Mathematical Biosciences and Engineering

MBE, 20(2): 2890–2907.
DOI: 10.3934/mbe.2023136
Received: 16 September 2022
Revised: 10 November 2022
Accepted: 13 November 2022
Published: 01 December 2022

http://www.aimspress.com/journal/MBE

Research article

Accurate prediction of glioma grades from radiomics using a multi-filter and multi-objective-based method

Jingren Niu[1,2], Qing Tan[1,2], Xiufen Zou[1,2] and Suoqin Jin[1,2,*]

[1] School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
[2] Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan 430072, China

*** Correspondence:** Email: sqjin@whu.edu.cn.

Abstract: Radiomics, providing quantitative data extracted from medical images, has emerged as a critical role in diagnosis and classification of diseases such as glioma. One main challenge is how to uncover key disease-relevant features from the large amount of extracted quantitative features. Many existing methods suffer from low accuracy or overfitting. We propose a new method, Multiple-Filter and Multi-Objective-based method (MFMO), to identify predictive and robust biomarkers for disease diagnosis and classification. This method combines a multi-filter feature extraction with a multi-objective optimization-based feature selection model, which identifies a small set of predictive radiomic biomarkers with less redundancy. Taking magnetic resonance imaging (MRI) images-based glioma grading as a case study, we identify 10 key radiomic biomarkers that can accurately distinguish low-grade glioma (LGG) from high-grade glioma (HGG) on both training and test datasets. Using these 10 signature features, the classification model reaches training Area Under the receiving operating characteristic Curve (AUC) of 0.96 and test AUC of 0.95, which shows superior performance over existing methods and previously identified biomarkers.

Keywords: radiomics; medical image analysis; glioma grades; biomarkers; multi-filters; multi-objective optimization

1. Introduction

As a common central nervous system (CNS) tumor, gliomas can be classified as low-grade gliomas (LGG) including grades I and II and high-grade gliomas (HGG) including grades III and IV

according to the World Health Organization (WHO) classification of CNS tumors based on the malignancy [1]. In the radiology reports, the classification of gliomas into LGG and HGG is an important standard of care [2–4]. Patients with different grades of gliomas require different surgical, radiotherapy, and chemotherapy plans [5]. Glioblastoma multiforme (GBM) is the most aggressive type of glioma (WHO grade IV), leading to a poor prognosis with a median survival time of around 15 months even after multimodal treatment [6]. Therefore, accurate prediction of LGG and HGG is crucial for improving treatment.

Medical imaging, in particular Magnetic Resonance Imaging (MRI), shows the characteristics of human cancer non-invasively and is widely used in clinical practice to aid diagnosis of brain tumor. By transforming medical images into quantitative data and extracting thousands of quantitative features as imaging markers, radiomics is becoming increasingly more important in analyzing tumor phenotypes [7–10]. It is increasing clear that radiomic features can greatly help to predict grades of glioma [11–13]. Basic MRI modalities available from any clinical scanner, including native T1-weighted pre-contrast (T1), T1-weighted post-contrast (T1-Gd), T2, and T2-FLAIR sequences, provide key details related to each part of the tumor. The combination of radiomic features derived from multiparametric MRI had good classification performance in glioma grading [14,15]. Cui et al. [16] compared prediction performance of features extracted from four modalities of MRI images (i.e., T1, T1-Gd, T2, and FLAIR) and found features derived from T1 and T1-Gd showed better prediction accuracy. Other studies either lacked prediction validation using external test datasets [17] or were not capable of uncovering radiomic biomarkers for glioma grading [14,18]. The large mass of quantitative features from complex clinical imaging arrays requires sophisticated machine learning algorithms to uncover their relevance with the tumor phenotypes. Therefore, identifying predictive and robust biomarkers and developing efficient models for glioma grading still remain challenging.

Image filters are important for feature extraction and further prediction performance. Recently, Gao et al. [19] applied six image filters in glioma MRI images to extract six categories of features, including Wavelet, Laplacian of Gaussian (LoG), Square, Square Root, Logarithm and Exponential features. By directly mixing all categories of features, a mathematical model was built for predicting the grades of glioma tumors. Different image filters highlight the information of the MRI images of glioma patients from different angles, and it is worth trying to analyze each category of filter features individually. In addition, existing computational models usually employed a single-objective optimization-based model for feature selection [16–20]. While single-objective-based feature selection can well control the number of selected features, it limits diversity of feature selection and easily leads to unreliable results.

To address these limitations, we developed a Multi-Filter-based feature extraction and Multi-Objective-based feature selection (MFMO) model to identify critical radiomic features for disease diagnosis and classification. By taking glioma grades as a case study, the multi-filter feature extraction was performed using eight different image filters in both T1 and T1-Gd MRI images. A novel multi-objective feature selection algorithm was developed to get robust features with high diversity from each category of the eight filters. A multi-step feature selection approach including the use of SHapley Additive exPlanations (SHAP) method was further employed to select important features with high contribution to the model prediction as the radiomic biomarkers. By applying to glioma images including 167 patients collected from the Cancer Imaging Archive (TCIA), 30 patients from MICCAI Brain Tumor Segmentation 2013 Challenge (BraTS 2013) and another 34 patients from

Ivy Glioblastoma Atlas (lvyGAP), we demonstrated the effectiveness of MFMO in distinguishing glioma grades over existing methods and revealing predictive and robust radiomic biomarkers.

2. Materials and methods

2.1. Framework of MFMO model

The proposed framework of MFMO included three major components: multi-filter-based feature extraction, multi-objective-based feature selection, and identification of important filters and features. We describe each of them in detail below.

2.1.1. Multi-filter-based feature extraction

We extracted radiomic features by considering the whole tumors (including peritumoral and intratumoral regions) as regions of interest (ROI). Among the four modalities including T1, T1-Gd, T2, FLAIR, we focused on T1 and T1-Gd modalities of MRI images and extracted multi-filter radiomic features from images in T1 and T1-Gd modalities. For MRI image of a patient in one modality, seven image filters were used to process the image. These seven image filters were Wavelet (including eight channels: HHH, HHL, HLH, HLL, LHH, LHL, LLH, and LLL), LoG (Laplacian of Gaussian, with kernel size being 2, 3, 4, and 5), Square, Square Root, Logarithm, Exponential, and Gradient. By including the original images that were not processed by any filter, we can extract eight categories of image filters from corresponding images.

Table 1. The number of features from each category of filters in T1 and T1-Gd modalities.

Image filters	# of features
Wavelet	688
LoG	344
Original	100
Exponential	86
Gradient	86
Logarithm	86
Square	86
SquareRoot	86
Total	1562

For a patient with MRI images in T1 and T1-Gd modalities, we finally got sixteen categories of filter features for a patient. Each category of features consisted of the shape, the first-order and the high-order texture features (including gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM) and gray level dependence matrix (GLDM) from ROI of the corresponding transformed MRI images. Table 1 showed the number of features in each category. In sum, we applied different image filters in MRI images and extracted radiomic features from transformed MRI images. The image filter and feature extraction were accomplished through an open-source Python package PyRadiomics 3.0 [21]. PyRadiomics assigns a value to each feature, which describes the distribution of intensities within the image region.

Definitions of image filters and features were publicly available at PyRadiomics website (https://pyradiomics.readthedocs.io/en/latest/features.html).

2.1.2. Multi-objective-based feature selection

Our multi-objective-based feature selection model aims at identifying a subset of features $S \subset X$, $X \in \mathbb{R}^{m \times n}$, which not only can accurately predict glioma grades, but also exhibit less redundancy. In the feature matrix X, m is the number of patients and n is the number of features. S is mainly obtained by solving the following multi-objective optimization problem:

$$\begin{cases} \max\limits_{S \subset X} : \frac{1}{|S|} \sum_{s_i \in S} NMI(s_i, c) & (1) \\ \min\limits_{S \subset X} : \frac{1}{|S|^2} \sum_{s_i, s_j \in S} NMI(s_i, s_j) & (2) \end{cases}$$

where $NMI(A, B)$ is the normalized mutual information between two random variables A and B, which is defined as [22]:

$$NMI(A, B) = 2 \frac{MI(A,B)}{H(A) + H(B)} \quad (3)$$

Given the entropy of A (denoted by $H(A)$), the mutual information $MI(A, B)$ is defined as:

$$MI(A, B) = H(A) + H(B) - H(A, B) \quad (4)$$

The NMI value ranges from 0 to 1, where the larger values represent the high relevance between the two variables A and B. Here, $s_i \in \mathbb{R}^{m \times 1}$ and $s_j \in \mathbb{R}^{m \times 1}$ are the calculated intensity value vectors of the i-th and j-th feature across m patients, and $|S|$ is the number of selected features in $S \in \mathbb{R}^{m \times |S|}$. s_i is the ith column of S. Therefore, $NMI(s_i, c)$ represents the relevance of the selected feature s_i with the prior label information of patients (e.g., the glioma grade c) and $NMI(s_i, s_j)$ represents the relevance of two selected features s_i and s_j. The first objective function Eq (1) maximizes the relevance of candidate features to the glioma grades and the second objective function Eq (2) minimizes redundancy among the selected features.

The above multi-objective optimization problem can be solved the classical Nondominated Sorting Genetic Algorithm-II (NSGAII) algorithm [23]. When running NSGAII, the size of the solution set was set to 100, and the maximum generation number was set to 50. For the mutation operator, the mutation probability was set to 0.2. The z-score-normalized values of features were used as inputs.

The multi-objective optimization model finds a Pareto front in the solution set and each solution on the front represents a combination of individual features, which exhibits a good balance the relevance and the redundancy. For each feature group on the Pareto front, we selected no more than ten features by utilizing a feature selection method named MIC-mRMR, which used maximal information coefficient (MIC) [24] as the correlation metric. Among all feature groups on the Pareto front, we further identified the optimal feature group with the highest training AUC on the training dataset, where a linear regression (LR) model was used for classifying glioma grades and the AUC was calculated by comparing with the true labels of patients in the training dataset.

2.1.3. Identification of radiomic biomarkers via a multi-step approach

To further reduce the number of radiomic features and identify the important and reliable radiomic biomarkers for glioma grading, we proposed a multi-step approach. First, for each category of the filter features, we applied the multi-objective-based feature selection method to obtain a set of ten features, and utilized a LR classifier to assess the performance of each filter in distinguishing between HGG and LGG. We retain the selected features from the top five filters with the highest training AUC. For each category of the filter features, we chose the number ten as the number of obtained features due to the following two reasons: 1) a small number of biomarkers are easier for their evaluation during diagnostic process and application as therapeutic targets; 2) previous studies also indicated that 10–20 features are likely sufficient for glioma diagnosis and classification [16,19]. In addition, we found that the performance of our model was relatively robust when varying the number of obtained features (Table S1).

Second, for the ten features selected from each category of the five filter features, we computed the SHAP values of each feature, and took the average of the absolute values of the SHAP values of each feature as a measurement of feature importance. We selected the top three features with the highest average values of the absolute values of the SHAP values within each category, leading to a total of 15 features across the five categories of filter features. Of note, the number three here is not fixed and can be changed. The model performance was relatively robust with respect to different number of features from each filter (Table S2). The average of the absolute value of the SHAP value of each feature is the average impact on output magnitude of each feature. In the following, we will refer the average of the absolute value of the SHAP value as SHAP value for short. The higher the value, the more important the feature is. SHAP value is shown to more effectually discriminate among model output classes [25] and can be used to explain a large variety of models including linear models (e.g., linear regression) and tree-based models (e.g., random forest), while other metrics can only be used to explain limited model types. Therefore, we choose the SHAP value as the measure of feature importance. In our study, SHAP [25] measures the contribution of each feature of the LR model and assigns each feature an importance value for the LR model prediction, which is useful for interpreting and improving model prediction.

Third, we built a LR classifier using these 15 features, and selected 10 important features with the highest SHAP value.

This multi-step feature filtering method was performed in training dataset. The first step selected the different filters with better performance to ensure the diversity of radiomic features, and the last two steps reduced the number of features to reduce the risk of overfitting.

2.2. Classification and evaluation

To assess the effectiveness of the selected features, we used the LR classifier to classify HGG and LGG patients and employed four metrics to evaluate the performance of the LR classifier, including the area under receiver operating characteristic (ROC) curve (AUC), Accuracy, Sensitivity and Specificity. Linear regression classifier was achieved by Scikit-learn [26]. Sensitivity represents the proportion of positive sample that are correctly identified as such while specificity represents the proportion of negative sample that are correctly classified.

2.3. Method comparisons

We compared the performance of our MFMO method against four previously proposed methods for predicting glioma grades [14,19,20,27]. Briefly, in Cho et al. [20], mRMR algorithm and RF model were used for feature selection and classification, respectively. In Gao et al. [19], PCA algorithm and RF model were used for feature selection and classification, respectively. In Chen et al. [14], partial least squares (PLS) regression and support vector machine (SVM) were used for reducing the number of the extracted features, and predicting glioma grades. In Kha et al. [27], two steps of features selection were performed, including Spearman's Correlation Coefficient (SCC) and the combination of XGBoost-SHAP analysis. For the datasets used in our study, we kept features with SCC greater than 0.8 and discarded features with SCC less than 0.8. XGBoost model was employed for predicting glioma grades. Because these methods did not provide source codes, we implemented the methods based on their description in the original paper. The best model and parameters described in the original papers were selected for comparison analysis.

2.4. Training and test datasets

The training datasets are the TCGA-LGG and TCGA-GBM Segmentation Datasets [28,29] from The Cancer Imaging Archive (TCIA), which are available from the public website (http://wiki.cancerimagingarchive.net/x/LIZyAQ, http://wiki.cancerimagingarchive.net/x/KoZyAQ). Computer-aided and manually corrected segmentation labels for the preoperative multi-institutional scans of 102 HGG and 65 LGG patients with the corresponding skull-stripped and co-registered multimodal MRI data were collected for training. The test dataset was collected from MICCAI Brain Tumor Segmentation 2013 Challenge (BraTS 2013) [30,31]. Of note, there is not any duplicated sample between the training and test datasets. The clinical image data of BraTS 2013 test dataset consists of multi-contrast MRI scans of 20 HGG and 10 LGG patients. The clinical image data of BraTS 2013 dataset and data providers were described in [30]. Briefly, the clinical data was acquired at four different centers—Bern University, Debrecen University, Heidelberg University, and Massachusetts General Hospital—over the course of several years, using MR scanners from different vendors and with different field strengths (1.5T and 3T) and implementations of the imaging sequences (e.g., 2D or 3D) [30]. Computer-aided segmentation was applied. Then the clinical images were manual annotated and segmented as HGG and LGG by experts. BraTS 2013 dataset was identical to the BraTS 2012 challenge with the omission of the synthetic dataset [32].

We also used Ivy Glioblastoma Atlas (lvyGAP) MRI dataset from TCIA as another test dataset [33], which is available from the public website (https://wiki.cancerimagingarchive.net/x/64MvB). However, IvyGAP only contains HGG images of 34 subjects segmented. We thus combined HGG images (34 patients) from lvyGAP and LGG images (10 patients) from BraTS 2013 dataset as a new test dataset. Similar to a previous study [34], we applied Synthetic Minority Over-sampling Technique (SMOTE) to the new test dataset, leading a new balanced dataset (34 HGG cases and 34 LGG cases). This new test dataset was denoted by IvyGAP-BraTS. We used two modalities of MRI images: T1 and T1-Gd. We used PyRadiomics for feature extraction, and conducted and reported radiomics analyses based on International Biomarker Standardization Initiative (IBSI) guidelines [35]. To be clearer, the IBSI checklist was provided in Table S3.

3. Results

3.1. Overview of MFMO for glioma grading

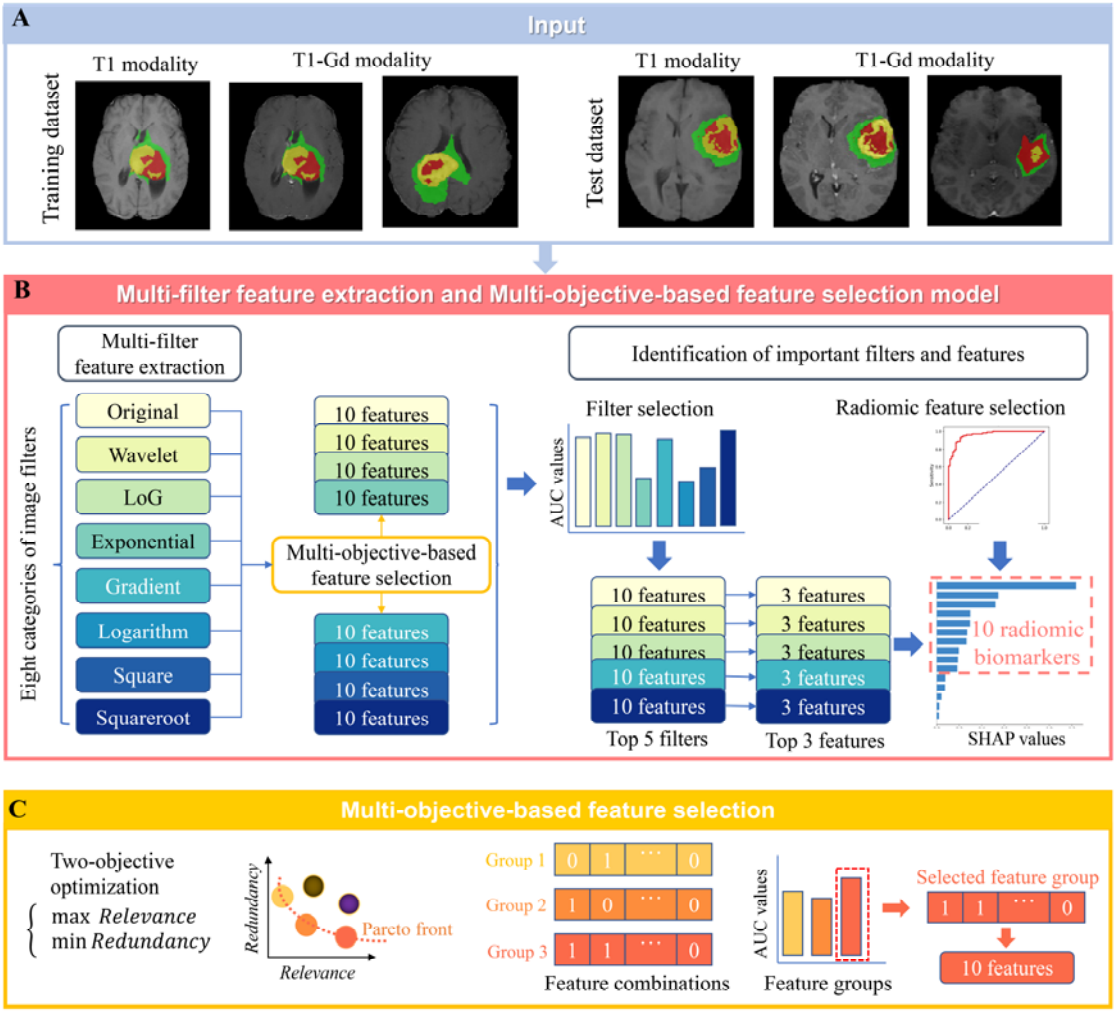

Figure 1. Framework of MFMO. (A) The MFMO model takes T1 and T1-Gd MRI images as inputs. Images are divided into training and test datasets. (B) Upon receiving inputs, MFMO applies multi-filter feature extractions in the images. For each category of the eight filter features, a multi-objective-based feature selection method is employed and the feature group with the highest AUC is selected. Top 10 features are selected for each filter. The top five filters are further selected based on the computed AUC of feature groups. The top three features from each category of the five filters are selected as candidate biomarkers for glioma grading. The final 10 radiomic biomarkers are determined based on the classification performance and the SHAP values of each feature. (C) The framework of the proposed multi-objective-based feature selection method.

To identify the important and robust radiomic features for glioma grading, MFMO took MRI images as inputs (Figure 1A) and first extracted multi-filter radiomic features using an open-source Python package PyRadiomics 3.0 [21]. Each filter represented one category of radiomic features and

contained tens to hundreds of radiomic features (Materials and Methods, Table 1). By default, eight filters were applied including the original images that were not processed by any filter, leading to a total of 1562 radiomic features across all filters. Second, for each category of the eight filters, a two-objective-based feature selection method was proposed and applied (Figure 1B,C). Based on the Pareto front of the two-objective optimization model, MFMO identified different combinations of individual features that not only accurately predicted glioma grades, but also exhibited less redundancy. For each filter, the feature group with the highest AUC based on a linear regression model was selected. Third, the number of filters and features were further reduced by using a multi-step approach, including selecting the top five filters with the highest AUC and the top three features with the highest SHAP value within each filter. The SHAP value of each feature quantified feature importance for classification (Materials and Methods). Finally, 10 important features were identified by computing the SHAP values. The overall framework of our proposed MFMO method was presented in Figure 1.

3.2. Identification of important filters for glioma grading

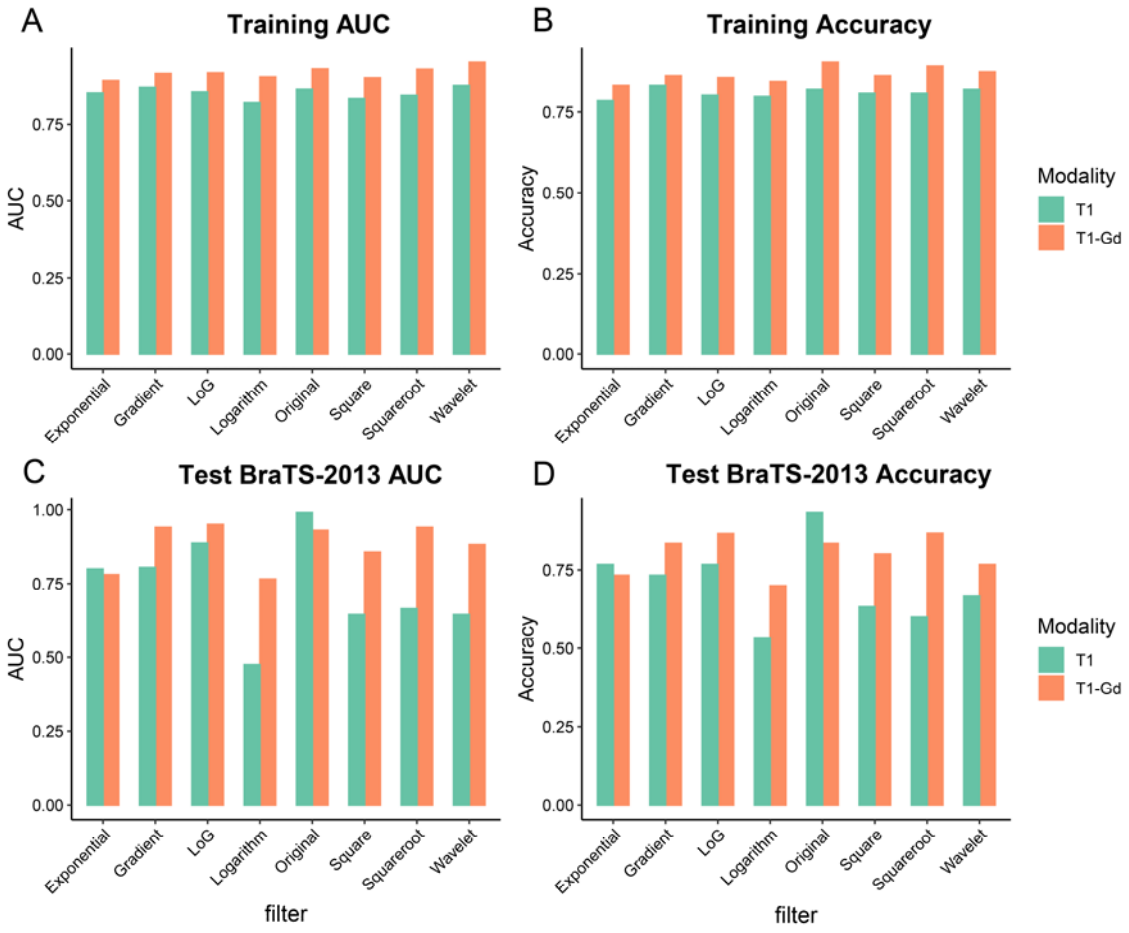

Figure 2. Prediction performance across eight categories of filter features in T1 and T1-Gd modalities. Two metrics were calculated. (A) AUC in the training dataset, (B) Accuracy in the training dataset, (C) AUC in the BraTS 2013 test dataset, (D) Accuracy in the BraTS 2013 test dataset.

For each of the two MRI modalities including T1-Gd and T1, we selected a set of ten features from each category of filter features by using the multi-objective-based feature selection method in MFMO, and then built a LR classifier to grade glioma. By computing AUC using the selected filter features from each modality in both training and test datasets, we observed that the eight sets of selected T1-Gd filter features had a training AUC of 0.918 and a BraTS 2013 test AUC of 0.882 on average, while those of selected T1 filter features had a training AUC of 0.851 and a BraTS 2013 test AUC of 0.739 on average (Figure 2A,C). By evaluating the performance using another metric called Accuracy, we also observed an overall higher Accuracy value in the T1-Gd modality compared to T1 modality (Figure 2B,D). These results showed that radiomic features extracted from T1-Gd MRI images generally had better classification performance than those from T1 MRI images. In the subsequent analysis, we focused on the T1-Gd MRI images.

We next identified the important filters based on the T1-Gd modality. Based on the calculated AUC values in the training phase, the top five categories of filter features were Wavelet, Original, SquareRoot, LoG and Gradient. The training AUCs of these five categories of selected features were 0.952, 0.930, 0.929. 0.917, and 0.915, and the BraTS 2013 test AUCs of these selected features were 0.880, 0.930, 0.940, 0.950, and 0.950, respectively.

These five categories of selected filter features had a training AUC of 0.928 and a BraTS 2013 test AUC of 0.93 on average. These results not only indicated that the feature selection algorithm can accurately predict the glioma grades, but also suggested that our approach was not likely overfitted and underfitted due to the comparable prediction performance across training and test datasets.

3.3. MFMO identified previously unrecognized radiomic features for glioma grading

Table 2. The selected 10 signature features.

Filter	Feature class	Feature name	Short feature name
Original	GLSZM	SmallAreaEmphasis	SAE
Original	GLSZM	SizeZoneNonUniformityNormalized	SZNUN
Gradient	GLRLM	RunLengthNonUniformity	RLNU
SquareRoot	GLDM	DependenceEntropy	DE
Gradient	GLSZM	GrayLevelNonUniformity	GLNU
Wavelet-HLH	GLCM	Idm	IDM
SquareRoot	GLSZM	ZoneEntropy	ZE
Wavelet-HLL	first-order	RootMeanSquared	RMS
SquareRoot	shape	SmallAreaEmphasis	SAE
LoG	GLDM	DependenceNonUniformity	DNU

Using the multi-step feature filtering method (Materials and Methods), we obtained 10 important radiomic features across the five categories of filter features (Table 2). Using the combination of all 10 features for predicting glioma grades via LR classifier, the AUC is 0.96, 0.95, and 0.955 on the training, BraTS-2013 test and lvyGAP-BraTS test datasets, respectively (Table 3; Figure 3A,B). The high accuracy, sensitivity and specificity were also achieved by using these 10 features in the training and test datasets (Figure 3A,B; Table 3). The consistent high AUC values

between the training and test datasets suggested that the 10 features selected by our model were stable and reliable in classifying LGG and HGG.

Table 3. Performance of 10 signature features in the training and test datasets.

Dataset	AUC	Accuracy	Specificity	Sensitivity
Training	0.962	0.910	0.877	0.931
Test BraTS 2013	0.95	0.867	1	0.75
Test lvyGAP-BraTS	0.955	0.887	0.941	0.912

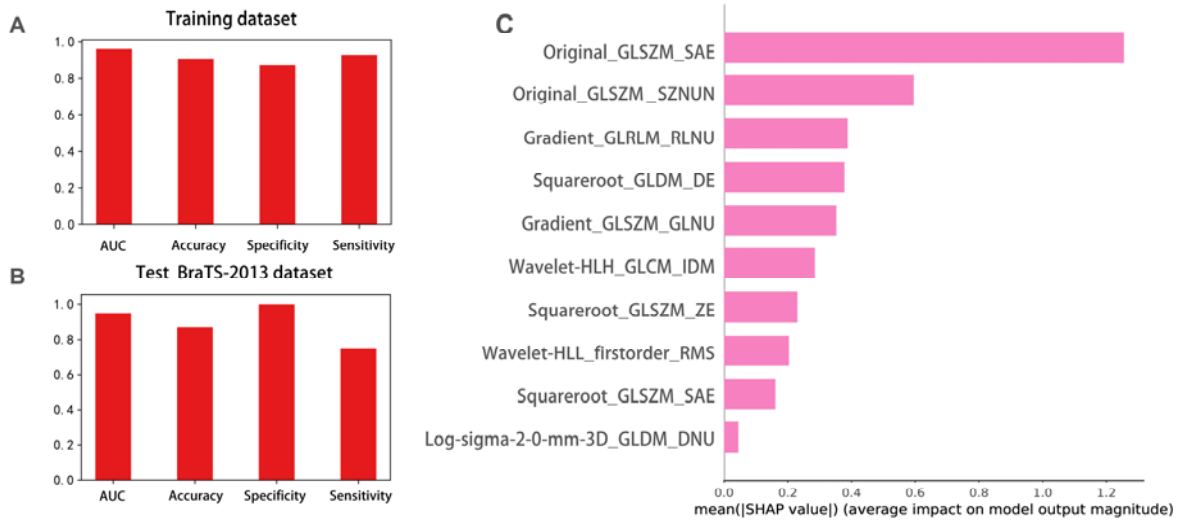

Figure 3. Comparison of prediction performance across different classifiers. (A) Performance of 10 signature features in the training dataset. (B) Performance of 10 signature features in the BraTS 2013 test dataset. (C) Contribution of each of the 10 features to the model prediction. The contribution was quantified using the mean of absolute values of SHAP values of each feature.

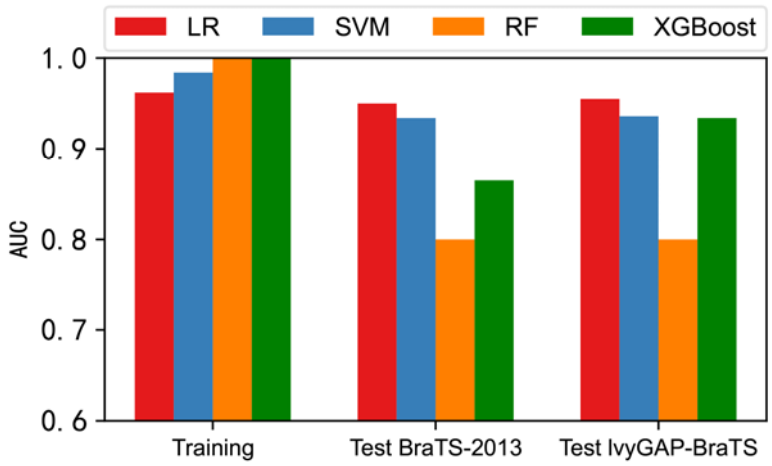

Figure 4. Comparison of prediction performance across different classifiers.

Moreover, we compared LR classifier with other classifiers including Support Vector Machine (SVM), random forest (RF) and XGBoost based on the 10 signature features identified by MFMO. We found that SVM, RF and XGBoost had higher training AUC but lower test AUCs on BraTS 2013 and lvyGAP-BraTS test dataset compared to LR (Figure 4), which indicates that these classifiers had higher risk of overfitting. These results demonstrate that although LR classifier is simple, it has an advantage of reducing overfitting while achieving relatively high prediction accuracy. This observation is also consistent with the well-known knowledge that overfitting often occurs when the model is too complex.

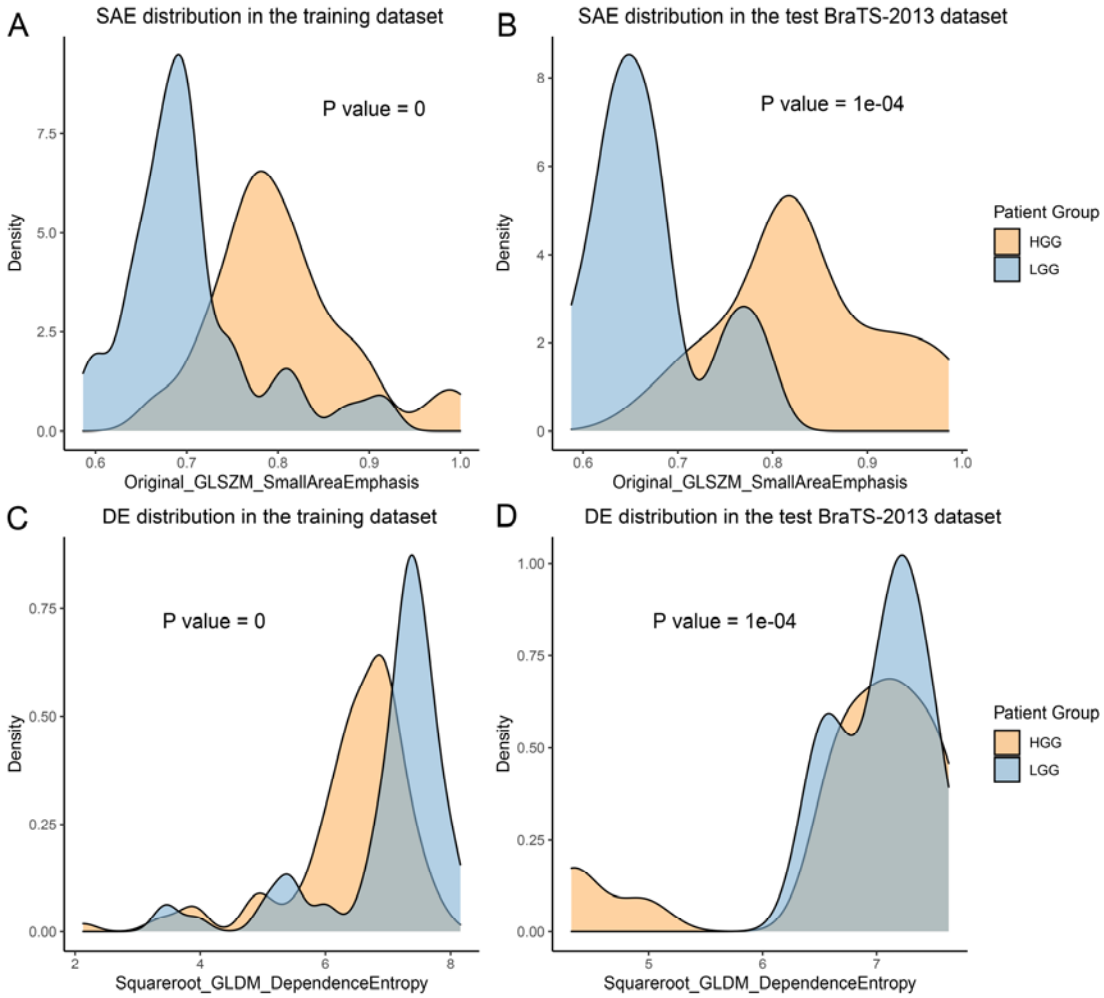

Figure 5. Distribution of image intensity values of the two selected features in HGG and LGG groups. The feature SAE quantifies the distribution of small size zones and DE measures gray level dependencies. (A) SAE feature in the training dataset, (B) SAE feature in the BraTS 2013 test dataset, (C) DE feature in the training dataset, (D) DE feature in the BraTS 2013 test dataset. P-values from Kolmogorov-Smirnov test were indicated.

We then assessed the contribution of each of the 10 features to the LR model prediction, which was measured by a SHAP value. The DependenceEntropy (DE) feature from the SquareRoot filter had relatively high SHAP values (Figure 3C), suggesting the importance in distinguishing HGG from LGG patients. DE measures the uncertainty and randomness of gray level dependencies of the

images. The smaller the DE value, the stronger the dependencies of the gray levels SAE. SZNUN and DE describe the ROI heterogeneity from connection density, variability and dependencies of gray level size zones perspectives, respectively, suggesting the importance of gray levels in quantifying heterogeneity. Among the 10 signature features, five of them were only found by our method [19]: Gradient_GLRLM_RLNU, SquareRoot_GLDM_DE, Wavelet-HL_GLCM_IDM, Wavelet-HLL_first-order_RMS and LoG-sigma-2-0-mm-3D_GLDM_DNU.

For the common signature Original_GLDM_SAE, we examined the distribution of the calculated intensity values from the T1-Gd MRI images, and observed significantly higher SAE values in the HGG patients compared to LGG patients in both training and test datasets (Figure 5A,B; Kolmogorov-Smirnov test). For the specific signature identified by our method such as SquareRoot_GLDM_DE, we also observed significant differences in the distributions between HGG and LGG patients in both training and test datasets (Figure 5C,D). However, smaller DE values were observed in the HGG patients compared to LGG patients. These results showed that SAE and DE feature was indeed effective in distinguishing HGG and LGG.

3.4. MFMO showed superior performance over existing methods and previously reported biomarkers

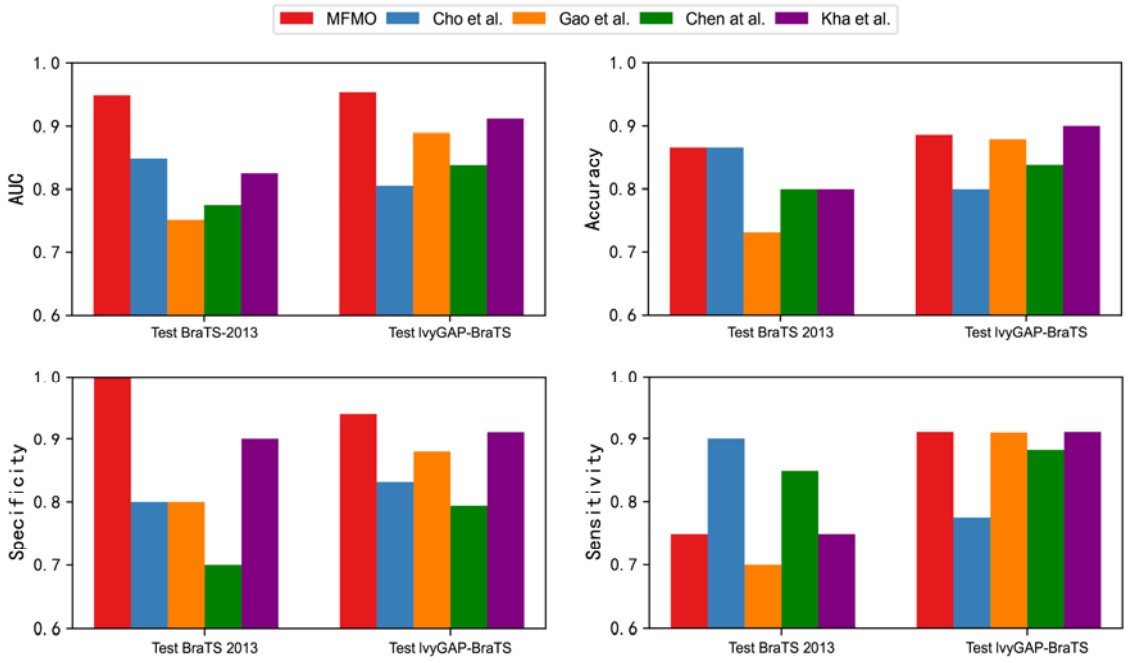

Figure 6. Comparison of MFMO against other methods in BraTS 2013 and IvyGAP-BraTS test datasets.

We next compared our MFMO method against four previously machine learning methods for predicting glioma grades, including Cho et al. [20], Gao et al. [19], Chen et al. [14], and Kha et al. [27]. Although Cho et al. [20] also considered both the relevance of candidate features to the glioma grades and the redundancy among the selected features, they performed feature selection using a single-objective model. After training these methods on the training dataset, we compared their performance in two test datasets. On the BraTS 2013 and lvyGAP-BraTS test datasets, MFMO reached the AUC of 0.95 and 0.955 respectively, which was higher than other methods (Figure 6). In

terms of Accuracy, MFMO exhibited better results than Gao et al and Chen et al, and comparable results with Cho et al on the BraTS 2013 test dataset and with Kha et al on the lvyGAP-BraTS test dataset. In addition, MFMO exhibited highest Specificity values and moderate Sensitivity. The high Specificity suggested the good ability of MFMO in identifying LGG patients. Together, these results showed the overall better performance of our proposed method MFMO in glioma grading.

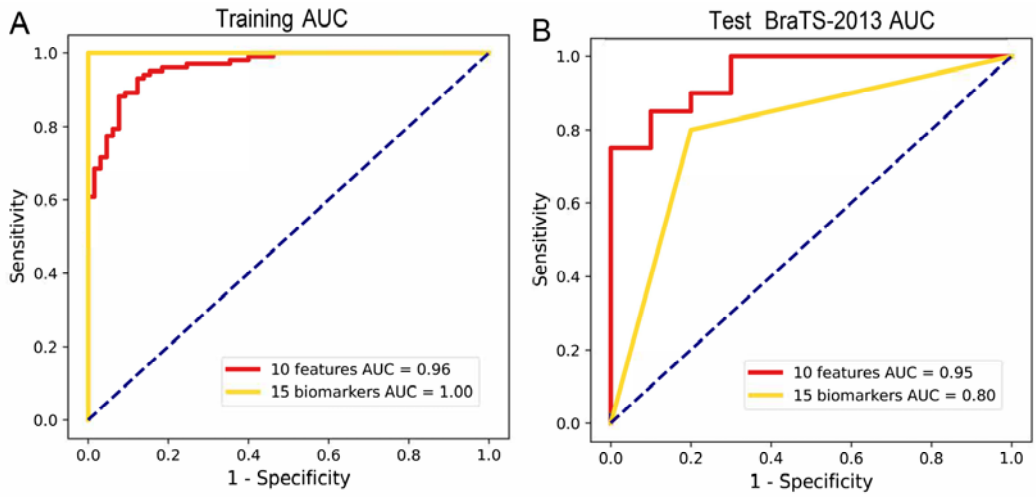

Figure 7. Comparison of prediction performance using the 10 features identified using MFMO and the previously reported 15 features. ROC and its associated AUC in (A) the training dataset and (B) the BraTS 2013 test dataset.

Gao et al. [19] identified 15 important features by using multiple-filter features as biomarker in glioma grading and found RF classifier had the best performance. We utilized these 15 biomarkers and built the RF classifier to predict grades of glioma in our datasets. Figure 7 showed the performance comparison between the RF classifier built using the identified 15 biomarkers and the LR classifier built using our proposed 10 signature features. We found that RF classifier built by 15 biomarkers had the risk of overfitting, where the training dataset had AUC of 1.00 and the BraTS 2013 test dataset had AUC of 0.80. On the contrast, the 10 signature features selected by our model achieved AUC of 0.96 on the training dataset and AUC of 0.95 on the BraTS 2013 test dataset, suggesting that our model achieved more robust and better classification performance in a smaller feature number. The identified 10 signature features can be considered as important biomarkers in classifying LGG and HGG.

4. Discussion

In this study, we applied several different filters (Wavelet, LoG, Square, Square Root, Logarithm, Exponential, and Gradient) in T1 and T1-Gd MRI images and extracted more diversified radiomic features. Different image filters highlight the information of the MRI images of glioma patients from different angles. Simple mixing of different filter features might cause information confusion and reduce the sensitivity of image filters to images. We analyzed radiomics features from different modalities MRI images and image filters individually and provided evidence for the effectiveness of different filters in T1 and T1-Gd MRI modalities. By exploring different combinations of the four modalities (T1, T1-Gd, T2, FLAIR), Cui et al. [16] found that while the

addition of more modalities increased the prediction accuracy overall, T1-Gd alone was as good as most combinations of modalities, suggesting that T1-Gd contains some information that is more informative in the classification task. Cui et al. also pointed out the potential disadvantages of adding more data, which could increase computational cost, introduces unknown sources of error, and increases patient burden both financially and timewise. Therefore, we hope to find candidate biomarkers from one modality that can also achieve relatively good prediction accuracy. In agreement with the findings from Cui et al., we observed that T1-Gd had better classification performance than T1 modality, and thus focused on T1-Gd filters to select candidate signature features.

Most of existing studies employed single-objective feature selection algorithm [18–20]. On contrast, we consider feature selection as a multi-objective optimization problem with the two objectives of maximizing the relevance of candidate features to the glioma grades and minimizing redundancy among the selected features. We applied multi-objective genetic algorithm NSGAII to solve this optimization model. Previous study by Lam et al. [34] also showed the advantages of genetic algorithm in glioma radiomic feature selection. Different from single-objective feature selection algorithm, multi-objective-based feature selection can find diversified feature groups on the Pareto front. Given that NSGA-II is able to find much better spread of solutions and better convergence near the true Pareto-optimal front [36], we combined the NMI and NSGAII to search for excellent feature groups which had good prediction performance and employed MIC-mRMR to restrict the number of selected features in each feature group. By using NMI-NSGAII and MIC-mRMR simultaneously, we bring the advantages of the multi-objective and single-objective feature selection algorithm together to better select radiomic features.

Particularly, the two GLSZM features extracted from original T1-Gd images, including Small Area Emphasis (SAE) and Size-Zone Non-Uniformity Normalized (SZNUN) showed the highest contributions to the model prediction. These two categories of features have been were found to have a good discrimination ability in glioma grading when using the Exponential filter [19] and our study indicated that SAE and SZNUN extracted from original T1-Gd images still had the good discrimination ability. GLSZM quantifies gray level zones in the image. The more uniform the image texture, the larger the GLSZM matrix width and the flatter. SAE quantifies the distribution of small-sized areas and SZNN measures the variability of the zone size volume across the image. Moreover, our method also identified other previously unrecognized features, including Gradient_GLRLM_RLNU, SquareRoot_GLDM_DE, Wavelet-HL_GLCM_IDM, Wavelet-HLL_first-order_RMS and LoG-sigma-2-0-mm-3D_GLDM_DNU. Together, our proposed MFMO method shows its great potential in searching for informative biomarkers and offers novel insights into the glioma grading, including signature features that were not previously identified.

In this study, we employed LR classifier for feature selection in order to classify HGG and LGG. Interestingly, we found that SVM, RF and XGBoost have higher training AUC and lower test AUC compared to LR, suggesting the potential higher overfitting risk of SVM, RF and XGBoost (Figure 4). These observations are consistent with the well-known knowledge that overfitting often occurs when the model is too complex. The better performance of LR is also likely due to the fact that the distinction between HGG and LGG is a binary classification problem. In this scenario, it might be more appropriate to use a simple model, such as LR. However, when it comes to the multi-classification problem of distinguishing glioma subtypes, complex models such as RF and XGBoost are likely more effective, as observed in previous studies [34,37]. Overfitting is a common problem in machine learning methods. In addition to the simple LR model we used, the

aforementioned multi-filter-based feature extraction and multi-objective-based feature selection are likely helpful for reducing overfitting. These strategies lead to the observation that our method produced comparable AUC values between training and test datasets. Other strategies such as cross-validation during the training process may also reduce overfitting, which had been employed in previous studies [27].

Although our MFMO method exhibits its superior performance in predicting radiomic biomarkers for classifying HGG and LGG, as mentioned by the previous study [34], the final diagnosis is a combination of clinical symptoms, medical images and pathological results. Le et al. [37] and Lam et al. [34] addressed an important question on the subtype classification of glioblastoma and low-grade gliomas from MRI, which has not well addressed by other existing studies including our work. Kha et al. [27] proposed XGBoost-based model to identify the 1p/19q codeletion status in LGG-diagnosed patients, which addressed the drawbacks of invasive gold-standard tests in clinical practice. The further extension of our work could focus on these important aspects. Further research could also correlate prognostic gene signatures of glioma with radiomic features for better prediction.

5. Conclusions

In conclusion, our proposed MFMO method that combined multi-filter radiomic features with multi-objective optimization model exhibits good performance in glioma grading and has the potential to advance clinical decision-making by systematically analyzing medical images. The multi-filter feature extraction enables comprehensive characterizations of MRI images and the multi-objective optimization model allows the identification of important features with high diversity and prediction power.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant (No. 11831015) and the Tian Yuan Mathematical Foundation (No.12126355). The datasets used in this study was publicly available. The training datasets are from Cancer Imaging Archive (TCIA) [28,29]. The test dataset was from MICCAI Brain Tumor Segmentation 2013 Challenge (BraTS 2013) [30,31] and Ivy Glioblastoma Atlas (lvyGAP) [33]. Our source codes are freely available at https://github.com/xiaoxiaozhihuishu/MFMO. SJ and XZ conceived the project. JN and QT conducted the research. JN, SJ and XZ wrote the paper. SJ and XF supervised the research. All authors approved the final manuscript.

Conflict of interest

The authors declare no competing interests.

References

1. P. Y. Wen, J. T. Huse, 2016 World Health Organization classification of central nervous system tumors, *Continuum (Lifelong Learn. Neurol.)*, **23** (2017), 1531–1547. https://doi.org/10.1212/con.0000000000000536

2. S. C. Thust, S. Heiland, A. Falini, H. R. Jager, A. D. Waldman, P. C. Sundgren, et al., Glioma imaging in Europe: A survey of 220 centres and recommendations for best clinical practice, *Eur. Radiol.*, **28** (2018), 3306–3317. https://doi.org/10.1007/s00330-018-5314-5
3. C. Gui, J. C. Lau, S. E. Kosteniuk, D. H. Lee, J. F. Megyesi, Radiology reporting of low-grade glioma growth underestimates tumor expansion, *Acta Neurochir.*, **161** (2019), 569–576. https://doi.org/10.1007/s00701-018-03783-3
4. A. Bink, J. Benner, J. Reinhardt, A. De Vere-Tyndall, B. Stieltjes, N. Hainc, et al., Structured reporting in neuroradiology: Intracranial tumors, *Front. Neurol.*, **9** (2018), 32. https://doi.org/10.3389/fneur.2018.00032
5. L. B. Nabors, J. Portnow, M. Ahluwalia, J. Baehring, H. Brem, S. Brem, et al., Central nervous system cancers version 3.2020, NCCN clinical practice guidelines in oncology, *JNCCN*, **18** (2020), 1537–1570. https://doi.org/10.6004/jnccn.2020.0052
6. Z. Y. Lin, R. W. Yang, K. S. Li, G. Z. Yi, Z. Y. Li, J. L. Guo, et al., Establishment of age group classification for risk stratification in glioma patients, *BMC Neurol.*, **20** (2020), 310. https://doi.org/10.1186/s12883-020-01888-w
7. P. Grossmann, O. Stringfield, N. El-Hachem, M. M. Bui, E. R. Velazquez, C. Parmar, et al., Defining the biological basis of radiomic phenotypes in lung cancer, *ELife*, **6** (2017), e23421. https://doi.org/10.7554/elife.23421
8. H. H. Cho, C. K. Kim, H. Park, Overview of radiomics in prostate imaging and future directions, *Br. J. Radiol.*, **95** (2022), 1131. https://doi.org/10.1259/bjr.20210539
9. J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, B. Baessler, Radiomics in medical imaging-"how-to" guide and critical reflection, *Insights Imaging*, **11** (2020), 91. https://doi.org/10.1186/s13244-020-00887-2
10. P. Lambin, R. T. H. Leijenaar, T. M. Deist, J. Peerlings, E. E. C. de Jong, J. van Timmeren, et al., Radiomics: the bridge between medical imaging and personalized medicine, *Nat. Rev. Clin. Oncol.*, **14** (2017), 749–762. https://doi.org/10.1038/nrclinonc.2017.141
11. T. Asklund, R. Birgander, P. Brynolfsson, A. Garpebring, J. Hauksson, R. Henriksson, et al., ADC texture—An imaging biomarker for high grade glioma, *Neuro-Oncol.*, **16** (2014), ii78. https://doi.org/10.1093/neuonc/nou174.299
12. H. H. Cho, H. Park, Classification of low-grade and high-grade glioma using multi-modal image radiomics features, in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2017), 3081–3084. https://doi.org/10.1109/embc.2017.8037508
13. H. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, *Nat. Commun.*, **5** (2014), 4006. https://doi.org/10.1038/ncomms5006
14. Q. J. Chen, L. H. Wang, L. Wang, Z. Y. Deng, J. Zhang, Y. M. Zhu, Glioma grade prediction using wavelet scattering-based radiomics, *IEEE Access*, **8** (2020), 106564–106575. https://doi.org/10.1109/ACCESS.2020.3000895
15. Y. W. Park, Y. S. Choi, S. S. Ahn, J. H. Chang, S. H. Kim, S. K. Lee, Radiomics MRI phenotyping with machine learning to predict the grade of lower-grade gliomas: A study focused on nonenhancing tumors, *Korean J. Radiol.*, **20** (2019), 1381–1389. https://doi.org/10.3348/kjr.2018.0814

16. G. Cui, J. J. Jeong, B. Press, Y. Lei, H. K. G. Shu, T. X. Liu, et al., Machine-learning-based classification of lower-grade gliomas and high-grade gliomas using radiomic features in multi-parametric MRI, preprint, arXiv:1911.10145.
17. K. L. C. Hsieh, C. Y. Chen, C. M. Lo, Quantitative glioma grading using transformed gray-scale invariant textures of MRI, *Comput. Biol. Med.*, **83** (2017), 102–108. https://doi.org/10.1016/j.compbiomed.2017.02.012
18. T. H. Xiao, W. Q. Hua, C. Li, S. S. Wang, Glioma grading prediction by exploring radiomics and deep learning features, in *Proceedings of the Third International Symposium on Image Computing and Digital Medicine*, (2019), 208–213. https://doi.org/10.1145/3364836.3364877
19. M. Gao, S. Y. Huang, X. Q. Pan, X. Liao, R. Yang, J. Liu, Machine learning-based radiomics predicting tumor grades and expression of multiple pathologic biomarkers in gliomas, *Front. Oncol.*, **10** (2020), 1676. https://doi.org/10.3389/fonc.2020.01676
20. H. H. Cho, S. H. Lee, J. Kim, H. Park, Classification of the glioma grading using radiomics analysis, *PeerJ*, **6** (2018), e5982. https://doi.org/10.7717/peerj.5982
21. J. J. M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, et al., Computational radiomics system to decode the radiographic phenotype, *Cancer Res.*, **77** (2017), e104–e107. https://doi.org/10.1158/0008-5472.can-17-0339
22. A. Lancichinetti, S. Fortunato, J. Kertész, Detecting the overlapping and hierarchical community structure in complex networks, *New J. Phys.*, **11** (2009), 033015. https://dx.doi.org/10.1088/1367-2630/11/3/033015
23. K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.*, **6** (2002), 182–197. https://dx.doi.org/10.1109/4235.996017
24. D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, et al., Detecting novel associations in large data sets, *Science*, **334** (2011), 6062. https://doi.org/10.1126/science.1205438
25. S. M. Lundberg, S. I. Lee, A unified approach to interpreting model predictions, in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, (2017), 4768–4777. https://dl.acm.org/doi/10.5555/3295222.3295230
26. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, **12** (2011), 2825–2830. https://dl.acm.org/doi/10.5555/1953048.2078195
27. Q. H. Kha, V. H. Le, T. N. K. Hung, N. Q. K. Le, Development and validation of an efficient MRI radiomics signature for improving the predictive performance of 1p/19q Co-deletion in lower-grade gliomas, *Cancers*, **13** (2021), 5398. https://doi.org/10.3390/cancers13215398
28. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, et al., Data descriptor: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features, *Sci. Data*, **4** (2017), 170117. https://doi.org/10.1038/sdata.2017.117
29. K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, et al., The cancer imaging archive (TCIA): Maintaining and operating a public information repository, *J. Digital Imaging*, **26** (2013), 1045–1057. https://doi.org/10.1007/s10278-013-9622-7
30. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., The multimodal brain tumor image segmentation benchmark (BRATS), *IEEE Trans. Med. Imaging*, **34** (2015), 1993–2024. https://doi.org/10.1109/tmi.2014.2377694

31. M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, P. Buchler, The virtual skeleton database: An open access repository for biomedical research and collaboration, *J. Med. Int. Res.*, **15** (2013), e245. https://doi.org/10.2196/jmir.2930
32. M. Ghaffari, A. Sowmya, R. Oliver, Automated brain tumor segmentation using multimodal brain scans: A survey based on models submitted to the BraTS 2012–2018 Challenges, *IEEE Rev. Biomed. Eng.*, **13** (2020), 156–168. https://doi.org/10.1109/rbme.2019.2946868
33. S. Pati, R. Verma, H. Akbari, M. Bilello, V. B. Hill, C. Sako, et al., Reproducibility analysis of multi-institutional paired expert annotations and radiomic features of the Ivy Glioblastoma Atlas Project (Ivy GAP) dataset, *Med. Phys.*, **47** (2020), 6039–6052. https://doi.org/10.1002/mp.14556
34. L. H. T. Lam, D. T. Do, D. T. N. Diep, D. L. N. Nguyet, Q. D. Truong, T. T. Tri, et al., Molecular subtype classification of low-grade gliomas using magnetic resonance imaging-based radiomics and machine learning, *NMR Biomed.*, **35** (2022), e4792. https://doi.org/10.1002/nbm.4792
35. A. Zwanenburg, M. Vallieres, M. A. Abdalah, H. Aerts, V. Andrearczyk, A. Apte, et al., The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping, *Radiology*, **295** (2020), 328–338. https://doi.org/10.1148/radiol.2020191145
36. T. M. Hamdani, J. M. Won, A. M. Alimi, F. Karray, Multi-objective feature selection with NSGA II, in *International Conference on Adaptive and Natural Computing Algorithms*, (2007), 240–247. https://doi.org/10.1007/978-3-540-71618-1_27
37. N. Q. K. Le, T. N. K. Hung, D. T. Do, L. H. T. Lam, L. H. Dang, T. T. Huynh, Radiomics-based machine learning model for efficiently classifying transcriptome subtypes in glioblastoma patients from MRI, *Comput. Biol. Med.*, **132** (2021), 104320. https://doi.org/10.1016/j.compbiomed.2021.104320

AIMS Press