

Differentiation Between Ependymoma and Medulloblastoma in Children with Radiomics Approach

Jie Dong, PhD candidate, Lei Li, PhD, Shengxiang Liang, PhD, Shujun Zhao, PhD, Bin Zhang, PhD, Yun Meng, MS, Yong Zhang, MD, Suxiao Li, MS

Rationale and Objectives: Ependymoma (EP) and medulloblastoma (MB) of children are similar in age, location, manifestations and symptoms. Therefore, it is difficult to differentiate them through visual observation in clinical diagnosis. The aim of this study is to investigate the effectiveness of radiomics and machine-learning techniques on multimodal magnetic resonance imaging (MRI) in distinguish EP from MB.

Materials and Methods: Three dimensional (3D) tumors were semi-automatic segmented by radiologists from postcontrast T1-weighted images and apparent diffusion coefficient maps in 51 patients (24 EPs, 27 MBs). Then, we extracted radiomics features and further reduced them by three feature selection methods. For each feature selection method, 4 classifiers were adopted which yield 12 different models. After extensive crossvalidation, pairwise test were carried out in receiver operating characteristic curves to explore performance of these models.

Results: The radiomics model built with multivariable logistic regression as feature selection method and random forests as classifier had the best performance, area under the curve achieved 0.91 (95 % confidence interval 0.787–0.968). Five relevant features were highly correlated to discriminate EP and MB, which may used as imaging biomarkers to predict the kinds of tumors.

Conclusion: The combination of radiomics and machine-learning approach on 3D multimodal MRI could well distinguish EP and MB of childhood, which assistant doctors in clinical diagnosis. Since there is no uniform model to obtained best performance for every specific data set, it is necessary to try different combination methods.

Key Words: Pediatric posterior fossa tumors; Ependymoma; Medulloblastoma; Machine-learning; Radiomics.

© 2020 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

INTRODUCTION

In children under 15 years of age, brain tumor is the second leading cause of death after acute lymphoblastic leukemia (1). It is reported that more than 1500 children in America per year and 1859 children in Britain per year (2014–2016) were diagnosed with cancer, 15% of whom subsequently died (2,3). About 55%–70% are pediatric posterior fossa tumor (PPFT). Among which, about 15% brain

tumors are medulloblastoma (MB), 4% tumors are ependymoma (EP) (2). There are many similarities between EP and MB. First, EP and MB are usually in the fourth ventricle, which are similar in location (2). Second, EP and MB are similar in terms of morphological. Magnetic resonance imaging (MRI) intensities of EP and MB appear heterogeneous enhancement of the solid portion due to coarse/gross calcification, hemorrhage, or necrosis (4). Third, EP and MB are similar in terms of symptoms and signs. Tumors in the posterior fossa are characterized by headache and vomiting, followed by unsteadiness, double vision, somnolence, irritability, and papilledema. Besides, EP and MB often show morning headaches because they are usually in the fourth ventricle near the vomiting area (5). Therefore, it is difficult to distinguish them in clinical diagnosis.

At present, the gold diagnosis standard of tumor classification is still pathological analysis, following biopsy or surgical resection. However, there are also disadvantages, including sampling error and variability in interpretation (6). Besides, biopsies may cause high-risk morbidity and mortality (7–10). The purpose of preoperative imaging is to establish a

Acad Radiol 2020; ■:1–10

From the School of Physics and Microelectronics, Zhengzhou University, NO.101 Kexue Road, Zhengzhou 450001, PR China (J.D., L.L., S.Z., B.Z., S.L.); National-Local Joint Engineering Research Center of Rehabilitation Medicine Technology, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian, PR China (S.L.); Traditional Chinese Medicine Rehabilitation Research Center of State Administration of Traditional Chinese Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian, PR China (S.L.); Department of Magnetic Resonance, Zhengzhou University First Affiliated Hospital, Zhengzhou, Henan PR China (Y.M., Y.Z.). Received November 16, 2019; revised January 31, 2020; accepted February 13, 2020. **Address correspondence to:** S. L. e-mail: hnsuxiao@zzu.edu.cn

© 2020 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.
<https://doi.org/10.1016/j.acra.2020.02.012>

diagnosis, differential diagnosis, or to describe the anatomical location of the tumor.

Multimodal MRI combines multiple sequences to provide additional characteristics at the tissue level, which improve the performance of classification, grading, survival prediction, and treatment response (11–15). Diffusion-weighted imaging (DWI) provides quantitative evaluation of water diffusion in brain tumors and reflects the formation of tumor cells. Apparent diffusion coefficient (ADC) map represents absolute measure of average diffusion, which has been proved to differentiate and grading brain tumors (16,17). Although most of the MRI experiments reported in pediatric brain tumor literature have focused on the analysis of conventional anatomic MRI sequences such as precontrast T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), and postcontrast T1-weighted (T1C) imaging (5,18), recent studies have tried to extend to other functional MR images (19,20). Posterior fossa tumor in children based on T1C and ADC needs in-depth research.

As an emerging and promising technique, radiomics extracts a large amount of quantitative information from medical images to describe tumor physiology and phenotypes, which could transform medical images into minable data (21–25). Using advanced mathematical algorithms, radiomics takes advantages in exploiting more tumor features that cannot be recognized by the naked eye (26). Several studies have reported that machine-learning analysis of tumor medical images could enhance the accuracy of tumor classification and grading (27–30). Pediatric brain tumors have special characteristics in tissue, morphology and texture, which can cause errors if they are carried out in accordance with adult standards (5,31,32). Recently, a number of studies applied radiomics to evaluate adult intracranial brain tumors (23,24,33), and several studies used texture features to research the classification of pediatric tumors (32,34). Radiomics provides additional descriptive information over and above the texture features, such as shape and intensity characters. However, few studies distinguished EP and MB of pediatric by means of machine-learning and radiomics method on multimodal MRI. No model has better performance for all different data sets, so the sub goal of this paper is to compare radiomics models constructed by different feature selection methods and various classifiers.

The main purpose of this study is to investigate the effectiveness of radiomics and machine-learning techniques on multimodal MR images in EP and MB of childhood. Up to now, radiomics and machine-learning are still research tools to a large extent, and multiple challenges need to be solved before they can be integrated into common clinical care.

MATERIALS AND METHODS

Patient Selection

This study was conducted following approval by the Research Ethics Board, From May 2008 to November 2017,

a database of 51 patients with histologically proven PPFT was retrospectively retrieved. Approval for the study was obtained from the research ethics committee, and informed consent was taken from guardians. Table 1 provides summary characteristics of the patient cohort. There were 24 cases of EP, 27 cases of MB.

We retrospectively identified patients who met the following criteria: (1) histopathological-proven PPFT; (2) known EP and MB; (3) patients aged between 0 and 15; (4) 3T MRI performed before resection operation; and (5) available pre-operative MRI. Patients with these situations were excluded from this study. (1) previous treatment for PPFT (i.e., bevacizumab therapy, radiation or chemotherapy); (2) those who had imaging artefacts making the segmentation of cancer lesions impossible. Our final cohort included 51 patients with EP ($n = 24$) and MB ($n = 27$). See Table 1 for details.

MR Imaging Acquisition

MRI was performed with a MAGNETOM Skyra 3.0T scanner (Siemens Medical Solutions, Erlangen, Germany) and standard head coil. In our study, imaging protocol for PPFT included the following sequences: (1) axial T1C with gadopentetate dimeglumine (Magnevist, Bayer Healthcare) was administered by injection through a peripheral venous catheter at a dose 0.2 mmol/kg. Echo time (TE) = 2.5 ms, repetition time (TR) = 434 ms, the slice thickness was 5 mm with 1 mm intersection gap. (2) Axial DWI with b values of 0 and 1000 s/mm². TE = 80 ms, TR = 4600 ms, the slice thickness was 5 mm with 1 mm intersection gap. The reconstructed ADC maps were generated automatically by the software package on a PACS workstation.

Image Processing and Tumor Segmentation

First, N4 bias correction was applied to correct the low-frequency intensity nonuniformity from the images (35). After isotropic voxel resampling, ADC maps were registered to T1C images using affine transformation. To minimize heterogeneity in image intensity, the images were subjected to signal intensity normalization by means of WhiteStripe method (36). Then, two neuroradiologists (Y.M., Y.Z., with 9 and 10 years' experience in radiology) used 3D slicer 4.7.0 (<http://www.slicer.org/>) to semi-automatically draw the

TABLE 1. Characteristics of Children Tumor Cohort

Characteristics	EP	MB
No. of patients	24 (47.06%)	27 (52.94%)
Sex	Male	14 (58.33%)
	Female	10 (41.67%)
Age(y)	Range	0–12
	Mean ± STD	4.84 ± 3.06
Grade	Low (I II)	8 (33.33%)
	High (III IV)	16 (66.64%)

EP, ependymoma; MB, medulloblastoma.

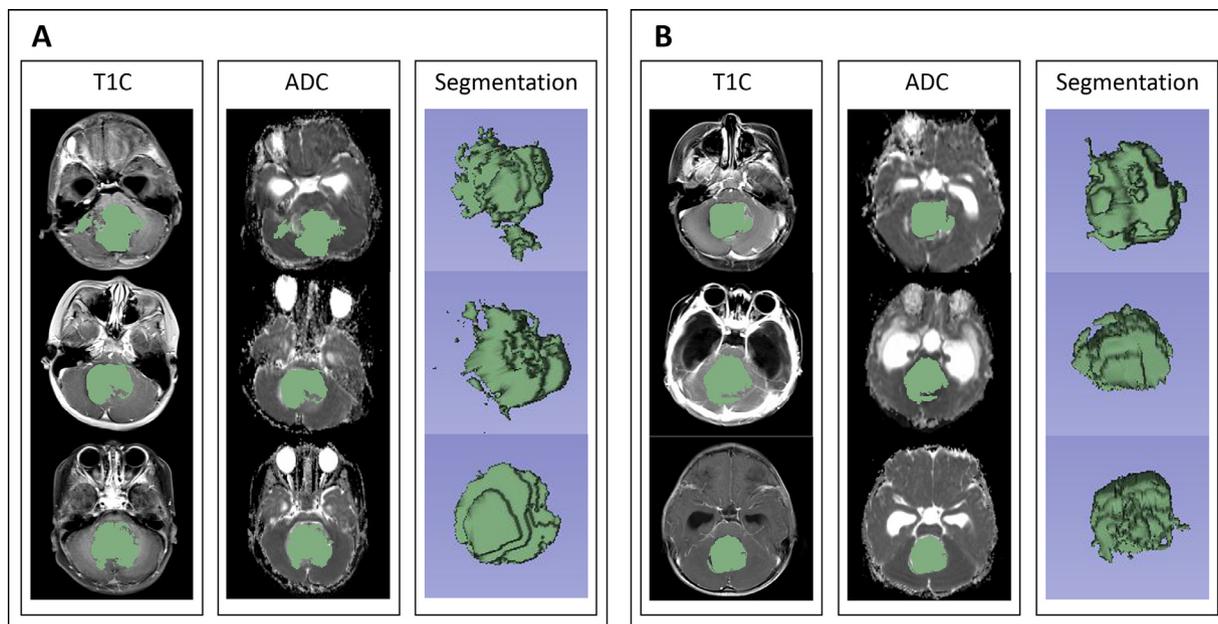


Figure 1. Example of children tumors definition performed on the T1C (left), ADC (middle) and their corresponding contrast-enhancing tumors (right). A represent 3 ependymoma patients; B represent 3 medulloblastoma patients.

three-dimensional (3D) volumes of interest (VOIs) of each patient layer by layer on axial T1C images. 3D slicer software is a user-driven manual motion contour segmentation tool for tumor segmentation (37–39). VOIs were defined as contrast-enhancement tumors, which were selected automatically by 2 neuroradiologists using a simple region growing segmentation algorithm implemented in 3D Slicer. After that, two neuroradiologists manually added pixels to the tumor area that is not included in the preliminary contour or remove pixels of the nontumor area contained in the preliminary contour. Subsequently, two experts accomplished the results of the consistency segmentation based on T1C. Finally, VOIs masks on T1C were registered to ADC maps. The segmentation results are shown in Figure 1. Necrosis, cystic, and edema were carefully avoided to minimize mistaking them for the solid tumor. The following radiomics features were carried out on the VOIs.

Feature Extraction

Radiomics features were conducted with 3D Slicer, which included three parts. First, shape-based features. Sixteen features were extracted, which were used to describe the morphological characteristics. Second, histogram-based features. Nineteen first-order features were calculated, which describe the distribution of gray values within an image (40). Third, the texture-based features were extracted, which included 27 grey level co-occurrence matrix (GLCM) features, 16 grey-level run length matrix (GLRLM) features and 16 grey-level size zone matrix (GLSZM) features. Different tumors have phenotype difference. In order to describe the tumor heterogeneity, we also need to consider the texture characteristics (41,42). GLCM describes the joint distribution of two pixel grayscale with a certain spatial position. GLRLM

features can be calculated from run length matrices, which represent the structure of an image, GLSZM is needed to provided effective information in representing texture consistency, nonperiodicity, or speckle. The features extracted are summarized in Table 2. Ninety-four features for each modality, a total of 188 features were extracted from T1C and ADC maps of each patient.

Feature Selection

Normalization and feature reduction were performed before classification. All the radiomics features were transformed to a standard range. Data scaling is to scale the data down to a small specific interval. It is often used in the processing of certain comparison and evaluation indicators to remove the unit limit of the data and convert it into a dimensionless pure value, so that indicators of different units or magnitudes can be compared and weighted. Therefore, before further feature analysis, we usually need to standardize the feature and then use the standardized feature for analysis. In our study, we mapped the features to the interval [0,1] for normalization.

Many of the features have no effect on distinguishing between EP and MB, and even redundancy. In addition, the features we extracted ($n = 188$) are much higher than our number of cases ($n = 51$). Therefore, the dimensionality reduction and the task-specific features selection are useful (12,43). We tried three methods for feature selection using SPSS 23.0 (SPSS Inc., Chicago, IL). The first way was univariate analysis (UA), which used the Kolmogorov-Smirnov method to test the distribution of the data. Normal image features were subjected to Student's t test, and non-normal image features were evaluated by the Mann-Whitney U-test (44). Selected features with $p < 0.05$ in the results indicates that these features are significant in distinguishing between EP and MB.

TABLE 2. Summary of the Radiomics Used and Their Corresponding Features

Feature Classes	Feature Names
Shape features	Maximum 3D diameter, Compactness1, Maximum 2D diameter slice, Sphericity, Minor axis, Compactness2, Elongation, Surface Volume ratio, Volume, Spherical disproportion, Major axis, Least axis, Flatness, Surface area, Maximum 2D diameter column, Maximum 2D diameter row.
Histogram features	Interquartile range, Skewness, Uniformity, Mean absolute deviation, Energy, Robust mean absolute deviation, Median, Total energy, Maximum, Root mean squared, 90 percentile, Minimum, Entropy, Standard deviation, Range, variance, 10 percentile, Kurtosis, Mean.
Texture features	GLCM features
	Sum variance, Homogeneity1, Homogeneity2, Cluster shade, Maximum probability, Idmn, Contrast, Difference entropy, Inverse variance, Dissimilarity, Sum average, Difference variance, Idn, Idm, Correlation, Auto-correlation, Sum entropy, Average intensity, Energy, Sum squares, Cluster prominence, Entropy, Information meast Difference average, Id, Cluster tendency.
	GLRLM features
	Short run low gray level emphasis, Sray level variance, Low gray level run emphasis, Gray level nonuniformity normalized, Run variance, Gray level nonuniformity, Long run emphasis, Short run high gray level emphasis, Run length nonuniformity, Short run emphasis, Long run high gray level emphasis, Run percentage, Long run low gray level emphasis, Run entropy, High gray level run emphasis, Run length nonuniformity normalized.
	GLSZM features
	Gray level variance, Small area high gray level emphasis, Gray level nonuniformity normalized, Size zone nonuniformity normalized, Size zone nonuniformity, Gray level nonuniformity, Large area emphasis, Zone variance, Zone percentage, Large area low gray level emphasis, Large area high gray level emphasis, High gray level zone emphasis, Small area emphasis, Low gray level zone emphasis, Zone entropy, Small area low gray level emphasis.

GLCM, gray level co-occurrence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix.

Another method was to use multivariable logistic regression (MLR) to select independently predictive features. All radiomics features were modeled with MLR method, the results of pathological tests were classified as the gold standard. Forward stepwise regression method was performed to remove redundant features, which base on the likelihood ratio probability test. It started with no variables in the model and added the variable that most significant improvement of the fit, repeated the process until none variable improves the model (44).

The third one was univariate analysis screening (UAS) method, which selected significant variables according to univariate analysis, than the selected variables were included in multivariable logistic regression model. Before MLR analysis, if the number of samples is small and the variables are large, the usual practice is to investigate the relationship between all independent variables and dependent variables through UA (*t* test, chi-square test, etc.), remove some variables that may be not such significant, and then carry out MLR analysis, so as to ensure more reliable results. This is so called UAS method, which has been widely published in the top medical journals (45–47). But some scholars reject this conclusion, Wang H, etc. pointed out that UAS method as a popular approach is not reliable because some variable unselected may also be significant according to univariate analysis, it would cause bias, even lead to a wrong conclusion (48). We conducted a comparative research to explore which method is more suitable for the subjects of this experiment.

Classification and Validation

The classification algorithm analyzes the data and derives a hypothesis function that can be used to predict the labels of

unseen observations. In our study, the selected subset of features were analyzed using four machine-learning classifiers, which were k-nearest neighbour (kNN), adaptive boosting (AdaBoost), random forests (RF), and support vector machines (SVM). Ten-fold cross-validation was carried out to evaluate generalizability and accuracy of the model. It divided the data set into 10 parts and took turns to use 9 parts as training data and 1 part as testing data to conduct experiments. Each time yielded the correct answer. The 10-fold crossvalidation was repeated 10 times, finally the average value was taken as the estimation of the accuracy of the algorithm. Evaluate the classification effect by using four indicators: area under the receiver operating characteristic curve (AUC), accuracy (Acc), sensitivity (Sens), and precision (Prec). Finally, we used pairwise test to compare the receiver operating characteristic (ROC) curves and investigated the difference between ROC curves with various classifiers. The $p < 0.05$ indicates that the two ROC curves are statistically significant different according to DeLong analysis. The workflow of the image processing and machine-learning is shown in Figure 2.

RESULTS

Feature Selection

After feature extraction, 188 radiomics features were extracted from each patient. Fifty-six of the available 188 features were selected by UA method, which including 12 for Shape features, 8 for T1C_based features, 36 for ADC_based

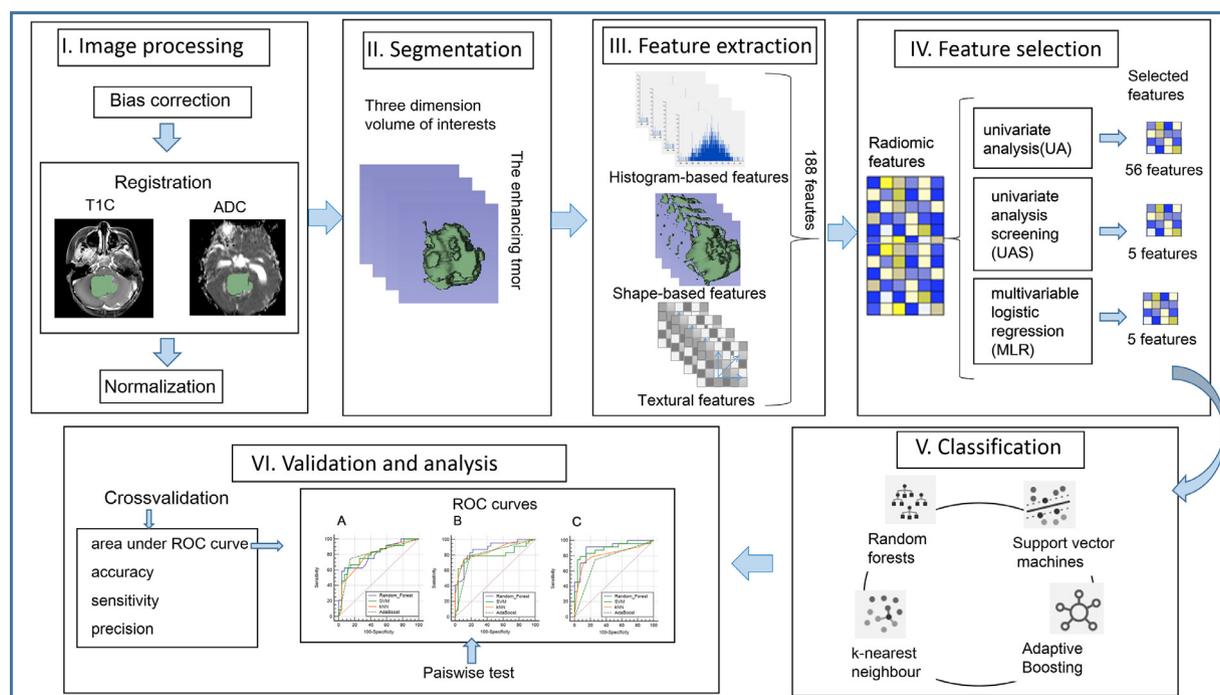


Figure 2. Workflow of the image processing and machine-learning procedure. (I) N4 bias correction was applied in postcontrast T1-weighted (T1C) images and apparent diffusion coefficient (ADC) maps. Then, ADC maps were registered to T1C images. Intensity normalization processes were performed for the registered images. (II) The three-dimensional (3D) volumes of interests (VOIs) were semisegmented by neuro-radiologists. (III) Radiomics features were extracted from VOIs, including shape, first order, and texture based features. (IV) Features were selected by the three feature selection methods separately. (V) Four machine-learning based algorithm was used as classifier to build the prediction model. (VI) 10-fold crossvalidation was repeated 10 times, area under the receiver operating characteristic curve (AUC), accuracy, sensitivity and precision were the indicators to assess the performance of our study. Then pairwise test was performed to compare the receiver operating characteristic (ROC) curves.

features. It should be noted that these 56 features could significantly distinguished EP and MB with $p < 0.05$.

The optimal features selected by UAS and MLR methods are shown in Tables 3 and 4. Table 3 shows the five features extracted by UAS, while Table 4 shows the five features extracted by MLR. These features were highly correlated to discriminate EP and MB, which may be used as imaging biomarkers in EP and MB, as detailed in the Discussion section.

TABLE 3. A Summary of the Features Selected by UAS Method

Feature Selection Method	No.	Selected Feature Name	Type
UAS method	n1	Spherical disproportion	Shape
	n2	Median	ADC_First order
	n3	Information measure of correlation	T1C_GLCM
	n4	Root mean squared	ADC_First order
	n5	Run length non uniformity	T1C_GLRLM

ADC, apparent diffusion coefficient; T1C, postcontrast T1-weighted; GLCM, gray level co-occurrence matrix; GLRLM, gray-level run length matrix; UAS, univariate analysis screening.

Classification Results and Model Validation

Table 5 gives detailed outcomes of the stratified 10-fold cross validation results. Three feature selection methods and four classification methods were used to compare EP and MB in terms of AUC, accuracy, sensitivity, and precision. For UA feature selection method, the optimal performance was yield

TABLE 4. A summary of the Features Selected by MLR Method

Feature Selection Method	No.	Selected Feature Name	Type
MLR method	r1	Spherical disproportion	Shape
	r2	Median	ADC_First order
	r3	Information measure of correlation	T1C_GLCM
	r4	Low gray level zone emphasis	ADC_GLSZM
	r5	Inverse difference moment normalized	T1C_GLCM

ADC, apparent diffusion coefficient; GLCM, gray level co-occurrence matrix; GLSZM, gray-level size zone matrix; MLR, multivariable logistic regression; T1C, postcontrast T1-weighted.

TABLE 5. Compare the Performance of Different Feature Selection and Classification Methods to Distinguish MB From EP

Feature Selection Method	Classification Method	EP		MB		Overall			
		Sens (%)	Prec (%)	Sens (%)	Prec (%)	AUC	Acc (%)	95% CI of AUC	<i>p</i>
UA	RF	62.5	68.2	74.1	69.0	0.80	68.6	0.659–0.896	<i>p</i> < 0.001
	SVM	66.7	76.2	81.5	73.3	0.80	74.5	0.660–0.896	<i>p</i> < 0.001
	KNN	66.7	69.6	74.1	71.4	0.79	70.6	0.645–0.886	<i>p</i> < 0.001
	AdaBoost	75.0	81.8	85.2	79.3	0.80	80.4	0.665–0.900	<i>p</i> < 0.001
UAS	RF	79.2	82.6	85.2	82.1	0.88	82.4	0.753–0.951	<i>p</i> < 0.001
	SVM	75.0	85.7	88.9	80.0	0.80	82.4	0.658–0.895	<i>p</i> < 0.001
	KNN	75.0	85.7	88.9	80.0	0.85	82.4	0.727–0.937	<i>p</i> < 0.001
	AdaBoost	79.2	79.2	81.5	81.5	0.80	80.4	0.668–0.901	<i>p</i> < 0.001
MLR	RF	91.7	81.5	81.5	91.7	0.91	86.3	0.787–0.968	<i>p</i> < 0.001
	SVM	83.3	83.3	85.2	85.2	0.89	84.3	0.766–0.958	<i>p</i> < 0.001
	KNN	70.8	77.3	88.9	77.4	0.82	80.4	0.681–0.910	<i>p</i> < 0.001
	AdaBoost	75.0	73.5	74.1	76.9	0.75	74.5	0.604–0.857	<i>p</i> < 0.001

p < 0.001 indicate that it's statistically significant different.

Acc, accuracy; AdaBoost, adaptive boosting; AUC, area under the receiver operating characteristic curve; CI, confidence interval; EP, ependymoma; kNN, k-nearest neighbour; MB, medulloblastoma; MLR, multivariable logistic regression; Prec, precision; RF, random forests; Sens, sensitivity; SVM, support vector machines; UA, univariate analysis; UAS, univariate analysis screening.

by Adaboost, achieved AUC of 0.80 (95% confidence interval [CI] 0.665–0.9000) and an accuracy of 80.4%. For UAS method, the best performance was obtained by RF classifier, which achieved AUC of 0.88 (95%CI 0.753–0.951) and an accuracy of 82.4%. For MLR method, the best performance was obtained by RF classifier, which achieved AUC of 0.91 (95% CI 0.787–0.968) and an accuracy of 86.3%. Table 5 shows the performance of EP and MB in detail. The optimal performance was yield by radiomics model built with MLR feature selection method and RF classifier, which gained the highest AUC and accuracy, achieved sensitivity of 91.7% and precision of 81.5% in EP, achieved sensitivity of 81.5% and precision of 91.7% in MB. To further reveal the relevance of the different feature selection methods and classification methods, the ROC curves are presented in Figure 3. It is

worth noting that, in general the performance of MLR method is better than the other two. We can see that in the ROC curves obtained by MLR method and different classifiers, the performance of RF classifiers in blue is better than that of SVM, KNN, and Adaboost classifiers, as shown in Figure 3(C). The DeLong analysis found a significant difference (*p* < 0.05).

The final five features selected by MLR with RF got the best performance in our study, which is listed in Table 4. By inspecting the box plots in Figure 4, we can see that after comparison of mean and median of five features in EP and MB, four features have significant differences in distinguishing EP and MB (*p* < 0.05), which are spherical disproportion, median (ADC), information measure of correlation (T1C), low gray level zone emphasis (ADC). Meanwhile, it is worth

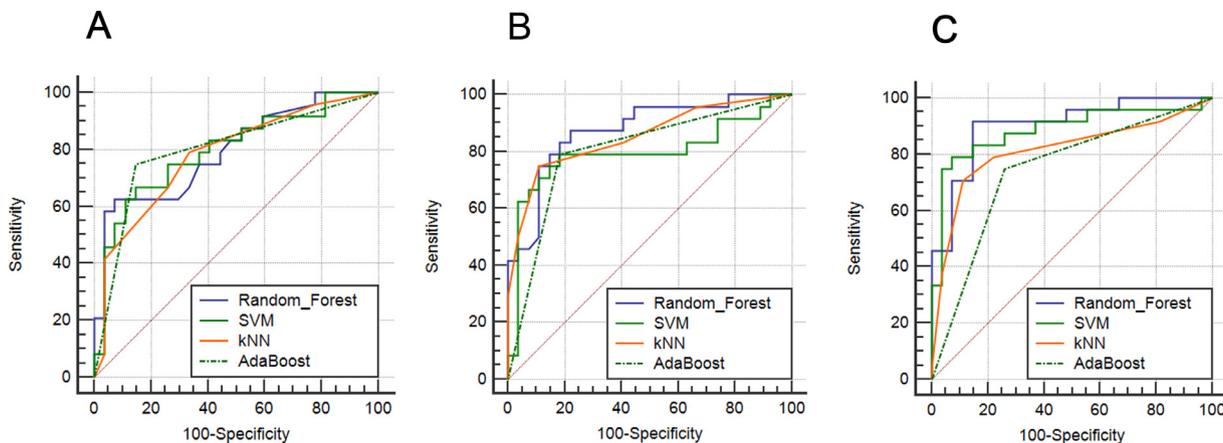


Figure 3. Receiver operating characteristic (ROC) curves of the radiomics models built with three feature selection methods and four classification methods. (A) ROC curves for radiomics model built with univariate analysis (UA) method. (B) ROC curves for radiomics model built with univariate analysis screening (UAS) method. (C) ROC curves for radiomics model built with multivariable logistic regression (MLR) method. In each figure, different color curves represent different classifiers, blue curve represents random forests (RF), green solid curve represents support vector machines (SVM), orange curve represents k-nearest neighbour (kNN), and green dotted curve represents adaptive boosting (AdaBoost). (Color version of figure is available online.)

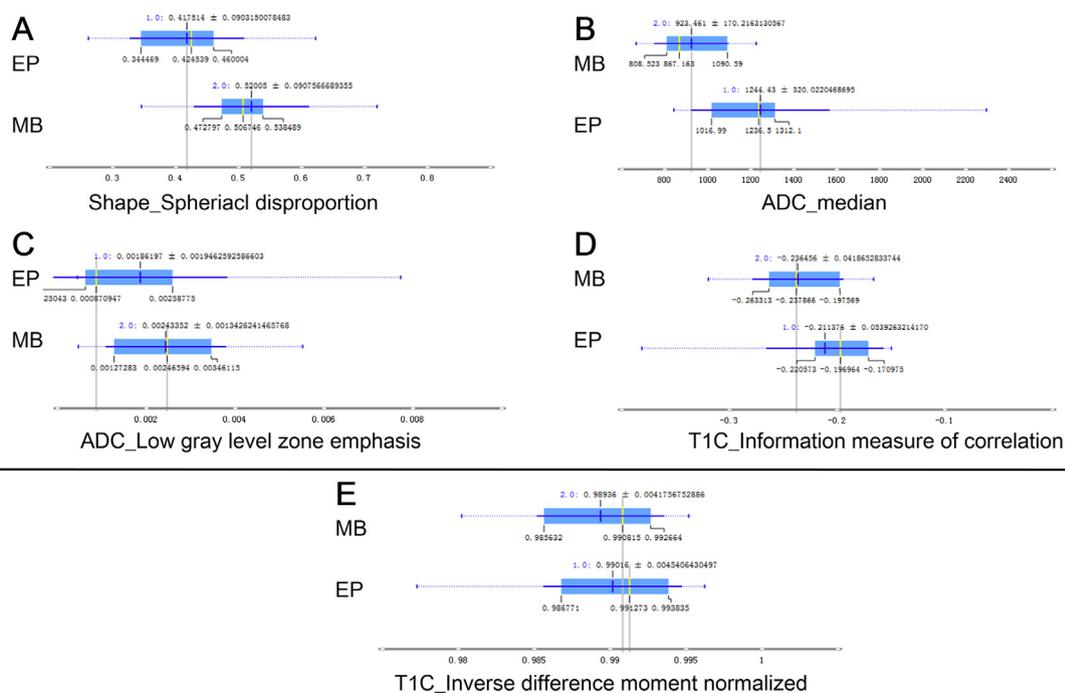


Figure 4. Box plots of the five most relevant features that distinguish ependymoma (EP) from medulloblastoma (MB). (A–D) These four features have significant differences in distinguishing EP and MB ($p < 0.05$), which are shape_spherical disproportion, ADC_median, T1C_information measure of correlation, ADC_low gray level zone emphasis. (E) T1C_inverse difference moment normalized was considered to be redundant features ($p > 0.05$) by univariate analysis (UA) and univariate analysis screening (UAS) method.

noting that the features that have important contributions to the results may not have significant differences. It is unreliable to directly perform univariate analysis in the feature selection stage, and many important features may be weed out. See Discussion section for details.

Statistical Findings

Although the highest AUC value was obtained when RF was carried out with five features selected by MLR (AUC of 0.91), we want to explore whether there are significant differences in results under different classifiers. For MLR feature selection method, the result show that there were significant differences between the following groups. Comparing RF with Adaboost ($p = 0.0031$), SVM with Adaboost ($p = 0.0133$), KNN with SVM ($p = 0.0320$), the p values in the two-tailed test were all less than 0.05. But there were not obvious significance between ROCs curves by others classifier, For instance, comparing KNN with RF ($p = 0.0914$), the details are listed in Table 6.

DISCUSSION

EP and MB in children are both tumors of the posterior cranial fossa. They are similar in age, location, MRI phenotype, and clinical manifestations, therefore it is difficult to distinguish them (5,18). The aim of quantitative medical image analysis is to assist doctors for diagnosis and clinical decision-making. Shape, intensity, location, and texture are commonly concerned by radiologists. While radiomics use images to

mine more features, provide more comprehensive information, and digitize the information. In this study, effective radiomics features were extracted from MRI images, and a classification model was established by machine-learning technology to distinguish EP from MB. We found that the model established by radiomics has a good effect on differentiation between the kinds of tumors.

In recent years, there have been studies using radiomics to evaluate adult intracranial brain tumors (22,49), and also research on the MRI texture analysis of pediatric tumors (32,50,51). However few researches explored PPFT via radiomics. Rodriguez Gutierrez et al. investigated the performance of SVM classifier with texture features extracted from T1WI, T2WI, and ADC maps in PPFT, the results showed

TABLE 6. Pairwise Comparison of ROC Curves

	SE	95% CI	p
Adaboost – RF	0.0533	0.0530–0.262	0.0031
Adaboost - SVM	0.0567	0.0292–0.252	0.0133
Adaboost - KNN	0.0641	–0.0562 to 0.195	0.2788
KNN – RF	0.0521	–0.0142 to 0.190	0.0914
KNN – SVM	0.0331	0.00611–0.136	0.0320
RF – SVM	0.0350	–0.0515 to 0.0855	0.6272

The $p < 0.05$ gray indicates that the two ROC curves are significant different.

AdaBoost, adaptive boosting; CI, confidence interval; kNN, k-nearest neighbour; RF, random forests; ROC, receiver operating characteristic; SE, standard error; SVM, support vector machines.

that classification rates of EP and MB ranged between 69.7% and 81.9% (51). Eleni Orphanidou-Vlachou et al. explored the effect of probabilistic neural network as a classifier to train the texture features extracted from T1WI and T2WI, showed that sensitivity of EP and MB ranged from 63% to 94% (18). Fetit et al. conducted a multicenter study on PPFT with texture analysis. Two from three centers were used as training set and the remaining one as validation set, Entropy-MDL, ReliefF, combination of Entropy-MDL, and ReliefF method as feature selection methods with SVM classifier were used to construct the model, the optimal AUC ranged between 76% and 86% (5). It is worth noting that, these studies focus on the application of texture analysis in PPFT. Although texture could describe the spatial variation of pixel intensities within an image, radiomics tries to provide more comprehensive information based on data mining, which include textures, intensity, morphological characters, geographic information, etc. Besides, the conclusions of these studies are drawn with only one classifier (5,18,51), which is random. Towards specific data, different feature selection and classifier methods may result in different conclusion. We did some research on these issues. In our study, radiomics analysis was used to extract 188 features of each patient from T1C and ADC maps. Meanwhile, we believe that different classification performance will be made by different methods towards different types of tumors. So we experimentally studied the effect of three feature selection and four classification methods, combined them to get the optimal result. The result shows the optimal value of AUC was 0.91 (95%CI 0.787–0.968) when the radiomics model built with MLR as the feature selection method and RF as the classifier.

Feature selection is a crucial step in discovering predictor from high-throughput characters. An interesting observation is in the feature reduction procedure, our study showed that using MLR alone was better than UA and UAS method, which is inconsistent with previous knowledge. They argue that univariate analysis is the initial step in feature selection stage. UAS method firstly conducts univariate analysis, and then variables with significant differences are included to establish multivariate logistic regression model, rather than directly using all variables for multivariable logistic regression analysis. UAS method is widely used and often appears in top medical journals (45–47). UAS method believes that there might be confounding factors from selected variables by univariate analysis, the purpose of multivariate analysis is to eliminate this interference, the covariates selected by multivariable logistic regression after univariate analysis are considered to be independent risk factors for the disease. By inspecting Table 5, the result shows that the MLR performance was best, UAS method took the second place, UA was the worst. This result shows that whether or not to do univariate analysis as the first step in feature selection depends on the specific situation. In our study, the features excluded by univariate analysis may have overlooked potential correlations or interactions between the features, which may be important to differentiate EP from MB.

To further investigate the effect of relevant features and non-redundant features, this study compared the mean and median of the features that contribute the most to distinguish EP and MB (Fig 4). These five features are selected by MLR, which we call relevant features. Four of them have significant differences in distinguishing EP and MB ($p < 0.05$), which are spherical disproportion, median (ADC), information measure of correlation (T1C), low gray level zone emphasis (ADC), they are listed in Figure 4(A–D). But the last one was considered to be redundant features ($p > 0.05$), excluded by UA and UAS method, which is listed in Figure 4(E). Again, the result illustrates that some variables have important contributions to the results may not have significant differences, and vice versa. UAS method uses univariate analysis as an initial step to screen variables, which may mistakenly excluded some important variables and lead to extremely deviation. Some scholars have confirmed this point of view through experiments. They even think that this approach should be removed from the tool kits of biomedical researchers and even some PhD statisticians (48). We believe that no matter what the case may be, the analysis of a problem must be done on a case-by-case basis. We have confirmed the result that it is better to include all variables into MLR than UAS method towards discrimination of EP and MB by experiment.

Explored the optimal machine-learning methods for specific data is a vital step towards radiomics applications. Parmar et al. drew a conclusion that the choice of classification method is the most important factor for performance variation via comparative experiment (34.21% of total variance) (49). Our study established a radiomics model through five most relevant features and four classifiers, by comparing their AUC values, RF performance was the best (AUC = 0.91, 95%CI: 0.787–0.968), SVM and KNN followed, Adaboost was the worst (AUC = 0.75, 95%CI: 0.604–0.857) (Table 5 and Fig 3C). Pairwise test was performed in order to compare the difference of ROC curves (Table 6), the results indicate that there were significant difference when use various classifiers ($p < 0.05$). Therefore, the choice of classifier has a significant impact on the overall performance of the model. For this study, the radiomics model established by MLR feature selection method and RF classifier has the best performance and could best distinguish EP and MB.

Based on the radiological hypothesis, the imaging phenotype may be a potential expression of biological or genetic heterogeneity (52). The selected five most relevant features consisted of one first order feature from ADC, three texture features from T1C and ADC, as well as one shape feature (Table 4). Specifically, r1 describes that there is a difference in the spherical shape of the 3D contrast-enhancement tumors between EP and MB, which is related to the fact that MB often presents as a round solid mass. r2 represents the difference between EP and MB in ADC median, which is related to ADC reflecting the diffusion ability of water molecules in tissues. The cell density in the malignant tumor is high, and it may be necrotic due to hypoxia, so the ADC value is low. MB is a grade IV malignant tumor, with a higher degree of

malignancy than EP, so the ADC value is lower (see Fig 4B), while EP cells proliferate vigorously, with a higher ADC value, therefore, ADC value can be used to reflect tumor heterogeneity. r_3 and r_5 measures the spatial dependencies of grey levels on T1C images. The value of EP are higher than MB, which is due to the different enhancement performance of EP and MB on T1C (see Fig 4D). EP shows moderate and obvious enhancement, MB shows slight and moderate enhancement, and a few cases have no enhancement of MB on T1C. r_4 characterized how many pixels of a given grey value are connected in a single group, that is to describe spatial intensity change in ADC. Interpretation of the radiological relevance remains challenging, although our selected features were contributed to classify EP and MB. Five relevant features selected by MLR obtained the best performance, which are highly correlated to discriminate EP and MB. It can also be used as imaging biomarkers to predict the classification of tumor types.

In this paper, we used T1C and ADC images, which have important clinical application value in the diagnosis and evaluation of brain tumors. Multiregion research also have many benefits for tumor classification. Different subregion is concerned with different tumor characteristics, T1C focus on the area of enhancing and necrosis of tumor, ADC reflects the magnitude of water motion and is the most commonly used DWI metric in clinical practice. This study achieved good automatic classification performance, some limitations should also be considered. First, although data of EP and MB in children have been collected for nearly 10 years, there is still a small sample in this study due to the rarity of brain tumors in children. In the future, we hope to do some multicenter research to put up with this problem. Second, in this paper, T1C and ADC maps were applied, and only enhancing-tumor region was studied. If other sequences and other regions can be added, it will be more convincing. Third, lack of independent validation set due to the small sample size of our study. However, in order to ensure that our method can be generalized to independent data, further research and validation are needed in independent data sets.

CONCLUSIONS

In conclusion, in this paper we attempted to discriminate EP from MB of pediatric in multimodal MRI by means of radiomics as well as machine-learning techniques. Through three feature selection methods and four classifiers, 12 radiomics models were established. The results showed that the combination of MLR and RF has the best performance in distinguishing EP and MB (AUC of 0.91, 95%CI: 0.787–0.968).

In addition, we also proved that logistic regression does not have to follow the “univariate-multivariate” (multivariable logistic regression after univariate analysis) through experiments, the method should depend on the specific target data. The results of our research showed that MLR has better performance than UAS method in discrimination of EP and MB. The choice of classifiers has an important influence on

the overall performance of the model. The radiomics model established by MLR feature selection method and RF classifier had the best performance in our study. Five relevant features, including spherical disproportion (Shape), median (ADC), information measure of correlation (T1C), low gray level zone emphasis (ADC), and inverse difference moment normalized (T1C) may be regarded as imaging biomarkers to predict classification of the kinds of tumors. The study could help with clinical diagnosis.

ACKNOWLEDGMENTS

The authors declare that they have no conflict of interest. The authors would like to express appreciation to the Zhengzhou University First Affiliated Hospital for providing the childhood brain tumor images for this work, in addition, thank the National Natural Science Foundation of China (81171410) for financially supporting this research.

REFERENCES

- Bright C, Reulen R, Fidler M, et al. Cerebrovascular complications in 208,769 5-year survivors of cancer diagnosed aged 15-39 years using hospital episode statistics: the population-based Teenage and Young Adult Cancer Survivor Study (TYACSS). *Eur J Cancer Care* 2015; 241:9.
- Albright AL. Pediatric brain tumors. *CA* 1993; 43:272–288.
- Cancer Research UK. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers/incidence#heading=Five>. Accessed September 2017.
- Gaudino S, Martucci M, Russo R, et al. MR imaging of brain pilocytic astrocytoma: beyond the stereotype of benign astrocytoma. *Child's Nervous Syst* 2017; 33:35–54.
- Fetit AE, Novak J, Rodriguez D, et al. Radiomics in paediatric neuro-oncology: a multicentre study on MRI texture analysis. *Nmr Biomed* 2018; 31:e3781.
- Aronen HJ, Gazit IE, Louis DN, et al. Cerebral blood volume maps of gliomas: comparison with tumor grade and histologic findings. *Radiology* 1994; 191:41–51.
- Lunsford LD, Martinez AJ. Stereotactic exploration of the brain in the era of computed tomography. *Surg Neurol* 1984; 22:222–230.
- Hall WA, Liu H, Martin AJ, et al. Comparison of stereotactic brain biopsy to interventional magnetic-resonance-imaging-guided brain biopsy. *Stereotact Funct Neurosurg* 1999; 73:148–153.
- Field M, Witham TF, Flickinger JC, et al. Comprehensive assessment of hemorrhage risks and outcomes after stereotactic brain biopsy. *J Neurosurg* 2001; 94:545–551.
- Bernstein M, Parrent AG. Complications of CT-guided stereotactic biopsy of intra-axial brain lesions. *J Neurosurg* 1994; 81:165–168.
- Jin G, An N, Jacobs MA, et al. The role of parallel diffusion-weighted imaging and apparent diffusion coefficient (ADC) map values for evaluating breast lesions. *Acad Radiol* 2010; 17:456–463.
- Suh HB, Choi YS, Bae S, et al. Primary central nervous system lymphoma and atypical glioblastoma: Differentiation using radiomics approach. *Eur Radiol* 2018; 28:3832–3839.
- Ho Sung Kim M, Myeong Ju Goh M, Namkug Kim P, et al. Which combination of mr imaging modalities is best for predicting recurrent glioblastoma study of diagnostic accuracy and reproducibility 1 *Radiology* 2014; 273:831–842.
- Lee J, Jain R, Khalil K, et al. texture feature ratios from relative CBV maps of perfusion MRI are associated with patient survival in glioblastoma. *Am J Neuroradiol* 2016; 37:37–43.
- Kickingeder P, Go Tz M, Muschelli J, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clin Cancer Res* 2016; 22:5765–5771.
- Bull JG, Saunders DE, Clark CA. Discrimination of paediatric brain tumours using apparent diffusion coefficient histograms. *Eur Radiol* 2012; 22:447–457.

17. Marupudi NI, Altinok D, Goncalves L, et al. Apparent diffusion coefficient mapping in medulloblastoma predicts non-infiltrative surgical planes. *Child's Nervous Syst* 2016; 32:2183–2187.
18. Orphanidou-Vlachou E, Vlachos N, Davies NP, et al. Texture analysis of T1-and T2-weighted MR images and use of probabilistic neural network to discriminate posterior fossa tumours in children. *Nmr Biomed* 2014; 27:632–639.
19. Koob M, Girard N, Ghattas B, et al. The diagnostic accuracy of multiparametric MRI to determine pediatric brain tumor grades and types. *J Neuro-Oncol* 2016; 127:345–353.
20. Zitouni S, Koc G, Doganay S, et al. Apparent diffusion coefficient in differentiation of pediatric posterior fossa tumors. *Jpn J Radiol* 2017; 35:448–453.
21. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging* 2012; 30:1234–1248.
22. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48:441–446.
23. Aerts H, Velazquez ER, Leijenaar R, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; 1–9.
24. Zhou H, Vallières M, Bai HX, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro-Oncology* 2017; 19:862–870.
25. Liang C, Cheng Z, Huang Y, et al. An MRI-based radiomics classifier for preoperative prediction of Ki-67 status in breast cancer. *Acad Radiol* 2018; 25:1111–1117.
26. Aerts SSFY. Applications and limitations of radiomics. *Phys Med Biol* 2016; 61:150–166.
27. Litjens G, Sánchez CI, Timofeeva N, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep-Uk* 2016; 6:26286.
28. Zhao L, Jia K. Multiscale CNNs for brain tumor segmentation and diagnosis. *Comput Math Method M* 2016; 1–7.
29. Emblem KE, Pinho MC, Zollner FG, et al. A generic support vector machine model for preoperative glioma survival associations. *Radiology* 2015; 275:228–234.
30. Macyszyn L, Akbari H, Pisapia JM, et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. *Neuro Oncol* 2016; 18:417–425.
31. Goya-Outi J, Orlhac F, Calmon R, et al. Computation of reliable textural indices from multimodal brain MRI: suggestions based on a study of patients with diffuse intrinsic pontine glioma. *Phys Med Biol* 2018; 63:105003.
32. Fetit AE, Novak J, Peet AC, et al. Three-dimensional textural features of conventional MRI improve diagnostic classification of childhood brain tumours. *Nmr Biomed* 2015; 28:1174–1184.
33. Li Z, Bai H, Sun Q, et al. Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. *Eur Radiol* 2018; 28:3640–3650.
34. Raja R, Sinha N, Saini J, et al. Assessment of tissue heterogeneity using diffusion tensor and diffusion kurtosis imaging for grading gliomas. *Neuroradiology* 2016; 58:1217–1231.
35. Tustison NJ. ABCP. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 2010; 1310–1320.
36. Shinohara RT, Sweeney EM, Goldsmith J, et al. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage* 2014; 6:9–19.
37. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. In, 2012; 1323–1341.
38. Pieper S, Lorensen B. The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D slicer as an open platform for the medical image computing community. *IEEE* 2006; 698–701.
39. Egger J, Kapur T, Fedorov A, et al. GBM volumetry using the 3D Slicer medical image computing platform. *Nat Scient Rep* 2013; 1–7.
40. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48:441–446.
41. Weszka JS, Dyer CR, Rosenfeld A. A comparative study of texture measures for terrain classification. *IEEE Trans Syst Man Cybern* 1976; 269–285. SMC-6.
42. Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. *IEEE Trans Syst Man Cybern* 1978; 8:460–473.
43. Dong F, Li Q, Xu D, et al. Differentiation between pilocytic astrocytoma and glioblastoma: a decision tree model using contrast-enhanced magnetic resonance imaging-derived quantitative radiomic features. *Eur Radiol* 2018.
44. Gigliotti E. *Discovering statistics using SPSS. Second Edition* Oxford, UK: Blackwell Publishing Ltd, 2007:303.
45. Karcutskie CA, Meizoso JP, Ray JJ, et al. Association of mechanism of injury with risk for venous thromboembolism after trauma. *Jama Surg* 2017; 152:35–40.
46. Templin C, Ghadri JR, Diekmann J, et al. Clinical features and outcomes of takotsubo (stress) cardiomyopathy. *N Engl J Med* 2015; 373:929–938.
47. Nor AM, Davis J, Davis M, et al. The Recognition of Stroke in the Emergency Room (ROSIER) scale: development and validation of a stroke recognition instrument. *Lancet Neurol* 2005; 4:727–734.
48. Wang H, Peng J, Wang B, et al. Inconsistency between univariate and multiple logistic regressions. *Shanghai Arch Psychiatry* 2017; 29:124–128.
49. Parmar C, Grossmann P, Bussink J, et al. Machine learning methods for quantitative radiomic biomarkers. *Nat Scient Rep* 2015; 1–11.
50. Mengmeng L, Zhigang S, Yonghui D, et al. Application of MRI texture analysis in the study of the posterior fossa tumors growing trend in children. *Conf Proc IEEE Eng Med Biol Soc* 2017; 2017:620–623.
51. Rodriguez Gutierrez D, Awwad A, Meijer L, et al. Metrics and textural features of MRI diffusion to improve classification of pediatric posterior fossa tumors. *Am J Neuroradiol* 2014; 35:1009–1015.
52. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology* 2013; 269:8–15.