

Research Article

Computer-Aided Grading of Gliomas Combining Automatic Segmentation and Radiomics

Wei Chen , Boqiang Liu , Suting Peng , Jiawei Sun , and Xu Qiao

Department of Biomedical Engineering, School of Control Science and Engineering, Shandong University, Shandong, China

Correspondence should be addressed to Boqiang Liu; bqliu@sdu.edu.cn

Received 5 January 2018; Revised 6 March 2018; Accepted 2 April 2018; Published 8 May 2018

Academic Editor: Chun-Chih Liao

Copyright © 2018 Wei Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gliomas are the most common primary brain tumors, and the objective grading is of great importance for treatment. This paper presents an automatic computer-aided diagnosis of gliomas that combines automatic segmentation and radiomics, which can improve the diagnostic ability. The MRI data containing 220 high-grade gliomas and 54 low-grade gliomas are used to evaluate our system. A multiscale 3D convolutional neural network is trained to segment whole tumor regions. A wide range of radiomic features including first-order features, shape features, and texture features is extracted. By using support vector machines with recursive feature elimination for feature selection, a CAD system that has an extreme gradient boosting classifier with a 5-fold cross-validation is constructed for the grading of gliomas. Our CAD system is highly effective for the grading of gliomas with an accuracy of 91.27%, a weighted macroprecision of 91.27%, a weighted macrorecall of 91.27%, and a weighted macro-F1 score of 90.64%. This demonstrates that the proposed CAD system can assist radiologists for high accurate grading of gliomas and has the potential for clinical applications.

1. Introduction

Gliomas are the most common primary brain tumors, characterized by the uncontrolled proliferation of abnormal brain cells. This disease is one of the most common causes of cancer death in men and women [1]. According to classification of the World Health Organization (WHO), gliomas can be subdivided by their malignancy into grade II (lower grade) to grade IV (high grade) [2]. The clinical patients with high-grade gliomas, such as glioblastomas, have the median overall survival rate of 23.1 months, the 2-year survival rate of 47.4%, and the 4-year survival rate of 18.5% [3]. On the contrast, the slower growing low-grade gliomas, such as astrocytomas and oligodendrogliomas, come with an overall 10-year survival rate of 57% [4]. Therefore, early detection is considered as an effective way to get a hopeful prognosis.

Modern imaging techniques allow clinicians and radiologists to evaluate the progression of tumors and choose optimal treatment strategy, without invasive neurosurgery. There are many imaging modalities that can be used to study the brain, such as computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI).

These imaging modalities can provide effective and reliable information about brain tissues. In view of the advantages of high soft tissues contrast and high spatial resolution, MRI is widely used to evaluate the tumor heterogeneity [5]. However, there are several common MRI modalities including T1-weighted (T1), T2-weighted (T2), gadolinium enhanced T1-weighted (T1c), and Fluid-Attenuated Inversion Recovery (FLAIR), which generate a large number of medical images. This has become a huge burden for radiologists, resulting in inaccurate detection or misinterpretation. Therefore, with the development of computer technology, there is an increasing demand for computer-aided diagnosis.

Recently, computer-aided detection diagnosis (CAD) has gradually become a research hotspot in the area of medical imaging. The main idea of CAD is to assist radiologists in making clinical decision by using the report of computer system as a “second opinion” [6]. El-Dahshan et al. [6] developed a CAD system that used feedback pulse-coupled neural network for image segmentation, discrete wavelet transform for features extraction, the principal component analysis for feature reducing, and the feed forward backpropagation neural network for classifying inputs into normal or

abnormal. Zacharaki et al. [7] suggested that MRI texture and shape features in a machine learning scheme can help to evaluate the malignancy of brain tumor and identify certain tumor. The CAD system based on the use of automatic segmentation and ensemble classification techniques can accurately classify brain tumor as benign or malignant [8]. All of these studies have shown that CAD can help to improve the accuracy and efficiency of the diagnostic process and reduce the burden of work.

In the meantime, a technology called radiomics has developed rapidly, which explores the correlation between medical images and underlying genetic characteristics using a large number of automated data characterization algorithms to transform the region of interest (ROI) into high-throughput quantitative features. In 2008, Diehn et al. [9] found that the image features and genetic characteristics are highly correlated through the study of glioblastoma multiforme. The degree of enhancement within the tumor can indirectly reflect the expression level of epithelial growth factor receptor (EGFR). Patients with an infiltrative imaging phenotype have more tendency to have multiple tumor foci and show significantly shorter survival than the corresponding counterparts. In 2012, Dutch scholar Lambin et al. [10] formally proposed the concept of radiomics; that is, high-throughput extraction of a large number of image features from radiological images and the adoption of a large number of automated data characterization algorithms transform the image data in the ROI to high-resolution and exploitable spatial data. Hsieh et al. [11] proposed a computer-aided diagnosis system based on local and global MRI features to predict the malignancy of diffuse gliomas. The proposed image features have great potential in distinguishing glioblastomas from lower-grade gliomas. Lao et al. [12] proposed a deep learning-based radiomics model that has better performance than conventional models relying on explicitly designed handcrafted features for survival prediction in glioblastoma multiforme.

However, many CAD systems ignore the meaning of the features in the process of classification. Meanwhile, most radiomics studies focus on manual segmentation based ROI, which largely limit the development of radiomics. In this paper, we present an automatic computer-aided diagnosis for gliomas grading that combines automatic segmentation and radiomics. We use the MRI data provided by MICCAI Brain Tumor Segmentation Challenge 2015 containing 220 high-grade gliomas and 54 low-grade gliomas to evaluate our system. A multiscale 3D convolutional neural network is trained to segment whole tumor regions. Compared with the ground truth, satisfying segmentations of the whole tumor regions are obtained. A wide range of radiomic features including first-order features, shape features, and texture features is extracted from segmented tumor regions. Using support vector machines with recursive feature elimination for feature selection, we construct a 20-feature radiomic signature and train an extreme gradient boosting classifier. Robustness and accuracy results are obtained via 5-fold cross-validation. The experimental results show that the proposed CAD system is highly effective in the grading of gliomas. We further analyze the selected feature subset for gliomas grading. We also test our system using ground truth segmentation and achieve

comparable results. The system is automated and general and, if a significant number of subjects exist, has the potential for clinical applications.

2. Method

In this section, the methods involved in the CAD system are mainly introduced, which are organized as 5 parts: data preprocessing, automated tumor segmentation, radiomic features extraction, feature selection, and data classification.

2.1. Preprocessing. One difficulty in dealing with MRI scans is to deal with the artifacts caused by the inhomogeneity of the magnetic field and the small motions that the patient produces during scanning. The existence of artifacts has a great impact on the segmentation performance. N4ITK [13] is a widely used algorithm for removing bias field. In addition, the intensities may vary significantly between different scans because MRI scans were acquired from multiple institutions. Therefore, histogram matching algorithm [14] is used to transform each image to a specified histogram to ensure that all the images have the similar gray level ranges.

2.2. Segmentation. Segmentation is the most important and challenging part of a CAD system. The purpose of segmentation is to extract the ROI that is of great value to the diagnosis. This is a challenging task because of the unclear boundary of gliomas. Furthermore, gliomas may appear anywhere in the brain with different shapes and sizes. A lot of automated methods have been developed for brain tumor segmentation problem, including generative modeling method and discriminative model method. The former method relies on the prior knowledge of health and oncology tissue and the latter method relies on a large number of features extracted from the raw image such as local histograms and texture features. Relying on traditional machine learning methods, these segmentation methods have achieved great success. However, these methods may not be able to take full advantage of the training data due to the complexity of medical images. Recently, deep learning methods, especially convolutional neural networks (CNN), have led to great success in the field of medical image processing, such as classification and segmentation of medical images. These approaches incorporate feature engineering steps into learning steps that allow learning a wider range of features, thus contributing to a more robust representation of image data.

The CNN-based methods are performed by stacking several convolutional layers to form a hierarchy of features. The stack of input planes is fed into the convolutional layers, producing an output containing several feature maps by the application of kernels. The kernels, often interchangeably called filters, are arrays of values representing some sort of feature.

The automatic brain tumor segmentation method we used for CAD is DeepMedic, which is a 11-layer deep, multiscale 3D CNN first presented by Kamnitsas et al. [15] for brain lesion segmentation. Figure 1 is a simplified version of DeepMedic. As can be seen from the figure, the network

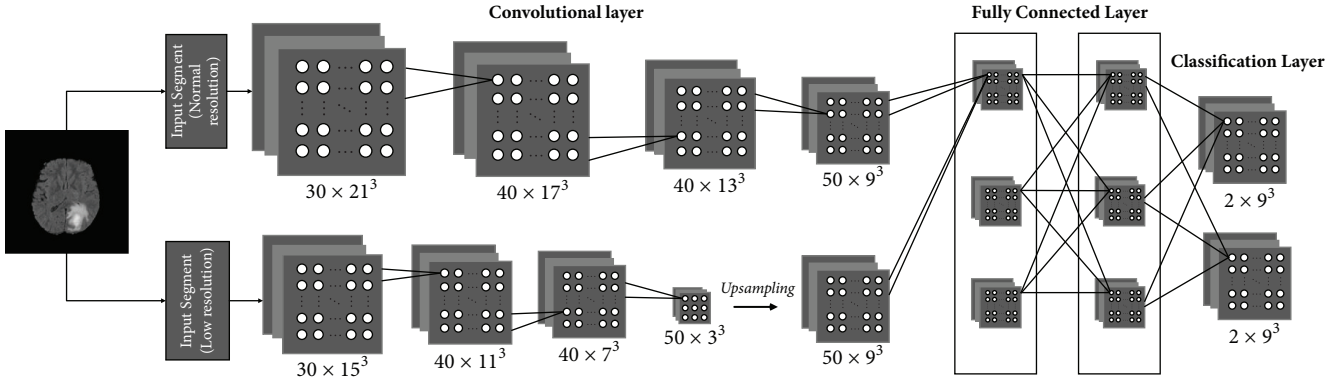


FIGURE 1: The basic architecture of DeepMedic.

consists of two pathways, one with normal resolution input segments and the remaining one with the lower resolution input segments. These segments then pass through a series of convolutional layers and two fully connected layers and finally a classification layer to produce soft segmentation maps.

One of the most prominent features of DeepMedic is the use of 3D CNNs as a basis for its architecture, which are characterized by convolving each layer's channels with a 3-dimensional kernel rather than traditional 2-dimensional kernel. This can be thought as a rectangular prism traversing the volume of the image at hand. Using these 3D CNNs can better represent volumetric data but will lead to increased computational costs. In order to overcome this problem, the dense inference technique was adopted. When the input image patch is larger than the CNN's receptive field, the classification layer can output multiple predictions.

In general, the deeper network with more convolutional layers has better discrimination power. This is attributed to the richer structures captured by the deeper models. However, with the deepening of the model, the computational cost is also getting higher and higher. To solve this problem, the small 3^3 kernels with less weights are adopted. These smaller kernels have the advantage of faster convolving speed and fewer parameters, thus contributing to a significant decrease in computational cost.

Deeper networks are also harder to train. As the network becomes deeper and deeper, preservation of loss functions becomes more and more difficult due to the multiplication of the variance of the loss function as it propagates through each layer. In order to alleviate this obstacle, DeepMedic initializes the kernel weights with the normal distribution $\mathcal{N}(0, \sqrt{2/n_l^{\text{in}}})$ where n_l^{in} is the number of weights through which a neuron of layer l is connected to its input. Batch normalization is also used to deal with the issue of covariate shift.

2.3. Radiomic Features Extraction. From the perspective of brain tumor medical characteristics and clinical cognition, doctors mainly determine the tumor grade from a lot of points. The brain tumor intensities of MRI scans with different modalities may have high signal, low signal, or equal signal due to different internal components. According

to the degree of intratumoral vessels, the amplitude of the signal enhancement after MRI enhanced scan is either high or low. The size and margins of the tumor may reflect the growth pattern of the tumor. Most brain tumors with uniform signal intensity are benign. In contrast, most brain tumors with nonuniform signals intensity are malignant. Higher-grade tumors usually show irregular shapes with larger size than lower-grade tumors. The high-grade tumors are more likely to appear to be significant enhancement of the signals, suggesting the destruction of the blood brain barrier, while low level tumors generally have no enhanced signals. The morphological and marginal features of the tumor can also provide some information to distinguish between benign and malignant tumors.

As mentioned above, conventional magnetic resonance imaging has been able to provide diagnostic information related to tumor grade, such as the location, shape, size, signal characteristics, mass effect, and peritumoral edema of brain tumors. However, these characteristics are often described subjectively and qualitatively. Recent advances in image analysis and machine learning facilitate the establishment of objective and accurate quantitative imaging descriptors that may be used as prognostic biomarkers [16].

Consistent with the features described by Imaging Biomarker Standardization Initiative (IBSI) [17], a wide range of radiomic features including first-order features, shape features, and texture features was extracted from the segmented brain tumor regions. First-order features (energy, entropy, standard deviation, skewness, kurtosis, etc.) describe the characteristics about voxel intensities and are calculated according to the intensity distribution and histogram. Shape features describe the three-dimensional size and shape of the tumors. These features can provide important diagnostic information; for example, the extent of tumor diffusion is usually measured by the longest diameter. Texture analysis contains a series of techniques that can quantify the gray-level pattern and the relationship of pixels through a wide range of statistical measures. Several studies have demonstrated the importance of texture features in evaluating the malignancy of gliomas [7, 11]. An extensive discussion of texture features can be seen in [18]. Inspired by them, we extracted a wider range of texture features as shown in Table 1. The detailed description of these features can be found in [17].

TABLE I: Texture features.

Texture feature groups	Abbreviation
Gray level cooccurrence matrix	GLCM
Gray level size zone matrix	GLSZM
Gray level run length matrix	GLRLM
Neighbouring gray tone difference matrix	NGTDM
Gray level dependence matrix	GLDM

2.4. Feature Selection. Too many features will increase the computational cost, and the redundancy between features will reduce the accuracy of the classification. Furthermore, the number of features is more than the number of samples in this work, which will increase the probability of overfitting. Therefore, feature selection is essential. There are two main types of feature selection algorithms: filtering methods and wrapping methods. Filter methods use a statistical measure based on the inherent information of samples to compute a score for each feature. The features are then ranked according to their scores, and the top-ranked features are kept as the final features for classification. However, the selected features may contain many redundant features because the scores are calculated independently for each feature, completely ignoring the dependencies on other features. On the contrary, the wrapping methods consider a subset of features that have best discrimination power by evaluating the performance of the classifier using this feature subset. Therefore, wrapping-based feature selection methods is more suitable for selecting radiomic features because of their obvious correlation.

A quintessential example of wrapping methods is Support Vector Machine Recursive Feature Elimination (SVM-RFE) algorithm, which was proposed by Guyon et al. [19] for selecting gene data for cancer classification. SVM-RFE is a feature selection algorithm utilizing support vector machine (SVM) methods based on recursive feature elimination (RFE). SVM is a widely used classification technique based on the idea of finding a separating hyperplane that best divides a dataset into two classes [20]. The RFE method selects features by recursively considering smaller and smaller sets of features according to their weights assigned by an external estimator. To be specific, given training examples $\mathbf{X}_0 = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_l]^T$ and class labels $y = [y_1, y_2, \dots, y_k, \dots, y_l]^T$; let $\mathbf{s} = [1, 2, \dots, n]$ be the subset of surviving features. In each loop, an SVM classifier is trained on \mathbf{s} , and the corresponding weight vector \mathbf{w} is calculated. The features are then ranked based on \mathbf{w} , and the features with smallest ranking criterion are removed from \mathbf{s} . This procedure is repeated until $\mathbf{s} = []$. Thus, we can obtain the feature ranked list \mathbf{r} using feature removing sequence, as the features removed later are more important. The features removed in the last iteration have the highest rankings. In order to find the optimal number of features to be selected, cross-validation can be used. The SVM-RFE with cross-validation starts with all the features, computes the cross-validated performance score, and removes the lowest k features. These procedures are repeated until all the features are eliminated. Finally, a feature subset with the best performance score can be obtained.

2.5. Classification. In order to solve the problem of a relatively small dataset in our work, an ensemble learning model is used to trade off the approximation error and estimation error. Ensemble learning model incorporates a variety of machine learning algorithms and typically achieves significantly superior generalization performance over a single learner. Gradient boosting tree (XGBoost) [21] is a kind of ensemble learning models for regression and classification problems, which generates a prediction model in the form of an ensemble of decision trees.

Given a dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$, the tree ensemble model can be seen as an additive model consisting of K trees.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F}, \quad (1)$$

where \mathcal{F} is the tree space containing all possible regression trees. The function f_k , which we need to learn, corresponds to the tree structure and leaf weights. In contrast to traditional decision tree, every leaf in the tree consists of a continuous score, and the final prediction is based on the summation of these scores. To learn the model, the following regularized objective will be minimized.

$$\mathcal{L}(\phi) = \sum_i^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

where $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$.

In (2), l is restricted to be a differentiable convex loss function that measures the difference between the prediction \hat{y}_i and the target y_i . The second term $\Omega(f)$ penalizes the complexity of tree in terms of the number of leaves in the tree T with coefficient γ and the vector of scores on leaves w with coefficient λ . Two additional techniques, shrinkage and column subsampling, are further used to reduce overfitting. Shrinkage reduces the influence of each tree by scaling newly added weights and thus leaves space for the following trees to improve the model while reducing the possibility of overfitting. Similar to random forest, column subsampling only considers the attributes of a random subset to establish a tree. This will not only improve the speed of training, but also help to reduce the overfitting.

3. Result

3.1. MRI Data. Experiments were carried out on the MR images obtained from BraTS'15 challenge [22], which contains 220 high-grade gliomas (HGG) and 54 low-grade gliomas (LGG). Each subject contains four modalities: native T1-weighted (T1), postcontrast T1-weighted (T1c), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). All subjects were confirmed pathologically by experts. The HGG group consists of anaplastic astrocytomas and glioblastoma multiforme tumors, and the LGG group consists of astrocytomas and oligoastrocytomas. All images were skull stripped and registered to the T1c image and resampled to 1 mm isotropic resolution within a standard axial orientation. All images were labeled into five parts: necrosis, edema, active-enhanced tumor, nonenhanced

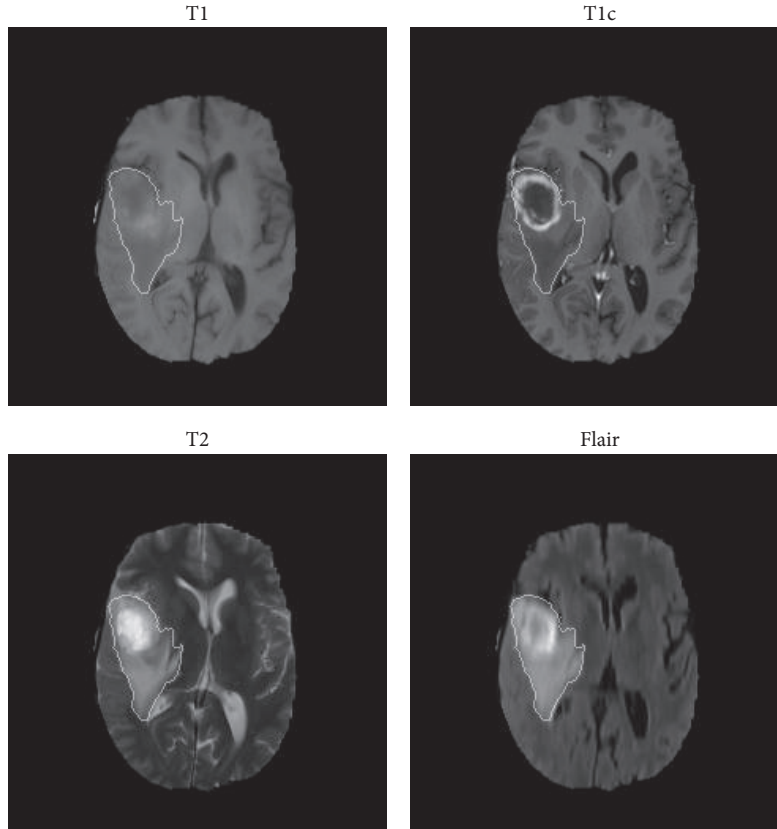


FIGURE 2: The whole tumor region with ground truth segmentation.

tumor, and normal region. In this work, we merged all tumor subregions into the whole tumor region as shown in Figure 2.

3.2. Preprocessing. The N4ITK bias correction method was applied to all scans using advanced normalization tools (ANTS) [23]. Then the histogram matching algorithm was applied to transform each scan to a specified histogram using SimpleITK [24]. The images of “brats_2013_pat0001_1” were selected as reference samples. Then all images were normalized to the zero-mean and unit variance through subtracting the mean and being divided by the standard deviation. Some preprocessing results are shown in Figure 3. Column (a) shows the reference images, column (b) shows the images before preprocessing of a random selected subject, and column (c) shows the results after preprocessing.

3.3. Segmentation. The detailed architecture of the network is shown in Table 2. We have four modalities, that is, T1, T1c, T2, and FLAIR. Therefore, a total of four channels are fed into the network. The predicted labels are classified into two sets: whole tumor (including necrosis, edema, active-enhanced tumor, and nonenhanced tumor) and nontumor region. There are two parallel identical pathways, and the number of feature maps of each convolutional layer is [30, 30, 40, 40, 40, 40, 50, 50]. The input dimensions of the normal resolution path is $25 * 25 * 25$, and the input dimension of the low resolution path is $19 * 19 * 19$. Two hidden

TABLE 2: Network architectures.

Structure name	Value
Input channels	T1, T1c, T2, FLAIR
Output classes	Tumor, normal
Pathways	2
FMs/layer	30, 30, 40, 40, 40, 40, 50, 50
FMs/Hidd	150, 150
Seg. norm	$25 * 25 * 25$
Seg. low	$19 * 19 * 19$
Batch size	10

layers with 150 feature maps follow the concatenation of the pathways. The network was trained on a GeForce GTX 1080 Ti GPU with a batch size of 10. RMSProp was used as the optimizer. The start learning rate was set to 0.001.

Multiple criteria were computed as performance metrics to quantify the segmentation results. Dice coefficient (see (3)) is the most commonly used metric for evaluating medical image segmentations [25]. P is the area that is predicted to be tumor and G is true tumor area. It measures the overlap between the automatic segmentations and ground truth with a value between 0 and 1. The higher the Dice score, the better the segmentation performance.

$$\text{Dice}(P, G) = \frac{P \cap G}{(|P| + |G|) / 2}. \quad (3)$$

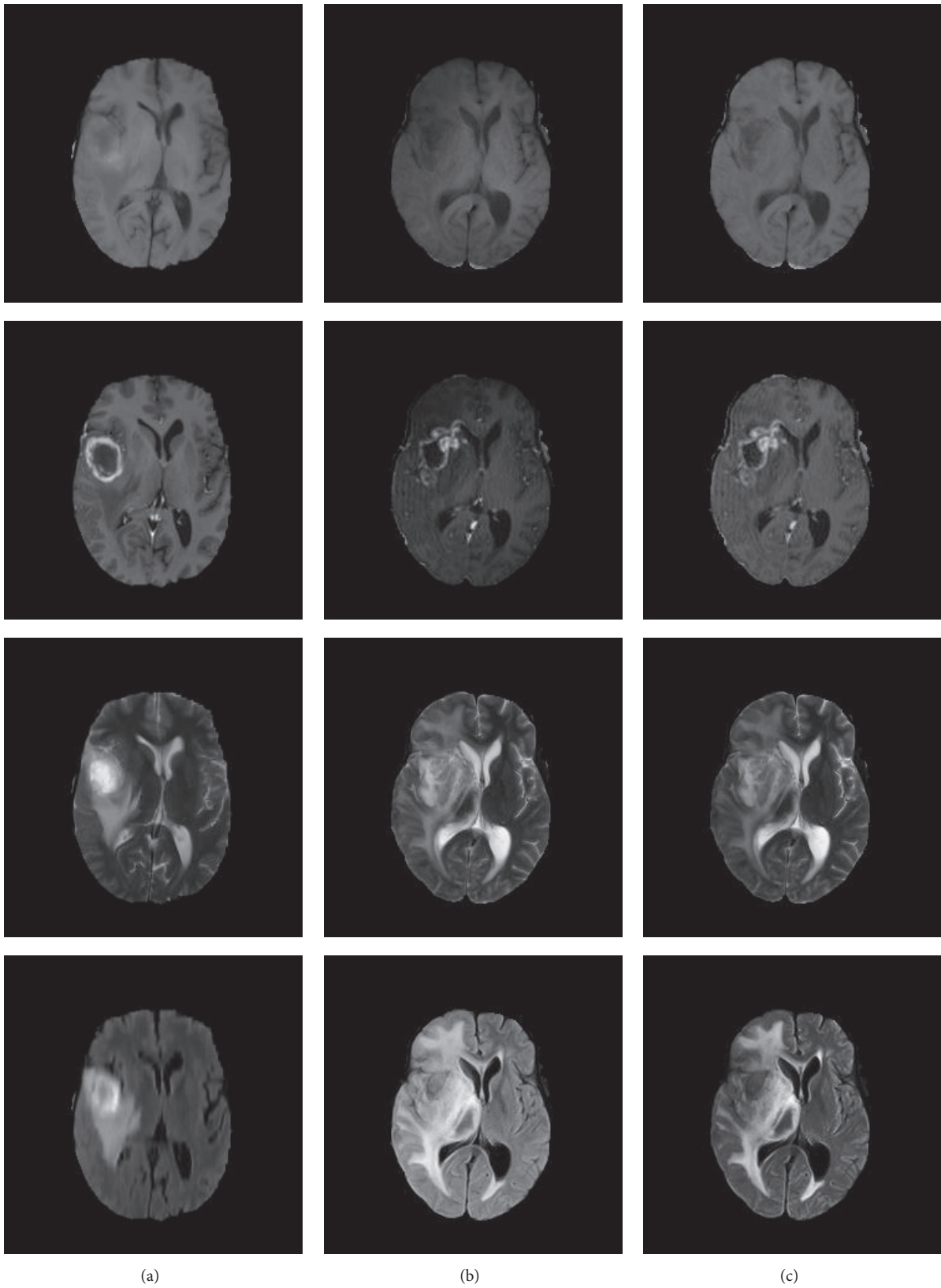


FIGURE 3: The results of preprocessing.

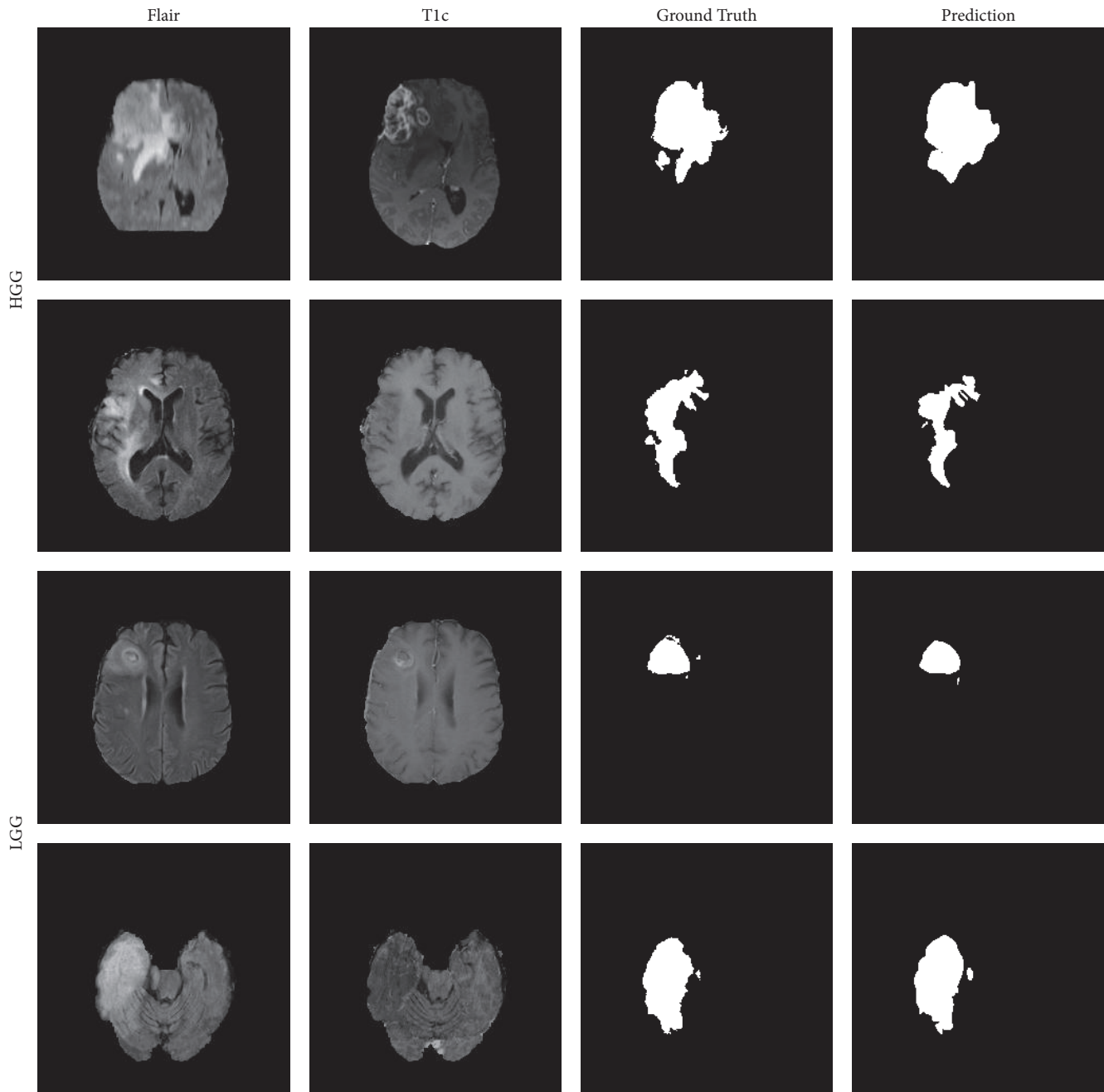


FIGURE 4: Examples of segmentation results.

Sensitivity and specificity are also commonly used statistical measures. The sensitivity, also called true positive rate, is defined as the proportion of positives that are correctly predicted. It measures the portion of tumor regions in the ground truth that are also predicted as tumor regions by the automatic segmentation method. The specificity, also called true negative rate, is defined as the proportion of negatives that are correctly predicted. It measures the portion of normal tissue regions in the ground truth that are also predicted as normal tissue regions by the automatic segmentation method.

We trained and evaluated our network on all the 274 subjects via 5-fold cross-validation and achieved a Dice

score of 0.89, a sensitivity of 0.89, and a specificity of 0.90. Four examples of the automatic segmentations on testing data are shown in Figure 4. The first two rows depict the segmentations from HGG subjects, and the next two rows depict the segmentations from LGG subjects. Compared with the ground truth, satisfying segmentations of the whole tumor regions are obtained.

3.4. Radiomic Features Extraction and Selection. A wide range of radiomic features was extracted from the whole tumor regions. This step was performed by Pyradiomics [26], which

TABLE 3: The feature subset based on automated segmentation.

Selected feature	Group	Modality
Cluster Shade	GLCM	T1
90 Percentile	First-order	T1c
Energy	First-order	T1c
Mean Absolute Deviation	First-order	T1c
Minimum	First-order	T1c
Root Mean Squared	First-order	T1c
Total Energy	First-order	T1c
Small Dependence Emphasis (SDE)	GLDM	T1c
Small Dependence Low Gray Level Emphasis (SDLGLE)	GLDM	T1c
High Gray Level Run Emphasis (HGLRE)	GLRLM	T1c
Low Gray Level Run Emphasis (LGLRE)	GLRLM	T1c
Gray Level Variance (GLV)	GLSZM	T1c
Large Area Emphasis (LAE)	GLSZM	T1c
Maximum	First-order	T2
Correlation	GLCM	T2
Informal Measure of Correlation 1 (Imc1)	GLCM	T2
Informal Measure of Correlation 2 (Imc2)	GLCM	T2
Maximum	First-order	FLAIR
Informal Measure of Correlation 1 (Imc1)	GLCM	FLAIR
Large Area Emphasis (LAE)	GLSZM	FLAIR

can extract a wide range of radiomic features from brain tumor MRI. For each modality, 105 3D-Radiomic features were extracted, including 18 first-order features, 13 shape features, and 74 texture features. Thus, a total of 420 radiomic features were extracted from each subject.

Because of too many features, a feature selection step should be required prior to establishing a classification model. This step not only reduces overfitting but also improves generalization of the classification model. In order to select the most discriminating feature subset, we used 5-fold cross-validation. For each fold, SVM-RFE with leave-one-out cross-validation algorithm was used to find the optimal features that best predict the malignancy of gliomas. Finally, 20 features were commonly selected for all folds, as shown in Table 3. A detailed description of these features can be found in Pyradiomics documentation website (<http://pyradiomics.readthedocs.io>).

As we can see from Table 3, 1 feature from the T1 image, 12 features from the T1c image, 4 features from the T2 image, and 3 features from the FLAIR image were selected. The features from T1 image make up the majority, suggesting that the T1c image best reflects the difference between HGG and LGG. In terms of feature groups, none of the shape features were selected. These results are consistent with the research in [27].

3.5. Classification. Implementation of XGBoost was performed in python using the open-source code provided in [21]. In addition to XGBoost, two other pattern classification methods were investigated for comparison: Extremely Randomized Trees (ERT) [28] and support vector machine (SVM) [20]. Extremely Randomized Trees are similar to

random forest but differ in two ways. Extremely Randomized Trees do not use the bagging procedure when building a tree. Like random forest, a random set of attributes were selected at every node. Extremely Randomized Trees go one step further and select the splitting attribute randomly, rather than finding the best split among random subset of variables. A SVM is a discriminative classifier based on the idea of finding a separating hyperplane that best divides a dataset into two classes. SVM is widely used for different tasks due to its effectiveness in high dimensional spaces.

In order to assess the robustness and effectiveness of the proposed CAD system, we also carried out our method based on the ground truth segmentation. The classification results are presented in Table 4. The first row in Table 4 shows four performance indices: accuracy, weighted macroprecision, weighted macrorecall, and weighted macro-F1 score. It is worth noting that “weighted” means that it takes class imbalance into account and can result in an F -score that is not between precision and recall. The values in bold show the best classification rank. The results show that the classification performance is better when using XGBoost, as expected. We also use the receiver operation characteristic (ROC) curve as another method to evaluate the overall performance (Figure 5). For XGBoost, the area under the ROC curve based on automated segmentation is 0.95, while that based on ground truth segmentation is 0.96. From the results above, it can be seen that classification results based on different segmentations have no significant differences.

We further investigate the selected feature subset based on ground truth segmentation, as shown in Table 5. A detailed description of these features can be found in Pyradiomics documentation website (<http://pyradiomics.readthedocs.io>).

TABLE 4: Average performance via 5-fold cross-validation.

	Accuracy	Precision	Recall	F1 score
XGBoost + automatic	91.27%	91.27%	91.27%	90.64%
XGBoost + Ground truth	91.25%	91.63%	91.25%	91.06%
ERT + automatic	90.98%	90.94%	90.89%	90.21%
ERT + Ground truth	90.52%	90.50%	90.52%	89.67%
SVM + automatic	90.16%	90.12%	90.16%	89.24%
SVM + Ground truth	90.16%	90.20%	90.16%	89.12%

TABLE 5: The feature subset based on ground truth segmentation.

Selected feature	Group	Modality
90 Percentile	First-order	T1
Cluster Shade*	GLCM	T1
Maximum Probability (MP)	GLCM	T1
90 Percentile*	First-order	T1c
Kurtosis	First-order	T1c
Mean	First-order	T1c
Mean Absolute Deviation*	First-order	T1c
Root Mean Squared*	First-order	T1c
Skewness	First-order	T1c
Dependence Nonuniformity (DN)	GLDM	T1c
Small Dependence Emphasis (SDE)*	GLDM	T1c
Small Dependence Low Gray Level Emphasis (SDLGLE)*	GLDM	T1c
High Gray Level Run Emphasis (HGLRE)*	GLRLM	T1c
Low Gray Level Run Emphasis (LGLRE)*	GLRLM	T1c
Run Length Nonuniformity (RLN)	GLRLM	T1c
Large Area High Gray Level Emphasis (LAHGLE)	GLSZM	T1c
Small Area High Gray Level Emphasis (SAHGLE)	GLSZM	T1c
Maximum*	First-order	T2
Correlation*	GLCM	T2
Informal Measure of Correlation 1 (Imc1)*	GLCM	T2
Informal Measure of Correlation 2 (Imc2)*	GLCM	T2
Dependence Nonuniformity Normalized (DNN)	GLDM	T2
Low Gray Level Zone Emphasis (LGLZE)	GLSZM	T2
Cluster Shade	GLCM	FLAIR
Large Area High Gray Level Emphasis (LAHGLE)	GLSZM	FLAIR

As we can see, 3 features from the T1 image and 14 features from the T1c image and 6 features from the T2 image and 2 features from the FLAIR image were selected. Consistent with the selected features based on automatic segmentation, the features of T1c image best reflect the difference between HGG and LGG, and none of the shape features were selected. The features with “*” are also selected in the radiomic signature based on automatic segmentation. There are 12 such features in total, including 1 feature from the T1 image, 7 features from the T1c image, and 4 features from the FLAIR image. This demonstrates that CAD system based on automatic segmentation achieved consistent results with the system based on ground truth segmentation. This is probably because most of the features are based on spatial averaging and highly accurate segmentation results may not be necessary.

4. Conclusion

This paper presents an automated scheme using multiparametric MRI scans for differentiating the malignancy of gliomas. A deep learning-based method, called DeepMedic, is used to automatically detect tumor regions. A wide range of radiomic features is extracted from segmentation regions. Using SVM-RFE for feature selection, we construct a 20-feature radiomic signature and train a XGBoost classifier. The results demonstrate that XGBoost-based classification is an objective and promising approach for evaluating the malignancy of gliomas.

Radiomics has been greatly developed in recent years, and its goal is to extract quantitative features as a clinical decision support tool. However, most of these studies are based on manually segmented ROIs, which limit the development

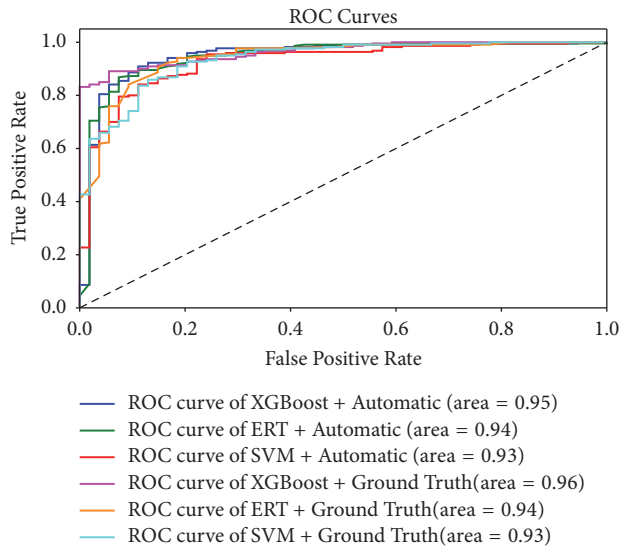


FIGURE 5: Receiver operating characteristic curve.

of radiomics. Recently, CNN-based approaches, such as DeepMedic, U-net [29], and V-net [30], have achieved great success in the field of medical image segmentation. All of these motivate us to develop a fully automatic method of combining automated brain tumor segmentation with radiomics to develop a fully automated CAD system.

The features that make up the signature contain all MRI modalities. This means all MRI modalities play important roles in our classification task. However, only first-order and texture features are contained in the signature, probably because the shape features of the whole tumor have poor discrimination ability of MR images. In the future work, we will explore the discrimination ability of brain tumor subregions, such as necrosis, edema, active-enhanced tumor, and nonenhanced tumor.

Our work demonstrates that the automatic segmentation of brain tumors can be integrated into computer-aided diagnosis. The classification results based on automated segmentation and ground truth segmentation have no significant differences. This is probably because most of the features are based on spatial averaging; highly accurate segmentation results may not be necessary. Our work provides an alternative idea that we can use automatic segmentation for radiomic features extraction. This will reduce the difficulty of radiomics research since manual segmentation is a time-consuming task.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Department of Science and Technology of Shandong Province (Grants nos. 2015ZDXX0801A01, ZR2014HQ054, and 2017CXGC1502)

and National Natural Science Foundation of China (Grant no. 61603218).

References

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 1, pp. 7–30, 2017.
- [2] D. N. Louis, A. Perry, G. Reifenberger et al., "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathologica*, vol. 131, no. 6, pp. 803–820, 2016.
- [3] M. Glas, C. Happold, J. Rieger et al., "Long-term survival of patients with glioblastoma treated with radiotherapy and lomustine plus temozolomide," *Journal of Clinical Oncology*, vol. 27, no. 8, pp. 1257–1261, 2009.
- [4] R. Ramakrishna, A. Hebb, J. Barber, R. Rostomily, and D. Silbergeld, "Outcomes in Reoperated Low-Grade Gliomas," *Neurosurgery*, vol. 77, no. 2, pp. 175–184, 2015.
- [5] M. O. Leach, K. M. Brindle, J. L. Evelhoch et al., "The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations," *British Journal of Cancer*, vol. 92, no. 9, pp. 1599–1610, 2005.
- [6] E. A.-S. El-Dahshan, H. M. Mohsen, K. Revett, and A.-B. M. Salem, "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014.
- [7] E. I. Zacharaki, S. Wang, S. Chawla et al., "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.
- [8] M. P. Arakeri and G. R. M. Reddy, "Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images," *Signal, Image and Video Processing*, vol. 9, no. 2, pp. 409–425, 2013.
- [9] M. Diehn, C. Nardini, D. S. Wang et al., "Identification of non-invasive imaging surrogates for brain tumor gene-expression modules," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 13, pp. 5213–5218, 2008.
- [10] P. Lambin, E. Rios-Velazquez, R. Leijenaar et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012.
- [11] K. L.-C. Hsieh, C.-M. Lo, and C.-J. Hsiao, "Computer-aided grading of gliomas based on local and global MRI features," *Computer Methods and Programs in Biomedicine*, vol. 139, pp. 31–38, 2017.
- [12] J. Lao, Y. Chen, Z.-C. Li et al., "A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme," *Scientific Reports*, vol. 7, no. 1, Article ID 10353, 2017.
- [13] N. J. Tustison, B. B. Avants, P. A. Cook et al., "N4ITK: improved N3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.
- [14] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 143–150, 2000.
- [15] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

- [16] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar et al., “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach,” *Nature Communications*, vol. 5, article no. 4006, 2014.
- [17] A. Zwanenburg, L. Stefan, M. Vallières, L. Steffen et al., “Image biomarker standardisation initiative-feature definitions,” *Computer Vision and Pattern Recognition*, 2016.
- [18] A. Kassner and R. E. Thornhill, “Texture analysis: a review of neurologic MR imaging applications,” *American Journal of Neuroradiology*, vol. 31, no. 5, pp. 809–816, 2010.
- [19] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [20] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [21] T. Chen and C. Guestrin, “XGBoost: a scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2016*, pp. 785–794, 2016.
- [22] H. Bjoern Menze, J. Andras, B. Stefan et al. et al., “The multimodal brain tumor image segmentation benchmark (brats),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [23] B. B. Avants, N. Tustison, and G. Song, “Advanced normalization tools (ants),” *Insight Journal*, vol. 2, pp. 1–35, 2009.
- [24] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, “The design of simpleITK,” *Frontiers in Neuroinformatics*, vol. 7, 2013.
- [25] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool,” *BMC Medical Imaging*, vol. 15, no. 1, p. 29, 2015.
- [26] J. J. M. Van Griethuysen, A. Fedorov, C. Parmar et al., “Computational radiomics system to decode the radiographic phenotype,” *Cancer Research*, vol. 77, no. 21, pp. e104–e107, 2017.
- [27] H.-H. Cho and H. Park, “Classification of low-grade and high-grade glioma using multi-modal image radiomics features,” in *Proceedings of the Engineering in Medicine and Biology Society (EMBC), 39th Annual International Conference of the IEEE*, pp. 3081–3084, IEEE, 2017.
- [28] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [29] O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.
- [30] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: fully convolutional neural networks for volumetric medical image segmentation,” in *Proceedings of the 3D Vision (3DV), Fourth International Conference on*, pp. 565–571, IEEE, 2016.