AJNR

Imaging-Based Algorithm for the Local Grading of Glioma

E.D.H. Gates, J.S. Lin, J.S. Weinberg, S.S. Prabhu, J. Hamilton, J.D. Hazle, G.N. Fuller, V. Baladandayuthapani, D.T. Fuentes and D. Schellingerhout

AJNR Am J Neuroradiol 2020, 41 (3) 400-407
doi: https://doi.org/10.3174/ajnr.A6405
http://www.ajnr.org/content/41/3/400

This information is current as of August 10, 2025.

Imaging-Based Algorithm for the Local Grading of Glioma

E.D.H. Gates, J.S. Lin, J.S. Weinberg, S.S. Prabhu, J. Hamilton, J.D. Hazle, G.N. Fuller, V. Baladandayuthapani, D.T. Fuentes, and D. Schellingerhout

ABSTRACT

BACKGROUND AND PURPOSE: Gliomas are highly heterogeneous tumors, and optimal treatment depends on identifying and locating the highest grade disease present. Imaging techniques for doing so are generally not validated against the histopathologic criterion standard. The purpose of this work was to estimate the local glioma grade using a machine learning model trained on preoperative image data and spatially specific tumor samples. The value of imaging in patients with brain tumor can be enhanced if pathologic data can be estimated from imaging input using predictive models.

MATERIALS AND METHODS: Patients with gliomas were enrolled in a prospective clinical imaging trial between 2013 and 2016. MR imaging was performed with anatomic, diffusion, permeability, and perfusion sequences, followed by image-guided stereotactic biopsy before resection. An imaging description was developed for each biopsy, and multiclass machine learning models were built to predict the World Health Organization grade. Models were assessed on classification accuracy, Cohen κ, precision, and recall.

RESULTS: Twenty-three patients (with 7/9/7 grade II/III/IV gliomas) had analyzable imaging-pathologic pairs, yielding 52 biopsy sites. The random forest method was the best algorithm tested. Tumor grade was predicted at 96% accuracy ($\kappa = 0.93$) using 4 inputs (T2, ADC, CBV, and transfer constant from dynamic contrast-enhanced imaging). By means of the conventional imaging only, the overall accuracy decreased (89% overall, $\kappa = 0.79$) and 43% of high-grade samples were misclassified as lower-grade disease.

CONCLUSIONS: We found that local pathologic grade can be predicted with a high accuracy using clinical imaging data. Advanced imaging data improved this accuracy, adding value to conventional imaging. Confirmatory imaging trials are justified.

ABBREVIATIONS: DCE = dynamic contrast-enhanced; K^{trans} = transfer constant from dynamic contrast-enhanced imaging; NAWM = normal-appearing white matter; ROC = receiver operating characteristic; TIC = T1 post-gadolinium; WHO = World Health Organization; *IDH* = *Isocitrate dehydrogenase*

Gliomas are the most common central nervous system malignancy. They are graded according to the World Health Organization (WHO) grading scale, which represents the overall malignant potential of the tumor.[1] The difference in prognosis for gliomas varies with grade, from 5–12 years (WHO II) to <14 months (WHO IV),[2-4] and nearly all treatment decisions rest critically on the grade of the disease.

In routine clinical care, the overall grade of a tumor is assigned as the highest grade found in ≥1 small biopsy sample. Thus, it is crucial that these samples represent the highest grade present within the tumor so that patients with high-grade disease are not "undergraded." Several studies have highlighted the intratumoral heterogeneity of gliomas,[5,6] which makes sampling the highest grade areas of the tumor a challenging task. The problem of undergrading can occur in as high as 30% of cases.[7]

Received October 22, 2019; accepted after revision December 16.

From the Departments of Imaging Physics (E.D.H.G., J.S.L., J.D.H., D.T.F.), Neurosurgery (J.S.W., S.S.P.), Pathology (G.N.F.), Neuroradiology (D.S.), and Cancer Systems Imaging (D.S.), University of Texas MD Anderson Cancer Center, Houston, Texas; University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences (E.D.H.G.), Houston, Texas; Baylor College of Medicine (J.S.L.), Houston, Texas; Department of Bioengineering (J.S.L.), Rice University, Houston, Texas; Radiology Partners (J.H.), Houston, Texas; and Department of Computational Medicine and Bioinformatics (V.B.), University of Michigan School of Public Health, Ann Arbor, Michigan.

This work was supported by the National Institutes of Health and National Cancer Institute grants R01-CA194391 and R01-CA160736 and the Cancer Center Support Grant P30 CA016672 (Biostatistics Shared Resource) to V.B. Partial support for this study was provided by the Dunn Chair funds (to Dr Bill Murphy), the MD Anderson Cancer Center Internal Research Grant and Clinical Research Support mechanisms for physician-sponsored clinical trials, and the Greenspun neurosurgical research fund.

Please address correspondence to Dawid Schellingerhout, MD, Neuroradiology and Cancer Systems Imaging, University of Texas MD Anderson Cancer Center, 1400 Pressler St, Unit 1482, Houston, TX 77030; e-mail: Dawid.Schellingerhout@mdanderson.org

Indicates open access to non-subscribers at www.ajnr.org
Indicates article with supplemental on-line appendix and tables.
Indicates article with supplemental on-line photo.
http://dx.doi.org/10.3174/ajnr.A6405

There is a gap in our knowledge regarding the use of imaging to identify the highest grade portions of gliomas on the local scale. In this study, we sought to measure the strength of imaging correlations with WHO grade on a per-biopsy level. Our goal was to estimate the WHO pathologic grade based on imaging data input.

Our study is characterized by the use of the latest imaging technologies, state-of-the-art surgical and neuropathologic techniques, and very close spatial imaging-pathologic correlations. We found that imaging data can predict tumor grade to clinically useful accuracies and that advanced imaging (perfusion, permeability, and diffusion) adds value to conventional imaging and improves the accuracy of grade estimates.

MATERIALS AND METHODS

Patients

Data were acquired in a prospective clinical imaging trial (clinicaltrials.gov, NCT03458676), for which this article serves as an interim report. The study was approved by the University of Texas MD Anderson Cancer Center institutional review board, complied with all Health Insurance Portability and Accountability Act regulations, and required informed consent of each participant. We recruited from the pool of previously untreated adult patients with gliomas scheduled for surgical resection in the neurosurgical service at our institution. Patients for whom MR imaging or intravenous contrast was contraindicated were excluded from the study. We previously reported a study estimating the cellular proliferation marker Ki-67 in this same patient population.[8] The current study is based on a separate analysis using a different histologic outcome measure, tumor grade.

Biopsy Sites: Selection and Pathology Analysis

Biopsy target locations were planned before the operation and used either conventional (areas of contrast enhancement) or advanced imaging features (high CBV and/or the transfer constant from dynamic contrast enhanced imaging [K^{trans}] and/or low ADC) to locate sites of suspected high-grade disease. At least 2 biopsy sites were prospectively located before an operation. Biopsy sites were subject to surgeon approval and could be modified as surgically dictated, provided that the altered coordinates were documented. During craniotomy, a neurosurgeon collected ≥1 biopsy using a stereotactic technique before starting resection. As the samples were collected, coordinates of the sampling location were recorded using neuronavigation software ("iPlan", Brainlab, Munich, Germany). This allowed precise, unambiguous identification of the sampling location on the preoperative imaging.[9] Tissue specimens were sectioned and stained using H&E. A board-certified neuropathologist graded each sample independently according to the WHO criteria while blinded to the imaging data.

Although our patients were evenly distributed among final WHO grades II–IV and biopsies were targeted toward areas of increased malignancy, we ultimately collected relatively few high-grade samples. For statistical reasons, we further grouped our samples into 3 categories: normal tissue; lower grade, composed of grade II samples; and higher grade, consisting of grouped grades III and IV samples. *Isocitrate dehydrogenase* (*IDH*) mutation status was not considered when grouping samples into lower- and higher-grade groups.

Imaging

All patients were scanned on a Signa HDxt or Discovery MR750 3T (GE Healthcare, Milwaukee, Wisconsin) clinical scanner using an 8-channel head coil. We collected conventional anatomic MR imaging sequences such as T1-weighted, T1-postgadolinium (T1C), T2-weighted, susceptibility-weighted angiography (SWAN) and T2 FLAIR, as well as advanced diffusion-weighted (DWI/DTI), DSC, and dynamic contrast-enhanced (DCE) sequences. The advanced imaging series were processed into parametric or pharmacokinetic maps using the Advantage Workstation (Version 4.5; GE Healthcare) and NordicICE (NordicNeuroLab, Bergen, Norway). Specific acquisition parameters are given in On-line Tables 1 and 2.

Diffusion-weighted images (4 b-values from 0 to 2000) were processed to maps of ADCs and exponential ADC and diffusion tensor imaging (27 encoding directions) provided maps of fractional anisotropy. DSC and DCE imaging used separate boluses of 0.1 mmol/kg of gadolinium contrast at 5 mL per second. The DCE bolus served as a preload for DSC imaging and was used for T1 postcontrast imaging. DCE time-series were processed into transfer constants and voxel fractions (K^{trans}, k_{ep}, v_p, v_e). We also computed slopes (wash-in, wash-out), TTP, peak enhancement, and curve area voxelwise from the time-series. DSC data were similarly processed into maps of relative CBF and CBV, delay time and MTT, and leakage parameter with a cutoff of 0.01. We did not apply motion correction and spatial or temporal smoothing to DSC or DCE time-series before processing.

Brain-extracted images were coregistered using rigid (6 *df*), followed by affine (12 *df*) mutual information–based registration.[10] The T2-weighted image was used as the reference for each patient. Anatomic images were normalized using ROIs manually placed in CSF, deep GM, or normal-appearing white matter (NAWM). Each image was linearly scaled so that the darkest and brightest ROIs had mean intensities of 0 and 1, respectively. For example, each patient's T2-weighted image was independently scaled to have a mean WM intensity of 0 and CSF intensity of 1. We found that this scaling greatly reduced interpatient variability.[11] DWI, DTI, DSC, and DCE quantitative parameters were used without normalization.

We recorded the average intensity in a 5-mm diameter spheric VOI centered on the biopsy coordinates for each sample. Ultimately, each sample had 23 imaging values associated with it (On-line Table 3), one for each imaging parameter. We also selected coordinates mirrored across the midline contralateral to each biopsy in NAWM to represent normal tissue. An oncologic neuroradiologist reviewed these placements to ensure that they were completely in NAWM. Corresponding imaging values were extracted for these "virtual biopsy" sites to serve as controls for tumor biopsies.

These contralateral virtual biopsies are intended to help the model discern the imaging features of normal brain versus tumor, and omitting these might yield a model good only for distinguishing grades from one another, but failing to distinguish normal brain from tumor (On-line Table 4). The ability to distinguish normal brain from tumor and various local grades from one another stands as equally desirable for clinical applications. Clearly, the best solution would be to acquire real,

histologically normal samples from peritumoral and normal brain regions, but ethical constraints prohibited us.

Modeling of Image Features to Predict Tumor Grade

Modeling and analysis were implemented using R, Version 3.4.2.[12] We used random forest,[13] support vector machine, and neural network classifiers for prediction of multiclass output of tumor grade (normal, lower-grade, higher-grade). The results of all models are given in On-line Table 5, and descriptions of model parameters are given in On-line Table 6. Although deep convolutional networks are powerful models for image-based prediction tasks, we elected not to use one in this case. The strength of this dataset is in the spatial specificity of the tissue samples, which means our training data are only the very small region around each sampling location. This is generally incompatible with convolutional networks, which require either large segmented regions or whole-image classifications for training. Our models were assessed using five-fold cross-validation, with the proportion of samples of each grade maintained between each fold. The classifier performance was measured using the average classification accuracy of the model over the testing set (20% of biopsy data not used for model training) and the Cohen κ.[14] The κ metric measures the accuracy relative to the expected agreement based on random guessing. Similarly to overall accuracy, $\kappa = 1$ for perfect classification. However, unlike accuracy, $\kappa = 0$ means that the classifier is no better than chance, even though some samples may be correctly classified (for an observed proportion of agreement p_{obs} and expected agreement p_{exp}: $\kappa = \frac{p_{obs} - p_{exp}}{1 - p_{exp}}$).

We focused our presentation on the results of our best-performing approach, the random forest. Performance of other models tested is listed in On-line Table 5. We found that an accurate model could be made using only 1 imaging parameter from each family of sequences. In each fold of cross-validation, we selected the best predictor from each family and used that reduced 4-variable set to make predictions on the testing set. We also aggregated the dominant imaging predictors into a single fixed variable set and repeated the cross-validation to estimate the performance of this final fixed model. Finally, we repeated the variable selection and cross-validation using conventional imaging only to investigate the benefit of diffusion, perfusion, and permeability imaging.

The primary benefit in predicting sample grades is localizing areas of high-grade disease. We further analyzed the ability of the classifier to separate the higher-grade samples from the pooled normal and lower-grade samples using receiver operating characteristic (ROC) and precision recall curves.

Table 1: Enhancement status—the overall enhancement characteristics of each tumor separated by clinical WHO grade on a patient-by-patient basis[a]

	Patient Clinical Grade		
Tumor Type	II	III	IV
Enhancing	0	1	6
Nonenhancing	7	8	1

[a] As expected, grade II gliomas are nonenhancing and grade IV gliomas are almost all enhancing.

Table 2: Enhancement status—tabulates which samples were collected from enhancing-versus-nonenhancing regions[a]

		Sample Grade (52 Total)				
Tumor Type	Sampling Region	Normal	II	II/III	III	IV
Enhancing tumor	Enhancing volume	0	4	0	2	3
	Nonenhancing volume	1	4	1	0	2
Nonenhancing tumor	Nonenhancing volume	2	31	2	0	0

[a] In this table, the grade represents the individual grade of the sample, on a sample-by-sample basis, which is not necessarily equal to the patient's clinical grade. For enhancing tumors, about half the samples were collected from the enhancing portion of the tumor.

Table 3: Model accuracy—accuracy for random forest models trained on conventional (anatomic imaging) and conventional-plus-advanced (diffusion, perfusion, permeability) imaging[a]

Metric	Variables	Conventional	Conventional-Plus-Advanced
Accuracy	All Variables	0.904	0.952
	RF Rank	0.875	0.933
	RF (fixed variables)	0.885 (T2, T1C, FLAIR, T1)	0.962 (T2, ADC, CBV, K^{trans})
κ	All Variables	0.822	0.912
	RF Rank	0.768	0.877
	RF (fixed variables)	0.788	0.930

Note:—RF indicates random forest.

[a] Performance using all 23 variables is compared with the performance averaged over 5 folds of cross-validation as well as the results using a fixed-variable subset. The Cohen κ is larger for more accurate models but is scaled so that a no-information classifier would have κ equal to zero.

RESULTS

Patients

Thirty-one patients were initially recruited; surgical complexity prohibited tissue harvest in 5 cases. Among the 26 patients with successful tissue harvest, a total of 64 biopsies were collected. Additional patient exclusion occurred due to missing DCE imaging (1 patient, 3 biopsies) and missing histologic values due to lack of analyzable tissue (2 patients, 4 total biopsies). Further exclusion of biopsies occurred due to poor VOI placement ($n = 3$ biopsies), insufficient quality of tissue for pathologic analysis ($n = 1$), and missing grades ($n = 1$). This left 23 patients, including 7 patients with grade II, 9 with grade III, and 7 with grade IV gliomas (final clinical grade) for use in the final analysis, with 52 real biopsies and 52 paired virtual biopsies. When a biopsy site was excluded, its corresponding imputed virtual biopsy site was also excluded. Full details of sample exclusions are given in the On-line Figure.

Among the 23 patients used in the final analysis, 11 (25 samples) had *IDH* wild-type tumors and the remainder (12 patients, 27 samples) had *IDH*-mutant tumors. *MGMT* promoter methylation was present in 21/23, and 1p/19q codeletion, in 9/23.

Pathology and Imaging Analysis

Samples were graded II–IV using the WHO criteria on a per-biopsy basis (ie, only features of that particular

Table 4: Model predictions by sample grade—confusion matrix for the random forest model trained on 4 fixed variables[a]

Predicted	Reference			
	Normal	Lower	Higher	Total
Conventional plus advanced				
Normal	53	1	0	54
Lower	2	41	1	43
Higher	0	0	6	7
Total	$n = 55$	$n = 42$	$n = 7$	104
Conventional only				
Normal	52	5	0	57
Lower	3	36	3	42
Higher	0	1	4	5
Total	$n = 55$	$n = 42$	$n = 7$	104

[a] A perfect classifier would have all entries along the main diagonal. The random forest using conventional-plus-advanced imaging had 96.2% (100/104) overall accuracy with $\kappa = 0.930$, whereas the random forest using only conventional imaging had 88.5% (92/104) accuracy with $\kappa = 0.788$. The conventional-only classifier also misclassified 43% (3/7) of high-grade samples.

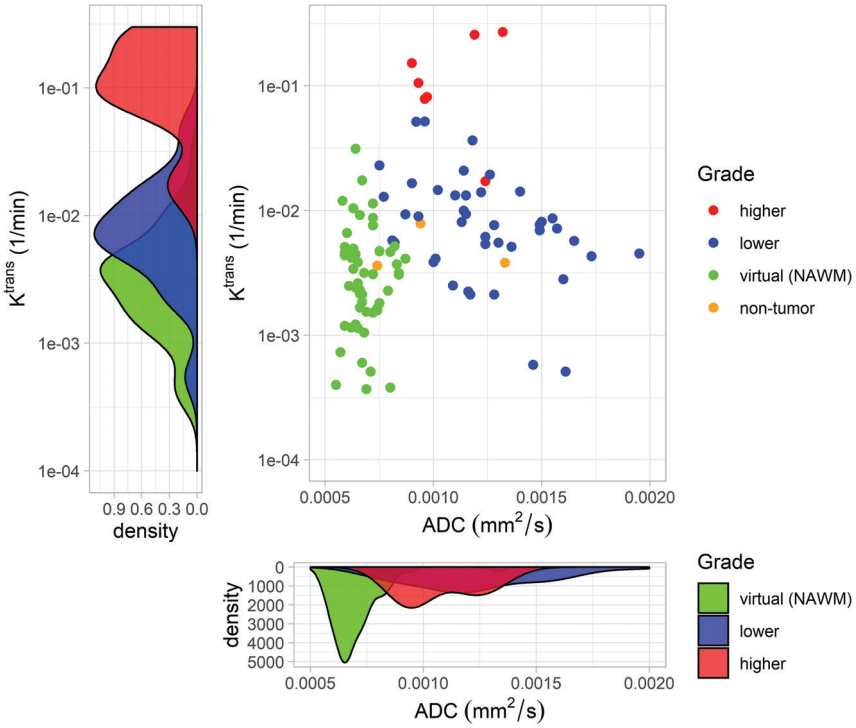

FIG 1. Synergistic properties of imaging sequences with "orthogonal" information. Two representatives of the variable inputs used to generate the grade map seen in Fig 3 are shown. The scatterplot shows the ADC (square millimeters/second) from DWI and K^{trans} (minute^{-1}) from DCE for normal, lower-grade, and higher-grade samples. The normal samples are further identified as being either virtual biopsies in NAWM or nontumor real tissue samples as designated by a pathologist. Each imaging parameter roughly separates 2 of the classes as seen by histograms. Combined, they form 3 distinct clusters that are identified by the random forest. K^{trans} alone distinguishes lower- and higher-grade samples but does not differentiate lower-grade tumor samples from those of healthy controls, as shown by the degree of overlap in the left density plot. However, the ADC from diffusion-weighted imaging separates these healthy controls from tumor samples but differs little among different grades (lower density plot). Combining these successfully identifies all 3 sample grades (normal, lower-grade, and higher-grade) simultaneously. This separation is further increased by adding CBV and T2-weighted signal intensity.

biopsy as judged on H&E staining were used to assign a grade to that biopsy). Note that this research methodology is at variance with the clinical practice of assigning a clinical grade corresponding to the maximum local grade found to a patient. The per-biopsy local grading used in this article is different from the conventional per-patient clinical grading. Some biopsy samples were graded II/III as an intermediate between grades II and III, due to our pathologist's clinical assessment that their malignant potential exceeded that of regular grade II samples.[15] Of the 52 real biopsies, 3 were normal brain, 39 grade II, 3 grade II/III, 2 grade III, and 5 grade IV. The lack of grade I disease is expected in a nonpediatric population. For the final analysis, we had 55 biopsies with normal results (3 real and 52 virtual), 42 lower-grade samples, and 7 higher-grade samples.

Among the 23 patients, nearly all the grade IV tumors were contrast-enhancing and nearly all of the grade II and III tumors were nonenhancing (see Table 1 for specifics). We collected both lower- and higher-grade samples in regions of contrast enhancement. Five of 7 higher-grade samples and 4 of 42 lower-grade samples were collected from the enhancing volume of an enhancing tumor. Table 2 lists the number of samples of each grade taken from regions of enhancement. This table shows that while enhancement is a good surrogate for local grade, it does not perfectly discriminate.

For 5 biopsies among 2 patients, missing susceptibility-weighted angiography imaging values were imputed as the median values among the corresponding real ($n = 46$) or virtual biopsies ($n = 47$). All imaging sequences were available for the remaining patients.

Modeling Results

We trained a random forest model to predict the grade of individual samples using imaging values. We chose to use the random forest because it had the best average performance and provides some resistance to overfitting.[13] Because many of the 23 parameters, especially those from the same family of imaging sequences, contain mostly redundant information, we used only the top predictor from each family for predictive modeling. We repeated this procedure for each fold of cross-validation (On-line Table 7) and

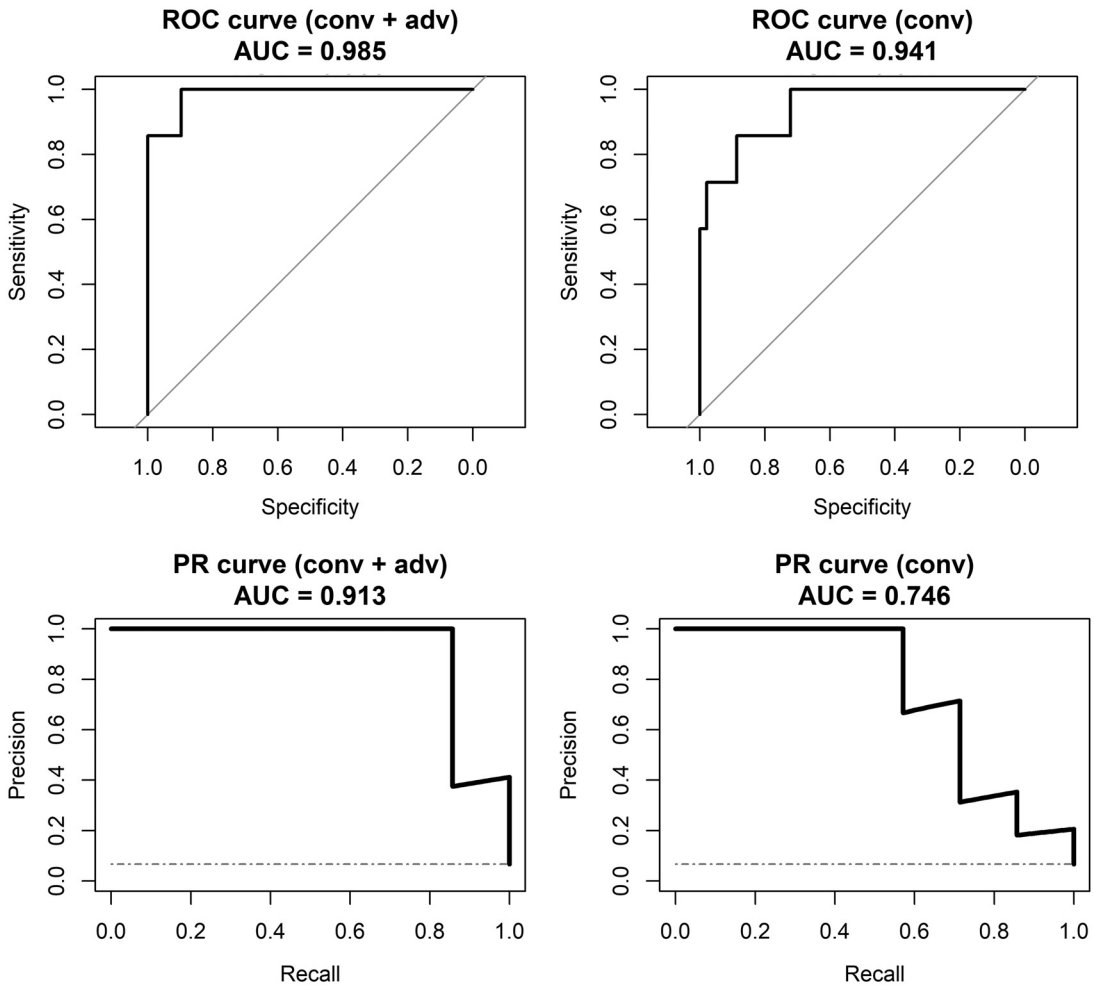

FIG 2. ROC and precision-recall (PR) curves for distinguishing high-grade samples versus normal tissue and low-grade disease samples. Plots in the left columns are based on the predictions from the random forest model using conventional and advanced imaging, whereas plots on the right are based on models using conventional imaging only. Although the area under the ROC curves is similar, the area under the PR curve is much smaller for conventional imaging only. The *diagonal lines* on the ROC curves and *horizontal dashed lines* on PR curves show the performance of a no-information or random classifier. Conv indicates conventional; adv, advanced; AUC, area under the curve.

selected the final 4-variable predictor set on the basis of consensus. This 4-variable set was T2, ADC, CBV, and K^{trans}.

With these 4 inputs, the random forest correctly classified the grade of individual samples with 96% accuracy (Cohen κ = 0.930) as shown in Table 3. Furthermore, none of the high-grade biopsies were classified as normal brain or vice versa (Table 4). The positive predictive value for high-grade disease is 1.0 (all predictions of high-grade are correct), and the negative predictive value is 0.990. Fig 1 intuitively shows how the combination of imaging from different families like diffusion and DCE are able to separate samples of different grades. The random forest leverages this type of separation to classify unknown samples. The normal, lower-, and higher-grade samples separate even more with the addition of conventional (T2) and perfusion (CBV) imaging.

By means of conventional imaging only (T2, T1C, FLAIR, T1), the model still managed to achieve 88.5% accuracy (κ = 0.788), but the error rate for high-grade samples was >40% (Table 4). In other words, the conventional imaging was generally unable to differentiate high-grade disease from low-grade disease using only anatomic MR imaging sequences. This issue highlights the importance of advanced and functional imaging in determining the grade of individual samples. For reference, a classifier with no information would achieve 53% accuracy by classifying every sample as normal, the most frequent class.

While our patient population is fairly well-distributed among clinical WHO grades II, III, and IV, the higher-grade samples represent a minority in the sample population. To analyze the ability of the model to specifically identify higher-grade disease in light of this class imbalance, we analyzed the ROC and precision-recall curves for the final models. Specifically, we used the estimated probability of each model that a given sample was higher-grade (WHO grade III or IV) disease to create a binary output. The areas under the ROC curves were very high for both conventional and conventional-plus-advanced models at 0.94 and 0.99, respectively (Fig 2). However, the area under the precision-recall curve was considerably lower for the conventional-only model at 0.75 versus 0.91 for the conventional-plus-advanced model. The decreased precision or positive predictive value reiterates the difficulty for the conventional-only model in identifying higher-grade disease.

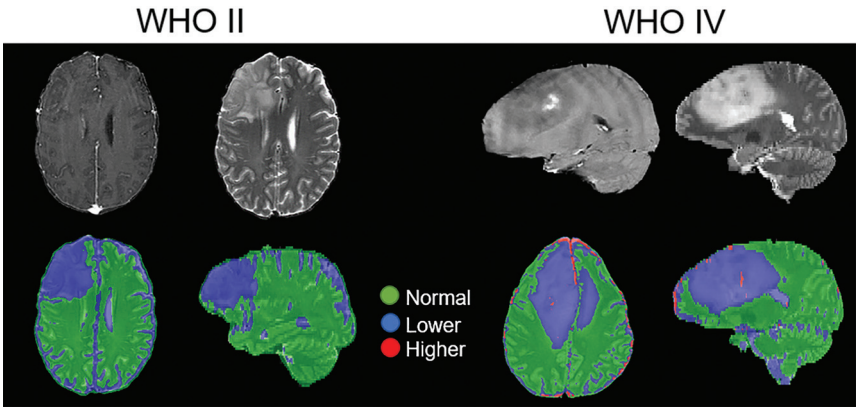

FIG 3. A clinically relevant map of the predicted tumor grade using the highest probability grade and smoothed by a median filter (2-voxel radius) superimposed on a T2-weighted image. Green, blue, and red correspond to predicted normal tissue and lower-grade and higher-grade disease, respectively. Shown on the left is a WHO grade II oligodendroglioma with a T2-weighted image for reference, and on the right is a WHO grade IV glioblastoma with a TI postcontrast image for reference (*upper part*). The classifier identifies the oligodendroglioma as containing only lower-grade disease, while the glioblastoma has a central region of higher grade near the enhancing-focus disease surrounded by lower-grade disease. Areas of normal brain falsely identified as tumor (ie, sulcus) are unlikely to confuse a clinician.

Virtual biopsy regions in contralateral normal-appearing white matter constituted a majority of the normal tissue class and half of the overall training data. To ensure that this did not unnecessarily bias the model against normal samples in the peritumoral region, we re-trained the model using only the real biopsy samples and found that the model retained a high overall accuracy (>90%) and 86% sensitivity to high-grade disease. More details are provided in On-line Table 4.

DISCUSSION

We found the following: 1) imaging could be used to predict the WHO grade in glioma to clinically useful accuracies using a random forest model, 2) advanced imaging (diffusion, perfusion, and permeability) outperformed conventional anatomic imaging alone, 3) the best anatomic imaging sequences for estimating grade were T2-weighted, FLAIR, and T1-weighted both pre- and postcontrast, 4) the best overall imaging sequences for estimating grade were T2-weighted, ADC, CBV from DSC, and K^{trans}, and 5) the algorithm developed could be used to derive graphic grade maps to visually present the information for imaging guidance.

Image-guided brain tissue sampling is costly and technically demanding. Furthermore, imaging technology development is rapid, meaning the literature on tissue correlations between modern imaging sequences and neuropathologic techniques is continually evolving.[16-19] Previous studies include predicting likely areas of infiltrative tumor and recurrence in glioblastoma using support vector machines and MR imaging.[20] However, the authors acknowledged that a limitation of their work was the lack of histopathologic validation. Indeed, many studies correlating image-based metrics with glioma grades are limited by this heterogeneity and lack of local histologic validation.[19,21,22] In other work, Barajas et al[23] demonstrated that pathologic features of aggressiveness correlated with increased relative CBV and decreased relative ADC. Other recent work using hand-crafted image features or neural networks to predict *IDH*-mutation status or glioma grade shows good overall accuracy but only evaluates the tumor as a whole and provides no regional information.[24,25]

We constructed a random forest model using imaging inputs to predict the individual WHO grade of brain biopsies with reasonable success. The ranking of imaging input also demonstrates which imaging sequences were most valuable in predicting clinical information. Our work agrees with the literature showing that perfusion, K^{trans},[26] and ADC can all distinguish low- and high-grade tumors.[27,28] CBV has also been shown to better differentiate tumor grades compared with conventional MR imaging.[19,29,30]

Our best cross-validated model for predicting tumor grade was extrapolated voxelwise across the whole-brain volume. This extrapolation provided a map of normal, lower-grade, and higher-grade disease for each patient using only a small number of imaging inputs (Fig 3). For the high-grade glioblastoma case shown, we can see a central region of high-grade disease surrounded primarily by low-grade disease. Indeed, one of the biopsies from this patient near the tumor periphery was graded as WHO grade II, while another sample near the center of the tumor was graded as WHO grade IV.

Maps generated with these models contain some obvious errors (higher-grade assignment in the sulci and sinuses). A skilled clinician using such maps should not find such errors confusing and would be able to recognize these as artifacts of processing. Our future work will remove these spurious signals. While voxelwise validation of maps is impractical, the predictive accuracy at known biopsy sites is sufficient justification for use as a clinical tool to guide procedures like biopsies. While we see qualitative agreement between predicted regions of high-grade disease and contrast enhancement, which is the clinical standard for targeting biopsies, in our data, the contrast-enhanced T1 was not sufficient to discriminate between all high- and low-grade samples. This result suggests that the predictive models may be able to identify high-grade disease outside the enhancing volume.

As of 2016, the WHO classification of gliomas heavily incorporates genetic and molecular factors.[1] Factors like *IDH1* mutation and *MGMT* promoter methylation are known to be highly prognostic.[31] However, many of these prognostic factors, particularly *IDH1* mutation, are homogeneously expressed throughout the tumor. Thus, there is no risk of undersampling *IDH1* mutation status with clinical biopsy. Our work focuses on the heterogeneous tissue characteristics that are at risk for undersampling, hence, our categorization of lower- and higher-grade samples based on histologic grading.

Limitations

Although our patients were evenly distributed among final WHO grades II–IV and biopsies were targeted toward areas of increased malignancy, we collected relatively few high-grade samples. In our final analysis, we combined WHO grade III and IV samples into a higher-grade category as an attempt to balance our dataset. This feature removes the ability to predict grade III versus IV but still allows us to separate high-grade from low-grade disease. The lack of high-grade samples also highlights the potential for undergrading high-grade tumors using a single biopsy sample. Future trials and larger datasets would allow finer distinctions like these.

The use of virtual biopsies in contralateral NAWM is a limitation, and while we would clearly prefer to have normal histologic samples and imaging parameters sufficient to balance out our tumor histology data, we faced ethical constraints. We chose contralateral regions because in the absence of true histologic validation, they had the highest likelihood of truly being normal. There very likely is tumor-infiltrated brain that appears normal on imaging and the opposite case of imaging of abnormal brain that actually is nontumoral; but reliably characterizing these cases will require a much greater number of well-chosen biopsies, something we would like to address in future work, using our current data as a springboard.

Future Work

In addition to improved discriminatory ability among sample grades, we believe that "grade maps" like Fig 3 could provide a useful tool for surgical guidance. Such maps could help surgeons identify areas of highest grade[20] or identify as eloquent that only containing low-grade disease and could be less aggressively treated to avoid neurologic deficits. These models could also be used to help plan biopsies[7] or radiation therapy, specifically by developing probability maps for tumor presence or severity,[32,33] especially in the peritumoral brain zone.[34] Prospective evaluation of the derived models in further imaging and surgical trials is justified, along with further imaging-directed biopsy trials to refine our results even more in the peritumoral area.

CONCLUSIONS

Individual biopsy grades can be predicted to useful accuracies using noninvasive MR imaging. Advanced imaging (diffusion, perfusion, and permeability) improves predictive results over conventional imaging alone.

Disclosures: Evan D. H. Gates—*RELATED*: a training fellowship from the Gulf Coast Consortia, the National Library of Medicine Training Program in Biomedical Informatics and Data Science (T15LM007093). A scholarship from the American Legion Auxiliary. Jonathan S. Lin—*RELATED*: the Baylor College of Medicine Medical Scientist Training Program and the Cullen Trust for Higher Education Physician/Scientist Fellowship Programs for the duration of this research.

REFERENCES

1. Louis DN, Perry A, Reifenberger G, et al. **The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary.** *Acta Neuropathol* 2016;131:803–20 CrossRef Medline
2. Stupp R, Mason WP, Van Den Bent MJ, et al; National Cancer Institute of Canada Clinical Trials Group. **Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.** *N Engl J Med* 2005;352:987–96 CrossRef Medline
3. Okamoto Y, Di Patre PL, Burkhard C, et al. **Population-based study on incidence, survival rates, and genetic alterations of low-grade diffuse astrocytomas and oligodendrogliomas.** *Acta Neuropathol* 2004;108:49–56 CrossRef Medline
4. Ohgaki H, Kleihues P. **Epidemiology and etiology of gliomas.** *Acta Neuropathol* 2005;109:93–108 CrossRef Medline
5. Parker NR, Khong P, Parkinson JF, et al. **Molecular heterogeneity in glioblastoma: potential clinical implications.** *Front Oncol* 2015;5:55 CrossRef Medline
6. Aum DJ, Kim DH, Beaumont TL, et al. **Molecular and cellular heterogeneity: the hallmark of glioblastoma.** *Neurosurg Focus* 2014;37:E11 CrossRef Medline
7. Jackson RJ, Fuller GN, Abi-Said D, et al. **Limitations of stereotactic biopsy in the initial management of gliomas.** *Neuro Oncol* 2001;3:193–200 CrossRef Medline
8. Gates EDH, Lin JS, Weinberg JS, et al. **Guiding the first biopsy in glioma patients using estimated Ki67 maps derived from magnetic resonance imaging: conventional versus advanced imaging.** *Neuro Oncol* 2019;21:527–36 CrossRef
9. Ringel F, Ingerl D, Ott S, et al. **VarioGuide: a new frameless image-guided stereotactic system: accuracy study and clinical assessment.** *Oper Neurosurg* 2009;64:(5 Suppl 2):365–71 CrossRef Medline
10. Avants BB, Tustison NJ, Song G, et al. **A reproducible evaluation of ANTs similarity metric performance in brain image registration.** *Neuroimage* 2011;54:2033–44 CrossRef Medline
11. Leung KK, Clarkson MJ, Bartlett JW, et al.; Alzheimer's Disease Neuroimaging Initiative. **Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection.** *Neuroimage* 2010;50:516–23 CrossRef Medline
12. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing; 2019 https://www.R-project.org/
13. Breiman L. **Random forests.** *Mach Learn* 2001;45:5–32 CrossRef
14. Cohen J. **A coefficient of agreement for nominal scales.** *Educational and Psychological Measurement* 1960;20:37–46 CrossRef
15. Pedeutour-Braccini Z, Burel-Vandenbos F, Gozé C, et al. **Microfoci of malignant progression in diffuse low-grade gliomas: towards the creation of an intermediate grade in glioma classification?** *Virchows Arch* 2015;466:433–44 CrossRef Medline
16. Zhao M, Guo LL, Huang N, et al. **Quantitative analysis of permeability for glioma grading using dynamic contrast enhanced magnetic resonance imaging.** *Oncol Lett* 2017;14:5418–26 CrossRef Medline
17. Prah MA, Al-Gizawiy MM, Mueller WM, et al. **Spatial discrimination of glioblastoma and treatment effect with histologically-validated perfusion and diffusion magnetic resonance imaging metrics.** *J Neurooncol* 2018;136:13–21 CrossRef Medline
18. Hu LS, Ning S, Eschbacher JM, et al. **Multi-parametric MRI and texture analysis to visualize spatial histologic heterogeneity and tumor extent in glioblastoma.** *PLoS One* 2015;10:e0141506 CrossRef Medline
19. Law M, Yang S, Wang H, et al. **Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging.** *AJNR Am J Neuroradiol* 2003;24:1989–98 Medline
20. Akbari H, Macyszyn L, Da X, et al. **Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma.** *Neurosurg* 2016;78:572–80 CrossRef Medline
21. Miloushev VZ, Chow DS, Filippi CG. **Meta-analysis of diffusion metrics for the prediction of tumor grade in gliomas.** *AJNR Am J Neuroradiol* 2015;36:302–08 CrossRef Medline
22. Jakab A, Molnár P, Emri M, et al. **Glioma grade assessment by using histogram analysis of diffusion tensor imaging-derived maps.** *Neuroradiology* 2011;53:483–91 CrossRef Medline

23. Barajas RF Jr, Hodgson JG, Chang JS, et al. **Glioblastoma multiforme regional genetic and cellular expression patterns: influence on anatomic and physiologic MR imaging.** *Radiology* 2010;254:564–76 CrossRef Medline
24. Zhang X, Yan LF, Hu YC, et al. **Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features.** *Oncotarget* 2017;8:47816–30 CrossRef Medline
25. Chang K, Bai HX, Zhou H, et al. **Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging.** *Clin Cancer Res* 2018;24:1073–81 CrossRef Medline
26. Law M, Yang S, Babb JS, et al. **Comparison of cerebral blood volume and vascular permeability from dynamic susceptibility contrast-enhanced perfusion MR imaging with glioma grade.** *AJNR Am J Neuroradiol* 2004;25:746–55 Medline
27. Sugahara T, Korogi Y, Kochi M, et al. **Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas.** *J Magn Reson Imaging* 1999;9:53–60 CrossRef Medline
28. Lee S, Choi SH, Ryoo I, et al. **Evaluation of the microenvironmental heterogeneity in high-grade gliomas with IDH1/2 gene mutation using histogram analysis of diffusion-weighted imaging and dynamic-susceptibility contrast perfusion imaging.** *J Neurooncol* 2015;121:141–50 CrossRef Medline
29. Yan H, Parsons DW, Jin G, et al. **IDH1 and IDH2 mutations in gliomas.** *N Engl J Med* 2009;360:765–73 CrossRef Medline
30. Noguchi T, Yoshiura T, Hiwatashi A, et al. **Perfusion imaging of brain tumors using arterial spin-labeling: correlation with histopathologic vascular density.** *AJNR Am J Neuroradiol* 2008;29:688–93 CrossRef Medline
31. Hartmann C, Hentschel B, Wick W, et al. **Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas.** *Acta Neuropathol* 2010;120:707–18 CrossRef Medline
32. Shah V, Turkbey B, Mani H, et al. **Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging.** *Med Phys* 2012;39:4093–103 CrossRef Medline
33. Durst CR, Raghavan P, Shaffrey ME, et al. **Multimodal MR imaging model to predict tumor infiltration in patients with gliomas.** *Neuroradiology* 2014;56:107–15 CrossRef Medline
34. Lemee JM, Clavreul A, Menei P. **Intratumoral heterogeneity in glioblastoma: don't forget the peritumoral brain zone.** *Neuro Oncol* 2015;17:1322–32 CrossRef Medline