Generic Contrast Agents



Our portfolio is growing to serve you better. Now you have a *choice*.



This information is current as of May 30, 2025.

Stable and Discriminatory Radiomic Features from the Tumor and Its Habitat Associated with Progression-Free Survival in Glioblastoma: A Multi-Institutional Study

R. Verma, V.B. Hill, V. Statsevych, K. Bera, R. Correa, P. Leo, M. Ahluwalia, A. Madabhushi and P. Tiwari

AJNR Am J Neuroradiol 2022, 43 (8) 1115-1123 doi: https://doi.org/10.3174/ajnr.A7591 http://www.ajnr.org/content/43/8/1115

0<u>.</u> =

Stable and Discriminatory Radiomic Features from the Tumor and Its Habitat Associated with Progression-Free Survival in Glioblastoma: A Multi-Institutional Study

🕫 R. Verma, 🕫 V.B. Hill, 🕫 V. Statsevych, 🏮 K. Bera, 🕫 R. Correa, 🕫 P. Leo, 🕫 M. Ahluwalia, 🕫 A. Madabhushi, and 🏮 P. Tiwari

ABSTRACT

BACKGROUND AND PURPOSE: Glioblastoma is an aggressive brain tumor, with no validated prognostic biomarkers for survival before surgical resection. Although recent approaches have demonstrated the prognostic ability of tumor habitat (constituting necrotic core, enhancing lesion, T2/FLAIR hyperintensity subcompartments) derived radiomic features for glioblastoma survival on treatment-naive MR imaging scans, radiomic features are known to be sensitive to MR imaging acquisitions across sites and scanners. In this study, we sought to identify the radiomic features that are both stable across sites and discriminatory of poor and improved progression-free survival in glioblastoma tumors.

MATERIALS AND METHODS: We used 150 treatment-naive glioblastoma MR imaging scans (Gadolinium-TIw, T2w, FLAIR) obtained from 5 sites. For every tumor subcompartment (enhancing tumor, peritumoral FLAIR-hyperintensities, necrosis), a total of 316 three-dimensional radiomic features were extracted. The training cohort constituted studies from 4 sites (n = 93) to select the most stable and discriminatory radiomic features for every tumor subcompartment. These features were used on a hold-out cohort (n = 57) to evaluate their ability to discriminate patients with poor survival from those with improved survival.

RESULTS: Incorporating the most stable and discriminatory features within a linear discriminant analysis classifier yielded areas under the curve of 0.71, 0.73, and 0.76 on the test set for distinguishing poor and improved survival compared with discriminatory features alone (areas under the curve of 0.65, 0.54, 0.62) from the necrotic core, enhancing tumor, and peritumoral T2/FLAIR hyperintensity, respectively.

CONCLUSIONS: Incorporating stable and discriminatory radiomic features extracted from tumors and associated habitats across multisite MR imaging sequences may yield robust prognostic classifiers of patient survival in glioblastoma tumors.

ABBREVIATIONS: AUC = area under the curve; CoLlAGe = Co-occurrence of Local Anisotropic Gradient Orientations; GBM = Glioblastoma; Gd = Gadolinium; LOSO = leave-one-site-out; PI = preparation-induced instability; PFS = progression-free survival; \hat{T} = training set; TCIA = The Cancer Imaging Archive; VS = validation set; TS = test set

Blioblastoma (GBM) is an aggressive brain tumor in which survival without treatment is typically 3 months. A conven-

Received April 29, 2021; accepted after revision June 13, 2022.

From the Department of Biomedical Engineering (R.V., K.B., R.C., P.L.), Case Western Reserve University, Cleveland, Ohio; Department of Neuroradiology (V.B.H.), Feinberg School of Medicine, Northwestern University, Chicago, Illinois; Brain Tumor and Neuro-Oncology Center (V.S.), Cleveland Clinic, Cleveland, Ohio; Alberta Machine Intelligence Institute (R.V.), Edmonton, Alberta; Miami Cancer Institute (M.A.), Miami, FL and Herbert Wertheim College of Medicine, Florida International University, Florida; Department of Biomedical Engineering (A.M.), Emory University, Atlanta Veterans Administration Medical Center; Departments of Radiology and Biomedical Engineering (P.T.), University of Wisconsin Madison, Wisconsin. tional first-line treatment regimen involves maximally safe surgical resection followed by concomitant chemotherapy and radiation therapy. Despite this aggressive treatment, patient outcomes are highly variable with >40% of patients with GBM inevitably developing recurrence within 6 months following treatment initiation. While GBM heterogeneity may be reflected in the MR

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the US Department of Veterans Affairs, the Department of Defense, or the US Government.

Please address correspondence to Ruchika Verma, PhD, Alberta Machine Intelligence Institute, 10065 Jasper Ave, Edmonton, AB, T5J3B1; e-mail: ruchika@amii.ca; @Ruchika Verma

Om Indicates open access to non-subscribers at www.ajnr.org

Indicates article with online supplemental data. http://dx.doi.org/10.3174/ajnr.A7591

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award Nos. IU24CA199374-01, R01CA202752-01A1, R01CA208236-01A1, R01 CA216579-01A1, R01 CA220581-01A1, R01CA264017, IU01 CA239055-01, IU01CA248226-01; the National Institute for Biomedical Imaging and Bioengineering (IR43EB028736-01); the National Center for Research Resources under award No. 1 C06 RR12463-01; the VA Merit Review Award (IBX004121A) from the US Department of Veterans Affairs Biomedical Laboratory Research Arogram Breakthrough Level 1 Award No. W81XWH-19-1-0668; the Department of Defense Prostate Cancer Idea Development Award (W81XWH-15-1-0558); the Department of Defense Lung Cancer Investigator-Initiated Translational Research Award (W81XWH-18-1-0440); the Department of Defense Peer Reviewed Cancer Research Program (W81XWH-16-1-0329); the Kidney Precision Medicine Project Glue Grant (5T32DK747033); the Case Western Reserve University Nephrology

Training Grant; the Neptune Career Development Award; the Ohio Third Frontier Technology Validation Start-up Fund; the Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering and the Clinical and Translational Science Award Program at Case Western Reserve University (NCI 1U01CA248226-01); the Department of Defense Peer Reviewed Cancer Research Program (W81XWH-18-1-0404); the Dana Foundation David Mahoney Neuroimaging Program: and the V Foundation Translational Research Award.



FIG 1. Overview of the methodology and overall workflow. MR imaging scans (Gd-Tlw, T2w, and FLAIR) were preprocessed and segmented into 3 tumor subcompartments (necrosis, enhancing tumor, and peritumoral edema). Next, a set of radiomic features was obtained for each of the MR imaging scans across the 3 tumor subcompartments. The most stable and discriminatory features were identified using the PI score and AUC measures across every subcompartment and every MR imaging protocol. The identified stable and discriminatory features were used for stability assessment across tumor subcompartments (with the associated habitat), individual imaging protocols (Gd-TIw, T2w, FLAIR), and MR imaging acquisition parameters (magnetic field strength, pixel resolution, and slice spacing).

imaging phenotypes with different intensity profiles across the enhancing tumor and its associated habitat (comprising necrotic core, peritumoral T2/FLAIR hyperintensity), these image-based differences may not be visually appreciable to build prognostic models of survival in patients with GBM using routine MR imaging scans (Gadolinium-enhanced T1w [Gd-T1w], T2w, FLAIR).

Radiomic approaches (ie, high-throughput extraction of quantitative features) have provided a surrogate mechanism for capturing image-based phenotypes of disease heterogeneity on routine MR imaging (ie, Gd-T1w, T2w, FLAIR) toward outcome prediction in GBM and other tumors. Specifically, in GBM tumors, the availability of large, multi-institutional imaging cohorts such as The Cancer Imaging Archive-GBM (TCIA-GBM) has enabled development of prognostic models using a variety of radiomic feature families (ie, Gray, Haralick,¹ Gradient,² Laws,³ Gabor,⁴ Co-occurrence of Local Anisotropic Gradient Orientations [CoLlAGe]⁵). These prognostic models have leveraged radiomic features extracted from the tumor habitat (necrotic core, enhancing tumor, peritumoral T2/FLAIR hyperintensity) using routine MR imaging toward outcome prediction in GBM tumors.^{6,7}

these radiomic approaches is to demonstrate their generalizability to variations in image-acquisition protocols across scanners and sites. Sources of variations in MR imaging acquisition can include differing slice thicknesses, image contrast, voxel resolutions, magnetic field strengths, echo times, and repetition times.⁸⁻¹⁰ This issue is of particular concern when including data from publicly available repositories such as TCGA-GBM, where imaging scans are pooled-in from across different institutions (Fig 1*A*). While preprocessing steps including bias field correction and

However, a key challenge in enabling the clinical utility of

while preprocessing steps including bias neut correction and intensity standardization, to some extent, have allowed correction of site- and scanner-specific variations in MR imaging, multiple studies have demonstrated^{9,11,12} that these steps may not be sufficient to ensure stability of radiomic features during downstream analysis. More recently, studies have used stability measures such as the preparation-induced instability (PI) score¹³ to identify radiomic features that are stable (across sites) in the context of lung¹⁴ and prostate cancer.^{13,15,16} However, in the context of GBM, most studies have focused on "controlled" test-retest studies¹² or identifying stable features across lesion segmentations obtained from different expert readers.^{9,17,18} No work, to our knowledge, has explicitly

Table 1: Summary of	patient	demographics	and	the most	commonly	varying	MR	parameters	across	our	cohort	curated	from	5
different sites (S1-S5)													

	S1, MD Anderson Cancer Center.	S2, Ivy Glioblastoma Atlas	S3, Henry Ford Hospital. Detroit.	S4, University of California, San	S5. Cleveland
Imaging Source Site	Houston, Texas	Project	Michigan	Francisco, California	Clinic, Ohio
No. of cases	17	33	28	15	57
Magnetic field strength					
1.5T	11	9	16	14	48
3T	6	24	12	1	9
Sex					
Female	7	17	5	6	23
Male	10	16	23	9	34
Scanner					
GE Healthcare	17	30	28	15	57
Siemens	0	3	0	0	0
Slice thickness (range) (mm)	6.5	0.9–6.5	1–6	1.5–6	0.9–7.5
Pixel spacing (range) (mm)	0.47-0.86	0.43–1.05	0.43-0.94	0.45–1.02	0.23–1

interrogated the stability of radiomic features across inherent sitespecific acquisition variations within a multisite retrospective cohort in GBM tumors. Additionally, existing approaches have not explicitly interrogated the stability of radiomic features from across different subcompartments of the GBM tumor habitat. In this study, we sought to address 2 specific questions in the context of building robust prognostic image-based markers for GBM tumors–Can we identify the following: 1) a subset of radiomic features from across imaging protocols such as Gd-T1w, T2w, and FLAIR scans that are most stable across sites; and 2) a set of cross-site stable radiomic features from each tumor subcompartment that are also discriminatory of progression-free-survival (PFS) in GBM tumors?

MATERIALS AND METHODS

Data Description

A total of 150 retrospectively analyzed treatment-naive multiparametric GBM MR imaging scans (Gd-T1w, T2w, FLAIR) along with their PFS information were obtained from 5 different institutions, including studies from TCGA-GBM (n = 60)¹⁹ and the Ivy Glioblastoma Atlas Project (n = 33)²⁰ cohorts, as illustrated in Table 1. We used median PFS obtained from our training cohort to segregate the studies into improved (PFS > 6.5 months) and poor (PFS ≤ 6.5 months) GBM survivors. Additional details regarding the data curation and preprocessing are provided in the Online Supplemental Data.

Segmentation of the Tumor Habitat. Following preprocessing, each 2D MR imaging slice with visible tumor was annotated by expert readers (V.B.H. with >15 years of experience, V.S. with 12 years of experience, and K.B. with >3 years of experience in neuroradiology) into 3 subcompartments: 1) enhancing tumor, 2) FLAIR hyperintensities (edema and nonenhancing tumor), and 3) necrosis. Tumor lesion annotations were also used to select the nontumor ROI for every patient volume. The nontumor region was defined as the region obtained by mirroring the tumor lesion (all 3 subcompartments in conjunction) onto the contralateral hemisphere. The nontumor region for each study was manually inspected by expert readers to verify that it did not contain any proportion of the tumor region or skull. An overview of the entire radiomics analysis workflow (including preprocessing and segmentation) is provided in Fig 1.

Radiomic Feature Extraction. For every tumor subcompartment (enhancing tumor, peritumoral FLAIR hyperintensities, and necrosis), a set of 1264 radiomic features were extracted across all 3 multiparametric MR imaging (Gd-T1w, T2w, and FLAIR) scans. These radiomic features included Laws energy,³ Haralick,¹ Gabor,⁴ CoLlAGe,⁵ Gradient,² and first-order gray level co-occurrence matrices (additional details provided in the Supplemental Online Data, A.3). A total of 316 three-dimensional radiomic textural features were extracted on a per-voxel basis with window sizes of 3 and 5 from each tumor subcompartment per the MR imaging scan. Four first-order statistics (median, variance, skewness, kurtosis) per feature descriptor were computed within each tumor subcompartment, yielding 1264 statistical features per region (enhancing tumor, peritumoral FLAIR hyperintensity, and necrosis) per the MR imaging protocol. When statistical features were concatenated across all 3 MR imaging sequences (Gd-T1w, T2w, and FLAIR), a total of 3792 texture features were obtained for analysis from each tumor subcompartment. Additionally, to identify stable features across multiple sites, we extracted 3792 statistical features from the nontumor brain parenchyma (mirroring the tumor mask on the opposite brain hemisphere). Following feature extraction, z score feature normalization²¹ was applied to ensure that radiomic features extracted from different sites lie within a comparable range of values for leave-one-site-out (LOSO) analysis. The mean and standard deviation (SD) feature values obtained from the training data were further used to apply zscore normalization on the independent holdout validation set.

Computation of Feature Instability across Sites. As described in Leo et al,¹³ the PI uses a statistical *t* test to compare the number of feature distributions that are significantly different across data sets from different sites. The PI score ranges between 0 and 1 with a high PI score for a feature representing low stability across site-specific variations, while a low PI score (closer to 0) indicates that the feature may not be affected by acquisition variations and is likely stable across sites.¹³



FIG 2. The strategy used for segregating our cohort into training $(\hat{T}_1 - \hat{T}_4)$, validation (VS₁–VS₄), and TSs. The training and the validation sets were created using the LOSO scheme.

Statistical Analysis

To perform a rigorous evaluation of feature stability across sites, our multi-institutional GBM cohort was segregated into 3 groups: training set (\hat{T}) , validation set (VS), and test set (TS) (Fig 2). Using these 3 groups, we assessed radiomic feature stability across subcompartments of the tumor and its associated habitat, individual imaging protocols (Gd-T1w, T2w, and FLAIR), and common sources of variance in MR imaging parameters (magnetic strength, pixel spacing, slice thickness). Specifically, we used sites S1-S4 in a LOSO fashion so that studies from 3 sites at a time consisted of \widehat{T} where $i \in (1,...,4)$, while studies from the remaining fourth site were used as VS_i. Finally, the TS consisted of studies obtained from our hold-out site (Cleveland Clinic Foundation), which was used for independent evaluation. For every training set \hat{T} , $i \in (1,...,4)$, we computed the PI score to measure the variability in radiomic feature values engendered due to acquisition variations across the 3 sites within the \hat{T} . The features with PI > 0 indicated that they were significantly different in the nontumor regions across different sites in the TS and hence were excluded from further analysis. After this triage, the remaining stable feature set was used to identify the most discriminatory features using minimum redundancy maximum relevance (mRMR) feature selection within a linear discriminant analysis classifier across 100 iterations of bootstrapping.

This LOSO analysis was performed separately for each of the 3 tumor subcompartments (enhancing tumor, peritumoral hyperintensity, necrotic core) to identify the most stable and discriminatory features from the training set for every MR imaging protocol (Gd-T1w, T2w, and FLAIR). The average area under the curve (AUC) across different iterations of LOSO was used to obtain discriminability values in identifying poor and improved survival across each of the tumor subcompartments. Furthermore, for each tumor subcompartment, we concatenated and rank-ordered stable and discriminatory features in terms of their frequency of occurrence across 400 trials (100 bootstrapped iterations across 4 runs of LOSO evaluation as shown in Fig 2) of cross-validation run on our training set. The 5 most-frequently selected features using \widehat{T} and VS were obtained and evaluated on the TS for their efficacy in distinguishing poor from improved survival in GBMs using the AUC and accuracy measures. Additionally, for comparison, we selected radiomic features using only minimum Redundancy-Maximum Relevance (discriminatory features) without inclusion of the PI

1118

metric for feature pruning. We performed the LOSO runs on discriminatory features using a linear discriminant analysis classifier to evaluate their performance (for distinguishing poor and improved survival) in terms of the AUC and accuracy on our hold-out validation set.

Lastly, we interrogated the feature distributions of stable and discriminatory radiomic features and compared them with discriminatory features using box-and-whisker plots. Specifically, the comparisons were made using the Wilcoxon rank-sum test in the context of evaluating variability in feature distri-

butions across 3 commonly varying MR imaging parameters: 1) slice thickness ($\leq 2 \text{ mm}$, 2-5 mm, $\geq 5 \text{ mm}$), 2) pixel spacing (< 0.5 mm, $\geq 0.5 \text{ mm}$), and 3) magnetic field strength (1.5T versus 3T), across our multi-institutional cohort.

RESULTS

Experiment 1: Assessing the Stability of Radiomic Features across Different Subcompartments of the Tumor and Its Associated Habitat

Figure 3*A*-*C* shows 2D scatterplots of discriminability (average AUC on the y-axis) versus instability (PI on the x-axis), where each data point in the 2D space corresponds to 1 of the 3792 radiomic features across every tumor subcompartment (necrosis, enhancing tumor, and FLAIR hyperintensity, respectively). The size of the data point on the scatterplot was used to represent the SD in AUC values for a specific feature across different LOSO runs. Our results suggested that the features with PI closer to zero also had a consistent diagnostic performance (in terms of AUC values) across sites as evident from the small size of the data points (reflecting a small deviation in the AUC) in Fig 3. The top 5 most frequent stable and discriminatory radiomic features per subcompartment selected using LOSO subexperiments are listed in Table 2.

Figure 4 illustrates the AUC values for training $(\hat{T}_1 - \hat{T}_4)$ and validation (VS₁-VS₄) cohorts using the top 5 stable and discriminatory radiomic features versus discriminatory features alone (without inclusion of the stability metric) from each of the 3 tumor subcompartments using LOSO analysis. Confidence intervals derived from across the bootstrapped experiments for the AUC values in the training set are available in the Online Supplemental Data. The AUC values on the training set did not show much improvement between stable and discriminatory and discriminatory classifiers.

However, on the validation set (VS₁–VS₄), the compartmentspecific linear discriminant analysis classifiers trained using the top 5 stable and discriminatory radiomic features yielded at least 5%–10% improvement in AUC values (0.78, 0.64, 0.66, 0.71 for VS₁–VS₄ respectively), compared with the linear discriminant analysis classifiers trained using discriminatory features alone (0.57, 0.51, 0.52, 0.32 for VS₁–VS₄, respectively) in the enhancing tumor and T2/FLAIR hyperintensity subcompartments (0.76,



FIG 3. 2D plots of discriminability (average AUC, y-axis) versus instability (PI score, x-axis) for each tumor subcompartment (A, Necrosis. B, Enhancing tumor. C, T2/FLAIR hyperintensity). Each radiomic feature is shown as a bubble where its size represents the SD between the AUC values across different LOSO runs. D, The stacked barplot of the total number of extracted features, the features that were identified as stable (PI = 0) from across different feature families per MR imaging sequence (Gd-TIw, T2w, and FLAIR), and the overlapping stable features across multiparametric MR imaging sequences. Different colors represent different feature families.

0.69, 0.73, 0.77 using stable and discriminatory features compared with 0.51, 0.59, 0.53, 0.70 from discriminatory features alone across VS_1-VS_4). For the necrosis subcompartment, the AUC values for VS were identified as 0.68, 0.52, 0.58, and 0.75 using discriminatory-alone across VS_1-VS_4 and 0.69, 0.68, 0.73, and 0.70 for stable and discriminatory features, respectively.

Our compartment-specific classifiers trained using the top 5 stable and discriminatory radiomic features yielded AUCs of 0.71, 0.73, and 0.76, while discriminatory-alone features yielded AUCs of 0.65, 0.54, and 0.62 for necrotic core, enhancing tumor, and FLAIR hyperintensities, respectively, on the hold-out testing cohort. Additionally, combining stable and discriminatory features from across the tumor habitat yielded an AUC of 0.78 compared

with using discriminatory features alone, which yielded an AUC of 0.69 on the hold-out test set.

Experiment 2: Assessing the Stability of Radiomic Features across Individual Imaging Protocols (Gd-T1w, T2w, and FLAIR)

Figure 3*D* shows stacked barplots corresponding to the total number of features (basic set) along with the proportion of features that were identified as cross-site stable with a PI = 0 (stability-informed set) from each feature family across Gd-T1w, T2w, and FLAIR sequences. Of a total of 3792 features for every protocol (1264 features per subcompartment), a total of 240 features from Gd-T1w, 302 features from T2w and 306 features from

Table 2: A list of the top 5 features identified to reliably differentiate poor from improved survival in patients with GBM from the training set for every tumor subcompartment (necrosis, enhancing tumor, peritumoral T2/FLAIR hyperintensity, and nonenhancing tumor)

No.	Feature Family	Descriptor	Window Size	MRI Sequence	Statistic
Top 5 features selected from necrosis					
1	Laws	Edge-edge-level	5	TI	Skewness
2	Laws	Level-level-level	3	TI	Skewness
3	Laws	Wave-wave-level	5	FLAIR	Median
4	CoLIAGe	Sum variance	3	FLAIR	Skewness
5	CoLIAGe	Sum variance	3	FLAIR	Kurtosis
Top 5 features selected from					
enhancing tumor					
1	CoLIAGe	Sum average	3	T2	Median
2	CoLIAGe	Sum average	5	T2	Median
3	CoLIAGe	Energy	3	FLAIR	Median
4	CoLIAGe	Sum average	3	FLAIR	Median
5	Laws	Spot-level-ripple	5	TI	Skewness
Top 5 features selected from edema					
and nonenhancing tumor					
1	Laws	Level-edge-level	5	TI	Median
2	CoLIAGe	Sum variance	5	T2	Variance
3	CoLIAGe	Sum variance	5	T2	Skewness
4	Haralick	Sum entropy	5	FLAIR	Skewness
5	Gabor	$-\theta = 1.178$, XZ $-\theta = 0$,	5	T2	Kurtosis
		$\lambda =$ 1.482, BW $=$ 1			

Note:-BW indicates bandwidth.

FLAIR were found to be stable (using PI = 0) across the studies from the training set (\hat{T}) sets. While ~20% of features from every feature family were identified as stable across all the 3 MR imaging protocols, the greatest number of stable features was found to belong to Laws, CoLlAGe, and Gabor feature families (Online Supplemental Data).

Experiment 3: Assessing the Radiomic Stability across Common Sources of Variance in MR Imaging Parameters (Magnetic Strength, Pixel Resolution, Slice Spacing)

Figure 2 demonstrates the boxplots representing feature distributions for a stable and discriminatory radiomic feature (CoLIAGe sum average feature) along with a corresponding discriminatory feature (Haralick info2), both extracted from enhancing tumor of FLAIR MR imaging across test sets, dichotomized across different slice thicknesses (millimeter): ≤ 2 , (n =34), and >2 (n = 23). Similarly, Fig 3A, -B and Fig 4 represent feature distributions of the corresponding features acquired at different pixel spacing categorized such as < 0.5 (n = 16) and $\geq 0.5 \,\mathrm{mm}$ (n = 41), as well as across different magnetic strengths (1.5T [n = 48] versus 3T [n = 9]), respectively. We observed the range and median values of stable and discriminatory radiomic features at different slice thicknesses, pixel spacing, and magnetic strengths to be more consistent (Online Supplemental Data) than the feature distributions obtained for discriminatory features (Online Supplemental Data). Similarly, we observed statistically significant separations (demonstrated with asterisks in the boxplots of Online Supplemental Data) in feature distributions between poor and improved survivors for stable and discriminatory radiomic features, compared with the discriminatory features, across slice thicknesses, pixel spacing, and magnetic strength variations in our training and testing cohorts.

DISCUSSION

Estimation of PFS often serves as a surrogate marker for predicting therapeutic efficacy in GBM tumors. Previous approaches have investigated radiomic features extracted from different subcompartments of the tumor habitat (constituting necrotic core, enhancing tumor, and T2/FLAIR hyperintensity subcompartments) on routine MR imaging in prognosticating PFS in patients with GBM. However, most of these studies do not explicitly take into account the sensitivity of radiomic features to site- and scanner-specific variations in MR imaging parameters, including variations in slice thicknesses, image contrast, voxel resolutions, and magnetic field strengths, which ultimately impact the generalizability and clinical applicability of these approaches.

More recently, a few approaches have focused on identifying stable and discriminatory features (using preparation-induced instability score¹³) across multi-institutional studies in the context of other tumor sites (lung,¹⁴ prostate^{13,15,16}). Our work leveraged a similar approach to identify radiomic features that are stable as well as discriminatory of poor- and improved-survival groups in the context of brain tumors. To rigorously evaluate the reproducibility of our selected radiomic features, we performed a comparison of stable and discriminatory features with discriminatoryalone features in a LOSO fashion (across 4 training sites) for distinguishing poor and improved GBM survivors on the basis of their PFS. This analysis was performed across each of the tumor subcompartments of the tumor habitat on routine multiparametric pretreatment MR imaging scans (Gd-T1w, T2w, FLAIR). Lastly, for each tumor subcompartment, the most frequent stable and discriminatory radiomic features in the training set were used to risk-stratify patients with GBM into poor and good survivors on an independent cohort obtained from a collaborating institute (Cleveland Clinic). Our findings on training and test



FIG 4. Barplots showing training and validation AUCs to distinguish patients with improved and poor outcomes using the top 5 stable and discriminatory radiomic versus discriminatory-only features. *Blue barplots* depict the AUC values using discriminatory features, while *orange bars* represent AUC values using stable and discriminatory radiomic features from the respective training and validation cohorts.

sets are in line with those in previous work,^{13,14} in which the use of PI score within feature selection allowed building improved diagnostic and prognostic classifiers using stable and discriminatory features.

In the context of GBM prognosis, a few studies that have interrogated variations in radiomic features have focused on the reproducibility across segmentations of the tumor obtained from across different experts/institutions^{9,12,17,18} or, in terms of repeatability, using test-retest studies.¹² For instance, Tixier et al¹⁷ explored the robustness of radiomic features extracted from the TCGA-GBM data set of 90 studies across segmentations obtained from different experts. They reported that the features computed from the histogram and co-occurrence matrices were the most robust with an intraclass correlation coefficient value of ≥ 0.8 . Similarly, Lee et al²² compared radiomic features (textural, morphometric, and first-order) across segmentations of the tumor habitat obtained from 2 semiautomated segmentation software programs on 45 studies and identified the first-order statistic-feature family (across all 3 tumor subcompartments) to be the most stable across the automated segmentations. In a recent study, Shiri et al¹² evaluated the repeatability of radiomic features across 17 patients with GBM using T1- and T2-weighted MR imaging scans obtained within the same imaging unit on 2 consecutive days. By means of the intraclass correlation coefficient, the study demonstrated that the textural features from the gray-level run length matrix and gray-level dependence matrix were highly repeatable (intraclass correlation coefficient, >95%) with respect to image preprocessing, different image-registration algorithms, and test-retest analysis. Similar features, including difference variance, inverse different moment, fraction, and long- and short-run emphasis were found to be highly reproducible among different field sizes and phantom positions in Rastegar et al.²³ In Suter et al,²⁴ the authors applied multiple perturbations, including variability in the imaging protocols (voxel size and axial slice spacing), artificial deformation in the segmentation labels, k-space subsampling on MR imaging, and selected robust features based on the intraclass correlation coefficient from a single site. Selected robust features along with machine learning classifiers evaluated on multicenter data demonstrated an improvement in the prognostic power of the models compared with the machine learning classifiers trained with nonrobust features.

Our work is different from previous related studies²⁵⁻²⁹ in a few ways. First, our work presented the first approach in the context of GBM tumors to jointly explore the stability and discriminability of radiomic features. Our results suggest that the radiomic features from T2/FLAIR hyperintensity are the most stable and discriminatory across different tumor subcompartments. Second, we carefully interrogated a set of stable features from across different MR imaging sequences (Gd-T1w, T2w, FLAIR). Finally, we leveraged retrospective MR imaging scans acquired from 5 different institutions to build a robust prognostic classifier that incorporated both stable and discriminatory features. Uniquely, we demonstrated that feature distributions across different slice thicknesses, pixel spacing, and magnetic field strengths were more consistent across stable and discriminatory radiomic features than the feature distributions obtained for the discriminatory features.

While the performance accuracy across stable and discriminatory features was similar to that of discriminatory features alone on the training set, there are 2 key points worth noting: 1) The model trained on stable and discriminatory was robust to the variability in training data because the SD of the AUC values was much higher for discriminatory-alone compared with stable and discriminatory features (Online Supplemental Data). 2) We observed an improvement in AUC values on the independent hold-out test set, suggesting that the model trained on stable and discriminatory features may be more generalizable and hence more amenable to providing reliable AUC assessments on the hold-out set, compared with the model trained on discriminatory features alone.

Our work did have some limitations. While our analysis included 150 patients from 5 different institutions, it was limited to a small cohort of studies per site. The retrospective nature of our cohort led to an unequal number of patients per site, which may have further impacted our analysis of radiomic variations in slice thickness, pixel spacing, and magnetic strengths across sites. Notably, our model yielded the lowest AUCs when validated on VS2 (the Ivy Glioblastoma Atlas Project cohort). A possible reason may be the high variability in acquisition parameters across scanners (GE Healthcare and Siemens), variations in the magnetic field strength, and a high range of slice thicknesses (0.9-6.5 mm) and pixel spacing (0.43-1.05 mm), compared with data sets from the other cohorts. Additional work is warranted to rigorously validate the reproducibility of the identified features across multiple different imaging parameters on a much larger multisite cohort. We will additionally also investigate the repeatability of the radiomic features from across the tumor habitat via a prospectively collected test-retest study for the GBM cohort.

CONCLUSIONS

Our approach demonstrated that the radiomic features from the T2/FLAIR hyperintentensity subcompartment were the most stable and discriminatory of PFS in GBM tumors compared with the features from enhancing tumor and necrotic core. Identification of stable and discriminatory radiomic features across multisite multiparametric MR imaging (Gd-T1w, T2w, FLAIR) sequences from within the tumor and its associated habitat may yield robust prognostic classifiers for patient survival in GBM tumors.

Disclosures: Virginia Hill—UNRELATED: Other: Google,* Comments: I am our departmental liaison with Google. They hire some of our radiologists to consult. Kaustav Bera-UNRELATED: Employment: Case Western Reserve University. Patrick Leo-RELATED: Grant: National Science Foundation,* Comments: Graduate Research Fellowship Program (CON501692). Manmeet Ahluwalia-RELATED: Consulting Fee or Honorarium, Comments: None, UNRELATED: Consultancy: Bayer, Novocure, Kiyatec, Insightec, GSK, Xoft, Nuvation, Cellularity, SDP Oncology, Apollomics, Prelude, Janssen, Tocagen, Voyager Therapeutics, Viewray, Caris Lifesciences, Pyramid Biosciences, Varian Medical Systems, Cairn Therapeutics, Anheart Therapeutics; Grants/Grants Pending: Astrazeneca, Abbvie, BMS, Baver, Incyte, Pharmacyclics, Novocure, Merck, Mimiyax, Novartis, Roswell Park Cancer Foundation, Velosano; Royalties: Comments: Wiley; Stock/Stock Options: Cytodyn, Doctible, Mimivax, Medinnovate Advisors LLC: Other: MedInnovate Advisors LLC. Anant Madabhushi-UNRELATED: Board Membership: Aiforia; Consultancy: Aiforia; Grants/Grants Pending: Bristol Myers Squibb, Astrazeneca, Boehringer-Ingelheim;* Stock/Stock Options: Elucid Bioimaging. Pallavi Tiwari—RELATED: Grant: National Institute of Health/National Cancer Institute U01 (1U01CA248226-01), Department of Defense Peer Reviewed Cancer Research Program (W81XWH-18-1-0404), Dana Foundation David Mahoney Neuroimaging Grant, the CCCC Brain Tumor Pilot Award, the CWRU Technology Validation Start-Up Fund (CTP), Johnson and Johnson WiSTEM2D Research Scholar Award, and The V Foundation Translational Research Award* --- UNRELATED: Employment: Case Western Reserve University. *Money paid to the institution.

REFERENCES

- Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* 1973;SMC-3:610–21 CrossRef
- Huang LL, Shimizu A, Hagihara Y, et al. Gradient feature extraction for classification-based face detection. *Pattern Recognition* 2003;36:2501–11 CrossRef
- Laws KI. Rapid texture identification. Image Processing for Missile Guidance 1980;0238:376–81 CrossRef
- Manjunath BS, Ma WY. Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell 996;18:837–42 CrossRef

- 5. Prasanna P, Tiwari P, Madabhushi A. Co-occurrence of Local Anisotropic Gradient Orientations (CoLlAGe): a new radiomics descriptor. *Sci Rep* 2016;6:37241 CrossRef Medline
- Coroller TP, Grossmann P, Hou YC, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol* 2015;114:345–50 CrossRef Medline
- Lao J, Chen Y, Li Z-C, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep* 2017;7:10353 CrossRef Medline
- Styner MA, Charles HC, Park J, et al. Multisite validation of image analysis methods: assessing intra- and intersite variability. *Proc* SPIE Int Soc Opt Eng 2002;46841:278–86 CrossRef
- 9. Um H, Tixier F, Bermudez D, et al. Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets. *Phys Med Biol* 2019;64:165011 CrossRef Medline
- Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 2016;6:23428 CrossRef Medline
- 11. Bakas S, Akbari H, Sotiras A, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data* 2017;4:170117 CrossRef Medline
- Shiri I, Hajianfar G, Sohrabi A, et al. Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: test-retest and image registration analyses. *Med Phys* 2020;47:4265–80 CrossRef Medline
- Leo P, Lee G, Shih NC, et al. Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images. J Med Imaging (Bellingham) 2016;3:047502 CrossRef Medline
- Khorrami M, Bera K, Leo P, et al. Stable and discriminating radiomic predictor of recurrence in early stage non-small cell lung cancer: multi-site study. *Lung Cancer* 2020;142:90–97 CrossRef Medline
- 15. Hiremath A, Shiradkar R, Merisaari H, et al. Test-retest repeatability of a deep learning architecture in detecting and segmenting clinically significant prostate cancer on apparent diffusion coefficient (ADC) maps. *Eur Radiol* 2021;31:379–91 CrossRef Medline
- Merisaari H, Taimen P, Shiradkar R, et al. Repeatability of radiomics and machine learning for DWI: short-term repeatability study of 112 patients with prostate cancer. *Magn Reson Med* 2020;83:2293– 309 CrossRef Medline

- Tixier F, Um H, Young RJ, et al. Reliability of tumor segmentation in glioblastoma: impact on the robustness of MRI-radiomic features. *Med Phys* 2019;46:3582–91 CrossRef
- Pati S, Verma R, Akbari H, et al. Reproducibility analysis of multiinstitutional paired expert annotations and radiomic features of the Ivy Glioblastoma Atlas Project (Ivy GAP) dataset. *Med Phys* 2020;47:6039–52 CrossRef Medline
- Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 2013;26:1045–57 CrossRef Medline
- 20. Shah N, Feng X, Lankerovich M, et al. Data from Ivy GAP. The Cancer Imaging Archive 2016 CrossRef
- Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems. Pattern Recognition 2005;38:2270–85 CrossRef
- 22. Lee M, Woo B, Kuo MD, et al. Quality of radiomic features in glioblastoma multiforme: impact of semi-automated tumor segmentation software. *Korean J Radiol* 2017;18:498–509 CrossRef Medline
- 23. Rastegar S, Beigi J, Saeedi E, et al. Radiographic image radiomics feature reproducibility: a preliminary study on the impact of field size. *J Med Imaging Radiat Sci* 2020;51:128–36 CrossRef Medline
- 24. Suter Y, Knecht U, Alão M, et al. Radiomics for glioblastoma survival analysis in pre-operative MRI: exploring feature robustness, class boundaries, and machine learning techniques. Cancer Imaging 2020;20:55 CrossRef Medline
- 25. Moradmand H, Aghamir SMR, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. J Appl Clin Med Phys 2020;21:179–90 CrossRef Medline
- 26. Aerts HJ, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006 CrossRef Medline
- 27. Verma R, Correa R, Hill VB, et al. Tumor habitat-derived radiomic features at pretreatment MRI that are prognostic for progression-free survival in glioblastoma are associated with key morphologic attributes at histopathologic examination: a feasibility study. *Radiol Artif Intell* 2020;2:190168 CrossRef Medline
- Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and prognosis of quantitative features extracted from CT images. *Transl Oncol* 2014;7:72–87 CrossRef Medline
- Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005;27:1226–38 CrossRef Medline