time course in the sagittal sinus, was utilized in this study.

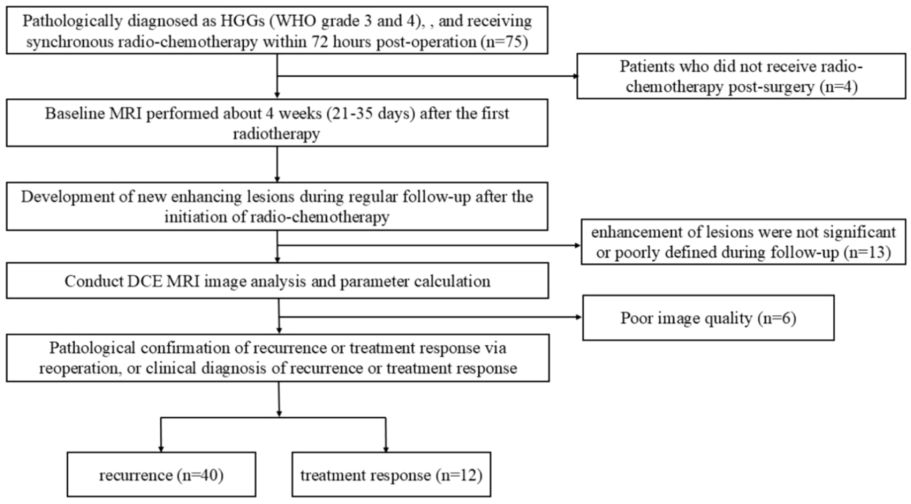


FIGURE 1
Flowchart of patient inclusion.

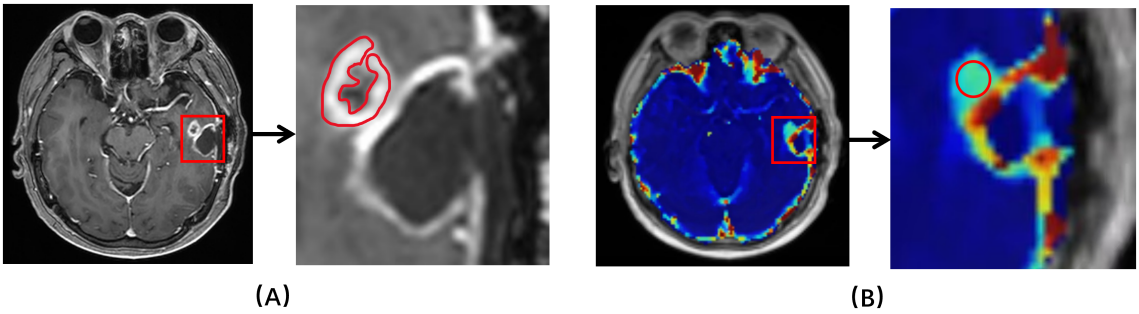


FIGURE 2
An example of ROIs drawn. **(A)** Based on structural T1 contrast-enhanced MRI. **(B)** Based on kinetic parameter maps.

The DCE software provided comprehensive tools for DCE data analytics, including semi-quantitative analysis, conventional TKMs and advanced TKMs. In this study, the following four TKMs were employed for comparison. For completeness, the equations of four TKMs were described as follows. Let $C_{tiss}(t)$ and $C_a(t)$ denote the concentration of contrast agent in the tissue of interest and in the arterial plasma respectively. By assuming that the capillary-tissue system is stationary and linear, these two variables can be related as follows:

$$C_{tiss}(t) = R(t) \otimes C_a(t)$$

where R(t) stands for the impulse residue function and $\otimes$ the convolution operator. The residue functions of four TKMs were listed below.

ETM:

$$R_{ETM}(t) = v_p \delta(t) + K^{trans} \exp\left(-\frac{K^{trans}}{v_e} t\right)$$

Brix (Equations 1a–1c):

$$R_{Brix}(t) = F_p[A \ \exp(\alpha t) + (1 - A)\exp(\beta t)] \tag{1a}$$

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{2}\left[-\left(\frac{PS}{v_p} + \frac{PS}{v_e} + \frac{F_p}{v_p}\right) \pm \sqrt{\left(\frac{PS}{v_p} + \frac{PS}{v_e} + \frac{F_p}{v_p}\right)^2 - 4\frac{PS}{v_e}\frac{F_p}{v_p}}\right] \tag{1b}$$

$$A = \frac{\alpha + \frac{PS}{v_p} + \frac{PS}{v_e}}{\alpha - \beta} \tag{1c}$$

where F_p stands for the flow rate of blood plasma through the IVPS and is generally denoted as CBF when applied to cerebral perfusion imaging.

DP model (Equation 2):

$$R_{DP}(t) = F_p\left\{\begin{array}{l} u(t) - u\left(t - \frac{v_p}{F_p}\right) + \\ u\left(t - \frac{v_p}{F_p}\right)\left\{1 - \exp\left(-\frac{PS}{F_p}\right)\left[1 + \int_0^{t - \frac{v_p}{F_p}} \exp\left(-\frac{PS}{v_e}\tau\right)\sqrt{\frac{PS}{v_e}\frac{PS}{F_p}\frac{1}{\tau}} I_1\left(2\sqrt{\frac{PS}{v_e}\frac{PS}{F_p}\tau}\right) d\tau\right]\right\} \end{array}\right\} \tag{2}$$

where $u(t)$ denotes the Heaviside unit-step function and I_1 is the modified Bessel function.

TH model (Equation 3):

$$R_{TH}(t) = F_p\left\{\left[1 - \exp\left(-\frac{PS}{F_p}\right)\right]\exp\left\{-\frac{F_p}{v_e}\left[1 - \exp\left(-\frac{PS}{F_p}\right)\right](t - \frac{F_p}{v_p})\right\}\right\} \tag{3}$$

After the analysis of a TKM, a few other parameters can be derived as follows. The tracer mean transit time (MTT) can be given by the central volume principle (Equation 4)

$$MTT = \frac{v_p + v_e}{F_p} \tag{4}$$

The first-pass extraction fraction (E) from the IVPS to EES can be evaluated as Equation 5

$$E = 1 - \exp\left(-\frac{PS}{F_p}\right) \tag{5}$$

The primary difference between ETM and another three advanced TKMs lies in that the latter utilizes parameter F_p to account for tracer intravascular transport and parameter PS for tracer exchange between IVPS and EES, whereas the former describes tracer transport using one parameter K^{trans}, which is in principle a mixture between F_p and PS. As for the difference among the advanced TKMs, it largely pertains to how the tracer distributes within a compartment, which is generally described as compartmental (meaning that tracer is well-mixed in the compartment) or distributed (indicating that the tracer distribution is a function of both time and space). Brix assumes both IVPS and EES to be compartmental; DP assumes both IVPS and EES to be distributed; and TH assumes EES to be compartmental and IVPS to be distributed. Interested readers can refer to the review paper (19) for more details of the different tracer kinetic models.

2.4 Statistical analysis

The data was partitioned into tumor recurrent group and treatment response group, based on histopathologic and follow-up imaging and clinical results. For each patient, the parameter values of all voxels within the tumor ROIs on multiple slices were pooled together and the median was determined. Besides the absolute feature values, the relative feature values were also evaluated, where the parameter value was normalized using the

median of contralateral normal-appearing white matter ROI. Interobserver consistency was assessed using intraclass correlation coefficient (ICC) with ICC >0.80, excellent; 0.61–0.80, good; 0.41–0.60, moderate; 0.21–0.40, fair; and <0.20, poor agreement (20). The normality of the distribution of all parameters was analyzed by the Kolmogorov–Smirnov test. Mann–Whitney U test was used to assess the difference of parameter values between recurrent tumor group and treatment response group. The receiver operating characteristic (ROC) curves of all parameters were obtained and the areas under the ROC curves (AUC) were evaluated to determine the discriminating power of DCE parameters between recurrent tumor group and treatment response group. Optimal cut-off values were chosen using the Youden index on the ROC curves, and the corresponding statistical metrics (sensitivity, specificity, accuracy) were computed. To account for the unbalanced issue in the data, the bootstrap re-sampling technique was employed in the study, where re-sampling with replacement was utilized to create a new dataset from original dataset but with predesigned balanced data size, followed by ROC analysis for the resampled dataset, and the process was repeated 200 times, with AUCs (mean ± standard deviation) being calculated. The false discovery rate (FDR) was used to obtain adjusted P values which correct for multiple testing when comparing the various parameters. P<0.05 indicates statistical significance. Statistical analyses were performed using MATLAB (2020b, MathWorks, Natick, MA).

3 Results

3.1 Baseline characteristics of the participants

A total of 52 patients were included, where 40/52 (76.9%) were recurrent and 12/52 (23.1%) had treatment response. Of all patients, 40 were confirmed by follow-up and 12 were confirmed by surgery. The demographic and clinical characteristics of the patient cohort were shown in Table 1.

3.2 Intraclass correlation coefficients

Supplementary Table SA2 showed the ICC values for the measured parameters of DP with delineation in anatomical images and parameter heat maps, where most ICC values in both delineation methods were greater than 0.9, indicating excellent agreement between measurements from two observers. Hence, the parameter values as measured by two observers were averaged and utilized in the subsequent analysis.

3.3 Comparison of kinetic parameter values between MRI scanners

To compare the difference between MRI scanners, Supplementary Table SA3 presented the measured values of kinetic parameters (median followed by interquartile range) by different TKMs, where *P* values of DP-derived parameters were mostly greater than 0.05 except for V_e, indicating only the difference of V_e was statistically significant between two scanners in DP model. Nevertheless, *P* values of other kinetic parameters by other TKMs were largely less than 0.05, indicating significant difference between scanners.

The normalized kinetic parameter values were listed in Table 2, where *P* values of all kinetic parameters by all TKMs were greater than 0.05, suggesting that the difference on the normalized kinetic parameter value between two scanners was insignificant. Hence, the following analysis was largely based on the normalized kinetic parameter values.

3.4 Differential diagnosis between recurrent tumor and treatment response

The normalized kinetic parameter values of recurrent tumor and treatment response by four TKMs were shown in Table 3. Compared with treatment response, lesions with tumor recurrence had higher MTT and V_p using advanced TKMs (DP, TH, Brix) ($P<0.05$). As for ETM, V_e ($P = 0.03$) and V_p ($P = 0.02$) were lower for patients with treatment response compared with patients with tumor recurrence. Figures 3, 4 illustrated the parameter maps of a cases of postoperative recurrence and a case of treatment response of glioblastoma based on DP model respectively, where tumor recurrence was manifested on the parameter maps as higher perfusion and higher permeability compared to treatment response.

Quantitative diagnostic metrics of normalized kinetic parameters derived by four TKMs were shown in Figure 5 (plot of the ROC curves) and Table 4 (optimal cutoff, AUC values, sensitivity, specificity, accuracy), where ROIs were delineated based on parameter heat maps. MTT(DP) attained the best performance in all TKMs' parameters with AUC 0.88, optimal threshold 2.64, specificity 0.92, accuracy 0.77 and sensitivity 0.73. MTT(TH), V_p (DP) and V_p (Brix) had AUCs around 0.80 (0.81, 0.80, 0.79 respectively), with optimal thresholds of 2.25, 2.02, and 8.90 respectively. The best performance in ETM was V_p with AUC 0.73.

Figure 6 and Supplementary Table SA4 presented the ROC curves and corresponding metrics of normalized kinetic parameters derived by four TKMs, where ROIs were delineated based on structural images. MTT(DP) exhibited the largest AUC (0.80) in all TKMs' parameters, with optimal cutoff 2.27, specificity 0.83, accuracy 0.73 and sensitivity 0.70. AUCs of ETM parameters were less than 0.66.

The results of bootstrapping test were summarized in Supplementary Table SA5, where table entries represented AUCs (mean ± standard deviation) of normalized kinetic parameters. The DP-derived MTT showed the largest AUC (0.88 ± 0.05). TH derived MTT, V_p in DP and Brix attained AUCs around 0.80 separately. AUCs by ETM derived parameters were less than 0.75.

10.3389/fonc.2025.1536122

TABLE 1 The demographic and clinical characteristics of the patients.

Variable	Recurrent (n=40)	Treatment response (n=12)	*P* value
Grouping criteria			0.433
Follow-up	30	10	
Pathology	10	2	
Age	56.7 ± 8.33	52.9 ± 14.58	**0.011**
Sex			0.746
Male	20	5	
Female	20	7	
Cerebral lobe			0.304
Frontal lobe	15	5	
Parietal lobe	5	4	
Occipital lobe	7	0	
Temporal lobe	12	3	
Others	1	0	
Location			1.000
Left	19	6	
Right	21	6	
Lesion number			0.743
Single	19	7	
Multiple	21	5	
WHO grade			**0.011**
Grade 3	3	5	
Grade 4	37	7	
Integrated classification			0.121
Glioblastoma, IDH wild-type	33	7	
Astrocytoma, IDH wild-type	2	0	
Astrocytoma, IDH mutant	2	2	
Astrocytoma, NOS	1	2	
Oligodendroglioma, IDH wild-type	1	0	
Oligodendroglioma, IDH mutant	1	0	
Oligodendroglioma, NOS	0	1	
IDH mutation status			**0.018**
Mutant	3	2	
Wild-type	36	7	
NA	1	3	

Bold *P* values less than 0.05 indicate a statistically significant difference. IDH, isocitrate dehydrogenase; NOS, not otherwise specified.

4 Discussion

This paper presented application of advanced TKMs to the differential diagnosis of tumor recurrence and treatment response in HGGs. MTT and V_p of advanced TKMs were higher in recurrent tumor than in treatment response. MTT(DP) attained the largest AUC (0.88). No statistical significance was observed on permeability parameters. Comparatively, advanced TKMs demonstrated advantages over ETM in differentiating glioma recurrence and treatment response.

10.3389/fonc.2025.1536122

TABLE 2 Comparison between MRI scanners on normalized kinetic parameter values in recurrent glioma tissue.

Parameters	United Imaging (n = 30)	Philips (n = 10)	*P* value
DP			
CBF	1.03 (0.88: 1.20)	0.99 (0.84: 1.22)	0.79
MTT	4.85 (2.56: 6.12)	5.43 (2.52: 9.48)	0.57
V_p	4.14 (2.40: 6.54)	4.44 (2.82: 6.95)	0.59
V_e	41.89 (26.53: 69.46)	74.30 (55.37: 106542.51)	0.11
PS	42.35 (17.83: 112.47)	38.46 (20.67: 27965.72)	0.24
E	43.64 (19.51: 94.95)	40.04 (16.37: 27777.53)	0.14
TH			
CBF	1.52 (1.06: 2.54)	1.34 (1.03: 1.87)	0.83
MTT	4.11 (1.65: 5.98)	6.60 (3.50: 9.34)	0.11
V_p	4.72 (3.10: 7.91)	7.20 (5.87: 8.57)	0.37
V_e	40.16 (26.10: 79.80)	64.23 (43.14: 13248755.01)	0.09
PS	25.64 (10.53: 56.83)	54.54 (16.26: 926318.71)	0.17
E	14.59 (7.37: 23.85)	25.02 (10.39: 1447508.26)	0.11
Brix			
CBF	1.54 (1.14: 1.89)	1.28 (1.06: 1.98)	0.25
MTT	11.11 (7.76: 17.98)	17.94 (11.83: 26.70)	0.50
V_p	18.60 (9.03: 30.59)	22.94 (12.61: 39.74)	0.74
V_e	49.09 (27.33: 85.92)	46.61 (24.20: 86.11)	0.11
PS	32.60 (12.90: 54.18)	37.33 (23.29: 72.01)	0.09
E	20.22 (11.38: 39.10)	27.97 (16.82: 46.87)	0.11
ETM			
K^{trans}	1.95 (0.79: 5.71)	3.96 (1.04: 13.15)	0.55
V_e	245.42 (46.55: 7925.95)	133.37 (53.31: 18122.45)	0.94
K_{ep}	0.14 (0.01: 0.27)	0.17 (0.01: 0.26)	0.37
V_p	3.17 (1.60: 6.00)	4.09 (2.10: 5.86)	0.33

CBF, cerebral blood flow; MTT, mean transit time; V_p, plasma volume; V_e, extravascular volume; PS, vascular permeability; E, first-pass extraction ratio.

Previous studies (21–23) on differential diagnosis between recurrent glioma and treatment response using DCE-MRI were largely based on conventional TKMs, which characterize the transport of contrast agent using K^{trans} and are recommended in the Quantitative Imaging Biomarkers Alliance (QIBA) Profile (24), where it is claimed that measured change in K^{trans} of a brain lesion of 21% or larger from DCE-MRI data at 1.5T indicates that a true change has occurred with two-sigma confidence (95%) confidence. In theory, K^{trans} is defined as the exchange rate of contrast agent from the blood vessels into the surrounding interstitial space and represents an important parameter in conventional TKMs. Zahra et al. reviewed 29 studies (total of 1194 patients) that correlate DCE-MRI with histopathological or clinical outcome data relevant to radiotherapy, and found the apparent discrepancy among the reported outcomes, which could be attributed to the heterogeneity in the methods, including the selection of the ROIs and the acquisition and analysis of the DCE-MRI data, as well as the small numbers of patients recruited in some studies (25). O'Connor and coauthors reviewed the role of DCE-MRI for decision making during the drug-development process in about 100 early-phase clinical trials and investigator-led studies of targeted antivascular therapies and found that, unlike serological assays, K^{trans} often had variable meanings between different clinical studies and within one study at different time points, which hindered wider application and acceptance of DCE-MRI in clinical practice (26). The precise meaning of K^{trans} has been

10.3389/fonc.2025.1536122

TABLE 3 Summary of normalized kinetic parameter values (median and interquartile range) of recurrent tumor and treatment response by four TKMs.

Parameters	Recurrent tumor (n=40)	Treatment response (n=12)	*P* value
DP			
CBF	1.00 (0.87: 1.21)	0.95 (0.81: 1.15)	0.38
MTT	4.85 (2.54: 6.79)	1.55 (1.31: 2.37)	**<0.01**
V_p	4.14 (2.41: 6.63)	1.51 (1.27: 3.16)	**<0.01**
V_e	47.01 (29.99: 77.22)	33.85 (9.81: 8400.57)	0.23
PS	42.35 (19.25: 133.96)	37.95 (15.72: 314.18)	0.72
E	42.04 (17.94: 114.67)	46.04 (17.95: 2655.28)	0.99
TH			
CBF	1.42 (1.05: 2.39)	1.74 (1.10: 1.95)	0.82
MTT	4.40 (2.19: 7.38)	1.49 (1.07: 2.22)	**<0.01**
V_p	5.83 (3.53: 8.49)	3.19 (1.68: 4.47)	**<0.01**
V_e	46.10 (29.25: 88.10)	35.63 (12.49: 144.96)	0.31
PS	28.62 (11.04: 66.55)	29.33 (3.44: 79.16)	0.45
E	14.59 (8.11: 30.77)	10.27 (2.93: 46.27)	0.39
Brix			
CBF	1.48 (1.13: 1.92)	1.17 (0.83: 1.85)	0.24
MTT	11.98 (8.33: 19.48)	4.53 (3.35: 14.39)	**0.02**
V_p	18.92 (9.57: 31.25)	7.92 (4.11: 10.67)	**<0.01**
V_e	49.09 (26.09: 86.02)	18.63 (12.10: 40429.65)	0.10
PS	35.19 (15.16: 61.08)	19.04 (3.47: 294.36)	0.37
E	23.54 (12.73: 42.52)	35.92 (4.39: 415.56)	0.94
ETM			
K^{trans}	2.87 (0.89: 7.69)	2.08 (0.25: 4.88)	0.38
V_e	144.59 (45.94: 12718.69)	38.90 (5.69: 129.81)	**0.03**
V_p	3.17 (1.82: 6.18)	1.33 (0.75: 3.48)	**0.02**
K_{ep}	0.17 (0.01: 0.28)	0.17 (0.04: 1.05)	0.43

Bold *P* values less than 0.05 indicate a statistically significant difference. CBF, cerebral blood flow; MTT, mean transit time; V_p, plasma volume; V_e, extravascular volume; PS, vascular permeability; E, first-pass extraction ratio.

theoretically investigated in (27, 28). It is understood that the physiological significance is tissue dependent; if the contrast uptake is flow limited, then K^{trans} will indicate the tissue perfusion, whereas if the uptake is permeability limited, then K^{trans} indicates the permeability. In most cases, K^{trans} indicates a combination of the blood flow and vessel wall permeability properties of tissue. The AUC of normalized K^{trans} (ETM) was 0.59 in differentiating recurrent tumor from treatment response in our study, which corroborated the previous studies (AUCs 0.62 and 0.51, respectively) (29, 30).

The primary mechanisms underlying treatment-related responses in HGGs, including RN and PsP, involve radiation-induced direct damage, injury to vascular endothelial cells, and excessive vascular proliferation or rupture leading to hemorrhage and plasma protein extravasation, which disrupt the blood-brain barrier (BBB), and cell death releases cytokines (e.g., IL-6, TNF-α) and mediators, triggering an inflammatory response that activates cells (31), all of which may contribute to increased vascular permeability. Due to the infiltrative growth characteristics of gliomas, residual tumor cell proliferation can lead to tumor recurrence. Tumor cells secrete pro-angiogenic factors, inducing abnormal vascular proliferation with incomplete basement membranes, resulting in contrast agent extravasation. In summary, the enhancing lesions in treatment-related responses (PsP and RN) primarily arise from therapy-induced inflammatory reactions and vascular permeability changes, whereas tumor recurrence-driven enhancement is driven by tumor cell proliferation and abnormal angiogenesis (32, 33). Consequently, both tumor recurrence and treatment-related responses in HGGs can exhibit elevated permeability parameter values on DCE MRI, which explains the lack of significant statistical differences in permeability parameters between the two groups in this study. Larsen et al.'s study also indicated that BBB permeability parameters could not effectively distinguish between PsP and recurrence (34), which was consistent with our study. Besides, Manual delineation of enhancing regions may include areas of coexisting treatment response and tumor recurrence, particularly in infiltrative gliomas, obscuring true permeability differences. The non-significant permeability results highlight the complexity of PsP/recurrence pathophysiology and the limitations of parameter imaging. This underscores the necessity of combining permeability data with systemic inflammatory markers (e.g., NLR, SII) and volumetric analyses to enhance diagnostic precision (6).

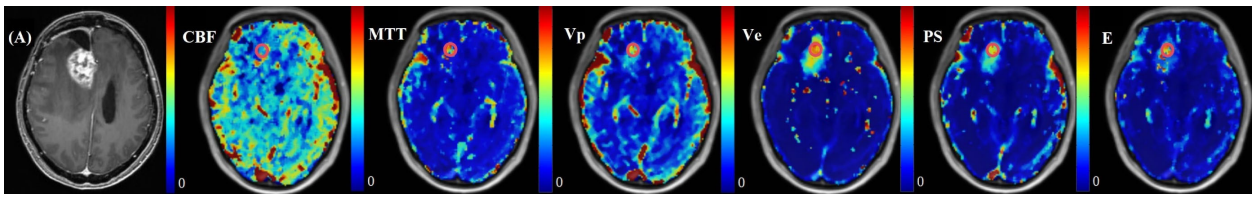


FIGURE 3
A 36-year-old female with WHO grade 4 glioblastoma of the right frontal lobe. A new enhanced lesion emerged during the follow-up. Postoperative pathology confirmed the enhanced lesion as recurrence. Parametric maps of cerebral blood flow CBF, mean transit time MTT, plasma volume V_p, extravascular volume V_e, vascular permeability PS, and first-pass extraction ratio E as derived using DP model.

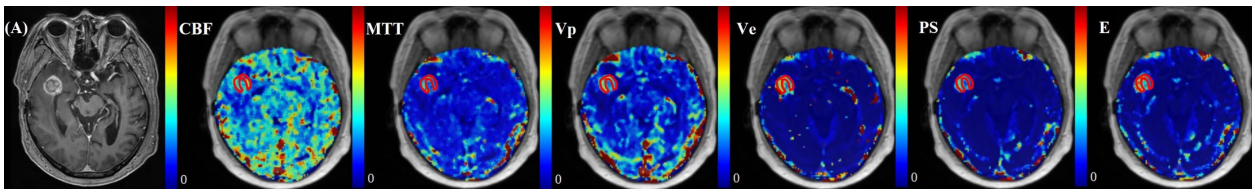


FIGURE 4
A 57-year-old male with WHO grade 3 oligodendroglioma of the right temporal lobe. Postoperative pathology confirmed the enhanced lesion as radiation necrosis (RN). Parametric maps of blood flow CBF, mean transit time MTT, fractional volume of intravascular space V_p, fractional volume of interstitial space V_e, vessel permeability PS, and extraction ratio E as derived using DP model.

Quantitative interpretation of kinetic parameter maps has two approaches in practice, namely, delineation based on anatomical images or parameter heat maps. The former approach defines lesions and their boundaries from correlative routine scans which have higher spatial resolution in interpreting tissue structures and are acquired in the same imaging plane as DCE-MRI (with similar FOV and spatial coverage) such as T_2WI and T_1CE images. This approach has been recommended by the committee of QIBA (24) for reproducibility. In the recent guidelines of both National Comprehensive Cancer Network (NCCN) and European Association of Neuro-Oncology (EANO), perfusion maps (in particular, the map of cerebral blood volume) are recommended to define metabolic hotspots for specific tumor tissue sampling (35, 36). This study compared these two approaches to lesion ROI delineation and demonstrated that parameter heat maps could be more accurate in distinguishing recurrent tumor from treatment response in high-

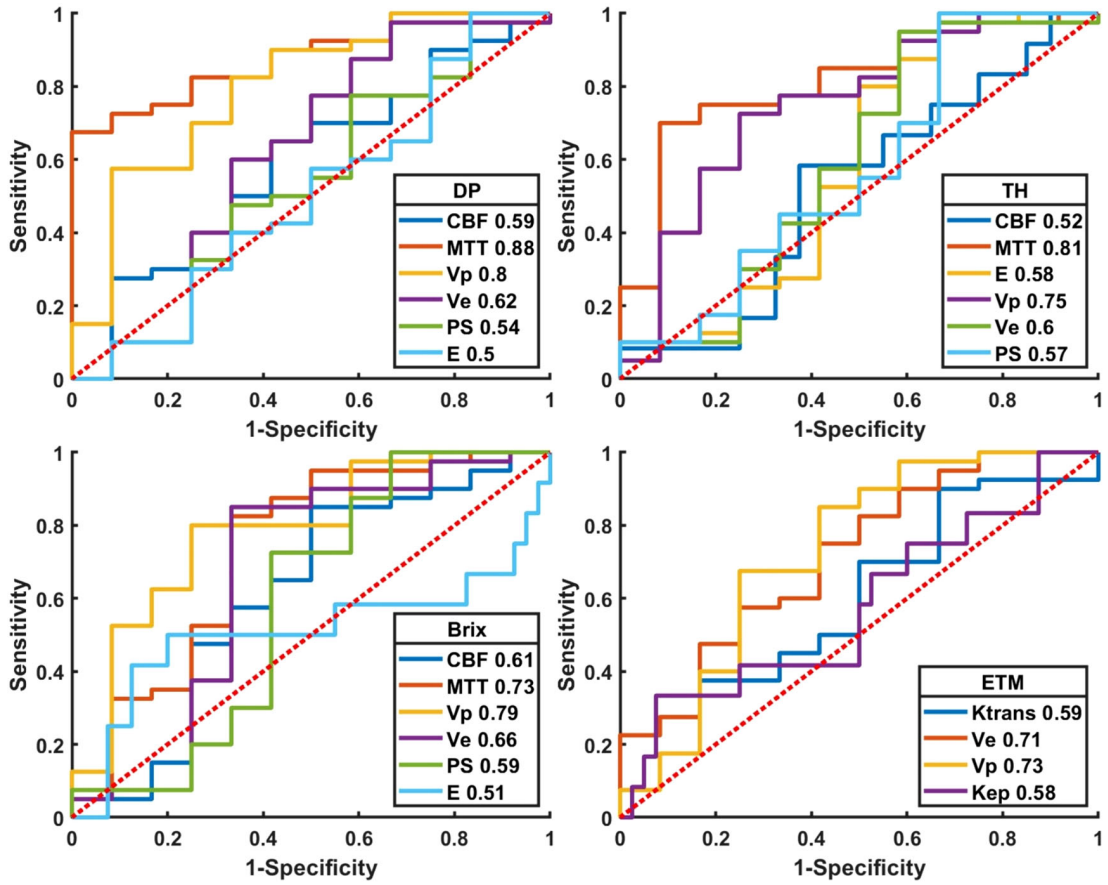


FIGURE 5
Plot of ROCs of normalized kinetic parameters derived by four tracer kinetic models in differentiating HGGs recurrence from treatment response. DP: distributed parameter model, TH: tissue homogeneity model, Brix: Brix's conventional two-compartment model, ETM: extended-Tofts model. Regions of interest (ROI) were delineated on kinetic parameter heat maps.

10.3389/fonc.2025.1536122

TABLE 4 ROC quantitative parameters of normalized kinetic parameters derived by four TKMs in differential diagnosis between recurrent HGG and treatment response, where ROIs were delineated on kinetic parameter heat maps.

Parameters	AUC	Threshold	Sensitivity	Specificity	Accuracy
DP					
CBF	0.59	0.97	0.65	0.58	0.63
MTT	**0.88**	2.64	0.73	0.92	0.77
V_p	0.80	2.02	0.83	0.67	0.79
V_e	0.62	40.7	0.60	0.67	0.62
PS	0.54	50.49	0.48	0.67	0.52
E	0.50	38.49	0.58	0.50	0.56
TH					
CBF	0.52	1.71	0.58	0.63	0.62
MTT	0.81	2.25	0.75	0.83	0.77
V_p	0.75	3.92	0.73	0.75	0.73
V_e	0.60	31.91	0.73	0.50	0.67
PS	0.57	38.33	0.45	0.67	0.50
E	0.58	7.27	0.80	0.50	0.73
Brix					
CBF	0.61	1.05	0.85	0.50	0.77
MTT	0.73	7.72	0.83	0.67	0.79
V_p	0.79	8.90	0.80	0.75	0.79
V_e	0.66	20.57	0.85	0.67	0.81
PS	0.59	20.78	0.73	0.58	0.69
E	0.51	50.00	0.50	0.80	0.73
ETM					
K^{trans}	0.59	1.20	0.70	0.50	0.65
V_e	0.71	46.17	0.75	0.58	0.71
K_{ep}	0.58	0.14	0.67	0.48	0.52
V_p	0.73	2.44	0.68	0.75	0.69

Bold MTT (DP) attained the best performance with the best AUC = 0.88.
CBF, cerebral blood flow; MTT, mean transit time; V_p, plasma volume; V_e, extravascular volume; PS, vascular permeability; E, first-pass extraction ratio.

grade glioma patients. The reason might be the heterogeneity of suspicious lesion as delineated in anatomical images, which could compromise the subsequent differential diagnosis where the information was based on the measured parameter values of the registered ROI, likely a mixture of heterogeneous tissue. Comparatively, delineation based on parameter heat maps could yield an ROI with more homogeneous tissue.

As discussed in the latest review paper (16), both MR imaging hardware and the theory of DCE tracer kinetic modeling have undergone significant advances over the years, thereby allowing acquisition of DCE images with higher temporal resolution, better signal-to-noise ratio, wider brain coverage and increased spatial resolution, and enabling separate quantification of CBF and PS in advanced TKMs. This study demonstrated clearly the advantages of advanced TKMs over ETM in differential diagnosis of recurrent tumor and treatment response.

A long-standing challenge in DCE-MRI is the reproducibility of quantitative results across imaging platforms. Standardization of imaging protocol and data post-processing is essential to achieve the purpose. Towards that end, QIBA has recommended the following protocol: (1) using 3D fast spoiled gradient recalled echo sequence; (2) using variable flip angle for T1-mapping measurement; (3) scanning parameters stay constant; (4) dynamic scan duration up to 6 mins; (5) temporal resolution less than 5

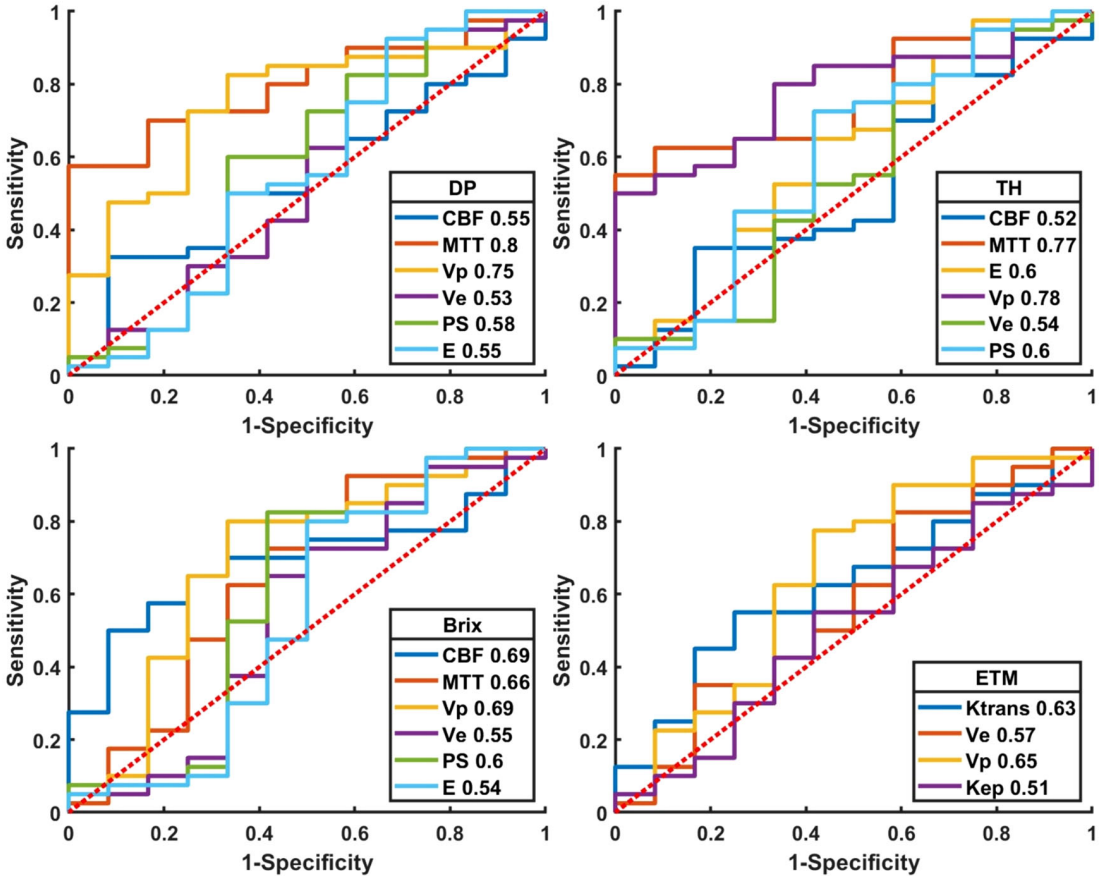


FIGURE 6
Plot of ROCs of normalized kinetic parameters derived by four tracer kinetic models in differential diagnosis between recurrent HGG and treatment response. DP: distributed parameter model, TH: tissue homogeneity model, Brix: Brix's conventional two-compartment model, ETM: extended-Tofts model. Regions of interest (ROI) were delineated on structure images.

seconds. This study acquired data using MR scanners from two vendors, but the imaging protocol has been attempted to follow the same standard, as largely recommended by QIBA, with the equivalent sequence, the same temporal resolution and the same brain coverage. Though some measured values in some TKMs showed significant difference between scanners, it turned out that the difference of the normalized value is statistically insignificant, which indicated the potential for the imaging protocol and the current DCE-MRI processing flow to fulfill the promise of using DCE-MRI as a clinically useful tool.

There are several limitations in our study. First, this was a single-center retrospective study with a moderate sample size. Second, the delineation of ROI was subjective, and the results might be biased, especially for lesions with unclear enhancement. Third, portion of data was evaluated based on follow-up results of imaging and clinical signs, which might be different from histopathological results.

5 Conclusion

In differentiating recurrence and post-treatment response in HGGs, DP demonstrated the best performance, with parameter MTT having the highest diagnostic performance. Moreover, MTT (TH) and V_p (DP, Brix) could also serve as potential quantitative imaging biomarkers. The kinetic parameters derived by advanced TKMs yielded superior performance compared to those by conventional ETM. Interpretation of TKM parameters in terms of treatment response assessment was best performed in the heat maps of kinetic parameters.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by the Medical Research Ethics Committee and the Institutional Review Board of the Nanjing Drum Tower Hospital (Protocol number 2023-563-02). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

JZ: Conceptualization, Funding acquisition, Methodology, Project administration, Writing – original draft, Writing – review & editing. ZH: Conceptualization, Methodology, Writing – review & editing. XG: Conceptualization, Methodology, Writing – review & editing. ZZ: Conceptualization, Methodology, Writing – original draft. WH: Formal Analysis, Writing – review & editing. CW: Formal Analysis, Writing – review & editing. WL: Formal Analysis, Writing – review & editing. CT: Conceptualization, Data curation, Methodology, Writing – review & editing. HY: Writing – review & editing. MY: Writing – review & editing. SC: Writing – review & editing. XZ: Funding acquisition, Writing – review & editing. BZ: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by National Science and Technology Innovation 2030 – Major program of "Brain Science and Brain-Like Research" (2022ZD0211800); Fundings for Clinical Trials from the Affiliated Drum Tower Hospital, Medical School of Nanjing University (2022-LCYJ-MS-25, 2021-LCYJ-PY-20); Postgraduate Research & Practice Innovation Program of Jiangsu Province (JX22014155).

Conflict of interest

Authors XG and WL were employed by the company FISCA Healthcare Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1536122/full#supplementary-material

References

1. Segura PP, Quintela NV, García MM, del Barco Berrón S, Sarrió RG, Gómez JG, et al. SEOM-GEINO clinical guidelines for high-grade gliomas of adulthood (2022). *Clin Trans Oncol.* (2023) 25:2634–46. doi: 10.1007/s12094-023-03245-y

2. Karschnia P, Young JS, Dono A, Häni L, Sciortino T, Bruno F, et al. Prognostic validation of a new classification system for extent of resection in glioblastoma: A report of the RANO resect group. *Neuro Oncol.* (2022) 25:940–54. doi: 10.1093/neuonc/noac193

3. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *New Engl J Med.* (2005) 352:987–96. doi: 10.1056/NEJMoa043330

4. McKinnon C, Nandhabalan M, Murray SA, Plaha P. Glioblastoma: clinical presentation, diagnosis, and management. *BMJ.* (2021) 374:n1560. doi: 10.1136/bmj.n1560

5. Wen PY, van den Bent M, Youssef G, Cloughesy TF, Ellingson BM, Weller M, 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. doi: 10.1200/JCO.23.01059

6. Satragno C, Schiavetti I, Cella E, Picichè F, Falcitano L, Resaz M, et al. Systemic inflammatory markers and volume of enhancing tissue on post-contrast T1w MRI images in differentiating true tumor progression from pseudoprogression in high-grade glioma. *Clin Trans Radiat oncol.* (2024) 49:100849. doi: 10.1016/j.ctro.2024.100849

7. Brandsma D, Stalpers L, Taal W, Sminia P, van den Bent MJ. Clinical features, mechanisms, and management of pseudoprogression in Malignant gliomas. *Lancet Oncol.* (2008) 9:453–61. doi: 10.1016/S1470-2045(08)70125-6

8. Taal W, Brandsma D, de Bruin HG, Bromberg JE, Swaak-Kragten AT, Sillevis Smitt PAE, et al. Incidence of early pseudo-progression in a cohort of Malignant glioma patients treated with chemoirradiation with temozolomide. *Cancer.* (2008) 113:405–10. doi: 10.1002/cncr.v113:2

9. Brandes AA, Franceschi E, Tosoni A, Blatt V, Pession A, Tallini G, et al. MGMT promoter methylation status can predict the incidence and outcome of pseudoprogression after concomitant radiochemotherapy in newly diagnosed glioblastoma patients. *J Clin Oncol.* (2008) 26:2192–7. doi: 10.1200/JCO.2007.14.8163

10. Chaskis C, Neyns B, Michotte A, De Ridder M, Everaert H. Pseudoprogression after radiotherapy with concurrent temozolomide for high-grade glioma: clinical observations and working recommendations. *Surg Neurol.* (2009) 72:423–8. doi: 10.1016/j.surneu.2008.09.023

11. Gunjur A, Lau E, Taouk Y, Ryan G. Early post-treatment pseudo-progression amongst glioblastoma multiforme patients treated with radiotherapy and temozolomide: A retrospective analysis. *J Med Imaging Radiat Oncol.* (2011) 55:603–10. doi: 10.1111/j.1754-9485.2011.02319.x

12. Young RJ, Gupta A, Shah AD, Graber JJ, Zhang Z, Shi W, et al. Potential utility of conventional MRI signs in diagnosing pseudoprogression in glioblastoma. *Neurology*. (2011) 76:1918–24. doi: 10.1212/WNL.0b013e31821d74e7

13. Blakstad H, Mendoza Mireles EE, Heggebø LC, Magelssen H, Sprauten M, Johannesen TB, et al. Incidence and outcome of pseudoprogression after radiation therapy in glioblastoma patients: A cohort study. *Neuro-Oncol Pract*. (2024) 11:36–45. doi: 10.1093/nop/npad063

14. Knudsen-Baas KM, Moen G, Fluge Ø, Storstein A. Pseudoprogression in high-grade glioma. *Acta Neurol Scand*. (2013) 127:31–7. doi: 10.1111/ane.2012.127.issue-s196

15. Tsakiris C, Siempis T, Alexiou GA, Zikou A, Sioka C, Voulgaris S, et al. Differentiation between true tumor progression of glioblastoma and pseudoprogression using diffusion-weighted imaging and perfusion-weighted imaging: systematic review and meta-analysis. *World Neurosurge*. (2020) 144:e100–e9. doi: 10.1016/j.wneu.2020.07.218

16. Zhou J, Hou Z, Tian C, Zhu Z, Ye M, Chen S, et al. Review of tracer kinetic models in evaluation of gliomas using dynamic contrast-enhanced imaging. *Front Oncol*. (2024) 14. doi: 10.3389/fonc.2024.1380793

17. Rohrer M, Bauer H, Mintorovitch J, Requardt M, Weinmann HJ. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest radiol*. (2005) 40:715–24. doi: 10.1097/01.rli.0000184756.66360.d3

18. Shen Y, Goerner FL, Snyder C, Morelli JN, Hao D, Hu D, et al. T1 relaxivities of gadolinium-based magnetic resonance contrast agents in human whole blood at 1.5, 3, and 7 T. *Invest Radiol*. (2015) 50:330–8. doi: 10.1097/RLI.0000000000000132

19. Koh TS, Bisdas S, Koh DM, Thng CH. Fundamentals of tracer kinetics for dynamic contrast-enhanced MRI. *J Magnet Resonance Imag*. (2011) 34:1262–76. doi: 10.1002/jmri.22795

20. Zou X, Luo Y, Li Z, Hu Y, Li H, Tang H, et al. Volumetric apparent diffusion coefficient histogram analysis in differentiating intrahepatic mass-forming cholangiocarcinoma from hepatocellular carcinoma. *J Magnet Resonance Imag*. (2018) 49:975–83. doi: 10.1002/jmri.26253

21. Nael K, Bauer AH, Hormigo A, Lemole M, Germano IM, Puig J, et al. Multiparametric MRI for differentiation of radiation necrosis from recurrent tumor in patients with treated glioblastoma. *Am J Roentgenol*. (2018) 210:18–23. doi: 10.2214/AJR.17.18003

22. Shin KE, Ahn KJ, Choi HS, Jung SL, Kim BS, Jeon SS, et al. DCE and DSC MR perfusion imaging in the differentiation of recurrent tumour from treatment-related changes in patients with glioma. *Clin Radiol*. (2014) 69:e264–e72. doi: 10.1016/j.crad.2014.01.016

23. Bisdas S, Naegele T, Ritz R, Dimostheni A, Pfannenberg C, Reimold M, et al. Distinguishing recurrent high-grade gliomas from radiation injury: a pilot study using dynamic contrast-enhanced MR imaging. *Acad Radiol*. (2011) 18:575–83. doi: 10.1016/j.acra.2011.01.018

24. Anon. QIBA MR Biomarker Committee. *MR DCE Quantification, Quantitative Imaging Biomarkers Alliance*. Profile Stage: Public Comment. 2020–10-12. Available at: https://qibawiki.rsna.org/index.php/Profiles.

25. Zahra MA, Hollingsworth KG, Sala E, Lomas DJ, Tan LT. Dynamic contrast-enhanced MRI as a predictor of tumour response to radiotherapy. *Lancet Oncol*. (2007) 8:63–74. doi: 10.1016/S1470-2045(06)71012-9

26. O'Connor JPB, Jackson A, Parker GJM, Roberts C, Jayson GC. Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. *Nat Rev Clin Oncol*. (2012) 9:167–77. doi: 10.1038/nrclinonc.2012.2

27. Sourbron SP, Buckley DL. On the scope and interpretation of the Tofts models for DCE-MRI. *Magnet Resonance Med*. (2011) 66:735–45. doi: 10.1002/mrm.22861

28. Koh TS, Hennedige TP, Thng CH, Hartono S, Ng QS. Understanding Ktrans: a simulation study based on a multiple-pathway model. *Phys Med Biol*. (2017) 62:n297–319. doi: 10.1088/1361-6560/aa70c9

29. Seeger A, Braun C, Skardelly M, Paulsen F, Schittenhelm J, Ernemann U, et al. Comparison of three different MR perfusion techniques and MR spectroscopy for multiparametric assessment in distinguishing recurrent high-grade gliomas from stable disease. *Acad Radiol*. (2013) 20:1557–65. doi: 10.1016/j.acra.2013.09.003

30. Zakhari N, Taccone MS, Torres CH, Chakraborty S, Sinclair J, Woulfe J, et al. Prospective comparative diagnostic accuracy evaluation of dynamic contrast-enhanced (DCE) vs. dynamic susceptibility contrast (DSC) MR perfusion in differentiating tumor recurrence from radiation necrosis in treated high-grade gliomas. *J Magnet Resonance Imag*. (2019) 50:573–82. doi: 10.1002/jmri.26621

31. Thust SC, van den Bent MJ, Smits M. Pseudoprogression of brain tumors. *J Magnet Resonance Imag*. (2018) 48:571–89. doi: 10.1002/jmri.26171

32. Mahase S, Rattenni RN, Wesseling P, Leenders W, Baldotto C, Jain R, et al. Hypoxia-mediated mechanisms associated with antiangiogenic treatment resistance in glioblastomas. *Am J Pathol*. (2017) 187:940–53. doi: 10.1016/j.ajpath.2017.01.010

33. Parvez K, Parvez A, Zadeh G. The diagnosis and treatment of pseudoprogression, radiation necrosis and brain tumor recurrence. *Int J Mol Sci*. (2014) 15:11832–46. doi: 10.3390/ijms150711832

34. Larsen VA, Simonsen HJ, Law I, Larsson HB, Hansen AE. Evaluation of dynamic contrast-enhanced T1-weighted perfusion MRI in the differentiation of tumor recurrence from radiation necrosis. *Neuroradiology*. (2013) 55:361–9. doi: 10.1007/s00234-012-1127-4

35. Nabors LB, Portnow J, Ahluwalia M, Baehring J, Brem H, Brem S, et al. Central nervous system cancers, version 3.2020, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw*. (2020) 18:1537–70. doi: 10.6004/jnccn.2020.0052

36. Weller M, van den Bent M, Preusser M, Le Rhun E, Tonn JC, Minniti G, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol*. (2020) 18:170–86. doi: 10.1038/s41571-020-00447-z

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 21 May 2025
DOI 10.3389/fonc.2025.1530332

OPEN ACCESS

EDITED BY
Camilla Russo,
Santobono-Pausilipon Children's Hospital, Italy

REVIEWED BY
Lanchun Lu,
The Ohio State University, United States
Joel Poder,
St George Hospital, Australia

*CORRESPONDENCE
Mark De Ridder
mark.deridder@uzbrussel.be

†These authors share last authorship

RECEIVED 18 November 2024
ACCEPTED 21 April 2025
PUBLISHED 21 May 2025

CITATION
Belloeil-Marrane T, Gutierrez A, Boussaer M, Teixeira C, Gevaert T and De Ridder M (2025) Virtual phantom methodology for assessment of MRI distortion correction in high-precision stereotactic radiosurgery treatment planning. *Front. Oncol.* 15:1530332. doi: 10.3389/fonc.2025.1530332

Virtual phantom methodology for assessment of MRI distortion correction in high-precision stereotactic radiosurgery treatment planning

Tristan Belloeil-Marrane, Adrian Gutierrez, Marlies Boussaer, Cristina Teixeira, Thierry Gevaert† and Mark De Ridder*†

Department of Radiotherapy, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Brussels, Belgium

Introduction: The accuracy of stereotactic treatment planning is primarily limited by the least accurate process in the whole chain of events, and is particularly important in cranial radiosurgery. Ameliorating this process can improve treatment targeting, providing additional reliability for these indications. Quality assurance (QA) in radiotherapy is often performed on the dose delivery and planning section rather than the localization. Magnetic Resonance Images (MRI) are notably subject to distortions, due to the nonlinearity of gradient fields, potentially source of geometric errors. This study aimed to analyze the impact of a patient-specific algorithm, rather than manufacturer-specific, to correct spatial distortion in cranial MRI by using a novel software-only paradigm.

Material and methods: An unbiased simulated T1-Weighted MRI validated dataset is utilized to create a synthetic CT (sCT). By introducing controlled distortion in simulated datasets, we can evaluate the influence of noise and intensity non-uniformity ("RF") ranging from 0 to 9% noise and 0 to 40% RF. These MRIs were corrected using the sCT as base modality for distortion correction. To evaluate the impact of the distortion correction, each corrected/non-corrected image set was compared to the unbiased MRI using Root-mean-square-error (RMSE) as a full-image reference comparison metric.

Results: The distortion correction allows for an improvement based on the RMSE correlation between baseline and distorted MRIs. The amelioration of average RMSE in corrected versus non-corrected MRI is up to 42.22% for the most distorted datasets.

Conclusion: The distortion correction results show a proportional improvement with increased noise and intensity non-uniformity. This provides additional robustness and reliability to the accuracy of SRS treatment planning using MR T1-W sequences as imaging reference for target definition and organ delineation, remaining consistent independently from the variability of the non-uniformity gradient values. This virtual phantom methodology primarily aims to provide a simple/robust evaluation metric in radiotherapy for MR distortion correction solutions, providing an additional/complement QA procedure to dedicated

hardware phantoms, comparatively costly in time and resources. This approach is also designed to assist with an easily implementable secondary QA for validation during commissioning of distortion correction software, focusing on this feature, to better isolate and identify sources of geometric errors resulting from MR distortions.

KEYWORDS

MRI distortion correction, stereotactic radiosurgery (SRS), cranial indications, target positioning accuracy, synthetic CT (sCT), magnetic resonance imaging (MRI), quality assurance (QA)

1 Introduction

Magnetic Resonance Imaging (MRI) is one of the most utilized imaging modalities in Cranial Stereotactic Radiosurgery (SRS). Computed Tomography (CT) images are the gold standard in radiation oncology and are used for diagnostic, contouring, and dose calculation. This modality is considered to have superior spatial accuracy compared to MRI, but does not provide sufficient anatomical information for target definition and delineation of the organs at risk (OAR). MRI is required to enrich the treatment targeting in SRS treatments with sufficient anatomical data, thanks to an excellent soft tissue contrast (1).

During the treatment planning workflow, specifically within the contouring process, CT and MR images are rigidly fused to allow the projection of structures defined in one dataset to be displayed in the other. This registration is the key component to allow for simultaneous, accurate cranial structure contouring and dose calculation. MR images are prone to intrinsic distortion introduced during their acquisitions and might not be fully corrected by the MR scanner's post-processing software. These remaining distortions can be a source of inaccuracy, resulting in potential incorrect target definition, sub-optimal protection of critical structures, and/or increased dose to normal tissue (2). The displacement of the treatment target linked to MRI distortions, mispositioning can lead to a geometric miss during delivery. This potential error is particularly crucial in SRS indications focusing on high-dose irradiation to small lesions, usually with a diameter inferior to 1cm. In these cases, a geometric deviation of 1mm or more could significantly impact the dose coverage of the target and, as well, increase the dose to the normal tissues. Moreover, the stereotactic target volume margins can be adjusted by increasing their size to compensate for potential geometric misses and guarantee sufficient dose coverage, leading to more than doubling the additional normal tissue volume receiving high doses for each 1mm margin increment. As a reference, for a sphere with a diameter of 1 cm, a 1mm margin will expand its initial volume by 33%, 73% with a 2mm margin, and 120% with a 3mm margin (3).

To verify accurate lesion targeting throughout SRS and SBRT treatments, the AAPM-RSS Medical Physics Practice Guideline 9.a (4). recommends an End-to-End (E2E) localization assessment "hidden target test" using an SRS frame and/or IGRT/SGRT system of 1 mm additionally stating that when developing the E2E tests, all aspects of the treatment process should be considered, including immobilization, simulation, respiratory management, treatment planning, and treatment delivery using a clinically relevant image guidance method. Systematic submillimeter E2E testing is necessary for SRS and requires continuous patient-specific quality assurance (QA), including discrete MRI correction distortion QA, considering the amplitude of potential displacement. For end-to-end testing, dedicated SRS-specific or anthropomorphic phantoms are typically used to define the overall error from image acquisition to radiation delivery However, the accuracy of a stereotactic treatment is primarily limited by the least accurate process in the whole chain of events. QA is often performed on the dose delivery and planning section rather than the target localization. The AAPM Task Group 284 (5) recommends a geometrical accuracy of ≤ 2 mm across a 25 cm field of view (FOV) for SRS and radiotherapy with MR-only planning.

During the SRS treatment process, the influence of the MRI inaccuracies associated with distortions persisting after initial scanner-level correction in the imaging QA is often overlooked and included in the broader treatment planning error. The principal challenge in asserting the accuracy of the MRI is that not all the distortions follow a linear gradient. Some distortions, the result of a gradient non-linearity, are referred to as B-spline distortions (6). Scanner manufacturers include some image reconstruction and correction processes during the post-processing of the images. Multiple software manufacturers have developed and validated elastic fusion in radiation oncology planning to improve the accuracy of target and critical organs definition. Institutions tend not to include distortion correction in their protocols, as the process remains a "black box" with little to no tools to assess the quality of the correction.

This study aims to define a new methodology based on a novel software-only paradigm. For this, we want to be software-agnostic and provide a robust and effortless technique that can be easily replicated in clinical institutions without requiring specific hardware and saving time and resources, particularly on medical imaging devices. This approach provided an adequate method to

evaluate and prove the quality and effectiveness, as well as validate the clinical use of post-acquisition scanner processing distortion correction software using non-biased data along with appropriate metrics to comprehend the influence of the process on different defined distortion variables and how they are correlated.

2 Materials and methods

2.1 MR images and distortions generation

Since the geometrical displacement is non-linear, the incremental parameter should reflect it. For this purpose, BrainWeb MRI (7), an online interface that generates a set of realistic simulated brain databases (SBD), was utilized for a non-biased MRI simulation. This quantitative analysis of the image data approach was developed to provide a ground truth for such image sets to resolve the issue of validation for these sequences. The different parameters and values were estimated and validated to provide a realistic range of values to emulate the potential distortions generated during MR acquisition and provide a gradual set of determined variables to better quantify how distortions affect cerebral anatomy radiomics. The MR datasets anatomically encompass the integrality of the cranium, cerebrum, cerebellum, and brainstem from the top of the scalp to the base of the foramen magnum. To generate a high SNR ratio model, 27 high-resolution MR datasets of the same individual with normal anatomy were acquired, subsampled, and intensity averaged, resulting in a single simulated dataset (8–11).

MRI distortions consist of various hardware-related factors, such as magnetic field inhomogeneity and gradient non-linearity, along with patient-related aspects like chemical shift and magnetic susceptibility. Given the multiple variables, there is a consequent challenge in accurately identifying and attributing whether each one significantly impacts each specific image set and, moreover, in selecting optimal MR imaging parameters. To simulate a standard approach that does not depend on unique anatomical or environmental conditions, this study focuses on gradient non-linearity as the primary source of geometric distortions in this imaging modality (3). For that purpose, this study utilizes MRI datasets that were artificially generated using 2 variables: noise or percent noise (PN) (Gaussian noise percent multiplied by the brightest tissue intensity) and intensity non-uniformity (INU or radiofrequency (RF)) were introduced with defined incremental values in MR images to simulate the effects of distortions. The advantage of that technique is that the MR T1-Weighted (T1WI) dataset (PN=0%, RF=0%) used as the reference for correction distortion is considered "ground truth" (or gold standard) for the modality and is simulated from normal brain anatomy without distortion.

The INU fields were estimated from patient MRI scans. They are not linear but are slowly varying fields of a complex shape. The % value specifies the intensity non-uniformity level. For a 20% level, the multiplicative INU field has a range of values of [0.90 - 1.10] over the brain area. For other INU levels, the field is linearly scaled accordingly (i.e. to a range of [0.80 - 1.20] for a 40% level). INU distortions can be spatially smooth, not fully reproducing heterogeneous biological interactions within large anatomical structures. As a consequence, they can be easily interpreted and/or interpolated, which does not represent the clinical reality in MR imaging. To emulate the complexity of the imaging in the human head, incremental inhomogeneity magnitudes can be introduced to prevent automatic registration or segmentation from simply anticipating the INU distortions. For that purpose, Gaussian noise is specifically used to provide an additional layer of convolution to the INU (12). The noise parameter utilizes Rayleigh statistics in the background and Rician statistics in the signal regions. The PN number represents the percent ratio of the standard deviation of the white Gaussian noise versus the signal for a reference tissue. For the MR T1WI, the reference tissue for the noise computation used was the white matter.

For this study, all noise and RF values available in the BrainWeb model were analyzed: 17 simulated brain anatomy MRI T1WI datasets with 1x1x1 mm resolution and 181 slices using values ranging from 0 to 9% noise and from 0 to 40% RF.

2.2 Synthetic CT Conversion

In clinical routine, CT images are fused to the MRI for target and tissue density accuracy. However, they are subject to their intrinsic distortions and motion/reconstruction artifacts. The use of a synthetic CT (sCT) based on the validated non-distorted MRI was introduced to ensure that the CT images used as a base for correction distortion were not introducing any additional error in the process, particularly since using newly acquired CT from physical phantoms could still lead to some minor image processing geometrical errors during the acquisition and CT images would not match exactly with the MR images resulting in minor differences during the rigid registration. The sCT dataset was generated from the MRI baseline dataset (PN=0%, RF=0%) using a 1:1 voxel equivalence (13). The MRI baseline was segmented semi-automatically using ITK-SNAP 3.6.0 (14, 15) to differentiate grey and white matter, CSF, eyes, bone, air, and the rest of the soft tissues (16). The MR Arbitrary Units in each segment were then converted to Hounsfield Units (17) using bulk assignments and directly or inverse linear ranges as described in Table 1. To cope with the segmentation irregularities, a series of filters was applied, including a median filter, to ensure that the root fusion CT would mirror exactly the initial MR without distortion.

2.3 MRI distortion correction

As part of the SRS treatment-specific planning solution currently utilized clinically in our institution, Elements Cranial Distortion Correction© software version 4.0 (Brainlab AG, München, Germany) was selected and employed to correct the MR distortions remaining post-acquisition.

TABLE 1 Conversion table and methods from MR arbitrary units to Hounsfield Units for all defined tissue types used to generate a synthetic CT.

Tissue Type	MR Abitratry Units/HU Values Conversion Method	Hounsfield units (HU)
Bone	Inverse Linear	500 to 1100
Brain	Linear	35 to 75
Water (CSF, eyes)	Bulk Assignement	0
Soft tissues	Inverse Linear	−80 to −30
Air	Bulk Assignement	−1000

These values were slightly adapted from Yu et al. (15) and are in agreement with observed and published tissue density ranges.

The sCT and simulated MRI datasets were imported in DICOM format and losslessly converted into a proprietary format. Within the software, the sCT is fused rigidly independently to all the MR datasets using mutual information registration. This process aligns the overall position of each MRI dataset to the CT positioning (no specific region of interest defined). The results of the fusion were verified visually using anatomical landmarks (e.g. ventricles, hemispherical midline, sulci,…) (Figure 1).

Following the initial rigid fusion, multiple rigid unit registrations of 3×3×3 cm3 overlapping units were performed within the images to locally improve the fusion. These units are then aggregated into a single deformation field that maps one of the datasets onto the other (18). This type of elastic registration can be referred to as "Multi-rigid" since multiple rigid fusions are applied at once, providing a blending of local registration results.

The multi-rigid fusions were applied independently for each MR Dataset in the distortion correction software using the sCT as a base for the distortion correction (Figure 2). The results of the registration were verified visually once more. The elastic registration software generates a corrected MRI dataset that can be exported while retaining the original MRI data. All the MR image datasets were exported in DICOM RT 3.0 format.

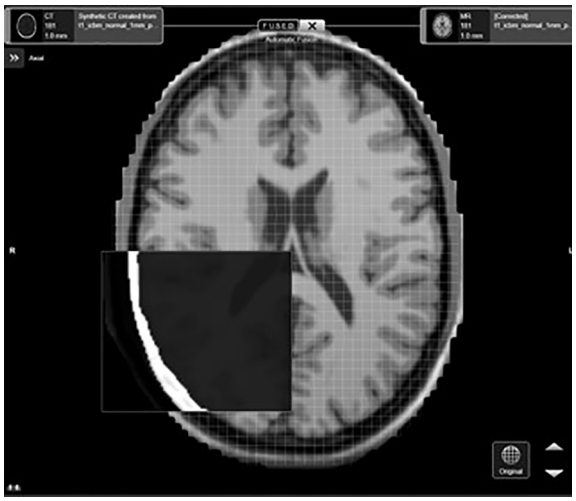


FIGURE 1
Display of the post-processed results of an MRI distorted dataset based on the synthetic CT using the cranial distortion correction software.

2.4 Evaluation of the correction

The root mean square error (RMSE) calculation is used to compare all the voxels (19) in the baseline MR against the corrected MR and the non-corrected MR for defined PN and RF values in the corresponding slice (Figure 3). In this study, the average RMSE was calculated using custom code in the MATLAB® software (MathWorks©, Natick, Massachusetts) to compare all the slices of the corrected to the non-corrected MR dataset according to the following equation:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

where y_i is the value of a specific voxel in a selected MR image, $\hat{y}_i$ is the value of the aligned voxel in the baseline undistorted MR image and n is the number of voxels compared in the image.

2.5 Statistical analysis

A student t-test was performed to test differences between two groups/variables (corrected/non-corrected). Statistical significance was considered for $p < 0.05$.

3 Results

On average, the distortion correction software allows for an improvement based on the RMSE correlation between the baseline MRI (PN=0%, RF=0%) and the distorted MRI corrected and non-corrected (Figure 4). For the corrected datasets, the RMSE ranges between 8.83 ± 0.27 and 55.58 ± 1.06 and averages at 33.36 ± 3.51. For the non-corrected datasets, the RMSE ranges between 7.70 ± 0.03 and 83.14 ± 0.52 and averages at 44.18 ± 5.61. A strong, statistically significant difference was found for the Average RMSE Corrected compared to Non-corrected, as observed in Table 2.

As a control, we measured the RMSE between the MRI baseline dataset (PN=0%, RF=0%) and its associated corrected MRI dataset. The average RMSE value was 14.01 (standard deviation SD 6.27) and serves as the baseline RMSE value comparing the undistorted reference images to other MR image sets.

The % difference of average RMSE in non-corrected versus corrected MRI ranges from 0 to 42.22%. A negative outlier (-14.56%) for the comparison at PN1, RF0 was found which was inconsistent with the rest of the data. This can be linked to the influence of the baseline value for lower values of PN and RF. The overall average, including the outlier value, was 21.08 ± 4.16%. This

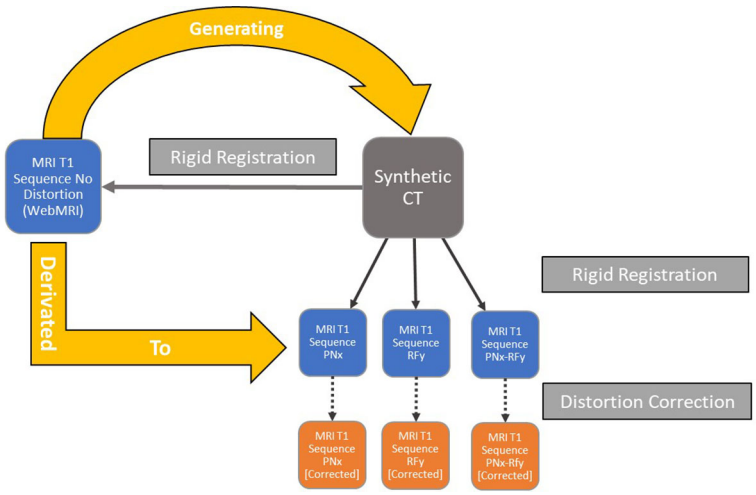


FIGURE 2
Workflow of the datasets' generations and semi-rigid/rigid registrations for each iteration of the simulated data with different noise and intensity non-uniform parameters.

confirms the overall increased accuracy, and the linear trendline (Table 2) shows a positive correlation with PN and RF values (R2 = 0.67).

3.1 RF correction

The combined averaged PN values from 0 to 9% for each RF value were calculated to evaluate the influence of distortion correction on the RF% %. For the corrected datasets, the RMSE ranges between 20.73 (SD 15.69) and 44.02 (SD 7.99) and averages at 31.46 ± 6.79 (SD 11.45). For the non-corrected datasets, the RMSE ranges between 33.52 (SD 28.33) and 55.25 (SD 17.98) and averages at 43.56 ± 6.33 (SD 22.98). The results between non-corrected and corrected RMSE average percentages showed a significant improvement for all RF values: RF0% = 38.18%, RF20% = 31.72%, and RF40% = 22.30% with an average of 29.27% ± 5.16%. Statistical significance (p=0.014) was found for the compared RF RMSE values.

3.2 PN correction

The combined averaged RF=0% RMSE for each PN value was calculated to evaluate the influence of distortion correction on the PN% %. For the corrected datasets, the RMSE ranges between 8.83 and 55.58, and averages at 20.73 ± 6.41 (SD 15.69). For the non-corrected datasets, the RMSE ranges between 7.70 and 72.96, and averages at 33.52 ± 11.57 (SD 28.33). The results between non-corrected and corrected RMSE average percentages showed significant improvements (p= 0.026) for PN values equal and superior to 1%: PN0 = 0% PN1 = -14.56%, PN3 = 34.11%, PN5 = 39.74%, PN7 = 40.75% and, PN9 = 42.22% with an average of 23.71 ± 10.04%.

4 Discussion

The premises of our study were formulated on the lack of a consistent and easy-to-implement clinical method to assess the quality of the distortion correction in MRI. The currently available

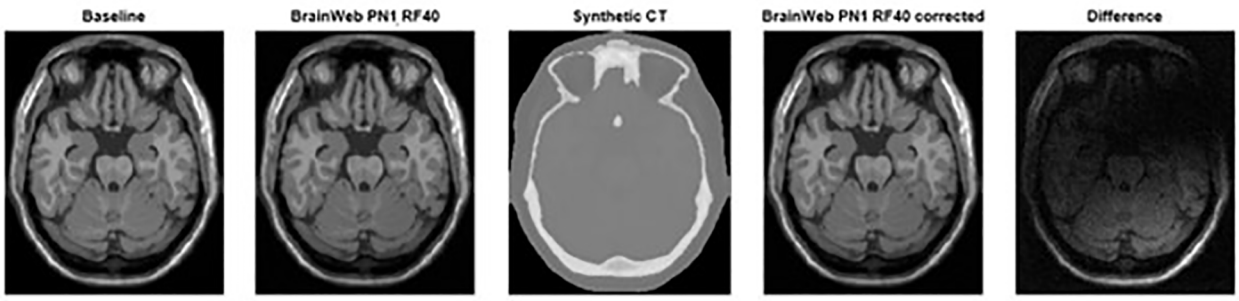


FIGURE 3
From left to Right, baseline MRI (PN0%, RF0%), distorted MRI (PN1%, RF40%), synthetic CT, and corrected distorted MRI (PN1%, RF40% - based on sCT correction) datasets including the difference observed using image subtraction between corrected and non-corrected datasets.

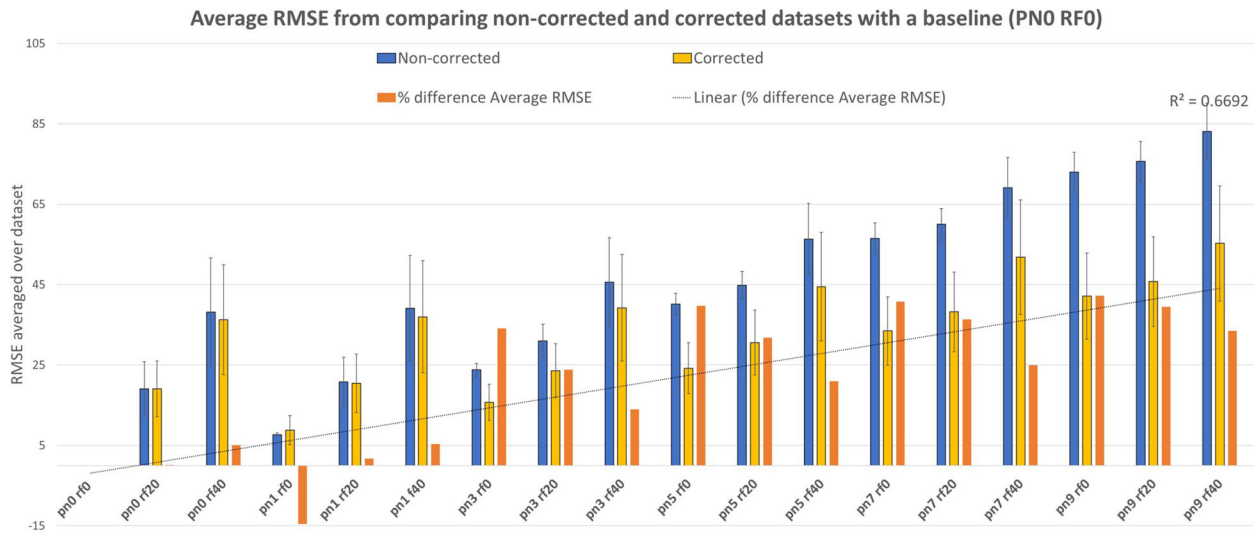


FIGURE 4
Average RMSE with linear trendline (R-squared = 0.67) comparing non-corrected and corrected datasets for PN% ranging from 0 to 9% and RF values ranging from 0 to 40% using a non-corrected baseline MRI (PN=0%, RF=0%).

options require having a dedicated MR phantom and acquiring the image data on CT and MRI scanners (20). This methodology was implemented to address this specific issue and independently assess the specific process of MRI distortion correction.

Accurate target definition is always crucial for cranial SRS as it will influence the overall treatment accuracy since the targeted volumes are small, typically ranging from 0.3cc to 50cc (21, 22), and are heavily dependent on MRI. This is even more predominant for functional SRS: Luo et al. (23) compared the positioning of the SRS treatment isocenter with the ventral intermediate (VIM) nucleus of the thalamus during thalamotomies and the tremor treatment response post-irradiation, concluding on the critical importance of submillimetric accuracy for these specific indications.

That is one of the reasons assessing the effect of the distortion correction is primordial to avoid inaccurate treatment delivery leading to diminished response and normal tissue toxicity.

Our results have shown that the Elements Cranial Distortion Correction© software was improving the accuracy of target and critical structures delineation in MR T1 sequences by mitigating geometric misses resulting from gradient nonlinearity distortions. According to the full-image RMSE results, the distortion correction is positively correlated to the simultaneous increment of the PN and RF. Moreover, the more the datasets are distorted, the more efficient the software will be. These correction effects were also observable for respectively augmenting the PN (average: 23.71% for Non-Corrected vs. Corrected datasets) and RF values (average: 29.27% for Non-Corrected vs. Corrected datasets) in the MR images. This supports an added reliability to the quality of the correction independently from the type of distortion in MR T1WI images.

Our results align with the results of the clinical validation of the software by the manufacturer where the spatial correlation between rigidly and elastically fused images was assessed through Euclidean distance, They found an improvement in fusion accuracy of the 1.5/3.0T MRI by 0.41 ± 0.95 mm (18).

The results of our study showed non-significant improvements for lower PN and RF % values correlated to the notable baseline observed using the RMSE value (14.01) from the MRI baseline datasets (PN=0%, RF=0%), inferior to the average standard deviation for both corrected and non-corrected datasets with their respective RMSE values of 14.06 and 25.93. This is most likely related to the limit of the distortion correction accuracy. Karger et al. (2006) (24) reported the radial distance correction with device-specific 2D and 3D image distortion correction algorithms in multiple scanners with different magnetic B-field strengths in Tesla (T) for cranial indications. They concluded that the image distortions superior to 2 mm were significantly reduced, but not significantly for distortions inferior to 2 mm due to gradient non-linearities. Bagherimofidi et al. (2019) (25) have reported similar findings using their specific distortion correction algorithm in a head phantom: the average error varied from 0.258 to 0.557 mm with a maximum error of 1.492 mm with diameter distances from 20 to 80 mm from the isocenter, confirming a baseline error post- correction distortion. The study of Retif et al. also aligns with our results as it was demonstrated in phantom and clinical patient data that the Elements Distortion Correction software was able to reduce the mean and standard deviation datasets, particularly in the maximum distortion in heavily distorted images, significantly reducing the number of control points with > 0.5-mm distortion. Furthermore, these results were consistent across acquisitions from different scanner makes, models, and magnetic field strengths (26). Image quality metrics (IQMs) such as RMSE and structural similarity index (SSIM) were the core processes for assessing the correction direction using a full-reference quality metric (27). These are commonly used in the evaluation and optimization of MRI acquisition and reconstruction strategies, including MRI distortion measurements. Root Mean Square

TABLE 2 Values of the average RMSE in 1 and percentage difference comparing non-corrected and corrected datasets for PN% ranging from 0 to 9% and RF values ranging from 0 to 40% using a non-corrected baseline MRI (PN=0%, RF=0%).

MRI Dataset	pn0 rf0	pn0 rf20	pn0 rf40	pn1 rf0	pn1 rf20	pn1 rf40	pn3 rf0	pn3 rf20	pn3 rf40	pn5 rf0	pn5 rf20	pn5 rf40	pn7 rf0	pn7 rf20	pn7 rf40	pn9 rf0	pn9 rf20	pn9 rf40
RMSE between non-corrected datasets																		
Average Value	0	19.11	38.21	7.7	20.78	39.09	23.81	31.02	45.63	40.14	44.87	56.31	56.53	60.02	69.12	72.96	75.69	83.14
Standard Deviation	0	6.74	13.48	0.44	6.14	13.13	1.58	4.1	11.05	2.72	3.39	8.95	3.81	3.93	7.54	4.94	4.96	6.97
Minimum Value	0	6.89	13.75	7.05	10.89	16.13	21.74	24.58	29.65	37.05	38.98	42.3	52.41	54.01	56.3	67.98	69.02	70.95
Maximum Value	0	39.97	79.93	8.5	40.57	80.24	26.9	45.66	82.65	45.45	54.26	87.37	64.46	67.07	93.74	82.87	84.9	102.53
RMSE between corrected datasets																		
Average Value	0	19.07	36.28	8.83	20.43	37.01	15.69	23.61	39.25	24.19	30.61	44.51	33.49	38.22	51.86	42.16	45.81	55.25
Standard Deviation	0	6.92	13.68	3.62	7.27	13.94	4.49	6.66	13.28	6.34	8.07	13.48	8.52	9.91	14.26	10.72	11.15	14.3
Minimum Value	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum Value	0	34.65	66.85	22.77	38.58	67.23	27.36	37.19	67.95	35.61	44.57	69.97	40.97	49.6	75.02	53.32	57.67	75.37
% difference RMSE between non-corrected and corrected																		
Average Value	0	0.17	5.04	-14.56	1.68	5.32	34.11	23.87	13.98	39.74	31.79	20.95	40.75	36.32	24.97	42.22	39.48	33.55
Minimum Value	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Maximum Value	0	13.31	16.37	-167.77	4.91	16.21	-1.72	18.53	17.79	21.65	17.86	19.92	36.44	26.04	19.97	35.66	32.07	26.49

10.3389/fonc.2025.1530332

Error (RMSE) is utilized in diverse fields of study to compare 2 images and has been a standard metric in medical images (28). The advantages and justifications of this measurement tool in the scope of our study were its ease of use, rapid calculation time, and robustness of results, permitting for more simple and accessible testing. More importantly, it allows us to compare the quality of the entire image to a baseline and, therefore is more inclusive than other methods. This is particularly important as the distortions are non-linear and comparison of points or structures could be misleading since they do not provide an overview of the image and might lead to misinterpretation of the quality of the dataset.

This is also a better fit for the verification of the MRI distortion correction as distortions are not uniform and punctual evaluation can lead to misinterpretation of the absolute local values and/or large amplitude between the minimum and maximum deviations. In the absence of fiducials or implanted markers, comparing anatomical landmarks or delineations of specific organs, due to the non-uniformity nature of the MR distortions, these values can be significantly different depending on the anatomical position of the sampled landmark or contour and do not reflect the influence of distortions for the entire image set or even in other anatomical cranial regions. Moreover, the operator can introduce additional errors when comparing or delineating organs, which can be further accentuated when multiple operators are involved. Intra- and inter-operator variabilities are common issues in assessing the quality of organ contouring and can be avoided by using a full-image index of similarity such as the RMSE.

Regarding the choice of metrics, out of all the IQMs, the RMSE was the most efficient and simple to implement. The other commonly used IQM: SSIM is also a strong measuring metric, that, however, requires a more complex implementation and calculating power. Based on the study of Mason et al. (28), the simple and rapid algorithms of RMSE demonstrated short calculation times (all less than 2 seconds). SSIM has slightly longer calculation times (less than 20 seconds). It also demonstrated that SSIM does not show a significantly stronger correlation with radiologists' observation of diagnostic image quality than RMSE.

This assessment method is not limited to the virtual phantom and can be applied similarly to hardware phantoms. These highly specialized apparatuses may already be accepted as the standard for process validation and periodic verification in the established QA protocol; the technique described in this study would grant an additional level of reliability and a secondary check to supplement the existing approach. This would further ensure the accuracy and consistency of the measurement. In addition, both virtual and hardware can be used concurrently following this method to the same purpose, adding the benefit of a strong correlation as a result of relying on a common metric to establish the validity and effectiveness of the distortion correction.

In the radiation oncology treatment planning, the CT and MRI T1WI are the 2 imaging modalities mandatorily included for all cranial indications. This study focused on T1-weighted imaging for this reason. For specific SRS cranial indications, other sequences are required, such as T2-weighted images (29). We aim to evaluate the quality of the distortion correction in subsequent studies using this measurement method and validate the software for all possible cranial MRI sequences.

Papas et al. (2017) (30) quantified the influence of distortion correction in RT planning by employing a phantom study. They concluded that for targets inferior to 20 mm in diameter, spatial disposition of the order of 1 mm could significantly affect plan acceptance/quality indices. For targets with a diameter greater than 2cm, the corresponding disposition was found to be greater than 1.5mm. It underlines the relevance of target accuracy in SRS treatment delivery. This effect is magnified in treatments of simultaneous multiple lesions with a single isocenter; translational and rotational deviations of isocenter as small as 0.5 mm and 0.5 deg. in the treatment delivery could lead to significant dosimetric impact as suggested by literature (31–33). For such indications, the distortion correction process would help to improve the accuracy and, logically, decrease the margin needed to treat the metastasis to preserve normal brain tissue.

This current study has reviewed a specific distortion algorithm based on multi-rigid registration. Other commercial software is available and makes use of different methods of elastic fusion. The continuation goal of our study is to use the newly defined method to further evaluate other distortion correction algorithms or techniques and provide a comprehensive comparison with different software and modalities for cranial SRS and MR-only treatment planning. We aim that these kinds of distortion correction algorithms will become more and more important with the emergence of 7T scanners, providing higher signal-to-noise ratio, spatial resolution, and contrast for clinical applications in SRS and neurosurgery (34). It has been shown in the literature that increasing the magnetic B-field strength (T) was correlated to increased distortions (35).

The ease of use and availability of the data for the implementation of this QA method can be further utilized in parallel to other available QA options, including those relying on an MR-dedicated phantom, offering the possibility of a secondary validation and providing additional details on the distortion correction process. Future studies will aim to assess the correlation between the different approaches and further investigate the complementarity of different QA techniques to improve the commissioning, validation, and daily verification of distortion correction processes in SRS treatment planning as well as introduce other MRI acquisition variables to assess their influence in the quality of the distortion correction and ensure that all QA methods align on results.

5 Conclusion

The distortion correction of MR T1-weighted images is a requirement to add robustness and reliability to the target definition ensuring accurate and consistent cranial treatment planning, particularly for SRS indications. The described distortion correction evaluation method based on non-biased datasets with defined parameter values and relying on standard medical image quality metrics has demonstrated the facility of isolating and assessing the quality of this specialized process with simple

software tools available for every institution. With this novel approach using a simulated virtual phantom, we are able to provide additional validation related to the image datasets' accuracy needed for dedicated cranial indications in radiosurgery, MR-only treatments as well as neurosurgical functional indications, and can be further utilized alongside other QA methods to add a secondary validation/verification method.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author/s.

Author contributions

TB-M: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Investigation, Methodology, Software. AG: Writing – original draft, Writing – review & editing, Data curation, Methodology, Software. MB: Writing – review & editing, Investigation, Methodology, Software, Validation. CT: Writing – review & editing, Data curation, Investigation, Methodology, Validation. TG: Writing – original draft, Writing – review & editing, Conceptualization, Investigation, Methodology, Supervision, Validation. MD: Writing – review & editing, Conceptualization, Methodology, Resources, Supervision.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

TB is financially supported by Brainlab Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest..

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

1. Schmidt MA, Payne GS. Radiotherapy planning using MRI. *Phys Med Biol*. (2015) 60:R323–61. doi: 10.1088/0031-9155/60/22/R323

2. Walker A, Liney G, Metcalfe P, Holloway L. MRI distortion: considerations for MRI-based radiotherapy treatment planning. *Australas Phys Eng Sci Med*. (2014) 37:103–13. doi: 10.1007/s13246-014-0252-2

3. Seibert TM, White NS, Kim GY, Moiseenko V, McDonald CR, Farid N, et al. Distortion inherent to magnetic resonance imaging can lead to geometric miss in radiosurgery planning. *Pract Radiat Oncol*. (2016) 6:e319–28. doi: 10.1016/j.prro.2016.05.008

4. Halvorsen PH, Cirino E, Das IJ, Garrett JA, Yang J, Yin F-F, et al. AAPM-RSS medical physics practice guideline 9.a. for SRS-SBRT. *J Appl Clin Med Phys*. (2017) 18:10–21. doi: 10.1002/acm2.12146

5. Glide-Hurst CK, Paulson ES, McGee K, Tyagi N, Hu Y, Balter J, et al. Task Group 284 report: magnetic resonance imaging simulation in radiotherapy: considerations for clinical implementation, optimization, and quality assurance. *Med Phys*. (2021) 48:e636–70. doi: 10.1002/mp.14695

6. Mullen M, Garwood M. Contemporary approaches to high-field magnetic resonance imaging with large field inhomogeneity. *Prog Nucl Magn Reson Spectrosc*. (2020) 120-121:95–108. doi: 10.1016/j.pnmrs.2020.07.003

7. BrainWeb. Simulated Brain Database (2024). Available online at: http://www.bic.mni.mcgill.ca/brainweb/ (Accessed October 20, 2024).

8. Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, et al. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging*. (Orlando Fla., USA: Academic Press) (1998) 17(3):463–8. doi: 10.1109/42.712135

9. Kwan RK-S, Evans AC, Pike GB. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans Med Imaging*. (1999) 18:1085–97. doi: 10.1109/42.816072

10. Kwan R, Evans A, Pike B. An extensible MRI simulator for post-processing evaluation. *Visualization in Biomedical Computing (VBC'96)*. (1996) 1131. doi: 10.1007/BFb0046947

11. Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, et al. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging*. (1998) 17:463–8. doi: 10.1109/42.712135

12. Hadjidemetriou S, Studholme C, Mueller S, Weiner M, Schuff N. Restoration of MRI data for intensity non-uniformities using local high order intensity statistics. *Med Image Anal*. (2009) 13:36–48. doi: 10.1016/j.media.2008.05.003

13. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol*. (2017) 12:28. doi: 10.1186/s13014-016-0747-y

14. ITK-SNAP. (2024). Available online at: http://www.itksnap.org/ (Accessed October 20, 2024).

15. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*. (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015

16. Yu H, Oliver M, Leszczynski K, Lee Y, Karam I, Sahgal A. Tissue segmentation-based electron density mapping for MR-only radiotherapy treatment planning of brain using conventional T1-weighted MR images. *J Appl Clin Med Phys*. (2019) 20:11–20. doi: 10.1002/acm2.12654

17. Broder J, Preston R. Chapter 1 - Imaging the Head and Brain. In: Broder J, editor. *Diagnostic Imaging for the Emergency Physician*. Philadelphia, PA, USA: W.B. Saunders (2011), p. 1–45. doi: 10.1016/B978-1-4160-6113-7.10001-8

18. Gerhardt J, Sollmann N, Hiepe P, Kirschke JS, Meyer B, Krieg SM, et al. Retrospective distortion correction of diffusion tensor imaging data by semi-elastic image fusion - Evaluation by means of anatomical landmarks. *Clin Neurol Neurosurg*. (2019) 183:105387. doi: 10.1016/j.clineuro.2019.105387

19. Weavers PT, Tao S, Trzasko JD, Shu Y, Tryggestad EJ, Gunter JL, et al. Image-based gradient non-linearity characterization to determine higher-order spherical harmonic coefficients for improved spatial position accuracy in magnetic resonance imaging. *Magn Reson Imaging*. (2017) 38:54–62. doi: 10.1016/j.mri.2016.12.020

20. Paštyková V, Novotný J, Veselský T, Urgošík D, Liščák R, Vymazal J. Assessment of MR stereotactic imaging and image co-registration accuracy for 3 different MR

scanners by 3 different methods/phantoms: phantom and patient study. *J Neurosurg*. (2018) 129:125–32. doi: 10.3171/2018.7.GKS181527

21. Jackie Wu Q-R, Wessels BW, Einstein DB, Maciunas RJ, Kim EY, Kinsella TJ. Quality of coverage: Conformity measures for stereotactic Radiosurgery. *J Of Appl Clin Med Phys*. (2003) 4:374–81. doi: 10.1120/jacmp.v4i4.2506

22. Andrews DW, Scott CB, Sperduto PW, Flanders AE, Gaspar LE, Schell MC, et al. Whole brain radiation therapy with or without stereotactic radiosurgery boost for patient with one to three brain metastases: phase III results of the RTOG 9508 randomised trial. *Lancet*. (2004) 363:1665–72. doi: 10.1016/S0140-6736(04)16250-8

23. Luo G, Cameron BD, Wang L, Yu H, Neimat JS, Hedera P, et al. Targeting for stereotactic radiosurgical thalamotomy based on tremor treatment response. *J Neurosurg*. (2021) 29:1–8. doi: 10.3171/2021.7.JNS21160

24. Karger CP, Höss A, Bendl R, Canda V, SChad L. *Accuracy of device-specific 2D and 3D image distortion correction algorithms for magnetic resonance imaging of the head provided by a manufacturer* Vol. 51. Bristol, England: IOP Publishing Ltd- Physics in Medicine & Biology (2006).

25. Bagherimofidi SM, Yang CC, Rey-Dios R, Kanakamedala MR, Fatemi A. Evaluating the accuracy of geometrical distortion correction of magnetic resonance images for use in intracranial brain tumor radiotherapy. *Rep Pract Oncol Radiother*. (2019) 24:606–13. doi: 10.1016/j.rpor.2019.09.011

26. Retif P, Djibo Sidikou A, Mathis C, Letellier R, Verrecchia-Ramos E, Dupres R, et al. Evaluation of the ability of the Brainlab Elements Cranial Distortion Correction algorithm to correct clinically relevant MRI distortions for cranial SRT. *Strahlenther Onkol*. (2022) 198:907–18. doi: 10.1007/s00066-022-01988-1

27. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. (2004) 13:600–12. doi: 10.1109/TIP.2003.819861

28. Mason A, Rioux J, Clarke SE, Costa A, Schmidt M, Keough V, et al. Comparison of objective image quality metrics to expert radiologists' Scoring of diagnostic quality of MR images. *IEEE Trans Med Imaging*. (2020) 39:1064–72. doi: 10.1109/TMI.2019.2930338

29. Tsien C, Drzymala RE, Rich K. Imaging advances in stereotactic radiosurgery. *Mo Med*. (2015) 112:373–8.

30. Pappas EP, Alshanqity M, Moutsatsos A, Lababidi H, Alsafi K, Georgiou K, et al. MRI-related geometric distortions in stereotactic radiotherapy treatment planning: evaluation and dosimetric impact. *Technol Cancer Res Treat*. (2017) 16:1120–9. doi: 10.1177/1533034617735454

31. Agazaryan N, Tenn S, Lee C, Steinberg M, Hegde J, Chin R, et al. Simultaneous radiosurgery for multiple brain metastases: technical overview of the UCLA experience. *Radiat Oncol*. (2021) 16:221. doi: 10.1186/s13014-021-01944-w

32. Gevaert T, Steenbeke F, Pellegri L, Engels B, Christian N, Hoornaert MT, et al. Evaluation of a dedicated brain metastases treatment planning optimization for radiosurgery: a new treatment paradigm? *Radiat Oncol*. (2016) 11:13. doi: 10.1186/s13014-016-0593-y

33. Winey B, Bussiere M. Geometric and dosimetric uncertainties in intracranial stereotactic treatments for multiple nonisocentric lesions. *J Appl Clin Med Phys*. (2014) 15:122–32. doi: 10.1120/jacmp.v15i3.4668

34. Kim HY, Lee SI, Jin SJ, Jin SC, Kim JS, Jeon KD. Reliability of stereotactic coordinates of 1.5-tesla and 3-tesla MRI in radiosurgery and functional neurosurgery. *J Korean Neurosurg Soc*. (2014) 55:136–41. doi: 10.3340/jkns.2014.55.3.136

35. Duchin Y, Abosch A, Yacoub E, Sapiro G, Harel N. Feasibility of using ultra-high field (7 T) MRI for clinical surgical targeting. *PloS One*. (2012) 7:e37328. doi: 10.1371/journal.pone.0037328

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact