Original Article

Developing and validating a deep learning and radiomic model for glioma grading using multiplanar reconstructed magnetic resonance contrast-enhanced T1-weighted imaging: a robust, multi-institutional study

Jialin Ding[1,2], Rubin Zhao[3], Qingtao Qiu[2], Jinhu Chen[2], Jinghao Duan[2], Xiujuan Cao[2], Yong Yin[2]

[1]School of Physics and Electronics, Shandong Normal University, Jinan, China; [2]Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China; [3]Department of Radiation Oncology and Technology, Linyi People's Hospital, Linyi, China

Contributions: (I) Conception and design: J Ding, Q Qiu, Y Yin; (II) Administrative support: Q Qiu, Y Yin; (III) Provision of study materials or patients: J Ding, R Zhao, Q Qiu; (IV) Collection and assembly of data: J Ding, Q Qiu, J Chen, J Duan; (V) Data analysis and interpretation: J Ding, Q Qiu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Qingtao Qiu; Yong Yin. Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan 250117, China. Email: qiuqingt@126.com; yinyongsd@126.com.

Background: Although surgical pathology or biopsy are considered the gold standard for glioma grading, these procedures have limitations. This study set out to evaluate and validate the predictive performance of a deep learning radiomics model based on contrast-enhanced T1-weighted multiplanar reconstruction images for grading gliomas.

Methods: Patients from three institutions who diagnosed with gliomas by surgical specimen and multiplanar reconstructed (MPR) images were enrolled in this study. The training cohort included 101 patients from institution 1, including 43 high-grade glioma (HGG) patients and 58 low-grade glioma (LGG) patients, while the test cohorts consisted of 50 patients from institutions 2 and 3 (25 HGG patients, 25 LGG patients). We then extracted radiomics features and deep learning features using six pretrained models from the MPR images. The Spearman correlation test and the recursive elimination feature selection method were used to reduce the redundancy and select most predictive features. Subsequently, three classifiers were used to construct classification models. The performance of the grading models was evaluated using the area under the receiver operating curve, sensitivity, specificity, accuracy, precision, and negative predictive value. Finally, the prediction performances of the test cohort were compared to determine the optimal classification model.

Results: For the training cohort, 62% (13 out of 21) of the classification models constructed with MPR images from multiple planes outperformed those constructed with single-plane MPR images, and 61% (11 out of 18) of classification models constructed with both radiomics features and deep learning features had higher area under the curve (AUC) values than those constructed with only radiomics or deep learning features. The optimal model was a random forest model that combined radiomic features and VGG16 deep learning features derived from MPR images, which achieved AUC of 0.847 in the training cohort and 0.898 in the test cohort. In the test cohort, the sensitivity, specificity, and accuracy of the optimal model were 0.840, 0.760, and 0.800, respectively.

Conclusions: Multiplanar CE-T1W MPR imaging features are more effective than features from single planes when differentiating HGG and LGG. The combination of deep learning features and radiomics features can effectively grade glioma and assist clinical decision-making.

Keywords: Multiplanar reconstruction (MPR); gliomas; radiomics; deep learning

Submitted Jul 12, 2021. Accepted for publication Oct 01, 2021.
doi: 10.21037/qims-21-722
View this article at: https://dx.doi.org/10.21037/qims-21-722

Introduction

Glioma is the most common primary tumour of the central nervous system (1,2). According to the diagnosis and treatment criteria raised by the World Health Organisation (WHO), gliomas can be classified into grades I–IV depending on the degree of malignancy. This classification can be simplified by grouping grades II and III as low-grade gliomas (LGGs) and grade IV as high-grade gliomas (HGGs) (3,4). Usually, patients with LGGs have a better prognosis than those with HGGs (5,6). Guidelines issued by the National Comprehensive Cancer Network (NCCN) recommend surgical resection combined with chemotherapy/radiotherapy as the main treatment strategy for gliomas (7). Treatment strategies vary greatly depending on the grade of the glioma, and preoperative grading of gliomas can guide decision-making for treatment delivery (8). Although surgical pathology or biopsy is considered the gold standard for glioma grading, these procedures have limitations, such as sampling error or delayed diagnosis (9). Therefore, the development of a non-invasive, accurate preoperative grading method is essential for the treatment and prognosis of patients with glioma (10).

Magnetic resonance imaging (MRI), particularly gadolinium-based contrast-enhanced T1-weighted imaging (CE-T1WI), is a routine clinical imaging modality that can be used to characterise gliomas according to their radiologic characteristics (11,12). However, the accuracy of glioma grading may suffer from radiologist inexperience and inter- or intra- observer variations (13). The quantitative features of MRI-derived tumours can provide supplemental information to decode the heterogeneity of the tumour and assist clinical decision-making, which can aid in glioma grading and prognosis (14). Numerous studies have investigated the glioma grade classification, some of which only used CE-T1W sequences or multiple sequences (11,15-17). However, in these studies, only the axial plane was used, not the coronal and sagittal planes, which can be observed using multiplanar reconstruction (MPR). MPR is the process of reformatting the volume of the original 3D projection data into three two-dimensional image planes. By displaying views in the axial plane (from top to bottom), coronal plane (from front to back), and sagittal plane (from left to right) of the 3D image dataset, this process can be used in many applications, as documented in the engineering field (18-22). We hypothesise that the integration of three-plane imaging features from MPR can be used to decode the three-dimensional information of the tumour, which may improve the performance of prediction models. However, there is a concern that MPR imaging can only be used to quantitatively analyse radiomics features while ignoring higher-level features, possibly including tumour heterogeneity information, which cannot be parsed by radiomics features. Therefore, the in-depth exploration of high-order three-dimensional MPR imaging features may improve the prediction accuracy of classification models (23).

In recent years, convolutional neural networks (CNNs) have developed rapidly and achieved widespread use in the medical image segmentation, classification, and detection fields; many CNN-based deep learning models can provide a diagnosis that is close to the diagnosis given by a doctor (24,25). Nevertheless, a notable limitation is that large datasets during training are required to achieve satisfactory accuracy, especially for medical imaging (26). Transfer learning using a pretrained CNN model on large natural image datasets for extracting high-order features from medical images may be beneficial for glioma grade classification (27). In this study, we aimed to develop and validate a deep learning radiomics (DLR) model based on three-plane MPR images to distinguish HGG from LGG. We present the following article in accordance with the TRIPOD reporting checklist (available at https://dx.doi.org/10.21037/qims-21-722).

Methods

Study design

Figure 1 shows our workflow, including image acquisition and segmentation, radiomics and deep learning feature extraction, and modelling. We first collected the axial, coronal, and sagittal data of enrolled patients from the three institutions, and used two feature extraction methods to extract the radiomics features and deep learning features of these images. Then, we used the Spearman correlation

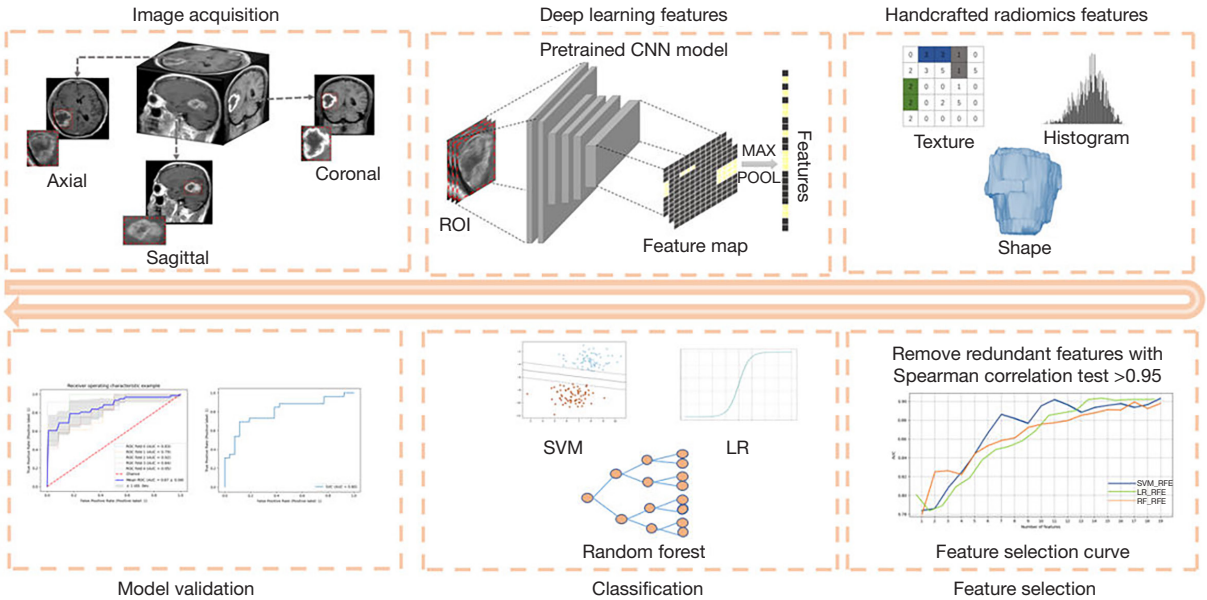

Figure 1 Study flow chart. First, the data were collected, and the radiomics features and deep learning features were extracted. Then, the features were selected, and modelling and model validation were carried out. LR, logistic regression; SVM, support vector machine.

test to remove redundant features, and the recursive feature elimination (RFE) feature selection method to select the optimal number of features. Finally, we used three machine learning algorithms support vector machine (SVM), logical regression (LR) and random forest to train and validate the models.

Datasets

In this study, 101 patients from Shandong Cancer Hospital and Institute and 30 patients from Linyi People's Hospital were selected from July 2, 2014, to December 30, 2020, as the training and test1 cohorts, respectively. Twenty patients from The Cancer Imaging Archive (TCIA) were selected as test2 cohorts (28-30). Test1 and test2 were merged together as a test set. The inclusion criteria were as follows: (I) availability of preoperative CE-T1W MPR MR images; (II) classification into grade II–IV glioma according to WHO diagnostic criteria; (III) high image quality (the image has a high signal-to-noise ratio and high contrast, and there are no imaging artifacts or image distortion). The exclusion criteria were as follows: (I) poor image quality (low signal-to-noise ratio, low contrast, image artifacts, and image distortion); and (II) a history of other malignancies. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective study was approved by the review committees of the Shandong Cancer Hospital and Institute and Linyi People's Hospital, and the request for informed consent was waived due to the retrospective nature of the study.

MRI data acquisition and tumour segmentation

The details of the MRI acquisition can be found in the Supplementary file. The gross tumour volumes (GTVs) were manually delineated on the axial MPR images by an oncologist using the 3D Slicer software (version 4.8.1, http://www.slicer.org, USA), after which the contoured GTVs were examined slice-by-slice by another oncologist with 5 years of work experience. To ensure the consistency of the three MPR planes, the axial MPR images and corresponding contours were reconstructed into coronal and sagittal images and contours as reference, after which the oncologists continued to delineate the original coronal and sagittal images.

To extract deep learning features, for axial, coronal, and sagittal sections, we first defined the region of interest (ROI, tumour area), then located the largest tumour areas and two adjacent slices by ROI, and then covered the ROI with a bounding box. Because the pretraining model had three RGB channels and our image was in grayscale, we put three slices into three channels, and the colour range of the

original input image of the pretraining model was 0–255. Therefore, we normalised the image to the same range, and then, to meet the input size of the pretraining model, we resized the image size to 224×224.

Radiomics features

For radiomics features, we delineated all axial, coronal, and sagittal slices and extracted the features with PyRadiomics (version 3.0.1) (31). PyRadiomics is an open-source Python package that can automatically extract quantitative features from medical images. In this study, 851 radiomic features were extracted from axial, coronal, and sagittal images, including: (I) shape features (n=14); (II) first-order features (n=162); (III) second-order features: grey-level cooccurrence matrix features (GLCM, n=216); (IV) grey-level run-length matrix features (GLRLM, n=144); (V) grey-level size zone matrix features (GLSZM, n=144); (VI) neighbouring grey-tone difference matrix features (NGTDM, n=45); and (VII) grey-level dependence matrix features (GLDM, n=126). Detailed descriptions of these features can be found in the Supplementary file.

Deep learning features

In this study, we use pretrained models, including VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, and Xception (32-36), on large ImageNet datasets to extract their image features. We removed the fully connected layer and output the features of the last convolutional layer, extracting 512, 512, 2,048, 2,048, 1,536, and 2,048 features. We also used guided gradient-weighted class activation mapping (guided Grad-CAM) to visualise the output of the last convolution layer and to show which locations in the tumour image were important for extracting the deep learning features.

Feature selection and model construction

Both deep learning and radiomic feature selection involved a two-step process. First, we eliminated redundant features with a correlation coefficient higher than 0.95 according to the Spearman correlation test. Second, we evaluated the classification performance of the models using the remaining feature sets by using the corresponding recursive feature elimination (RFE) feature selection method for various models (37,38). This method ranks the features according to their importance using an iterative process, and removes the features with low weights that make limited contributions to classifier performance.

SVM, RF, and LR were separately used to establish classifiers to distinguish HGG from LGG. SVM uses four kinds of kernel functions for classification (linear kernel function, polynomial kernel function, radial basis kernel, and sigmoid kernel function). The maximum depth of the random forest classifier that was composed of multiple decision trees was 12. The logical regression classifier used L2 regularisation. In addition, a combined model was established by using the features from the three MPR planes, and the DLR model was established by using both the deep learning and radiomics features to explore whether the performance of the MPR-based model could be further improved. During the training process, five-fold cross-validation was performed 500 times to avoid overfitting. The principle of five-fold cross-validation is detailed in the Supplementary file. After training, independent test datasets were used to evaluate the robustness and accuracy of the model. In this study, the model was evaluated by calculating the area under the curve (AUC), sensitivity, specificity, precision, negative predictive value (NPV), and accuracy. The definitions of these indicators are listed in detail in Table S1. The code for feature selection and model construction is available on GitHub (https://github.com/ljljlj02/deep-learning-radiomics-feature-extraction).

Statistical analysis

The Wilcoxon rank-sum test and the chi-square test were used to compare the patients' characteristics, where appropriate. A two-sided P<0.05 was considered statistically significant. The code for statistical analysis is available on GitHub (https://github.com/ljljlj02/deep-learning-radiomics-feature-extraction).

Results

Patients' characteristics

As shown in *Table 1*, there was no difference in age (P=0.464), pathological grade (χ^2=0.745, P=0.388) or sex (χ^2=0.267, P=0.605) between the training cohort and the test cohort. *Table 2* shows that there was no significant difference in the tumour volume among the axial, coronal, and sagittal sections between the training cohort and the test cohort (P=0.153–0.992).

Table 1 Demographic and clinical characteristics of the enrolled patients in the training and test cohorts

Characteristics	Train cohort (n=101), institution 1	Test cohort (n=50)			P value 2
		Institution 2	Institution 3	P value 1	
Age, mean ± SD, years	48.17±12.44	43.23±7.47	45.67±13.65	0.648	0.464
Gender				0.295	0.605
Male	61 (60%)	15 (30%)	13 (26%)		
Female	40 (40%)	15 (30%)	7 (14%)		
Grades of gliomas (%)				0.248	0.388
HGG	43 (43%)	17 (34%)	8 (16%)		
LGG	58 (57%)	13 (26%)	12 (24%)		

SD, standard deviation; HGG, high grade glioma; LGG, low grade glioma.

Table 2 Tumour volumes of patients in the training and test cohorts segmented from three planes of MPR images

MPR images	Training cohort		Test cohort	
	LGG	HGG	LGG	HGG
Axial	53.20±31.94	49.68±43.26	45.61±36.68	54.84±52.62
Coronal	51.81±30.38	49.78±42.41	46.09±36.17	54.11±52.46
Sagittal	51.43±31.51	46.56±39.21	45.44±36.78	52.73±51.23

All data are volumes, with the standard deviation in parentheses. No difference was found between the training and validation cohorts (P=0.153–0.992). MPR, multiplanar reconstruction; LGG, low-grade glioma; HGG, high-grade glioma.

Radiomics features

A total of 851 radiomic features were extracted from the three MPR planes, and redundant features in the axial, coronal, and sagittal planes were removed by the Spearman correlation coefficient method, leaving 247, 261, and 269 features, respectively. The number of selected features is illustrated in Figure S1. There were 194 repetitive features among the three feature categories, including 8 shape features, 38 first-order features, and 148 second-order features. Three different feature selection methods and the corresponding modelling methods were used to compare and distinguish HGG from LGG. As shown in *Table 3*, for random forest, the AUCs with the axial, coronal, and sagittal features in the test cohort were 0.742, 0.753, and 0.710, respectively. After combining the features of the three planes, the AUC, accuracy, sensitivity, specificity, precision, and NPV in the test cohort were 0.822, 0.740, 0.800, 0.680, 0.714, and 0.773, respectively. For LR, the axial, coronal, and sagittal AUC values in the test cohort were 0.811, 0.758, and 0.744, respectively. The AUC, accuracy, sensitivity, specificity, precision, and NPV of the combined model in the test cohort were 0.822, 0.680, 0.760, 0.600, 0.655, and 0.714, respectively. For SVM, the axial, coronal, and sagittal AUC values in the test cohort were 0.790, 0.788, and 0.686, respectively. The AUC, accuracy, sensitivity, specificity, precision, and NPV of the combined model in the test cohort were 0.822, 0.740, 0.760, 0.720, 0.731, and 0.750, respectively. All selected radiomics features are listed in Table S2.

Deep learning features

The deep learning features selected from six different pretrained models were used to compare the performance in classifying HGG and LGG. By combining the different feature selection methods and modelling methods, we built 18 models; their prediction results can be found in the Tables S3,S4.

Table S4 shows that, for the SVM and random forest models, the optimal models were constructed with the features extracted by the pretrained VGG19 and Xception models. The AUCs of the training cohort for the combined

Table 3 Performance of the optimal combined model (radiomics + VGG16 features) with the random forest model using the test cohort

Source of features	Test cohort					
	AUC (95% CI)	Accuracy	Sensitivity	Specificity	Precision	NPV
Radiomics-axial	0.742 (0.606–0.877)	0.640	0.560	0.720	0.667	0.621
Radiomics-coronal	0.753 (0.609–0.897)	0.560	0.880	0.240	0.537	0.667
Radiomics-sagittal	0.710 (0.564–0.857)	0.720	0.600	0.840	0.789	0.677
Radiomics-combine	0.822 (0.706–0.937)	0.740	0.800	0.680	0.714	0.773
VGG16-axial	0.674 (0.515–0.832)	0.700	0.520	0.880	0.813	0.647
VGG16-coronal	0.734 (0.594–0.874)	0.680	0.760	0.600	0.655	0.714
VGG16-sagittal	0.602 (0.442–0.763)	0.580	0.720	0.440	0.563	0.611
VGG16-combine	0.712 (0.569–0.855)	0.600	0.680	0.520	0.586	0.619
(Radiomics + VGG16)-combine	0.898 (0.809–0.986)	0.800	0.840	0.760	0.778	0.826

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

models were 0.945, and 0.898, and the AUCs of the test cohort were 0.782 and 0.746, which were higher than the corresponding values for the three single planes. In the LR model, although a relatively high AUC was obtained by using Inception-ResNetV2 (training cohort 0.946, test cohort 0.706), the results showed that the model could not distinguish between valid samples, because the specificity was too low (0.280), which may have caused misclassification.

Combination of radiomics and deep learning features

We further explored whether the model performance could be improved by merging and selecting different classifiers using a combination of radiomics features and deep learning features. The SVM model was constructed with a combination of Xception deep learning and radiomics features, and the AUC value, accuracy, sensitivity, specificity, and precision were 0.867, 0.760, 0.760, 0.760, and 0.760 in the test cohort, respectively. All the performance indicators were well balanced. The LR model was constructed by combining VGG19 deep learning features and radiomics features, and the sensitivity and specificity were quite different (0.920 *vs.* 0.600). This shows that most of the samples were predicted as positive, which is disadvantageous to the generalisation performance of the model. The random forest model was built from a combination of VGG16 deep learning features and radiomics features, and its feature selection curve is shown in *Figure 2*. Seventeen features were selected by the RFE method across the radiomics model, the deep learning model, and the combined model. *Table 3* shows the random forest combined model performance, which achieved relatively high values for the performance indicators with the test cohort (AUC 0.898, accuracy 0.800, sensitivity 0.840, specificity 0.760, precision 0.778, NPV 0.826). The receiver operating characteristic (ROC) curve (shown in *Figure 3*) indicated that the model has strong generalisability.

Discussion

In this study, deep learning and radiomic features derived from axial, coronal, and sagittal plane MPR images were used to construct models for LGG and HGG classification. For the test cohort, 62% (13 out of 21) of the classification models constructed with multiple planes MPR images outperformed those constructed with single-plane MPR images from single planes, and 44% (8 out of 18) of classification models constructed with both radiomics features and deep learning features had higher AUCs than those constructed with only radiomics or deep learning features. The results of this study show that the features combined from three-plane MPR images can better distinguish HGGs from LGGs than the features from single-plane MPR images using the SVM and LR models. The sensitivity and specificity of the model that was based only on deep learning features were low, and the model's generalisability was poor. By using a combination of deep learning and radiomics features, some models demonstrated a greater ability to distinguish between HGGs and LGGS,

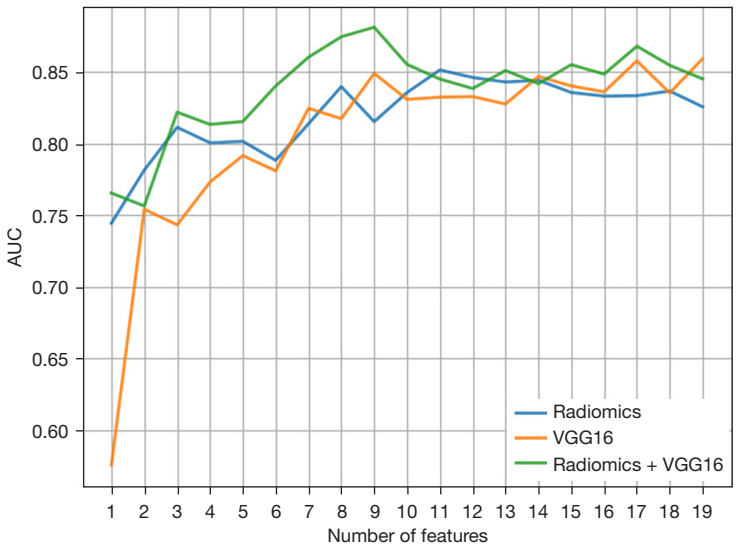

Figure 2 The RFE method was used to evaluate the feature selection process to determine the model AUC value and the optimal number of features. The X-axis represents the different number of features selected by the model, and the Y-axis indicates the AUC values calculated with the training cohort for the given number of features, each of which was averaged over five experiments. The different-coloured curves represent different feature sources (blue represents radiomics features, orange represents deep learning features extracted using the pretrained VGG16 model, and green represents their combination). RFE, recursive feature elimination; AUC, area under the curve.

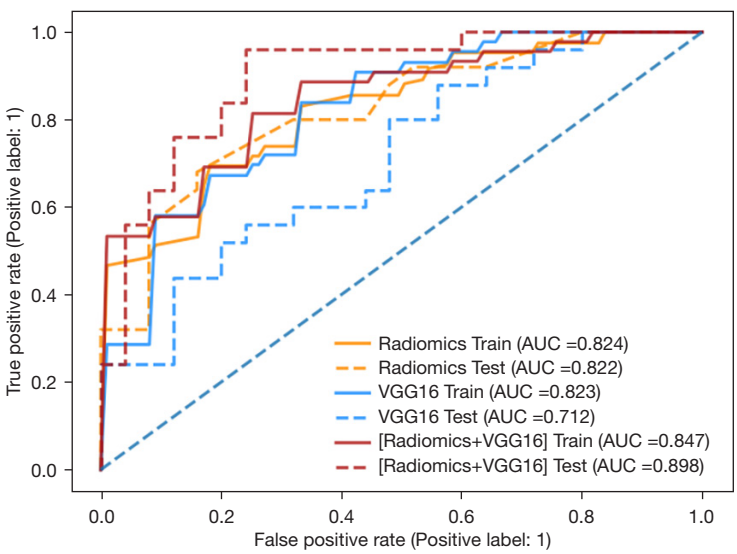

Figure 3 The ROC curve of the prediction model were built with different features following five-fold cross-validation with the original dataset. ROC, receiver operating characteristic.

which made them more robust.

The results also indicate that, if the modelling performance with the axial, coronal, or sagittal plane is poor, or if the three-plane model does not perform well in the test cohort, it will be difficult to improve the model performance using both sets of features. Additionally, we noted that the three-plane MPR-based model constructed with only radiomic features performed better than the model constructed with only deep learning features. For the test cohort, the sensitivity and specificity of the models built

with only deep learning features were usually low, which frequently led to misclassification. This may be because the deep learning models were pretrained on a large (~ millions) natural image dataset. Since the size of this dataset is much larger than the medical image dataset used in this study, the transfer learning step resulted in overfitting, as shown in *Figure 2*. Although the deep learning feature selection curve had a higher trend, we carefully selected more deep learning features to model, and we controlled the number of all features between 10 and 20. Among the models constructed with a combination of deep learning features and radiomics features, the most stable was the model constructed with radiomics features + VGG16 deep learning features with random forest (test cohort: AUC 0.898, accuracy 0.800, sensitivity 0.840, specificity 0.760), in which the radiomics features accounted for 88.2% (15 radiomics features/17 deep learning radiomics features). Despite the high AUC value of the original group, the AUC value of the model constructed only with the VGG16 features was 0.712 in the test cohort, which was smaller than the AUC of 0.822 of radiomics feature modelling. This may indicate that individual radiomics features are more competitive than VGG16 features in random forest modelling. Therefore, the proportion of radiomics features in the combined model was high, and the result was good. In most of our models, the performance of either the deep learning model or the radiomics model was insufficient, making it difficult to improve the performance of the combined model, which was consistent with the conclusions of previous research (39).

In Table S2, SVM_RFE, RF_RFE and LR_RFE were used to filter out a common feature, "Wavelet-HLL-firstorder_Mean", among radiomics features. According to the Pyradiomics manual, this feature denotes "the average grey-level intensity within the ROI" (31). We calculated the average and standard deviation of this feature, which was 0.75±0.12 in LGG and 0.53±0.21 in HGG. This may indicate that the higher the grey intensity value, the more likely it is to be LGG. *Figure 4* shows that the area around the tumour in the VGG16 images had a feature extraction value. The features around the tumour may be more valuable than those within the tumour, which was also reported in previous research (39). This may be related to the microenvironment around the tumour, which is still important in tumour progression, survival, and prognosis (40,41).

In recent years, there have been many studies on glioma grading. In research using multifunctional images for analysis, the importance of the imaging features from CE-T1W images in glioma classification cannot be ignored. One study used multiple imaging modalities [CE-T1WI, fluid-attenuated inverse recovery (FLAIR) imaging, diffusion-weighted imaging (DWI)] for analysis. After feature selection modelling, only CE-T1W texture features were found to participate in the identification of high- and low-grade gliomas, with an AUC value of up to 0.90 (16). Other studies have shown that high-order texture features can effectively predict the grade of malignant gliomas (grades III and IV). Six optimal features were ultimately extracted, of which five were derived from CE-T1W images and one from T2-weighted (T2W) images, and the resulting model achieved an AUC of 0.902±0.024 in the training cohort and 0.75 in the independent test cohort (17). In addition, another study used multiple imaging modalities, including CE-T1WI, apparent diffusion coefficient (ADC), T2-weighted imaging (T2WI), and cerebral blood flow (CBF) imaging, and found that the features extracted by CE-T1WI accounted for the largest weights in glioma grading. The authors concluded that the texture features extracted with conventional anatomical MRI, especially those from CE-T1WI, may lead to more accurate classification than those extracted via other methods (11). However, the above studies only focused on the axial plane from CE-T1W images. Accordingly, we explored whether combining the axial, coronal, and sagittal features from MPR images can achieve better results. According to our results, the combination of features from all three planes in CE-T1W MPR imaging performed much better in distinguishing HGG from LGG than only the axial CE-T1W MPR imaging features. We found this in our optimal models, including those constructed with only radiomics features, only deep learning features, or a combination of both. Therefore, future research should consider analysing the features from the other two planes, which may lead to unexpected findings that may shed further light on the results of the present study. Because of the difficulty in obtaining data from other imaging sequences, we only explored single- and three-plane CE-T1W MPR imaging features to distinguish HGG from LGG. In the future, we plan to add comparative studies of three-plane and single-plane MPR imaging using other sequences.

In addition, some studies have examined glioma grading using deep learning imaging, and we noted that these studies use the BraTs dataset (42-45), an open challenge MRI glioma dataset. The dataset contains T1, gd-enhanced T1, T2, and FLAIR sequences of LGG or HGG patients, and these datasets are pre-processed. All clinically obtained multiparameter MRI scans were registered into a common

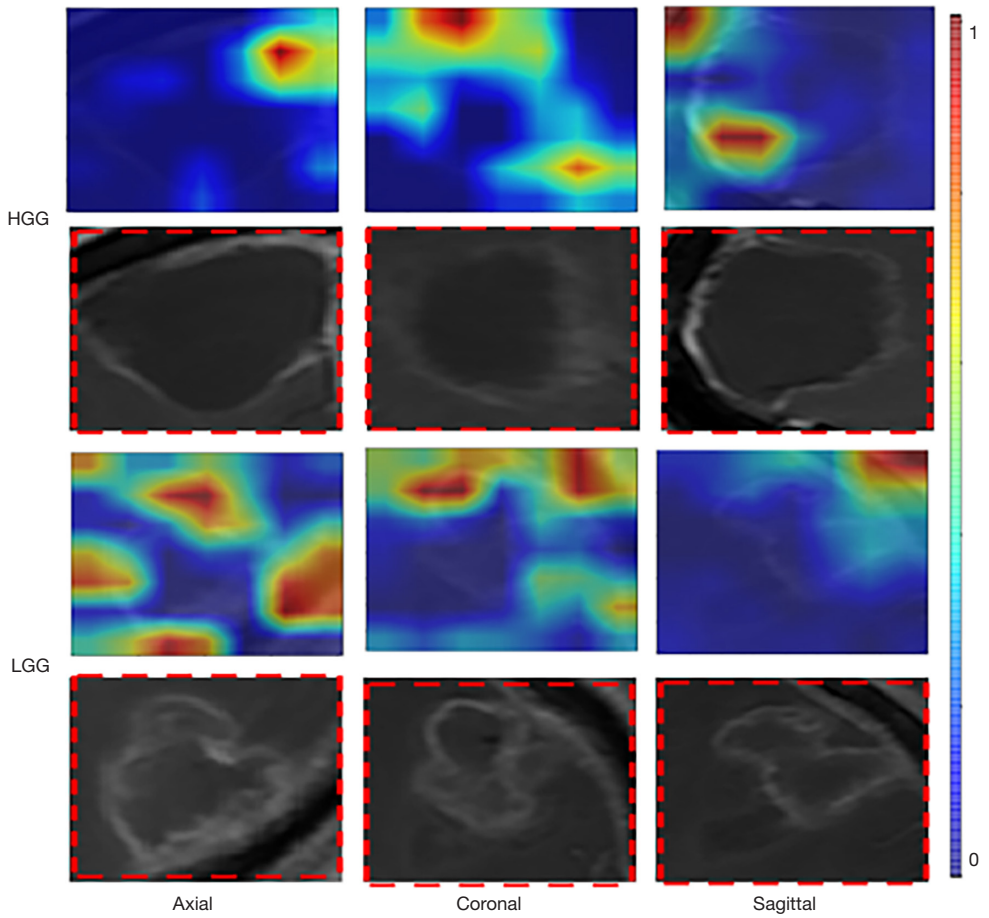

Figure 4 The CE-T1W images and corresponding deep learning feature heat maps of HGG and LGG patients were generated by a pretrained VGG16 model based on Guided Grad-CAM. The colour bar represents the proportional weights of the deep learning features. HGG, high-grade glioma; LGG, low-grade glioma.

anatomical template, resampled to 1 mm^3, and skull-stripped. We attempted to obtain the BraTs dataset, however, we found that all its images were axial images, and we could not perform MPR three-phase research, so we did not use this dataset. Some studies have extracted imaging features, and then used the deep neural network as a classifier to classify gliomas and achieved good results (46). In these studies, the fixed input of the convolutional network in our study was images rather than features. In the future, we may try this new approach.

Our research has some additional limitations. First, although our models demonstrated satisfactory performance, the pretrained models were constructed using millions of natural images. In the future, we expect to develop a very large medical database to achieve transfer learning to avoid overfitting, potentially leading to the extraction of unexpected high-order features that are different from those in the current study to assess tumour heterogeneity. Second, the clinical information was not included in the model. Furthermore, the DLR model only required three tumour slices to be input into the pretrained model, which may not provide a good overview of the tumour, and therefore should be addressed in future studies.

Conclusions

We developed and validated deep learning and radiomics models based on CE-T1W images to distinguish between HGG and LGG and found that they outperformed models constructed with only radiomics features. Furthermore, models constructed by integrating all three planes from MPR images were better at distinguishing between HGG

and LGG than those constructed using only single-plane MPR images.

Acknowledgments

Funding: This study received funding by the National Natural Science Foundation of China (Grant No. 82001902 and No. 82072094), the Natural Science Foundation of Shandong Province (Grant No. ZR2020QH198 and No. ZR2019LZL017), and the Taishan Scholars Project of Shandong Province (Grant No. ts201712098).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://dx.doi.org/10.21037/qims-21-722

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/qims-21-722). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective study was approved by the review committees of the Shandong Cancer Hospital and Institute and Linyi People's Hospital, and the request for informed consent was waived due to the retrospective nature of the study.

References

1. Weller M, van den Bent M, Tonn JC, Stupp R, Preusser M, Cohen-Jonathan-Moyal E, et al. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. Lancet Oncol 2017;18:e315-29.
2. Takacs GP, Flores-Toro JA, Harrison JK. Modulation of the chemokine/chemokine receptor axis as a novel approach for glioma therapy. Pharmacol Ther 2021;222:107790.
3. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol 2016;131:803-20.
4. Takahashi S, Takahashi W, Tanaka S, Haga A, Nakamoto T, Suzuki Y, Mukasa A, Takayanagi S, Kitagawa Y, Hana T, Nejo T, Nomura M, Nakagawa K, Saito N. Radiomics Analysis for Glioma Malignancy Evaluation Using Diffusion Kurtosis and Tensor Imaging. Int J Radiat Oncol Biol Phys 2019;105:784-91.
5. Lapointe S, Perry A, Butowski NA. Primary brain tumours in adults. Lancet 2018;392:432-46.
6. Tan AC, Ashley DM, López GY, Malinzak M, Friedman HS, Khasraw M. Management of glioblastoma: State of the art and future directions. CA Cancer J Clin 2020;70:299-312.
7. 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 Canc Netw 2020;18:1537-70.
8. Reni M, Mazza E, Zanon S, Gatta G, Vecht CJ. Central nervous system gliomas. Crit Rev Oncol Hematol 2017;113:213-34.
9. Amraei R, Moradi A, Zham H, Ahadi M, Baikpour M, Rakhshan A. A Comparison between the Diagnostic Accuracy of Frozen Section and Permanent Section Analyses in Central Nervous System Asian Pac J Cancer Prev 2017;18:659-66.
10. Sanai N, Berger MS. Surgical oncology for gliomas: the state of the art. Nat Rev Clin Oncol 2018;15:112-25.
11. Tian Q, Yan LF, Zhang X, Zhang X, Hu YC, Han Y, Liu ZC, Nan HY, Sun Q, Sun YZ, Yang Y, Yu Y, Zhang J, Hu B, Xiao G, Chen P, Tian S, Xu J, Wang W, Cui GB. Radiomics strategy for glioma grading using texture features from multiparametric MRI. J Magn Reson Imaging 2018;48:1518-28.
12. Geethanath S, Vaughan JT Jr. Accessible magnetic resonance imaging: A review. J Magn Reson Imaging 2019;49:e65-77.

13. Crowe EM, Alderson W, Rossiter J, Kent C. Expertise Affects Inter-Observer Agreement at Peripheral Locations within a Brain Tumor. Front Psychol 2017;8:1628.
14. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, Cook G. Introduction to Radiomics. J Nucl Med 2020;61:488-95.
15. Skogen K, Schulz A, Dormagen JB, Ganeshan B, Helseth E, Server A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. Eur J Radiol 2016;85:824-9.
16. Ditmer A, Zhang B, Shujaat T, Pavlina A, Luibrand N, Gaskill-Shipley M, Vagal A. Diagnostic accuracy of MRI texture analysis for grading gliomas. J Neurooncol 2018;140:583-9.
17. Nakamoto T, Takahashi W, Haga A, Takahashi S, Kiryu S, Nawa K, Ohta T, Ozaki S, Nozawa Y, Tanaka S, Mukasa A, Nakagawa K. Prediction of malignant glioma grades using contrast-enhanced T1-weighted and T2-weighted magnetic resonance images based on a radiomic analysis. Sci Rep 2019;9:19411.
18. Koonsanit K, Thongvigitmanee S, Narkbuakaew W, Yampri P, Sinthupinyo W, Oblique Multi-Planar Reformation Using Viewport Translation Technique on Visualization Toolkit Library. Seoul, Korea: 2010 International Conference on Information Science and Applications, 2010.
19. Kanezaki A, Matsushita Y, Nishida Y. RotationNet for Joint Object Categorization and Unsupervised Pose Estimation from Multi-View Images. IEEE Trans Pattern Anal Mach Intell 2021;43:269-83.
20. Yan C, Gong B, Wei Y, Gao Y. Deep Multi-View Enhancement Hashing for Image Retrieval. IEEE Trans Pattern Anal Mach Intell 2021;43:1445-51.
21. Chen Y, Li D, Zhang X, Jin J, Shen Y. Computer aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning. Med Image Anal 2021;67:101819.
22. Qu Y, Li X, Yan Z, Zhao L, Zhang L, Liu C, Xie S, Li K, Metaxas D, Wu W, Hao Y, Dai K, Zhang S, Tao X, Ai S. Surgical planning of pelvic tumor using multi-view CNN with relation-context representation learning. Med Image Anal 2021;69:101954.
23. Han W, Qin L, Bay C, Chen X, Yu KH, Miskin N, Li A, Xu X, Young G. Deep Transfer Learning and Radiomics Feature Prediction of Survival of Patients with High-Grade Gliomas. AJNR Am J Neuroradiol 2020;41:40-8.
24. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. Radiology 2019;290:590-606.
25. Attanasio S, Forte SM, Restante G, Gabelloni M, Guglielmi G, Neri E. Artificial intelligence, radiomics and other horizons in body composition assessment. Quant Imaging Med Surg 2020;10:1650-60.
26. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. Nat Med 2019;25:24-9.
27. Hu Y, Xie C, Yang H, Ho JWK, Wen J, Han L, Lam KO, Wong IYH, Law SYK, Chiu KWH, Vardhanabhuti V, Fu J. Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma. Radiother Oncol 2021;154:6-13.
28. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013;26:1045-57.
29. Pedano N, Flanders AE, Scarpace L, Mikkelsen T, Eschbacher JM, Hermes B, Ostrom Q. Radiology Data from The Cancer Genome Atlas Low Grade Glioma [TCGA-LGG] collection. The Cancer Imaging Archive. 2016. Available online: http://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK
30. Beers A, Gerstner E, Rosen B, Clunie D, Pieper S, Fedorov A, Kalpathy C. DICOM-SEG Conversions for TCGA-LGG and TCGA-GBM Segmentation Datasets. TCGA-GBM Segmentation Datasets. The Cancer Imaging Archive. 2018. Available online: https://doi.org/10.7937/TCIA.2018.ow6ce3ml
31. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77:e104-7.
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 2014. Available online: https://arxiv.org/abs/1409.1556
33. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. Vegas, NV, USA: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
34. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z.

Rethinking the inception architecture for computer vision. Vegas, NV, USA: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:2818-26.
35. Chollet F, Xception: Deep learning with depthwise separable convolutions, Proceedings of the IEEE conference on computer vision and pattern recognition. 2017; 1251-1258
36. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning, Proceedings of the AAAI Conference on Artificial Intelligence. San Francisco, California, USA: AAAI Press, 2017:4278-84.
37. Lopez-Rincon A, Mendoza-Maldonado L, Martinez-Archundia M, Schönhuth A, Kraneveld AD, Garssen J, Tonda A. Machine Learning-Based Ensemble Recursive Feature Selection of Circulating miRNAs for Cancer Tumor Classification. Cancers (Basel) 2020;12:1785.
38. Dai H, Lu M, Huang B, Tang M, Pang T, Liao B, Cai H, Huang M, Zhou Y, Chen X, Ding H, Feng ST. Considerable effects of imaging sequences, feature extraction, feature selection, and classifiers on radiomics-based prediction of microvascular invasion in hepatocellular carcinoma using magnetic resonance imaging. Quant Imaging Med Surg 2021;11:1836-53.
39. Wu X, Dong D, Zhang L, Fang M, Zhu Y, He B, Ye Z, Zhang M, Zhang S, Tian J. Exploring the predictive value of additional peritumoral regions based on deep learning and radiomics: A multicenter study. Med Phys 2021;48:2374-85.
40. Raviraj R, Nagaraja SS, Selvakumar I, Mohan S, Nagarajan D. The epigenetics of brain tumors and its modulation during radiation: A review. Life Sci 2020;256:117974.
41. Bao Z, Wang Y, Wang Q, Fang S, Shan X, Wang J, Jiang T. Intratumor heterogeneity, microenvironment, and mechanisms of drug resistance in glioma recurrence and evolution. Front Med 2021;15:551-61.
42. Cho HH, Lee SH, Kim J, Park H. Classification of the glioma grading using radiomics analysis. PeerJ 2018;6:e5982.
43. Xiao T, Hua W, Li C, Wang S. Glioma grading prediction by exploring radiomics and deep learning features. New York: Proceedings of the Third International Symposium on Image Computing and Digital Medicine, 2019.
44. Banerjee S, Mitra S, Masulli F, Rovetta S. Glioma classification using deep radiomics. SN Comput Sci 2020;1:1-14.
45. Kobayashi K, Miyake M, Takahashi M, Hamamoto R. Observing deep radiomics for the classification of glioma grades. Sci Rep 2021;11:10942.
46. Çinarer G, Emiroğlu BG, Yurttakal AH. Prediction of Glioma Grades Using Deep Learning with Wavelet Radiomic Features. Appl Sci 2020;10:6296.

Cite this article as: Ding J, Zhao R, Qiu Q, Chen J, Duan J, Cao X, Yin Y. Developing and validating a deep learning and radiomic model for glioma grading using multiplanar reconstructed magnetic resonance contrast-enhanced T1-weighted imaging: a robust, multi-institutional study. Quant Imaging Med Surg 2022;12(2):1517-1528. doi: 10.21037/qims-21-722

Methods

MR image acquisition

MR images from patients from Shandong Cancer Hospital and Research Institute were obtained using a Philips 3.0 Tesla magnetic resonance scanner (Philips Medical Systems, the Netherlands) and a 6-channel head coil. Fast spin-echo sequences were routinely used to obtain high-resolution T1-weighted MPR sequences with the following parameters: TR =2.5 ms, TE =2.3 ms, slice thickness =5.0 mm, matrix size =512×512, and in-plane resolution =1.56×1.56 mm^2. The contrast-enhanced scan was performed 3–5 min after the intravenous injection of 0.1 mmol/kg contrast medium at a speed of 2 mL/s. The axial scanning parameters were as follows: TR =3.8 ms, TE =1.7 ms, matrix size =512×512, layer thickness =3.2 mm, and in-plane resolution =1.5×1.50 mm^2. The images from the axial scan were used to reconstruct the coronal and sagittal planes. The reconstruction parameters were as follows: matrix size =512×512, slice thickness =3.5 mm, and in-plane resolution =1.5×1.50 mm^2.

The MR images from the patients from Linyi People's Hospital were obtained by a 3.0 Tesla (Magnetom Tim/Trio; Siemens, Germany) magnetic resonance scanner and a 6-channel head coil. The scanning parameters for the T1-weighted sequence were as follows: TR =7.5 ms, TE =3.2 ms, thickness =3 mm, matrix size =512×512, and in-plane resolution =1.39×1.39 mm^2. Contrast medium (0.2 mL/kg) was injected intravenously at a speed of 2 mL/s; 2–4 min later, a contrast-enhanced scan was performed. The axial scanning parameters were as follows: TR =3.6 ms, TE =1.5 ms, matrix size =512×512, layer thickness =3.3 mm, and in-plane resolution =1.49×1.49 mm^2. The axial scan images were then used to reconstruct the coronal and sagittal planes using the following reconstruction parameters: matrix size =512×512, slice thickness =3.0 mm, and in-plane resolution =1.37×1.37 mm^2. However, the imaging parameters of the MR images from the TCIA varied, with the use of 1.5 Tesla and 3.0 Tesla MR scanners and various TR, TE, thickness, and in-plane resolution settings.

The principle of five-fold cross-validation

The principle of five-fold cross-validation is that the training set is randomly divided into 5 subsamples; 4 samples are used for model training, while a single subsample is retained to verify the trained model. The cross-validation was repeated 5 times, each subsample was validated once, and the average result was calculated (47).

Definition of indicators

As shown in Table S1, the following calculated indicators can be obtained:

(I) True negative (TN): the sample number indicates that it is a negative sample and is predicted to be a negative sample;

(II) False positive (FP): the sample number indicates that it is a negative sample and is predicted to be a positive sample;

(III) False negative (FN): the sample number indicates that it is a positive sample and is predicted to be a negative sample;

(IV) True positive (TP): the sample number indicates that it is a positive sample and is predicted to be a positive sample;

(V) $Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$, the proportion of all correctly judged samples in all samples;

(VI) $Sensitivity = \frac{TP}{TP+FN}$, indicates the proportion of pairs in the positive sample;

(VII) $Specificity = \frac{TP}{TP+FP}$, indicates the proportion of pairs in a negative sample;

(VIII) $Precision / positive\ prediction\ value\ (PPV) = \frac{TP}{TP+FP}$, refers to the proportion of the true positive class among all the people who are judged to be positive;

(IX) $Negative\ predictive\ value\ (NPV) = \frac{TP}{TP+FN}$, refers to the proportion of the true negative class among all the people who are judged to be negative.

https://dx.doi.org/10.21037/qims-21-722

Radiomics features

Before the feature extraction, images were discretised the method "fixed bin size". Besides, GLCM, GLRLM, GLSZM, GLDM, and NGTDM are 2D features, while shape features are 3D. During feature calculation, directions and pixel distance were employed with default settings. First-order texture statistics were based on the first-order histogram that describes distribution of voxel intensities in an image (31).

Table S1 Confusion matrix

Predicted class	Acted Class	
	1	0
1	TP	FP
0	FN	TN

Table S2 Three kinds of feature recursive elimination methods (SVM_RFE, LR_RFE and RF_RFE) were used to select the image group features extracted from the three single-plane MPR images, which were then combined and again subjected to the feature selection method

SVM_Combine	LR_Combine	Random forest_Combine
original_shape_Sphericity	original_glcm_Imc2	wavelet-LHH_glcm_Correlation
wavelet-HLL_firstorder_Mean	original_shape_Sphericity	wavelet-HLL_firstorder_Mean
wavelet-LHL_firstorder_Mean	wavelet-HLL_firstorder_Mean	wavelet-LLH_ngtdm_Contrast
wavelet-HHL_glcm_Imc2	original_shape_SurfaceVolumeRatio	wavelet-HLH_glcm_Correlation
wavelet-LLL_firstorder_Minimum	wavelet-HHL_glcm_Imc1	wavelet-LHL_firstorder_Mean
wavelet-HHH_firstorder_Median	wavelet-HLH_firstorder_Median	wavelet-LLH_ngtdm_Contrast
wavelet-LLH_glcm_Correlation	wavelet-HHH_gldm_DependenceNonUniformityNormalised	wavelet-LHL_firstorder_Mean
wavelet-HLH_firstorder_Median	wavelet-HLH_glcm_MCC	wavelet-HLH_glszm_LargeAreaEmphasis
original_glszm_GrayLevelNonUniformity	wavelet-LLL_firstorder_Skewness	wavelet-HLL_firstorder_Mean
original_glszm_GrayLevelNonUniformity	wavelet-LLL_firstorder_Minimum	wavelet-HLH_glcm_Correlation
wavelet-LHL_gldm_DependenceVariance	original_gldm_SmallDependenceLowGrayLevelEmphasis	wavelet-HLL_glrlm_RunEntropy
original_shape_Elongation	original_glszm_GrayLevelNonUniformity	
wavelet-HHH_firstorder_Median	wavelet-LHH_firstorder_Skewness	
wavelet-HHL_ngtdm_Contrast	original_glszm_GrayLevelNonUniformity	
wavelet-LHL_firstorder_Mean	original_shape_Elongation	
wavelet-LLH_firstorder_Kurtosis	wavelet-HHH_firstorder_Median	
wavelet-LLH_glcm_Idmn	wavelet-HHL_ngtdm_Contrast	
wavelet-LLH_firstorder_Skewness	wavelet-LLH_firstorder_Kurtosis	
wavelet-LHH_firstorder_Median	wavelet-LLH_glcm_Idmn	
Common features: wavelet-HLL_firstorder_Mean		

H: high pass filter; L: low pass filter. LR, logistic regression; SVM, support vector machine; RFE, recursive feature elimination.

https://dx.doi.org/10.21037/qims-21-722

Table S3 Different single-plane MPR images, feature selection methods, classifiers, and deep learning feature extractors were employed to generate glioma grading models with cross-combinations of different deep learning and radiomics features, and their results were assessed with the training cohort

Model	Source of features	Train cohort					
		AUC	Accuracy	Sensitivity	Specificity	Precision	NPV
SVM	Radiomics-Axial	0.866±0.09	0.841	0.791	0.879	0.829	0.850
	Radiomics-Coronal	0.812±0.04	0.811	0.837	0.793	0.75	0.868
	Radiomics-Sagittal	0.927±0.03	0.871	0.884	0.862	0.826	0.909
	Radiomics-Combine	0.960±0.03	0.900	0.860	0.931	0.902	0.900
	VGG16-Axial	0.920±0.06	0.832	0.651	0.966	0.933	0.789
	VGG16-Coronal	0.895±0.038	0.772	0.767	0.776	0.717	0.818
	VGG16-Sagittal	0.963±0.016	0.921	0.837	0.983	0.973	0.891
	VGG16-Combine	0.967±0.030	0.911	0.837	0.966	0.947	0.889
	Radiomics+VGG16	0.927±0.055	0.832	0.744	0.897	0.842	0.825
	VGG19-Axial	0.859±0.097	0.811	0.581	0.983	0.962	0.760
	VGG19-Coronal	0.920±0.032	0.842	0.674	0.966	0.935	0.800
	VGG19-Sagittal	0.961±0.024	0.871	0.744	0.966	0.941	0.836
	VGG19-Combine	0.945±0.028	0.832	0.791	0.862	0.810	0.847
	[Radiomics+VGG19]-Combine	0.913±0.082	0.822	0.698	0.914	0.857	0.803
	Resnet-Axial	0.913±0.037	0.842	0.791	0.879	0.829	0.850
	Resnet-Coronal	0.942±0.054	0.881	0.860	0.897	0.840	0.897
	Resnet-Sagittal	0.950±0.034	0.881	0.814	0.931	0.897	0.871
	Resnet-Combine	0.933±0.060	0.901	0.837	0.948	0.923	0.887
	[Radiomics+Resnet]-Combine	0.896±0.053	0.703	0.349	0.966	0.882	0.667
	Xception-Axial	0.977±0.016	0.921	0.884	0.948	0.927	0.917
	Xception-Coronal	0.993±0.004	0.970	0.977	0.966	0.955	0.982
	Xception-Sagittal	0.978±0.015	0.970	0.977	0.966	0.955	0.982
	Xception-Combine	0.993±0.004	1.0	1.0	1.0	1.0	1.000
	[Radiomics+Xception]-Combine	0.983±0.017	0.931	0.884	0.966	0.950	0.918
	InceptionV3-Axial	0.977±0.020	0.941	0.907	0.966	0.951	0.933
	InceptionV3-Coronal	0.982±0.019	0.950	0.930	0.966	0.952	0.949
	InceptionV3-Sagittal	0.993±0.004	0.950	0.930	0.966	0.952	0.949
	InceptionV3-Combine	0.968±0.031	0.921	0.907	0.931	0.907	0.931
	[Radiomics+InceptionV3]-Combine	0.975±0.025	0.881	0.907	0.862	0.830	0.926
	InceptionResnetV2-Axial	0.770±0.055	0.732	0.767	0.707	0.660	0.804
	InceptionResnetV2-Coronal	0.960±0.032	0.901	0.884	0.914	0.884	0.914
	InceptionResnetV2-Sagittal	0.954±0.029	0.871	0.814	0.914	0.875	0.869
	InceptionResnetV2-Combine	0.957±0.042	0.891	0.837	0.931	0.900	0.885
	[Radiomics+InceptionResnetV2]-Combine	0.959±0.046	0.960	0.930	0.983	0.976	0.950
LR	Radiomics-Axial	0.898±0.069	0.792	0.630	0.914	0.844	0.768
	Radiomics-Coronal	0.893±0.088	0.832	0.814	0.845	0.795	0.860
	Radiomics-Sagittal	0.888±0.079	0.792	0.767	0.810	0.750	0.825
	Radiomics-Combine	0.919±0.054	0.881	0.837	0.914	0.878	0.883
	VGG16-Axial	0.897±0.062	0.812	0.791	0.828	0.773	0.842
	VGG16-Coronal	0.768±0.077	0.663	0.419	0.845	0.667	0.662
	VGG16-Sagittal	0.860±0.102	0.673	0.372	0.897	0.727	0.658
	VGG16-Combine	0.917±0.027	0.851	0.860	0.845	0.804	0.891
	[Radiomics+VGG16]-Combine	0.947±0.032	0.861	0.814	0.897	0.854	0.867
	VGG19-Axial	0.800±0.028	0.743	0.581	0.862	0.758	0.735
	VGG19-Coronal	0.872±0.144	0.792	0.744	0.828	0.762	0.814
	VGG19-Sagittal	0.933±0.015	0.822	0.791	0.844	0.791	0.845
	VGG19-Combine	0.937±0.037	0.871	0.791	0.931	0.895	0.857
	[Radiomics+VGG19]-Combine	0.947±0.047	0.911	0.884	0.931	0.905	0.915
	Resnet-Axial	0.849±0.068	0.752	0.674	0.810	0.725	0.770
	Resnet-Coronal	0.879±0.108	0.802	0.721	0.862	0.795	0.806
	Resnet-Sagittal	0.909±0.062	0.832	0.791	0.862	0.810	0.847
	Resnet-Combine	0.937±0.035	0.851	0.744	0.931	0.889	0.831
	[Radiomics+Resnet]-Combine	0.940±0.062	0.832	0.837	0.828	0.783	0.873
	Xception-Axial	0.967±0.026	0.901	0.837	0.948	0.923	0.887
	Xception-Coronal	0.958±0.039	0.910	0.860	0.948	0.925	0.902
	Xception-Sagittal	0.932±0.052	0.891	0.884	0.897	0.864	0.912
	Xception-Combine	0.952±0.053	0.901	0.860	0.931	0.902	0.900
	[Radiomics+Xception]-Combine	0.976±0.029	0.931	0.907	0.948	0.929	0.932
	InceptionV3-Axial	0.905±0.046	0.822	0.767	0.862	0.805	0.833
	InceptionV3-Coronal	0.950±0.024	0.842	0.767	0.897	0.846	0.839
	InceptionV3-Sagittal	0.889±0.087	0.812	0.698	0.897	0.833	0.800
	InceptionV3-Combine	0.959±0.448	0.901	0.907	0.897	0.867	0.929
	[Radiomics+InceptionV3]-Combine	0.972±0.041	0.921	0.884	0.948	0.927	0.917
	InceptionResnetV2-Axial	0.759±0.126	0.663	0.627	0.690	0.600	0.714
	InceptionResnetV2-Coronal	0.887±0.061	0.743	0.535	0.897	0.793	0.722
	InceptionResnetV2-Sagittal	0.827±0.136	0.772	0.744	0.793	0.727	0.807
	InceptionResnetV2-Combine	0.946±0.049	0.812	0.721	0.879	0.816	0.810
	[Radiomics+InceptionResnetV2]-Combine	0.984±0.011	0.931	0.907	0.948	0.929	0.932
Random Forest	Radiomics-Axial	0.824±0.073	0.752	0.651	0.828	0.737	0.762
	Radiomics-Coronal	0.811±0.095	0.772	0.674	0.845	0.763	0.778
	Radiomics-Sagittal	0.860±0.072	0.743	0.651	0.810	0.718	0.758
	Radiomics-Combine	0.824±0.084	0.733	0.628	0.810	0.711	0.746
	VGG16-Axial	0.678±0.084	0.634	0.465	0.759	0.588	0.657
	VGG16-Coronal	0.790±0.112	0.733	0.698	0.759	0.682	0.772
	VGG16-Sagittal	0.799±0.110	0.703	0.534	0.828	0.697	0.706
	VGG16-Combine	0.823±0.081	0.693	0.605	0.759	0.650	0.721
	Radiomics+VGG16	0.847±0.049	0.812	0.791	0.828	0.773	0.842
	VGG19-Axial	0.790±0.073	0.713	0.651	0.759	0.667	0.746
	VGG19-Coronal	0.818±0.112	0.743	0.651	0.810	0.718	0.758
	VGG19-Sagittal	0.690±0.107	0.703	0.651	0.741	0.651	0.741
	VGG19-Combine	0.865±0.086	0.802	0.674	0.897	0.829	0.788
	[Radiomics+VGG19]-Combine	0.845±0.084	0.772	0.698	0.828	0.750	0.769
	Resnet-Axial	0.906±0.074	0.861	0.767	0.931	0.892	0.844
	Resnet-Coronal	0.762±0.085	0.683	0.558	0.776	0.649	0.703
	Resnet-Sagittal	0.846±0.075	0.802	0.698	0.879	0.811	0.797
	Resnet-Combine	0.840±0.069	0.802	0.651	0.914	0.848	0.779
	[Radiomics+Resnet]-Combine	0.826±0.068	0.792	0.721	0.845	0.775	0.803
	Xception-Axial	0.839±0.091	0.792	0.698	0.862	0.789	0.794
	Xception-Coronal	0.884±0.058	0.792	0.698	0.862	0.789	0.794
	Xception-Sagittal	0.818±0.039	0.802	0.744	0.845	0.780	0.817
	Xception-Combine	0.898±0.048	0.792	0.744	0.828	0.762	0.814
	[Radiomics+Xception]-Combine	0.858±0.071	0.812	0.767	0.845	0.786	0.831
	InceptionV3-Axial	0.800±0.052	0.723	0.558	0.845	0.727	0.721
	InceptionV3-Coronal	0.858±0.064	0.802	0.674	0.897	0.829	0.788
	InceptionV3-Sagittal	0.664±0.142	0.663	0.372	0.879	0.696	0.654
	InceptionV3-Combine	0.811±0.064	0.713	0.581	0.810	0.694	0.723
	[Radiomics+InceptionV3]-Combine	0.846±0.068	0.782	0.721	0.828	0.756	0.800
	InceptionResnetV2-Axial	0.807±0.119	0.802	0.791	0.810	0.756	0.839
	InceptionResnetV2-Coronal	0.667±0.079	0.574	0.186	0.862	0.500	0.588
	InceptionResnetV2-Sagittal	0.832±0.057	0.733	0.581	0.845	0.735	0.731
	InceptionResnetV2-Combine	0.787±0.123	0.733	0.651	0.793	0.700	0.754
	[Radiomics+InceptionResnetV2]-Combine	0.873±0.078	0.762	0.721	0.793	0.721	0.793

LR, logistic regression; SVM, support vector machine; NPV, negative predictive value.

https://dx.doi.org/10.21037/qims-21-722

Table S4 Different single-plane MPR images, feature selection methods, classifiers, and deep learning feature extractors were employed to generate glioma grading models with cross-combinations of different deep learning and radiomics features, and their results were assessed with the test cohort

Model	Source of features	Test cohort							
		AUC	Accuracy	Sensitivity	Specificity	Precision	NPV	Number of features	Radiomics ratio
SVM	Radiomics-Axial	0.790	0.720	0.600	0.840	0.789	0.677	12	
	Radiomics-Coronal	0.788	0.780	0.720	0.840	0.818	0.750	19	
	Radiomics-Sagittal	0.686	0.680	0.800	0.560	0.645	0.737	19	
	Radiomics-Combine	0.822	0.740	0.760	0.720	0.731	0.750	19	
	VGG16-Axial	0.641	0.580	0.400	0.760	0.625	0.559	19	
	VGG16-Coronal	0.550	0.440	0.480	0.400	0.444	0.435	18	
	VGG16-Sagittal	0.612	0.480	0.440	0.520	0.478	0.481	19	
	VGG16-Combine	0.622	0.620	0.440	0.800	0.688	0.588	19	
	Radiomics+VGG16	0.792	0.680	0.640	0.720	0.696	0.667	17	10/17 (58.8%)
	VGG19-Axial	0.684	0.620	0.520	0.720	0.650	0.600	18	
	VGG19-Coronal	0.699	0.620	0.360	0.880	0.750	0.579	17	
	VGG19-Sagittal	0.568	0.480	0.280	0.680	0.447	0.486	19	
	VGG19-Combine	0.782	0.740	0.720	0.760	0.750	0.731	19	
	[Radiomics+VGG19]-Combine	0.760	0.680	0.360	1.00	1.00	0.610	19	11/19 (57.9%)
	Resnet-Axial	0.541	0.500	0.400	0.600	0.500	0.500	19	
	Resnet-Coronal	0.582	0.640	0.800	0.480	0.606	0.706	19	
	Resnet-Sagittal	0.662	0.580	0.400	0.760	0.625	0.559	19	
	Resnet-Combine	0.749	0.680	0.760	0.600	0.655	0.714	17	
	[Radiomics+Resnet]-Combine	0.811	0.720	0.640	0.800	0.762	0.670	16	10/16 (62.5%)
	Xception-Axial	0.528	0.520	0.200	0.840	0.556	0.512	19	
	Xception-Coronal	0.683	0.660	0.560	0.760	0.700	0.633	18	
	Xception-Sagittal	0.628	0.640	0.520	0.760	0.684	0.613	19	
	Xception-Combine	0.672	0.700	0.520	0.880	0.813	0.647	18	
	[Radiomics+Xception]-Combine	0.867	0.760	0.760	0.760	0.760	0.760	17	4/17 (23.5%)
	InceptionV3-Axial	0.529	0.600	0.760	0.440	0.576	0.647	18	
	InceptionV3-Coronal	0.533	0.460	0.400	0.520	0.455	0.464	19	
	InceptionV3-Sagittal	0.613	0.540	0.440	0.640	0.550	0.533	19	
	InceptionV3-Combine	0.707	0.660	0.600	0.720	0.682	0.643	16	
	[Radiomics+InceptionV3]-Combine	0.862	0.760	0.920	0.600	0.697	0.882	12	4/12 (33.3%)
	InceptionResnetV2-Axial	0.574	0.520	0.560	0.480	0.519	0.522	19	
	InceptionResnetV2-Coronal	0.602	0.540	0.160	0.920	0.667	0.523	18	
	InceptionResnetV2-Sagittal	0.702	0.520	0.040	1.00	1.00	0.510	19	
	InceptionResnetV2-Combine	0.686	0.600	0.320	0.880	0.727	0.564	19	
	[Radiomics+InceptionResnetV2]-Combine	0.849	0.740	0.680	0.800	0.773	0.714	19	7/19 (36.8%)
LR	Radiomics-Axial	0.811	0.740	0.720	0.760	0.750	0.731	14	
	Radiomics-Coronal	0.758	0.620	0.880	0.360	0.579	0.750	19	
	Radiomics-Sagittal	0.744	0.740	0.640	0.840	0.800	0.700	11	
	Radiomics-Combine	0.822	0.680	0.760	0.600	0.655	0.714	19	
	VGG16-Axial	0.690	0.640	0.600	0.680	0.652	0.630	14	
	VGG16-Coronal	0.710	0.720	0.600	0.840	0.789	0.677	18	
	VGG16-Sagittal	0.651	0.580	0.400	0.760	0.625	0.559	19	
	VGG16-Combine	0.692	0.600	0.480	0.720	0.632	0.581	19	
	[Radiomics+VGG16]-Combine	0.782	0.720	0.800	0.640	0.690	0.762	19	10/19 (52.6%)
	VGG19-Axial	0.669	0.600	0.520	0.680	0.619	0.586	19	
	VGG19-Coronal	0.704	0.660	0.600	0.720	0.682	0.643	19	
	VGG19-Sagittal	0.603	0.600	0.600	0.600	0.600	0.600	17	
	VGG19-Combine	0.597	0.540	0.400	0.680	0.556	0.531	17	
	[Radiomics+VGG19]-Combine	0.830	0.760	0.920	0.600	0.697	0.882	18	10/18 (55.6%)
	Resnet-Axial	0.602	0.580	0.520	0.640	0.591	0.571	19	
	Resnet-Coronal	0.566	0.560	0.640	0.480	0.552	0.571	19	
	Resnet-Sagittal	0.570	0.540	0.560	0.520	0.538	0.542	18	
	Resnet-Combine	0.508	0.500	0.560	0.440	0.500	0.500	18	
	[Radiomics+Resnet]-Combine	0.755	0.720	0.720	0.720	0.720	0.720	19	10/19 (52.6%)
	Xception-Axial	0.506	0.480	0.240	0.720	0.462	0.486	19	
	Xception-Coronal	0.651	0.660	0.480	0.840	0.750	0.618	19	
	Xception-Sagittal	0.632	0.620	0.520	0.720	0.650	0.600	17	
	Xception-Combine	0.678	0.600	0.320	0.880	0.727	0.564	17	
	[Radiomics+Xception]-Combine	0.712	0.700	0.560	0.840	0.778	0.656	19	4/19 (21.1%)
	InceptionV3-Axial	0.630	0.620	0.840	0.400	0.583	0.714	17	
	InceptionV3-Coronal	0.610	0.560	0.440	0.680	0.579	0.548	17	
	InceptionV3-Sagittal	0.682	0.680	0.440	0.920	0.846	0.622	19	
	InceptionV3-Combine	0.651	0.580	0.440	0.720	0.611	0.563	19	
	[Radiomics+InceptionV3]-Combine	0.688	0.580	0.520	0.640	0.590	0.571	18	3/18 (16.7%)
	InceptionResnetV2-Axial	0.450	0.420	0.080	0.760	0.250	0.452	14	
	InceptionResnetV2-Coronal	0.624	0.540	0.160	0.920	0.667	0.523	17	
	InceptionResnetV2-Sagittal	0.718	0.540	0.120	0.960	0.750	0.522	18	
	InceptionResnetV2-Combine	0.706	0.600	0.280	0.920	0.778	0.561	12	
	[Radiomics+InceptionResnetV2]-Combine	0.827	0.700	0.800	0.600	0.667	0.750	17	9/17 (52.9%)
Random Forest	Radiomics-Axial	0.742	0.640	0.560	0.720	0.667	0.621	15	
	Radiomics-Coronal	0.753	0.560	0.880	0.240	0.537	0.667	15	
	Radiomics-Sagittal	0.710	0.720	0.600	0.840	0.789	0.677	14	
	Radiomics-Combine	0.822	0.740	0.800	0.680	0.714	0.773	11	
	VGG16-Axial	0.674	0.700	0.520	0.880	0.813	0.647	15	
	VGG16-Coronal	0.734	0.680	0.760	0.600	0.655	0.714	19	
	VGG16-Sagittal	0.602	0.580	0.720	0.440	0.563	0.611	10	
	VGG16-Combine	0.712	0.600	0.680	0.520	0.586	0.619	19	
	Radiomics+VGG16	0.898	0.800	0.840	0.760	0.778	0.826	17	15/17 (88.2%)
	VGG19-Axial	0.649	0.580	0.560	0.600	0.583	0.577	15	
	VGG19-Coronal	0.675	0.580	0.840	0.320	0.553	0.667	18	
	VGG19-Sagittal	0.660	0.660	0.800	0.520	0.625	0.722	12	
	VGG19-Combine	0.734	0.680	0.760	0.600	0.655	0.714	17	
	[Radiomics+VGG19]-Combine	0.802	0.780	0.800	0.760	0.769	0.792	16	16/16 (100%)
	Resnet-Axial	0.604	0.580	0.240	0.920	0.750	0.750	11	
	Resnet-Coronal	0.621	0.580	0.600	0.560	0.577	0.583	12	
	Resnet-Sagittal	0.609	0.600	0.640	0.560	0.593	0.593	16	
	Resnet-Combine	0.626	0.640	0.520	0.760	0.684	0.684	12	
	[Radiomics+Resnet]-Combine	0.872	0.820	0.840	0.800	0.808	0.833	13	12/13 (92.3%)
	Xception-Axial	0.597	0.580	0.480	0.680	0.600	0.567	19	
	Xception-Coronal	0.702	0.620	0.720	0.520	0.600	0.650	16	
	Xception-Sagittal	0.657	0.600	0.720	0.480	0.581	0.632	13	
	Xception-Combine	0.746	0.680	0.840	0.520	0.636	0.636	11	
	[Radiomics+Xception]-Combine	0.819	0.780	0.800	0.760	0.769	0.792	13	10/13 (76.9%)
	InceptionV3-Axial	0.569	0.540	0.600	0.480	0.536	0.545	14	
	InceptionV3-Coronal	0.600	0.560	0.560	0.560	0.560	0.560	16	
	InceptionV3-Sagittal	0.610	0.600	0.640	0.560	0.593	0.593	15	
	InceptionV3-Combine	0.617	0.620	0.600	0.640	0.625	0.625	15	
	[Radiomics+InceptionV3]-Combine	0.806	0.780	0.920	0.640	0.719	0.889	19	14/19 (73.7%)
	InceptionResnetV2-Axial	0.551	0.540	0.440	0.640	0.550	0.533	17	
	InceptionResnetV2-Coronal	0.659	0.620	0.560	0.680	0.636	0.607	15	
	InceptionResnetV2-Sagittal	0.618	0.520	0.160	0.880	0.571	0.512	17	
	InceptionResnetV2-Combine	0.695	0.660	0.760	0.560	0.633	0.700	15	
	[Radiomics+InceptionResnetV2]-Combine	0.851	0.840	0.880	0.800	0.815	0.870	19	14/19 (73.7%)

LR, logistic regression; SVM, support vector machine; NPV, negative predictive value.

https://dx.doi.org/10.21037/qims-21-722

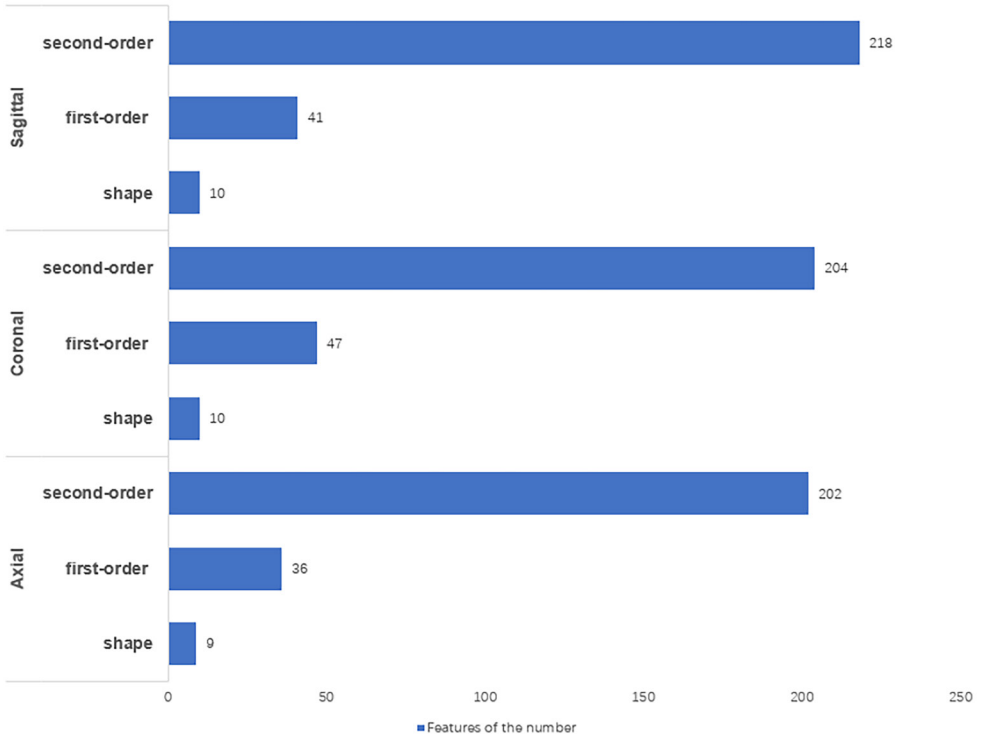

Figure S1 The remaining feature figures of axial, coronal, and sagittal images after removing redundant features by the Spearman correlation test.

References

47. Bengio Y, Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation. J Mach Learn Res 2004;5:1089-105.

https://dx.doi.org/10.21037/qims-21-722