RESEARCH ARTICLE

Differential diagnosis of low- and high-grade gliomas using radiomics and deep learning fusion signatures based on multiple magnetic resonance imaging sequences

Jian Pan*

School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China.

Received: February 21, 2024; **accepted:** April 19, 2024.

Glioma, the most prevalent type of brain tumor, is categorized into low-grade gliomas (LGGs) and high-grade gliomas (HGGs). Accurate differentiation between these categories is crucial for selecting the appropriate therapeutic approach. Magnetic resonance imaging (MRI) technology offers a noninvasive and rapid diagnostic method for visual interpretation by clinicians. Machine learning plays a pivotal role in intelligently analyzing multiple MRI images, enhancing diagnostic efficiency, and providing an objective diagnosis. Radiomics and deep learning (DL) techniques contribute distinctively to glioma grading by extracting pertinent features from various MRI images. This study aimed to assess the efficacy of fusing radiomic and DL features from multiple MRI images to distinguish between LGGs and HGGs. Image data from four MRI scanning sequences including T1-weighted, postcontrast T1-weighted, T2-weighted, and T2 fluid-attenuated inversion recovery images were collected from various institutions with different instrumental magnetic field strengths. The radiomic features from each MRI sequence were used to construct respective single-sequence radiomic models, while the slice images from each MRI sequence were used to construct corresponding single-sequence DL models. A radiomic fusion model was created by amalgamating signatures from the four single-sequence radiomic models, and a DL fusion model was constructed by integrating information from all four MRI sequences. The radiomic and DL (RDL) model was constructed using two fusion signatures generated by the radiomic and DL fusion models, respectively. The RDL model demonstrated excellent diagnostic performance, achieving an area under the curve (AUC) of 0.959 on a testing cohort with the radiomic fusion signature proving pivotal in differentiating LGGs from HGGs. The RDL model exhibited superior generalization performance compared to the radiomic or DL models constructed solely based on MRI sequences. The results confirmed the feasibility and effectiveness of combining radiomic and DL features through fusion signatures, thereby highlighting the potential of integrating radiomic and DL fusion signatures to enhance preoperative glioma grading performance.

Keywords: gliomas; radiomics; deep learning; MRI; glioma grading.

***Corresponding author:** Jian Pan, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China. Email: ppw020306@outlook.com.

Introduction

Gliomas are a prevalent form of brain cancer [1, 2], accounting for approximately 80% of all malignant brain tumors [3, 4], and can be categorized as low-grade gliomas (LGGs) and high-grade gliomas (HGGs) [5-8]. In clinical practice, different treatment approaches are used for LGGs and HGGs. HGGs require surgery combined with radiotherapy or chemotherapy,

whereas LGGs necessitate post-surgery monitoring without immediate intervention [9]. Thus, accurate differentiation between LGGs and HGGs is vital for selecting a precise therapeutic approach. Pathological examination serves as the gold standard for diagnosis [10]. However, it is an invasive procedure that requires a biopsy sample and carries associated risks. By contrast, magnetic resonance imaging (MRI) offers a noninvasive and rapid diagnostic method that plays a crucial role in the diagnosis of gliomas [7, 11, 12]. MRI examination includes multiple scanning sequences, among which T1-weighted (T1), postcontrast T1-weighted (T1C), T2-weighted (T2), and T2 fluid-attenuated inversion recovery (FLAIR) images have been widely used in brain tumor classification [13]. Different MRI sequences can offer various valuable insights for clinicians, each playing significant roles in glioma grading. However, in clinical practice, glioma grading based on MRI relies on visual interpretation by clinicians. Visual interpretation, a traditional diagnostic method, is subjective and labor intensive [14]. Fortunately, machine learning (ML) has been increasingly employed in the intelligent analysis of medical images to improve diagnosis [15].

Radiomics and deep learning (DL) are two major branches of ML [16]. In contrast to visual interpretation, radiomics methods possess the capability to extract multiple disease-related features from a defined region of interest (ROI) [17, 18]. Previous studies have employed radiomic features based on multiple MRI sequences for glioma grading [19-21]. A recent study underscored that DL models outperformed radiomic models [22]. DL, a relatively recent ML methodology, can automatically extract intricate and abstract features from entire medical images. Currently, DL models based on multiple MRI sequences have been employed for glioma grading [23, 24]. Radiomics and DL, utilizing distinct feature extraction techniques, offer diverse sets of information that can be valuable in glioma grading. Consequently, integrating DL and radiomics based on multiple MRI sequences holds the potential to enhance the accuracy of distinguishing between LGGs and HGGs. Previous studies have demonstrated that combining radiomics and DL features is a promising approach in cancer studies [25-27]. Zhang *et al*. employed the combination of radiomics and DL features to classify benign and malignant breast lesions [28]. Tian *et al*. employed the combination of radiomics and DL features to predict the PD-L1 expression level in preoperative MRI of hepatocellular carcinoma (HCC) patients [29]. These studies demonstrated that the model constructed using a combination of radiomics and DL features outperformed models constructed using either radiomics or DL features. However, few studies have employed the combination of radiomics and DL features extracted from multiple MRI sequences in glioma grading, particularly in discriminating between LGGs and HGGs. To explore the feasibility of combining radiomics and DL features from multiple MRI sequences in glioma grading, a pioneering study combined radiomic features and DL features extracted from multiple MRI sequences to construct a classifier for predicting glioma grade [30]. Compared with models constructed solely using radiomics or deep features, the model constructed using the combination of radiomic and DL features demonstrated superior performance. However, that study did not employ the T1WI and T2WI sequences, which were also important in glioma grading. Additionally, the merging of radiomics and DL features may lead to an excessive number of features, potentially impacting the performance of the constructed model. Thus, further validation is needed to assess the performance of the combination of DL and radiomics features based on multiple MRI sequences.

This study evaluated the effectiveness of a combination of radiomics and DL utilizing fusion signatures extracted from multiple MRI sequences in distinguishing LGGs and HGGs. Five radiomic and DL models were constructed, and a radiomic and DL (RDL) model was developed for discriminating between LGGs and HGGs,

leveraging radiomic and DL fusion signatures derived from multimodal MR images. In addition, a multimodal attention module, named the channel-spital-global (CSG) attention module, was developed through this study.

Materials and methods

Data resource

The patients' image data and corresponding diagnoses were obtained from the Cancer Imaging Archive (TCIA) with the permission of TCIA (https://www.cancerimagingarchive.net) [31]. All data were from BraTS-TCGA-LGG and BraTS-TCGA-GBM projects [32, 33], including 243 patients with 76 and 167 patients in the testing and training datasets, respectively. The training dataset was further divided into training (n = 133) and validation (n = 34) cohorts. The patients' clinical data from BraTS-TCGA-LGG project can be accessed online at https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=5309188#530918 8c26029d5d06d47298421bc97a34ba253, and from BraTS-TCGA-GBM project at the website of https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966258#19662585384a856aefe4daf9184d99f46990e0b. The demographic characteristics of the training and testing datasets were shown in Table 1.

Image description and preprocessing

For each patient, T1, T1C, T2, and FLAIR images were obtained. Sequences were scanned before surgery. MRI images utilized in this study were collected from multiple institutions with varying magnetic field strengths. The image data were obtained online from the websites of https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=24282668 for BraTS-TCGA-LGG project and https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=24282666 for BraTS-TCGA-GBM project. According to the description on the TCIA website, these images underwent skull stripping, interpolation at the same resolution, and co-registration to the same anatomical template. The preprocessed images were then segmented. The segmentation data was obtained from TCIA with three segmentation labels as necrotic region, peritumoral region, and enhancing tumor. All MRI images were normalized using the following equation.

$$x_{normalized} = \frac{x_{raw} - x_{min}}{x_{max} - x_{min}},$$

where x_{raw} and $x_{normalized}$ represented the original and normalized intensity of each voxel, respectively. x_{min} and x_{max} were the minimum and maximum intensities, respectively, of all voxels in the corresponding MRI image.

Radiomics-based predictive model construction

1. Radiomic feature extraction

In this study, three regions including necrosis, the peritumoral region, and the enhancing tumor region were collectively considered as a whole and utilized as ROIs [34, 35]. The open-source PyRadiomics package (https://pyradiomics.readthedocs.io) was used to extract radiomic features from the ROIs [36]. Prior to feature extraction, the MRI images were preprocessed using eight filters including square root, gradient, Laplacian of Gaussian with four sigma levels (2.0, 3.0, 4.0, and 5.0), 3D local binary pattern, wavelet, logarithm, exponential, and square (Figure 1A). The preprocessed and original images were used to extract textural, shape, and first-order features, resulting in 1,967 features from each MRI sequence. Detailed information on these radiomic features can be found in the PyRadiomics documentation (https://pyradiomics.readthedocs.io/en/latest/features.html).

2. Single-sequence radiomics model construction

The process of constructing single-sequence radiomic models by utilizing corresponding MRI

Table 1. Patient demographics in the training and testing sets.

	Training dataset			Testing cohort* (*n* = 76)
	Overall (*n* = 167)	Training cohort* (*n* = 133)	Validation cohort (*n* = 34)	
Age (year, mean ± SD)	52.36±15.46	53.19±15.11	49.15±16.62	53.65±14.87
Age range	18~84	20~84	18~74	17~80
Gender				
Male	90	68	22	43
Female	76	64	12	32
Race				
White	147	118	29	64
Asian	6	3	3	1
African American	9	9	0	7
Not available	5	3	2	4
Grade				
LGG	65	49	16	43
HGG	102	84	18	33

Note: * The age and gender of one patient were unknown.

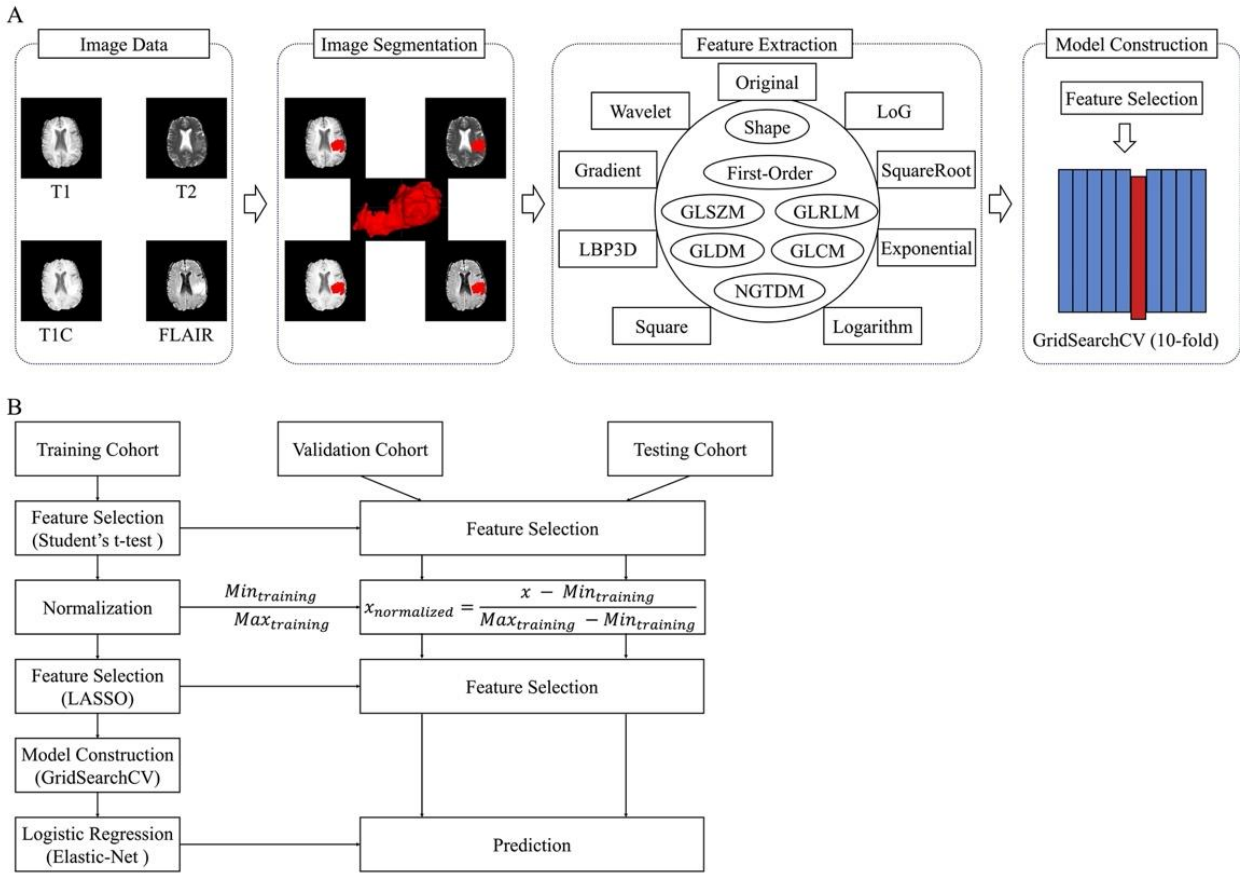

Figure 1. Single-sequence radiomic model construction. Radiomic features were extracted and employed to construct a single-sequence radiomic model using GridSearchCV with ten folds. **A.** Framework of single-sequence radiomic model construction. **B.** Training, validation, and testing of the single-sequence radiomic models. LASSO: least absolute shrinkage and selection operator. LoG: Laplacian of Gaussian. LBP3D: 3D local binary pattern.

radiomic features was shown in Figure 1B. In this process, a classifier in the form of a logistic regression (LR) model with elastic-net regularization was employed to generate a single-sequence radiomic signature. To enhance the model's generalizability, Student's *t*-test was applied for feature analysis [37]. Subsequently, the feature values in the training cohort were normalized to a range of 0 to 1 based on the minimum and maximum values of features in the training cohort. The same normalization was applied to the corresponding features in the validation and testing cohorts. The least absolute shrinkage and selection operator (LASSO) was employed to further reduce the dimensionality of the radiomic features [37]. The imbalance in the number of patients with LGGs and HGGs necessitated the use of class weights to mitigate this limitation in LR construction. The parameters of each LR were optimized during the training process using the GridSearchCV tool in Scikit-learn (https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html#sklearn.model_selection.GridSearchCV). For LR with elastic-net regularization, the λ and *l*_1 ratios were selected from 21 values [2^{-10}, 2^{-9}, …, 2^{9}, 2^{10}] and 10 values [0.1, 0.2, …, 0.9, 1], respectively. Finally, LR models with their corresponding optimal hyperparameters were constructed using the training cohort. The constructed models were validated using radiomic features of the corresponding MRI sequence in the validation cohort and tested using radiomic features of the corresponding MRI sequence in the testing cohort. The prediction probability was defined as a single-sequence radiomic signature.

3. Radiomic fusion model construction

Figure 2 provided an overview of the workflow for constructing the radiomic fusion model with LR with L2 regularization used to generate a radiomic fusion signature. The process involved combining the signatures of the four single-sequence models from the training cohort and using them as input to train an LR model with L2 regularization. This step resulted in the creation of the radiomic fusion model. Before constructing the LR models, the signatures were normalized using a *z*-score transformation. During the training process, the LR parameters were optimized using GridSearchCV. To address the imbalance in the number of patients between LGGs and HGGs, class weights were applied. In the case of LR with L2 regularization, the parameter λ was selected from 21 values [2^{-10}, 2^{-9}, …, 2^{9}, 2^{10}]. The final LR model with optimal hyperparameters was constructed using signatures from the four single-sequence radiomic models from the training cohort. The LR model was evaluated using signatures from both the validation and testing cohorts. The constructed radiomic fusion model was validated using the four single sequence radiomic signatures of the validation cohort and tested using the four single-sequence radiomic signatures of the testing cohort. Furthermore, the weighted values associated with the corresponding single radiomic signatures in the radiomic fusion model denoted the contributions of various parameters in distinguishing between LGGs and HGGs. The prediction probability was defined as the radiomic fusion signature.

Building the deep learning model

Due to the complexity of 3D convolutional neural networks and the limited size of the patient dataset in this study, a different approach was adopted by utilizing slice images to create a 2D DenseNet121 model. Specifically, slice images with a cubic ROI were cropped to exclude non-tumor areas. Each MRI sequence resulted in 9,000, 2,265, and 4,931 slice images in the training cohort, validation cohort, and testing cohort, respectively. Four individual DL models were established using the four single MRI sequences. Additionally, a multimodal DL fusion model was developed by combining all four MRI sequences to generate a DL fusion signature. The slice images from the training cohort were employed for constructing these five DL models. The slice images of the validation cohort were fed into the constructed model to evaluate the performance of the corresponding constructed

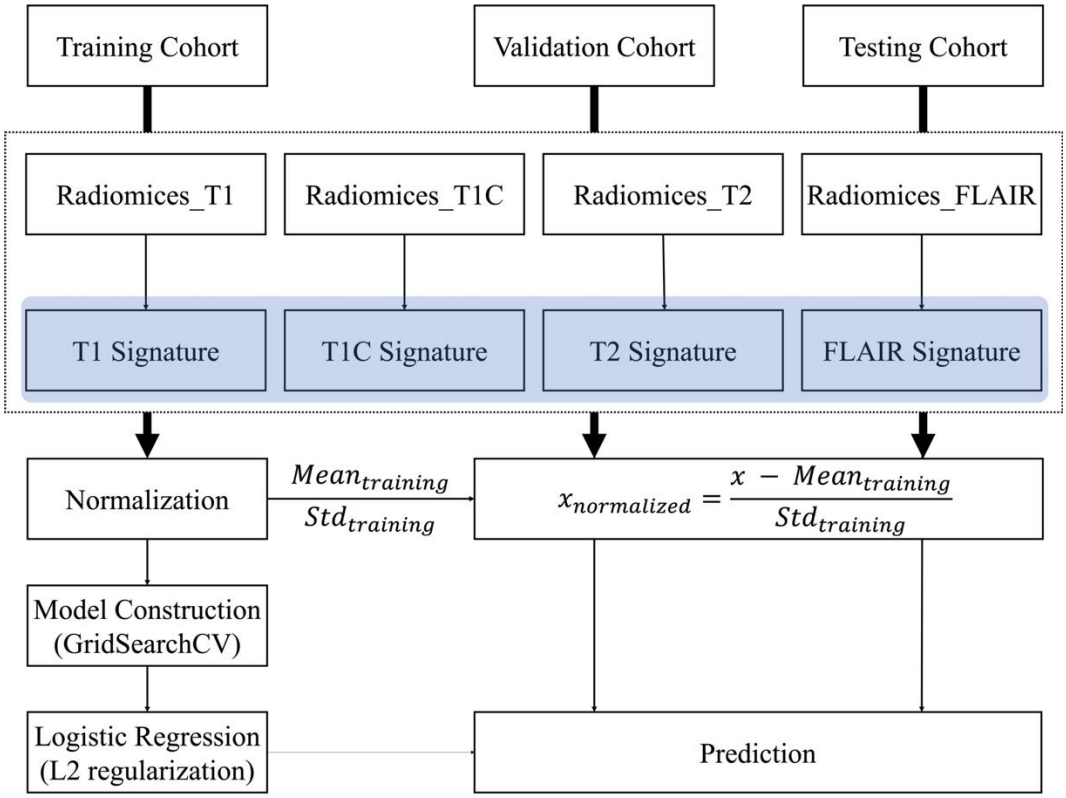

Figure 2. Training, validation, and testing of the radiomic fusion model. Four single-sequence signatures were used as features to construct the radiomic fusion model. Std: Standard deviation.

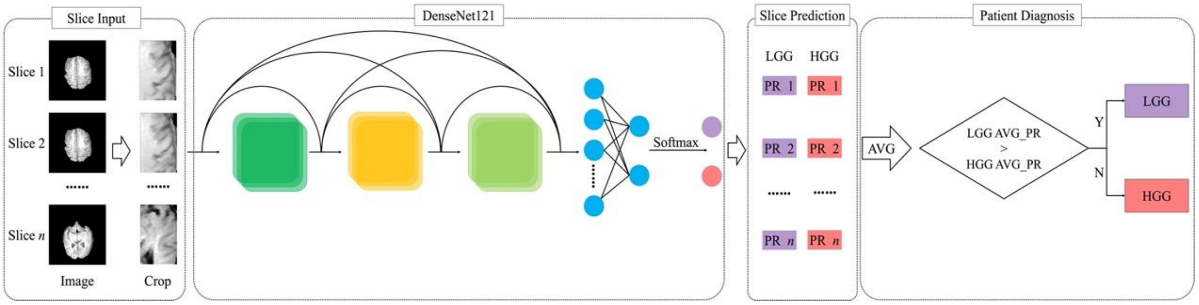

Figure 3. Framework of single-sequence DL model construction, based on slice images. Patient-level predictions were made using the average PRs at the slice level. PR: probability.

model and select the corresponding best model. Finally, the slice images from the testing cohort were utilized to evaluate the performance of the selected DL model.

1. Single-sequence deep learning model construction

A DenseNet121 model was constructed using the corresponding slice images for each MRI sequence (Figure 3) [38]. To accommodate the varying dimensions of the slice images, the convolution layer of each DenseNet121 model was configured to accept a single input channel. Given the inconsistency in the dimensions of the slice images, they were resized to a uniform size of 64 × 64 pixels. Because the task involved binary classification to distinguish between LGGs and HGGs, the output dimensions of the fully

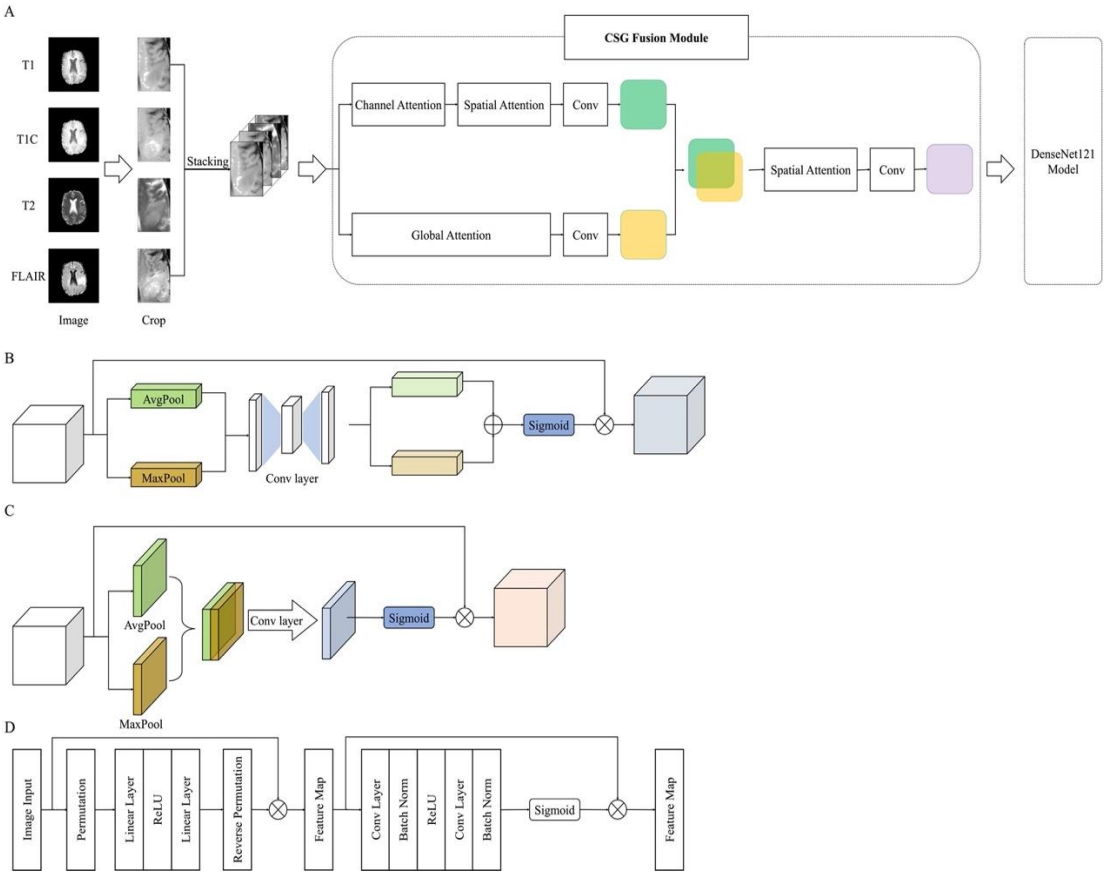

Figure 4. Structure of CSG-DenseNet121. **A.** CSG fusion module was included in CSG-DenseNet121. The CSG fusion module consists of a channel attention module (**B**), a spatial attention module (**C**), and a global attention module (**D**).

connected layers were set to two. A patient-level diagnostic strategy was employed. For each patient, the prediction probabilities (PRs) of LGGs and HGGs for all slice images were generated using the DenseNet121 model. The average prediction PRs of LGGs (LGG AVG_PR) and HGGs (HGG AVG_PR) across all slice images of a patient were calculated, which indicated the prediction PRs of LGGs and HGGs at the patient level, respectively. Therefore, if the LGG AVG_PR was larger than the HGG AVG_PR, the patient was identified as having LGG, otherwise, the patient was identified as having HGG.

2. Multimodal deep learning fusion model

Four MRI sequences were used to construct a multimodal DL fusion model (Figure 4). Slice images of the four MRI sequences were stacked before being fed into the multimodal DL fusion model. Within this model, a novel fusion module (Figure 4A) was devised, incorporating channel, spatial, and global attention modules. This fusion module, known as the CSG fusion module, was responsible for amalgamating information from multimodal MRI images. The structures of the channel attention, spatial attention, and global attention modules were shown in Figures 4B, 4C, and 4D, respectively [39, 40]. In the CSG fusion module, a feature map with four channels was generated by the channel and spatial attention modules. The map was then compressed into a feature map with one channel using a convolutional layer. In addition, another feature map with one channel was generated using the global attention module and a convolutional layer. The two feature maps were stacked and processed using a spatial attention module [39]. Finally, a feature map with one channel was

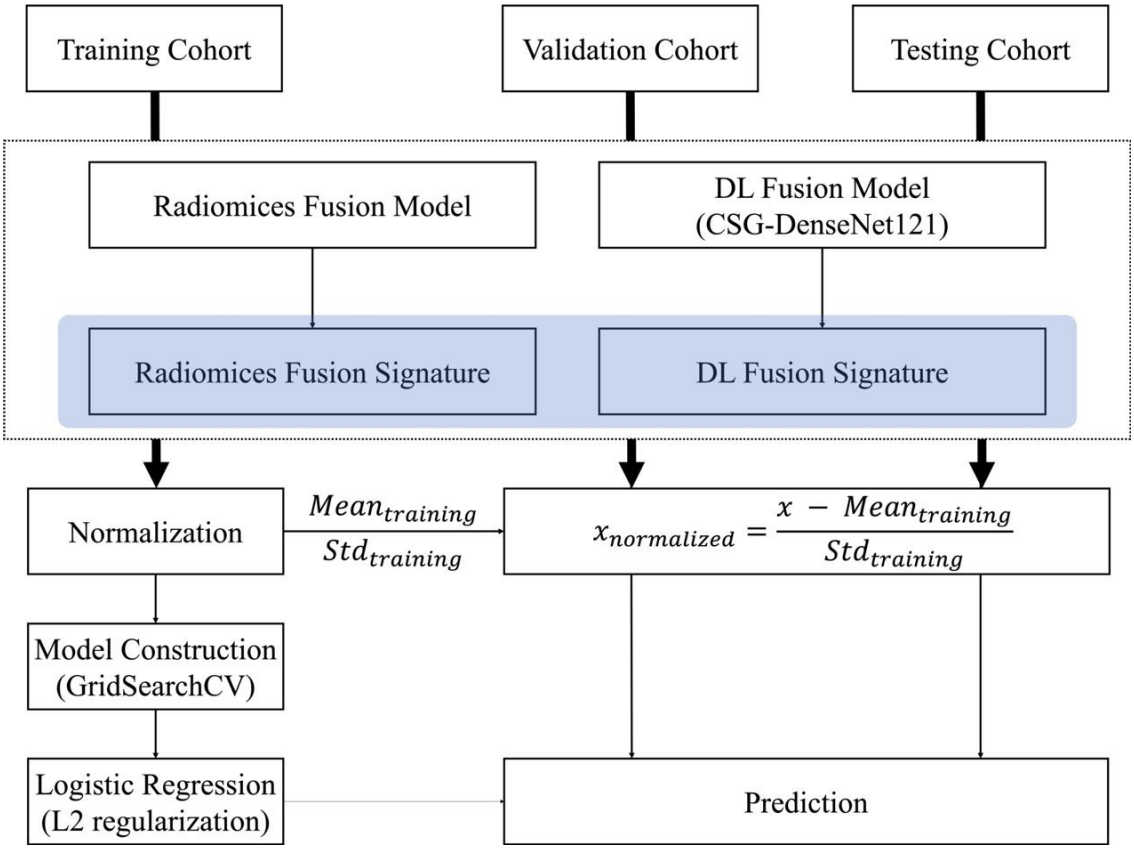

Figure 5. Workflow of the RDL model construction. The radiomics and DL fusion signatures generated from four MRI sequences were used to construct the RDL model.

generated and fed into DenseNet121. A patient-level diagnostic strategy was also employed in the DL fusion model. The DL fusion signature was defined as the prediction probability of a patient. Five DL models were constructed using PyTorch (version 1.8.1) (https://pytorch.org) on a Windows computer. The number of epochs, batch size, and initial learning rate were set to 100, 32, and 0.0001, respectively. After 20 iterations, the learning rate was multiplied by 0.1. A stochastic gradient descent optimization strategy was used. In addition, the gradient-weighted class activation mapping (Grad-CAM) (https://arxiv.org/abs/1610.02391) was employed for visual investigation of the CSG-DenseNet121 based on slice images [41].

RDL model construction

The construction workflow of the RDL model using LR with L2 regularization was illustrated in Figure 5. The fusion signatures of the radiomic and CSG-DenseNet121 models from the training cohort were stacked to train an LR model with L2 regularization to construct the RDL model. Before constructing the LRs, the signatures were normalized using a z-score transformation. The LR parameters were optimized during the training process using GridSearchCV. The class weight was used in the training process to overcome the patient number imbalance between LGGs and HGGs. The regularization parameter λ was selected from 21 values [2^{-10}, 2^{-9}..., 2^{9}, 2^{10}]. The final LR with optimal hyperparameters was constructed using the fusion signatures from the training cohort. The constructed RDL model was validated using the radiomic and DL fusion signatures of the validation cohort and tested using the radiomic and DL fusion signatures of the testing cohort.

Table 2. Performance of the radiomic, DL, and RDL models in the validation cohort.

Image data	AUC	*P* value	Accuracy	Sensitivity	Specificity
Radiomics_T1	0.670	<0.05*	61.76%	61.11%	56.25%
Radiomics_T1C	0.924	0.080	88.24%	88.89%	81.25%
Radiomics_T2	0.851	<0.05*	76.47%	77.78%	62.50%
Radiomics_FLAIR	0.830	<0.05*	67.65%	61.11%	75.00%
Radiomics_Fusion	0.972	0.245	94.12%	94.44%	93.75%
DL_T1	0.854	<0.05*	76.47%	72.22%	81.25%
DL_T1C	0.979	0.251	91.18%	88.89%	93.75%
DL_T2	0.858	<0.05*	79.41%	88.89%	68.75%
DL_FLAIR	0.771	<0.05*	70.59%	61.11%	81.25%
DL_Fusion	0.986	0.362	94.12%	94.44%	93.75%
RDL	1.000	Reference	97.06%	100.00%	87.50%

Note: * indicated significant difference.

Table 3. Performance of the radiomic, DL, and RDL models on the testing cohort.

Image data	AUC	*P* value	Accuracy	Sensitivity	Specificity
Radiomics_T1	0.880	0.069	76.32%	90.91%	65.12%
Radiomics_T1C	0.933	0.346	85.53%	87.88%	83.72%
Radiomics_T2	0.708	<0.05*	65.79%	75.76%	58.14%
Radiomics_FLAIR	0.727	<0.05*	72.37%	69.70%	74.42%
Radiomics_Fusion	0.921	<0.05*	88.16%	90.91%	86.05%
DL_T1	0.808	<0.05*	71.05%	69.70%	72.09%
DL_T1C	0.899	0.057	77.63%	69.70%	83.72%
DL_T2	0.791	<0.05*	69.74%	69.70%	69.77%
DL_FLAIR	0.681	<0.05*	63.16%	57.58%	67.44%
DL_Fusion	0.918	0.107	82.89%	90.91%	76.74%
RDL	0.959	Reference	90.79%	90.91%	90.70%

Note: * indicated significant difference.

Results

Patient characteristics

There were no significant differences in age, gender, and race between the training dataset and testing cohorts or between the training and validation cohorts. However, the numbers of LGGs and HGGs differed significantly between the training dataset and testing cohort ($P < 0.05$). Conversely, there was no noticeable disparity in the prevalence of LGGs or HGGs between the training and validation cohorts.

Classification performance of radiomics, DL, and RDL models in the validation cohort

The performance of the radiomic, DL, and RDL models for glioma grading on the validation cohort were shown in Table 2. The results clearly indicated that the RDL model achieved the highest accuracy and AUC among all models. Specifically, the RDL model attained an AUC of 1, an accuracy of 97.06%, a sensitivity of 100%, and a specificity of 87.50%. Among the five radiomic models, the radiomic fusion model exhibited superior performance compared to the four single-sequence models. Likewise, within the five DL models, the DL fusion model (CSG-DenseNet121) outperformed the other four single-sequence models.

Classification performance of the radiomics, DL, and RDL models on the testing cohort

The performance of the radiomic, DL, and RDL models in grading gliomas within the testing

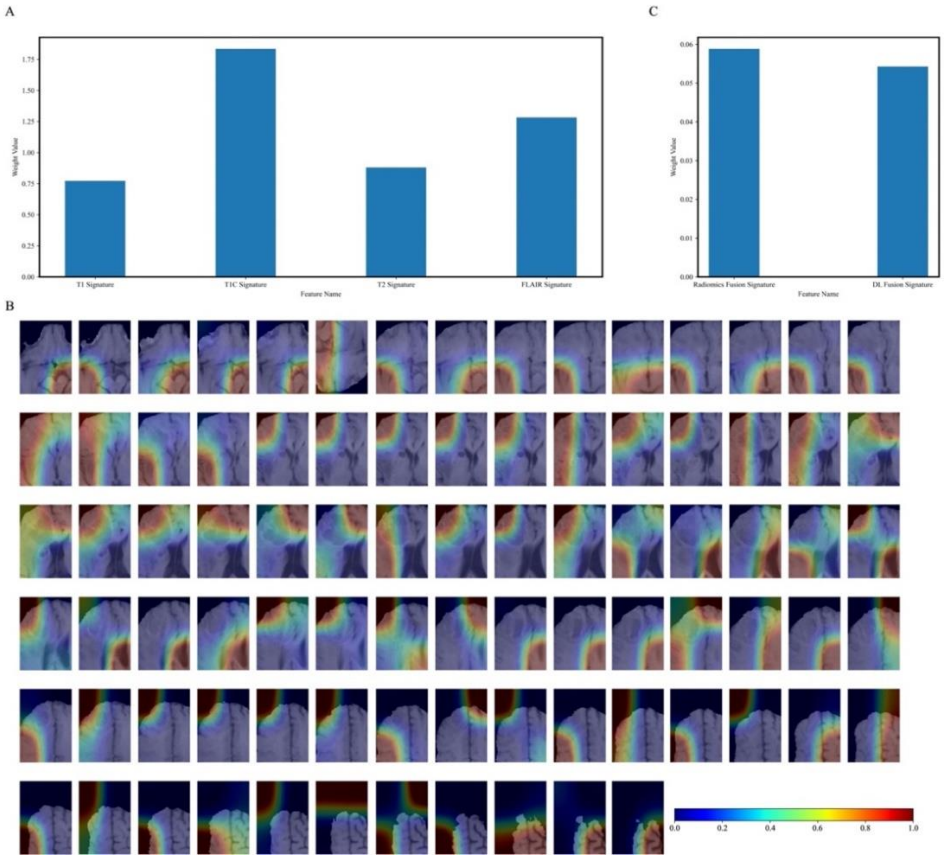

Figure 6. Important features for radiomic fusion, DL fusion, and RDL models. **A.** Weight values corresponding to single radiomic signatures in the radiomic fusion models. **B.** Example of a heatmap of slice images for a patient. **C.** Weight values corresponding to the fusion signatures in the RDL.

cohort was shown in Table 3. The results demonstrated that the RDL model achieved the highest accuracy and AUC compared with the other models. Specifically, the RDL model attained an AUC of 0.959, an accuracy of 90.79%, sensitivity of 90.91%, and specificity of 90.70%. Among the five radiomic models, the radiomic fusion model outperformed the four single-sequence models. Additionally, the DL fusion model (CSG-DenseNet121) surpassed the other four single-sequence DL models.

Important features of the radiomic fusion, DL fusion, and RDL models

For the radiomic fusion model, the weight value associated with the T1C signature was the highest (Figure 6A), which indicated that T1C played a pivotal role in differentiating between LGGs and HGGs among all the evaluated signatures. Figure 6B provided insights into the DL fusion model, where Grad-CAM was employed based on the slice images. The Grad-CAM heatmap demonstrated the DL fusion model's ability to detect tumors. In the case of the RDL model, the weight value corresponding to the radiomic fusion signature was slightly greater than that of the DL fusion signature (Figure 6C). Therefore, the radiomic fusion signature made the most substantial contribution to the discrimination between LGGs and HGGs among all the evaluated signatures.

Discussion

An RDL method to enhance the discrimination between LGGs and HGGs was developed in this study. A radiomic fusion model and a DL fusion

model incorporating a CSG fusion module were constructed before radiomic and DL fusion signatures from these models were generated. An RDL model, combining the radiomic and DL fusion signatures, was eventually established. Notably, the RDL model demonstrated remarkable performance with AUCs of 1 in the validation cohort and 0.959 on the testing cohort. Furthermore, the radiomic and DL fusion model outperformed their respective single-sequence models.

A radiomic fusion signature by combining four individual radiomic signatures was generated in this study. The radiomic signature was created through LR, a multivariate analysis method that considered all features as a cohesive entity [42], to discriminate LGGs and HGGs based on the distinctive patterns among these features. Consequently, the original radiomic features were amalgamated into a comprehensive radiomic signature, effectively reducing feature dimensions for the RDL model while preserving the essence of the original features. The proposed method of this study excelled in performance compared to the four single-sequence radiomic models. The radiomic fusion model harnessed information from the four MRI sequences *via* radiomic signatures with minimal feature dimensions. However, note that the original radiomic features were extracted from the ROI based on empirical knowledge, limiting the information to the interior of tumors delineated using this method. By contrast, DL could extract image features through convolution operations, encompassing the entire input image without relying on specific medical expertise. Consequently, the DL fusion signature encapsulated global information. In this research, an RDL model by combining radiomic and DL fusion signatures was constructed, enabling the model to access both local and global information related to gliomas, which significantly improved discrimination between LGGs and HGGs. Comparatively, Ning *et al*. merged radiomics and DL features to assess glioma grading based on T1C and FLAIR images [30]. By contrast, this study employed radiomic and DL signatures to construct a model for discriminating between LGGs and HGGs. Notably, this study achieved better performance than the study by Ning *et al*. [30]. Two potential reasons for these performance differences could be the incorporation of additional MRI sequences and variations in feature fusion methodologies.

The DL fusion signature was derived from a specific DL fusion model known as CSG-DenseNet121, which exhibited the highest performance in discriminating between LGGs and HGGs. Within the CSG-DenseNet121 model, a CSG fusion module was employed before applying DenseNet121, which was designed to effectively integrate diverse information provided by different MRI sequences. The CSG fusion module operated from two distinct perspectives to extract valuable features. The first perspective utilized channel and spatial attention modules, enabling the extraction of features that were not only important in individual channels but also in their spatial context. The channel attention module played a vital role in selecting crucial MRI sequences, while the spatial attention module directed the classifier's focus toward tumor-related regions. Consequently, this combination of channel and spatial attention empowered the classifier to prioritize tumor regions within critical MRI sequences without disregarding information from other imaging modalities, ultimately enhancing prediction accuracy. Additionally, a global attention module was incorporated to generate a feature map from another standpoint, which reduced information redundancy while amplifying global dimension-interactive features [40]. By combining the feature maps generated from these two perspectives, a fusion feature map was created to encapsulate comprehensive information from the four original MRI sequences, thereby improving the accuracy of distinguishing between LGGs and HGGs. However, according to the Grad-CAM findings, although the DL fusion model could identify tumors, it was not perfectly focused on them. This observation might explain

the relatively lower weight assigned to the DL fusion signature in the RDL model, which was constructed by incorporating both radiomic and DL fusion signatures. The present study employed slice images to construct DL models, following similar methods as previous studies [43], thereby increasing the number of samples available for constructing the DL models. Compared with 3D DenseNet121, 2D DenseNet121 had fewer parameters. The number of patients included in this study was small, making it challenging to train the 3D DenseNet121 model using 3D MRI images. Therefore, 2D DenseNet121 models were constructed using image slices to discriminate between LGGs and HGGs. In addition, we found that the T1C-based single-sequence radiomic and DL models performed better than the T1-, T2-, and FLAIR-based single-sequence models. Further, in the radiomic fusion model, the weight of the T1C signature was greater than that of the other MRI signatures. These weight values indicated the contribution of the corresponding signature when discriminating between LGGs and HGGs. The results were consistent with the results of previous studies [44], which might be attributed to the breakdown of the blood–brain barrier. HGGs often disrupt the blood–brain barrier, whereas LGGs do not exhibit breakdown of the blood–brain barrier, which can be detected using T1C images. Therefore, the T1C level could be a potential diagnostic tool for glioma grading.

The current study was subject to two noteworthy limitations. First, the patient sample size in this study was relatively small. To assess the performance of the RDL model more comprehensively, future investigations should encompass a larger and more diverse patient cohort drawn from multiple healthcare institutions. Second, cropped slice images were used in DL models and the cropping process was reliant on tumor segmentation. In future research, it is advisable to employ entire images rather than cropped ones when constructing DL models. This approach would eliminate the dependence on tumor segmentation and could potentially enhance the model's performance and applicability.

References

1. Li T, Li J, Chen Z, Zhang S, Li S, Wageh S, *et al*. 2022. Glioma diagnosis and therapy: Current challenges and nanomaterial-based solutions. J Control Release. 352:338-370.
2. Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, *et al*. 2014. The epidemiology of glioma in adults: A "state of the science" review. Neuro Oncol. 16(7):896-913.
3. Goodenberger ML, Jenkins RB. 2012. Genetics of adult glioma. Cancer Genet. 205(12):613-621.
4. Alhasan AS. 2021. Clinical applications of artificial intelligence, machine learning, and deep learning in the imaging of gliomas: A systematic review. Cureus. 13(11):e19580.
5. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, *et al*. 2021. The 2021 WHO classification of tumors of the central nervous system: A summary. Neuro Oncol. 23(8):1231-1251.
6. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, *et al*. 2016. The 2016 World Health Organization classification of tumors of the central nervous system: A summary. Acta Neuropathol. 131(6):803-820.
7. Luo J, Pan M, Mo K, Mao Y, Zou D. 2023. Emerging role of artificial intelligence in diagnosis, classification and clinical management of glioma. Semin Cancer Biol. 91:110-123.
8. Hummel TR, Chow LM, Fouladi M, Franz D. 2013. Pharmacotherapeutic management of pediatric gliomas: Current and upcoming strategies. Paediatr Drugs. 15(1):29-42.
9. Diwanji TP, Engelman A, Snider JW, Mohindra P. 2017. Epidemiology, diagnosis, and optimal management of glioma in adolescents and young adults. Adolesc Health Med Ther. 8:99-113.
10. Abdel Razek AAK, Alksas A, Shehata M, AbdelKhalek A, Abdel Baky K, El-Baz A, *et al*. 2021. Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging. Insights Imaging. 12(1):152.
11. Villanueva-Meyer JE, Mabray MC, Cha S. 2017. Current clinical brain tumor imaging. Neurosurgery. 81(3):397-415.
12. Hooper GW, Ansari S, Johnson JM, Ginat DT. 2023. Advances in the radiological evaluation of and theranostics for glioblastoma. Cancers (Basel). 15(16):4162.
13. Xie Y, Zaccagna F, Rundo L, Testa C, Agati R, Lodi R, *et al*. 2022. Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): Review, challenges, and future perspectives. Diagnostics (Basel). 12(8):1850.
14. Le WT, Maleki F, Romero FP, Forghani R, Kadoury S. 2020. Overview of machine learning: Part 2: Deep learning for medical image analysis. Neuroimaging Clin N Am. 30(4):417-431.
15. Currie G, Hawk KE, Rohren E, Vial A, Klein R. 2019. Machine learning and deep learning in medical imaging: Intelligent imaging. J Med Imaging Radiat Sci. 50(4):477-487.

16. Park HJ, Park B, Lee SS. 2020. Radiomics and deep learning: Hepatic applications. Korean J Radiol. 21(4):387-401.
17. Gillies RJ, Kinahan PE, Hricak H. 2016. Radiomics: Images are more than pictures, they are data. Radiology. 278(2):563-577.
18. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, *et al*. 2014. Decoding tumor phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 5:4006.
19. Qin JB, Liu Z, Zhang H, Shen C, Wang XC, Tan Y, *et al*. 2017. Grading of gliomas by using radiomic features on multiple magnetic resonance imaging (MRI) sequences. Med Sci Monit. 23:2168-2178.
20. Ubaldi L, Saponaro S, Giuliano A, Talamonti C, Retico A. 2023. Deriving quantitative information from multiparametric MRI *via* radiomics: Evaluation of the robustness and predictive value of radiomic features in the discrimination of low-grade versus high-grade gliomas with machine learning. Phys Med. 107:102538.
21. Hwan-Ho C, Hyunjin P. 2017. Classification of low-grade and high-grade glioma using multi-modal image radiomics features. Annu Int Conf IEEE Eng Med Biol Soc. 2017:3081-3084.
22. Gutta S, Acharya J, Shiroishi MS, Hwang D, Nayak KS. 2021. Improved glioma grading using deep convolutional neural networks. Am J Neuroradiol. 42(2):233-239.
23. Ge C, Gu IY, Jakola AS, Yang J. 2018. Deep learning and multi-sensor fusion for glioma classification using multistream 2D convolutional networks. Annu Int Conf IEEE Eng Med Biol Soc. 2018:5894-5897.
24. Decuyper M, Bonte S, Deblaere K, Van Holen R. 2021. Automated MRI based pipeline for segmentation and prediction of grade, IDH mutation and 1p19q co-deletion in glioma. Comput Med Imaging Graph. 88:101831.
25. Gao W, Wang W, Song D, Yang C, Zhu K, Zeng M, *et al*. 2022. A predictive model integrating deep and radiomics features based on gadobenate dimeglumine-enhanced MRI for postoperative early recurrence of hepatocellular carcinoma. Radiol Med. 127(3):259-271.
26. Chang R, Qi S, Wu Y, Yue Y, Zhang X, Qian W. 2023. Nomograms integrating CT radiomic and deep learning signatures to predict overall survival and progression-free survival in NSCLC patients treated with chemotherapy. Cancer Imaging. 23(1):101.
27. Wang Y, Shao Q, Luo S, Fu R. 2021. Development of a nomograph integrating radiomics and deep features based on MRI to predict the prognosis of high grade Gliomas. Math Biosci Eng. 18(6):8084-8095.
28. Zhang X, Liang C, Zeng D, Jiang X, Zhong R, Lan Y, *et al*. 2021. Pattern classification for breast lesion on FFDM by integration of radiomics and deep features. Comput Med Imaging Graph. 90:101922.
29. Tian Y, Komolafe TE, Zheng J, Zhou G, Chen T, Zhou B, *et al*. 2021. Assessing PD-L1 expression level *via* preoperative MRI in HCC based on integrating deep learning and radiomics features. Diagnostics (Basel). 11(10):1875.
30. Ning Z, Luo J, Xiao Q, Cai L, Chen Y, Yu X, *et al*. 2021. Multi-modal magnetic resonance imaging-based grading analysis for gliomas by integrating radiomics and deep features. Ann Transl Med. 9(4):298.
31. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, *et al*. 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging. 26(6):1045-1057.
32. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, *et al*. 2017. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. The Cancer Imaging Archive. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
33. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, *et al*. 2017. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
34. Cho HH, Lee SH, Kim J, Park H. 2018. Classification of the glioma grading using radiomics analysis. Peer J. 6:e5982.
35. Li Y, Ammari S, Lawrance L, Quillent A, Assi T, Lassau N, *et al*. 2022. Radiomics-based method for predicting the glioma subtype as defined by tumor grade, IDH mutation, and 1p/19q codeletion. Cancers (Basel). 14(7):1778.
36. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, *et al*. 2017. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 77(21):e104-e107.
37. Ma C, Wang L, Song D, Gao C, Jing L, Lu Y, *et al*. 2023. Multimodal-based machine learning strategy for accurate and non-invasive prediction of intramedullary glioma grade and mutation status of molecular markers: A retrospective study. BMC Med. 21(1):198.
38. Huang G, Liu Z, Laurens VDM, Weinberger KQ. 2017. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:4700-4708.
39. Woo S, Park J, Lee J-Y, Kweon IS. 2018. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision. 2018:3-19.
40. Liu Y, Shao Z, Hoffmann N. 2021. Global attention mechanism: Retain information to enhance channel-spatial interactions. arXiv preprint arXiv:2112.05561.
41. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. 2017. Grad-CAM: Visual explanations from deep networks *via* gradient-based localization. IEEE International Conference on Computer Vision. 2017:618-626.
42. Hebart MN, Baker CI. 2018. Deconstructing multivariate decoding for the study of brain function. Neuroimage. 180(Pt A):4-18.
43. Pan J, Lv R, Wang Q, Zhao X, Liu J, Ai L. 2023. Discrimination between leucine-rich glioma-inactivated 1 antibody encephalitis and gamma-aminobutyric acid B receptor antibody encephalitis based on ResNet18. Vis Comput Ind Biomed Art. 6(1):17.
44. Lin K, Cidan W, Qi Y, Wang X. 2022. Glioma grading prediction using multiparametric magnetic resonance imaging-based radiomics combined with proton magnetic resonance spectroscopy and diffusion tensor imaging. Med Phys. 49(7):4419-4429.