Computer-aided Grading of Gliomas Based on Local and Global MRI Features

Abstract

Background and Objectives: A computer-aided diagnosis (CAD) system based on quantitative magnetic resonance imaging (MRI) features was developed to evaluate the malignancy of diffuse gliomas, which are central nervous system tumors.

Methods: The acquired image database for the CAD performance evaluation was composed of 34 glioblastomas and 73 diffuse lower-grade gliomas. In each case, tissues enclosed in a delineated tumor area were analyzed according to their gray-scale intensities on MRI scans. Four histogram moment features describing the global gray-scale distributions of gliomas tissues and 14 textural features were used to interpret local correlations between adjacent pixel values. With a logistic regression model, the individual feature set and a combination of both feature sets were used to establish the malignancy prediction model.

Results: Performances of the CAD system using global, local, and the combination of both image feature sets achieved accuracies of 76%, 83%, and 88%, respectively. Compared to global features, the combined features had significantly better accuracy (p=0.0213). With respect to the pathology results, the CAD classification obtained substantial agreement κ=0.698, p<0.001.

Conclusions: Numerous proposed image features were significant in distinguishing glioblastomas from lower-grade gliomas. Combining them further into a malignancy prediction model would be promising in providing diagnostic suggestions for clinical use.

Keywords: brain tumor, diffuse glioma, glioblastoma, computer-aided diagnosis, image moment, gray-level co-occurrence matrix, magnetic resonance imaging

Introduction

Gliomas are central nervous system (CNS) tumors formed of neoplastic cells that display glial cell differentiation. According to the World Health Organization (WHO) classification of tumors of the CNS, diffuse gliomas can be subdivided by the degree of malignancy into WHO grade II (lower grade) to grade IV (high malignancy) [1, 2]. Glioblastomas (GBMs), WHO grade IV tumors, are the most aggressive tumor type with a dismal prognosis despite advances in therapeutic management [3]. In contrast to GBMs, diffuse lower-grade gliomas (LGGs, grades II and III) have more-favorable outcomes and shared many similar histopathologic and genomic signatures [2, 4]. Since the therapeutic approach of them are also different [5], distinguishing GBM from LGG is a very critical clinical issue. Determining the tumor grade depends on several pathological features including cytological atypia, mitotic activity, angiogenesis, and necrosis. However, there are still some pitfalls in the histopathological analysis which can lead to ambiguity in glioma grading. For example, interpretation of some criteria can vary because their definitions are semiquantitative or imprecise [6, 7]. Moreover, the heterogeneous expressions of aggressive cellular features make unguided surgical biopsies prone to sampling error, resulting in misgrading in up to 30% of cases [7-11].

With the development of diagnostic imaging technologies, the accuracy of estimating the malignancy of brain tumors has greatly increased by applying magnetic resonance (MR) imaging (MRI) features [12, 13]. MRI is commonly used because it provides a wide range of physiologically meaningful contrasts to distinguish different tissues by imaging, and therefore improves evaluations of heterogeneous patterns of tissue compositions within diffuse gliomas [14]. In addition to conventional sequences, several MRI techniques including diffusion-weighted imaging (DWI), MR spectroscopy (MRS), and perfusion-weighted imaging (PWI), are also applied to non-invasively differentiate LGGs from GBMs [15-18]. A previous study supported MRI scans being highly specific for diagnosing brain stem gliomas and can replace biopsies before radiotherapy in most patients [19]. To avoid unnecessary operations, the role of MRI in the diagnostic imaging of brain tumors is especially crucial.

Computer-aided diagnosis (CAD) systems based on quantitative image features and artificial intelligence classifiers were developed to assist radiologists in determining tumor types and grades [20-22]. With machine learning schemes, textural features extracted from MRI scans are used to classify different tissue types which can assist clinical decision-making regarding initial and evolving treatment strategies [23]. CAD systems can quantitatively combine numerous imaging features to estimate the likelihood of tumor malignancy by percentages. Efficient and consistent procedures can provide reliable suggestions to radiologists to avoid invasive procedures for which risks outweigh benefits.

In this study, local and global imaging features extracted from the entire tumor area on MRI scans were quantified to reveal levels of heterogeneity. Quantified image features were combined in a logistic regression classifier to generate a prediction model for each case. The performances of an individual image feature set and the combination of both local and global features were evaluated in the experiment. As a second viewer, the CAD can provide suggestions of tumor grading to the radiologists on clinical examinations.

Materials and Methods

Patient information

The Cancer Genome Atlas (TCGA) and the Cancer Imaging Archive (TCIA)

MRI datasets for 34 GBM and 73 LGG patients were obtained from TCIA (*http://cancerimagingarchive.net/*) of the National Cancer Institute, a portal containing images of TCGA patients for image analysis. The collection of original materials and data provided by TCGA project was conducted in compliance with all applicable laws, regulations, and policies for the protection of human subjects. All necessary approvals, authorizations, human subject assurances, informed consent documents, and IRB approvals were obtained [24]. The images used in this research were generated from three institutes: Henry Ford Hospital, Thomas Jefferson University, and Case Western hospitals as shown in Table 1. **All images used in this research were created before any operative procedure including surgical biopsy.**

There were totally 34 GBMs (grade 4) (http://dx.doi.org/10.7937/K9/TCIA.2016.RNYFUYE9) and 73 LGGs (grades 2 and 3) (http://dx.doi.org/10.7937/K9/TCIA.2016.L4LTD3TK) included in this study. In the LGG group, there were 33 oligodendrogliomas, 16 oligoastrocytomas, and 24 astrocytomas. Nineteen oligodendrogliomas were classified into grade 2, and 14 cases were classified into grade 3. Seven cases of oligoastrocytoma were classified into grade 2, and nine cases were classified into grade 3. Among astrocytomas, four cases were classified into grade 2, and 20 cases were classified into grade 3. Therefore, we had totals of 30 grade 2 and 43 grade 3 gliomas in the LGG group.

Image analysis

The MRI sequence used for the analysis was the contrast-enhanced axial T1-weighted image (T1WI). Imaging features were quantitatively analyzed by procedures described herein. A board-certified neuroradiologist (K.H., with 12 years of experience) who was blinded to the clinical information selected the most representative 2D image of each tumor. Intensity normalization which extended the gray-level distribution of each MRI image to the whole value range (0-255) was performed to enhance the contrast between tumor and background tissues for contour delineation. Regions-of-interests (ROIs) were then outlined manually using OsiriX in the selected contrast-enhanced T1WI. Pixels encircled in the ROI were used for feature analysis.

Image features

Global statistics

Observing the gray-scale distribution of the tumor region, the composition of pixel values in the region can be presented by a probability distribution. The regional distribution formed a histogram which contained global statistics of the tissue properties which can be characterized by the histogram moments [25, 26]. Quantification of the moments provided objective measures of the shape which were used to express the difference between LGGs and GBMs in the experiment. The first-, second-, third-, and fourth-order central moments of the gray-scale histograms were calculated as the global statistical features, i.e., the mean, variance, skewness, and kurtosis.

$$\text{Mean} = \frac{1}{N}\sum_{i=1}^{N} P_i \qquad (1)$$

$$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N} (P_i - Mean)^2 \qquad (2)$$

$$\text{Skewness} = \frac{1}{N}\sum_{i=1}^{N}(P_i - Mean)^3 \quad (3)$$

$$\text{Kurtosis} = \frac{1}{N}\sum_{i=1}^{N}(P_i - Mean)^4 \quad (4)$$

P_i is the gray-scale pixel value. The mean is the center of a distribution obtained by summarizing all pixel values and dividing this by the number of pixels in a tumor region. Variance measures how far the gray-scale values are spread out. Skewness estimates the symmetry of a distribution such as a bias to the left or right side. Compared to a normal distribution, kurtosis is a single-peaked shape with heavily weighted tails.

Local statistics

Detailed correlations between adjacent image pixels were the local statistics of tumor characteristics. For pattern recognition, local statistics were used to describe textures to identify different objects. Because the compositions of MRI scans are intensities with gray-level values, the gray-level co-occurrence matrix (GLCM) [27] which presents the local statistics can be calculated and are features distinguishing LGGs and GBMs. An original image was first quantified into an image, *G*, with intensity bins. From *G*, co-occurrence matrices $P=[p(i,j|d,\theta)]$ were generated to express the frequencies of each pixel (gray value *i*) and its neighboring pixels (gray value *j*) at distance *d* and direction θ. As shown in Fig. 3, d=1 and θ=0°, 45°, 90°, and 135° were used in the experiment for the defined local area. From the matrices, the GLCM features were extracted:

$$\textit{Autocorrelation} = \sum_{i}\sum_{j} (p_x - \mu_x)(p_y - \mu_y) / \sigma_x\sigma_y \quad (5)$$

Contrast= $$\sum_{n} n^2 \left\{ \sum_{i} \sum_{j} p\,(i,j) \right\}, |i-j| = n \qquad (6)$$

Correlation = $$\frac{\sum_i \sum_j (i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x \sigma_y} \qquad (7)$$

Cluster Prominence = $$\sum_{i} \sum_{j} (i+j-\mu_x-\mu_y)^4 p(i,j) \qquad (8)$$

Cluster Shade = $$\sum_{i} \sum_{j} (i+j-\mu_x-\mu_y)^3 p(i,j) \qquad (9)$$

Dissimilarity= $$\sum_{i} \sum_{j} p(i,j)|i-j| \qquad (10)$$

Energy = $$\sum_{i} \sum_{j} p(i,j)^2 \qquad (11)$$

Entropy = $$-\sum_{i} \sum_{j} p(i,j)\log(p(i,j)) \qquad (12)$$

Homogeneity = $$-\sum_{i} \sum_{j} \frac{1}{1+i-j} p(i,j) \qquad (13)$$

Difference variance= $$\sum_{i} i^2 p_{x-y}(i) \qquad (14)$$

Difference entropy= $$-\sum_{i} p_{x+y}(i)\log(p_{x+y}(i)) \qquad (15)$$

Information measure of correlation= (16)

$$\frac{HXY - HXY1}{max\{HX, HY\}}$$

$$HXY = (8),$$

$$HXY1 = -\sum_{i} \sum_{j} p(i,j)\log(p_x(i)p_y(j))$$

$$HX = entropy\ of\ p_x,$$

$$HY = entropy\ of\ p_y$$

Inverse difference normalized= $$\sum_i \sum_j \frac{1}{1+|i-j|} p(i,j) \quad (17)$$

Inverse difference moment = $$\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j) \quad (18)$$

where μ_x, μ_y, σ_x and σ_y are the mean and standard deviation (SD) of the marginal distributions of $p(i,j|d,\theta)$.

$$\mu_x = \sum_i i \sum_j p(i,j), \mu_y = \sum_j j \sum_i p(i,j) \quad (19)$$

$$\sigma_x^2 = \sum_i (i-u_x)^2 \sum_j p(i,j), \sigma_y^2 = \sum_j (j-u_y)^2 \sum_i p(i,j) \quad (20)$$

Statistical analysis

The image features proposed above, including global and local statistics, were evaluated as to whether they could distinguish between LGG and GBM tumors. The feature value distributions were first evaluated by the Kolmogorov-Smirnov test [28] to determine their normalities. Normal image features were subjected to Student's *t*-test [28], and non-normal image features were evaluated by the Mann-Whitney U-test [28]. Resulting *p* values of <0.05 indicated that features were statistically significant in distinguishing between LGG and GBM tumors.

Another evaluation method was the prediction performance of these image features. Using a binary logistic regression as the classifier, global and local image features were combined into respective feature sets. First, the performance of an individual feature set was generated. Then, the two feature sets were combined to see the complementary power. When establishing a prediction model, biopsy-proven pathology results were acquired as the gold standard in the classifier. Step-wise backward elimination removed redundant features based on their abilities, and the most relevant features with the smallest error rates were selected. Leave-one-out cross-validation [28] was used to evaluate the generalizability of the selected features. In the iteration loop, one case was separated from the total *n* cases and was used to test the trained model from the remaining *n*-1 cases.

According to the pathology results, the performance of the prediction model can be presented using five general performance indices: accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). In the determination of an LGG or GBM, cases with a predicted probability of >0.5 were regarded as GBMs to obtain the best tradeoff between the sensitivity and specificity. Different points of tradeoff combinations were also calculated and illustrated using a receiver operating characteristic (ROC) curve. To provide an overall performance evaluation, the area under the ROC curve, Az, was formulated using ROCKIT software (C. Metz, University of Chicago, Chicago, IL, USA).

The agreement between the prediction model of the CAD system and the pathology results was obtained by Cohen's kappa statistic (κ) [28]. Generally, the agreement was slight if the κ value was <0.20; fair if κ was in the range of 0.21~0.40; moderate if κ was in the range of 0.41~0.60; substantial if κ was in the range 0.61~0.80; and almost perfect, if κ was in the range of 0.81~1.00. The test and correlation analyses were carried out using SPSS software (vers. 16 for Windows; SPSS, Chicago, IL, USA).

Results

According to distributions of feature values, the proposed global and local image features were tested by either Student's *t*-test (for those with a normal distribution) or the Mann-Whitney U-test (for those with a non-normal distribution). Tables 2 and 3 show the statistical data and *p* values, respectively, of significant features in distinguishing LGG from GBM tumors. Three of four global image features achieved *p* values of <0.001, and nine local image features had *p* values of <0.05.

Taking the pathology results as the standard for tumor grading, performances of the global image feature sets achieved an accuracy of 76%, a sensitivity of 68%, a specificity of 79%, and an Az of 0.78, while local image feature sets achieved an accuracy of 83%, a sensitivity of 79%, a specificity of 85%, and an Az of 0.89 (Table 4). Overall, the local image feature set performed better than the global image feature set. However, differences in performances were not significant. Combining both global and local image features together for the tumor classification achieved even better performance: an accuracy of 88%, a sensitivity of 82%, a specificity of 90%, and an Az of 0.89. Compared to the global image features set, the combined features achieved significantly better accuracy (p=0.0213) and Az (p=0.0197) (Table 5).

Trade-offs between sensitivity and specificity are illustrated as ROC curves in Fig. 4 to show the performances with different cutoff points. Compared to the pathology results, the classification results of the proposed CAD system obtained substantial agreement κ=0.698, p<0.001. Figure 5 shows a successfully classified GBM tumor by the combined image features, but it was misclassified by both the global and local image feature sets.

Discussion

Brain MRI provides an advanced diagnostic imaging technology to interpret tumor characteristics for evaluating tumor type and grade. Based on the gray-scale distribution of tissues in the tumor area, CAD systems can perform malignancy estimations using numerous quantitative image features to provide more-objective and -reliable suggestions. In this study, global image features as statistics of the image moment describing the histogram shape were quantified to express the overall brightness distribution in the tumor area. Local image features were textural patterns describing correlations among neighboring pixels. Benefiting from the complementary power, the combination of both global and local image features achieved an accuracy of 88%, a sensitivity of 82%, a specificity of 90%, and an Az of 0.89. Originally, local image features performed better than global image features without significance. Nevertheless, the combined features achieved significantly better accuracy (p=0.0213) and Az (p=0.0197) than the global image features set. This shows that global image features interpret some characteristics which local features cannot reveal. Previous studies [29-31] which only used GLCM features as local image features for tumor classification might have been insufficient. Also, too many features may induce additional computational complexity. Whether the image features truly interpret the underlying tissue characteristics should reasonably be discussed. For this study, some misclassified cases seemed to have irregular enhancement rings surrounding central necrosis according to the image features used in the CAD system and the conventional diagnosis criteria in clinical use. The dimension of this kind of characteristic is regional rather than pixel-wise. More regional features should be developed via the separation of the enhancement regions and the other regions in tumors for the performance improvement. Besides, although many of the proposed features were formulated using relative intensity distributions such as *Variance* in global features and *Contrast* in local features, more intensity-invariant image features can be developed to reduce the effect of intensity variation in the next study. **For the acquired database, different patients have different settings for the same MR sequence, even they were all scanned in the same MR machine. Since there is wide-variation of the parameters used in both groups, we don't think this is the cause of our statistically valid differences of computed features between LGG and GBM.** Completely quantifying characteristics in tumor area is also important. In this experiment, proposed image features were extracted from the entire tumor area, which should provide more-reliable tissue characteristics and possibly be reproducible in clinical use compared to some studies [23, 32] using one or more squares or circles as the ROI to define tumor tissues.

With respect to the classifier, artificial neural networks (ANN) was also used for comparison. Generally, using one kind of classifier to be the technique of choice in all circumstances is unlikely. ANN is particularly useful if complex nonlinearities existed in a data set. On the other hand, logistic regression provides a clear choice to understand the relationships between the diagnostic result and the predictor variables. Based on logistic regression, tumor malignancy can be divided by using different weights on different characteristics to express the individual importance. The diagnostic result based on ANN with back-propagation achieved an accuracy of 84%, a sensitivity of 79%, and a specificity of 86% which is slightly lower than that of logistic regression (accuracy: 88%, sensitivity: 82%, and specificity: 90%) as shown in Table 6. According to the result and purpose, logistic regression is considered to be appropriate to provide accurate and meaningful malignancy estimation in brain tumor classification.

In this study, only contrast-enhanced T1WIs were used instead of complete MR sequences to estimate the tumor grading. The obvious shortcoming of this design is that peri-tumoral edema might not be well depicted on T1WIs. However, key determinants for differentiating grades II and III from grade IV gliomas are necrosis and/or angiogenesis. Necrosis is an area of a non-enhanced region within the neoplasm with a signal similar to that of cerebrospinal fluid, which can always be clearly demonstrated in contrast-enhanced T1WIs [13]. Also, the degree of contrast enhancement was found to be associated with the activity of the angiogenesis module within the tumor [33, 34]. Since both necrosis and angiogenesis are important criteria applied in histopathology to differentiate GBM from LGG; therefore, we believe that measurements of signal intensities on CET1WI can be key determinants to differentiate GBM from LGG. Nevertheless, further investigation of the role of other important sequences like fluid-attenuated inversion recovery (FLAIR), PWI, DWI, and MRS is warranted.

One limitation of this study is that only two-dimensional tumor areas were delineated for feature extraction and subsequent classification. Using the three-dimensional volume for malignancy evaluation would be more convincing. However, contour delineation would be a time-consuming task. Automatic tumor segmentation is a better way to save time. With respect to the anatomical structures in the brain, normal tissues with various gray-scale intensities surrounding the tumors can barely be separated. A more-sophisticated method would be helpful such as a learning model with prior knowledge about the anatomical structures in the brain. Second, the LGG group contained both grade 2 and 3 gliomas with three different histological cell types. It is possible that tumors belonging to each subset may have different MR imaging signatures. Further researches about distinguishing the grades and types of glioma are warranted. Currently, the proposed CAD system could rapidly provide suggestions about glioma malignancy to radiologists based on preoperative clinical examinations.

Using CAD with the quantitative approach, the diagnostic procedure can be speeded up with reduced diagnostic errors. The consistent estimation can also provide reliable suggestions to radiologists to avoid invasive procedures for which risks outweigh benefits. Whether CAD can improve radiologists' performances is absolutely the most meaningful utility on clinical examinations. The next experiment would be an observers' study.

Conclusions

Twelve proposed MR image features were significant in distinguishing glioblastomas from diffuse lower-grade gliomas ($p<0.05$). Combining them further into a malignancy prediction model was very promising (accuracy: 88%, $\kappa=0.698$, $p<0.001$) in providing diagnostic suggestions for clinical use.

Conflict of interest statement

The authors declare that they have no financial or personal relationships with other people or organizations that could inappropriately have influenced their work.

Acknowledgments

The authors would like to thank the Ministry of Science and Technology in Taiwan (MOST 104-2218-E-038-004 and MOST 103-2314-B-038-067) and Taipei Medical University (TMU 104-AE1-B04) for financially supporting this research.

References

[1] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, *et al.*, "The 2007 WHO classification of tumours of the central nervous system," *Acta neuropathologica*, vol. 114, pp. 97-109, 2007.

[2] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, *et al.*, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathol,* vol. 131, pp. 803-20, Jun 2016.

[3] H. Ohgaki and P. Kleihues, "Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas," *Journal of Neuropathology & Experimental Neurology,* vol. 64, pp. 479-489, 2005.

[4] D. J. Brat, R. Verhaak, K. D. Aldape, W. Yung, S. R. Salama, L. Cooper, *et al.*, "Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas," *The New England journal of medicine,* vol. 372, pp. 2481-2498, 2015.

[5] J. Gallego Perez-Larraya and J. Y. Delattre, "Management of elderly patients with gliomas," *Oncologist,* vol. 19, pp. 1258-67, Dec 2014.

[6] P. C. Burger, F. S. Vogel, S. B. Green, and T. A. Strike, "Glioblastoma multiforme and anaplastic astrocytoma pathologic criteria and prognostic implications," *Cancer*, vol. 56, pp. 1106-1111, 1985.

[7] S. W. Coons, P. C. Johnson, B. W. Scheithauer, A. J. Yates, and D. K. Pearl, "Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas," *Cancer,* vol. 79, pp. 1381-1393, 1997.

[8] P. Kleihues, F. Soylemezoglu, B. Schäuble, B. W. Scheithauer, and P. C. Burger, "Histopathology, classification, and grading of gliomas," *Glia*, vol. 15, pp. 211-221, 1995.

[9] F. H. Gilles, W. D. Brown, A. Leviton, C. J. Tavaré, L. Adelman, L. B. Rorke, *et al.*, "Limitations of the World Health Organization classification of childhood supratentorial astrocytic tumors," *Cancer*, vol. 88, pp. 1477-1483, 2000.

[10] R. A. Prayson, D. P. Agamanolis, M. L. Cohen, M. L. Estes, B. Kleinschmidt-DeMasters, F. Abdul-Karim, *et al.*, "Interobserver reproducibility among neuropathologists and surgical pathologists in fibrillary astrocytoma grading," *Journal of the neurological sciences*, vol. 175, pp. 33-39, 2000.

[11] S. H. Kim, W. Chang, J. P. Kim, Y. Minn, J. Choi, J. Chang, *et al.*, "Peripheral compressing artifacts in brain tissue from stereotactic biopsy with sidecutting biopsy needle: a pitfall for adequate glioma grading," *Clinical neuropathology*, vol. 30, pp. 328-332, 2010.

[12] M. S. Mahaley Jr, C. Mettlin, N. Natarajan, E. R. Laws Jr, and B. B. Peace, "National survey of patterns of care for brain-tumor patients," *Journal of neurosurgery*, vol. 71, pp. 826-836, 1989.

[13] J. A. Guzmán-De-Villoria, J. M. Mateos-Pérez, P. Fernández-García, E. Castro, and M. Desco, "Added value of advanced over conventional magnetic resonance imaging in grading gliomas and other primary brain tumors," *Cancer Imaging*, vol. 14, pp. 1-10, 2014.

[14] M. O. Leach, K. Brindle, J. Evelhoch, J. R. Griffiths, M. R. Horsman, A. Jackson, *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, pp. 1599-1610, 2005.

[15] X. Bai, Y. Zhang, Y. Liu, T. Han, and L. Liu, "Grading of supratentorial astrocytic tumors by using the difference of ADC value," *Neuroradiology,* vol. 53, pp. 533-539, 2011.

[16] A. Jackson, J. P. O'Connor, G. J. Parker, and G. C. Jayson, "Imaging tumor vascular heterogeneity and angiogenesis using dynamic contrast-enhanced magnetic resonance imaging," *Clinical Cancer Research*, vol. 13, pp. 3449-3459, 2007.

[17] R. G. Blasberg, "Imaging update: new windows, new views," *Clinical Cancer Research,* vol. 13, pp. 3444-3448, 2007.

[18] H. Arvinda, C. Kesavadas, P. Sarma, B. Thomas, V. Radhakrishnan, A. Gupta, *et al.*, "RETRACTED ARTICLE: Glioma grading: sensitivity, specificity, positive and negative predictive values of diffusion and perfusion imaging," *Journal of neuro-oncology,* vol. 94, pp. 87-96, 2009.

[19] A. L. Albright, R. J. Packer, R. Zimmerman, L. B. Rorke, J. Boyett, and G. D. Hammond, "Magnetic resonance scans should replace biopsies for the diagnosis of diffuse brain stem gliomas: a report from the Children's Cancer Group," *Neurosurgery,* vol. 33, pp. 1026-1030, 1993.

[20] C.-M. Lo, Y.-C. Lai, Y.-H. Chou, and R.-F. Chang, "Quantitative breast lesion classification based on multichannel distributions in shear-wave imaging," *Computer methods and programs in biomedicine,* vol. 122, pp. 354-361, 2015.

[21] C.-M. Lo, W. K. Moon, C.-S. Huang, J.-H. Chen, M.-C. Yang, and R.-F. Chang, "Intensity-invariant texture analysis for classification of bi-rads category 3 breast masses," *Ultrasound in medicine & biology,* vol. 41, pp. 2039-2048, 2015.

[22] W. K. Moon, C.-M. Lo, N. Cho, J. M. Chang, C.-S. Huang, J.-H. Chen*, et al.*, "Computer-aided diagnosis of breast masses using quantified BI-RADS findings," *Computer methods and programs in biomedicine*, vol. 111, pp. 84-92, 2013.

[23] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem*, 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, pp. 1609-1618, 2009.

[24] R. McLendon, A. Friedman, D. Bigner, E. G. Van Meir, D. J. Brat, G. M. Mastrogianakis*, et al.*, "Comprehensive genomic characterization defines human glioblastoma genes and core pathways," *Nature,* vol. 455, pp. 1061-1068, 2008.

[25] R. A. Groeneveld and G. Meeden, "Measuring skewness and kurtosis," *The Statistician,* pp. 391-399, 1984.

[26] H. J. Baek, H. S. Kim, N. Kim, Y. J. Choi, and Y. J. Kim, "Percent change of perfusion skewness and kurtosis: a potential imaging biomarker for early treatment response in patients with newly diagnosed glioblastomas," *Radiology,* vol. 264, pp. 834-843, 2012.

[27] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on,* pp. 610-621, 1973.

[28] A. P. Field, *Discovering statistics using SPSS, 3rd ed*. Los Angeles: SAGE Publications, 2009.

[29] D. M. Joshi, N. Rana, and V. Misra, "Classification of brain cancer using artificial neural network," in *Electronic Computer Technology (ICECT), 2010 International Conference on*, 2010, pp. 112-116.

[30] D. Singh and K. Kaur, "Classification of abnormalities in brain MRI images using GLCM, PCA and SVM," *International Journal of Engineering and Advanced Technology (IJEAT) ISSN*, pp. 2249-8958, 2012.

[31] S. Jain, "Brain Cancer Classification Using GLCM Based Feature Extraction in Artificial Neural Network," *International Journal of Computer Science & Engineering Technology*, vol. 4, 2013.

[32] S. Herlidou-Meme, J. Constans, B. Carsin, D. Olivie, P. Eliat, L. Nadal-Desbarats, *et al.*, "MRI texture analysis on texture test objects, normal brain and intracranial tumors," *Magnetic resonance imaging*, vol. 21, pp. 989-993, 2003.

[33] M. Diehn, C. Nardini, D. S. Wang, S. McGovern, M. Jayaraman, Y. Liang, *et al.*, "Identification of noninvasive imaging surrogates for brain tumor gene-expression modules," *Proceedings of the National Academy of Sciences,* vol. 105, pp. 5213-5218, 2008.

[34] W. B. Pope, J. H. Chen, J. Dong, M. R. Carlson, A. Perlina, T. F. Cloughesy, *et al.*, "Relationship between gene expression and enhancement in glioblastoma multiforme: exploratory dna microarray analysis 1," *Radiology,* vol. 249, pp. 268-277, 2008.

Figure Captions

Fig. 1. Examples selected from the acquired database showing the challenge of distinguishing between lower-grade gliomas (a, b) and glioblastomas (c, d).

Fig. 2. Examples of delineated tumor areas and corresponding gray-scale distributions of histograms shown in Fig. 1.

Fig. 3 Co-occurrence matrices established with distance=1 and directions=0°, 45°, 90°, and 135° for each pixel and its neighboring pixels.

Fig. 4. Trade-offs between the sensitivity and specificity of tumor classification illustrated by receiver operating characteristic (ROC) curves.

Fig. 5. A malignant glioblastoma (GBM) tumor which was misclassified by both the global (malignancy likelihood=33%) and local image features (malignancy likelihood=22%) but correctly classified by the combined image features (malignancy likelihood=58%). (a) Original MRI image and (b) the delineated tumor area.

Table 1. Common parameters of contrast enhanced T1WI in three institutions*.

	Henry Ford Hospital	Thomas Jefferson University	Case Western
MR Machine	GE Signa HDxt	Siemens Magnetom Vision	Siemens Avanto
Magnetic field strength	1.5T	1.5T	1.5T
TE (ms)	13	3.5	2.81
TR (ms)	500	7.6	2160
Slice thickness (mm)	2.5	1.5	1
Flip angle	90	15	15
FOV(mm)	240	280	250
Matrix	256X192	512X256	256X256
Contrast medium	Gadolinium-based contrast medium	Gadolinium-based contrast medium	Gadolinium-based contrast medium

* The detailed parameters of each image varied from case to case. Here lists the common imaging parameters of the representative cases from three institutions.

Table 2. Significant global image features and corresponding *p* values evaluated using Student's *t*-test (for those with a normal distribution, mean values) or the Mann-Whitney U-test (for those with a non-normal distribution, median values)

Feature	Lower-grade gliomas		Glioblastomas		*p* value
	Mean±SD	Median	Mean±SD	Median	
Mean	85.58±44.32		125.23± 28.63		<0.001*
Variance		256.32		1412.15	<0.001*
Kurtosis		3.85		2.76	<0.001*

* A *p* value of <0.05 indicates a statistically significant difference.

Table 3. Significant local image features and corresponding *p* values evaluated using Student's *t*-test (for those with a normal distribution, mean values) or the Mann-Whitney U-test (for those with a non-normal distribution, median values)

Feature	Lower-grade gliomas		Glioblastomas		*p* value
	Mean±SD	Median	Mean±SD	Median	
Contrast		0.02		0.04	<0.001*
Correlation		0.95		0.92	<0.001*
Dissimilarity	0.021±0.007		0.026±0.008		<0.01*
Homogeneity	1.00±0.01		0.99±0.01		<0.05*
Difference variance		0.02		0.04	<0.001*
Difference entropy	0.04±0.02		0.06±0.02		<0.05*
Information measure of correlation	-0.81±0.05		-0.76±0.02		<0.001*
Inverse difference normalized	0.9989±0.0008		0.9985±0.0008		<0.01*
Inverse difference moment normalized	0.9996±0.0003		0.9994±0.0003		<0.001*

* A *p* value of <0.05 indicates a statistically significant difference.

PPV, positive predictive value; NPV, negative predictive value; Az, area under the curve.

Table 4. Performances of different image feature sets for the classification of lower-grade gliomas (LGGs) and glioblastomas (GBMs)

	Accuracy	Sensitivity	Specificity	PPV	NPV	Az
Global image features	76% (81/107)	68% (23/34)	79% (58/73)	61% (23/38)	84% (58/69)	0.78
Local image features	83% (89/107)	79% (27/34)	85% (62/73)	71% (27/38)	90% (62/69)	0.89
Combined features	88% (94/107)	82% (28/34)	90% (66/73)	80% (28/35)	92% (66/72)	0.89

Table 5. Statistical test results of performance differences between different image feature sets for the classification of lower-grade gliomas (LGGs) and glioblastomas (GBMs)

p value	Accuracy	Sensitivity	Specificity	PPV	NPV	Az
Local vs. Global	0.1760	0.2716	0.3869	0.3335	0.3120	0.0540
Combined vs. Global	0.0213*	0.1614	0.0642	0.0701	0.1654	0.0197*
Combined vs. Local	0.3315	0.7578	0.3140	0.3756	0.7101	0.8436

* A *p* value of <0.05 indicates a statistically significant difference.

Table 6. Performances of different classifiers for the classification of lower-grade gliomas (LGGs) and glioblastomas (GBMs)

	Accuracy	Sensitivity	Specificity	PPV	NPV	Az
Logistic Regression	88% (94/107)	82% (28/34)	90% (66/73)	80%(28/35)	92%(66/72)	0.89
ANN	84% (90/107)	79% (27/34)	86% (63/73)	73% (27/37)	90% (63/70)	0.83
p-value	0.4309	0.7578	0.4389	0.4829	0.7306	0.2036