

Machine Learning–Based Radiomics for Molecular Subtyping of Gliomas

Chia-Feng Lu^{1,2,3}, Fei-Ting Hsu^{2,4,5}, Kevin Li-Chun Hsieh^{2,4,5}, Yu-Chieh Jill Kao^{2,5}, Sho-Jen Cheng⁴, Justin Bo-Kai Hsu⁶, Ping-Huei Tsai^{2,7,8}, Ray-Jade Chen^{9,10}, Chao-Ching Huang^{11,12,13}, Yun Yen¹⁴, and Cheng-Yu Chen^{2,4,5}



Abstract

Purpose: The new classification announced by the World Health Organization in 2016 recognized five molecular subtypes of diffuse gliomas based on isocitrate dehydrogenase (IDH) and 1p/19q genotypes in addition to histologic phenotypes. We aim to determine whether clinical MRI can stratify these molecular subtypes to benefit the diagnosis and monitoring of gliomas.

Experimental Design: The data from 456 subjects with gliomas were obtained from The Cancer Imaging Archive. Overall, 214 subjects, including 106 cases of glioblastomas and 108 cases of lower grade gliomas with preoperative MRI, survival data, histology, IDH, and 1p/19q status were included. We proposed a three-level machine-learning model based on multimodal MR radiomics to classify glioma subtypes. An independent dataset with 70 glioma

subjects was further collected to verify the model performance.

Results: The IDH and 1p/19q status of gliomas can be classified by radiomics and machine-learning approaches, with areas under ROC curves between 0.922 and 0.975 and accuracies between 87.7% and 96.1% estimated on the training dataset. The test on the validation dataset showed a comparable model performance with that on the training dataset, suggesting the efficacy of the trained classifiers. The classification of 5 molecular subtypes solely based on the MR phenotypes achieved an 81.8% accuracy, and a higher accuracy of 89.2% could be achieved if the histology diagnosis is available.

Conclusions: The MR radiomics-based method provides a reliable alternative to determine the histology and molecular subtypes of gliomas. *Clin Cancer Res*; 24(18): 4429–36. ©2018 AACR.

Introduction

Recent studies on glioma based on The Cancer Genome Atlas (TCGA) database have uncovered the strong association of isocitrate dehydrogenase (IDH) mutation, 1p/19q codeletion, and

telomerase reverse transcriptase (TERT) mutation with the patient outcomes (1–3). The new classification announced by the World Health Organization (WHO) in 2016 recognized several new entities of diffuse gliomas based on genotypes in addition to the histologic phenotypes of tumors (4, 5). Among them, the mutations in the IDH gene and 1p/19q codeletion were selected as the critical genetic parameters to further classify the gliomas into five molecular subtypes: the oligodendroglioma and/or anaplastic oligodendroglioma with IDH mutation and 1p/19q codeletion, diffuse and/or anaplastic astrocytoma with IDH mutation, diffuse astrocytoma with wild-type IDH, glioblastoma (GBM) with IDH mutation, and GBM with wild-type IDH, where the former three belong to lower grade gliomas (LGGs, grade 2 and 3) and the latter two are GBMs (grade 4; refs. 4, 5).

Growing evidence has revealed the feasibility of using MRI phenotypes to probe the underlying genotypes, suggesting the potential application in differentiating tumor molecular profiles based on imaging traits (6). Radiomics, a recently developed high-throughput approach, can potentially characterize tumor phenotypes by using thousands of image features based on intensity histogram, geometry, and texture analyses covering the entire tumor volume (7, 8). By applying MR radiomics, substantial relations between imaging traits and genomic profiles were further discovered in GBM. To handle such a large amount of radiomic features in the characterization of tumor phenotypes, a machine-learning algorithm provides a reliable model for tumor classification and outcome prediction. A computer-aided diagnostic tool for the differentiation of GBMs from LGG based on the radiomic features of contrast-enhanced T1-weighted images was developed (9, 10). Recent attempt to predict IDH mutations in

¹Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei, Taiwan. ²Research Center of Translational Imaging, College of Medicine, Taipei Medical University, Taipei, Taiwan. ³Department of Anatomy and Cell Biology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan. ⁴Department of Medical Imaging, Taipei Medical University Hospital, Taipei, Taiwan. ⁵Department of Radiology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan. ⁶Department of Medical Research, Taipei Medical University Hospital, Taipei, Taiwan. ⁷Department of Medical Imaging and Radiological Sciences, Chung Shan Medical University, Taichung, Taiwan. ⁸Department of Medical Imaging, Chung Shan Medical University Hospital, Taichung, Taiwan. ⁹Department of Surgery, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan. ¹⁰Division of General Surgery, Department of Surgery, Taipei Medical University Hospital, Taipei, Taiwan. ¹¹Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan. ¹²Department of Pediatrics, College of Medicine, National Cheng Kung University, Tainan, Taiwan. ¹³Department of Pediatrics, Wan-Fang Hospital, Taipei Medical University, Taipei, Taiwan. ¹⁴Ph.D. Program for Cancer Biology and Drug Discovery, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Cheng-Yu Chen, Department of Radiology, School of Medicine, College of Medicine, Taipei Medical University, 250 Wuxing Street, Taipei 11031, Taiwan. Phone: 8862-2737-2181, ext. 1133; Fax: 8862-23780943; E-mail: sandy0932@gmail.com, sandychen@tmu.edu.tw

doi: 10.1158/1078-0432.CCR-17-3445

©2018 American Association for Cancer Research.

Translational Relevance

Machine learning–based radiomics provides the potential for noninvasive and efficient assessment of 2016 WHO classification of glioma subtypes. The advances in knowledge of this study include: (i) a three-level machine-learning model composed of 4 binary classifiers was proposed to stratify 5 molecular subtypes of gliomas; (ii) machine learning based on multimodal magnetic resonance (MR) radiomics allowed the classifications of the IDH and 1p/19q status of gliomas with accuracies between 87.7% and 96.1%; (iii) the complete classification of 5 molecular subtypes solely based on the MR radiomics achieved an 81.8% accuracy, and a higher accuracy of 89.2% could be achieved if the histology diagnosis is available. In conclusion, multimodal MR radiomics can effectively differentiate glioblastomas from lower grade gliomas and characterize the IDH and 1p/19q status using the machine-learning approach to benefit the diagnosis and treatment of gliomas in clinical practice.

higher grade gliomas based on MR radiomics has shown clinical implications (11, 12). On the other hand, multimodal MR radiomics that combines features from different imaging sequences, such as contrast enhancement, T2 fluid attenuation inversion recovery (FLAIR), and ADC, has also shown promise in the identification of tumor genotypes and in the prediction of patient survivals (11, 13).

In this study, we developed a full scale of a three-level machine-learning algorithm with 4 binary classifiers to characterize the histology, IDH, and 1p/19q status of gliomas based on multimodal MR radiomics. We aim to test the hypothesis that MR radiomics can classify five glioma subtypes according to the new WHO standard.

Materials and Methods

Study cohorts

This study was approved by the local Institutional Review Board. The image data of 456 subjects with gliomas were obtained from The Cancer Imaging Archive (14), including 257 GBM cases from the TCGA-GBM collection (15) and 199 LGG cases from the TCGA-LGG collection (16). The inclusion criteria for this study were as follows: (i) available histology, IDH, and 1p/19q status recorded in TCGA; (ii) preoperative MR image data; (iii) postcontrast T1-weighted images (T1 + C), T2 FLAIR, T2-weighted images (T2W), and diffusion-weighted images (DWI), where T2W and DWI are optional; and (iv) sufficient image quality without significant head motion or artifacts. A total of 214 subjects (106 GBM and 108 LGG subjects) were finally included for the subsequent analyses and training of machine-learning models (Supplementary Fig. S1). The detailed information of included subjects is given in Supplementary Table S1, and the MR data integrity is listed in Supplementary Table S2.

Based on the histology, driver gene mutations of IDH, and 1p/19q codeletion, gliomas can be classified into 5 subtypes (three are LGGs and two are GBMs), as follows: (i) LGG with IDH mutation and 1p/19q codeletion (*LGG-IDHmut-codel*); (ii) LGG with IDH mutation and 1p/19q non-codeletion (*LGG-IDH-*

mut-noncodel); (iii) LGG with wild-type IDH (*LGG-IDHwt*); (iv) GBM with IDH mutation (*GBM-IDHmut*); and (v) GBM with wild-type IDH (*GBM-IDHwt*; ref. 2). These 5 glioma subtypes exhibit distinct tumor characteristics and overall survival outcomes (Table 1).

An independent dataset, including 30 subjects recruited from local hospitals with approval of local Institutional Review Boards and 40 subjects downloaded from the REMBRANDT collection (17), was collected for the validation of model performances. All the included subjects were confirmed to have required multimodal MR image data with sufficient image quality. Please see Supplementary Table S3 for the full subject list of the validation dataset.

Image postprocessing and MR radiomics

Several postprocessing steps on the MR images were applied to reduce the discrepancy of imaging parameters that were employed in different hospitals. The adjustment of image resolution was first performed to resample all voxel size to $0.75 \times 0.75 \times 3.00 \text{ mm}^3$ without gaps between consecutive slices for each MR modalities. The T2 FLAIR, T2W images, and apparent diffusion coefficient (ADC) maps derived from DWI were then registered to the subject's T1 + C images using a six-parameter rigid body transformation and mutual information algorithm. Image intensity normalization was employed to transform MR imaging intensity into standardized ranges for each imaging modality among all subjects. The region of interest (ROI) covering the total tumor volumes (including the contrast enhancing, edema, and necrotic regions) was identified through a semiautomatic image process. Prime regions of contrast enhancing and edema portions were first detected by applying a threshold to extract the hyperintense voxels on the T1 + C images and T2 FLAIR, respectively. The region-growing segmentation algorithm was then implemented on the ROIs to remove the irrelevant voxels from the target regions. The necrotic regions (if existed) were delineated by the surrounding contrast-enhancing and edema portions. Finally, manual adjustment was performed if demanded by an experienced researcher in neuroradiology (C.F. Lu) and confirmed by two experienced neuroradiologists (K.L.-C. Hsieh and C.-Y. Chen). The diagram of image processing is displayed in Supplementary Fig. S2.

A discrete and undecimated wavelet transform was then applied for a multiscale representation of each MR image using the three-dimensional low- and high-spatial frequency filters (18). The 16 first-order and 1,073 texture features [including 22 gray-level cooccurrence matrix features (8), 11 gray-level run-length matrix features (8), 16 local binary pattern features (19), and 1,024 scale invariant feature transform features (20, 21)] were calculated on the raw MR images and 8 wavelet image sets to yield 9,801 features. The 8 shape and size features were calculated based on the three-dimensional geometry of the tumor volumes (8, 13). In total, 39,212 MR radiomic features (9,801 features \times 4 image contrasts + 8 shape and size features) at most were generated for each subject. The detailed calculations of MR radiomics are provided in the Supplementary Table S4. The imaging postprocessing and the calculation of MR radiomics employed in this study were carried out on a home-made software, MR Radiomics Platform (MRP, www.ym.edu.tw/~cflu/MRP_MLinglioma.html), with a graphic user interface built on MATLAB programming environment.

Table 1. The clinical characteristics of the training dataset

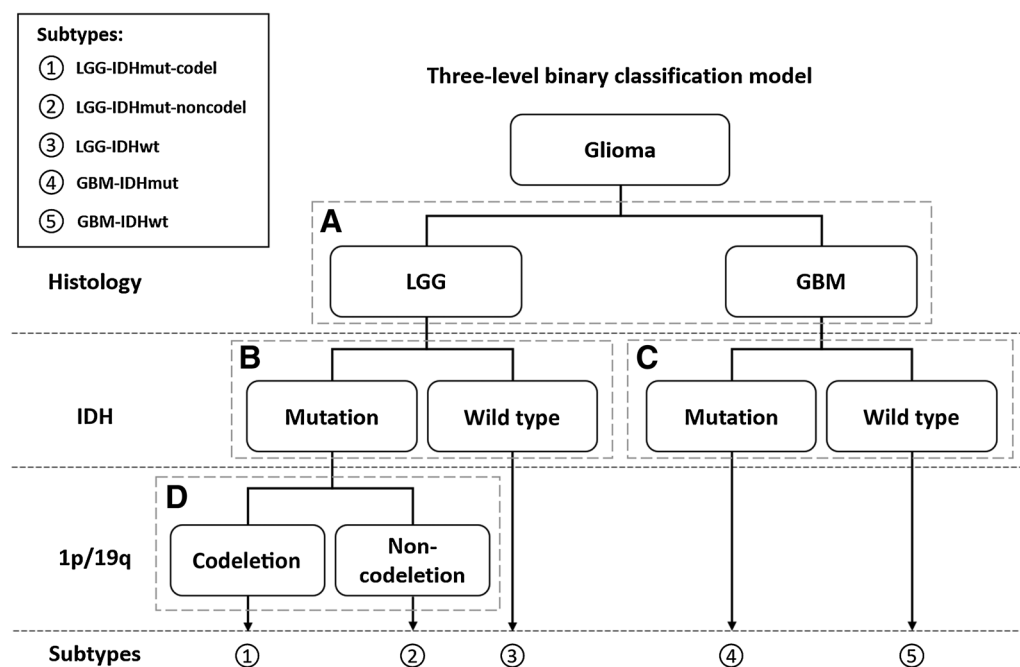
Subtypes	LGG IDH mut - codel	LGG IDH mut - noncodel	LGG IDH wt	GBM IDH mut	GBM IDH wt
Subject number	31 (28.7% of LGG)	56 (51.9% of LGG)	21 (19.4% of LGG)	8 (7.5% of GBM)	98 (92.5% of GBM)
2016 WHO entity	Oligodendroglioma/ anaplastic oligodendroglioma, IDH mut - codel	Diffuse/anaplastic astrocytoma, IDH mut	Diffuse astrocytoma, IDH wt; oligodendroglioma, NOS	GBM, IDH mut	GBM, IDH wt
Histology					
Astrocytoma	0 (0%)	22 (39.3%)	10 (47.6%)	0 (0%)	0 (0%)
Oligoastrocytoma	4 (12.9%)	19 (33.9%)	3 (14.3%)	0 (0%)	0 (0%)
Oligodendroglioma	27 (87.1%)	15 (26.8%)	8 (38.1%)	0 (0%)	0 (0%)
Glioblastoma	0 (0%)	0 (0%)	0 (0%)	8 (100%)	98 (100%)
ATRX status					
Wild type	30 (96.8%)	18 (32.1%)	22 (100.0%)	3 (37.5%)	53 (54.1%)
Mutation	1 (3.2%)	38 (67.9%)	0 (0%)	3 (37.5%)	1 (1.0%)
Unknown	0 (0%)	0 (0%)	0 (0%)	2 (25%)	44 (44.9%)
Age at diagnosis (years)					
Mean (SD)	51.7 (13.2)	40.2 (12.4)	52.5 (12.3)	39.0 (15.9)	60.8 (12.1)
Survival (months)					
Mean (95% CI)	57.8 (40.6–74.9)	90.0 (62.6–115.3)	48.0 (12.1–83.9)	32.7 (19.2–46.2)	15.0 (12.6–17.5)
Karnofsky performance scale					
100	3 (9.7%)	9 (16.0%)	1 (4.8%)	3 (37.5%)	12 (12.3%)
90	6 (19.4%)	17 (30.4%)	8 (38.1%)	0	1 (1.0%)
70–80	3 (9.7%)	7 (12.5%)	4 (19.0%)	4 (50.0%)	50 (51.0%)
<70	2 (6.5%)	2 (3.6%)	0 (0%)	0	17 (17.3%)
Unknown	17 (54.7%)	21 (37.5%)	8 (38.1%)	1 (12.5%)	18 (18.4%)

Abbreviations: Codel, 1p/19q codeletion; NOS, not otherwise specified.

Machine learning-based classification

We proposed a three-level binary classification model to classify gliomas into 5 molecular subtypes based on MR radiomic features (Fig. 1). The classification model was composed of 4 binary classifiers to differentiate patients with LGG or GBM (the first level, Fig. 1A), IDH mutation or wild type in LGGs/GBMs (the

second level, Fig. 1B and C), and codeletion or non-codeletion of 1p/19q in IDH mutation LGGs (the third level, Fig. 1D). The best model for each binary classification was selected from 6 support vector machines (SVM) and 3 ensemble learning approaches with the protection of overfitting using the 5-fold cross-validation. The 6 SVM models included the linear, quadratic, cubic, fine Gaussian,

**Figure 1.**

Three-level machine-learning architecture. The proposed three-level (histology, IDH mutation, and 1p/19q codeletion) binary classification model (a–d) to categorize the diffuse gliomas into 5 potential subtypes, that is, *LGG-IDHmut-codel*, *LGG-IDHmut-noncodel*, *LGG-IDHwt*, *GBM-IDHmut*, and *GBM-IDHwt*.

medium Gaussian, and coarse Gaussian methods (22), and the 3 ensemble learning approaches were the bootstrap-aggregated (bagged) tree algorithm with decision tree (23), the AdaBoost algorithm with decision tree (24), and the RUSBoost algorithm with decision tree (25). The SVM models have high computational efficiency and can achieve satisfactory performance when handling big feature sets, such as the radiomics applied in this study. Alternatively, ensemble learning approaches that combined several machine-learning techniques into one predictive model may have better performance when a single model fails. All the machine-learning algorithms were implemented using the Statistics and Machine Learning Toolbox on MATLAB environment (MathWorks, Inc.).

Statistical analysis

Even though the gigantic amount of radiomic features may provide a comprehensive model in revealing molecular profiles of gliomas, the process of feature selection that removes redundant features can potentially improve the model efficacy in the tumor classification (26). The radiomic features were first ranked by the *t* scores of two-sample *t* tests with a pooled variance estimate. Afterward, 0.05% to 5% top ranking features (i.e., 20–1,960 features) along with patient age and sex were then iteratively selected for the subsequent model training and performance evaluation. A 5-fold cross-validation approach was applied to validate the performance of the machine-learning models. Subjects were randomly divided into two subsets, 80% for model training and 20% for validation, and the process was repeated for 5 rounds to obtain averaged estimates of performance. The model

and feature selection was determined by the criteria of the highest overall accuracy and the AUC of the ROC curve among all tested combinations. The Matthews correlation coefficients (MCC), used as a measure of binary classification quality, were also calculated (27). The MCC is a balanced measure that takes into account full components of confusion matrix that can be used even if the classes are of very different sizes. The MCC represents a correlation coefficient between the observed and predicted binary classifications, where a coefficient of +1 represents perfect prediction and –1 indicates total disagreement between predictions and observations. The interpretations of MCC are given as follows: (i) a value higher than 0.7 represents a very strong agreement; (ii) between 0.5 and 0.7 indicates a moderate agreement; (iii) below 0.5 suggests a weak agreement (28–30).

Results

Clinical characteristics of the study cohort

Table 1 lists the clinical characteristics and the relevant subtypes of the 214 included glioma subjects in the training dataset. For LGG, the most prevalent subtype is *LGG-IDHmut-noncodel* (51.9%), followed by *LGG-IDHmut-codel* (28.7%) and *LGG-IDHwt* (19.4%). Most of the subjects with GBM had the *GBM-IDHwt* subtype (92.5%), which shows the poorest overall survival (average, 15.0 months) among all glioma subtypes. Only a small cohort of GBM subjects (7.5%) had the *GBM-IDHwt* subtype, which has a mean survival of 32.7 months. Most *LGG-IDHmut-codel* gliomas were oligodendroglioma (87.1%) with wild-type ATRX (30/31 cases, 96.8%). The included study cohort exhibited

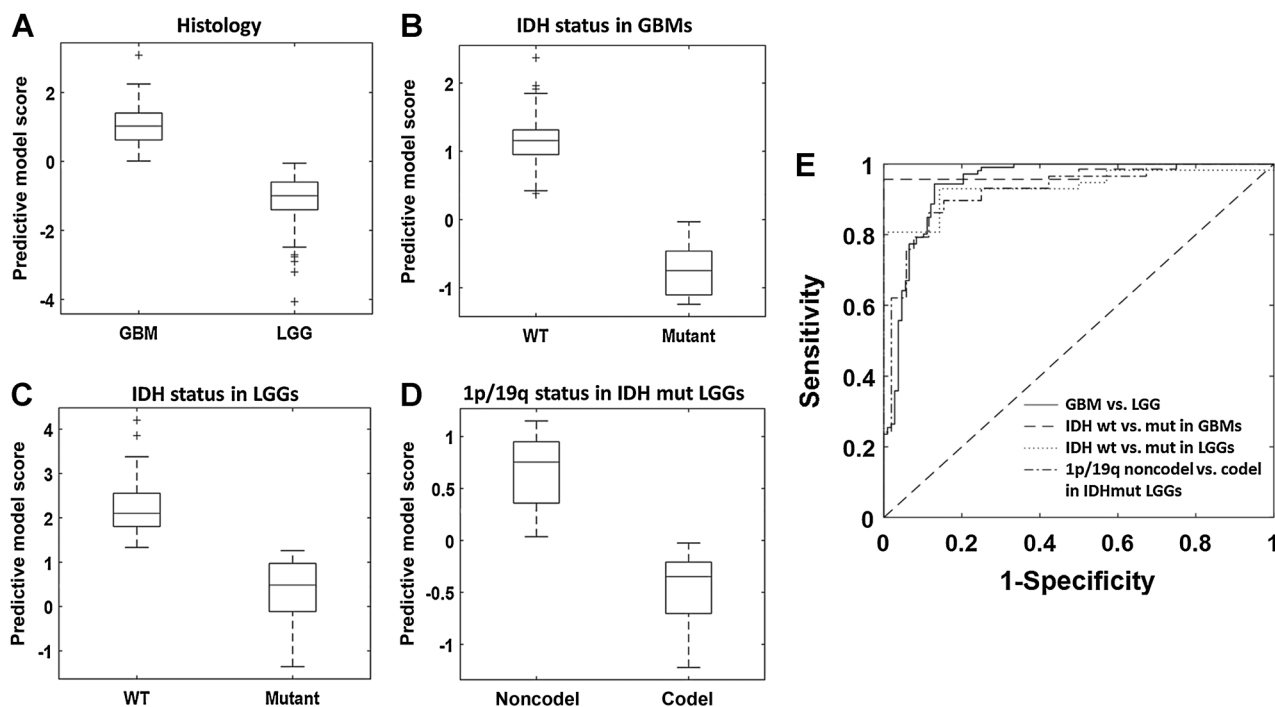


Figure 2.

Predictive model scores and ROC curves. The predictive model scores estimated by the binary classifier for GBM versus LGG (A), IDH wt versus mut in GBMs (B), IDH wt versus mut in LGGs (C), and 1p/19q noncodel versus codel in IDH mut LGGs (D). E, The areas under the ROC curves for the 4 binary classification models are between 0.922 and 0.975, representing the satisfactory results that can be achieved in the classification of histologic and molecular status based on the proposed method (please see Table 2 for details).

Table 2. Model performances for the 4 binary classifiers estimated on the training dataset

Classification (subject numbers)	Model/required image contrasts	AUC	Accuracy	Sensitivity	Specificity	MCC
GBM vs. LGG (214 subjects)	Linear SVM/T1+C, T2 FLAIR	0.944	90.7%	94.3% (true rate for GBM)	87.0% (true rate for LGG)	0.830
IDH wt vs. mut in GBMs (77 subjects)	Cubic SVM/T1+C, T2 FLAIR, T2W	0.975	96.1%	95.7% (true rate for wt)	100.0% (true rate for mut)	0.834
IDH wt vs. mut in LGGs (71 subjects)	Linear SVM/T1+C, T2 FLAIR, T2W, DWI	0.936	91.6%	85.7% (true rate for wt)	93.0% (true rate for mut)	0.769
1p/19q noncodel vs. codel in IDH mut LGGs (81 subjects)	Quadratic SVM/T1+C, T2 FLAIR, T2W	0.922	87.7%	88.5% (true rate for noncodel)	86.2% (true rate for codel)	0.768

consistent profiles with the full TCGA glioma dataset (974 subjects; refs. 2, 3).

Performance of the three-level binary classification model

Profiles of the selected radiomic features in the differentiation of LGG/GBM, IDH, and the 1p/19q status of gliomas are shown in Supplementary Fig. S3. The chosen machine-learning models were the linear SVM for the classification of histology (LGG vs. GBM, Fig. 1A), the linear SVM for the classification of IDH status in LGG (Fig. 1B), the cubic SVM for the classification of IDH status in GBM (Fig. 1C), and the quadratic SVM for the classification of 1p/19q status in IDH mutation LGG (Fig. 1D). The predictive model scores estimated by the selected machine-learning models are shown in Fig. 2A–D. The discrepancies between the predictive scores of the groups demonstrated the ability of the machine-learning models to transfer radiomic features into a differentiable value for effective classification. The machine-learning models can achieve satisfactory classifications with AUCs between 0.922 and 0.975 and MCCs between 0.768 and 0.834 estimated using the training dataset. The ROC curves for the four classifications are displayed in Fig. 2E. The detailed model performances are listed in Table 2.

The trained classifiers were then applied to the validation dataset, and the results are listed in Table 3. In general, the model performances are comparable with the estimates based on the training dataset, suggesting the satisfactory efficacy of classification on the new dataset. It is noted that the specificity in the classification of 1p/19q status in IDH-mutant LGGs is only 66.7%. This low specificity is due to the small testing size of only 5 subjects (2 subjects with non-codel and 3 subjects with codel) in this subgroup. Our model correctly classified the 1p/19q status in 4 of 5 subjects; only 1 of 3 subjects with codel was misclassified as non-codel resulting in a $2/3 \times 100\% = 66.7\%$ specificity.

In addition to the use of an individual classifier as proposed in the previous section, the proposed classification model can be applied in several circumstances, creating potential applications in clinical practice with specific combinations (Combi) of trained classifiers (Table 4). More specifically, the applications can be separated into two scenarios. In a scenario in which only MRI is available for patients with gliomas, Combi #1 listed in Table 4 can be used to differentiate the malignancy of glioma in the patients who receive MRI before surgery (achieving an accuracy of 90.7%). If further information regarding IDH status and full classification of the 5 molecular subtypes is required, Combi #2 and #3 can be employed with the accuracy of 85.1% and 81.8%, respectively. In a scenario in which both tumor histology and MRI are available (more likely in clinical practice), the first-level classifier can be excluded from the combination. Accordingly, a higher accuracy of 93.2% can be achieved in the differentiation of IDH status using Combi #4, and an accuracy of 89.2% can be achieved for the differentiation of IDH and 1p/19q status using Combi #5 (Table 4).

Discussion

We developed a three-level classification model with satisfactory performance to probe the histologic and genomic profiles of gliomas based on MR phenotypes. Based on the analysis results, we suggested that multimodal MR radiomics along with machine-learning models reflected glioma subtypes consistent with the new 2016 WHO classification. By employing a specific

Table 3. Model performances for the 4 binary classifiers estimated on the validation dataset

Classification (subject numbers)	Accuracy	Sensitivity	Specificity	MCC
GBM vs. LGG (70 subjects)	87.7%	82.6% (true rate for GBM)	90.5% (true rate for LGG)	0.830
IDH wt vs. mut in GBMs (18 subjects)	88.9%	88.2% (true rate for wt)	100.0% (true rate for mut)	0.542
IDH wt vs. mut in LGGs (12 subjects)	91.7%	85.7% (true rate for wt)	100.0% (true rate for mut)	0.845
1p/19q noncodel vs. codel in IDH mut LGGs (5 subjects)	80.0%	100.0% (true rate for noncodel)	66.7% (true rate for codel)	0.667

combination of the developed classifiers, several clinical applications for the detection of IDH and 1p/19q statuses in gliomas can be accomplished with or without tumor histology.

The proposed three-level binary classification design was inspired by the general strategy for reducing the problem of multiclass classification to multiple binary classifications and the tree structure of the hierarchical clustering. This design had several advantages compared with the traditional multiclass classification, namely classifying subjects into one of the 5 subtypes using a single classification learner. First, we incorporated the flowchart from the 2016 CNS WHO guideline in the differentiation of the histologic and genetic types of gliomas (4). Based on the designed structure, the binary classifier of 1p/19q status was applied to only the classified IDH-mutation LGG subgroup, reducing the model complexity. Second, feature selection was performed separately for each binary classification. This procedure specified the radiomic features extracted from specific image contrasts that exhibited significant difference between two classified conditions for each classifier and therefore ensured the classification performance. Third, we were able to separately select the best classifier from the 9 tested machine-learning models and perform the parameter optimization accordingly. As shown in our results, the best model varied between classifications based on the discrepant patterns of employed radiomic features.

The identifications of imaging features that can comprehensively describe the target condition are important in machine learning-based classification. Contrast enhancement observed on T1 + C, which suggests blood-brain barrier impairments with leakage of contrast agents, is generally associated with more aggressive lesions or high-grade gliomas (31). Therefore, T1 + C relevant features contributed predominantly to the classification between LGGs and GBMs (Supplementary Fig. S3A). However, some LGGs may also show contrast enhancement and one third of nonenhancing gliomas are malignant (32). The added values extracted from other image contrast, such as T2 FLAIR, to reflect infiltrative edema can further improve differentiation. Regarding the detection of IDH mutations, the radiomics of T1 + C and T2W have been reported to be useful imaging biomarkers

in the differentiation of IDH status in high-grade gliomas (11, 12). In addition to these biomarkers, we found that the features associated with T2 FLAIR were critical in the classification of IDH genotypes in GBMs (Supplementary Fig. S3C). We further established the classifiers for IDH genotype in LGGs and 1p/19q status in IDH mutation LGGs based on MR radiomics to identify the subgroup of LGGs with the IDH mutation and 1p/19q non-codelation (with a high prevalence of ATRX loss) that exhibited a favorable clinical outcome (2, 33). It is also noteworthy that more than 97.3% of the selected features belonged to texture category for the 3 classifiers of IDH and 1p/19q status (Supplementary Fig. S3F–S3H), and no shape and size feature played a role in all the classifiers (Supplementary Fig. S3E–S3H). Texture features quantify local image patterns and the inhomogeneity of signal intensities across the full tumor volume. Our results indicated that the texture measurements describing spatial variations of tumor intensity were the most illustrative for the IDH and 1p/19q genotypes.

Several issues and limitations are discussed as follows. First, the inclusion of advanced MR techniques in addition to the employed modalities should be considered to construct more comprehensive functional and metabolic radiomics in the characterization of gliomas. For instance, the MR perfusion-weighted images for the measurement of tumor vascular leakage and/or regional cerebral blood volume are associated with tumor malignancy and patient outcomes (34, 35). Recently, proton MR spectroscopy provided promising results in the detection of IDH mutation by quantifying the concentration of 2-hydroxyglutarate *in vivo* (36). With this *in vivo* 2-hydroxyglutarate indicator, the accuracy of IDH classification in LGGs may be further improved. Several studies have demonstrated that diffusion kurtosis imaging can differentiate glioma grades more effectively than the conventional ADC and fractional anisotropy (37, 38). Second, the recently highlighted deep-learning approach, such as 3D convolutional neural networks, can be applied for automatic lesion detection and pattern recognition to improve the prediction accuracy (39, 40). The technical concern of deep learning is the insufficient number of samples to train a reliable learner model (typically, at least 1,000

Table 4. Applications of the proposed three-level classification model

Combinations/applications	GBM vs. LGG (1st level)	IDH wt vs. mut in GBMs (2nd level)	IDH wt vs. mut in LGGs (2nd level)	1p/19q noncodel vs. codel in IDH mut LGGs (3rd level)	Accuracy ^a
Available MRI					
#1/Classification of GBM and LGG	✓				90.7%
#2/Prediction of IDH status	✓	✓	✓		85.1%
#3/Full classification of 5 molecular subtypes	✓	✓	✓	✓	81.8%
Available histology and MRI					
#4/Prediction of IDH status in histologically diagnosed GBMs or LGGs		✓	✓		93.2%
#5/Prediction of IDH and 1p/19q status in histologically diagnosed GBMs or LGGs		✓	✓	✓	89.2%

Abbreviation: Codel, codeletion.

^aAccuracies are estimated using the training dataset.

subjects for each molecular subtype are required). Transfer learning that applies a pretrained model in a similar problem domain and fine-tunes the parameters by approximately 100 subjects may be the alternative solution to overcome the limitation of sample size for glioma subtyping (41). Finally, the small sample size of IDH-mutant GBMs can cause an issue of imbalance sampling while training the classification model of the IDH status in GBMs. However, this small subgroup reflects the actual prevalence of IDH mutation, that is, around 7% to 8% in GBMs (3), and hence causing the difficulty for data collection. Similar to the enrolled training dataset that only 8 of 106 GBMs were IDH mutant, only 1 of the 18 GBMs in the validation dataset (recruited from local hospitals) exhibited IDH mutation. However, our results in Tables 2 and 3 show that the IDH-mutant GBM can always be classified in both the training and validation datasets (100% specificity) with a trade-off that the sensitivity (the correctness rates for detecting IDH-wt GBM) may be sacrificed in a certain level (88.2%–95.7% sensitivity). This phenomenon is relevant to the threshold selection when performing binary classification. Refinement of the proposed models with a larger and balanced population is encouraged.

We concluded that multimodal MR radiomics can effectively differentiate GBMs from LGGs and characterize the IDH and 1p/19q status of gliomas. The proposed image-based approach provides an alternative for the noninvasive and efficient identification of the molecular profiles, which can benefit the diagnosis and treatment of gliomas without increasing health care expenses.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

- Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sciotte H, et al. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N Engl J Med* 2015;372:2499–508.
- Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, Cooper LA, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med* 2015;372:2481–98.
- Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 2016;164:550–63.
- Louis DN, Ohgaki H, Wiestler OD, Cavenee WK. World Health Organization Histological Classification of Tumours of the Central Nervous System. Lyon, France: International Agency for Research on Cancer; 2016.
- Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol* 2016;131:803–20.
- Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci U S A* 2008;105:5213–8.
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441–6.
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.
- Hsieh KL-C, Chen C-Y, Lo C-M. Quantitative glioma grading using transformed gray-scale invariant textures of MRI. *Computers Biol Med* 2017; 83:102–8.
- Hsieh KL-C, Lo C-M, Hsiao C-J. Computer-aided grading of gliomas based on local and global MRI features. *Computer Methods Prog Biomed* 2017;139:31–8.
- Zhang B, Chang K, Ramkissoon S, Tanguturi S, Bi WL, Reardon DA, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro-oncol* 2017;19:109–17.
- Hsieh K, Chen C, Lo C. Radiomic model for predicting mutations in the isocitrate dehydrogenase gene in glioblastomas. *Oncotarget* 2017;8: 45888–97.
- Kickingereder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer HP, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology* 2016;280: 880–9.
- Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 2013;26:1045–57.
- Scarpace L, Mikkelsen T, Cha S, Rao S, Tekchandani S, Gutman D, et al. Radiology data from The Cancer Genome Atlas Glioblastoma Multiforme [TCGA-GBM] collection. The Cancer Imaging Archive; 2016.
- Pedano N, Flanders AE, Scarpace L, Mikkelsen T, Eschbacher JM, Hermes B, et al. Radiology Data from The Cancer Genome Atlas Low Grade Glioma [TCGA-LGG] collection. The Cancer Imaging Archive; 2016.
- Scarpace L, Flanders AE, Jain R, Mikkelsen T, Andrews DW. Data from REMBRANDT. The Cancer Imaging Archive; 2016.
- Starck J-L, Fadili J, Murtagh F. The undecimated wavelet decomposition and its reconstruction. *IEEE Trans Image Process* 2007;16:297–309.
- Ojala T, Pietikäinen M, Mäenpää T. Gray scale and rotation invariant texture classification with local binary patterns. Berlin/Heidelberg, Germany: Springer; 2000.
- Rister B, Horowitz MA, Rubin DL. Volumetric Image Registration From Invariant Keypoints. *IEEE Trans Image Process* 2017;26:4900–10.
- Cheung W, Hamarneh G. N-SIFT: N-Dimensional Scale Invariant Feature Transform. *IEEE Trans Image Process* 2009;18:2012–21.

Authors' Contributions

Conception and design: C.-F. Lu, K.L.-C. Hsieh, Y. Yen, C.-Y. Chen
Development of methodology: C.-F. Lu, C.-Y. Chen
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): C.-F. Lu, S.-J. Cheng, P.-H. Tsai, C.-Y. Chen
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): C.-F. Lu, K.L.-C. Hsieh, Y.-C.J. Kao, S.-J. Cheng, R.-J. Chen, C.-Y. Chen
Writing, review, and/or revision of the manuscript: C.-F. Lu, Y.-C.J. Kao, S.-J. Cheng, R.-J. Chen, C.-C. Huang, C.-Y. Chen
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C.-F. Lu, F.-T. Hsu, S.-J. Cheng, J.B.-K. Hsu, P.-H. Tsai
Study supervision: R.-J. Chen, C.-C. Huang, C.-Y. Chen

Acknowledgments

The authors thank Yung-Hsiao Chiang, Wan-Yuo Guo, Min-Hsong Chen, Liang-Wei Chen, Chih-Chun Wu, and Kuo-Chen Wei for the assistance in patient recruitment from local hospitals. This work was supported by the Ministry of Science and Technology, Taiwan (MOST106-2314-B-010-058-MY2, MOST105-2314-B-038-014, and MOST104-2314-B-038-051-MY3), Taipei Medical University (TMU103-AE1-B20), and National Health Research Institutes (MG-106-SP-07 and NHRI-EX107-10732NI). The funding sources had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 23, 2017; revised April 11, 2018; accepted May 17, 2018; published first May 22, 2018.

22. Schölkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. Cambridge, MA: MIT press; 2002.
23. Breiman L. Random Forests. *Machine Learn* 2001;45:5–32.
24. Rätsch G, Onoda T, Müller K-R. Soft margins for AdaBoost. *Machine Learn* 2001;42:287–320.
25. Seiffert C, Khoshgoftaar TM, Van Hulse J. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems* 2010;40: 185–97.
26. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Machine Learn Res* 2003;3:1157–82.
27. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 1975;405:442–51.
28. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Tech* 2011; 2:37–63.
29. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 2012;24:69–71.
30. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. Boston, MA: Houghton Mifflin College Division: 2003.
31. Upadhyay N, Waldman A. Conventional MRI evaluation of gliomas. *Br J Radiol* 2011;84:S107–11.
32. Scott J, Brasher PM, Sevick RJ, Rewcastle NB, Forsyth PA. How often are nonenhancing supratentorial gliomas malignant? A population study. *Neurology* 2002;59:947–9.
33. Wiestler B, Capper D, Holland-Letz T, Korshunov A, von Deimling A, Pfister SM, et al. ATRX loss refines the classification of anaplastic gliomas and identifies a subgroup of IDH mutant astrocytic tumors with better prognosis. *Acta Neuropathol* 2013;126:443.
34. Law M, Young RJ, Babb JS, Peccerelli N, Chheang S, Gruber ML, et al. Gliomas: predicting time to progression or survival with cerebral blood volume measurements at dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. *Radiology* 2008;247:490–8.
35. Law M, Yang S, Babb JS, Knopp EA, Golfinos JG, Zagzag D, et al. Comparison of cerebral blood volume and vascular permeability from dynamic susceptibility contrast-enhanced perfusion MR imaging with glioma grade. *Am J Neuroradiol* 2004;25:746–55.
36. Choi C, Ganji SK, DeBerardinis RJ, Hatanpaa KJ, Rakheja D, Kovacs Z, et al. 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. *Nat Med* 2012;18:624–9.
37. Van Cauter S, Veraart J, Sijbers J, Peeters RR, Himmelreich U, De Keyzer F, et al. Gliomas: diffusion kurtosis MR imaging in grading. *Radiology* 2012;263:492–501.
38. Raab P, Hattingen E, Franz K, Zanella FE, Lanfermann H. Cerebral gliomas: diffusional kurtosis imaging analysis of microstructural differences 1. *Radiology* 2010;254:876–81.
39. Pereira S, Pinto A, Alves V, Silva CA. Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI. In: Crimi A, Menze B, Maier O, Reyes M, Handels H, editors. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Lecture Notes in Computer Science*. Springer: Cham, Switzerland; 2015.
40. Nie D, Zhang H, Adeli E, Liu L, Shen D. 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science*. Springer: Cham, Switzerland; 2016.
41. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016;35: 1285–98.