Neuro-Oncology Advances

7(1), vdaf154, 2025 | https://doi.org/10.1093/noajnl/vdaf154 | Advance Access date 3 August 2025

Estimating overall survival of glioblastoma patients using clinical variables, tumor size, and location

Alexandros Ferles, Paulina Majewska, Ragnhild Holden Helland, Ivar Kommers, André Pedersen, Mario Tranfa, Hilko Ardon, Lorenzo Bello, Mitchel S. Berger, Tora Dunås, Marco Conti Nibali, Julia Furtner, Shawn L. Hervey-Jumper, Albert J. S. Idema, Barbara Kiesel, Rishi Nandoe Tewarie, Emmanuel Mandonnet, Pierre A. Robe, Marco Rossi, Tommaso Sciortino, Tom Aalders, Michiel Wagemakers, Georg Widhalm, Aeilko H. Zwinderman, Lisa Millgård Sagberg, Asgeir Store Jakola, Erik Thurin, Ingerid Reinertsen, David Bouget, Ole Solheim, Roelant S. Eijgelaar, Philip C. de Witt Hamer and , Frederik Barkhof

All author affiliations are listed at the end of the article

Corresponding Author: Alexandros Ferles, Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, location VUmc, De Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands (a.ferles@amsterdamumc.nl).

Abstract

Background. Accurate prognosis of glioblastoma is crucial for better-informed treatment decisions, potentially leading to improved disease management. We investigated whether clinical variables, tumor size, and location, can serve as prognostic factors.

Methods. A retrospective, multicenter study enrolled 1318 adult patients with histopathologically confirmed glioblastoma undergoing first-time surgery, with survival censored for 188 patients. Pre-operative brain MRIs were used to compute tumor size and derive advanced radiological features describing tumor location, later refined by expert-based opinion. Post-operative MRIs were used to measure the enhancing residual tumor volume. The prognostic quality of all variables, measurements, and features was assessed as inputs of three survival regression models (CoxPH, Random Survival Forests, DeepSurv) to predict overall survival, under five timepoints of patient treatment: onset presentation, assessment by multidisciplinary board, intervention planning, post-intervention evaluation, and chemoradiotherapy planning. Model evaluation was performed with the C-index, Brier Score over Time, and Integrated Brier Score.

Results. Multivariable Cox analysis identified most clinical variables and tumor size as strong predictors of patient survival, with varying hazard ratios across timepoints. DeepSurv was consistently the top performing model under all possible inputs and at all timepoints, yielding mean test C-index scores ranging from 61.71% to 70.29%, and mean Integrated Brier Scores ranging from 8.57% to 7.63%.

Conclusion. Clinical variables, tumor size, and location carry prognostic value for the overall survival of patients with glioblastoma. The best predictive performance was observed under a Deep Survival model using all variables at the stage of chemoradiotherapy planning.

Key Points

- An automated pipeline for predicting overall survival regression in glioblastoma was developed.
- Clinical variables, tumor size, and expert-selected features were assessed as prognostic factors.
- Deep Survival network yielded the best performance.

Importance of the Study

Accurate prognosis of glioblastoma is critical for creating realistic expectations to patients, and guiding treatment choices, allowing for improved disease management. This study develops a computational framework for predicting the overall survival of patients with glioblastoma. By integrating clinical and imaging data, our approach provides personalized prognosis at any point during the disease course, from the onset to post-operative chemoradiotherapy planning. Additionally, we provide insights on the prognostic value of clinical variables, tumor size, and location. Overall, our study design employs a fully-automated pipeline, flexible in integrating different data sources at various timepoints of patient treatment, to support informed decision-making, possibly improving quality of life outcomes. This could potentially lead to a generalized tool for survival prediction in glioblastoma.

Glioblastoma is the most aggressive brain tumor,[1] associated with low median survival rates.[2] Despite treatment development, significant challenges persist. The need for an accurate prognosis in patients with glioblastoma is of critical importance, as it facilitates better-informed treatment decisions, and guides decision making, potentially leading to improved disease management. Additionally, variations in patient populations hinder cross-study comparisons, and a reliable prognostic model can help alleviate such challenges by providing a better understanding of patient individual profiles, thereby mitigating the effects of case-mix variability. Despite advances in statistical learning methods and medical image processing, which have assisted to significant improvement of patient outcomes in many diseases, the prognosis of glioblastoma at the individual level is challenging due to the high heterogeneity of the disease. Glioblastomas are characterized by high variability of patient survival times and a limited set of prognostic factors associated with the overall survival of patients.

Survival regression methods model survival probability functions across time periods and identify factors that carry prognostic value for patient outcomes. In addition, they incorporate censored patients in such analyses, where the exact time of survival remains unknown, which is particularly common in clinical studies. A few clinical variables were identified through survival regression as predictors for the overall survival of patients diagnosed with glioblastoma, including but not limited to the patient age[3] and Karnofsky Performance Status (KPS) score[4] at disease presentation, extent of surgical resection,[5] MGMT (O6-methylguanine-DNA methyltransferase) promoter methylation,[6] radiotherapy administration and chemotherapy status,[7] and patient's sex.[8] Furthermore, by incorporating volumetric and spatial data from Magnetic Resonance Imaging (MRI) scans, both the tumor size[9] and location[10] were highlighted as prognostic factors, with the presence of necrotic tumor core[11] and peritumoral edema[12] associated with lower survival rates.

Radiological features obtained through imaging studies were also employed to model the survival of patients with glioblastoma,[13] where machine learning techniques were applied in conjunction with such features, due to their ability to model interactions between the variables of the input space, and inherently select those more relevant to the survival regression task. Two prominent feature extraction approaches are the radiomics,[14] which represent general features derived from multi-modal MR images, and the Raidionics[15,16] software, formerly known as Glioblastoma Surgery Imaging—Reporting And Data System (GSI-RADS). Raidionics provides segmentation masks and automatic reports for glioblastomas, lower-grade gliomas, meningiomas, and metastases. The automatic reporting component of Raidionics describes the tumor location and measures tumor overlap with atlas-based regions of the brain. Recent studies focused on subset feature selection from the set of radiomics,[17] where the time of survival was modeled either as a discrete[18] (short-, mid-, and long-term survival) or continuous[19] outcome. Ranging from statistical regression[20] to deep-learning-based feature extraction,[21] radiomics seemed to offer additive predictive value, compared to approaches using clinical variables exclusively. Despite their broad use, which extends beyond survival regression, vulnerabilities of radiomics have been discussed in terms of limited reproducibility[22] and instability attributed to high inter-rater variability[23] across different tumor sites, therefore questioning the generalization ability of such methods. In contrast, associating the information represented by the features extracted from Raidionics with the prognosis of patient survival in glioblastoma remains unexplored.

In this study, we pose the following questions: *Can Raidionics features effectively model survival of patients diagnosed with glioblastoma?* and *What is the additive value of Raidionics to the simple use of recorded patient clinical variables?* Additionally, *How do state-of-the-art survival regression methods compare in estimating overall survival of glioblastoma patients?* To address these questions, we combined advanced radiological features extracted from Raidionics with automatic tumor volumetric measurements and patient clinical variables to form three sets of input variables, and trained independent survival regression models on five consecutive timepoints: initial disease presentation, assessment by multidisciplinary board, intervention planning, post-intervention assessment, and chemoradiotherapy planning. We performed cross-validation using a multicenter dataset of glioblastoma patients with multiparametric MRI and clinical follow-up and assessed all combinations of input variables, prediction models, and timepoints.

Materials and Methods

Data

This multi-center, retrospective study was conducted in accordance with the Declaration of Helsinki. The study protocol was approved by the Medical Ethics Review Committee of VU University Medical Center (IRB00002991, 2014.336). Written informed consent was obtained from patients as required for each participating hospital.

We identified 1615 patients of at least 18 years old with a newly diagnosed glioblastoma at first-time surgery (tumor resection or biopsy) between January 2012 and December 2018, with the majority of patients stemming from a previous study, and including patients treated in the years of 2012 and 2013[15] originating from 12 different sites in Europe and the US. The patient selection process was based on the availability of a pre-operative MRI (for all patients) and a post-operative MRI (for tumor resection patients) acquired within 72 hours after surgery. Clinical variables were recorded, and pre-treatment structural MR scans were acquired for all study participants. For those undergoing surgical tumor resection, post-operative structural MR scans acquired within 72 hours after surgery were additionally included. This time window aligns with the National Comprehensive Cancer Network (NCCN) recommendations, to distinguish between enhancing residual tumor and enhancement caused by post-surgical changes. Overall, imaging data were acquired using MR scanners from various manufacturers, including Philips, Siemens, GE, and Toshiba and with various field strengths, including 1T, 1.5T, and 3T. Although scan protocols were standardized within each center, such variations may introduce differences in image quality and contrast characteristics, thereby reflecting a real-world clinical setting, and enhancing the generalizability of potential findings to routine practice. Survival outcomes were measured from surgical intervention (tumor resection or biopsy) to death or last follow-up. The acquisition protocols for all participating centers have been previously described in detail.[15]

Study-specific exclusion criteria focused on the feasibility of volumetric tumor measurements in the pre-intervention and post-operative scans, where we deployed two segmentation networks. The pre-intervention segmentation network[24] requires the availability of at least the post-contrast T1-weighted (T1ce) scan, whereas the post-operative segmentation network requires the concurrent availability of both the pre-contrast T1-weighted (T1w) and T1ce scans.[25] We excluded from any subsequent analyses all patients for whom the pre-intervention T1ce scan was missing, and all tumor resection patients who had any of the required post-operative scans missing. Totally, this exclusion process resulted in the removal of $n = 297$ patients, resulting in a final cohort of $N = 1318$ patients for analysis.

Clinical Variables

Recorded clinical variables of study participants included the patient's sex, age, and KPS score at disease presentation and post-intervention, the intervention type (surgery or biopsy), and whether patients were treated with chemotherapy and/or radiotherapy. Almost all clinical variables were categorical, except for patient age, which was treated as a continuous variable. KPS scores were initially reported on a 10-point scale. However, a unit change in such scores is non-linear[26] and does not have the same effect on describing the patient status. As in previous studies,[27] we aggregated the KPS scores on a 3-point scale. KPS scores were stratified into three groups: less than 50, between 50 and 70, and more than 70. The remaining categorical variables were considered binary.

Not all variables had been recorded for all patients, due to several reasons such as differences in the imaging and clinical variables acquisition process in different centers. To compensate for missing observations, we performed feature imputation. Patient age was missing for just 5 patients (0.4% of the dataset) and followed a non-normal distribution, therefore, we applied median value imputation. For missing categorical variables, we deployed independent Random Forest classifiers—using three-class models for KPS scores, and binary models for the remaining categorical variables—fitted on each case on the subset of patients for whom all variables were recorded.

Volumetric Tumor Measurements

The final predictive models for survival were employed in a fully-automated fashion to reduce the time required to annotate tumor segmentation masks and eliminate rater variability. A recent study has also demonstrated that there is no statistically significant difference between using manual or automatic tumor volumetric measurements on glioblastoma survival regression.[27] Therefore, we employed two independent nnU-Net-based[28] segmentation networks, segmenting the pre-intervention and post-operative tumor.

The pre-operative automated segmentation network[24] was trained with a sparsified learning approach,[29] offering flexibility under missing sequences. Up to four sequences can be used: the T1ce, T1w, T2-weighted (T2w), and T2-Fluid-Attenuated-Inversion-Recovery (FLAIR). However, for pre-operative scans, the only inclusion criterion was the availability of the T1ce scan. Segmentation masks are provided for three classes: enhancing tumor, necrotic core, and peritumoral non-enhancing tissue. For optimal enhancing residual tumor segmentation performance, the post-operative segmentation network[25] requires the availability of both the T1ce and T1w scans.

Similar pre-processing protocols were applied for input images of both networks. Pre-operatively, N4 bias field correction was followed by registration to the SRI atlas,[30] using the T1ce scan as reference, and skull-stripping was applied using the HD-BET method.[31] Lastly, Z-score normalization was applied based on voxel values within the brain mask. Post-operatively, the same steps were applied with the exception of bias field correction. Predicted segmentation masks were resampled back to native space, where volumetric measurements took place. Volumetric residual tumor measurements could not be calculated for patients who received a biopsy, given the absence of post-operative MRI scans. For those patients, we reused

the pre-intervention tumor volume to align the data in the same format as tumor resection patients.

Advanced Radiological Features

Raidionics[15,16] is a pre- and post-operative central nervous system tumor segmentation and standardized reporting software. In glioblastoma, Raidionics has shown robust performance in predicting segmentation masks,[32] afterwards used to determine features such as the expected extent of resection, based on glioblastoma resection probability maps,[33] tumor laterality, multifocality, and location. Location is defined via overlap with brain regions based on several brain coordinates systems and parcellations, including the MNI space,[34] Yeo7 resting-state networks,[35] and the BCB atlas,[36] allowing for a detailed analysis of how the tumor interacts with these regions and its consequent impact on brain function. While Raidionics can extract tumor segmentation masks and afterwards employ them for standardized reporting, a segmentation mask can be provided to the software beforehand, and feature computation is based on the submitted mask instead. In our study, we extracted Raidionics features using the pre-intervention T1ce scan in native space as reference, and the predicted segmentation masks of the enhancing tumor and necrotic tumor core, by combining these compartments to a single mask representing the volume of the glioblastoma tumor core.

Timepoints of Patient Treatment

Predictive models will be applicable at any stage of patient treatment, from initial diagnosis to chemoradiotherapy planning after surgical intervention. Moving deeper into this timeline, extra information becomes available with regards to the recording of clinical variables and acquisition of MRI scans.

The following timepoints were indentified for survival analysis:

1. **First visit**: demographic characteristics of the patient (sex, age) along with the pre-treatment KPS score are recorded.
2. **Board meeting**: a multidisciplinary board discusses the status of the patient, using pre-intervention MRI scans.
3. **Surgical planning**: a decision is made on the type of intervention: resection or biopsy.
4. **Surgical outcome**: for patients undergoing tumor resection, a post-operative MRI scan for patients is acquired. Post-intervention KPS score is recorded for all patients.
5. **Chemoradiotherapy planning**: a decision is made on whether patients will be treated with adjuvant therapies (chemotherapy and/or radiotherapy).

Three sets of inputs were used at all timepoints:

1. **Clinical variables.**
2. **Volumetric measurements** of the pre- and post-intervention tumor.
3. **Advanced radiological features** computed by Raidionics.

We considered an incremental use of inputs. Clinical variables were used as a base model. When using volumetric measurements, we assumed the availability of clinical variables too. When using Raidionics features, both clinical variables and volumetric measurements were assumed available. Clinical variables of "First visit" and "Board meeting" overlap, as no new information of patients' characteristics becomes available when moving between them. What changes is the availability of pre-intervention MR imaging, which allows to compute corresponding tumor volumes and Raidionics features.

Survival Regression Models

We used the following survival regression methods:

1. Cox Proportional Hazards (CoxPH).
2. Random Survival Forests (RSFs).
3. Deep Survival network (DeepSurv).

The CoxPH model is a semi-parametric model widely used in survival analysis.[37] It models a hazard function h that assumes the effect of each covariate in the model's input vector over time is constant. RSFs[38] is a non-parametric, non-linear model, serving as an extension of Random Forests, where an ensemble of "survival trees" is created using bootstrap and splitting criteria adapted for survival data. Survival probabilities are aggregated from all trees to estimate overall survival. DeepSurv[39] is a deep-learning-based extension of CoxPH.

By design, all three selected methods inherently handle censored data. CoxPH uses a partial-likelihood function during optimization to order the recorded survival of participants, and account for censorship by including samples up to the timeline they were studied for, to generate a risk function over time. RSFs incorporate censored data by creating individual survival trees that include them up to censorship time, while excluding them by survival trees that examine later dates. By modifying the CoxPH partial-likelihood function, DeepSurv allows it to be incorporated in the optimization process of a multi-layer perceptron.

Figure 1 shows an overview of our survival regression pipelines.

Statistical Analysis

We employed two commonly used evaluation metrics in survival regression: the concordance index (C-index) and Brier Score over time. The C-index is a ranking-based metric, where assigned risk scores of patients are compared together, whereas the Brier Score at a given timepoint t measures the distance between the predicted survival probability and true survival of a patient. Brier Scores can be aggregated across the whole timeframe of the study, reported as the Integrated Brier Score (IBS). To evaluate the prognostic value of clinical variables, volumetric measurements, and radiological features, we conducted Cox multivariable regression analysis (calculating hazard ratios and respective P-values) using the *lifelines* package, and implemented all survival regression models using the

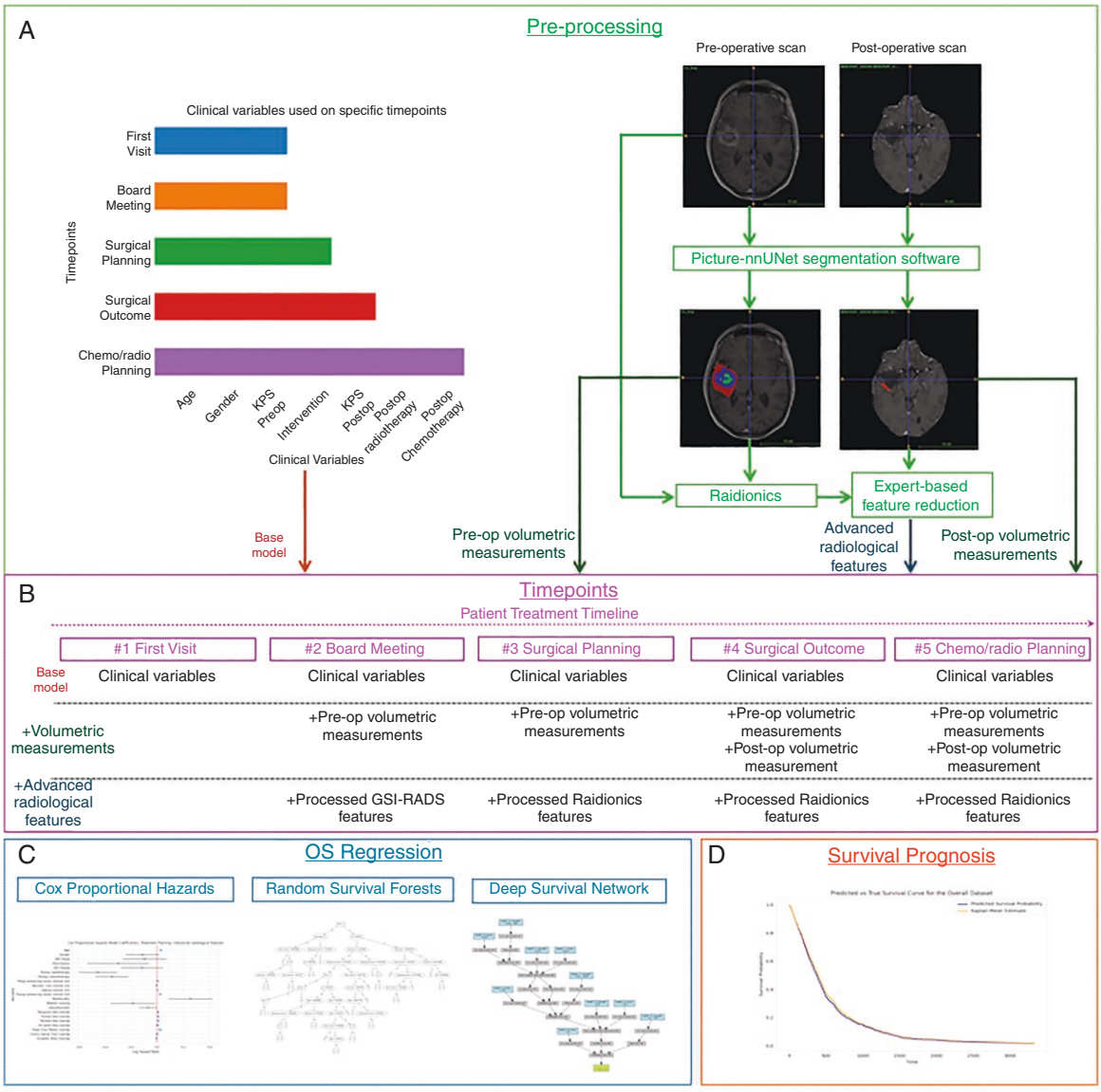

Figure 1. Survival regression pipeline overview. **Part A:** Pre-processing steps (green box) entail i) the selection of clinical variables depending on the timepoint under examination (left), ii) the automatic measurement of the pre and post-treatment tumor volumes with nnU-Nets (right), and iii) calculation of the Raidionics features on the pre-treatment scan using the predicted segmentation mask as reference followed by feature reduction based on expert criteria. **Part B:** We identify distinct yet incremental timepoints (purple box) of the patient treatment pipeline i.e. any information considered available at timepoint #1 is also considered available at timepoints #2-#5 etc. At each timepoint we apply three sets of inputs: a *base model* input including clinical variables only, a + *volumetric measurements* model input that incorporates the automatic measurements on top of the clinical variables, and a + *radiological features* model input that combines the clinical variables and volumetric measurements with the subset of selected Raidionics features, **Part C:** Three survival regression model (cyan box) are applied in a 5-fold cross-validation scheme to estimate the prognosis of patients with glioblastoma: The semi-parametric Cox proportional hazards model (CoxPH), the Random Survival Forests model (RSF), and the Deep Survival network (DeepSurv). Under these three model definitions, we display in respect the independent hazards of clinical variables, survival tree instance, and optimal network configuration decided via hyperparameter grid search for the Chemoradiotherapy Planning timepoint. **Part D:** Survival prognosis (orange box) on new unseen patients. We use the overall predicted survival curve of hold-out patients accumulated across all five folds which can be also seen in higher resolution in Figure 3, compared with the Kaplan-Meier estimate of the true survival for the same patients.

PySurvival package, both using Python 3.7. We also compared the differences in patient characteristics between the tumor resection and biopsy groups using Mann–Whitney *U* tests. Finally, we conducted Wilcoxon signed rank tests to examine whether there was a significant difference in predictive performance when using different sets of input features (for instance, when using advanced radiological features compared to using clinical variables and volumetric tumor measurements), irrespective of the chosen method or timepoint, across all repeats of cross-validation.

Table 1. Recorded clinical variables, and calculated measurements of study participants.

Category	Variables	Training/validation set (n = 1318)	# Missing observations n (%)
	Age (years), median (IQR)	63.7 (55-70)	5 (0.4%)
	Sex		3 (0.2%)
	Female n (%)	522 (39.6%)	
	Male n (%)	793 (60.2%)	
	KPS pre-op		366 (27.8%)
	>= 80 n (%)	624 (47.3%)	
	>=50, < 80 n (%)	316 (24%)	
	< 50 n (%)	12 (0.9%)	
	Intervention		0 (0%)
	Resection n (%)	1053 (79.9%)	
	Biopsy n (%)	265 (20.1%)	
	KPS post-op		162 (12.3%)
	>= 80 n (%)	684 (51.9%)	
	>=50, < 80 n (%)	407 (30.9%)	
	< 50 n (%)	65 (4.9%)	
	Post-op chemotherapy		61 (4.6%)
	Yes n (%)	697 (52.9%)	
	No n (%)	560 (42.5%)	
	Post-op radiotherapy		82 (6.2%)
	Yes n (%)	906 (68.7%)	
	No n (%)	330 (25.1%)	
	Pre-op peritumoral non-enhancing tissue volume (mL), median (IQR)	97.03 (51.88-146.83)	0 (0%)
	Pre-op necrotic core volume (mL), median (IQR)	8.58 (2.52-20.73)	0 (0%)
	Pre-op enhancing tumor volume (mL), median (IQR)	19.77 (10.21-33.04)	0 (0%)
	Post-op enhancing residual volume (mL), median (IQR)	2.15 (0.41-8.81)	265 (20.1%)
	Multifocality		0 (0%)
	Yes n (%)	239 (18.13%)	
	No n (%)	1079 (81.87%)	
	Midline crossing		0 (0%)
	Yes n (%)	359 (27.24%)	
	No n (%)	959 (72.76%)	
	Laterality Index mean ± SD	31.99 ± 34.04	0 (0%)
	Temporal lobe overlap (%) mean ± SD	32.08 ± 35.78	0 (0%)
	Frontal lobe overlap (%) mean ± SD	18.31 ± 27.22	0 (0%)
	Parietal lobe overlap (%) mean ± SD	2.89 ± 10.55	0 (0%)
	Occipital lobe overlap (%) mean ± SD	5.11 ± 8.79	0 (0%)
	Deep gray matter overlap (%) mean ± SD	6.58 ± 12.96	0 (0%)
	Cortico-spinal tract overlap (%) mean ± SD	49.88 ± 21.41	0 (0%)
	Yeo7 atlas overlap (%) mean ± SD	49.21 ± 39.53	0 (0%)
	Survival days, median (IQR)	365 (171-650)	188 (14.3%)
	Censorship		0 (0%)
	Yes n (%)	1130 (85.7%)	
	No n (%)	188 (14.3%)	

Abbreviations: KPS, Karnofsky Performance Status; SD, Standard Deviation; IQR, Interquartile Range.
The positive sex class is assigned to female patients. The positive class for intervention type is assigned to tumor resection.
Laterality Index may range from -1 to 1.

Predictive performance was compared both in terms of C-index and IBS(R1.10). We conducted extra Wilcoxon signed rank tests to examine whether there was a significant difference between the use of different models, irrespective of timepoint and input features, comparing CoxPH with RSF, CoxPH with DeepSurv, and CoxPH with DeepSurv, in terms of C-index performance. A significance level of 5% was used to assess statistical significance. More details on evaluation metrics and software implementation are provided in the Supplementary Material.

Advanced Radiological Features Dimensionality Reduction

In initial experiments, we observed that while using the raw set of the Raidionics features with CoxPH and RSFs, it was difficult to deliver performance comparable to the performance of sparser models (in terms of C-index or IBS) using only clinical variables and volumetric measurements. The same observation was made for DeepSurv. Theoretically, since DeepSurv is based on a multi-layer perceptron, hidden layers of appropriate size should model relationships between the raw input variables effectively, delivering a better estimate of patient survival. Performing a comprehensive hyperparameter tuning on the number of hidden layers and number of units per layer did not result in a DeepSurv instance that could outperform its counterparts trained without using Raidionics.

We thus considered dimensionality reduction of the input space, exploring Independent and Principal Component Analysis, and the "Maximum Relevance Minimum Redundancy" method. Lastly, we also asked for expert opinion in proposing features relevant with better or worse prognosis. Results showed that expert-selected features delivered the overall best performance in terms of C-index score, and thus opted for their use. Expert-selected features include the use of a tumor laterality index, ranging from –1 (left laterality) to + 1 (right laterality), and two binary indices describing tumor multifocality and midline crossing. In pre-intervention timepoints, the expected Resection Index in percentage was additionally selected, and was afterwards excluded in post-intervention timepoints, where the true measurement of the residual tumor volume becomes available. Additionally, seven measurements of tumor overlap with the (1) occipital, parietal, temporal, and frontal lobes (MNI space), (2) deep gray matter brain regions (MNI space), (3) cortico-spinal tract (BCB atlas), and (4) functional areas (Yeo7 atlas) were computed in percentage. In particular, we opted for the use of a single value, representing the total overlap of the tumor with such brain areas. To calculate overlaps, we aggregated the individual overlap values from Raidionics. For instance, for the total overlap between the tumor location and functional areas, we aggregated the overlap values of the tumor with each one of the seven resting-state networks defined by the Yeo7 atlas. Detailed statistics of these features can be found in Table 1 under 'Advanced radiological features.

Training Specifications for the Survival Regression Models

Given the absence of external datasets, we employed a repeated, stratified five-fold cross-validation, performed 5 times. Our stratification target was two-fold: for each train/test split, we ensured balanced ratios of censored/uncensored patients, and short-, mid-, and long-term survival patients.[40] Patients with survival up to 10 months were considered short-term survival patients, whereas patients who survived more than 15 months were considered long-term survival patients. Mid-term survival referred to patients with recorded survival between 10 and 15 months. While we focused on survival regression rather than classification, we used these categories in our stratification to maintain a distributed range of survival periods across the training data. Five train/test pairs were created per repeat, to simulate real-word data distributions.

For each experiment (unique combination of timepoint, inputs, and survival regression model), models were trained independently with each of the three input configurations, and afterwards assessed primarily by C-index performance in their respective testing data. Model instances yielding highest mean C-index performance are reported, afterwards assessed by IBS score performance in addition (Supplementary Tables 8 and 9).

Results

Patient Characteristics

Following application of the inclusion criteria, 1318 (81.3% of the original population) patients were selected for this study, out of whom 188 patients (14.3%) were censored. Median age of participants was 63.7 years (IQR 55–70) with median survival of 365 days (IQR 171–650). Median size of the pre-treatment tumor core (enhancing tumor and necrotic core) was 30.76 ml (IQR 14.88–54.17), whereas for patients receiving surgery the median enhancing residual tumor size was 2.15 ml (IQR 0.41–8.81). A subset of 603 patients were treated (45.8%) with both chemotherapy and radiotherapy. A total of 906 patients (68.7%) were treated with radiotherapy, with radiotherapy status missing for 61 (4.6%) patients of the dataset. A total of 697 patients (52.9%) were treated with chemotherapy, with chemotherapy status missing for 82 (6.2%) patients of the dataset. An additional sensitivity analysis (details given in the Supplementary Material) excluding one of the participating centers reporting a small percentage of patients receiving standard care showed the same performance trends as the ones reported for the original study. Detailed patient characteristics are presented in Table 1.

Tumor resection was performed on 1053 (79.9%) patients, whereas 265 (20.1%) patients received a biopsy. Patients in the resection group were younger than biopsy patients, with a median age of 63 compared to 66 years (IQR 54–69 vs. 57–73, $P < .05$). The resection group also had higher median pre-intervention tumor core volumes (32.68 mL vs. 24.74 mL, IQR 15.5–57.58 vs. 11.75–39.26, $P < .05$), and a significantly better survival prognosis, with a median survival of 411 days (IQR 228–657) compared to 158 days (IQR 66–284) for patients who received a biopsy ($P < .05$).

Prognostic Value of Individual Variables

We computed the hazard ratios HR of input variables separately at each timepoint, as the relative importance of variables may change over time, when more information becomes available.

Regarding clinical variables at "First Visit," older patient age at disease presentation was associated with shorter patient survival (HR = 1.03, 95% CI [1.02, 1.03], $P < .005$). Conversely, higher KPS scores (HR = 0.7, 95% CI [0.62, 0.79], $P < .005$) and female sex (HR = 0.85, 95% CI [0.75, 0.95], $P < .01$) were associated with longer survival prognosis. After pre-intervention volumetric tumor measurements were added at "Board Meeting," larger enhancing tumor volume (HR = 1.01, 95% CI [1.01, 1.02], $P < .005$) was associated with shorter survival, whereas larger necrotic core volume (HR =0.99, 95% CI [0.99, 1.00], $P < .005$) was associated with longer survival. Clinical variables retained the above-mentioned association, albeit with lower prognostic value, and additionally decreased statistical significance for patient sex. In "Surgical Planning," the type of intervention yielded the strongest prognostic value compared among all variables (HR = 0.36, 95% CI [0.31, 0.42], $P < .005$), demonstrating better prognosis for tumor resection patients. The remaining variables displayed lower prognostic values, while some displayed high *P*-values. This trend was subsequently observed across all timepoints. Post-intervention, larger enhancing residual volume (HR = 1.02, 95% CI [1.02, 1.03], $P < .005$) was associated with shorter survival, whereas higher post-treatment KPS score (HR = 0.78, 95% CI [0.67, 0.91], $P < .005$) was associated with longer survival. Lastly, at "Chemoradiotherapy Planning," undergoing chemotherapy (HR = 0.78, 95% CI [0.67, 0.91], $P < .005$), and radiotherapy (HR = 0.67, 95% CI [0.58, 0.77], $P < .005$) treatment were both associated with longer survival.

We examined hazard ratios of advanced radiological features when initially computed at the "Board Meeting" timepoint. Bigger overlap of the brain tumor with the functional networks defined by Yeo (HR = 0.99, 95% CI [0.99 0.99], $P < .005$) was associated with longer patient survival, whereas bigger overlap with the deep gray matter (HR = 1.02, 95% CI [1.00 1.04], $P = .01$) was associated with shorter patient survival and multifocal tumors were associated with shorter survival (HR = 1.49, 95% CI [1.27 1.75], $P < .005$). While for overlap-based variables such as the "Yeo7 atlas overlap" the changes in HR values may appear modest at first. Since overlap is treated as a continuous variable; such HR values reflect the effect per one-unit increase in overlap. Scaled to a larger change in overlap such as 10%, the corresponding HR values for the "Yeo7 atlas overlap" and "Deep gray matter overlap" become $0.99^{10} \approx 0.904$ and $1.02^{10} \approx 1.22$, indicating a 9.6% reduction and a 22% increase in hazard, respectively. Lastly, of the remaining features, a few displayed HR values equal to 1.00, whereas the Laterality and Resection indexes showed wide confidence intervals and high *p*-values, indicating non-statistically significant contribution. Supplementary Tables 2 to 7 provide detailed CoxPH model results for the prognostic value of all variables over timepoints.

Survival Regression Model Performance

Average C-index performance across all 36 experiments, repeated 5 times is summarized in Figure 2. DeepSurv outperformed both CoxPH ($P < .005$) and RSFs ($P < .005$) across all experiments, indicating that models more capable of capturing non-linear interactions between variables of the input space, can deliver better predictive performance. However, this difference was higher at the earliest timepoint, and reduced at following timepoints. Given that the set of radiological features was the same, the significance of the clinical variables to the prediction task is highlighted. Specifically, the simple addition of the "Intervention" variable at "Surgical Planning," increased the performance of both CoxPH and RSFs by more than 3%, reducing the performance gap between these models and DeepSurv. Post-intervention, adding the enhancing residual tumor measurement, and, later on, chemoradiotherapy status further added predictive power in the survival regression models.

In addition, we examined whether there was statistically significant difference in C-index performance, between using the Raidionics features on top of clinical variables and volumetric tumor measurements. We collected results across all repeats of cross-validation, timepoint, and model selection, generating 300 unique values in total for C-index and IBS for each input selection of *Clinical variables, + Volumetric measurements,* and + *Advanced radiological features.* We then performed Wilcoxon signed rank tests, comparing C-index and IBS results between *Clinical variables* and + *Volumetric measurements, Clinical variables* and + *Advanced radiological features,* and between + *Volumetric measurements* and + *Advanced radiological features.* The results of all rank tests showed statistically significant improvement when using volumetric measurements in addition to clinical variables ($P < .005$ for both C-index and IBS comparison), when using Raidionics features and volumetric measurements in addition to clinical variables ($P < .005$ for both C-index and IBS comparison), and when using Raidionics features in addition to volumetric measurements and clinical variables ($P < .005$ for both C-index and IBS comparison).

Quality of Predicted Survival Probability Functions

Figure 3 presents three panels describing the quality of the predicted survival probability functions generated from five independent DeepSurv networks trained during "Chemoradiotherapy Planning," using advanced radiological features. The mean survival probability with a 95% confidence interval of the predictions across the five test datasets is displayed, along with a comparison between predicted probabilities and the Kaplan–Meier estimate of the true survival probability function in the top row. Bottom panel displays the Brier Score over Time of the predictions aggregated on the five test folds, with each point representing the mean Brier Score at that time unit across each fold. By survival probability definition, predicted and ground-truth probabilities are equal to 1 on day 0, which is the beginning of the study. As time progresses, survival probabilities begin to diverge, leading to non-zero Brier Scores that capture the model's prediction error over time. The narrow width of the confidence intervals indicates low uncertainty/high confidence of model predictions. This uncertainty increases slightly midway through the study, which can be clearly observed for the same timeframe in

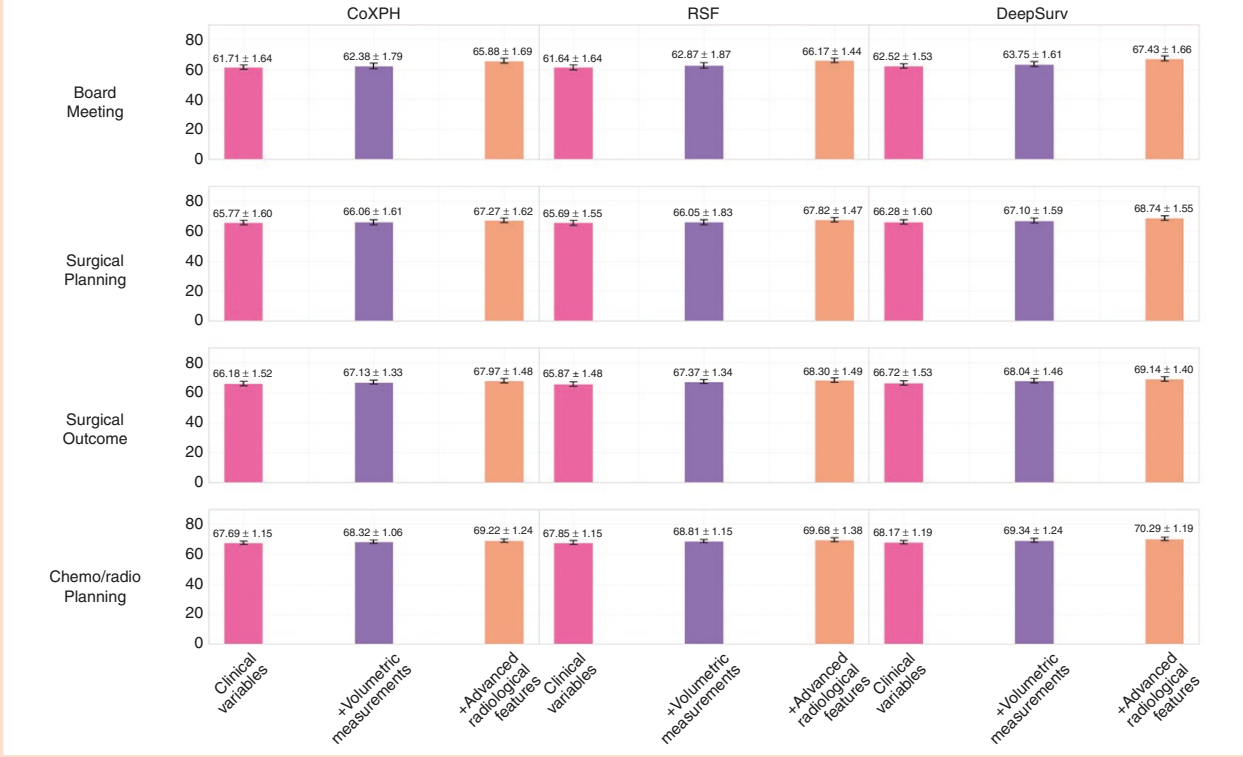

Figure 2. C-index performance for all experiments. Each row represents a unique timepoint, starting from'Board Meeting' at the top row which occurs earlier in a patient's treatment timeline and moving further in time till the last timepoint, i.e.'Chemoradiotherapy Planning' at the bottom. At each timepoint, three barplots of distinct colors are presented per survival regression model: Clinical variables (magenta), + Volumetric measurements (purple), and + Radiological features (orange). From left to right, we display the performance of the CoxPH, RSF, and DeepSurv models. Exact performance with two-digit precision, is displayed at the top of each bar.

Brier scores, where the distance between the predicted and true survival probability functions peaked slightly above 0.2(20%). This timeframe encompasses the majority of our participants, whereas in later days Brier scores decrease, indicating better survival prognosis for remaining patients. Overall, IBS performance (7.63%) indicates good predictive power of the DeepSurv networks. The overlap between the estimate of the true survival probability and the predicted survival probability functions shows that the models captured the general survival trends of the dataset well, further backing this observation. Examples of predicted survival probability functions for individual participants of the study can be seen in Figure 4, for "Board Meeting" and all inputs; one patient per survival category is included.

Discussion

The present study examined predictors of the overall survival of patients diagnosed with glioblastoma, using a combination of clinical variables and measurements of tumor volume and location in the brain at various patient treatment timepoints. A multi-center dataset was utilized, along with three commonly employed survival regression methods, and expert opinion was employed to propose a set of tumor location-based features, possibly associated with better or worse prognosis. This information was used in addition to the clinical variables and tumor size, where the additional prognostic value of these features was investigated.

The main findings of our study show that Raidionics features enhance the performance of predictive models, irrespective of the chosen method. Experiments incorporating advanced radiological features in model inputs consistently yielded the best performance at all timepoints. This effect was stronger at earlier stages of patient treatment, where three clinical variables with high prognostic value, that is, the intervention type, and post-operative chemotherapy and radiotherapy status, were unavailable. This suggests that clinical variables are extremely relevant to patient overall survival with glioblastoma, accounting for approximately 93% to 97% of the best predictive performance across timepoints, which is further backed by computed hazard ratios from multivariable Cox regression analysis. Thus, the sole use of clinical variables can prognosticate patient survival reasonably well in the absence of imaging-related information. However, adding tumor size and location leads to more accurate estimates of overall survival rates. Results also suggest that proper feature selection is more important than model selection, considering that models trained with expert-selected features yielded better prognostic performance than those trained with the raw Raidionics features.

Expanding on the findings of the Cox multivariable analysis, most computed values for the hazard ratios of

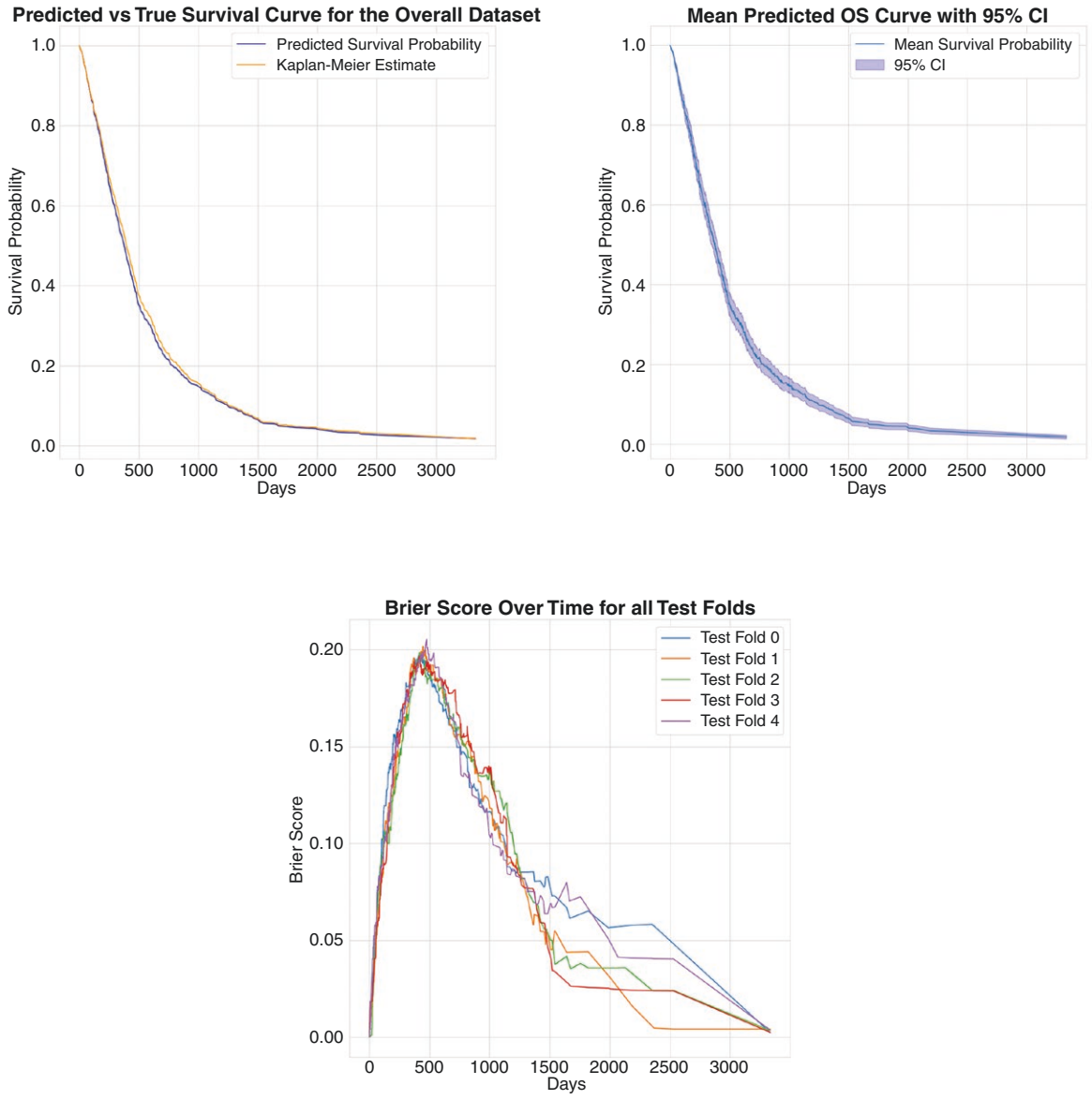

Figure 3. Quality of predicted survival probabilities on test patients. **Top row** displays the mean predicted survival curve across the five test folds, with the 95% confidence interval overlayed (left), and a comparison between the mean predicted survival curve and the Kaplan-Meier estimate of the mean recorded survival of the patients with glioblastoma. **Bottom row** displays the mean Brier score as time progresses during patient treatment on each individual test fold.

variables and features across timepoints, align well with expectations. Our results confirm previous findings that larger pre-intervention[9] and post-intervention[41] tumor volumes, multifocal tumors,[4] older patient age,[3] and higher tumor overlap with functionally relevant brain areas like the frontal lobe[42] and deep gray matter[43] are linked with shorter survival. Conversely, our results demonstrate that higher KPS scores (before[4] and after[44] intervention) and aggressive treatments (tumor resection[6] and/or chemotherapy,[6] and/or radiotherapy[45]) are associated with longer survival. However, two computed values stand out as they show counter-intuitive association with longer survival: larger size of the pre-intervention tumor necrotic core (despite the established link between higher necrosis to tumor volume ratio and worse prognosis[11,46]) and higher overlap with functional areas of the brain as defined by the total overlap with Yeo's resting-state networks. For the latter, brain tumors affecting these areas can potentially lead to earlier diagnosis due to patients seeking medical attention sooner, given earlier expression of symptoms. Additionally, treatment of tumors with higher overlaps with such areas might be more aggressive compared to the rest. We investigated this further by associating such higher overlap with smaller residual volumes and more patients being treated with chemotherapy and radiotherapy (additional details are provided in the

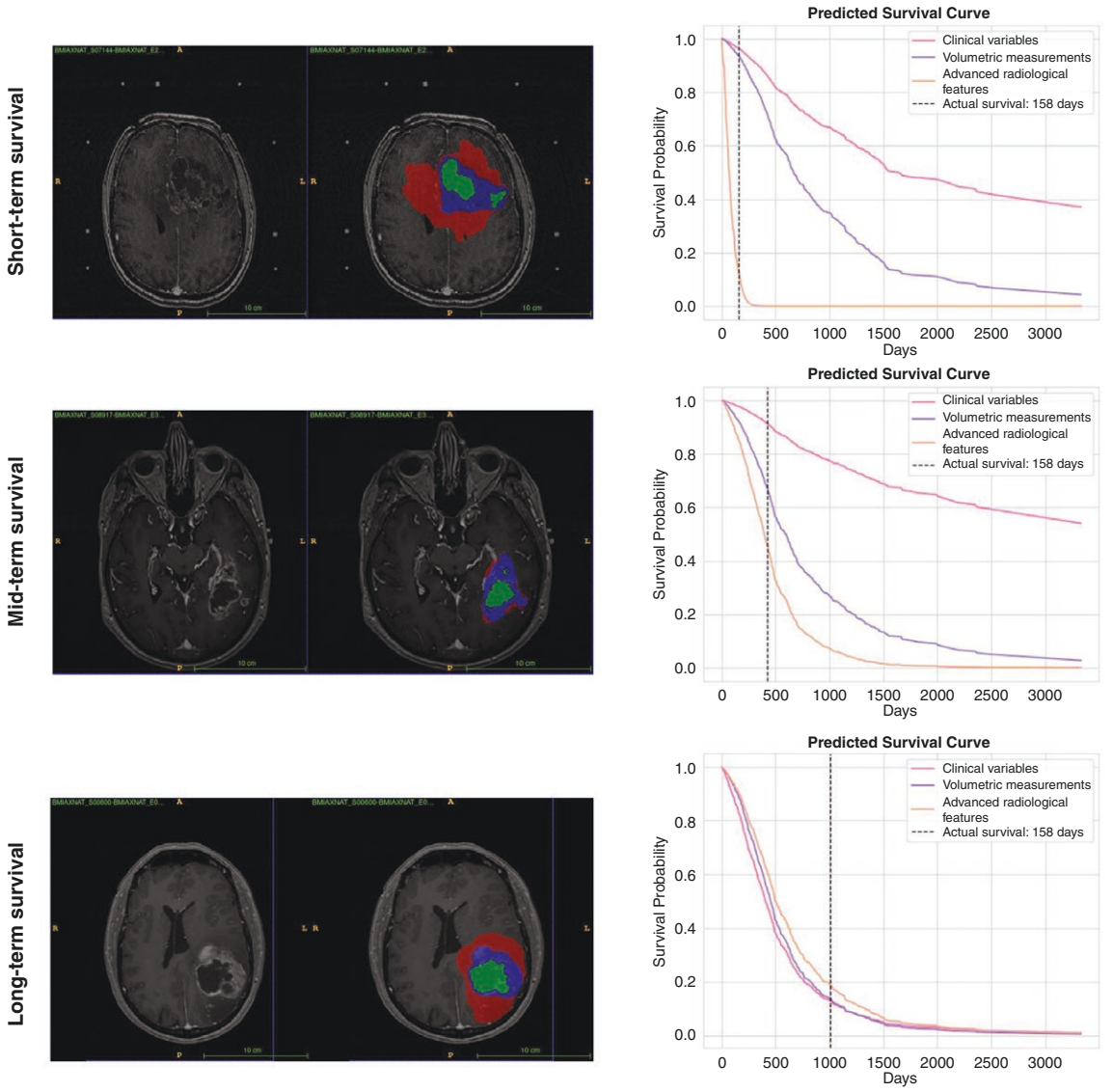

Figure 4. Prognosis of survival curves for patients recorded with short-term (top row), mid-term (middle row), and long-term (bottom row) survival. At each row, from left to right we display a 2D axial slice of the pre-intervention T1ce scan and the same slice with the predicted segmentation mask overlayed (red: peritumoral non-enhancing tissue, blue: enhancing tumor, and green: necrotic tumor core). At the right-most place of each row, we depict the predicted survival curves from all DeepSurv models trained on the 'Board Meeting' timepoint using different sets of input features (magenta: Clinical Variables, purple: + Volumetric measurements, and orange: + Advanced radiological features). **Short-term survival:** 63-year-old male biopsy patient with a frontal lobe tumor (72.62% overlap) and a pre-intervention KPS score of 60, pre-intervention enhancing tumor volume of 64.39 ml, and a survival time of 158 days. **Mid-term survival:** 69-year old male with a temporal lobe tumor, a pre-intervention KPS score of 90, pre-operative enhancing tumor volume of 24.86 ml, an enhancing residual tumor volume of 7.8 ml, and a survival time of 420 days. The patient was treated with both chemotherapy and radiotherapy after tumor resection **Long-term survival:** 65 year-old male patient with a pre-treatment KPS score of 100, pre-operative enhancing tumor volume of 10.53 ml, an enhancing residual tumor volume of 0.032 ml, and a survival time of 1008 days. The patient was treated with both chemotherapy and radiotherapy after tumor resection, while the pre-operative tumor had a 76.87% overlap with brain functional areas.

Supplementary Material). Regarding the link between better prognosis and larger pre-intervention necrotic core volume, we investigated whether the latter was inversely proportional to either pre-intervention enhancing tumor size and/or enhancing residual tumor size, which was not the case. We then proceeded with subset analysis and introducing interaction terms in the Cox multivariable analysis, yet the outcome did not explain this association either. Therefore, this remains an area for future exploration and research.

Survival studies on patients diagnosed with diffuse gliomas have mostly focused on the integration of clinical variables and radiomics features in the same predictive model. Similar to ours, a few studies have tried to associate tumor location with patient survival, either by fitting traditional survival regression and deep-learning-based models to the survival regression task,[47] or by associating specific atlas-based areas with short-, mid-, and long-term survival.[48] While results are not directly comparable due to the use of different datasets, which also vary in their handling of censored patients, most studies report C-indices in the vicinity between 60% and 70%[20] for glioblastoma datasets. With the additional inclusion of low-grade gliomas[47] and by predicting the survival probability in months,[49] reported C-indices might exceed 70%. Such modifications simplify the task of overall survival regression, as low-grade gliomas are associated with longer survival, whereas predicting survival in months allows for higher precision in predictions. Notably, the reported C-index should not be confused with the comparison between the predicted survival curve and the true survival curve. The C-index is a ranking-based metric that demonstrates how well we can order patient risk scores correctly, whereas the predicted survival curve demonstrates whether a model captures the survival trends correctly.

Our study also comes with certain limitations. One limitation was that for various reasons, including but not limited to different protocols in recording of clinical variables, individual observations were missing for subsets of study participants. Since we assumed full availability of clinical variables in our model design, we applied feature imputation to mitigate this situation, using independent models trained and validated on subsets without missing observations. While imputation can introduce potential biases and distortions (eg, loss of variability, overfitting), the survival regression models captured survival trends effectively on their respective, unseen test data. Furthermore, certain prognostic factors known to impact survival outcomes (MGMT status, IDH mutation, or detailed information on chemoradiotherapy regimens) were missing for the vast majority of study participants. In such cases, feature imputation from a few known observations to many unknown ones would entail the risk of imputing mostly incorrect values, severely affecting predictive performance. Notably, the predictive performance of our models increased in post-operative timepoints, suggesting that our choice of survival regression model training was effective, even in the absence of this information. Our methodological design allows for the incorporation of additional variables in the input space, thereby enabling future studies to examine them and potentially predict clinical outcomes with greater accuracy (R1.3). Another limitation was the absence of an external survival dataset. Prior to carrying out experiments, we did consider the exclusion of a single center from model training and its use as an internal test dataset. However, there was no clear candidate among the participating centers, since we could not guarantee that results would be unaffected by non-recorded clinical variables. The distribution of patients across centers was also non-uniform, which could potentially introduce bias when computing average survival regression metrics. Alternatively, selecting a random subset of patients to leave out as an independent validation cohort could introduce potential biases in our analysis. Consequently, a repeated five-fold cross-validation scheme was chosen to minimize the impact of any specific selection of data. Given the magnitude of the dataset, we regard it as representative of a real world setting, and deem our trained models generalizable to the task.

Apropos of future considerations, while for patients who underwent tumor resection the T1ce scans were at our disposal and we did extract the corresponding residual tumor masks, we did not consider the use of post-operative Raidionics features for our analysis, to avoid the convolution of the input space with repeated values for biopsy patients. Post-operative tumor location and overlap with brain areas have been discussed with relation to patient survival,[50] demonstrating promising results. Thus, by solely focusing on tumor resection patients, future studies can explore this information. Moreover, we did not investigate the direct use of MRI scans on the input space, while through the use of deep convolutional neural networks (CNNs) as feature extractors, raw MRI-based features could potentially carry prognostic value for the survival prediction task. In BraTS survival challenges,[40] MRI scans have been often used for survival category classification, but rarely for predicting survival probability functions. Even in such studies, they either focused on generating new predictors for survival[49] or used CNNs for extra preprocessing and feature extraction prior to employing RSFs. Thus, directly training CNN models in overall survival regression remains an open area for exploration.

In conclusion, our retrospective study harnessed clinical and imaging data, collected from multiple centers in Europe and the US, and demonstrated the value of incorporating advanced radiological features in overall survival prognosis for patients diagnosed with glioblastoma at various stages of treatment. Despite certain limitations, our results show that model predictions improve when imaging information from MRI scans is added, particularly at earlier stages of patient monitoring. DeepSurv was found to be the most accurate method across timepoints, yielding higher predictive performance and better capturing survival trends.

Supplementary Material

Supplementary material is available online at *Neuro-Oncology Advances* (https://academic.oup.com/noa).

Keywords:

survival analysis | glioblastoma | deep neural networks | magnetic resonance imaging.

Funding

A.F., R.S.E, and P.C.d.W.H are funded by the PICTURE project, which is sponsored by an unrestricted grant of Stichting Hanarth fonds, "Machine learning for better neurosurgical decisions in patients with glioblastoma"; a grant for public-private partnerships (Amsterdam UMC PPP-grant) sponsored by the

Dutch government (Ministry of Economic Affairs) through the Rijksdienst voor Ondernemend Nederland (RVO) and Topsector Life Sciences and Health (LSH), "Picturing predictions for patients with brain tumors"; a grant from the Innovative Medical Devices Initiative program, project number 10-10400-96-14003; The Netherlands Organisation for Scientific Research (NWO), 2020.027; a grant from the Dutch Cancer Society, VU2014-7113 and the Anita Veldman foundation, CCA2018-2-17. F.B. is funded by the National Institute for Health Research (NIHR) biomedical research centre at UCLH. R.H.H. is supported by a grant from The Research Council of Norway, grant number 323339. D.B., I.R., and O.S. are partly funded by the Norwegian National Research Center for Minimally Invasive and Image-Guided Diagnostics and Therapy. A.S.J holds grants from the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreements (ALFGBG-965622 and ALFGBG-1006089).

Conflict of interest statement: J.F. has received honoraria for lectures and consultations from the following for-profit companies: Novartis, Seagen, Sanova, and Servier. G.W. is on the advisory board of Servier. The remaining authors declare no conflict of interest.

Author Contributions

All authors revised and approved the final manuscript. *Conceptualization*: A.F., P.M., R.H.H., T.D., A.S.J., I.R., D.B., O.S., R.S.E., P.C.d.W.H., F.B. *Analyses*: A.F., R.S.E., P.C.d.W.H., F.B. *Review and Editing*: P.M., R.H.H., I.K., A.P., M.T., H.A., L.B., M.S.B., T.D., M.C.N., J.F., S.L.H-J., A.J.S.E., B.K, R.N.T., E.M., P.A.R., M.R., T.S., T.A., M.W., G.W., A.H.Z., L.M.S, A.S.J., E.T., I.R., D.B., O.S., R.S.E., P.C.d.W.H., F.B. *Supervision*: R.S.E., P.C.d.W.H., F.B.

Data availability

The primary patient dataset including clinical variables and MRI scans are not publicly available due to privacy regulations.

Affiliations

Cancer Center Amsterdam, Brain Tumor Center, Amsterdam University Medical Centers, Amsterdam, The Netherlands (A.F., I.K., R.S.E., P.C.W.H.); Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands (A.F., M.T., F.B.); Department of Neurosurgery, St. Olav's University Hospital, Trondheim, Norway (P.M., L.M.S., A.S.J., O.S.); Department of Neuromedicine and Movement Science, The Norwegian University of Science and Technology, Trondheim, Norway (P.M., O.S.); Department of Health Research, SINTEF Digital, Trondheim, Norway (R.H.H., A.P., I.R., D.B.); Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway (R.H.H., I.R.); Department of Neurosurgery, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands (I.K., R.S.E., P.C.W.H.); Application Solutions, Sopra Steria, Trondheim, Norway (A.P.); Department of Advanced Biomedical Sciences, University "Federico II," Naples, Italy (M.T.); Department of Neurosurgery, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands (H.A.); Neurosurgical Oncology Unit, Department of Oncology and Hemato-oncology, Humanitas Research Hospital and IRCCS Galeazzi Sant'Ambrogio, Universitá Degli Studi di Milano, Milano, Italy (L.B.); Department of Neurological Surgery, University of California San Francisco, California, USA (M.S.B., S.L.H.-J.); Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden (T.D., A.S.J., E.T.); IRCCS Ospedale Galeazzi Sant'Ambrogio, Milan, Italy (M.C.N., T.S.); Department of Biomedical Imaging and image-guided Therapy, Medical University Vienna, Wien, Austria (J.F.); Research Center for Medical Image Analysis and Artificial Intelligence (MIAAI), Faculty of Medicine and Dentistry, Danube Private University, Krems, Austria (J.F.); Department of Neurosurgery, Northwest Clinics, Alkmaar, The Netherlands (A.J.S.I.); Department of Neurosurgery, Medical University Vienna, Wien, Austria (B.K., G.W.); Department of Neurosurgery, Haaglanden Medical Center, The Hague, The Netherlands (R.N.T.); Hôpital Lariboisière, Department of Neurological Surgery, Paris, France (E.M.); Department of Neurology and Neurosurgery, University Medical Center Utrecht, Utrecht, The Netherlands (P.A.R.); Department of Medical Biotechnology and Translational Medicine, Universitá Degli Studi di Milano, Milano, Italy (M.R.); Department of Neurosurgery, Isala, 8025 AB Zwolle, The Netherlands (T.A.); Department of Neurosurgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands (M.W.); Department of Clinical Epidemiology and Biostatistics, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands (A.H.Z.); Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway (L.M.S.); Department of Radiology, Sahlgrenska University Hospital, Gothenburg, Sweden (E.T.); Institutes of Neurology and Healthcare Engineering, University College London, London, UK (F.B.)

References

1. Wesseling P, Capper D. WHO. 2016 classification of gliomas. *Neuropathol Appl Neurobiol.* 2018;44(2):139–150.
2. Stupp R, Hegi ME, Mason WP, et al; European Organisation for Research and Treatment of Cancer Brain Tumour and Radiation Oncology Groups. 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(5):459–466.
3. Jiang H, Cui Y, Wang J, Lin S. Impact of epidemiological characteristics of supratentorial gliomas in adults brought about by the 2016 World Health Organization classification of tumors of the central nervous system. *Oncotarget* 2017;8(12):20354–20361.
4. Ahmadipour Y, Jabbarli R, Gembruch O, et al. Impact of multifocality and molecular markers on survival of glioblastoma. *World Neurosurg* 2019;122(1):e461–e466.
5. Coburger J, Wirtz CR, König RW. Impact of extent of resection and recurrent surgery on clinical outcome and overall survival in a consecutive series of 170 patients for glioblastoma in intraoperative high field magnetic resonance imaging. *J Neurosurg Sci.* 2017;61(3):233–244.

6. Brown NF, Ottaviani D, Tazare J, et al. Survival outcomes and prognostic factors in glioblastoma. *Cancers* 3161;14(13):3161.
7. Zhao R, Zhuge Y, Camphausen K, Krauze AV. Machine learning based survival prediction in Glioma using large-scale registry data. *J Health Inform* 2022;28(4).
8. Kim M, Kim S, Park YW, et al. Sex as a prognostic factor in adult-type diffuse gliomas: an integrated clinical and molecular analysis according to the 2021 WHO classification. *J Neurooncol.* 2022;159(3):695–703.
9. Iliadis G, Kotoula V, Chatzisotiriou A, et al. Volumetric and MGMT parameters in glioblastoma patients: Survival analysis. *BMC Cancer* 2012;12(3):3.
10. Senders JT, Staples P, Mehrtash A, et al. An online calculator for the prediction of survival in glioblastoma patients using classical statistics and machine learning. *Neurosurgery.* 2020;86(2):E184–E192.
11. Henker C, Hiepel MC, Kriesen T, et al. Volumetric assessment of glioblastoma and its predictive value for survival. *Acta Neurochir.* 2019;161(8):1723–1732.
12. Wu CX, Lin GS, Lin ZX, et al. Peritumoral edema shown by MRI predicts poor clinical outcome in glioblastoma. *World J Surgical Oncol* 2015;13(97):97.
13. Sanghani P, Ang BT, King NKK, Ren H. Regression based overall survival prediction of glioblastoma multiforme patients using a single discovery cohort of multi-institutional multi-channel MR images. *Med Biol Eng Comput* 2019;57(8):1683–1691.
14. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48(4):441–446.
15. Kommers I, Bouget D, Eijgelaar RS, et al. Glioblastoma surgery imaging—reporting and data system: standardized reporting of tumor volume, location, and resectability based on automated segmentations. *Cancers* 8254;13(18):2021.
16. Bouget D, Alsinan D, Gaitan V, et al. Raidionics: an open software for pre- and postoperative central nervous system tumor segmentation and standardized reporting. *Sci Rep.* 2023;13(1):15570.
17. Hajianfar G, Haddadi Avval A, Hosseini SA, et al. Time-to-event overall survival prediction in glioblastoma multiforme patients using magnetic resonance imaging radiomics. *Radiol Med* 2023;128(12):1521–1534.
18. Chato L, Latifi S. Machine Learning and Radiomic Features to Predict Overall Survival Time for Glioblastoma Patients. *J. Pers. Med* 2021;11(12):1336.
19. Babaei Rikan S, Sorayaie Azar A, Naemi A, et al. Survival prediction of glioblastoma patients using modern deep learning and machine learning techniques. *Sci Rep.* 2024;14(1):2371.
20. Osman AF. A Multi-parametric MRI-based radiomics signature and a practical ML model for stratifying glioblastoma patients based on survival toward precision oncology. *Front Comput Neurosci.* 2019;13:58.
21. Fu J, Singhrao K, Zhong X, et al. Automatic deep learning-based workflow for glioblastoma survival prediction using preoperative multimodal MR images: a feasibility study. *Adv Radiat Oncol* 2021;6(5):100746.
22. Fiset S, Welch ML, Weiss J, et al. Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother Oncol.* 2019;135:107–114.
23. Pavic M, Bogowicz M, Würms X, et al. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol.* 2018;57(7):1070–1074.
24. Pemberton HG, Wu J, Kommers I, et al. Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms. *Sci Rep.* 2023;13(1):18911.
25. Helland RH, Ferles A, Pedersen A, et al. Segmentation of glioblastomas in early post-operative multi-modal MRI with deep neural networks. *Sci Rep.* 2023;13(1):18897.
26. Timmermann C. "Just give me the best quality of life questionnaire": the Karnofsky scale and the history of quality of life measurements in cancer trials. *Chronic Illn.* 2013;9(3):179–190.
27. Majewska P, Holden Helland R, Ferles A, et al. Prognostic value of manual versus automatic methods for assessing extents of resection and residual tumor volume in glioblastoma. *J Neurosurg.* 2025;142(5):1–9.
28. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2021;18(2):203–211.
29. Eijgelaar RS, Visser M, Müller DMJ, et al. Robust deep learning-based segmentation of glioblastoma on routine clinical MRI scans using sparsified training. *Radiol Artif Intell* 2020;2(5):e190103.
30. Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. *Hum Brain Mapp.* 2010;31(5):798–819.
31. Isensee F, Schell M, Pflueger I, et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Hum Brain Mapp.* 2019;40(17):4952–4964.
32. Bouget D, Eijgelaar RS, Pedersen A, et al. Glioblastoma surgery imaging-reporting and data system: validation and performance of the automated segmentation task. *Cancers (Basel)* 2021;13(18):4674.
33. Müller DMJ, Robe PAJT, Eijgelaar RS, et al. Comparing glioblastoma surgery decisions between teams using brain maps of tumor locations, biopsies, and resections. *JCO Clin Cancer Inform* 2019;2:1–12.
34. Collins DL, Zijdenbos AP, Baaré WF, Evans AC. ANIMAL+ INSECT: Improved Cortical Structure Segmentation. In: *Biennial International Conference on Information Processing in Medical Imaging.* Springer, Berlin/Heidelberg, Germany; 1999:p. 210–223. doi: 10.1007/3-540-48714-X_16
35. Yeo BT, Krienen FM, Sepulcre J, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol.* 2011;106(3):1125–1165.
36. Rojkova K, Volle E, Urbanski M, et al. Atlasing the frontal lobe connections and their variability due to age and education: a spherical deconvolution tractography study. *Brain Struct Funct.* 2016;221(3):1751–1766.
37. Cox DR. Regression Models and Life-Tables. In: Kotz S, Johnson NL, editors. New York: Springer; 1992:527–541. doi: 10.1007/978-1-4612-4380-9
38. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat* 2008;2(3):841–860.
39. Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Methodol.* 2018;18(24):1–11.
40. Bakas S, Reyes M, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv:1811.02629 [cs]* 2018.
41. Molinaro AM, Hervey-Jumper S, Morshed RA, et al. Association of Maximal Extent of Resection of Contrast-Enhanced and Non-Contrast-Enhanced Tumor With Survival Within Molecular Subgroups of Patients With Newly Diagnosed Glioblastoma [published correction appears in JAMA Oncol. 2020 Mar 1;6(3):444. doi: 10.1001/jamaoncol.2020.0360]. *JAMA Oncol.* 2020;6(4):495–503.
42. Carson KA, Grossman SA, Fisher JD, Shaw EG. Prognostic factors for survival in adult patients with recurrent glioma enrolled onto the new approaches to brain tumor therapy CNS consortium phase I and II clinical trials. *J Clin Oncol.* 2007;25(18):2601–2606.
43. Awad AW, Karsy M, Sanai N, et al. Impact of removed tumor volume and location on patient outcome in glioblastoma. *J Neurooncol.* 2017;135(1):161–171.
44. Chambless LB, Kistka HM, Parker SL, et al. The relative value of postoperative versus preoperative Karnofsky Performance Scale scores as a predictor of survival after surgical resection of glioblastoma multiforme. *J Neurooncol.* 2015;121(2):359–364.

45. Filippini G, Falcone C, Boiardi A, et al; Brain Cancer Register of the Fondazione IRCCS (Istituto Ricovero e Cura a Carattere Scientifico) Istituto Neurologico Carlo Besta. Prognostic factors for survival in 676 consecutive patients with newly diagnosed primary glioblastoma. *Neuro-Oncology.* 2008;10(1):79–87.
46. Wan Y, Rahmat R, Price SJ. Deep learning for glioblastoma segmentation using preoperative magnetic resonance imaging identifies volumetric features associated with survival. *Acta Neurochir.* 2020;162(12):3067–3080.
47. Ghanem M, Ghaith AK, Zamanian C, et al. Deep learning approaches for glioblastoma prognosis in resource-limited settings: a study using basic patient demographic, clinical, and surgical inputs. *World Neurosurg* 2023;175:e1089–e1109.
48. Osadebey M, Liu Q, Fuster-Garcia E, Emblem KE. Interpreting deep learning models for glioma survival classification using visualization and textual explanations. *BMC Med Inform Decis Mak.* 2023;23(1):225.
49. Lee JO, Ahn SS, Choi KS, et al. Added prognostic value of 3D deep learning-derived features from preoperative MRI for adult-type diffuse gliomas. *Neuro-Oncology.* 2024;26(3):571–580.
50. Pálsson S, Cerri S, Poulsen HS, et al. Predicting survival of glioblastoma from automatic whole-brain and tumor segmentation of MR images. *Sci Rep.* 2022;12(1):19744.