Predicting Isocitrate Dehydrogenase (IDH) Mutation Status in Gliomas Using Multiparameter MRI Radiomics Features

Hong Peng, MD,[1,2][†] Jiaohua Huo, BD,[3][†] Bo Li, BS,[4] Yuanyuan Cui, MD,[1,2]

Hao Zhang, MS,[1,2] Liang Zhang, PhD,[3]* and Lin Ma, PhD, MD[1,2]*

Background: Accurate and noninvasive detection of isocitrate dehydrogenase (IDH, including IDH1 and IDH2) status is clinically meaningful for molecular stratification of glioma, but remains challenging.
Purpose: To establish a model for classifying IDH status in gliomas based on multiparametric MRI.
Study Type: Retrospective, radiomics.
Population: In all, 105 consecutive cases of grade II–IV glioma with 50 IDH1 or IDH2 mutant (IDHm) and 55 IDH wildtype (IDHw) were separated into a training cohort ($n = 73$) and a test cohort ($n = 32$).
Field Strength/Sequence: Contrast-enhanced T_1-weighted (CE-T_1W), T_2-weighted (T_2W), and arterial spin labeling (ASL) images were acquired at 3.0T.
Assessment: Two doctors manually labeled the volume of interest (VOI) on CE-T_1W, then T_2W and ASL were coregistered to CE-T_1W. A total of 851 radiomics features were extracted on each VOI of three sequences. From the training cohort, all radiomics features with age and gender were processed by the Mann–Whitney U-test, Pearson test, and least absolute shrinkage and selection operator to obtain optimal feature groups to train support vector machine models. The accuracy and area under curve (AUC) of all models for classifying the IDH status were calculated on the test cohort. Two subtasks were performed to verify the efficiency of texture features and the Pearson test in IDH status classification, respectively.
Statistical Tests: The permutation test with Bonferroni correction; chi-square test.
Results: The accuracy and AUC of the classifier, which combines the features of all three sequences, achieved 0.823 and 0.770 ($P < 0.05$), respectively. The best model established by texture features only had an AUC of 0.819 and an accuracy of 0.761. The best model established without the Pearson test got an AUC of 0.747 and an accuracy of 0.719.
Data Conclusion: IDH genotypes of glioma can be identified by radiomics features from multiparameter MRI. The Pearson test improved the performance of the IDH classification models.
Level of Evidence: 4
Technical Efficacy Stage: 1

GLIOMA is a devastating brain tumor with poor prognosis, uncertain pathogenesis, various biological characters, and low median survival time.[1] Previous studies have shown that the mutation status of isocitrate dehydrogenase (IDH) is a critical factor for the diagnosis, treatment, and prognosis of gliomas[2,3]; preoperative identification of the IDH mutation status would be of great clinical significance in selecting potential patients. It benefits chemotherapy, and thus assists with planning pf the therapy regime.[4,5] In the 2016 World Health Organization (WHO) classification of central nervous system (CNS) tumors, IDH was identified as a basic biomarker for subtyping of gliomas,[6] as well as an important clue for prognosis. However, confirmation of IDH mutation status requires a tissue sample, usually obtained through surgery

or biopsy, which has the risk of inducing neurological deficits and reducing quality of life.[7,8] Magnetic resonance imaging (MRI) is a routine and noninvasive technique for evaluating gliomas. However, the sensitivity and specificity for identification of IDH mutation by MRI ranges from 56–100% and from 51–100%, respectively.[9] Therefore, more reliable methods to predict IDH mutation status in gliomas by MRI are desired.

Previous studies have evaluated the performance of various machine-learning algorithms on predicting the genotypes of gliomas.[10–12] High-throughput features derived from MRI have been shown to have a great advantage that helps to effectively predict the classification of IDH.[13,14] The ability of features to noninvasively predict –IDH mutation status of glioma from a single MR sequence, such as T_2 weighted imaging has been shown,[15] Dynamic contrast-enhanced MRI[16] and dynamic susceptibility contrast-enhanced MRI[17] have been evaluated. Combinations of features from multiple sequences have been shown to be more reliable to predict the genotypes related to the occurrence of gliomas.[18,19] Furthermore, deep-learning models were developed to predict IDH mutation status preoperatively based on MRI data.[20,21]

Thus, the purpose of this study was to investigate the utility of a radiomics signature based on multiparametric MRI as a preoperative and noninvasive biomarker of IDH status in gliomas.

Materials and Methods
Patients
This study was approved by the Ethics Committee of our hospital, and written informed consent was obtained from all patients. We retrospectively collected clinical information and the MRI data of all patients from January 2015 to December 2018, with pathologically diagnosed (after resection) grade II–IV gliomas according to the WHO Classification of Tumors of the Central Nervous System.[6]

The inclusion criteria were as follows: 1) patients with available pathological analysis report; 2) patients with preoperative MRI data; and 3) over 18 years old. A total of 208 patients were included. Then 103 patients were excluded due to the following conditions: i) patients lack of IDH gene expression status assessed by EnVision immunohistochemical staining (DAKO, Hamburg, Germany) with a standard protocol developed by the German Cancer Research Center ($n = 41$); ii) patients lacked at least one of the following MRI sequences: contrast-enhanced T_1-weighted images (CE-T_1W), T_2-weighted images (T_2W), and arterial spin labeling images (ASL) ($n = 53$); and iii) the MR images that had motion or other kinds of artifacts that may affect subsequent segmentation and analysis ($n = 9$). Finally, 105 subjects met the requirements and were included in this study (50 cases of IDHm [mutant] and 55 cases of IDHw [wildtype]). The clinical information (age and sex) and tumor characteristics (grade and location) were summarized in Table 1.

The whole cohort was randomly split into a training cohort (70%, $n = 73$) and a test cohort (30%, $n = 32$).

MRI Acquisition
MR images were acquired with a 3T MRI system (Discovery 750; GE Healthcare, Milwaukee, WI) using a receive-only 32-channel phased-array head coil. MRI included T_2W fast spin-echo (FSE), CE-T_1W FSE-IR (inversion recovery), and ASL. T_2W was obtained with repetition time / echo time (TR/TE) = 4252/103.7 msec, field of view (FOV) = 24 × 24 cm, matrix = 192 × 192, NEX (number of excitations) = 1.5; CE T_1W (TR/TE/TI [inversion time] =1750/24/780 msec, FOV = 24 × 24 cm, matrix = 320 × 320, NEX = 1) was acquired 2 minutes after intravenous administration of contrast agents (Magnevist, 0.1 mmol/kg, Bayer HealthCare Pharmaceuticals, Wayne, NJ). These images were obtained with identical section thickness (5 mm) and section space (1.5 mm). ASL was acquired using a background-suppressed 3D spiral FSE technique. The parameters were as follows: TR/TE = 4844/10.5 msec, post-labeling delay = 2025 msec, FOV = 24 × 24 cm, section thickness = 4.0 mm, number of sections = 36, NEX = 3.

Data Preprocessing
A volume of interest (VOI) was manually segmented on the CE-T_1W slice-by-slice by two radiologists (P.H. and L.B., 11 and 15 years of experience, respectively). Areas of edema and necrosis were excluded (Fig. 1a,b) so that only tumor was included (Fig. 1, green areas). The T_2W and the ASL perfusion images were co-registered to the CE-T_1W for each case using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). The perfusion images were automatically calculated by an MRI scanning system.

Finally, image intensity values were standardized using PyRadiomics,[22] an open source toolkit for medical image feature extraction. The standardization method for one image: first calculating the mean value and variance of image pixels, then each pixel value of the image is subtracted by the mean value and divided by the variance (Fig. 2).

Feature Extraction
Feature extraction was performed using PyRadiomics. For each image set in the training cohort, a total of 851 radiomics features were extracted from each VOI. The definitions and calculation formulas of all radiomics features are available on the Pyradiomics's website (https://pyradiomics.readthedocs.io/en/latest/features.html). Fourteen shape features were extracted based on the shape of the VOI. Eighteen first-order statistical features and 75 texture features were extracted from nine kinds of images (the original image and eight wave-let transform images). The 14 first-order statistics features describe the distribution of voxel intensities within the image region defined by the mask through commonly used and basic metrics. The texture features include 24 gray level cooccurrence matrix (GLCM) features, 16 gray level run length matrix (GLRLM) features, 16 gray level size zone matrix (GLSZM) features, five neighboring gray tone difference matrix (NGTDM) features, and 14 gray level dependence matrix (GLDM) features. The 851 radiomics features (14 + (18 + 24 + 16 + 16 + 5 + 14) × 9) plus age and gender, with a total of 853 dimensional features, were used for the classifier training. The age was used as a continuous variable and the gender was a categorical variable.

TABLE 1. Clinical and Tumor Characteristics of the Whole Cohort

	Total cases	Mutated	Wildtype	P
Case numbers	105	50 (47.6%)	55 (52.4%)	
Clinical characteristics				
Sex				0.706
Male	61 (58.1%)	30 (60.0%)	31 (56.4%)	
Female	44 (41.9%)	20 (40.0%)	24 (43.6%)	
Age				0.005
18–30	17 (16.2%)	9 (18.0%)	8 (14.5%)	
31–60	74 (70.5%)	40 (80.0%)	34 (61.8%)	
>60	14 (13.3%)	1 (2.0%)	13 (23.7%)	
Mean ± SD	45.4 ± 13.0	42.1 ± 11.3	48.5 ± 13.8	
Grade				0.000
II	45 (42.9%)	34 (68.0%)	11 (20.0%)	
III	16 (15.2%)	8 (16.0%)	8 (14.5%)	
IV	44 (41.9%)	8 (16.0%)	36 (65.5%)	
Tumor location				0.005
Frontal lobe	32 (30.6%)	25 (50.0%)	8 (14.6%)	
Temporal lobe	18 (17.1%)	4 (8.0%)	10 (18.2%)	
Parietal lobe	6 (5.7%)	3 (6.0%)	2 (3.6%)	
Occipital lobe	1 (0.9%)	0 (0%)	1 (1.8%)	
Central area	21 (20.0%)	7 (14.0%)	9 (16.4%)	
Cerebellum	2 (1.9%)	0 (0%)	1 (1.8%)	
Two or more	25 (23.8%)	11 (22.4%)	24 (43.6%)	

Feature Selection

A group of optimal features for each MR sequence was selected based on the training cohort before the training process. First, a Mann–Whitney U-test was conducted on the features, and the features with differences ($P < 0.05$) between two categories were preserved. Next, a Pearson test was used to assess the correlation between features and categories, and features with $P < 0.05$ were selected as potentially predictive. Finally, the LASSO (least absolute shrinkage and selection operator)[23] with default parameters was carried out, and the features with nonzero weight were selected as the optimal feature groups.

Classifier Modeling

The modeling of the support machine (SVM) was based on a machine-learning toolkit of scikit-learn (https://scikit-learn.org/stable/about.html#citing-scikit-learn). For the optimal feature groups from three sequences, the model with the best performance was chosen from four SVM classifiers, including default parameters SVM (def), SVM with radial basis function kernel (RBF), with linear kernel (linear), and with poly kernel (poly), which were built to predict IDH genotype, to represent them. To avoid overfitting, the parameters of the three kernel function models were determined using 5-fold cross-validation and grid search.

In addition, we combined the optimal feature groups of each sequence into a set of multiparametric MRI features to train classifiers, and experiments were performed on all possible combinations of two or three sequences. The training process of multiparametric MRI features was the same as that of single-sequence features. It should be noted that, for a single sequence in the same case, the age, gender, and shape features were the same, thus these repeated features were eliminated before training. This process is referred to as Method 1, below.

Classifier Modeling With Only Texture Features

In our study texture features accounted for 79.1% (75*9/853) features in one sequence. In this part of the study, only texture features (all features except age, gender, shape features, and first-order statistical features) were sent to three stages of feature selection
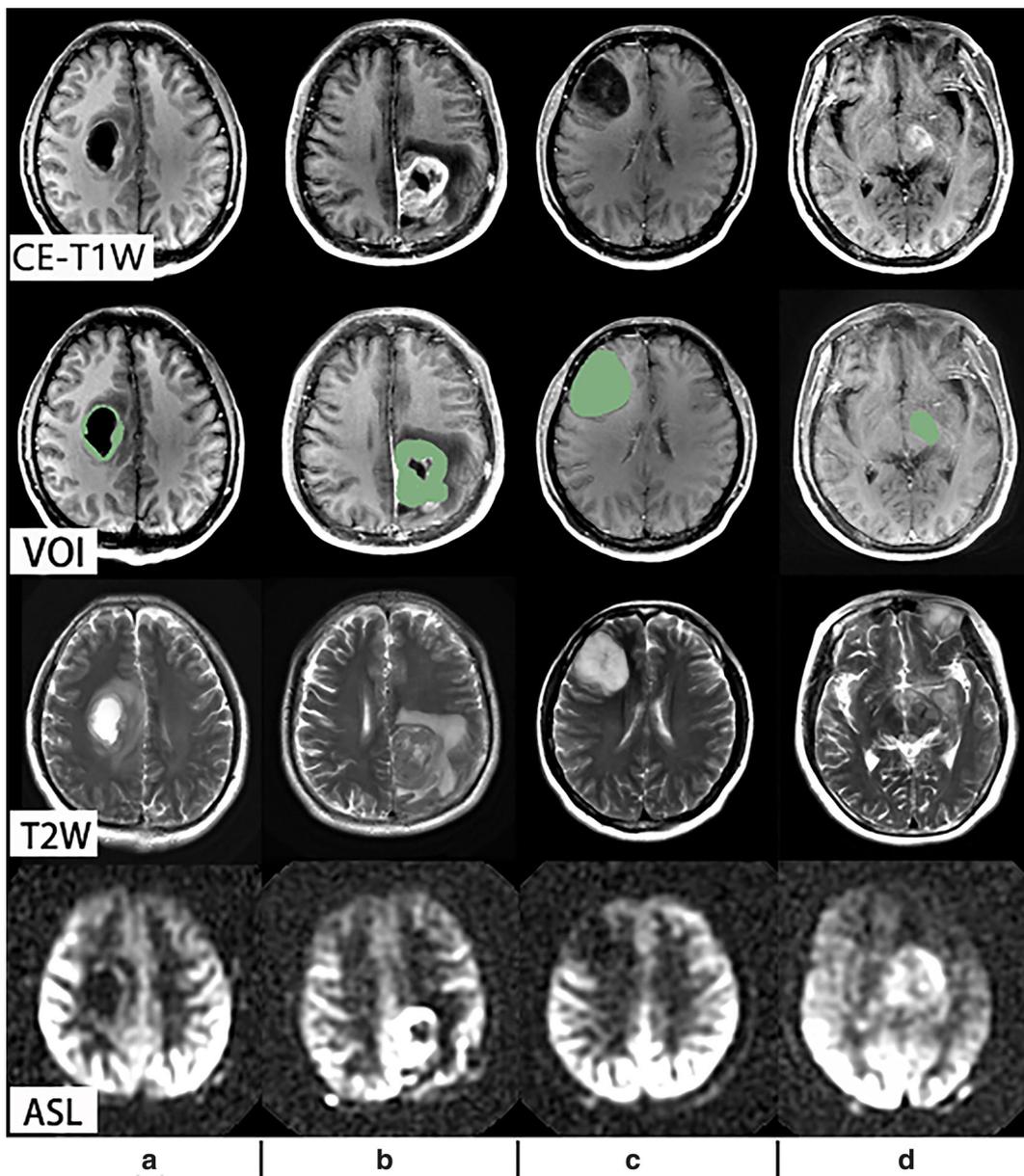
FIGURE 1: Examples of volume of interest (VOI) segmentation. The green areas represent the regions of interest segmented on CE-T₁W for (a) IDH1 or IDH2 mutant (IDHm) of high-grade glioma, (b) IDH wildtype (IDHw) of high-grade glioma, (c) IDHm of low-grade glioma, and (d) IDHw of low-grade glioma.

(Mann–Whitney U-test, Pearson test, and LASSO) to explore the predictive efficacy of texture features compared with that of all features. This process is referred to as Method 2, below.

Classifier Modeling Without a Pearson Test

To verify the effect of the Pearson test on the classification results, we designed an additional experiment. Specifically, as in Method 2, only texture features were used for training, but the optimal feature group was selected by Mann–Whitney U-test and LASSO. This process is referred to as Method 3.

Statistical Analysis

The chi-square test was performed using IBM SPSS Statistics (v. 22.0; Armonk, NY) to analyze whether there were statistically significant differences in clinical and tumor information between the IDHm and IDHw groups. The Mann–Whitney U-test and Pearson test were performed using SciPy[24] for feature selection, an open source mathematical, scientific, and engineering computing tool. The level of confidence for all the statistical analyses mentioned above was kept at 95% and results with $P < 0.05$ were significant.

To evaluate the models' ability to distinguish IDHm and IDHw, the values of accuracy (ACC), sensitivity (SEN), specificity (SP), and AUC in the test cohort were calculated using the pathological diagnosis of IDH status as the reference standard.

The permutation test was used to analyze whether the optimal feature group was an effective tool for distinguishing IDHm and IDHw or just approximate a random guess. Since we compared the performance of different models based on ACC, the ACC of the test set was assessed by the permutation test with 1000 epochs,[25] and Bonferroni correction was applied to the result of permutation test.
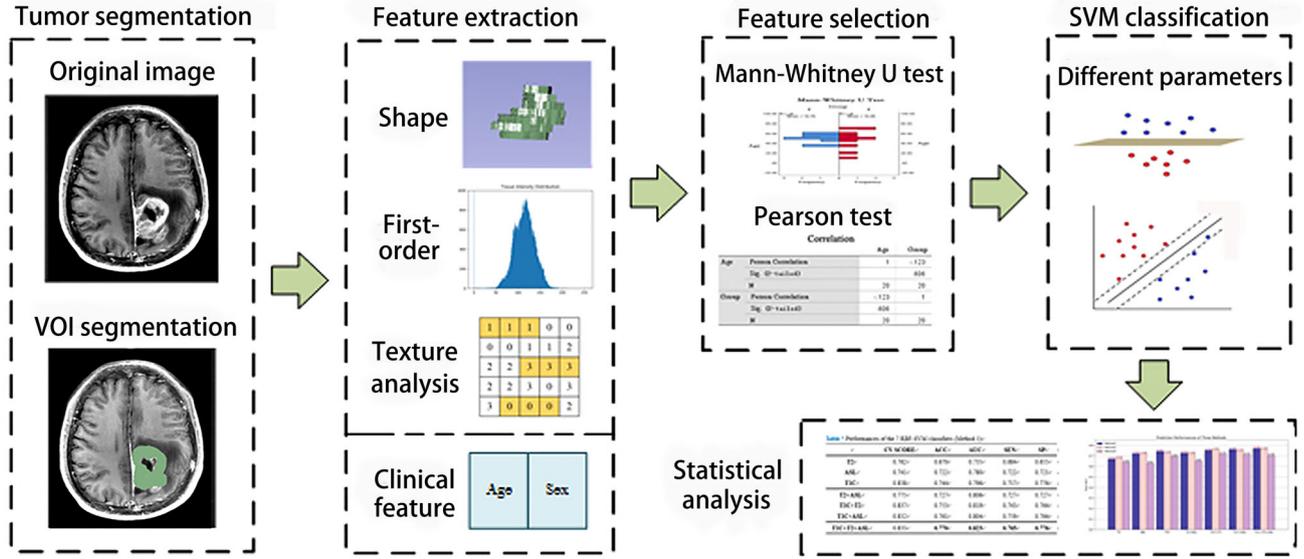
FIGURE 2: The pipeline for IDH status prediction. The VOIs were segmented on CE-T_1W. T_2W and ASL sequences needed to be registered to CE-T_1W after "Tumor segmentation."

When $P < 0.05$ for the permutation test, there was a significant difference between the real ACC and the random guess ACC. After Bonferroni correction, the statistical significance level should be $0.05/11 \approx 0.005$, where 11 is the number of repeated experiments.

Due to the small size of the dataset and the random way of split, this may lead to the high variance of classification results. Therefore, the randomly split and model training and testing process were repeated 11 times, all evaluation indexes of the test set were calculated for each split, and models were evaluated by the average value of 11 indexes. The models with the same average value were evaluated by standard deviation of 11 indexes. The smaller the standard deviation, the more stable the model.

Results

Clinical Characteristics of the Study Cohort
No significant difference was found in gender ($P = 0.706$). However, there were significant differences in age, grade, and tumor location between IDHm and IDHw. See Table 1.

Performance of the Classification Model
After feature selection, the number of remaining features belonging to four categories (clinical features, shape features,

first-order statistical features, and texture features) for the three MR sequences were counted (Table 2). For all sequences, texture features accounted for the majority (77.6%, 85.0%, and 81.5% for T_2W, ASL, and CE-T_1W, respectively); the number and categories of texture features varied greatly in 11 splits. The first-order statistical features are rarely found in the optimal feature groups for T_2W (4.0%) and ASL (0.8%), but represented a relatively large portion of the optimal features for CE-T_1W (13.9%). Only age was selected from clinical features and only the sphericity of VOI (sphericity is a measure of the roundness of the shape of the tumor region relative to a circle) was selected from shape features, but these two features appeared frequently in the feature selection results of T_2W and ASL. These two features appeared both seven times in 11 results of T_2W. In the 11 results of ASL, these two features appeared eight times and nine times, respectively.

By comparing the average performance of all classifiers, the best classifier is the one that used the RBF kernel and was trained by three sequences features (Fig. 3). The average ACC, AUC, SEN, SP, and P-value of seven RBF-SVM models in predicting IDH mutation status are listed in Table 3. The average AUC was 0.823 (95% confidence

---

TABLE 2. In Method 1, the Frequency of Four Types of Features in 11 Optimal Feature Groups of Three Sequences

	Clinical	Shape	First-order	Texture
T_2W	7 (9.2%)	7 (9.2%)	3 (4.0%)	59 (77.6%)
ASL	8 (6.7%)	9 (7.5%)	1 (0.8%)	102 (85.0%)
CE-T_1W	2 (1.8%)	3 (2.8%)	15 (13.9%)	88 (81.5%)

Some features appeared repeatedly in 11 splits. The data in parentheses are percentages.
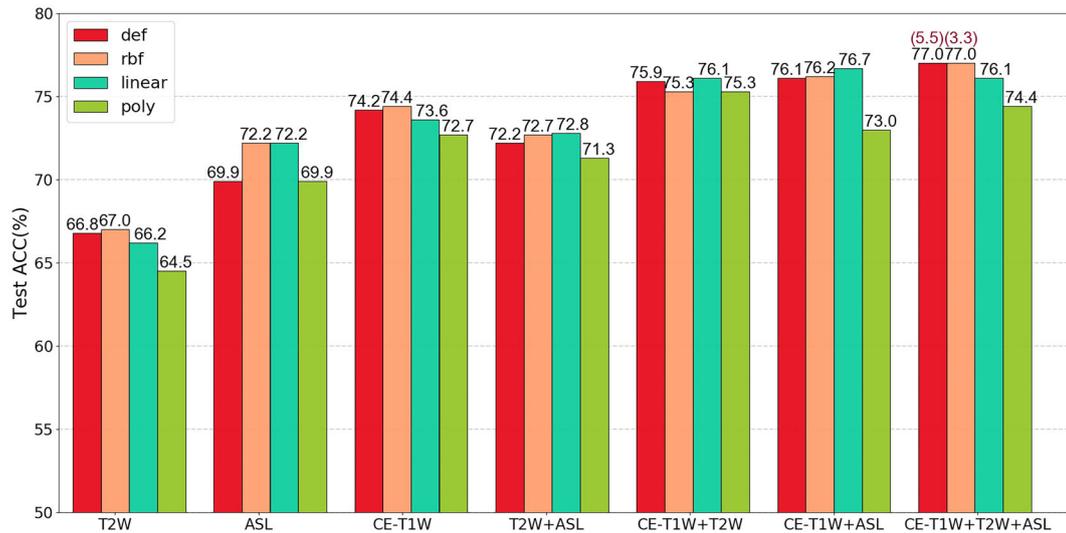
FIGURE 3: The prediction performances of all models in Method 1. The black numbers on the columns are the ACCs, and the red data are the standard deviations. The "def," "rbf," "linear," and "poly" represent four kinds of SVM. The RBF-SVM which trained by features from all three sequences had the highest accuracy with the lowest standard deviation.

interval [CI]: 0.715, 0.931), and the average ACC was 0.770 (95% CI: 0.704, 0.835) with a SEN of 0.765 (95% CI: 0.562, 0.967) and SP of 0.776 (95% CI: 0.626, 0.926) (Table 3). The parameters and classification indexes of all models are given in Supplement Tables S1 and S4. Among all the single sequences, CE-T$_1$W had the highest ACC of 0.744 (95% CI: 0.667, 0.822) and AUC of 0.796 (95% CI: 0.679, 0.913) with SEN of 0.717 (95% CI: 0.534, 0.899) and SP of 0.776 (95% CI: 0.636, 0.915) (Fig. 4).

Among the average results of the permutation test (*P*-value, Table 3), the models trained by features from CE-T$_1$W, CE-T$_1$W + T$_2$W, CE-T$_1$W + ASL, and CE-T$_1$W + T$_2$W + ASL, respectively, can effectively distinguished the two types of IDHs (*P*-value < 0.005). The model trained by features from three sequences had the lowest *P*-value of 0.001. In addition, the values of *P* for T$_2$W, ASL, and T$_2$W + ASL were 0.017, 0.009, and 0.006, respectively.

TABLE 3. Performances of the Seven RBF-SVM Classifiers (Method 1)

	ACC[a]	*P*-value[b]	AUC[c]	Sen[d]	Sp[e]
T$_2$W[f]	0.670 (0.568, 0.773)	0.017	0.735 (0.641, 0.829)	0.684 (0.433,0.936)	0.655 (0.407, 0.903)
ASL[g]	0.722 (0.590, 0.853)	0.009	0.780 (0.635, 0.924)	0.722 (0.481, 0.963)	0.721 (0.547, 0.896)
CE-T$_1$W[h]	0.744 (0.667, 0.822)	0.002	0.796 (0.679, 0.913)	0.717 (0.534, 0.899)	0.776 (0.636, 0.915)
T$_2$W + ASL	0.727 (0.605, 0.850)	0.006	0.806 (0.688, 0.923)	0.727 (0.523, 0.932)	0.727 (0.556, 0.899)
CE-T$_1$W + T$_2$W	0.753 (0.681, 0.824)	0.002	0.819 (0.708, 0.930)	0.743 (0.594, 0.893)	0.764 (0.699, 0.829)
CE-T$_1$W + ASL	0.761 (0.653, 0.870)	0.003	0.804 (0.677, 0.931)	0.759 (0.548, 0.971)	0.764 (0.603, 0.925)
CE-T$_1$W + T$_2$W + ASL	**0.770** (0.704, 0.835)	**0.001**	**0.823** (0.715, 0.931)	**0.765** (0.562, 0.967)	**0.776** (0.626, 0.926)

Data in parentheses are 95% confidence intervals.
[a]Accuracy.
[b]The *P*-value of permutation test for the accuracy of test cohort.
[c]Area under receiver operating characteristic curve.
[d]Sensitivity.
[e]Specificity.
[f]T$_2$-weighted image.
[g]Arterial spin labeling image.
[h]Contrast-enhanced T$_1$-weighted image.
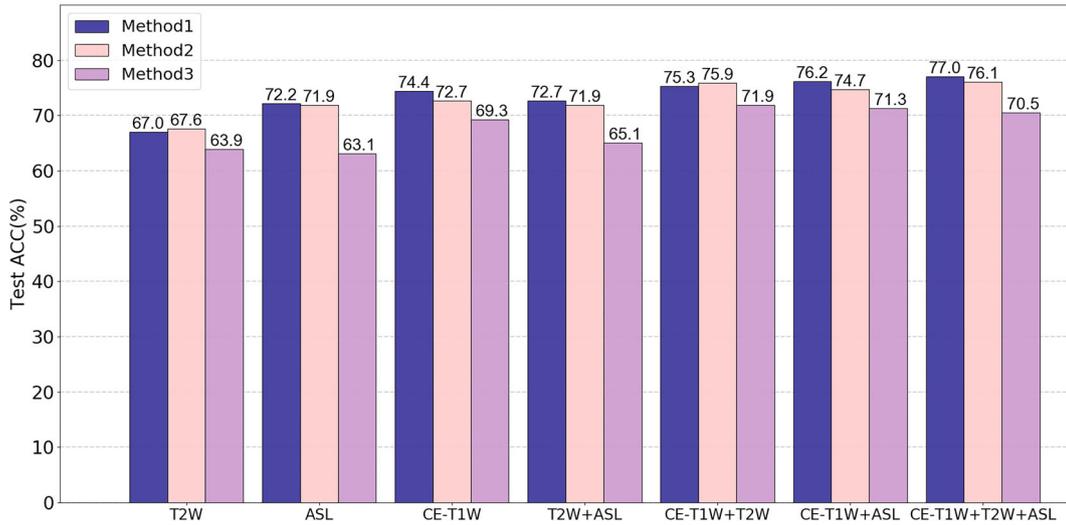
FIGURE 4: Visual comparison of the performances of the three methods. The black numbers on columns are the ACCs.

Performance of Classifier Only Using Texture Features

Table 4 shows the classification results of RBF-SVM models from Method 2. The other three kinds of SVM indexes (linear kernel, poly kernel, and default parameters) are presented in Supplement Table S2. The result accuracy with them are lower than that with the RBF kernel. For Method 2, the combined features of the three sequences (CE-T_1W + T_2W + ASL) had the best performance, with an ACC of 0.761 (95% CI: 0.696, 0.827) and AUC of 0.819 (95% CI: 0.762, 0.875).

Classifier Performance Without a Pearson Test

The classification results of RBF-SVM models from Method 3 are presented in Table 5. The classifier with the best

performance (ACC = 0.719, 95% CI: 0.605, 0.833) was CE-T_1W + T_2W. All classification indexes of this subtask are given in Supplement Table S3.

Discussion

We evaluated the ability of features from individual MR sequences to predict IDH mutation status. The performance of CE-T_1W features was the best, which was consistent across the three methods assessed in our study. As a noninvasive perfusion MRI technique, ASL obtains cerebral blood flow without administration of an exogenous contrast agent and provides clues to the prediction of molecular features.[10] In previous studies, ASLs were shown to be helpful in predicting IDH mutation.[26,27] This is consistent with our results. In the first

---

TABLE 4. Performances of the Classifiers With Only Texture Features (Method 2)

	ACC[a]	AUC[b]	Sen[c]	Sp[d]
T_2W[e]	0.676 (0.574, 0.778)	0.751 (0.593, 0.909)	**0.754** (0.468, 1)	0.588 (0.316, 0.860)
ASL[f]	0.719 (0.571, 0.866)	0.769 (0.604, 0.933)	0.727 (0.555, 0.899)	0.709 (0.458, 0.960)
CE-T_1W[g]	0.727 (0.674, 0.780)	0.766 (0.658, 0.873)	0.717 (0.579, 0.854)	0.739 (0.622, 0.857)
T_2W + ASL	0.719 (0.581, 0.857)	0.791 (0.654, 0.927)	0.743 (0.464, 1)	0.691 (0.496, 0.886)
CE-T_1W + T_2W	0.759 (0.657, 0.860)	0.799 (0.742,0.856)	0.743 (0.521, 0.965)	0.776 (0.597, 0.954)
CE-T_1W + ASL	0.747 (0.659, 0.836)	0.800 (0.727, 0.873)	0.738 (0.558, 0.918)	0.758 (0.555, 0.961)
CE-T_1W + T_2W + ASL	**0.761** (0.696, 0.827)	**0.819** (0.762, 0.875)	0.749 (0.564, 0.933)	**0.776** (0.589, 0.963)

[a]Accuracy.
[b]Area under receiver operating characteristic curve.
[c]Sensitivity.
[d]Specificity.
[e]T_2-weighted image.
[f]Arterial spin labeling image.
[g]Contrast-enhanced T_1-weighted image.

TABLE 5. Performances of the Classifiers With Two-Stage Feature Selection (Method 3)

	ACC[a]	AUC[b]	Sen[c]	Sp[d]
T_2W[e]	0.639 (0.559, 0.719)	0.698 (0.562, 0.834)	0.690 (0.367, 1)	0.582 (0.287, 0.877)
ASL[f]	0.631 (0.512, 0.750)	0.673 (0.517, 0.830)	0.690 (0.481, 0.899)	0.564 (0.300, 0.827)
CE-T_1W[g]	0.693 (0.583, 0.803)	0.727 (0.589, 0.864)	0.695 (0.519, 0.871)	0.691 (0.504, 0.878)
T_2W + ASL	0.651 (0.501, 0.800)	0.713 (0.483, 0.944)	0.668 (0.360, 0.977)	0.630 (0.405, 0.856)
CE-T_1W + T_2W	**0.719** (0.605, 0.833)	0.747 (0.633, 0.861)	0.722 (0.525, 0.919)	**0.715** (0.485, 0.945)
CE-T_1W + ASL	0.713 (0.597, 0.829)	**0.750** (0.626, 0.874)	**0.727** (0.578, 0.877)	0.697 (0.518, 0.876)
CE-T_1W + T_2W + ASL	0.705 (0.617, 0.793)	0.745 (0.633, 0.857)	0.717 (0.579, 0.854)	0.691 (0.551, 0.830)

[a]Accuracy.
[b]Area under receiver operating characteristic curve.
[c]Sensitivity.
[d]Specificity.
[e]T_2-weighted image.
[f]Arterial spin labeling image.
[g]Contrast-enhanced T_1-weighted image.

two methods of our study, ASL had a similar accuracy as CE-T_1W. It has been suggested that T_2W features play a more important role in distinguishing IDH mutation status than CE-T_1W[28]; however, that study counted the number of features belonging to each sequence in the optimal feature group, rather than training the single sequence classifiers separately.

It has been shown that classification performance can be improved by combining different sequences.[29] Our results showed that the ACCs and AUCs of all three sequences combined were higher than that of an individual sequence. Moreover, the classifier with the best performance and the lowest P-value is modeled by radiomics features from three sequences. This suggests that the image features of different sequences reflect a part of the visual features expressed by IDH and can complement each other.

Texture features are the most important feature type for glioma classification,[30] and have proved useful for predicting the IDH1-mutation status in gliomas.[31,32] In our study, texture features accounted for the largest portion in the optimal feature groups selected from each single sequence. Moreover, the model trained only by texture features also showed a good performance for identifying IDH mutation status. However, the models that considered all of the features (clinical, shape, first-order, and texture features) showed a much better diagnostic performance than assessments performed with only texture features. Although relatively few features were selected from clinical, shape, and first-order features, these were rather important for the complete description of IDH mutation status. This could be attributable to the significant correlation between these features and the incidence of IDH mutation.[33]

A large number of features extracted from the VOI can fully describe the properties of the tumor area, but they also include redundant and noise features that could increase the unnecessary consumption of computing resources and make a negative impact on the prediction efficiency. In order to improve the classification performance of our study, the Mann–Whitney U-test[34] was used to remove features with no statistical difference in the average value between two categories of IDHm and IDHw ($P > 0.05$); the Pearson test[35] was used to remove features with no significant correlation with categories ($P > 0.05$), and LASSO[23] was used to reduce redundant features.

A Pearson test is a commonly used statistical method for measuring the linear relationship between two variables.[35] The correlation analysis is typically used in genotypic diagnosis of glioma, either for result analysis[30] or for feature dimension reduction.[36] The Pearson test was applied for result analysis of IDH mutation status classification in a previous study,[37] and has not been used in feature selection. The performance of Method 3, which was the same as Method 2 but without the Pearson test, was lower than that of Methods 1 and 2. A potential reason may be that there is more noise in the features that have no linear relationship with the label and can disturb the modeling process. This suggests that the Pearson test should be included in the process of IDH mutation status prediction to obtain a better outcome in future studies.

Limitations

First, only three sequences were included in this study. Future studies may consider more sequences, such as diffusion-weighted imaging (DWI), diffusion tensor imaging (DTI), and so forth. And other critical biomarkers (such as 1p/19q, Ki67, P53) were not investigated. Second, our method has not been applied extensively in the classification of other histologic variables. Third, our model was trained and tested using relatively

small data collected from a single institution. A large-scale prospective and multicenter validation cohort collection should also be studied in future research.

Conclusion

In this study the classification model based on radiomics features from CE-T_1W, T_2W, and ASL, through three stages of feature selection (Mann–Whitney U-test, Pearson test, and LASSO), had the best performance in predicting the IDH status of gliomas. The result suggested that this method has promising prediction efficiency, and may become a helpful tool for the diagnosis, treatment, and prognostic prediction of gliomas.

REFERENCES

1. DeAngelis LM. Brain tumors. N Engl J Med 2001;344:114-123.

2. Wen PY, Kesari S. Malignant gliomas in adults. N Engl J Med 2008; 359:492-507.

3. Olar A, Wani KM, Alfaro-Munoz KD, et al. IDH mutation status and role of WHO grade and mitotic index in overall survival in grade II-III diffuse gliomas. Acta Neuropathol 2015;129:585-596.

4. Minniti G, Scaringi C, Arcella A, et al. IDH1 mutation and MGMT methylation status predict survival in patients with anaplastic astrocytoma treated with temozolomide-based chemoradiotherapy. J Neurooncol 2014;118:377-383.

5. Yu J, Shi Z, Lian Y, et al. Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. Eur Radiol 2017;27:3509-3522.

6. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: A summary. Acta Neuropathol 2016;131:803-820.

7. Walker DG, Kaye AH. Diagnosis and management of astrocytomas, oligodendrogliomas and mixed gliomas: A review. Australas Radiol 2001;45:472-482.

8. Sanai N, Martino J, Berger MS. Morbidity profile following aggressive resection of parietal lobe gliomas. J Neurosurg 2012;116:1182-1186.

9. Suh CH, Kim HS, Jung SC, Choi CG, Kim SJ. Imaging prediction of isocitrate dehydrogenase (IDH) mutation in patients with glioma: A systemic review and meta-analysis. Eur Radiol 2019;29:745-758.

10. Yoo RE, Yun TJ, Hwang I, et al. Arterial spin labeling perfusion-weighted imaging aids in prediction of molecular biomarkers and survival in glioblastomas. Eur Radiol 2020;30:1202-1211.

11. Kim M, Jung SY, Park JE, et al. Diffusion- and perfusion-weighted MRI radiomics model may predict isocitrate dehydrogenase (IDH) mutation and tumor aggressiveness in diffuse lower grade glioma. Eur Radiol 2020;30:2142-2151.

12. Liu Z, Zhang T, Jiang H, Xu W, Zhang J. Conventional MR-based preoperative nomograms for prediction of IDH/1p19q subtype in low-grade glioma. Acad Radiol 2019;26:1062-1070.

13. Tan Y, Zhang ST, Wei JW, et al. A radiomics nomogram may improve the prediction of IDH genotype for astrocytoma before surgery. Eur Radiol 2019;29:3325-3337.

14. Lee MK, Park JE, Jo Y, Park SY, Kim SJ, Kim HS. Advanced imaging parameters improve the prediction of diffuse lower-grade gliomas subtype, IDH mutant with no 1p19q codeletion: Added value to the T2/-FLAIR mismatch sign. Eur Radiol 2020;30:844-854.

15. Liu X, Li Y, Li S, et al. IDH mutation-specific radiomic signature in lower-grade gliomas. Aging 2019;11:673-696.

16. Lu HT, Xing W, Zhang YW, Qin HP, Wu RH, Ding JL. The value of DCE-MRI in predicting IDH gene mutation of high-grade gliomas. Zhonghua Yi Xue Za Zhi 2019;99:3105-3109.

17. Sudre CH, Panovska-Griffiths J, Sanverdi E, et al. Machine learning assisted DSC-MRI radiomics as a tool for glioma classification by grade and mutation status. BMC Med Inform Decis Mak 2020;20:149.

18. Lu CF, Hsu FT, Hsieh KL, et al. Machine learning–based radiomics for molecular subtyping of gliomas. Clin Cancer Res 2018;24:4429-4436.

19. Labreche K, Kinnersley B, Berzero G, et al. Diffuse gliomas classified by 1p/19q co-deletion, TERT promoter and IDH mutation status are associated with specific genetic risk loci. Acta Neuropathol 2018;135:743-755.

20. Matsui Y, Maruyama T, Nitta M, et al. Prediction of lower-grade glioma molecular subtypes using deep learning. J Neurooncol 2020;146:321-327.

21. Bangalore YCG, Shah BR, Yu FF, et al. A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas. Neuro Oncol 2020;22:402-411.

22. Griethuysen JJMV, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017; 77:104-107.

23. Tibshirani R. Regression shrinkage and selection via the LASSO. J R Stat Soc Ser B 1996;58:267-288.

24. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020; 17:261-272.

25. Golland P, Fischl B. Permutation tests for classification: Towards statistical significance in image-based studies. Inf Process Med Imaging 2003; 18:330-341.

26. Hong EK, Choi SH, Shin DJ, et al. Radiogenomics correlation between MR imaging features and major genetic profiles in glioblastoma. Eur Radiol 2018;28:4350-4361.

27. Yamashita K, Hiwatashi A, Togao O, et al. MR imaging-based analysis of glioblastoma multiforme: Estimation of IDH1mutation status. Am J Neuroradiol 2016;37:58-65.

28. Zhang X, Tian Q, Wang L, et al. Radiomics strategy for molecular subtype stratification of lower-grade glioma: Detecting IDH and TP53 mutations based on multimodal MRI. J Magn Reson Imaging 2018;48:916-926.

29. Xing Z, Zhang H, She D, et al. IDH genotypes differentiation in glioblastomas using DWI and DSC-PWI in the enhancing and peri-enhancing region. Acta Radiol 2019;60:1663-1672.

30. Li Y, Liu X, Qian Z, et al. Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. Eur Radiol 2018;28:2960-2968.

31. Han L, Wang S, Miao Y, et al. MRI texture analysis based on 3D tumor measurement reflects the IDH1 mutations in gliomas ⊠ A preliminary study. Eur J Radiol 2019;112:169-179.

32. Park YW, Han K, Ahn SS, et al. Whole-tumor histogram and texture analyses of DTI for evaluation of IDH1-mutation and 1p/19q-Codeletion status in World Health Organization grade II gliomas. Am J Neuroradiol 2018;39:693-698.

33. Shibahara I, Sonoda Y, Shoji T, et al. Malignant clinical features of anaplastic gliomas without IDH mutation. Neuro Oncol 2015;17:136-144.

34. Neuhuser M. Wilcoxon–Mann–Whitney test. In: Lovric M, editor. International encyclopedia of statistical science. Berlin: Springer; 2011. p 1649-1659.

35. Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. Am Stat 1988;42:59-66.

36. Kim D, Wang N, Ravikumar V, et al. Prediction of 1p/19q codeletion in diffuse glioma patients using pre-operative multiparametric magnetic resonance imaging. Front Comput Neurosci 2019;30:52.

37. Yan R, Xi Z, Wenting R, et al. Noninvasive prediction of IDH1 mutation and ATRX expression loss in low-grade gliomas using multiparametric MR radiomic features: MRI radiomics of IDH1 and ATRX in LGG. J Magn Reson Imaging 2018;49:808-817.