



Comparison of feature selection method and machine learning classifier for radiomics analysis in glioma evaluation

NISHITH NARAYAN SHETTY

PRATHAMESH VALMIK PATIL

Guide : Ms. Bindy Wilson

Keraleeya Samajam's Model College, Khambalpada Road, Thakurli,
Dombivli (East), Maharashtra

Abstract Radiomics-based studies have demonstrated predictive capabilities using machine learning approaches. However, it is still unclear whether different radiomics strategies affect prediction performance. The purpose of this study was to compare the predictive performance of commonly used radiographic feature selection and classification methods in glioma staging. Quantitative radiomics features were extracted from tumor regions of 210 High MRI glioblastoma (GBM) and 75 low-grade glioma (LGG) patients. Then, the diagnostic performance of 16 feature selection and 15 classification methods was evaluated using two different test modes: 10-fold cross-validation and percentage split. The receiver operating characteristics balanced accuracy and area under the curve (AUC) were used to evaluate the predictive performance. Additionally, the number of selected features, feature type, MRI modality, and role of tumor subregions were compared to optimize radiomics-based prediction. Results show that the combination of L1-based linear support vector machine (L1-SVM) feature selection method and classifier multilayer perceptron (MLPC) distinguishes between GBM and LGG in both 10-fold cross-validation (balanced) It was shown that the best performance was achieved. Accuracy: 0.944, AUC: 0.986) and Percent Split (Balanced Accuracy: 0.953, AUC: 0.981). Radiomics feature extraction combines the enhancing tumor region (ET) with necrotic tumor region and non-enhancing tumor region (NCR/NET) in the T1 post-contrast modality (T1-Gd), and other combinations of tumor regions and MRI modalities. A more significant tumor-related phenotype was obtained. In our comparative study, we found that both feature selection methods and machine learning classifiers influenced predictive performance in glioma staging. Furthermore, a cross-combination strategy comparing radiomics feature selection and classification methods provided a way to find the best machine learning model for future radiomics-based predictions.

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 can occur anywhere in the brain and exhibits strong spatiotemporal heterogeneity. According to the WHO, gliomas can be classified into grades I to IV based on their histological malignant behavior [2]. Patients with low-grade glioma (LGG, grades I and II) typically have a life expectancy of 5 years or more. The associate editor who coordinated the review of this manuscript and approved it for publication was Yi Zhang. The survival rate is 4,444, but only 3-5% of glioblastoma (GBM, grade IV) patients survive for more than 5 years, with a median survival of approximately 12 months [3

]. GBM is the most common histological type of glioma, accounting for 70% of primary brain tumors. Preoperative classification of gliomas, especially the distinction between GBM and LGG, is of great importance for diagnostic decisions in clinical practice [4].

MRI offers the possibility of non-invasive classification of gliomas with high spatial resolution and unique contrast between brain tissue and tumor [5], [6]. A comprehensive view of brain structures and tumors reveals a high degree of heterogeneity in histological tumor subregions. Quantitative analysis can then be performed on tumor regions of interest (ROIs) to explore the relationship between tumor

characteristics and clinical diagnosis. Radiomics is one of the analysis techniques that converts image data into resolvable features with high throughput [7], [8]. This is based on the hypothesis that these image-based features capture phenotypic differences in tumors and have the potential to be used as diagnostic features for clinical outcomes. Radiomics provides a non-invasive method to study the relationship between gliomas and image-based descriptors such as tumor appearance, shape, size, intensity, location, and texture [9].

Current quantitative radiomics-based diagnostic models have shown high clinical potential to predict glioma malignancy [10], gene expression patterns [11], and genetic mutations [12].

Current radiomics-based MRI grading of gliomas uses different radiomics features, feature selection methods, and classification or regression models that exhibit identical diagnostic performance. As the extracted radiomics feature types, histogram-based features [4], shape features [13], texture features [14]-[16], and wavelet features [17] are designed. Then, feature selection strategies such as filter-based [18] and embedding methods [19], [20] are applied to identify features that are valuable for glioma grading. Next is the machine learning classifier. Random forest [21], logistic regression [13], and support vector machine [22] are used to classify histological glioma types. Additionally, multiparameter MRI sequences are used to extract image features for grading gliomas [23], [24]. Most of the radiomics-based studies have demonstrated the predictive ability of a number of machine learning approaches [9], [25]. Few recent studies have compared the diagnostic performance of different radiomics feature selection and classification models [26]-[31]. However, it is still unknown whether different feature selection and classification methods affect the performance of radiomics-based predictions in glioma staging. In this regard, the use of reliable machine learning strategies for effective image feature extraction and comparison in glioma grading is desirable.

In this study, we investigated the diagnostic value of commonly used machine learning approaches and the inconsistency of different radiomics features for predicting glioma grade. Sixteen feature selection methods and 15 classification methods were evaluated for popularity, effectiveness, and complexity in the literature. To reduce bias, commonly available implementations of feature extraction, selection, and classification strategies were adopted. A total of 210 GBM subjects and 75 LGG subjects were used to obtain balanced accuracy and area under the curve (AUC) for receiver operating characteristics in both iterative 10-fold cross-validation mode and percentage-split test mode. The diagnostic performance was estimated.

Additionally, the role of selected feature number, feature type, MRI modality, and tumor subregion was evaluated to optimize radiomics-based glioma grade prediction.

II. Method and materials

A. PATIENTS AND IMAGE PREPROCESSING

The Multimodal Brain Tumor Segmentation Challenge (BraTS) 2018 public magnetic resonance imaging (MRI) dataset was used for this study [32], [33]. This dataset was provided by the Cancer Genome Atlas (TCGA) Glioma

Phenotyping Research Group. A total of 210 his GBM and 75 LGG patients were included. Four MRI modalities were provided to each subject, including native T1-weighted (T1), T1 post-contrast (T1-Gd), T2-weighted (T2), and T2 fluid-attenuated inversion recovery (FLAIR). The subject's tumor region and three types of tumor subregions (enhancing tumor region, ET, peritumoral edema region, ED, necrotic and non-enhancing tumor region, NCR/NET) were provided. Image processing details are described in a previous study [32]. All provided images were released from the skull and mapped to the same anatomical structure through a rigorous registration model using a mutual information similarity metric. All image volumes were then resampled to an isotropic resolution of 1 mm in the standardized axial direction using a linear interpolator.

The BraTS group did not use a nonparametric nonuniform intensity normalization algorithm to correct for the intensity nonuniformity caused by the scanner magnetic field inhomogeneity during image acquisition. Because it was observed that the T2 FLAIR signal decreases when applying such an algorithm.[32]. To ensure comparability of intensity-based features, we performed image intensity normalization using a hybrid white stripe approach. It has been shown to be robust to intensity normalization of MRI data [34]. Additionally, we performed intensity discretization using a fixed bin number method for the next texture feature extraction.

B. RADIOMICS FEATURE EXTRACTION

To extract radiomics features from tumor regions, the publicly available open-source pyradiomics feature extraction package (V1.3.0) was used [35]. This package was used to extract three categories of features from the original images, including shape-based features, first-order statistical features, and texture features. Texture features include gray level co-occurrence matrix (GLCM) features, gray level run-length matrix (GLRLM) features, gray level size zone matrix (GLSZM) features, adjacent gray tone difference matrix (NGTDM), and gray level dependence matrix.

(GLDM) included.) Features. Additionally, the aforementioned texture features (GLCM, GLRLM, GLDM, NGTDM) were also extracted from the images preprocessed with a Gaussian Laplace (LoG) bandpass filter and a wavelet filter. Definitions of specific features can be found in [35].

Overall, the radiomics features of each lesion were extracted from both the original and filtered images.

All features were available at that time Images extracted from four types of MRI modality images (T1, T1-Gd, T2, FLAIR).

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	gaussian 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		

To investigate the relationship between glioma grade and radiomic features that can be extracted from different lesion types and MRI modalities, tumor subregions were divided into seven ROIs (ET, NCR/NET, ED, ET + NCR/ NET, NCR)/NET + ED, ET + ED, ET + NCR/NET + ED. Radiomic features were then extracted from the ROIs in different MRI modalities. Since the extracted radiomics features are of multiple centers and scales, feature normalization was performed with a mean of 0 and a standard deviation of 1 (Z-score transformation).

C. FEATURE SELECTION METHODS

Considering the popularity, validity, and complexity reported in previous related studies [26]–[28], 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), L1 based Logistic Regression (L1-LGR), L1-based Linear Support Vector Machine (L1-SVM), Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (EN), and Principal Component Analysis (PCA). There are two types of feature selection methods: filter methods and embedding methods. Filtering methods are computationally efficient and classifier independent, whereas embedding methods involve feature selection as part of the training process. Filtering methods are highly versatile, and embedding methods have generally achieved high accuracy in previous research [28]. CHSQ, TSQ, WLCX, and VAR are univariate filter feature selection methods that evaluate feature relevance, while RELF, MI, and mRMRe are multivariate filters that examine both feature relevance and redundancy. Embedding methods mainly consist of penalty-based methods and tree-based methods. Typical penalty-based embedding methods include L1-LGR, L1-SVM, LASSO, and EN, and tree-based embedding methods include RF, ETE, GBDT, and XGB. Unlike other types, PCA is an important feature extraction strategy that can generate features of new specified dimensions and achieve high performance through dimensionality reduction [36]. Wrappers for feature selection methods were not considered in this paper due to their computationally complex nature [37]. B. Boruta has proven valuable in data preparation [38].

Another feature selection method called hybrid method, which can be formed by combining several different feature selection methods [39], was also not investigated in this work. This comparison is considered useful as the single feature selection method serves as a reference for the hybrid feature selection method.

D. CLASSIFICATION METHODS

We considered 15 machine learning classifiers: Gaussian Naive Bayes (GNB), Multinomial Naive Bayes (MNB), Bernoulli Naive Bayes (BNB), k-Nearest Neighbors (KNN), Random Forests (RF), and 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 Multilayer Perceptron (MLPC). Table 1 lists the acronyms for each feature selection and classification method. All feature selection and classification methods were implemented using the scikit-learn package in Python [40] (scikit-learn version 0.21, Python version 3.6.3).

We used a cross-combination strategy to compare the performance of feature selection and classification methods. Specifically, each feature selection method was combined with all 15 classification methods, and each classification method was combined with all 16 feature selection methods. In the end, we obtained 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 method was determined using two different testing modes: k-fold cross-validation and percentage splitting [41].

a) k-fold cross-validation: We used 10-fold cross-validation by splitting the data into 10 equal parts and alternating using 9 parts for training and the rest for testing.

b) Data percentage split criteria: The dataset was split into training data and testing data with a certain percentage ratio. Here, for a total of 210 his GBM subjects and 75 LGG subjects, 228 patients (147 GBM, 53 LGG) were assigned to the training set and 57 patients (63 GBM, 22 LGG) were assigned to the test set. assigned to. This corresponds to a ratio of 73. The Da dataset is unbalanced, with the GBM group being about 3 times larger than his LGG group, which may have been biased by the unbalanced distribution of the samples. We performed a synthetic minority oversampling technique (SMOTE) [42], [43] to oversample the LGG group to have the same number of instances as the HGG group in the training procedure. The balance accuracy and AUC defined by equation (3) were used as diagnostic indicators. True positives (TP, number of correctly predicted positive instances), false positives (FP, number of incorrectly predicted positive instances), false negatives (FN, number of incorrectly predicted negative instances), and true negatives (TN, number of correctly predicted positive instances). was used to calculate the indicator.

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

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

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

2) EVALUATION OF SELECTED FEATURE NUMBER IN PREDICTING GLIOMA

Prediction accuracy was used to evaluate the number of features selected by different feature selection methods.

To reduce the performance distortion caused by different classifiers, we evaluate the predictive performance of the classification method using balanced accuracy, select the top four classifiers, and calculate the average of the selected feature numbers. Obtained a balanced accuracy. For each type of feature selection method, a range of feature numbers from 10 to 160 in 5 intervals was selected. We then repeated the 10-fold cross-validation strategy using the top four classifiers to evaluate prediction accuracy.

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

To explore the diagnostic value of different feature types, normalized feature type importance (NFTI) coefficients were defined to describe the feature types selected by each feature selection method.

Specifically, we counted the number of selected features corresponding to the radiomics type as described in the "Radiomics feature extraction" section. This number was first normalized by the feature number extracted for each type and then by the feature number selected for all feature types. Finally, during feature selection, we obtain her NFTI coefficients for each feature type. Here, the averaged NFTI was collected with selected feature numbers between 40 and 80 in 5 intervals to reduce bias. Furthermore, we compared the predictive accuracy of the extracted features across four MRI modalities and seven combinations of tumor subregions to determine the predictive value of the extracted features. It should be noted, especially in LGG patients, that not all

patients clinically exhibit ET, ED, or NCR/NET subregions. In this study, we considered the radiomics features of the defective region as a specific feature, as the absence of specific subregions may also correlate with 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 the entire tumor region and multiparametric sequences. The diagnostic performance of the feature selection and classification methods was evaluated using an iterative 10-fold cross-validation and percentage split strategy. In this current study, diagnostic performance was quantified by balanced accuracy and AUC.

We investigated 240 combinations of feature selection and classification methods. Figures 1 and 2 show balanced accuracy and AUC results for 10-fold cross-validation iterations. Figures 3 and 4 show the results of the percentage split strategy. In 10 rounds of cross-validation, the feature selection method L1-SVM + classifier MLPC achieves the highest prediction accuracy (accuracy: 0.944, AUC: 0.986), followed by XGB + classification MLPC (accuracy: 0.932, AUC: 0.988). For the accuracy of percentage split test set, the feature selection method L1-SVM + classifier MLPC has the highest prediction accuracy (Accuracy: 0.953, AUC: 0.981), followed by LASSO + classifier LDA (accuracy: 0.942, AUC: 0.974). and L1-SVM+ClassifierLDA (accuracy: 0.936, AUC: 0.985).

L1-SVM, LASSO, XGB, and GBDT feature selection methods demonstrated valuable balanced accuracy and AUC performance for most classifiers.

Meanwhile, among the classifiers, XGB, LDA, LGR, and MLPC showed higher stability for most feature selection methods. However, the VAR feature selection method showed a lower average accuracy for the majority classifier, and the MNB classifier showed a lower average accuracy.

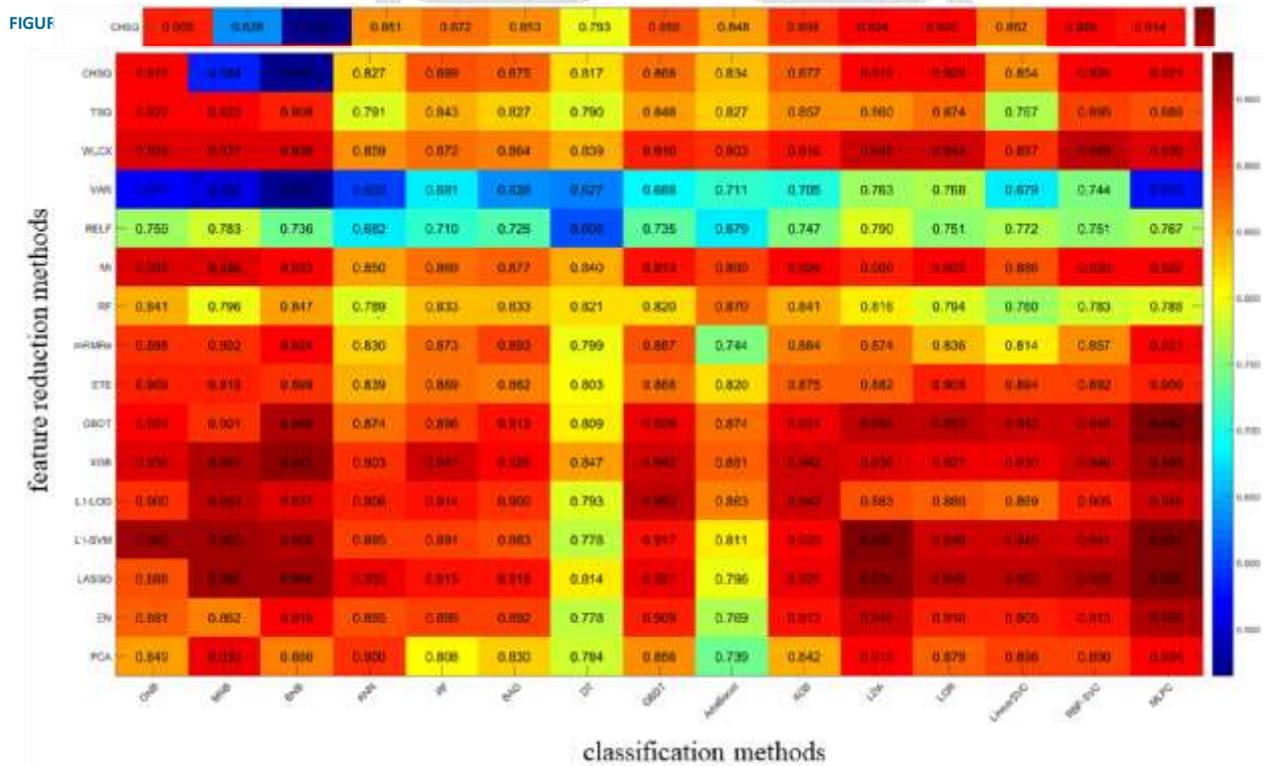
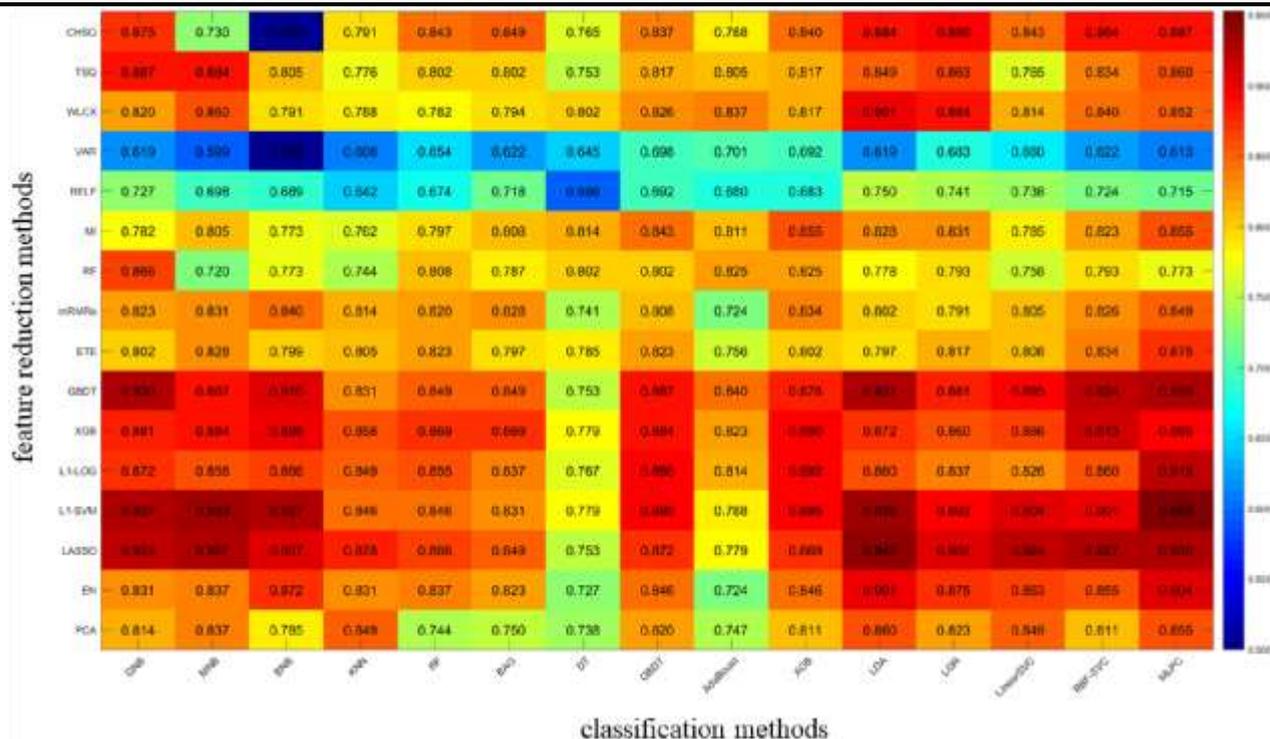


FIGURE 4. AUC heatmap of feature selection methods (in columns) and feature classification methods (in rows) in percentage split validation.

B. SELECTED FEATURE NUMBER SLIGHTLY INFLUENCED PREDICTION ACCURACY

To investigate the relationship between the number of selected features and diagnostic accuracy, we tuned the feature selection parameters of each method to obtain a set of selected features. In this study, the selected feature numbers ranged from 5 to 160 in 5 intervals, as shown in Figure 5.

The four most accurate classifiers LDA, LGR, MLPC, and XGB were used to reduce the bias in prediction accuracy

caused by the classifiers. selected in the diagnostic evaluation step. each subset

A selection of selected features was then trained by four classifiers with 10 iterations of cross-validation each. L1-SVM and LASSO outperformed other feature selection methods for most feature counts, with average balance accuracies of 0.951 ± 0.014 and 0.946 ± 0.015 , respectively. EN, L1-LGR, XGB, GBDT, ETE, RF, and MI achieved average accuracies above 0.9, while the remaining methods (e.g., B.VAR and RELF had lower average accuracies of).

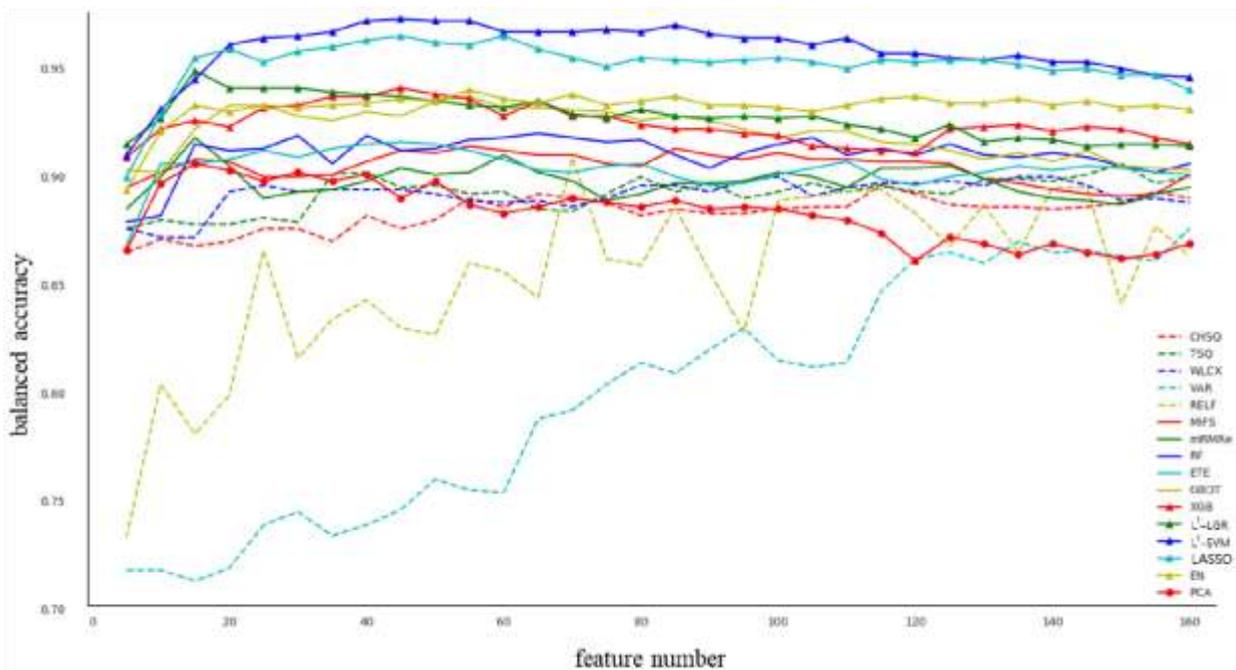


FIGURE 5. The balanced accuracy (in columns) for different feature selection methods with different selected feature numbers (in rows).

We also found that the predictive accuracy of most feature selection methods remained relatively stable as the number of selected features increased.

C. IDENTIFYING SELECTED FEATURE TYPES FOR PREDICTION

A total of 16 feature selection types were used in this article, as described in the “Radiomics Feature Selection” section. This evaluation excluded the PCA feature selection method. As can be seen from Figure 6, for feature selection methods with high average accuracy, e.g. L1-SVM and GBDT, the selected features included almost all types of feature types, while for low-performance For the method B.VAR, a partial object type was selected. Most feature selection methods often select first-order statistical, GLCM, and GLRLM texture feature types.

D. EXTRACTING PREDICTIVE RADIOMICS FEATURES

The diagnostic value was then evaluated for different MRI modalities and tumor subregions. To reduce the prediction accuracy bias caused by feature selection methods and classifiers, we used the four most accurate feature selection methods (L1-SVM, LASSO, XGB, and GBDT) and the four most accurate classifiers (LDA, LGR, MLPC, and XGB). For average performance, balanced diagnostic accuracy and AUC evaluation according to Table 2 in 10-fold cross-validation. The ET + NCR/NET tumor region of T1-Gd modality achieved the best diagnostic performance with a balanced accuracy of 0.901 and AUC of 0.953.

For each tumor subregion, T1-Gd had the highest average balance accuracy. For each MRI modality, the ET + NCR/NET tumor region showed the highest average balance accuracy.

IV. DISCUSSION

Radiomics is an emerging and rapidly growing field that converts medical images into quantitatively usable data [44].

In this study, we investigated different methods for selecting and classifying radiomics features to evaluate their inconsistent performance in predicting glioma malignancy.

Additionally, other controllable variables, such as number of selected features, type of features, MRI modality, and tumor ROI, are also discussed for optimal radiological prediction of glioma grade. We found that the L1-SVM + MLPC machine learning strategy achieved the best prediction performance in both iterative 10-fold cross-validation mode and percentage-split testing mode. Feature selection methods L1-SVM, LASSO, XGB, and GBDT showed valuable and balanced prediction accuracy and AUC performance for most classifiers. Meanwhile, among the classifiers, XGB, LDA, LGR, and MLPC showed higher stability for most feature selection methods. Regarding the extraction of radiomics features, the ET + NCR/NET region in T1-Gd modality provided important tumor-related phenotypes in glioma grading.

Sixteen feature selection methods and 15 classification methods were investigated for radiomics-based glioma grade prediction. Our results showed that the L1-SVM feature selection method combined with MLPC classification method had the highest diagnostic performance than other crossover methods in predicting the malignancy of glioma.

Then, among the feature selection methods, L1-SVM,

compared to his other two boosting methods, XGB and GBDT. We note that there are few studies comparing RF methods with high predictive performance classifiers for radiomics-based clinical prediction that we previously provided in the article. Furthermore, by using a cross-combination strategy for feature selection and classification method comparison, we were able to identify the most suitable radiomics-based framework for

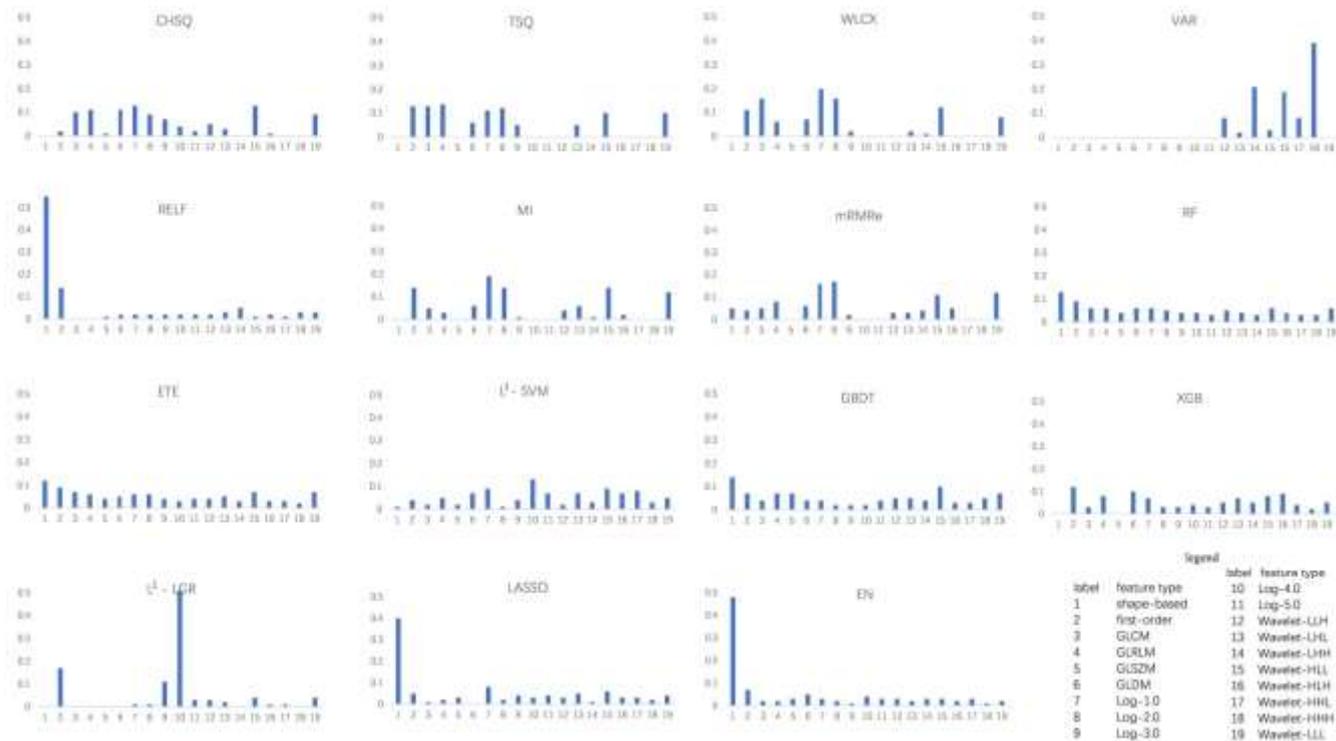


FIGURE 6. The NFTI coefficients of selected feature in each feature type for the fifteen feature selection methods.

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

LASSO, XGB, and GBDT showed relatively high predictive performance for most classifiers. It is noted that LASSO has previously been shown to be an efficient strategy for feature selection [26]. Among the classification methods, MLPC, LDA, LGR, and XGB performed better than other classifiers in most cases when combined with specific selected features.

In a previous evaluation of filter-based feature selection strategies for radiomics analysis, WLCX achieved satisfactory results in NSCLC survival prediction [28].

In our study, WLCX performed better than other filter feature selection methods, but worse than embedding feature selection methods. RF classifiers are widely used in machine learning [45] and have been proven to be efficient and powerful tree-based classification algorithms [30]. However, in this study, the prediction performance was not outstanding

glioma staging. Cross-combination strategies may help provide a framework for future radiomics-based analyses.

Only a few studies have investigated and compared the number of features and types of features in radiomics-based studies [46], [47].

In our study, the diagnostic performance of most feature selection and classification methods was consistent with the corresponding methods used in previous results [28]. The best performing characteristic numeric interval for predicting glioma grade was approximately 40–80. On the other hand, when the number of features exceeds 100, it tends to decrease slightly. This may indicate that the prediction results benefited from feature selection. In other words, feature selection is an effective strategy to improve radiomics-based prediction studies. Feature type analysis showed that features selected by

high-precision feature selection methods have relatively richer coverage feature types than low-precision feature selection methods. This may be because different feature types contain different tumor features [48] and a comprehensive feature extraction strategy is likely to improve the prediction of clinical outcomes.

These results provide important feature extraction aspects for feature selection and classification, and thus for overall clinical analysis. The results of MRI modality and tumor region analysis showed that the ET + NCR/NET region of T1-Gd modality achieved the most valuable tumor heterogeneity. This improved his diagnosis of GBM and LGG when using MRI-based non-invasive and cost-effective radiomics glioma grading. ET + NETs are routinely treated as solid tumor regions and have been widely considered as ROIs for radiomics feature extraction in many previous studies [4], [49]. Our results further support that solid tumor regions are more associated with tumor grade than other tumor subregions or their combinations. Furthermore, comparing the diagnostic performance results, we found that the diagnostic value of the combination of radiomics features in all multiparametric MRI modalities exceeds that of a single MRI modality.

This study also had some limitations. First, only four MRI modalities were used in this study. Quantitative imaging has benefited from innovations and advances in medical imaging hardware, imaging agents, standard protocols, and image analysis strategies. Newer parametric MRI modalities such as apparent diffusion coefficient (ADC) [50] and diffusion kurtosis imaging (DKI) [10] also have great potential for glioma grading, although they were not investigated in our study. It shows gender. Second, the predictive value of radiomics for other clinical glioma outcomes, such as progression survival, overall survival, and surgical recurrence, is expected to be discussed in future studies. Third, deep learning, one of the most recent representative developments in radiomics analysis, is not discussed in this study [51], [52]. Unlike traditional image features, deep imaging capabilities may have other diagnostic radiomics capabilities for glioma grading, and deep learning strategies can provide accurate classification of glioma grades.

may contribute to.

Finally, his MRI scans of glioma patients were collected from multiple institutions, increasing the heterogeneity of image quality. Future radiomics research should consider standardizing image acquisition and reconstruction or feature harmonization methods such as ComBat [53].

V. CONCLUSIONS

In conclusion, our study compared different radiomics strategies in terms of glioma grading. After comparing feature selection and classification methods, we found that the combination of L1-SVM feature selection and MLPC classification methods provided the highest diagnostic performance. On the other hand, feature selection methods such as L1-SVM, LASSO, XGB, and GBDT and feature classification methods such as LDA, MLPC, LGR, and XGB showed high diagnostic performance. Regarding the extraction of radiomics features, the ET + NCR/NET region in the T1-Gd modality may provide important tumor-related phenotypes in tumor grading, but other MRI modalities and tumor regions also causes heterogeneity. Our comparative study could serve as an important reference in identifying reliable and effective machine learning methods for radiology-based diagnostic analysis in non-invasive grading of gliomas. there is.

ACKNOWLEDGMENT

The authors would like to thank Drs. Wu-Tao Lu and Dr. Hongbao Li for reading and critiquing this manuscript.

You declare that you have no conflict of interest.

REFERENCES

- [1] D. Ricard, A. Idbay, F. Ducray, M. Lahutte, K. Huang Xuan and J.-Y. Delattre, "Primary Brain Tumors in Adults," *Lancet*, vol. 379, No. 9830, 1984-1996, May/June. 2012.
- [2] D. N. Lewis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "2016 World Health Organization Central Nervous System Tumor Classification: Overview," *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. Hoebbers, 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. Hoebbers, 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. ZÄüpfel, 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. Sofiras, 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, "Scikitlearn: 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. Upadhyaya, 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. Denoux, 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 Ruyscher, 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 School of Control Science and Engineering, Shandong University, Jinan, China, in 2013, and the M.S. In 2016, she graduated from the Department of Biomedical Engineering and Instrumentation Science, Zhejiang University, Hangzhou, China. She is currently pursuing her Ph.D.in. She received her degree from the Department of Medical Therapeutic Medicine

at the Chinese University of Hong Kong in China. Her research interests include medical image processing and machine learning.



DEFENG WANG received the B.S.degree from the School of Computer Science, Jilin University, and a Ph.D.She graduated from the Department of Computing, Hong Kong Polytechnic University.

He is currently a professor at Peking University School of Instrumentation Science and Optoelectronic Engine and Beijing Advanced Innovation Center for Big Database Precision Medicine. His research interests include medical imaging, statistical morphometric analysis, fMRI

paradigm design and post-processing, quantitative medical image analysis, and computational life sciences.



VINCENT CT MOK is currently a professor in the Department of Imaging and Interventional Radiology, Chinese University of Hong Kong.

He is the chair of the Department of Neurology and associate dean of the School of Medicine.

For his research on cerebrovascular diseases, he received the Higher Education Excellence Scientific Research Achievement Award (First Class) in Natural Sciences from the Ministry of Education of China in 2011, and the Excellence Research Award from the

Food and Health Bureau of the Hong Kong Special Administrative Region did. In 2017, he received the award for his research into vascular cognitive impairment. He has won the Teacher of the Year Award and the Master Teacher Award seven times.



He received Hong Kong Special Administrative Region Food and Health Department Research Award in 2017 for his research on vascular cognitive disorders. He has won the Teacher of the Year Award and the Master Teacher Award seven times. LIN SHI received his Ph.D. In 2008 he received his degree from

the Chinese University of Hong Kong. He is currently an assistant professor in the Department of Diagnostic Imaging and Interventional Radiology at the Chinese University of Hong Kong. Her research interests include neuroimaging, image analysis, the development of new image processing tools, and their application to facilitate the detection of imaging biomarkers in various neurological diseases.