Radiotherapy and Oncology 160 (2021) 132–139

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Deciphering the glioblastoma phenotype by computed tomography radiomics

Inge Compter [a,*], Maikel Verduin [b], Zhenwei Shi [a], Henry C. Woodruff [c,d], Robert J. Smeenk [e], Tom Rozema [f], Ralph T.H. Leijenaar [g], René Monshouwer [e], Daniëlle B.P. Eekers [a], Ann Hoeben [b], Alida A. Postma [d], Andre Dekker [a], Dirk De Ruysscher [a], Philippe Lambin [c,d], Leonard Wee [a]

[a] Department of Radiation Oncology (Maastro), GROW School for Oncology, Maastricht University Medical Centre+; [b] Department of Medical Oncology, GROW-School for Oncology, Maastricht University Medical Center; [c] The D-Lab, Dept of Precision Medicine, GROW - School for Oncology, Maastricht University; [d] Dept. of Radiology and Nuclear Medicine, GROW-School for Oncology, Maastricht University Medical Center, Maastricht; [e] Dept. of Radiation Oncology, Radboud University Nijmegen Medical Centre, Nijmegen; [f] Verbeeten Institute, Dept. of Radiotherapy, Tilburg, The Netherlands; [g] Oncoradiomics SA, Liège, Belgium

ARTICLE INFO

Article history:
Received 22 November 2020
Received in revised form 19 April 2021
Accepted 3 May 2021
Available online 10 May 2021

Keywords:
Glioblastoma
Radiomics
Computed tomography
Radiotherapy
Model development
Model validation

ABSTRACT

Introduction: Glioblastoma (GBM) is the most common malignant primary brain tumour which has, despite extensive treatment, a median overall survival of 15 months. Radiomics is the high-throughput extraction of large amounts of image features from radiographic images, which allows capturing the tumour phenotype in 3D and in a non-invasive way. In this study we assess the prognostic value of CT radiomics for overall survival in patients with a GBM.

Materials and methods: Clinical data and pre-treatment CT images were obtained from 218 patients diagnosed with a GBM via biopsy who underwent radiotherapy +/− temozolomide between 2004 and 2015 treated at three independent institutes (n = 93, 62 and 63). A clinical prognostic score (CPS), a simple radiomics model consisting of volume based score (VPS), a complex radiomics prognostic score (RPS) and a combined clinical and radiomics (C + R)PS model were developed. The population was divided into three risk groups for each prognostic score and respective Kaplan–Meier curves were generated.

Results: Patient characteristics were broadly comparable. Clinically significant differences were observed with regards to radiation dose, tumour volume and performance status between datasets. Image acquisition parameters differed between institutes. The cross-validated c-indices were moderately discriminative and for the CPS ranged from 0.63 to 0.65; the VPS c-indices ranged between 0.52 and 0.61; the RPS c-indices ranged from 0.57 to 0.64 and the combined clinical and radiomics model resulted in c-indices of 0.59–0.71.

Conclusion: In this study clinical and CT radiomics features were used to predict OS in GBM. Discrimination between low-, middle- and high-risk patients based on the combined clinical and radiomics model was comparable to previous MRI-based models.

Despite extensive treatment with surgery, radiotherapy, and concurrent and adjuvant chemotherapy, patients with glioblastoma (GBM) have a very poor prognosis with a median overall survival (mOS) of 15 months after diagnosis [1]. A small subgroup of patients (~5%) who survive for more than 3 years has been reported [2]. The wide range in OS underlines the need to estimate likely prognosis on the individual level in order to support personalized treatment.

Several clinical recursive partitioning analysis models (RPA) have previously been developed to compare glioma survival categories and obtain homogenous groups of patients to evaluate in clinical trials [3–6]. These models only include clinical parameters such as performance status (PS) and age. In order to refine these models, the glioma phenotype has received considerable attention, thus identification of different phenotypes may hold important prognostic and predictive information in addition to RPA.

Radiomics has emerged as a novel component of clinical decision support systems [7]. It refers to the automated extraction of large amounts of imaging features from radiographic images. The radiomics hypothesis is that image-derived features capture information that is otherwise not visible [8,9]. Previous radiomics studies have demonstrated that quantitative assessment of non-invasive biomarkers holds prognostic information in numerous

* Corresponding author at: MAASTRO Doctor Tanslaan, 12 6229 ET Maastricht, The Netherlands.
E-mail address: inge.compter@maastro.nl (I. Compter).

https://doi.org/10.1016/j.radonc.2021.05.002
0167-8140/

cancer types in addition to molecular and clinical characteristics [8,10,11].

Radiomics may improve clinical decision making in glioma in addition to the known RPA models as there is evidence that it can differentiate between tumour grade, identify druggable mutations, and assess tumour response [12–16]. Moreover, radiomics offers the opportunity to retrieve imaging biomarkers for clinical trials and can be obtained retrospectively. Currently, radiomics research in GBM has primarily focused on MRI. However, GBM patients regularly receive CT scans either at first presentation or as part of their work-up for radiotherapy treatment planning.

Translating radiomics-based treatment outcome models into routine clinical use requires objective assessment of its prognostic performance in independent datasets. However, the majority of radiomics studies to date are lacking either detailed external validation or direct access to the individual patient data [17,18]. The primary objective of this study was to examine the prognostic potential of clinical, tumour volume and CT-derived radiomics features for OS in adult patients with biopsy confirmed GBM, treated by radiotherapy with or without temozolomide.

Materials and methods

Study population

The study population was derived from routine care cases from the three participating institutions: MAASTRO Clinic, Radboud University Medical Centre (RadboudUMC) and the Verbeeten institute, all located in The Netherlands. Patient and treatment characteristics are summarized in Table 1. The inclusion criteria were: age at biopsy $\geq$ 18 years, pathologically confirmed GBM at diagnosis, CT imaging and RT structure set available for analysis and treated with radiotherapy between January 2004 and December 2014. Only patients without a tumour resection (i.e. biopsy only) were included to create a clinically homogenous group and allow for evaluation of the tumour still present on the CT images. Clinical parameters and survival intervals from start of radiotherapy were extracted from electronic patient records. Follow-up consisted of clinical review and imaging every three months until death.

Study design

This work has been approved by internal review boards of three participating institutions as a retrospective chart review-based observational study (*IRB/P0122*). The study is a Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Type 2b investigation [19].

We used internal-external cross-validation on multi-institutional data to investigate potential CT radiomics signature for OS in adult histopathologically biopsy confirmed GBM patients treated by radiotherapy. The "internal-external cross-validation" method of Steyerberg and Harrell offers a more rigorous validation than a purely randomized split into training and testing cohorts [20]. Due to the vast number of candidate features relative to the number of deaths, a step-by-step feature selection approach was used to reduce the number of features. We compared this "data-driven" radiomics model with (1) a knowledge-based clinical model consisting of well-known prognostic factors from an RPA model [6] and (2) a simple radiomics model consisting of only one feature – the primary tumour volume [21].

Image acquisition

Treatment planning simulation CTs acquired prior to radiotherapy were used for the analysis. All CT scanners were used entirely without modifications to standard clinical procedure. Image acquisition parameters are presented in supplementary Table 2.

Gross tumour volume identification

Each gross tumour volume (GTV) was manually delineated by experienced radiation oncologists (n = 9, all with at least 5- years of experience) using a radiotherapy planning MRI co-registered to the planning CT as reference image. The delineations were checked by a second experienced radiation oncologist. The planning MRI consisted of contrast-enhanced 1–3 mm slice T1-weighted images. The GTV was defined as the enhancing tumour in accordance with ESTRO-ACROP guidelines [22]. The delineation was performed using the participating institutes treatment planning systems (TPS) (XiO (version 4.0, updated to 4.51 in use in

Table 1
Patient characteristics.

	Centre 1 (n = 93)	Centre 2 (n = 62)	Centre 3 (n = 63)
Biological sex			
Female	37 (39.8%)	21 (33.9%) *(p = 0.56)*	26 (41.3%) *(p = 0.98)*
Age at biopsy			
Median (range)	63.9 (21–86)	60.8 (18–77) *(p = 0.37)*	63.6 (40–80) *(p = 0.98)*
Over 70 years of age	29 (31.2%)	12 (19.4%) *(p = 0.15)*	13 (20.6%) *(p = 0.20)*
WHO Performance			
0	29 (31.2%)	0 (0%)	0 (0%)
1	45 (48.4%)	19 (30.6%)	41 (65.1%)
2	16 (17.2%)	20 (32.2%)	21 (33.3%)
3 or 4	3 (3.2%)	3 (4.8%)	1 (1.6%)
(missing)	–	20	–
		(p < 0.01)	***(p < 0.01)***
Interval biopsy to RT			
Median (range) days	28 (0–86)	**33 (16–65) *(p = 0.017)**	**32 (5–402) *(p = 0.027)**
Concurrent temozolomide	59 (63.4%)	38 (61.3%) *(p = 0.92)*	31 (47.6%) *(p = 0.073)*
Tumor physical dose			
Median (range) in Gy	60 (2–60)	60 (0–60) *(p = 0.92)*	**45 (4–59.4) *(p < 0.01)***
Gross tumor volume			
Median (range) in cm^3	51.9 (0.5–194)	48.2 (1–222) *(p = 0.52)*	**31.9 (4–165) *(p = 0.01)***
Kaplan–Meier median (range) overall survival in days	224 (3–4221)	**162 (4–4055) *(p = 0.02)***	181 (32–1246) *(p = 0.25)*
Kaplan–Meier overall survival at 1 year	23.7%	17.9% *(p = 0.50)*	17.5% *(p = 0.47)*
(95% confidence interval)	(16.4–34.1)%	(10.2–31.4) %	(10.2–29.9)%

Table 2
Prognostic discrimination according to concordance index (c-index) with the 95% confidence interval provided between parentheses. Cross-validation is by the internal-external method quoted in the text and leaving one participating centre out in turn.

	Apparent c-index	Validation c-index	Optimism-corrected c-index
CPS model			
Train in 1 & 2, test in 3	0.652 (0.60–0.70)	0.648 (0.57–0.73)	–
Train in 1 & 3, test in 2	0.651 (0.60–0.70)	0.654 (0.56–0.74)	–
Train in 2 & 3, test in 1	0.659 (0.60–0.72)	0.633 (0.57–0.70)	–
Train in pooled data	0.657 (0.61–0.70)	–	0.651 (0.60–0.69)
VPS model			
Train in 1 & 2, test in 3	0.541 (0.49–0.60)	0.609 (0.53–0.69)	–
Train in 1 & 3, test in 2	0.567 (0.51–0.62)	0.515 (0.41–0.62)	–
Train in 2 & 3, test in 1	0.569 (0.50–0.63)	0.553 (0.49–0.62)	–
Train in pooled data	0.554 (0.49–0.62)	–	0.547 (0.48–0.61)
RPS model			
Train in 1 & 2, test in 3	0.659 (0.61–0.71)	0.598 (0.52–0.68)	–
Train in 1 & 3, test in 2	0.668 (0.62–0.71)	0.568 (0.48–0.66)	–
Train in 2 & 3, test in 1	0.641 (0.58–0.70)	0.639 (0.59–0.69)	–
Train in pooled data	0.655 (0.61–0.70)	–	0.643 (0.60–0.69)
(C + R)PS model			
Train in 1 & 2, test in 3	0.693 (0.65–0.74)	0.684 (0.62–0.75)	–
Train in 1 & 3, test in 2	0.703 (0.66–0.74)	0.589 (0.48–0.70)	–
Train in 2 & 3, test in 1	0.665 (0.61–0.72)	0.707 (0.66–0.76)	–
Train in pooled data	0.691 (0.65–0.73)	–	0.689 (0.65–0.73)

2014), Elekta, Stockholm, Sweden; Eclipse (version 10), Varian Medical Systems, California, USA; Pinnacle 3, Philips Medical Systems, Fitchburg, USA). The CT images and delineated structures were exported from each TPS in Digital Imaging and Communications in Medicine (DICOM) format.

Radiomics feature extraction

Radiomics features were extracted using an open-source Python library pyRadiomics (v2.1.2) [23]. Images were resampled to 1 mm isotropic voxels. A total of 1093 features were extracted. These consisted of 13 morphology (shape) features, 17 intensity-histogram (first-order) features and 73 textural which includes e.g., Haralick features (gray-level co-occurrence matrix (GLCM), grey level size zone matrix (GLSZM), grey level run length matrix (GLRLM), and neighborhood grey tone difference matrix (NGTDM). The intensity and textural features were re-computed after Laplacian of Gaussian (LoG) filter with three kernel dimensions – 1 mm (90 features), 2 mm (90 features) and 3 mm (90 features). The features were re-computed following wavelet (coif1) decomposition at 8 levels, for every possible combination of either a high or low pass filter in each of three cardinal axes (720 features). Morphology features in pyRadiomics are independent of pre-processing filters. Binary masks for the GTV were generated from DICOM structure sets using the O-RAW extension library to pyRadiomics to process DICOM inputs [24]. The mathematical definition of each feature can be found in the pyRadiomics documentation and the pyRadiomics extraction settings file is included in the Supplementary Materials (supplementary file 1). Divergences between PyRadiomics and the Image Biomarker Standardization Initiative (IBSI) have been documented by the developers [25].

Model development

A knowledge-based clinical prognostic score (CPS) was derived from a multivariable Cox regression of OS, using age in years and WHO PS as predictors. Cognitive function scores were not available in the present datasets [6]. Surgery type and tumour grade were not used because these were already applied as eligibility filters for the study. Categorical WHO-PS scores were re-coded in binary fashion as PS $\geq$ 1, PS $\geq$ 2, etc. Binary variables containing less than 10 occurrences were excluded.

Next, a GTV prognostic score (VPS) was evaluated using a single radiomics feature, tumour volume. This follows the recommendation of Welch et al [21]. to test accepted clinical factors that could yield similar prognostic performance as more complex signatures.

A radiomics prognostic score (RPS) was developed with a "data-driven" approach by sequentially reducing the number of candidate radiomics features within the training set. Each extracted radiomics feature was normalized by a Yeo-Johnson transformation [26] followed by centering (mean = 0) and scaling (sample standard deviation = 1) according to the training set. The same steps were applied in the validation set without any re-computation.

Finally, the CPS model and RPS model were combined in a clinical- and radiomics prognostic score ((C + R)PS) model.

Feature selection

Since the variables for the CPS were knowledge-based, and the VPS has one feature, no feature selection was applied to these models. A univariate association with OS was performed in order to confirm no important clinical features were missed.

As radiomics features are known to change significantly due to differences in scanner model, image acquisition or reconstruction settings, we used the post-reconstruction ComBat harmonisation method to harmonise features extracted from images acquired across the different institutes [27]. No test–retest to eliminate intensity-based and textural features with poor repeatability was performed as there was no GBM CT test–retest study available at time of writing.

The feature selection approach for the RPS is shown schematically in Fig. 1. One thousand unique bootstrap samples were drawn up from the entire cohort. In each bootstrap sample, the *findCorrelation* function in the R *caret* library [28] was used to minimize the number of pairwise feature correlations greater than 0.90 or less than −0.90; this removes the highly correlated features without looking at the survival outcome. Next, least absolute shrinkage (LASSO) [29] embedded with Cox regression was used to count how many times individual features were retained; this was done with 5-fold internal cross-validation with the default model tuning grid from R *glmnet* [30]. The top 10 features were arbitrarily retained, since the frequency of selected features decayed rapidly (Fig. 1a). With the top ten features, a stepwise backward Cox

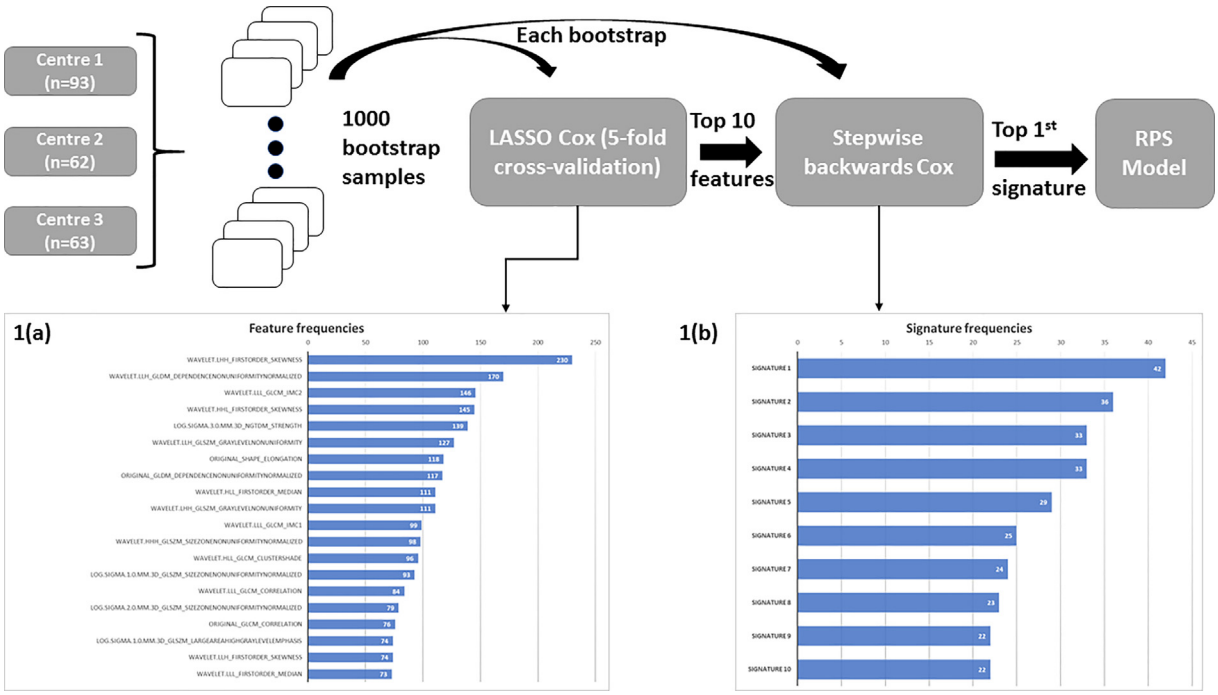

Fig. 1. (a-b). Feature selection approach for the RPS model.

regression was applied on the same 1000 bootstrap samples, and counted the frequency of unique radiomics signatures (combinations of one or more features) that were retained by the aforementioned stepwise regression. The top most frequent signature was arbitrarily selected as the RPS model, however there is no universal agreement on which signature to choose (Fig. 1b).

Model validation and risk grouping

The general procedure recommended in TRIPOD for model development and validation was applied [19]. Using data from the three centres the prognostic performance using the internal-external method, with each centre being left out in turn, was estimated. For prognostic signatures, the global coefficients using all three datasets combined were determined and an over-optimism correction by the bootstrapping method was applied [20]. The same 1000 bootstrap samples for feature selection were also used for optimism estimation. The Harrell concordance index (c-index) [31] was used as measure of discrimination.

For clinical relevance, prognostic scores were used to divide the population into three risk groups – the best prognosis 25%, the worst prognosis 25% and the remainder [32]. Survival curves for the risk group are presented as Kaplan–Meier (KM) plots.

Statistical analysis

All analyses were performed in R statistical software (V3.6.1, https://www.R-project.org/). For differences in patient characteristics, the appropriate two-sided hypothesis test was applied; standard Z-test for proportions of continuous variables, unpaired Wilcoxon test for median and range, and exact Fisher test for categorical variables. Univariate association with OS was used to confirm that important clinical features had not been overlooked. OS was analysed from commencement of radiotherapy until death or loss to follow-up (right censored) using the *coxph* function in the package *survival* [33] and c-indices were obtained from the *survcomp package* [34]. Survival outcomes were compared between risk groups using the Kaplan–Meier survival analysis (log-rank test with one-sided p value). In addition to the c-index and KM curves, we examined the distribution of prognostic scores generated by the model (namely, median and interquartile range) as well as regression of the prognostic scores themselves against OS as indication of the calibration slope, in accordance with Royston and Altman [32] to check for discrimination, calibration and model fit.

Results

Patient and treatment characteristics are summarized in Table 1. There were no statistically significant differences between the cohorts for gender, age at biopsy, or treatment. Significant and clinically relevant differences were observed for radiation dose, primary tumour volume and WHO-PS scores. The proportion of 1-year OS after the start of radiotherapy was not significantly different between cohorts. In univariate analysis, age, WHO-PS $\geq$ 1 and WHO-PS $\geq$ 2 were strongly associated with OS ($p < 0.005$ after Bonferroni correction). The treating clinic was marginally associated with OS (HR 1.2; 95% CI 0.99–1.3, $p = 0.0067$). Tumour location (e.g. frontal, parietal, occipital or temporal) was not strongly associated with OS (data not shown). There were no pairwise Spearman correlations stronger than ±0.33 between any of the features present in the RPS and CPS models.

Prognostic performance was quantified as c-indices for each of the four models, through internal-external cross-validation with each centre alternately left out, and with all subjects pooled together (Table 2). The CPS returned c-indices ranging from 0.63 to 0.65 in cross-validation, which was also consistent with a model trained on pooled subjects (apparent c-index 0.66, optimism-corrected c-index 0.65). The estimated 95% confidence interval (CI) was always above 0.50. The tumour volume model (VPS) was uniformly inferior to CPS with a c-index which fluctuated from 0.52 to 0.61 in cross-validation. The estimated 95% CI generally included 0.50. The radiomics model (RPS) resulted in an apparent c-index of 0.57 to 0.64 and an optimism-corrected c-index of 0.64. The (C + R)PS model returned c-

indices ranging from 0.59 to 0.71 in cross-validation. For both the RPS and (C + R)PS model the estimated 95% CI only included 0.50 when the model was trained in centres 1&3 and validated in centre 2.

The distribution of prognostic scores and calibration slopes of each model are shown in Supplementary Table 3. Similar discrimination performance and appropriateness of model fit of the CPS and RPS models was confirmed by the calibration slope and

Table 3
Model coefficients.

Model and variables	Cox regression coefficients	Hazard ratio (HR) estimate	95% confidence interval of HR	p-value
CPS (pooled)				
Age at biopsy	0.035	1.04	(1.02–1.05)	<0.001
WHO-PS ≥ 1	0.302	1.35	(0.88–2.08)	0.167
WHO-PS ≥ 2	0.510	1.66	(1.20–2.30)	0.002
VPS (pooled)				
Tumor volume	0.099	1.10	(0.96–1.27)	0.2
RPS (pooled)				
Wavelet-LHH: Intensity Skewness	0.263	1.30	(0.09–2.96)	0.003
Wavelet-LLH: GLDM DependenceNonUniformityNormalized	0.304	1.35	(0.11–2.85)	0.004
Wavelet-LLL: GLCM IMC2	0.291	1.34	(0.08–3.87)	<0.001
Wavelet-HHL: Intensity Skewness	0.396	1.49	(0.15–2.64)	0.008
Wavelet-LLH: GLSZM GrayLevelNonUniformity	0.261	1.30	(0.10–2.71)	0.006
Shape: Elongation	0.128	1.14	(0.08–1.57)	0.116
Wavelet-HLL: Intensity Median	−0.140	0.87	(0.07–1.98)	0.048

Abbreviations: WHO–World Health Organization; PS – Performance status; GLDM – Grey-level Difference Matrix; GLCM – Grey-level co-occurrence matrix; IMC – Information Measure of Content; GLSZM – Gray-level size zone matrix.

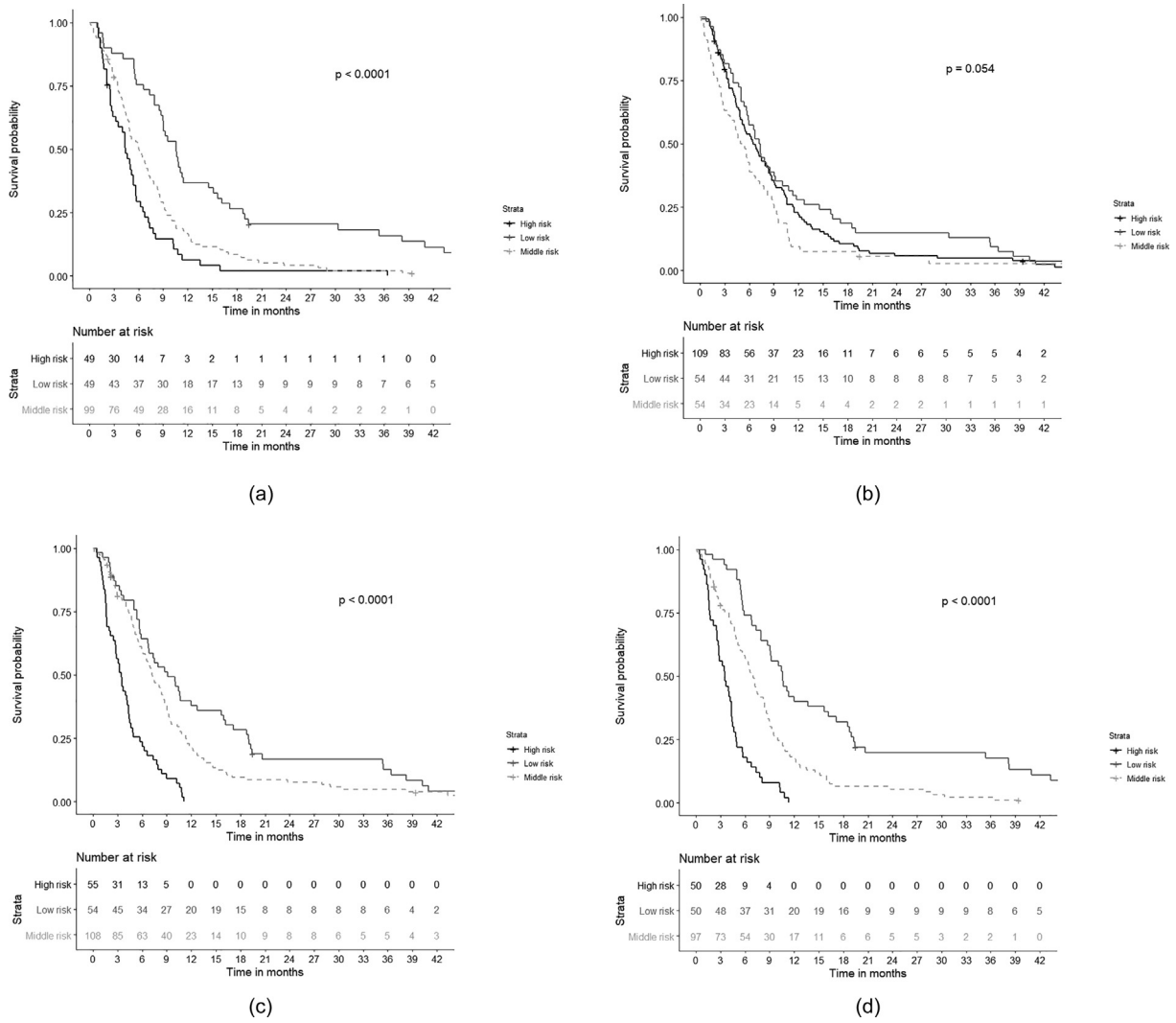

Fig. 2. (a-d). Kaplan–Meier analysis of overall survival of the complete study cohort. The population was divided into three risk groups for each prognostic score a) clinical prognostic score (CPS), b) volume based score (VPS), c) radiomics prognostic score (RPS), d) clinical and radiomics prognostic score (C + R)PS.

distribution of their prognostic indices. The median and interquartile range of prognostic scores are broadly overlapping in each of the three datasets, suggestive of adequate discrimination and calibration in each. For the CPS models the scores at centres 2 and 3 have less heterogeneity, as evidenced by smaller interquartile range, due to the lack of any WHO-PS 0 patients in those datasets. Poor performance of the VPS model was confirmed by large changes in the calibration between centres and error in the calibration slope, such that we cannot rule out the actual slope is zero and therefore overall lack of model fit. Inspection of Schoenfeld residuals of the calibration line shows no evidence of the proportional hazards assumption being violated in any combination of model and dataset.

Model coefficients, estimated hazard ratio (HR) with 95% confidence interval, and *p*-value for globally pooled models are shown in Table 3. Kaplan–Meier curves for the four different models (CPS, VPS, RPS, (C + R)PS) trained on pooled data are shown in Fig. 2.

Discussion

Identification of patients with a poor prognosis is of vital clinical importance as it may be questioned if it would be beneficial to offer these patient chemo-radiation, which may significantly decrease quality of life, or discuss best supportive care. This decision is currently primarily based on performance status, but this study demonstrates more factors, such as the GBM phenotype, should be considered. In the future these considerations should be implemented into decision support systems for GBM [35]. A steadily growing number of studies have been able to predict survival based on CT analysis in numerous types of cancer [8,36–39]. Skogen et al. have previously demonstrated that with a 2D CT texture analysis tumour heterogeneity correlated better with grade than size and attenuation did in glioma patients [40]. However, the present study is the first to investigate OS in patients with a GBM with CT-based radiomics. The CPS model based on age and WHO performance status and RPS model both demonstrated a moderate discriminative signature with an apparent c-index 0.66, and a combined c-index of 0.69, whereas the VPS only returned a c-index of 0.55. Fig. 2 demonstrates the discriminative power of the four models to separate the cohort into risk classes for survival. The assignment to the high-risk group according to the Cox models is associated with an increased Relative risk (RR) of mortality in the first 3 months compared to all the others (CPS: RR = 2.03; RPS: RR = 2.46; combined CPS + RPS: RR = 2.75). The first three months is an arbitrarily chosen timepoint, however it is clinically meaningful for decision-making as most physicians will be very hesitant to start treatment if the life-expectancy of the patient is three months or shorter.

Since discrimination is preserved in every validation for the CPS model, it may be possible to update the model with new clinical parameters such that the calibration is augmented. The lack of discriminative power for the VPS model is noteworthy as it is generally accepted that a larger tumour volume is correlated with a worse prognosis [41]. This indicates that other factors should be accounted for. The distribution of the PIs for the RPS model (supplementary Table 3) does not suggest there is a major problem with the radiomics features being defined differently or having an incompatible set of units. Potential reasons for the lack of fit are the considerable heterogeneity in clinical- and imaging parameters between the three investigated cohorts or the need to recalibrate the models specifically for the brain. The residuals of the regression of the PIs to the outcome of interest, i.e. a Schoenfeld test, did not indicate that the proportional hazards assumption had become untenable in any of the aforementioned datasets.

The historically available recursive partitioning analyses have shown their prognostic significance in the characterization of glioma subgroups with respect to survival [3–6]. These classifications were limited to tumour grade, age, PS, extent of surgery and mental status. Table 1 shows there is clinical heterogeneity present between the three datasets. Despite our best efforts, information on the WHO-PS was missing in a significant portion of centre 2 (32%). When comparing patient characteristics and OS between datasets we believe that there may be a selection bias present based on local neurosurgical clinical practice.

All of the CT scans used were acquired for simulation in routine image-guided radiotherapy planning (supplementary Table 2). However, the CT scanners and the imaging protocols used in this study differed in regards to the acquisition parameters and contrast enhancement. It has previously been demonstrated that both image quality and repeatability is highly dependent on imaging protocol and CT vendor, potentially influencing our results [42–44]. In order to minimize protocol differences additional CT scanner quality assurance by means of phantom studies may prove to be necessary for reproducibility [45]. The volumes-of-Interest required were manually delineated by the treating physicians based on the ESTRO guideline [22]. However, inter-observer variability cannot be excluded [46]. Automated segmentation based on HU for CT or artificial neural-network based methods for MR, could aid in reducing inter- and intra-observer variability and may also significantly influence feature reproducibility, but remains to be investigated. There are, to date, neither test–retest, multiple scanner series of the same subject nor inter-observer delineation CT datasets specifically for GBM that might otherwise have been used during feature selection.

Although MRI is the preferred imaging modality for characterizing GBM in clinical practice, this study has focussed on re-using CT scans acquired for radiotherapy treatment planning. A major advantage of CT is that the grey-level values are expressed in HU which can be correlated with tissue density. For the delineation of GBM a fusion of CT images with MRI is preferred due to high resolution images and enhanced contrast within soft tissues [47]. However, unlike CT, MR gray-level values in regularly acquired sequences do not always directly correlate to the physical properties of tissue, but require a pre-processing normalization step [48]. This renders standard MRI sequences such as T1, T2 and Fluid-attenuated inversion recovery (FLAIR) unsuitable for radiomics analysis without pre-processing, resulting in a potential loss of information. Methods to reduce these issues include intensity normalization (e.g. gaussian and Z-score normalisation) and voxel reslicing [49–51]. So far, several studies have investigated normalized MRI images by means of radiomics and demonstrated c-indices between 0.62 and 0.85 for multivariate models with either MR features alone or combined with clinical features such as age and performance status or molecular markers such as MGMT [52–60]. The combined (C + R)PS model presented in this study returned an apparent c-index of 0.69, which is comparable to the studies evaluating radiomics on MRI. More advanced MRI techniques such as Diffusion Weighted Imaging with Apparent Diffusion Coefficient maps, Diffusion Tensor Imaging, do directly allow for the assessment of the physiological properties of tissue and correlation with survival has been demonstrated [61–63].

This present study has several limitations. All three cohorts included retrospectively collected data, in which the risk of aforementioned selection bias cannot be excluded. Moreover, not all clinical factors could be retrieved in all three datasets. As images were collected over a longer time period differences in image quality may exist, in order to retain sample size all available images were included. A larger prospectively collected dataset, including standardized homogenized imaging protocols, may be required in order to improve the performance of radiomics models. Decentral-

ized methods for model development and model validation, where individual patient-level does not need to leave the institution, will be a boon to future radiomics modelling studies [64,65]. Known prognostic/predictive molecular markers such as *MGMT* (O-6-methylguanine-DNA methyltransferase) promotor methylation and *IDH1* (isocitrate dehydrogenase 1) mutation were not available in all patients. A cohort membership model has been used to distinguish external validation that tests for replication or transferability in models for dichotomous outcomes, but there is presently no analogous test procedure for survival time analysis [66,67]. Our study excluded patients with a tumour debulking in order to create a homogeneous patient cohort and investigate the entire tumour volume at time of the planning CT. Confounding effects due to blood remnants after the biopsy or an open vs stereotactic biopsy may have influenced the results.

A major advantage of radiomics is that it is capable of evaluating the entire tumour volume, whereas biopsies could result in sampling errors as genetic profiles from one section of the tumour may not accurately reflect its whole genetic profile. Presently the molecular status does not influence treatment choice in GBM. However, awaiting the results of studies investigating the efficacy of targeted anticancer drugs, radiomics may support personalized treatment for GBM. In order for radiomics-based models to be universally applicable new models should be validated on large and qualitatively high-level datasets. Making datasets publicly available through websites such as Cancerdata (https://www.cancerdata.org/), the radiomics Imaging Archive (www.radiomicsimagingarchive.info) or the Cancer imaging archive (https://www.cancerimagingarchive.net/) should be highly encouraged.

Conclusion

This study was a TRIPOD type 2b model development study in three independent datasets. Clinical and CT radiomics features were used to predict OS. Discrimination between low-, middle- and high-risk patients based on the combined clinical and radiomics model was comparable to previous MRI-based models. Future research should evaluate CT images of patients with a GBM longitudinally, compare CT features with MRI features, investigate a correlation between molecular markers and radiomics features, prospective collection of data and data sharing. Ultimately the objective of the radiomics field is to enhance clinical outcome by improved outcome prediction. This will most likely be achieved by combining all available information into one single model.

Role of funding source

The authors received no specific funding for this work

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2021.05.002.

References

[1] Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol 2009;10:459–66.
[2] Tykocki T, Eltayeb M. Ten-year survival in glioblastoma. A systematic review. J Clin Neurosci 2018;54:7–13.
[3] Curran WJ, Scott CB, Horton J, Nelson JS, Weinstein AS, Fischbach AJ, et al. Recursive partitioning analysis of prognostic factors in three Radiation Therapy Oncology Group malignant glioma trials. J Natl Cancer Inst 1993;85:704–10.
[4] Mirimanoff R-O, Gorlia T, Mason W, Van den Bent MJ, Kortmann R-D, Fisher B, et al. Radiotherapy and temozolomide for newly diagnosed glioblastoma: recursive partitioning analysis of the EORTC 26981/22981-NCIC CE3 phase III randomized trial. J Clin Oncol 2006;24:2563–9.
[5] Scott CB, Scarantino C, Urtasun R, Movsas B, Jones CU, Simpson JR, et al. Validation and predictive power of Radiation Therapy Oncology Group (RTOG) recursive partitioning analysis classes for malignant glioma patients: a report using RTOG 90–06. Int J Radiat Oncol Biol Phys 1998;40:51–5.
[6] Li J, Wang M, Won M, Shaw EG, Coughlin C, Curran WJ, et al. Validation and simplification of the Radiation Therapy Oncology Group recursive partitioning analysis classification for glioblastoma. Int J Radiat Oncol Biol Phys 2011;81:623–30.
[7] Lambin P, Zindler J, Vanneste BGL, De Voorde LV, Eekers D, Compter I, et al. Decision support systems for personalized and participative radiation oncology. Adv Drug Deliv Rev 2017;109:131–53.
[8] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
[9] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–77.
[10] Ou D, Blanchard P, Rosellini S, Levy A, Nguyen F, Leijenaar RTH, et al. Predictive and prognostic value of CT based radiomics signature in locally advanced head and neck cancers patients treated with concurrent chemoradiotherapy or bioradiotherapy and its added value to Human Papillomavirus status. Oral Oncol 2017;71:150–5.
[11] Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol 2015;5:272.
[12] Kickingereder P, Isensee F, Tursunova I, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. Lancet Oncol 2019:1474–5488.
[13] Tian Q, Yan LF, Zhang X-A-O, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. J Magn Reson Imaging 2018:1522–2586.
[14] Skogen K, Schulz A, Dormagen JB, Ganeshan B, Helseth E, Server A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. Eur J Radiol 2016;85:824–9.
[15] Bahrami N, Hartman SJ, Chang YH, et al. Molecular classification of patients with grade II/III glioma using quantitative MRI characteristics. J Neurooncol 2018:1573–7373.
[16] Rathore S, Akbari H, Rozycki M, Abdullah KG, Nasrallah MP, Binder ZA, et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. Sci, Rep 2018;8. https://doi.org/10.1038/s41598-018-22739-2.
[17] Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys 2018;102:1143–58.
[18] Yip SS, Aerts HJ. Applications and limitations of radiomics. Phys Med Biol 2016;61:R150–66.
[19] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Cancer 2015;112:251–9.
[20] Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245–7.
[21] Welch ML, McIntosh C, Haibe-Kains B, Milosevic MF, Wee L, Dekker A, et al. Vulnerabilities of radiomic signature development: the need for safeguards. Radiother Oncol 2019;130:2–9.
[22] Niyazi M, Brada M, Chalmers AJ, et al. ESTRO-ACROP guideline "target delineation of glioblastomas". Radiother, Oncol. 2016;Jan(1879-0887).
[23] van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017:1538–7445.
[24] Shi Z, Traverso A, Soest J, Dekker A, Wee L. Technical Note: Ontology-guided radiomics analysis workflow (O-RAW). Med Phys 2019;46:5677–84.
[25] Zwanenburg A-A-O, Vallières M-A-O, Abdalah M-A-O, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295(2):328–38.
[26] Yeo I-K, Johnson RA. A new family of power transformations to improve normality or symmetry. Biometrika 2000;87:954–9.
[27] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2006;8:118–27.
[28] Kuhn M. Caret package. J Stat Softw 2008;28:5.
[29] Tibshirani R. The lasso method for variable selection in the cox model. Stat Med 1997;16:385–95.
[30] Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw 2011;39:1–13.
[31] Harrell FE, Jr., Lee Kl Fau - Mark DB, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-387.
[32] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;13:33.
[33] M. TT, M. GP. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.
[34] Schroder MS, Culhane Ac Fau - Quackenbush J, Quackenbush J Fau - Haibe-Kains B, Haibe-Kains B. survcomp: an R/Bioconductor package for performance

assessment and comparison of survival models. Bioinformatics. 2011;27 (22):3206-3208.
[35] Walsh S, de Jong EEC, van Timmeren JE, Ibrahim A, Compter I, Peerlings J, et al. Decision support systems in oncology. JCO Clin Cancer Inform 2019:1–9.
[36] Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. PLoS ONE 2014;9:e102107.
[37] Larue R, Klaassen R, Jochems A, et al. Pre-treatment CT radiomics to predict 3-year overall survival following chemoradiotherapy of esophageal cancer. Acta, Oncol. 2018;(1651-226X).
[38] Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol. 2016;(1527-7755).
[39] Ding J, Xing Z, Jiang Z, Chen J, Pan L, Qiu J, et al. CT-based radiomic model predicts high grade of clear cell renal cell carcinoma. Eur J Radiol 2018;103:51–6.
[40] Skogen K, Ganeshan B, Good C, Critchley G, Miles K. Measurements of heterogeneity in gliomas on computed tomography relationship to tumour grade. J Neurooncol 2013;111:213–9.
[41] Ellingson BM, Abrey LE, Nelson SJ, et al. Validation of postoperative residual contrast-enhancing tumor volume as an independent prognostic factor for overall survival in newly diagnosed glioblastoma. Neuro Oncol. 2018;20:1240-1250.
[42] Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring computed tomography scanner variability of radiomics features. Invest Radiol 2015;50:757–65.
[43] Leijenaar RT, Carvalho S, Velazquez ER, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol. 2013;52(7):1391-1397.
[44] Clarke LP, Nordstrom RJ, Zhang H, Tandon P, Zhang Y, Redmond G, et al. The quantitative imaging network: NCI's historical perspective and planned goals. Transl Oncol 2014;7:1–4.
[45] Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017:1759–4782.
[46] Visser M, Müller DMJ, van Duijn RJM, Smits M, Verburg N, Hendriks EJ, et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. Neuroimage, Clin 2019;22:101727. https://doi.org/10.1016/j.nicl.2019.101727.
[47] Fiorentino A, Caivano R, Pedicini P, Fusco V. Clinical target volume definition for glioblastoma radiotherapy planning: magnetic resonance imaging and computed tomography. Clin Transl Oncol 2013;15:754–8.
[48] Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30:1234–48.
[49] Loizou CP, Petroudi S, Seimenis I, Pantziaris M, Pattichis CS. Quantitative texture analysis of brain white matter lesions derived from T2-weighted MR images in MS patients with clinically isolated syndrome. J Neuroradiol 2015;42:99–114.
[50] Depeursinge A, Foncubierta-Rodriguez A, Van De Ville D, Müller H. Three-dimensional solid texture analysis in biomedical imaging: review and opportunities. Med Image Anal 2014;18:176–96.
[51] Madabhushi A, Udupa JK. New methods of MR image intensity standardization via generalized scale. Med Phys 2006;33:3426–34.
[52] Xi YB, Guo F, Xu Z-L, Li C, Wei W, Tian P, et al. Radiomics signature: a potential biomarker for the prediction of MGMT promoter methylation in glioblastoma. J Magn Resonance Imaging: JMRI 2018;47:1380–7.
[53] Choi Y, Nam Y, Jang J, Shin N-Y, Lee YS, Ahn K-J, et al. Radiomics may increase the prognostic value for survival in glioblastoma patients when combined with conventional clinical and genetic prognostic models. Eur Radiol 2021;31:2084–93.
[54] Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings. Eur Radiol 2016.
[55] Chaddad A, Tanougast C. Extracted magnetic resonance texture features discriminate between phenotypes and are associated with overall survival in glioblastoma multiforme patients. Med Biol Eng Compu 2016;54:1707–18.
[56] Lao J, Chen Y, Li ZA-O, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Sci. Rep. 2017 (2045-2322).
[57] Chaddad A, Sabri S, Niazi T, Abdulkarim B. Prediction of survival with multi-scale radiomic analysis in glioblastoma patients. Med Biol Eng, Comput 2018;56:2287–300.
[58] Liu Y, Zhang X, Feng N, et al. The effect of glioblastoma heterogeneity on survival stratification: a multimodal MR imaging texture analysis. Acta Radiol 2018. 1600-0455.
[59] Verduin M, Primakov S, Compter I, et al. Prognostic and predictive value of integrated qualitative and quantitative magnetic resonance imaging analysis in glioblastoma. Cancers 2021;13:722.
[60] Chen X, Fang M, Dong D, Liu L, Xu X, Wei X, et al. Development and validation of a MRI-based radiomics prognostic classifier in patients with primary glioblastoma multiforme. Acad Radiol 2019;26:1292–300.
[61] Hu LS, Ning S, Eschbacher JM, et al. Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. Neuro Oncol. 2017;19(1):128-137.
[62] Kickingereder P, Götz M, Muschelli J, Wick A, Neuberger U, Shinohara RT, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. Clin Cancer Res 2016;22:5765–71.
[63] Abrol S, Kotrotsou A, Salem A, Zinn PO, Colen RR. Radiomic phenotyping in brain cancer to unravel hidden information in medical images. Top Magn Reson Imaging 2017;26:43–53.
[64] Bogowicz M, Jochems A, Deist TM, Tanadini-Lang S, Huang SH, Chan B, et al. Privacy-preserving distributed learning of radiomics to predict overall survival and HPV status in head and neck cancer. Sci Rep 2020;10. https://doi.org/10.1038/s41598-020-61297-4.
[65] Shi ZA-Ohoo, Zhovannik I, Traverso A, et al. Distributed radiomics as a signature validation study using the Personal Health Train infrastructure. Sci Data. 2019;6(1):218.
[66] Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015;68:279–89.
[67] van Soest J, Meldolesi E, van Stiphout R, Gatta R, Damiani A, Valentini V, et al. Prospective validation of pathologic complete response models in rectal cancer: transferability and reproducibility. Med Phys 2017;44:4961–7.