IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 19, 2019, accepted July 7, 2019, date of publication July 15, 2019, date of current version August 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928975

Comparison of Feature Selection Methods and Machine Learning Classifiers for Radiomics Analysis in Glioma Grading

PAN SUN[1], (Student Member, IEEE), DEFENG WANG[2], VINCENT CT MOK[1,3], AND LIN SHI[4,5]

[1]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong
[2]School of Instrumentation Science and Opto-electronics Engineering, Beihang University, Beijing 100191, China
[3]Institute of Innovative Medicine, The Chinese University of Hong Kong, Hong Kong
[4]Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Hong Kong
[5]BrainNow Research Institute, Shenzhen 518081, China

Corresponding authors: Defeng Wang (dfwang@buaa.edu.cn) and Lin Shi (shilin@cuhk.edu.hk)

This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project CUHK14204117, in part by the grants from the Innovation and Technology Commission under Project GHP-025-17SZ and Project GHP-028-14SZ, in part by the Scheme Double First Class Program, in part by the Ministry of Education, in part by the Ministry of Finance, and in part by the National Development and Reform Commission, China.

ABSTRACT Radiomics-based researches have shown predictive abilities with machine-learning approaches. However, it is still unknown whether different radiomics strategies affect the prediction performance. The aim of this study was to compare the prediction performance of frequently utilized radiomics feature selection and classification methods in glioma grading. Quantitative radiomics features were extracted from tumor regions in 210 Glioblastoma (GBM) and 75 low-grade glioma (LGG) MRI subjects. Then, the diagnostic performance of sixteen feature selection and fifteen classification methods were evaluated by using two different test modes: ten-fold cross-validation and percentage split. Balanced accuracy and area under the curve (AUC) of the receiver operating characteristic were used to evaluate prediction performance. In addition, the roles of the number of selected features, feature type, MRI modality, and tumor sub-region were compared to optimize the radiomics-based prediction. The results indicated that the combination of feature selection method L^1-based linear support vector machine (L^1-SVM) and classifier multi-layer perceptron (MLPC) achieved the best performance in the differentiation of GBM and LGG in both ten-fold cross validation (balanced accuracy:0.944, AUC:0.986) and percentage split (balanced accuracy:0.953, AUC:0.981). For radiomics feature extraction, the enhancing tumor region (ET) combined with necrotic and non-enhancing tumor (NCR/NET) regions in T1 post-contrast (T1-Gd) modality provided more considerable tumor-related phenotypes than other combinations of tumor region and MRI modality. Our comparative investigation indicated that both feature selection methods and machine learning classifiers affected the predictive performance in glioma grading. Also, the cross-combination strategy for comparison of radiomics feature selection and classification methods provided a way of searching optimal machine learning model for future radiomics-based prediction.

INDEX TERMS Glioma grade, machine learning, feature classification, feature selection, radiomics.

I. INTRODUCTION

Glioma is the most common primary intracranial tumor in adults [1]. It might occur anywhere in brain and appear highly spatial-temporal heterogeneity. According to WHO, glioma can be classified into grades I-IV based on histologically malignant behavior [2]. Low-grade glioma (LGG, grades I and II) patients typically have more than five years survival whereas only 3-5% of glioblastoma (GBM, grade IV) patients survive more than five years, with median survival about 12 months [3]. GBM is the most common glioma histology type, accounting for 70% of primary brain tumors. Preoperative glioma grading, particularly the differentiation between GBM and LGG, is of great importance for making diagnostic decisions in clinical [4].

MRI provides a way of grading glioma with high spatial resolution and unique contrast between brain tissues and tumor non-invasively [5], [6]. With a comprehensive

The associate editor coordinating the review of this manuscript and approving it for publication was Yi Zhang.

view of brain structure and tumor, the highly heterogeneity degree in the histological tumor sub-regions are revealed. Then quantitative analysis can be performed in tumor regions of interest (ROIs) to find the relationship between tumor characteristics and clinical diagnosis. Radiomics is one of the analysis methodologies which converts imaging data into high-throughput mineable features [7], [8]. It is based on the hypothesis that these image-based features could capture phenotypic differences of tumor and be potentially used as diagnostic features for clinical outcomes. Radiomics provides a non-invasive way of exploring the relationship between glioma and image-based descriptors such as tumor appearance, shape, size, intensity, position and texture [9]. Recent quantitative radiomics-based diagnostic models have shown highly potential clinical values in predicting glioma grades [10], gene expression patterns [11] and gene mutation [12].

In current radiomics-based MRI glioma grading, various radiomics features, feature selection methods and classification or regression models are employed, which demonstrates identical diagnostic performance. For the extracted radiomics feature types, histogram-based features [4], shape-features [13], texture features [14]–[16], wavelet features [17], are designed. Subsequently, feature selection strategies, e.g. filter-based [18] and embedded-based methods [19], [20], are applied to identify the valuable features for grading glioma; and then machine-learning classifiers, e.g. random forest [21], logistic regression [13], support vector machine [22] are utilized to classify glioma histological types. Besides, multiparametric MRI sequences are employed to extract imaging features for grading glioma [23], [24]. Most radiomics-based researches have shown predictive abilities with multitudinous machine-learning approaches [9], [25]. Only a few recent studies have compared diagnostic performance by different radiomics feature selection and classification models [26]–[31]. However, it is still unknown whether different feature selection and classification methods affects radiomics-based prediction performance in glioma grading. In these regards, extracting effective imaging features and engaging reliable machine-learning strategies are desired to compare in grading glioma.

In this study, we investigated the diagnostic value of frequently used machine-learning approaches as well as the discrepancy of different radiomics features for glioma grade prediction. Sixteen feature selection methods and fifteen classification methods were evaluated in terms of their popularity, effectiveness and complexity in literature. Publicly available implementations for feature extraction, selection and classification strategies were adopted to reduce bias. Totally 210 GBM subjects and 75 LGG subjects were applied and the diagnostic performance were estimated by balanced accuracy and area under the curve (AUC) of receiver operating characteristic in both repeated ten-fold cross-validation and percentage split test modes. Besides, the roles of selected feature number, feature type, MRI modality and tumor sub-region were evaluated to optimize the radiomics-based glioma grade prediction.

II. METHODS AND MATERIALS

A. PATIENTS AND IMAGE PREPROCESSING

The public Multimodal Brain Tumor Segmentation Challenge (BraTS) 2018 magnetic resonance imaging (MRI) dataset was utilized for this study [32], [33]. This dataset was provided by Cancer Genome Atlas (TCGA) glioma phenotype research group. Totally 210 GBM patients and 75 LGG subjects were included. Four MRI modalities including native T1-weighted (T1), T1 post-contrast (T1-Gd), T2-weighted (T2) and T2-Fluid Attenuated Inversion Recovery (FLAIR) were supplied for each subject. Tumor regions of the subjects were provided, as well as three types of tumor sub-regions (the enhancing tumor region, ET; the peritumoral edema region, ED; the necrotic and non-enhancing tumor region, NCR/NET).

The details of image processing were described in the previous study [32]. All the provided images were skull-stripped, co-registered to the same anatomical structure by a rigid registration model with the mutual information similarity metric. And all the image volumes were resampled to 1mm isotropic resolution in a standardized axial orientation with a linear interpolator. No non-parametric, non-uniform intensity normalization algorithm was used by the BraTS group to correct for intensity non-uniformities caused by the inhomogeneity of the scanner's magnetic field during image acquisition, as it was observed that application of such algorithm obliterated the T2-FLAIR signal [32]. To ensure the comparability of intensity-based features, we performed image intensity normalization using the hybrid white-stripe approach, which has been proven robust for MRI data intensity normalization [34]. Also, intensity discretization was performed for the following extraction of texture features with fixed bin number method.

B. RADIOMICS FEATURE EXTRACTION

The public open-source pyradiomics feature extraction package (V1.3.0) was utilized to extract radiomics features from the tumor regions [35]. With this package, three categories of features were extracted from original images, including shape-based features, first-order statistics features, and texture features. The texture features included gray level co-occurence matrix (GLCM) features, gray level run length matrix (GLRLM) features, gray level size zone matrix (GLSZM) features, neighbouring gray tone difference matrix (NGTDM), gray level dependence matrix (GLDM) features. Besides, the aforementioned texture features (GLCM, GLRLM, GLDM and NGTDM) were also extracted from the images preprocessed using the Laplacian of Gaussian (LoG) band pass filter and Wavelet filter. The concrete feature definition could be found in [35]. Overall, as for each lesion, the radiomics features were extracted from both original and filtered images. All the features were then

TABLE 1. Summary of the used feature selection and classification methods with the acronyms and full names.

Acronym	Feature selection method name	Acronym	Classification method name
CHSQ	chi-square score	GNB	gaussion naïve bayes
TSQ	t-test score	MNB	multinomial naïve bayes
WLCX	wilcoxon	BNB	bernoulli naïve bayes
VAR	variance	KNN	k-nearest neighborhood
RELF	relief	RF	random forest
MI	mutual information	BAG	bagging
mRMRe	minimum redundancy maximum relevance ensemble	DT	decision tree
RF	random forest	GBDT	gradient boosting decision tree
ETE	extra tree ensemble	Adaboost	adaptive boosting
GBDT	gradient boosting decision tree	XGB	xgboost
XGB	xgboost	LDA	linear discriminant analysis
L^1-LGR	L^1-based logistic regression	LGR	logistic regression
L^1-SVM	L^1-based linear support vector machine	Linear-SVM	linear support vector machine
LASSO	least absolute shrinkage and selection operator	RBF-SVM	radial basis function support vector machine classification
EN	elastic net	MLPC	multi-layer perceptron
PCA	principal component analysis		

extracted in the four types of MRI modality (T1, T1-Gd, T2, FLAIR) images.

To explore the relationship between glioma grades and radiomics features, which could be extracted from different lesion types as well as MRI modalities, we combined the tumor sub-regions into seven kinds of ROIs into ET, NCR/NET, ED, ET + NCR/NET, NCR/NET + ED, ET + ED, ET + NCR/NET + ED. Radiomics features were then extracted from the ROIs respectively in different MRI modalities. Because the extracted radiomics features were multiple centers and magnitudes, feature normalization was performed with a mean of zero and a standard deviation of one (z-score transformation).

C. FEATURE SELECTION METHODS

Sixteen feature selection methods were chosen considering their popularity, effectiveness and complexity reported in previous interrelated researches [26]–[28], which included chi-square score (CHSQ), t-test score (TSQ), wilcoxon (WLCX), variance (VAR), Relief (RELF), mutual information (MI), minimum redundancy maximum relevance ensemble (mRMRe), random forest (RF), extra tree ensemble (ETE), gradient boosting decision tree (GBDT), xgboost (XGB), L^1-based logistic regression (L^1-LGR), L^1-based linear support vector machine (L^1-SVM), least absolute shrinkage and selection operator (LASSO), elastic net (EN) and principal component analysis (PCA).

Filter method and embedded method are two kinds of feature selection methods. Filter method is computational efficiency and classifier independent, while embedded method incorporates feature selection as a part of training process. Filter method has high generalizability and embedded method generally achieves high accuracy in previous study [28]. CHSQ, TSQ, WLCX and VAR are univariate filter feature selection methods which score the feature by relevancy, whereas RELF, MI and mRMRe are multivariate filter investigating the feature relevancy as well as the redundancy. The embedded methods are mainly composed of penalty-based and tree-based methods. L^1-LGR, L^1-SVM, LASSO and EN are representative penalty-based embedded methods while RF, ETE, GBDT and XGB are tree-based embedded methods. Different from the other types, PCA is an important feature extraction strategy which could generate new specified dimension features and achieve high performance in a way of dimensional reduction [36].

Feature selection method wrapper was not investigated in this paper because of its expensive computation [37], although some wrapper methods e.g. Boruta, had been proven valuable in preparing data [38]. Another feature selection method named hybrid method, which could be formed by combing multiple different feature selection methods [39], was not investigated in this paper either. We believe the comparison for single feature selection method will be beneficial for providing the references for hybrid feature selection method.

D. CLASSIFICATION METHODS

We investigated fifteen machine-learning classifiers: gaussion naïve bayes (GNB), multinomial naï ve bayes (MNB), bernoulli naïve bayes (BNB), k-nearest neighborhood (KNN), random forest (RF), bagging (BAG), decision tree (DT), gradient boosting decision tree (GBDT), adaptive boosting (Adaboost), xgboost (XGB), linear discriminant analysis (LDA), logistic regression (LGR), linear support vector machine (Linear-SVM), radial basis function support vector machine (RBF-SVM) and multi-layer perceptron (MLPC). The acronym for each feature selection method and classification method was listed in Table 1.

All the feature selection and classification methods were implemented using scikit-learn package in python [40] (scikit-learn version 0.21, python version 3.6.3). Cross-combination strategy was utilized to compare the

performance of feature selection and classification methods. Specifically, each feature selection method was combined with all the fifteen classification methods respectively, and each classification method was combined with all the sixteen feature selection methods. Finally, we got 240 combinations of feature selection and classification strategies.

E. EXPERIMENTAL DESIGN

1) DEFINING PREDICTION PERFORMANCE MATRIX

In this radiomics study, the performance of the feature reduction and classification methods was obtained by using two different test modes: k-fold cross-validation and percentage split [41].

a) k-fold cross-validation: we used ten-fold cross-validation, which split the data into 10 equal parts then used 9 parts for training and the rest part for testing alternately.

b) data percentage split criteria: the dataset was split into training data and testing data by certain percentage ratio. Here, for the totally 210 GBM and 75 LGG subjects, we assigned 228 patients (147 GBM, 53 LGG) to the training set, and 57 patients (63 GBM, 22 LGG) to the testing set according to a ratio of 7:3.

Since the dataset was unbalanced, with the GBM group about three times as large as that of the LGG group, which might be biased by the unbalanced distribution of the sample. We performed synthetic minority over-sampling technique (SMOTE) [42], [43] for over-sampling the LGG group to have the same number of instances as the HGG group in the training procedure. Balance accuracy defined in the equation (3) and AUC were used to as diagnostic indicators. True positive (TP, the correctly predicted positive instances number), false positive (FP, the incorrectly predicted positive instances number), false negative (FN, the incorrectly predicted negative instances number) and true negative (TN, the correctly predicted positive instances number) were used to calculate the indicators.

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}) \quad (1)$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \quad (2)$$

$$\text{Balanced accuracy} = (\text{Sensitivity} + \text{Specificity})/2 \quad (3)$$

2) EVALUATION OF SELECTED FEATURE NUMBER IN PREDICTING GLIOMA

Prediction accuracy was utilized to evaluate the number of selected features for different feature selection methods. To reduce the performance bias caused by different classifiers, we used balanced accuracy to rank the prediction performance of the classification methods and then the top four classifiers were selected to obtain the averaged balanced accuracy of the selected feature number. For each kind of feature selection method, a range of feature number from 10 to 160 with an interval of 5 was selected. The top four classifiers were then utilized to evaluate the prediction accuracy with repeated ten-fold cross-validation strategy.

3) EVALUATION OF FEATURE TYPE, MRI MODALITY AND TUMOR REGION

To explore the diagnostic value of different feature types, the normalized feature type importance (NFTI) coefficient was defined to describe the selected feature types for each feature selection method. Concretely, we counted the selected feature number corresponding to their radiomics types as described in the radiomics feature extraction section; the number was firstly normalized by the extracted feature number in each type and then normalized by the selected feature number for all feature types. Finally, we get the NFTI coefficient for each feature type in feature selection. Here, averaged NFTI was acquired with selected feature number from 40 to 80 by interval of 5 to reduce bias. Also, the prediction accuracy of extracted features in the four MRI modalities and seven combination of tumor sub-regions were compared to get the predictive value of the extracted features. It should be noted that not all the patients have the ET, ED or NCR/NET sub-regions in clinical, especially for the LGG patients. In this paper, we regarded the radiomics features in the missing region as specific features as the missing of a certain sub-region could also be correlated with the glioma grading.

III. RESULTS

A. COMPARISON OF FEATURE SELECTION AND CLASSIFICATION METHODS

To compare different machine-learning methods for radiomics models of glioma patients, we extracted quantitative features from whole tumor region and the multiparametric sequences. The diagnostic performance of feature selection and classification methods were evaluated by repeated ten-fold cross-validation and percentage split strategies. In this present study, the diagnostic performance was quantified by balanced accuracy and AUC. We examined 240 combinations of feature selection and classification methods. FIGURE 1 and FIGURE 2 depict the balanced accuracy and AUC results in repeated ten-fold cross-validation. FIGURE 3 and FIGURE 4 depict the results in percentage split strategy. In the ten-fold cross validation, feature selection method L^1-SVM + classifier MLPC achieved the highest prediction accuracy (accuracy:0.944, AUC:0.986), followed by XGB + classifier MLPC (accuracy:0.932, AUC:0.977), and L^1-SVM + classifier LDA (accuracy:0.930, AUC:0.988). For accuracy in percentage split testing set, feature selection method L^1-SVM + classifier MLPC achieved the highest prediction accuracy (accuracy:0.953, AUC:0.981), followed by LASSO + classifier LDA (accuracy:0.942, AUC:0.974), and L^1-SVM + classifier LDA (accuracy:0.936, AUC:0.985). Feature selection methods L^1-SVM, LASSO, XGB, GBDT exhibited valuable balanced accuracy and AUC performance with the majority classifiers. Meanwhile, for classifiers, XGB, LDA, LGR and MLPC demonstrated higher stabilities with the majorities of feature selection methods. However, VAR feature selection method showed low mean accuracy with the majority classifiers and MNB classifier showed lower mean accuracy.

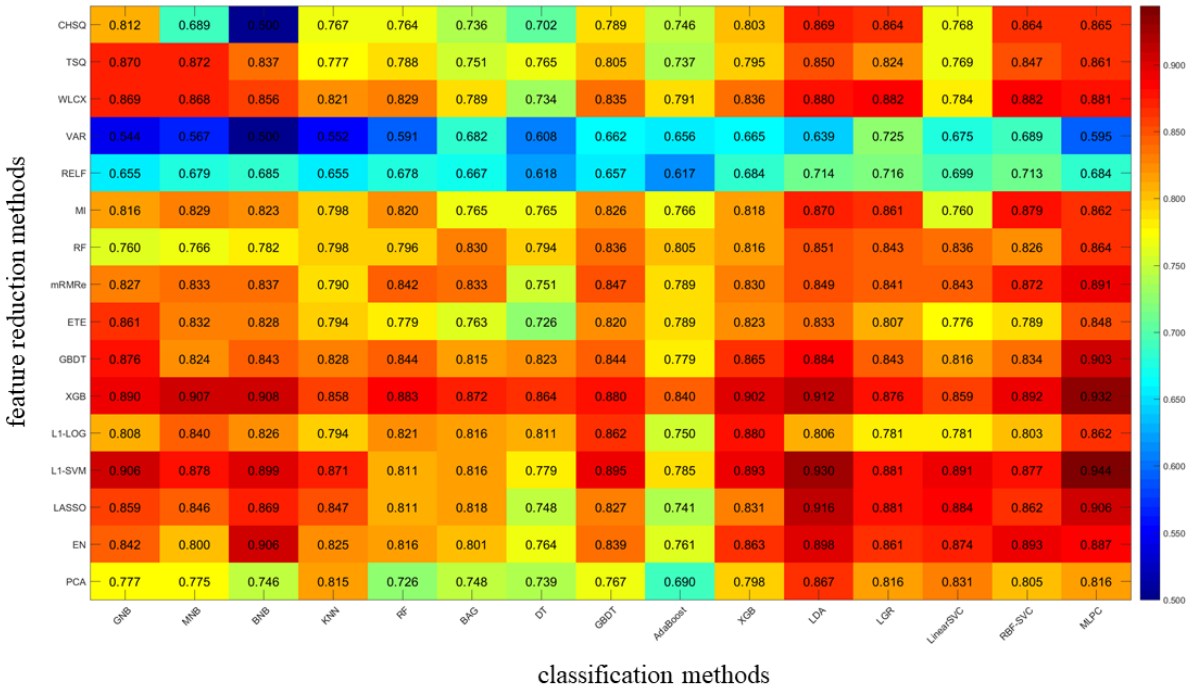


FIGURE 1. Balanced accuracy heatmap of feature selection methods (in columns) and feature classification methods (in rows) in ten-fold cross validation.

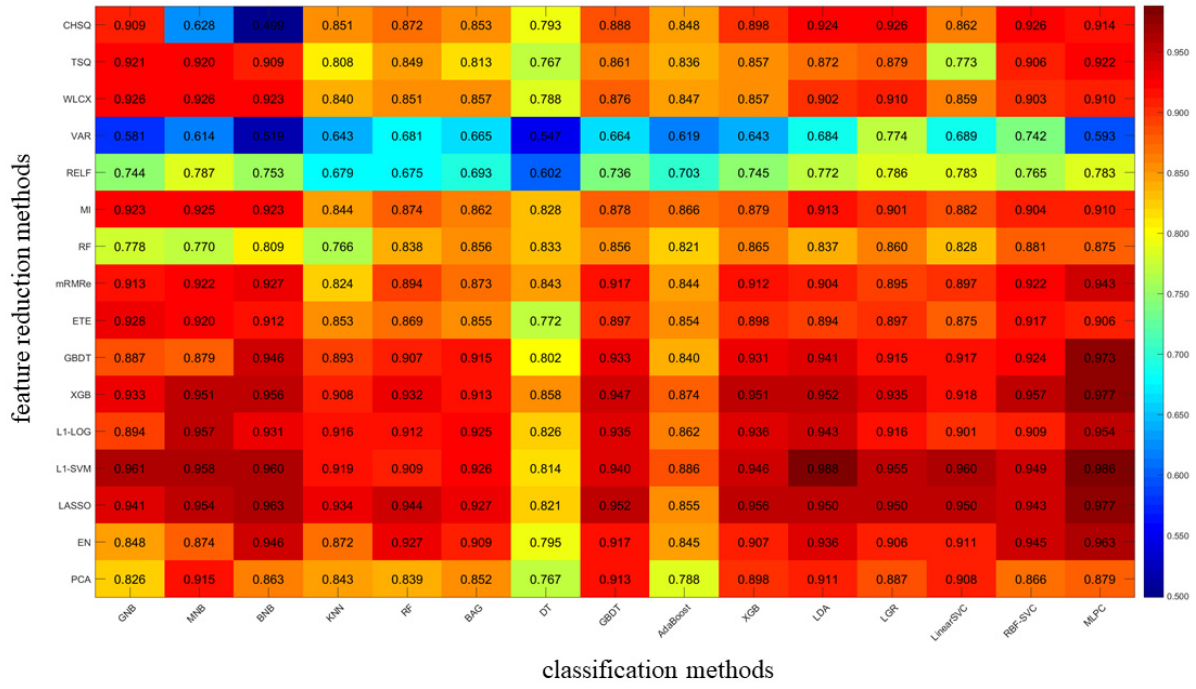


FIGURE 2. AUC heatmap of feature selection methods (in columns) and feature classification methods (in rows) in ten-fold cross validation.

B. SELECTED FEATURE NUMBER SLIGHTLY INFLUENCED PREDICTION ACCURACY

To explore the relationship between selected feature number and diagnostic accuracy, we adjusted feature selection parameters for each method to obtain a range of selected features. In this study, the range of selected feature numbers was from 5 to 160 with an interval of 5 as shown in FIGURE 5. To reduce the predicted accuracy bias caused by classifiers, the top four precise classifiers LDA, LGR, MLPC and XGB, were selected in the diagnostic evaluation step. Each subset

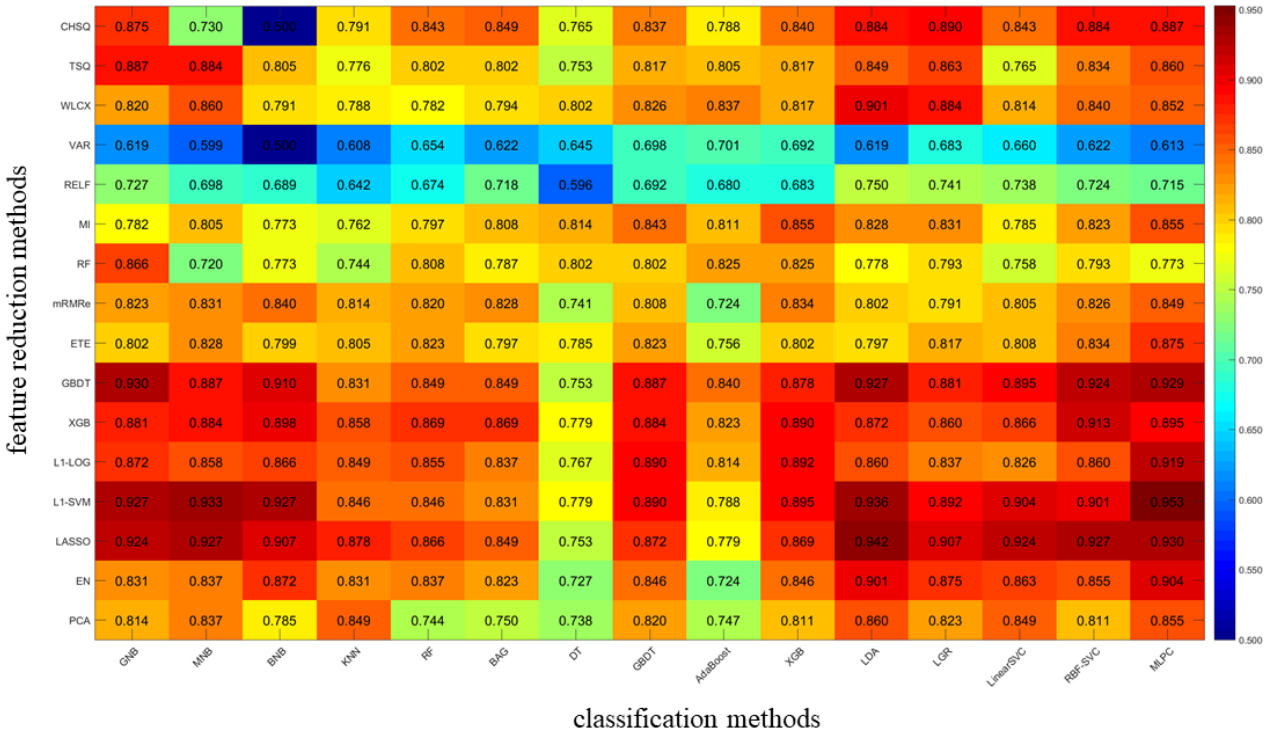


FIGURE 3. Balanced accuracy heatmap of feature selection methods (in columns) and feature classification methods (in rows) in percentage split.

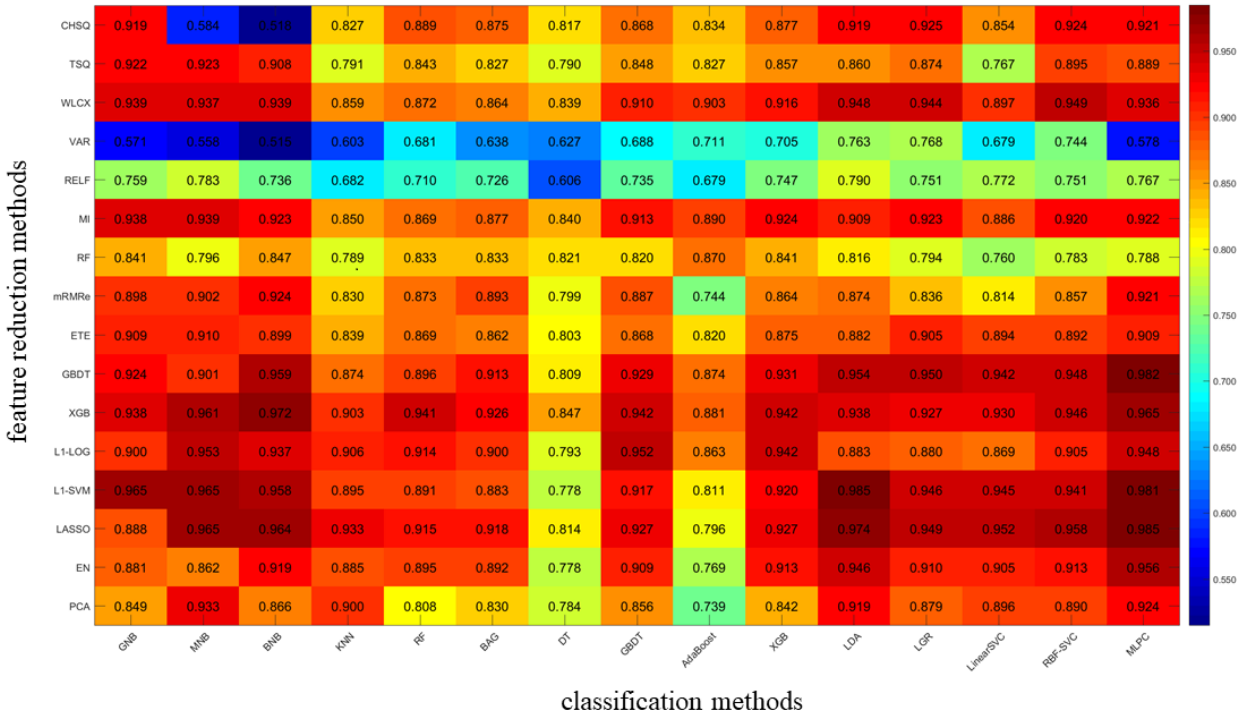


FIGURE 4. AUC heatmap of feature selection methods (in columns) and feature classification methods (in rows) in percentage split.

of selected features was then trained by the four classifiers respectively with repeated ten-fold cross-validation.

L^1-SVM and LASSO outperformed the other feature selection methods on the majority of feature numbers with the mean balanced accuracy 0.951 ± 0.014 and 0.946 ± 0.015 respectively. EN, L^1-LGR, XGB, GBDT, ETE, RF and MI achieved mean accuracy larger than 0.9 while the rest methods, e.g. VAR and RELF, showed lower mean accuracy.

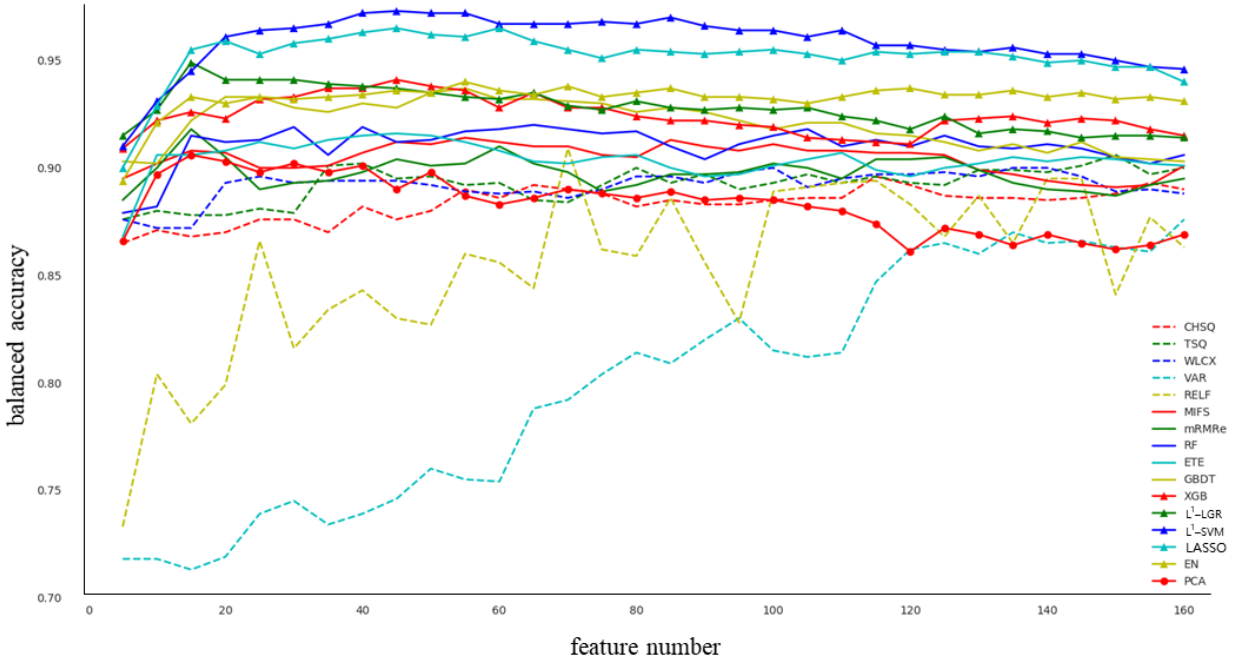


FIGURE 5. The balanced accuracy (in columns) for different feature selection methods with different selected feature numbers (in rows).

Also, we found that as increasing of selected feature number, the predicted accuracy was relatively stable for the majority of feature selection methods.

C. IDENTIFYING SELECTED FEATURE TYPES FOR PREDICTION

There were totally sixteen feature selection types used in this paper as described in the radiomics feature selection section. In this assessment, PCA feature selection method was excluded. As we can see from the FIGURE 6, for the high mean accuracy feature selection methods, e.g. L^1-SVM and GBDT, the selected features contained nearly all kinds of feature types while for the low performance methods, e.g. VAR, partial feature types were selected. The first-order statistics feature, GLCM and GLRLM texture features type were frequently selected for the majority of feature selection methods.

D. EXTRACTING PREDICTIVE RADIOMICS FEATURES

The diagnostic value for different MRI modalities and tumor sub-regions were then evaluated. To reduce the predicted accuracy bias caused by feature selection methods and classifiers, we selected top four precise feature selection methods (L^1-SVM, LASSO, XGB and GBDT) and top four precise classifiers (LDA, LGR, MLPC, and XGB) to get the average performance for the evaluation of the diagnostic balanced accuracy and AUC as shown in Table 2 in ten-fold cross validation. The ET + NCR/NET tumor region in T1-Gd modality achieved the highest diagnostic performance with balanced accuracy 0.901, AUC 0.953. For each tumor sub-region, T1-Gd had the highest mean balanced accuracy. For each MRI modality, the ET + NCR/NET tumor region had the highest mean balanced accuracy.

IV. DISCUSSION

Radiomics is an emerging and rapidly growing filed which converts medical images into quantitative mineable data [44]. In this study, different radiomics feature selection and classification methods were investigated to evaluate the discrepant performance for the prediction of glioma grade. Besides, other controllable variables e.g. the number of selected features, feature type, MRI modality and tumor ROI were discussed for optimal radiomics-based glioma grade prediction. We found that L^1-SVM + MLPC machine-learning strategy achieved the highest predictive performance in both repeated ten-fold cross-validation and percentage split test modes. Feature selection methods L^1-SVM, LASSO, XGB, GBDT exhibited valuable predictive balanced accuracy and AUC performance with the majority classifiers. Meanwhile, for classifiers, XGB, LDA, LGR and MLPC demonstrated higher stability with the majorities of feature selection methods. As for the extraction of radiomics features, the ET + NCR/NET region in T1-Gd modality provided the considerable tumor-related phenotypes in glioma grading.

Sixteen feature selection methods and fifteen classification methods were investigated for radiomics-based glioma grade prediction. Our results showed that L^1-SVM feature selection method combined with MLPC classification method yielded the highest diagnostic performance than the other crossed methods in glioma grade prediction. Subsequently, for the feature selection methods, L^1-SVM, LASSO, XGB and GBDT showed relatively higher predictive performance

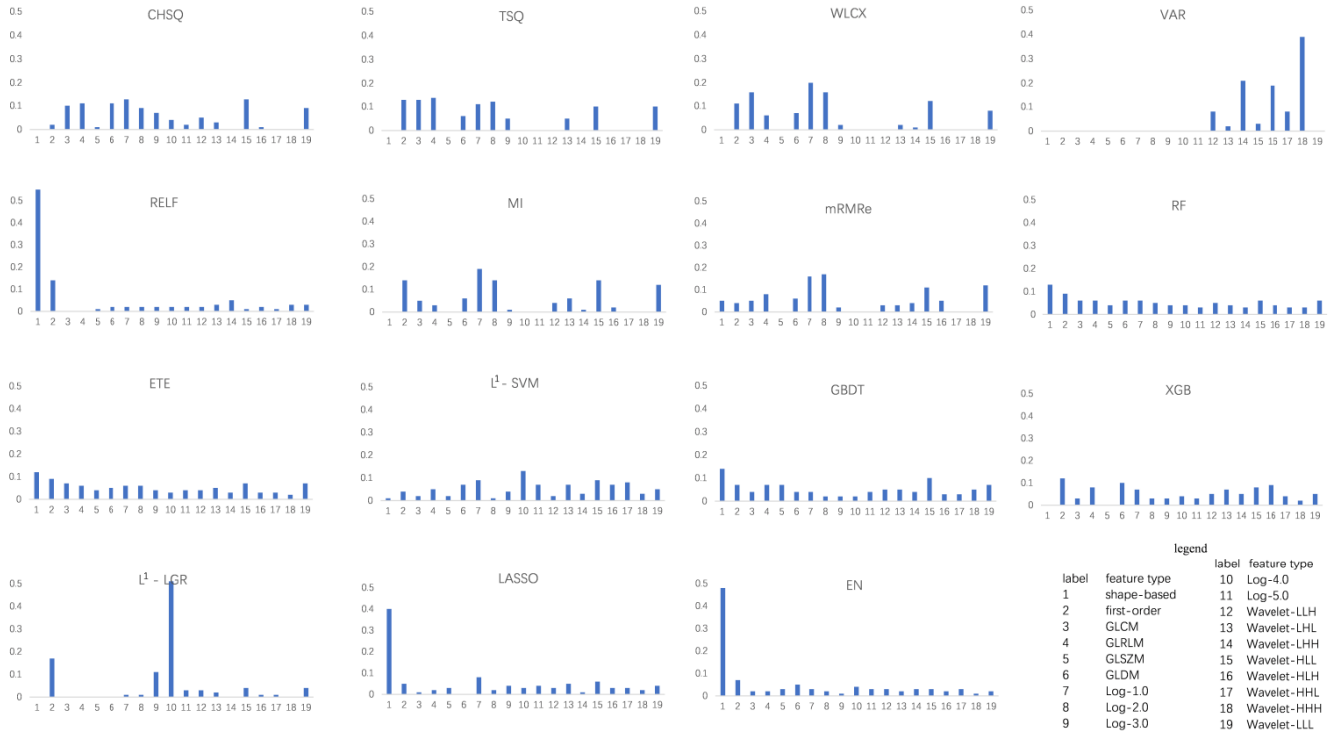


FIGURE 6. **The NFTI coefficients of selected feature in each feature type for the fifteen feature selection methods.**

TABLE 2. **Diagnostic evaluation for each MRI modality combined with different tumor sub-regions.**

Modality	Indicator	ET	NCR/NET	ED	ET+NCR/NET	NCR/NET+ED	ET+ED	ET+NCR/NET+ED
T1	balanced accuracy	0.770	0.748	0.775	0.852	0.832	0.732	0.845
	AUC	0.815	0.799	0.808	0.871	0.880	0.767	0.879
T1-Gd	balanced accuracy	*0.896*	*0.887*	*0.891*	*0.901*	*0.889*	*0.861*	*0.896*
	AUC	0.931	0.936	0.934	0.953	0.931	0.923	0.924
T2	balanced accuracy	0.839	0.811	0.828	0.878	0.856	0.798	0.848
	AUC	0.878	0.857	0.847	0.921	0.919	0.851	0.879
FLAIR	balanced accuracy	0.790	0.767	0.779	0.861	0.844	0.740	0.847
	AUC	0.841	0.814	0.828	0.872	0.898	0.774	0.890

with the majority of classifiers. Noted that LASSO has been proven to be an efficient feature selection strategy previously [26]. For the classification methods, MLPC, LDA, LGR, and XGB provided higher performance than other classifiers when combined with specified selected features in the most case. In previously evaluation of filter-based feature selection strategy for radiomics analysis, WLCX achieves satisfactory outcomes in NSCLC survival prediction [28]. In our study, although WLCX performed better than the other filter feature selection method, it was inferior to the embedded feature selection methods. RF classifier was frequently utilized in machine learning [45] and it was proven to be an efficient and powerful tree-based classification algorithm [30]. Whereas in this study its predictive performance was not outstanding than the other two boosting method XGB and GBDT. It should be noted that few researches have been compared between RF method and the high predictive performance classifiers provided in our paper previously for radiomics-based clinal predictions. Besides, by using the cross-combination strategy for feature selection and classification method comparison, we could filtrate the optimal radiomics-based framework for glioma grading. The cross-combination strategy might aid establishing framework for the future radiomics-based analysis.

Only a few studies have investigated and compared feature number and feature type in the radiomics based researches [46], [47]. In our study, for the majority of feature selection and classification methods, the diagnostic performance was consistent with corresponding utilized methods in previous results [28]. The highest performance feature

number interval was approximately between 40 and 80 for the glioma grade prediction. Meanwhile, there was a slightly decrease trend when the feature number exceeding 100. This might suggest that the predictive result benefited from feature selection. In other words, feature selection is an effective strategy to improve radiomics-based predictive studies. In the feature type analysis, we observed that for the high accuracy feature selection method, the selected features had relatively extensive coverage feature type than the low accuracy feature selection methods. This might be on account that different feature types contain different tumor characteristics [48] and a comprehensive feature extraction strategy is likely to improve the prediction of clinical outcome. These results provide a dimension for feature extraction which is crucial for the feature selection and classification and hence the overall clinical analysis.

The results for MRI modalities and tumor regions analysis showed that the ET + NCR/NET region in T1-Gd modality achieved the highest valuable tumor heterogeneity. This enhanced the diagnosis of GBM and LGG in the applications of non-invasive and cost-effective radiomics glioma grading based on MRI. ET + NET is routinely treated as solid tumor region, which has been frequently regarded as the ROI to extract radiomics features in a quantity of previous studies [4], [49]. Our results add verification that the solid tumor region is more relevant to the tumor grade than other tumor subregion or their combinations. Moreover, by comparing the results of the diagnostic performance, we found that the diagnostic value of combinative radiomics features in all multiparametric MRI modalities outperformed the single MRI modality.

There were also some limitations for this study. Firstly, only four kinds of MRI modalities were utilized in this paper. Quantitative imaging has been beneficial from the innovations and progresses in medical imaging hardware, imaging agents, standard protocol and imaging analysis strategies. And new parametric MRI modalities such as apparent diffusion coefficient (ADC) [50], diffusion kurtosis imaging (DKI) [10] have also shown tremendous glioma grading potential, while they are not investigated in our study. Secondly, the predictive value of radiomics for other glioma outcomes in clinical, which include progression survival, overall survival, recurrence of surgery, are expected to be discussed in future researches. Thirdly, deep learning, which is one of latest representative development of radiomics analysis, is not discussed in this study [51], [52]. Different from the traditional image features, the deep imaging features might possess different diagnostic radiomics features for glioma grading, and deep learning strategy might contribute to the precise classification of glioma grade. Finally, the MRI scans of glioma patients were collected from multi-institution, which increased the heterogeneity of the image quality. The standardization in image acquisition and reconstruction or feature harmonization method, e.g. ComBat [53], should be considered in future radiomics studies.

V. CONCLUSIONS

In conclusion, our study compared disparate radiomics strategies in the face of glioma grading. By comparing the feature selection and classification methods, we found that L^1-SVM feature selection method combined with MLPC classification method yielded the highest diagnostic performance. Meanwhile, feature selection methods, e.g. L^1-SVM, LASSO, XGB, GBDT and feature classification methods, e.g. LDA, MLPC, LGR, XGB demonstrated high diagnostic performance. As for the extraction of radiomics features, the ET + NCR/NET region in T1-Gd modality might provide the considerable tumor-related phenotypes in tumor grading, while other MRI modalities and tumor regions also exerted the tumor inhomogeneity. Our comparative investigation may be an important reference in identifying the reliable and effective machine-learning methods for radiomics-based diagnostic analysis in glioma grading non-invasively.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Wutao Lou and Dr. Hongbao Li for reading and critiquing this manuscript. They declare no conflict of interest.

REFERENCES

[1] D. Ricard, A. Idbaih, F. Ducray, M. Lahutte, K. Hoang-Xuan, and J.-Y. Delattre, "Primary brain tumours in adults," *Lancet*, vol. 379, no. 9830, pp. 1984–1996, May/Jun. 2012.

[2] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 world health organization classification of tumors of the central nervous system: A summary," *Acta Neuropathologica*, vol. 131, no. 6, pp. 803–820, Jun. 2016.

[3] Q. T. Ostrom, L. Bauchet, F. G. Davis, I. Deltour, J. L. Fisher, C. E. Langer, M. Pekmezci, J. A. Schwartzbaum, M. C. Turner, and K. M. Walsh, "The epidemiology of glioma in adults: A "state of the science," review," *Neuro Oncol.*, vol. 16, no. 7, pp. 896–913, Jul. 2014.

[4] X.-X. Qi, D.-F. Shi, S.-X. Ren, S.-Y. Zhang, L. Li, Q.-C. Li, and L.-M. Guan, "Histogram analysis of diffusion kurtosis imaging derived maps may distinguish between low and high grade gliomas before surgery," *Eur. Radiol.*, vol. 28, no. 4, pp. 1748–1755, Apr. 2018.

[5] H. Hyare, S. Thust, and J. Rees, "Advanced MRI techniques in the monitoring of treatment of gliomas," *Current Treat. Options Neurology*, vol. 19, no. 3, p. 11, Mar. 2017.

[6] C. H. Suh, H. S. Kim, S. C. Jung, J. E. Park, C. G. Choi, and S. J. Kim, "MRI as a diagnostic biomarker for differentiating primary central nervous system lymphoma from glioblastoma: A systematic review and meta-analysis," *J. Magn. Reson. Imag.*, vol. 14, pp. 560–572, Aug. 2019.

[7] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Commun.*, vol. 5, Jun. 2014, Art. no. 4006.

[8] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012.

[9] M. Zhou, J. Scott, B. Chaudhury, L. Hall, D. Goldgof, K. W. Yeom, M. Iv, Y. Ou, J. Kalpathy-Cramer, S. Napel, R. Gillies, O. Gevaert, and R. Gatenby, "Radiomics in brain tumor: Image assessment, quantitative feature descriptors, and machine-learning approaches," *Amer. J. Neuroradiol.*, vol. 39, no. 2, pp. 208–216, Feb. 2018.

[10] A. Falk Delgado, M. Nilsson, D. van Westen, and A. Falk Delgado, "Glioma grade discrimination with MR diffusion kurtosis imaging: A meta-analysis of diagnostic accuracy," *Radiology*, vol. 287, no. 1, pp. 119–127, Apr. 2018.

[11] X. Fan, C. Qi, X. Liu, Y. Wang, S. Liu, S. Li, L. Wang, and Y. Wang, "Regional specificity of matrix metalloproteinase-9 expression in the brain: Voxel-level mapping in primary glioblastomas," *Clin. Radiol.*, vol. 73, no. 3, pp. 283–289, Mar. 2018.

[12] Y. Ren, X. Zhang, W. Rui, H. Pang, T. Qiu, J. Wang, Q. Xie, T. Jin, H. Zhang, H. Chen, Y. Zhang, H. Lu, Z. Yao, J. Zhang, and X. Feng, "Noninvasive prediction of IDH1 mutation and ATRX expression loss in low-grade gliomas using multiparametric MR radiomic features," *J. Magn. Reson. Imag.*, vol. 49, no. 3, pp. 808–817, Mar. 2018.

[13] H.-H. Cho and H. Park, "Classification of low-grade and high-grade glioma using multi-modal image radiomics features," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Society. (EMBC)*, Jul. 2017, pp. 3081–3084.

[14] T. Xie, X. Chen, J. Fang, H. Kang, W. Xue, H. Tong, P. Cao, S. Wang, Y. Yang, and W. Zhang, "Textural features of dynamic contrast-enhanced MRI derived model-free and model-based parameter maps in glioma grading," *J. Magn. Reson. Imag.*, vol. 47, no. 4, pp. 1099–1111, Apr. 2018.

[15] S. Bisdas, C. Tisca, C. Sudre, E. Sanverdi, D. Roettger, and J. M. Cardoso, "Non-invasive *in vivo* prediction of tumour grade and IDH mutation status in gliomas using dynamic susceptibility contrast (DSC) perfusion- and diffusion-weighted MRI," *Amer. Soc. Clin. Oncol.*, vol. 36, no. 15, p. e24173, Jun. 2018.

[16] Q. Tian, L. F. Yan, X. Zhang, X. Zhang, Y. C. Hu, Y. Han, Z. C. Liu, H. Y. Nan, Q. Sun, Y. Z. Sun, Y. Yang, Y. Yu, J. Zhang, B. Hu, G. Xiao, P. Chen, S. Tian, J. Xu, W. Wang, and G. B. Cui, "Radiomics strategy for glioma grading using texture features from multiparametric MRI," *J. Magn. Reson. Imag.*, vol. 48, no. 6, pp. 1518–1528, Dec. 2018.

[17] C. Su, J. Jiang, S. Zhang, J. Shi, K. Xu, N. Shen, J. Zhang, L. Li, L. Zhao, J. Zhang, Y. Qin, Y. Liu, and W. Zhu, "Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour," *Eur. Radiol.*, vol. 29, no. 4, pp. 1986–1996, Apr. 2019.

[18] Y. Wu, B. Liu, W. Wu, Y. Lin, C. Yang, and M. Wang, "Grading glioma by radiomics with feature selection based on mutual information," *J. Ambient Intell. Hum. Comput.*, vol. 9, no. 5, pp. 1671–1682, 2018.

[19] A. Vamvakas, S. Williams, K. Theodorou, E. Kapsalaki, K. Fountas, C. Kappas, K. Vassiou, and I. Tsougos, "Imaging biomarker analysis of advanced multiparametric MRI for glioma grading," *Phys. Medica*, vol. 60, pp. 188–198, Apr. 2019.

[20] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "Computer-aided grading of gliomas combining automatic segmentation and radiomics," *Int. J. Biomed. Imag.*, vol. 2018, Apr. 2018, Art. no. 2512037.

[21] H.-H. Cho, S.-H. Lee, J. Kim, and H. Park, "Classification of the glioma grading using radiomics analysis," *PeerJ*, vol. 6, p. e5982, Nov. 2018.

[22] X. Chen, M. Fang, D. Dong, L. Liu, X. Xu, X. Wei, X. Jiang, L. Qin, and Z. Liu, "Development and validation of a MRI-based radiomics prognostic classifier in patients with primary glioblastoma multiforme," *Academic Radiol.*, to be published.

[23] J.-B. Qin, Z. Liu, H. Zhang, C. Shen, X.-C. Wang, Y. Tan, S. Wang, X.-F. Wu, and J. Tian, "Grading of gliomas by using radiomic features on multiple magnetic resonance imaging (MRI) sequences," *Med. Sci. Monitor*, vol. 23, pp. 2168–2178, May 2017.

[24] Q. Wang, Q. Li, R. Mi, H. Ye, H. Zhang, B. Chen, Y. Li, G. Huang, and J. Xia, "Radiomics nomogram building from multiparametric MRI to predict grade in patients with glioma: A cohort study," *J. Magn. Reson. Imag.*, vol. 49, no. 3, pp. 825–833, Mar. 2019.

[25] E. Lotan, R. Jain, N. Razavian, G. M. Fatterpekar, and Y. W. Lui, "State of the art: Machine learning applications in glioma imaging," *Amer. J. Roentgenology*, vol. 212, no. 1, pp. 26–37, Jan. 2019.

[26] P. Yin, N. Mao, C. Zhao, J. Wu, C. Sun, L. Chen, and N. Hong, "Comparison of radiomics machine-learning classifiers and feature selection for differentiation of sacral chordoma and sacral giant cell tumour based on 3D computed tomography features," *Eur. Radiol.*, vol. 29, no. 4, pp. 1841–1847, Apr. 2019.

[27] B. Zhang, X. He, F. Ouyang, D. Gu, Y. Dong, L. Zhang, X. Mo, W. Huang, J. Tian, and S. Zhang, "Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma," *Cancer Lett.*, vol. 403, pp. 21–27, Sep. 2017.

[28] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, "Machine learning methods for quantitative radiomic biomarkers," *Sci. Rep.*, vol. 5, Aug. 2015, Art. no. 13087.

[29] X. Zhang, L.-F. Yan, Y.-C. Hu, G. Li, Y. Yang, Y. Han, Y.-Z. Sun, Z.-C. Liu, Q. Tian, and Z.-Y. Han, "Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features," *Oncotarget*, vol. 8, no. 29, pp. 47816–47830, Jul. 2017.

[30] T. M. Deist, F. J. Dankers, G. Valdes, R. Wijsman, I. C. Hsu, C. Oberije, T. Lustberg, J. van Soest, F. Hoebers, and A. Jochems, "Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers," *Med. Phys.*, vol. 45, no. 7, pp. 3449–3459, Jul. 2018.

[31] S. Leger, A. Zwanenburg, K. Pilz, F. Lohaus, A. Linge, K. ZÃűphel, J. Kotzerke, A. Schreiber, I. Tinhofer, and V. Budach, "A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling," *Sci. Rep.*, vol. 7, no. 1, Oct. 2017, Art. no. 13206.

[32] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Sci. Data*, vol. 4, Sep. 2017, Art. no. 170117.

[33] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, and R. Wiest, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[34] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, and C. M. Crainiceanu, "Statistical normalization techniques for magnetic resonance imaging," *NeuroImage, Clin.*, vol. 6, pp. 9–19, Jan. 2014.

[35] J. J. M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H. Beets-Tan, J. C. Fillion-Robin, S. Pieper, and H. J. W. L. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017.

[36] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Adv. Bioinf.*, vol. 2015, May 2015, Art. no. 198363.

[37] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.

[38] M. B. Kursa, A. Jankowski, and W. R. Rudnicki, "Boruta—A system for feature selection," *Fundam. Inf.*, vol. 101, no. 4, pp. 271–285, 2010.

[39] Y. Peng, Z. Wu, and J. Jiang, "A novel feature selection approach for biomedical data classification," *J. Biomed. Inform.*, vol. 43, no. 1, pp. 15–23, 2010.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[41] M. N. Halgamuge, "Machine Learning for bioelectromagnetics: Prediction model using data of weak radiofrequency radiation effect on plants," *Mach. Learn.*, vol. 8, no. 11, pp. 223–235, 2017.

[42] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[43] Y. Zhang, A. Oikonomou, A. Wong, M. A. Haider, and F. Khalvati, "Radiomics-based prognosis analysis for non-small cell lung cancer," *Sci. Rep.*, vol. 7, Art. no. 46349, Apr. 2017.

[44] E. F. Jackson, "Quantitative imaging: The translation from research tool to clinical practice," *Radiology*, vol. 286, no. 2, pp. 499–501, Feb. 2018.

[45] T. Upadhaya, M. Vallières, A. Chatterjee, F. Lucia, P. A. Bonaffini, I. Masson, A. Mervoyer, C. Reinhold, U. Schick, J. Seuntjens, C. C. Le Rest, D. Visvikis, and M. Hatt, "Comparison of radiomics models built through machine learning in a multicentric context with independent testing: Identical data, similar algorithms, different methodologies," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 192–200, Mar. 2019.

[46] C. Lian, S. Ruan, T. Denœux, F. Jardin, and P. Vera, "Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction," *Med. Image Anal.*, vol. 32, pp. 257–268, Aug. 2016.

[47] U. R. Acharya, Y. Hagiwara, V. K. Sudarshan, W. Y. Chan, and K. H. Ng, "Towards precision medicine: From quantitative imaging to radiomics," *J. Zhejiang Univ.-Sci. B*, vol. 19, no. 1, pp. 6–24, Jan. 2018.

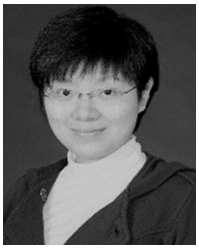
[48] R. T. Larue, G. Defraene, D. De Ruysscher, P. Lambin, and W. van Elmpt, "Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures," *Brit. J. Radiol.*, vol. 90, no. 1070, Feb. 2017, Art. no. 20160665.

[49] R. Jiang, J. Jiang, L. Zhao, J. Zhang, S. Zhang, Y. Yao, S. Yang, J. Shi, N. Shen, C. Su, J. Zhang, and W. Zhu, "Diffusion kurtosis imaging can efficiently assess the glioma grade and cellular proliferation," *Oncotarget*, vol. 6, no. 39, pp. 42380–42393, Dec. 2015.
[50] S. Vajapeyam, D. Brown, P. R. Johnston, K. I. Ricci, M. W. Kieran, H. G. W. Lidov, and T. Y. Poussaint, "Multiparametric analysis of permeability and ADC histogram metrics for classification of pediatric brain tumors by tumor grade," *Amer. J. Neuroradiol.*, vol. 39, no. 3, pp. 552–557, Mar. 2018.
[51] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[52] J. Lao, Y. Chen, Z.-C. Li, Q. Li, J. Zhang, J. Liu, and G. Zhai, "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Sci. Rep.*, vol. 7, no. 1, Sep. 2017, Art. no. 10353.
[53] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical Bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, Jan. 2007.

PAN SUN received the B.Sc. degree from the Department of Control Science and Engineering, Shandong University, Jinan, China, in 2013, and the M.Sc. degree from the Department of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China, in 2016. She is currently pursuing the Ph.D. degree with the Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China. Her research interests include medical image processing and machine learning.

DEFENG WANG received the B.Sc. degree from the Department of Computer Science, Jilin University, and the Ph.D. degree from the Department of Computing, The Hong Kong Polytechnic University. He is currently a Professor with the School of Instrumentation Science and Optoelectronics Engineering, Beihang University, and the Beijing Advanced Innovation Center for Big Data-Based Precision Medicine. His research interests include medical imaging, statistical morphometry analysis, fMRI paradigm design and post-processing, quantitative medical image analysis, and computational life science.

VINCENT CT MOK is currently a Professor with the Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong. He is the Head of the Division of Neurology, and the Assistant Dean of the Faculty of Medicine. He received the Higher Education Outstanding Scientific Research Output Award in Natural Sciences (1st Class) from the Ministry of Education, China, in 2011, for his research in cerebrovascular disease, and the Excellent Research Award from the Food and Health Bureau of the Hong Kong Special Administrative Region, in 2017, for his research in vascular cognitive disorder. He was 7-time Teacher of the Year Award recipient and the Master Teacher Award.



LIN SHI received the Ph.D. degree from The Chinese University of Hong Kong, in 2008. She is currently an Assistant Professor with the Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong. Her research interests include neuroimaging, image analysis and developing new image computing tools and applying them to facilitate the detection of imaging biomarkers for various neurological disorders.