NEURO

Machine learning and radiomic phenotyping of lower grade gliomas: improving survival prediction

Yoon Seong Choi[1,2] · Sung Soo Ahn[1] (iD) · Jong Hee Chang[3] · Seok-Gu Kang[3] · Eui Hyun Kim[3] · Se Hoon Kim[4] · Rajan Jain[5,6] · Seung-Koo Lee[1]

Abstract

Background and purpose Recent studies have highlighted the importance of isocitrate dehydrogenase (*IDH*) mutational status in stratifying biologically distinct subgroups of gliomas. This study aimed to evaluate whether MRI-based radiomic features could improve the accuracy of survival predictions for lower grade gliomas over clinical and *IDH* status.

Materials and methods Radiomic features ($n = 250$) were extracted from preoperative MRI data of 296 lower grade glioma patients from databases at our institutional ($n = 205$) and The Cancer Genome Atlas (TCGA)/The Cancer Imaging Archive (TCIA) ($n = 91$) datasets. For predicting overall survival, random survival forest models were trained with radiomic features; non-imaging prognostic factors including age, resection extent, WHO grade, and *IDH* status on the institutional dataset, and validated on the TCGA/TCIA dataset. The performance of the random survival forest (RSF) model and incremental value of radiomic features were assessed by time-dependent receiver operating characteristics.

Results The radiomics RSF model identified 71 radiomic features to predict overall survival, which were successfully validated on TCGA/TCIA dataset (iAUC, 0.620; 95% CI, 0.501–0.756). Relative to the RSF model from the non-imaging prognostic parameters, the addition of radiomic features significantly improved the overall survival prediction accuracy of the random survival forest model (iAUC, 0.627 vs. 0.709; difference, 0.097; 95% CI, 0.003–0.209).

Conclusion Radiomic phenotyping with machine learning can improve survival prediction over clinical profile and genomic data for lower grade gliomas.

Key Points

• *Radiomics analysis with machine learning can improve survival prediction over the non-imaging factors (clinical and molecular profiles) for lower grade gliomas, across different institutions.*

Keywords Glioma · Machine learning · Prognosis · Survival

Abbreviations

FLAIR	Fluid-attenuated inversion recovery
iAUC	Integrated area under the ROC curve
IDH	Isocitrate dehydrogenase
IRB	Institutional research board
LGG	Lower grade glioma
OS	Overall survival
RF	Random forest

✉ Sung Soo Ahn
SUNGSOO@yuhs.ac

1 Department of Radiology and Research Institute of Radiological Science, College of Medicine, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea

2 Department of Diagnostic Radiology, Singapore General Hospital, Singapore, Singapore

3 Department of Neurosurgery, Yonsei University College of Medicine, Seoul, South Korea

4 Department of Pathology, Yonsei University College of Medicine, Seoul, South Korea

5 Department of Radiology, Langone Medical Center, New York University School of Medicine, New York, NY, USA

6 Department of Neurosurgery, Langone Medical Center, New York University School of Medicine, New York, NY, USA

ROC Receiver operating characteristics
RSF Random survival forest
T1C T1-weighted contrast-enhanced
T2   T2-weighted
TCGA The Cancer Genome Atlas
TCIA The Cancer Imaging Archive
TE   Echo time
TR   Repetition time

Introduction

Glioma is the most common primary brain tumor, accounting for approximately 51% of all central nervous system tumors [1]. Lower grade gliomas (LGGs) are WHO grade II and III tumors; they show heterogeneous survival outcomes, with some LGGs exhibiting high therapeutic sensitivity [2]. Thus, prior studies have investigated important clinical predictors of survival for patients with LGG, such as age and extent of tumor resection [3, 4]. The importance of molecular biomarkers, such as isocitrate dehydrogenase (*IDH*) mutational status in survival prediction for LGG has been recognized [5]; *IDH* mutation is associated with better survival outcomes for patients with LGG [5–7]. Additionally, imaging biomarkers, such as enhancement pattern, have been investigated for survival prediction [8]. However, most imaging parameters in prior studies were heterogeneous and non-quantitative [8–10]. Here, we attempted to extract extensive quantitative information by radiomics analysis and apply it for survival prediction using machine learning tools.

Radiomics is based on the conversion of medical images to high-dimensional quantitative information that reflects underlying pathophysiology, especially tumor heterogeneity [11]. Radiomics has recently gathered attention for its role in predicting genetic status and survival in patients with brain tumors [12–14]. However, many prior studies have focused on the prediction of survival or genetic alterations such as *IDH* status of LGG by radiomics alone, and some have not performed external validation. However, decision-making for LGG patients depends on multi-level prognostic information including *IDH* status and clinical information such as age or extent of tumor resection; therefore, we focused on the role of radiomics in predicting survival within such multi-level prognostic information. We also conducted external validation using the TCGA/TCIA (The Cancer Genome Atlas/The Cancer Imaging Archive) dataset to assess the generalizability of radiomic phenotyping. The aim of this study was to evaluate whether MRI-based radiomic features could improve the accuracy of survival predictions for LGG over clinical and *IDH* status.

Materials and methods

This retrospective study was conducted with approval from our institutional research board (IRB). The need for acquiring informed consent was waived. The National Institute of Health/National Cancer Institute–approved TCGA and TCIA databases contain publicly available datasets in which all data are anonymized [2, 15]. Therefore, IRB approval from individual institutions was not required to use information from TCGA and TCIA databases in this present study.
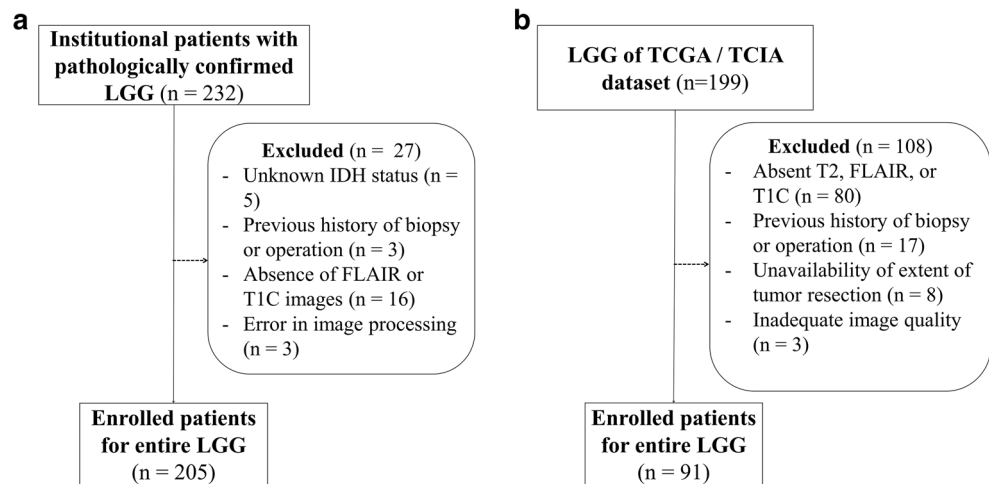
Study cohort

A flow chart of patient enrollment is shown in Fig. 1. Two cohorts were enrolled, i.e., an institutional cohort and a TCGA/TCIA cohort (hereinafter referred to as the "TCGA cohort"). In the institutional cohort, 232 patients were pathologically diagnosed with WHO grade II and III gliomas and underwent preoperative MRI between September 2007 and January 2017. The inclusion criteria were as follows: (1) histopathologically confirmed WHO grade II or III glioma; (2) known *IDH* mutation status; (3) no prior brain biopsy, surgery, or other treatments; (4) age ≥ 18 years; and (5) availability of preoperative T2-weighted (T2), fluid-attenuated inversion recovery (FLAIR), and T1-weighted contrast-enhanced (T1C) images. Based on these criteria, the following were excluded: (1) unknown *IDH* mutation status ($n = 5$); (2) previous history of biopsy or operation of brain tumor ($n = 3$); (3) absence of FLAIR or T1C images ($n = 16$); and (4) error in image processing ($n = 3$).

For TCGA cohort, the medical records of 199 patients with available MRI data from the LGG dataset of the TCGA/TCIA database [2] were retrospectively reviewed. The inclusion criteria were as follows: (1) pathologically confirmed WHO grade II and III gliomas; (2) known clinical information, including age, *IDH* mutation status, and extent of tumor resection; (3) no prior brain biopsy, surgery, or other treatments; and (4) availability of preoperative T2, FLAIR, and T1C images. Based on these criteria, the following were excluded: (1) absence of FLAIR, T2, or T1C images ($n = 80$); (2) previous history of biopsy or surgery ($n = 17$); (3) unavailability of information on extent of tumor resection ($n = 8$); and (4) inadequate image quality ($n = 3$). Thus, a total of 205 and 91 patients were selected for the institutional and TCGA cohorts, respectively.

IDH mutation status and clinical information

In the institutional cohort, the IDH R132H mutation was detected using peptide nucleic acid–mediated clamping polymerase chain reaction and immunohistochemical analysis. Along with *IDH* status, WHO grade and the clinical information including age, sex, resection extent, and overall survival

Fig. 1 Patient enrollment processes for the institutional (**a**) and TCGA (**b**) cohorts. *IDH*, isocitrate dehydrogenase; FLAIR, fluid-attenuated inversion recovery images; LGG, lower grade gliomas; T1C, T1-weighted contrast-enhanced images; T2, T2-weighted images; TCGA, The Cancer Genome Atlas; TCIA, The Cancer Imaging Archive

a

Institutional patients with pathologically confirmed LGG (n = 232)

Excluded (n = 27)
- Unknown IDH status (n = 5)
- Previous history of biopsy or operation (n = 3)
- Absence of FLAIR or T1C images (n = 16)
- Error in image processing (n = 3)

Enrolled patients for entire LGG (n = 205)

b

LGG of TCGA / TCIA dataset (n=199)

Excluded (n = 108)
- Absent T2, FLAIR, or T1C (n = 80)
- Previous history of biopsy or operation (n = 17)
- Unavailability of extent of tumor resection (n = 8)
- Inadequate image quality (n = 3)

Enrolled patients for entire LGG (n = 91)

(OS) information were recorded. Resection extent was categorized as subtotal (gross tumor removal ≥ 75% but < 100%), or partial (gross tumor removal, < 75%) or biopsy based on postoperative MRI findings. Details on *IDH* mutation status and clinical information are available in Table 1.

For the TCGA cohort, histological findings and grading of tissues submitted to TCGA were confirmed by neuropathology review, as previously described [2, 15]. The clinical information, WHO grade, and *IDH* status for the TCGA cohort were recorded, similar to the institutional cohort.

Image acquisition and postprocessing

The image processing workflow is shown in Fig. 2. MR images of the institutional cohort were acquired using a 3.0-T MRI (Achieva, Philips) and an 8-channel SENSE head coil.

The preoperative MRI protocol included acquisition of T1-weighted (repetition time [TR], 2000 ms; echo time [TE], 10 ms; field of view [FOV], 240 mm; slice thickness, 5 mm; and matrix size, 256 × 256); T2 (TR, 3000 ms; TE, 80 ms; FOV, 240 mm; slice thickness, 5 mm; and matrix size, 256 × 256); and FLAIR (TR, 10,000 ms; TE, 125 ms; FOV,

Table 1 Clinical characteristics of the patients

| Clinical characteristics	Institutional cohort (n = 205)	TCGA cohort (n = 91)	p value*
Median overall survival (days)	2986	1859	0.008[a]
Number of deaths observed	49 (23.4%)	22 (24.2%)	
Age, in years	44.6 (12.7)	46.8 (13.6)	0.195
Gender			0.389
Male	108 (52.7%)	43 (47.3%)	
Female	97 (47.3%)	48 (52.7%)	
Resection extent			0.189
Total	90 (43.9%)	48 (52.7%)	
Subtotal or partial	89 (43.4%)	37 (40.7%)	
Biopsy	26 (12.7%)	6 (6.6%)	
IDH mutation			0.059
Wild type	69 (33.7%)	20 (22.0%)	
Mutant	136 (66.3%)	71 (78.0%)	
WHO grade			
Grade II	122 (59.5%)	42 (46.2%)	0.045
Grade III	83 (40.5%)	49 (53.8%)	

Data are expressed as mean with standard deviation in parentheses or number with percentage in parentheses, except median overall and progression-free survival

NA not applicable

[a] Calculated from the log-rank test

*Calculated from the Student *t* test for continuous variables and chi-square test for categorical variables, if not otherwise mentioned
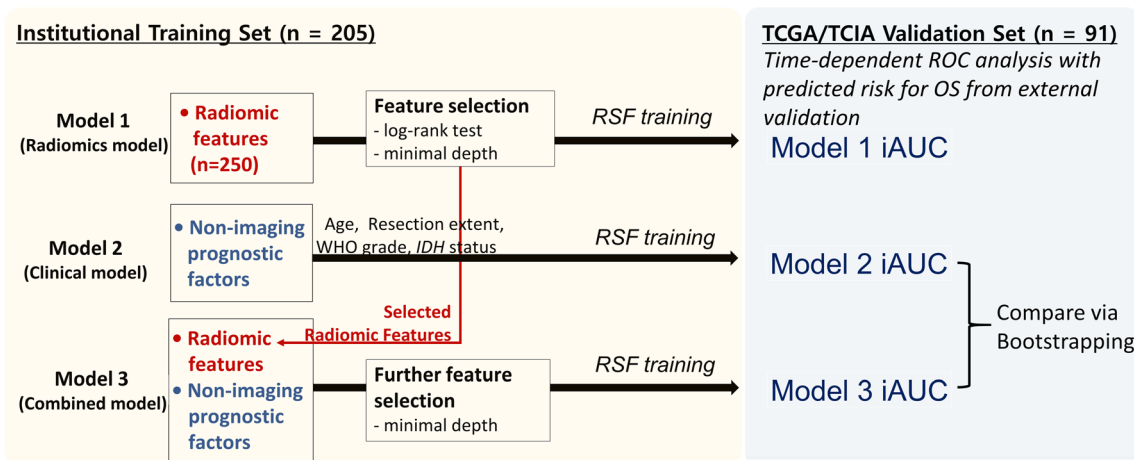
1) Image processing

Skull stripping & Registration

Normalization
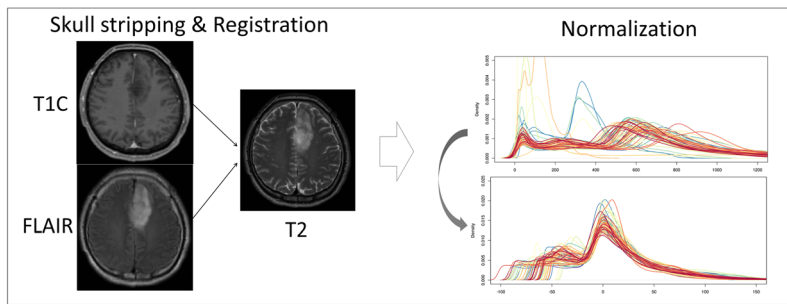
2) Feature Extraction

- Shape
- First order
- GLCM
- GLRLM
- GLSZM

3) Machine Learning

Fig. 2 Workflow of external validation of radiomics models for predicting OS. GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length matrix; GLSZM, gray-level size zone matrix; iAUC, integrated area under the time-dependent ROC curve; IDH, isocitrate dehydrogenase; OS, overall survival; ROC, receiver operating characteristic

240 mm; slice thickness, 5 mm; and matrix size, 256×256 images. Additionally, 3-D T1C images (TR, 6.3 ms; TE, 3.1 ms) were acquired after the administration of a gadolinium-based contrast agent (0.1 ml/kg gadobutrol, Gadovist, Bayer Schering Pharma).

Owing to the difference and heterogeneity of image protocol of TCGA cohort including spatial resolution, T2 images from the institutional and TCGA cohorts were resampled to 2-D spatial resolution of $1 \times 1 \times 5$ mm. FLAIR and T1C images were registered to the resampled T2 images after skull stripping, using the FMRIB software library (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL). After image registration, image signal intensity was normalized using the WhiteStripe R package [16]. Regions of interest (ROIs) were drawn to segment the entire tumor (which was defined as a hyperintense area on FLAIR images), including edema, using semiautomatic methods from an open-source application (Medical Image Processing, Analysis and Visualization, available at https://mipav.cit.nih). The ROIs were drawn by a neuroradiologist (Y.S.C., with 9 years of experience) and confirmed by an independent reviewer (S.S.A., with 12 years of experience).

A total of 250 radiomic features, including shape ($n = 16$), first order ($n = 19$), gray-level co-occurrence matrix (GLCM; $n = 27$), gray-level run length matrix (GLRLM; $n = 16$), and gray-level size zone matrix (GLSZM; $n = 16$), were extracted from ROIs on T1C, T2, and FLAIR images using a Python-based open-source module (Pyradiomics 1.2.0, http://www.radiomics.io/pyradiomics.html), with the following parameter setting: "binWidth", 1; "force2D", True [17]. Details of the radiomic features are available at Supplementary Material S1 and https://pyradiomics.readthedocs.io/en/latest/features.html.

Statistical analysis and machine learning

Models to predict OS based on radiomic features, clinical profile, and *IDH* mutations status were trained using the random survival forest (RSF) method. RSF is an extension method of random forest that intended to analyze right-censored survival data, which uses log-rank test for node splitting instead of logistic regression used in random forest [18]. We chose RSF over commonly used Cox regression–based

methods for survival analysis, because RSF does not rely on restrictive assumptions such as proportional hazards, and automatically handles non-linear effects and interactions of hundreds of radiomic features [18]. The feature selection and training of RSF models to predict OS were conducted completely within the institutional data; the TCGA/TCIA data was used for external validation only. The overall workflow of RSF machine learning and other statistical analyses are shown in Fig. 2. All statistical analyses were performed using R package (R version 3.2.4; R Foundation for Statistical Computing). A *p* value (*p*) < 0.05 was considered statistically significant.

To evaluate whether radiomic phenotyping improves survival prediction over models established from non-imaging prognostic factors, three RSF models were trained, as follows: (1) a radiomics RSF model based on radiomic features (model 1); (2) a clinical RSF model based on the non-imaging prognostic factors consisting of age, resection extent, WHO grade, and *IDH* status (model 2); and (3) a combined RSF model based on radiomic features and the non-imaging factors (model 3). For training of the radiomics RSF model (model 1), the 250 radiomic features were subjected to two steps of feature selection using the univariate log-rank test and minimal depth as follows: First, patients were dichotomized on the basis of each radiomic feature, using the optimal cutoff that yielded the most significant survival difference as calculated from the "cutp" function from the "survMisc" R package. Then, the OS from the dichotomized groups of the institutional cohort were compared using the log-rank test. Only those features with *p* values < 0.05 in the log-rank test were considered for the second step of feature selection using minimal depth. Minimal depth is a dimensionless statistic that measures the predictive ability of a variable in a survival tree of RSF, and has shown to be useful to select relevant features in high-dimensional problems using RSF [19]. The selected radiomic features based on the univariate log-rank test and minimal depth were used to train the radiomics RSF model (model 1). The minimal depth–based feature selection and RSF model training were performed using the "var.select" and "rfsrc" function from "randomForestSRC" of R package, respectively. For clinical RSF model (model 2) training, all the non-imaging prognostic factors consisting of age, resection extent, WHO grade, and *IDH* status were used without feature selection. Age was considered a continuous variable, while resection extent, WHO grade, and *IDH* status were considered categorical variables. For the combined RSF model (model 3), the selected radiomic features from model 1 and the non-imaging factors from model 2 were subjected to further feature selection based on minimal depth, where only radiomic features from model 1 were targeted for further feature selection based on minimal depth, and the non-imaging factors from model 2 were retained in the final version of model 3. This process was to remove the unhelpful radiomic features in

presence of the non-imaging factors and reduce the total number of features in model 3, while retaining the well-known non-imaging prognostic factors in the final model. As the RSF model training uses intrinsic randomization for selecting bagging samples and features, and prediction error from out of bag samples [18], cross-validation was not separately conducted for RSF model training.

After feature selection and RSF training of models 1–3 within the institutional cohort, models 1–3 were externally validated on TCGA cohort. To measure the performance of each RSF model from external validation, the OS predicted risks of models 1–3 from validation on TCGA cohort were entered into time-dependent receiver operating characteristic (ROC) analysis, and the integrated area under the ROC curve (iAUC) was calculated. To evaluate the incremental value of radiomic features in OS prediction, iAUC differences between model 2 (clinical RSF model) and model 3 (combined RSF model with radiomic features and non-imaging prognostic factors) were calculated with 1000 iterations of bootstrapped resampling. The iAUC difference was considered statistically significant if its 95% confidence interval (95% CI) did not contain a zero value. Instead of one *p* value, the iAUC differences and their 95% CI values were provided because it has been suggested that providing a bootstrapped CI is preferred whenever there is a cause to doubt the assumed underlying parametric CIs, such as CI for iAUC values [20, 21], and the interval estimation may be more informative than just one *p* value [22].

In addition, to evaluate whether the radiomics model (model 1) stratifies OS, patients were dichotomized on the basis of the optimal cutoff of the predicted risk in model 1 and OS was compared between these two dichotomized groups using the log-rank test. This optimal cutoff was determined to yield the most significant OS difference between the dichotomized groups and was calculated using the "cutp" function from the "survMisc" module of R package.

Results

Population characteristics

The clinical characteristics of the study cohort are summarized in Table 1. In the institutional cohort (total, *n* = 205; *IDH*-mutant gliomas, *n* = 136; *IDH* wild-type gliomas, *n* = 69), the mean age of patients was 44.6 ± 12.7 years, and the male-to-female ratio was 108:97. In TCGA cohort (total, *n* = 91; *IDH*-mutant gliomas, *n* = 71; *IDH* wild-type gliomas, *n* = 20), the mean age of patients was 46.8 ± 13.6 years, and the male-to-female ratio was 43:48. Patients from the institutional cohort showed significantly longer OS (median OS duration, 2986 vs. 1859 days; *p* = 0.008), a non-significant trend towards a lower proportion of *IDH* mutation (66.3% vs.

78.0%; $p = 0.059$), and a higher proportion of WHO grade II gliomas (59.5% vs. 46.2%; $p = 0.045$) compared with the TCGA cohort. The male-to-female ratio and resection extent were not significantly different between the institutional and TCGA cohorts ($p = 0.389$ and 0.189, respectively).

Improved survival prediction by radiomic phenotyping

For the radiomics model (model 1), 71 radiomic features were selected for predicting OS (Fig. 3 and Supplementary Material S2). The radiomics model (model 1) was successfully validated on the TCGA cohort with an iAUC of 0.620 (95% CI, 0.501–0.756). The predicted risk from model 1 significantly

stratified OS in the TCGA cohort (HR, 4.08; 95% CI, 1.40–11.91; log-rank $p = 0.007$), as well as in the institutional cohort (HR, 11.29; 95% CI, 4.44–28.67; log-rank $p < 0.001$; Fig. 3). The 71 selected radiomic features consisted of 9 shape features, and 25, 26, and 11 first-order or texture (i.e., GLCM, GLRLM, and GLSZM) features from T1C, T2, and FLAIR, respectively, among which the top 4 most relevant features consisted of shape features. The clinical model (model 2) with non-imaging prognostic factors yielded an iAUC of 0.627 (95% CI, 0.529–0.745) on validation on the TCGA cohort.

When radiomic features from model 1 were integrated into the clinical model (model 2), the newly established combined model 3 showed significantly improved OS prediction, as
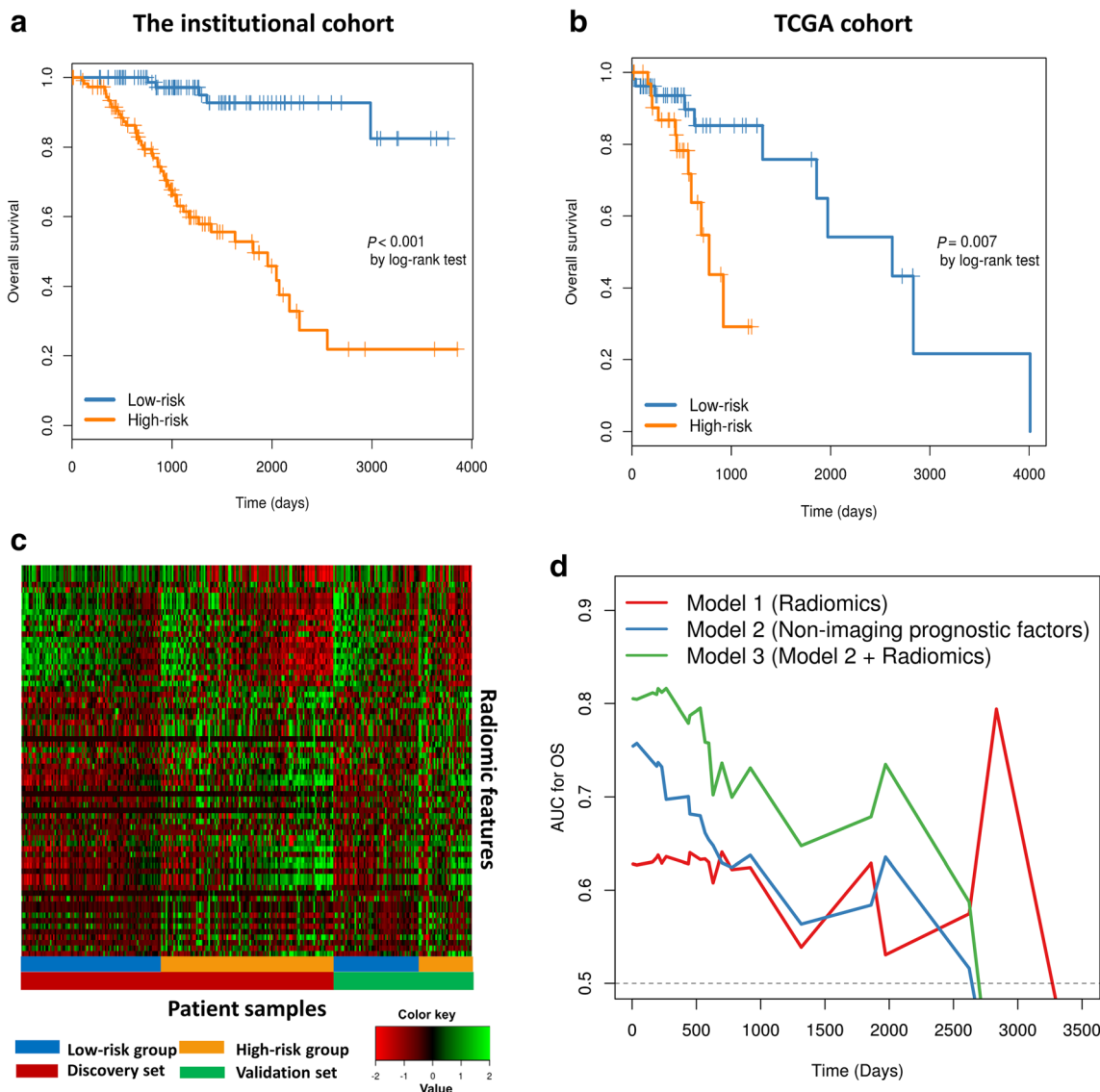
Fig. 3 Kaplan–Meier curves (**a**, **b**) and radiomic feature heat map (**c**) of the low-risk and high-risk groups stratified by a radiomics RSF model for predicting OS, and time-dependent ROC curve analysis (**d**) to show the incremental value of radiomic phenotyping over clinical profile and *IDH* status in the prediction of OS. IDH, isocitrate dehydrogenase; OS, overall survival; ROC, receiver operating characteristic; RSF, random survival forest

Table 2 The iAUCs to measure performances of RSF models in OS prediction

Model	Included parameters	iAUC (95% CI)	Difference in iAUC[a] (95% CI)
Model 1	Radiomic features	0.620 (0.501–0.756)	NA
Model 2	Non-imaging prognostic factors (age + resection extent + WHO grade + IDH status)	0.627 (0.529–0.745)	NA
Model 3	Non-imaging prognostic factors + radiomic features[b]	0.709 (0.623–0.843)	0.097 (0.003–0.209)

iAUC integrated area under the time-dependent ROC curve, *NA* not applicable

[a] Difference in iAUC caused by the addition of radiomics to model 2 with non-imaging prognostic factors, which was calculated from 1000 times bootstrapping

[b] The radiomics features from model 1 underwent further feature selection based on minimal depth, while other prognostic factors were kept in final model 3

compared with model 2 (Table 2 and Fig. 3; iAUC, 0.627 vs. 0.709; difference, 0.097; 95% CI, 0.003–0.209). In the further feature selection for model 3, 19 out of 71 radiomic features from model 1 were retained in the presence of the non-imaging prognostic factors. The 19 selected radiomic features consisted of 6 shape features, and 9, 3, and 1 first-order or texture features from T1C, T2, and FLAIR, respectively (Supplementary Material S3). Similar to model 1, the top 4 most relevant features out of 19 radiomic features were shape features. With regard to the non-imaging prognostic factors constituting model 3, the *IDH* status was the top important feature, and age, resection extent, and WHO grade were 8th, 16th, and 17th important features among total 23 features in model 3, respectively.

Discussion

We investigated the role of radiomic phenotyping to improve survival prediction for patients with LGG in presence of the non-imaging prognostic factors from molecular and clinical profiles. We performed external validation to test the model generalizability, and found that the accuracy of OS prediction was significantly improved by integrating radiomic features with the non-imaging prognostic factors.

Radiomic phenotyping has shown the potential to predict survival outcome of the patients with gliomas [13, 23–25]. However, many previous studies were based on a single institutional cohort [13, 23, 24]. As radiomic features are likely sensitive to differences in MRI protocols from different institutions, our study was conducted with external validation to assess the generalizability and robustness of the radiomic phenotyping models across heterogeneous MRI protocols and clinical composition of the patient population. Clinically, many patients with LGG undergo surgery, and survival outcome is predicted based on various non-imaging prognostic factors that was obtained before and after surgery, such as *IDH* mutation status, age, WHO grade, and extent of tumor resection. The *IDH* mutation is the most important prognostic biomarker in gliomas that recently integrated into the 2016 WHO

guidelines for classification of glioma [26], and is associated with better prognosis and response to chemoradiotherapy than *IDH* wild-type gliomas of the same histopathological grade [27]. Therefore, *IDH* status and other non-imaging prognostic information should be considered when investigating the role of imaging biomarkers in survival prediction. Previous studies [24, 28] reported that the radiomic features extracted from preoperative MRI can help improve survival prediction beyond molecular and clinical profiles in patients' glioblastoma. However, the role of radiomic features in LGG in consideration of non-imaging prognostic factors has not been well established yet; one study [25] established and externally validated a nomogram that was built upon clinicopathologic risk factors and a radiomic risk score from T2 images, to predict progression-free survival of LGG patients. However, the addition of radiomic risk score to the nomogram yielded relatively small improvement in survival prediction (C-index, 0.668 vs. 0.684 and 0.815 vs. 0.823 for the nomograms before and after the addition of radiomic risk score to the clinicopathologic risk factors, on the training and validation sets, respectively), and the significance of this improvement by radiomic risk score was not tested in this study [25]. On the contrary, our study focused on the incremental value of radiomic features within non-imaging prognostic factors, and allowed flexibility in radiomic feature selection in consideration of the non-imaging factors for model 3.

In the feature selection for models 1 and 3 of our study, shape features were constantly selected as the top relevant features. This result is in line with previous results that reported the association between tumor shape and prognosis [28–30]; one study [30] reported that tumor surface irregularity measuring how much the tumor surface deviates from a sphere of the sample volume, was a powerful predictor of survival outcome in glioblastoma patients. Another study [28] identified 18 radiomic features that predicted OS of glioblastoma patients, and 7 out of 18 features were shape features. Aside from shape features, it is also noteworthy that all the texture and first-order features that were retained in model 3 were from either T1C or T2, except one feature from FLAIR. This implies that FLAIR sequence may not be

mandatory to achieve compatible performance from the combination of radiomic features and non-imaging factors. In addition, given that 19 out of 71 radiomic features from model 1 were retained in model 3, our results imply that a subset of features from radiomic phenotyping might be independent from the non-imaging prognostic information, thus providing an added value for survival prediction. However, although better than random prediction, the performance of survival prediction by radiomic features alone (model 1) was not optimal probably owing to differences in MRI protocols and the clinical composition of the patient populations, including the different incidences of *IDH* mutation. In future investigations aimed at enhancing generalizability, the radiomic feature–based model should be trained on a more heterogeneous cohort that includes external cohort samples, or a prospective study can be conducted where protocols are specified, thereby mitigating potential differences in imaging protocols among institutions. As seen in the model with radiomic features, the model with non-imaging prognostic factors also showed suboptimal performance for OS prediction. These results suggest that compromised model performance from limited generalizability in different cohorts is a general issue across various non-imaging prognostic factors, as well as in radiomics. Nonetheless, our results indicate that the added prognostic value of radiomic features over non-imaging prognostic factors can be retained in an external cohort, despite the limited performance of individual models.

Our study has several limitations. First, the number of clinical prognostic factors was limited in the TCGA dataset, so that other important clinical factors such as Karnofsky performance status and postoperative treatment were not considered in the clinical RSF model; therefore, this clinical RSF model might not thoroughly reflect the performance of more comprehensive clinical factors in real-world clinical settings. Future investigations are necessary to confirm the added prognostic value of radiomic features when more comprehensive clinical prognostic information with uncompensated prognostic values are provided within a large and homogeneous cohort. Second, due to the heterogeneity of MRI protocols, 2-D spatial resolution (i.e., $1 \times 1 \times 5$ mm) was used for image resampling, which is below the resolution of 3-D isotropic T1C from current consensus recommendations [31]. This is because a considerable number of samples from TCGA data had 2-D T1C images, and we thought that resampling TCGA set images to a higher resolution than the original resolution may synthesize non-informative pixels, which may negatively affect the model performance from external validation on the TCGA set. The performance of the radiomics model may vary depending on MRI protocol, especially spatial resolution of the original images, which may limit the reproducibility of model performance. However, in our opinion, the strength of our study lies in that the potential

of radiomic phenotyping for improving survival prediction was demonstrated by resolving this heterogeneity of original images with postprocessing techniques.

Conclusions

We have shown that radiomic phenotyping using machine learning tools improves survival prediction in LGGs over clinical profile and *IDH* status. Our validation on a multi-institute TCGA cohort suggests that analytic tools developed using machine learning and radiomics analysis could be employed across different scanners and institutes.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Sung Soo Ahn.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry The statistical methodology of this study was reviewed by Kyunghwa Han, Yonsei University College of Medicine.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
• Retrospective
• Cross-sectional study
• Multi-center study

References

1. Hess KR, Broglio KR, Bondy ML (2004) Adult glioma incidence trends in the United States, 1977–2000. Cancer 101:2293–2299
2. Cancer Genome Atlas Research N, Brat DJ, Verhaak RG et al (2015) Comprehensive, integrative genomic analysis of diffuse lower-grade Gliomas. N Engl J Med 372:2481–2498
3. Pignatti F, van den Bent M, Curran D et al (2002) Prognostic factors for survival in adult patients with cerebral low-grade glioma. J Clin Oncol 20:2076–2084
4. Schomas DA, Laack NN, Rao RD et al (2009) Intracranial low-grade gliomas in adults: 30-year experience with long-term follow-up at Mayo Clinic. Neuro-Oncol 11:437–445

5. Zhao S, Lin Y, Xu W et al (2009) Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1α. Science 324:261–265

6. Yan H, Parsons DW, Jin G et al (2009) IDH1 and IDH2 mutations in gliomas. N Engl J Med 360:765–773

7. Parsons DW, Jones S, Zhang X et al (2008) An integrated genomic analysis of human glioblastoma multiforme. Science 321:1807–1812

8. Wang Y, Wang K, Wang J et al (2016) Identifying the association between contrast enhancement pattern, surgical resection, and prognosis in anaplastic glioma patients. Neuroradiology 58:367–374

9. Pope WB, Sayre J, Perlina A, Villablanca JP, Mischel PS, Cloughesy TF (2005) MR imaging correlates of survival in patients with high-grade Gliomas. AJNR Am J Neuroradiol 26:2466–2474

10. Wang YY, Wang K, Li SW et al (2015) Patterns of tumor contrast enhancement predict the prognosis of anaplastic Gliomas with IDH1 mutation. Am J Neuroradiol 36:2023–2029

11. Gillies RJ, Kinahan PE, Hricak H (2015) Radiomics: images are more than pictures, they are data. Radiology 278:563–577

12. Kickingereder P, Burth S, Wick A et al (2016) Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. Radiology 280:880–889

13. Macyszyn L, Akbari H, Pisapia JM et al (2015) Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. Neuro-oncology 18:417–425

14. Zhang B, Chang K, Ramkissoon S et al (2017) Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. Neuro-Oncology 19:109–117

15. Ceccarelli M, Barthel FP, Malta TM et al (2016) Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse Glioma. Cell 164:550–563

16. Shinohara RT, Sweeney EM, Goldsmith J et al (2014) Statistical normalization techniques for magnetic resonance imaging. Neuroimage Clin 6:9–19

17. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational Radiomics system to decode the radiographic phenotype. Cancer Res 77:e104–e107

18. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS (2008) Random survival forests. Ann Appl Stat 2:841–860

19. Ishwaran H, Kogalur UB, Chen X, Minn AJ (2011) Random survival forests for highdimensional data. Stat Anal Data Min 4:115–132

20. Carpenter J, Bithell J (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat Med 19:1141–1164

21. Heagerty PJ, Zheng Y (2005) Survival model predictive accuracy and ROC curves. Biometrics 61:92–105

22. Wasserstein RL, Lazar NA (2016) The ASA's statement on p-values: context, process, and purpose. Am Stat 70:129–133

23. Zhou H, Vallieres M, Bai HX et al (2017) MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro-Oncology 19:862–870

24. Kickingereder P, Neuberger U, Bonekamp D et al (2017) Radiomic subtyping improves disease stratification beyond key molecular, clinical, and standard imaging characteristics in patients with glioblastoma. Neuro-oncology 20:848–857

25. Liu X, Li Y, Qian Z et al (2018) A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas. NeuroImage: Clinical 20:1070–1077

26. Louis DN, Perry A, Reifenberger G et al (2016) The 2016 World Health Organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 131:803–820

27. Yan H, Parsons DW, Jin G et al (2009) IDH1 and IDH2 mutations in gliomas. N Engl J Med 360:765–773

28. Bae S, Choi YS, Ahn SS et al (2018) Radiomic MRI phenotyping of glioblastoma: improving survival prediction. Radiology 289:797–806

29. Chaddad A, Desrosiers C, Hassan L, Tanougast C (2016) A quantitative study of shape descriptors from glioblastoma multiforme phenotypes for predicting survival outcome. Br J Radiol 89:20160575

30. Pérez-Beteta J, Molina-García D, Ortiz-Alhambra JA et al (2018) Tumor surface regularity at MR imaging predicts survival and response to surgery in patients with glioblastoma. Radiology 288:218–225

31. Ellingson BM, Lai A, Harris RJ et al (2013) Probabilistic radiographic atlas of glioblastoma phenotypes. Am J Neuroradiol 34:533–540