



OPEN

## Predicting survival of glioblastoma from automatic whole-brain and tumor segmentation of MR images

Sveinn Pálsson<sup>1</sup>✉, Stefano Cerri<sup>1</sup>, Hans Skovgaard Poulsen<sup>2</sup>, Thomas Urup<sup>2</sup>, Ian Law<sup>3</sup> & Koen Van Leemput<sup>1,4</sup>

Survival prediction models can potentially be used to guide treatment of glioblastoma patients. However, currently available MR imaging biomarkers holding prognostic information are often challenging to interpret, have difficulties generalizing across data acquisitions, or are only applicable to pre-operative MR data. In this paper we aim to address these issues by introducing novel imaging features that can be automatically computed from MR images and fed into machine learning models to predict patient survival. The features we propose have a direct anatomical–functional interpretation: They measure the deformation caused by the tumor on the surrounding brain structures, comparing the shape of various structures in the patient’s brain to their expected shape in healthy individuals. To obtain the required segmentations, we use an automatic method that is contrast-adaptive and robust to missing modalities, making the features generalizable across scanners and imaging protocols. Since the features we propose do not depend on characteristics of the tumor region itself, they are also applicable to post-operative images, which have been much less studied in the context of survival prediction. Using experiments involving both pre- and post-operative data, we show that the proposed features carry prognostic value in terms of overall- and progression-free survival, over and above that of conventional non-imaging features.

Glioblastoma (GBM) is the most common malignant primary brain tumor in adults<sup>1</sup>. Prognosis is generally very poor, with a median overall survival (OS) of less than 15 months, and a 5-year OS rate of only 10%, even when aggressively treated<sup>2–6</sup>. The standard treatment consists of maximal surgical resection followed by radiation therapy and chemotherapy with temozolomide<sup>5</sup>. Following standard therapy, OS and progression-free survival (PFS) have been shown to correlate with several patient-specific features such as age, performance status and expression of O<sup>6</sup>-methylguanine-DNA-methyltransferase (MGMT)<sup>6–9</sup>. However, the prognostic value of these features is still too low to guide treatment choices in individual patients.

Magnetic resonance (MR) images of glioblastoma patients contain vast amounts of information about the disease, some of which may carry prognostic value. The literature on imaging biomarkers for glioblastoma survival prediction is currently dominated by radiomics<sup>10</sup>, an approach in which hundreds or even thousands of features are extracted from delineated tumor regions of MR images, each quantifying some shape, texture, wavelet or histogram property. This approach has shown good performance in predicting survival in many studies<sup>11–18</sup>, likely stemming from the correlation between the tumor’s texture in MR images and its intratumoral heterogeneity and aggressiveness<sup>19,20</sup>. However, despite good prediction performance, radiomics suffers from three issues impeding wide-scale practical adoption:

- Lack of interpretability: Radiomic features, instead of aiming to be interpretable, are designed to be many, to maximize the chance of some having correlation to the target variable. Consequently, many radiomic features are seemingly arbitrary and hard to connect in a meaningful way to the nature of the disease. However,

<sup>1</sup>Department of Health Technology, Technical University of Denmark, Lyngby, Denmark. <sup>2</sup>Department of Oncology, The Finsen Center, Rigshospitalet, Copenhagen, Denmark. <sup>3</sup>Department of Clinical Physiology, Nuclear Medicine and PET, Center of Diagnostic Investigation, Rigshospitalet, Copenhagen, Denmark. <sup>4</sup>Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, USA. ✉email: sveinnpals1@gmail.com

interpretability of features is important: If a model cannot give interpretable explanations of its predictions, clinicians may not trust the model enough to factor its predictions into their decisions, even if the model is accurate<sup>21</sup>. Interpretable models may also uncover patterns in the data that give valuable new insight into the disease, and inspire future research.

- Difficulties generalizing: The reproducibility of studies using radiomics has been shown to be less than ideal, with results failing to generalize well across scanners and software implementations<sup>22–25</sup>. Since many radiomic features depend directly on raw image intensities, they are sensitive to subtle changes in scanning equipment and image acquisition parameters. Furthermore, both textural and shape features depend on the segmentation mask that is used<sup>26</sup>, underlining the importance of using image segmentation methods that are robust with respect to such sources of variation.
- Focus on pre-operative data: Compared to pre-operative images of glioblastoma, relatively little attention has been given to radiomics and other biomarkers in post-operative images. The reason may be that post-operatively, tumor shape and textural features are less easily detectable, as a large part of the tumor is usually removed. Nevertheless, post-operative images are collected closer to the time of disease progression and contain information about the success of operation, making them important to consider in a survival model. While the volume of tumor in post-operative images has been shown to correlate with OS<sup>27,28</sup>, more advanced imaging biomarkers in post-operative images remain mostly unexplored.

In this paper, we propose a method that aims to start addressing these shortcomings. Rather than focusing on in-region radiomic features of the tumor itself, we look at out-of-region features that are more straightforward to interpret and that can readily be applied both to pre- and post-operative data. For this purpose, we take advantage of a recently proposed method to robustly segment dozens of neuroanatomical structures in the presence of tumors<sup>29</sup>. Because this method aims to be invariant to imaging variations, it can be directly applied to data acquired at different centers with different scanners and protocols.

We demonstrate the resulting survival prediction method on two fundamentally different datasets: one pre- and one post-operative dataset, each acquired with different scanners, MR contrasts, and pre-processing workflows. Our results show that the proposed features improve the performance of survival models for both overall- and progression-free survival, compared to models based only on several previously known prognostic factors. To the best of our knowledge, this is the first time a survival model for glioblastoma has been proposed that characterizes a detailed segmentation of the surrounding brain, although related registration-based work exists<sup>30–32</sup>.

## Methods

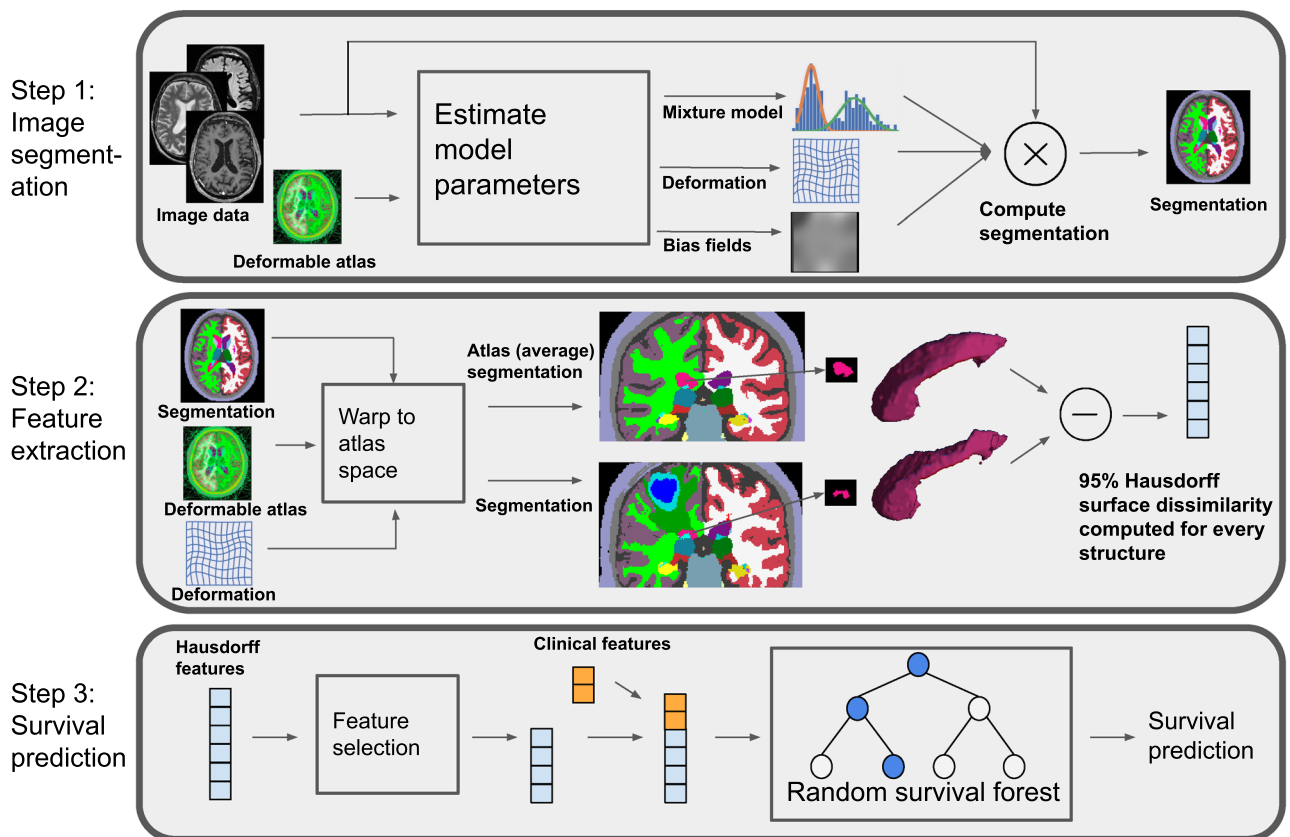
The method we propose for survival prediction consists of three steps, illustrated in Fig. 1. The first step is to segment the images with a contrast-adaptive *whole-brain* segmentation method, simultaneously segmenting dozens of brain structures and the tumor. In the second step, features are computed by comparing each segmented structure to its expected healthy shape using the 95% Hausdorff distance. In the third step, the extracted features are fed into a feature selector and a survival prediction model.

**Image segmentation.** For segmentation we use a method that we recently developed<sup>29</sup>, in which three tumor components (edema, enhancing core, and non-enhancing core) and dozens of neuroanatomical structures are automatically delineated from a patient's brain MR scan. For post-operative scans, another component is added to capture resection cavity. The method builds on a tool for whole-brain segmentation called Sequence Adaptive Multimodal SEGmentation (SAMSEG), which is distributed with the open-source software suite FreeSurfer<sup>33</sup>. It robustly segments head MR scans without any form of preprocessing, using an algorithm that can analyze multimodal data and adapt to variations in contrast due to differences in acquisition hardware or pulse sequences<sup>34</sup>.

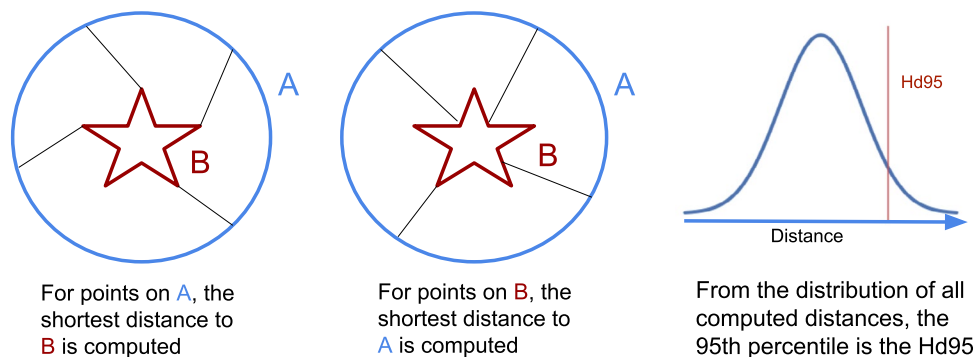
SAMSEG is centered on a probabilistic atlas that encodes the spatially varying voxel-wise prior probability of 41 different structures in an average-shaped head<sup>35</sup>. This atlas is augmented with a deformation model warping it to match the anatomy of individual subjects, along with models of the MR bias field and of structure-specific intensity profiles. At segmentation time, these models are fitted to the image being segmented and then used to compute an automatic segmentation (Fig. 1).

For the purpose of segmenting scans with brain tumors, the basic SAMSEG model is further augmented with a spatial regularization model of tumor shape using generative neural networks<sup>29</sup>. Although in its original formulation we used convolutional restricted Boltzmann machines<sup>36</sup> for this purpose, our current implementation has variational autoencoders<sup>37</sup> since these have a deeper structure and can therefore better represent lesion shape<sup>38</sup>.

**Feature extraction.** Once segmentations are available, we aim to extract features that can sensitively measure the effect the brain tumor has on the shape of the various neuroanatomical structures, compared to those seen in healthy individuals (Fig. 1 (Step 2)). To facilitate comparisons between individuals, we compute the features in atlas space, i.e., we warp the automatic segmentations back onto the average-shaped head model by applying the deformation fields that were estimated as part of the segmentation process. The resulting warped, subject-specific segmentations can then be compared to an “average” head segmentation that does not take any intensity information into account, obtained by assigning each voxel to the structure with the highest probability in the atlas. We will refer to this “average” head segmentation as *the atlas segmentation*. In healthy individuals, the subject-specific warped segmentations will be fairly close to the atlas segmentation in non-cortical structures after warping into atlas space, whereas in brain tumor patients the difference will often be much larger.

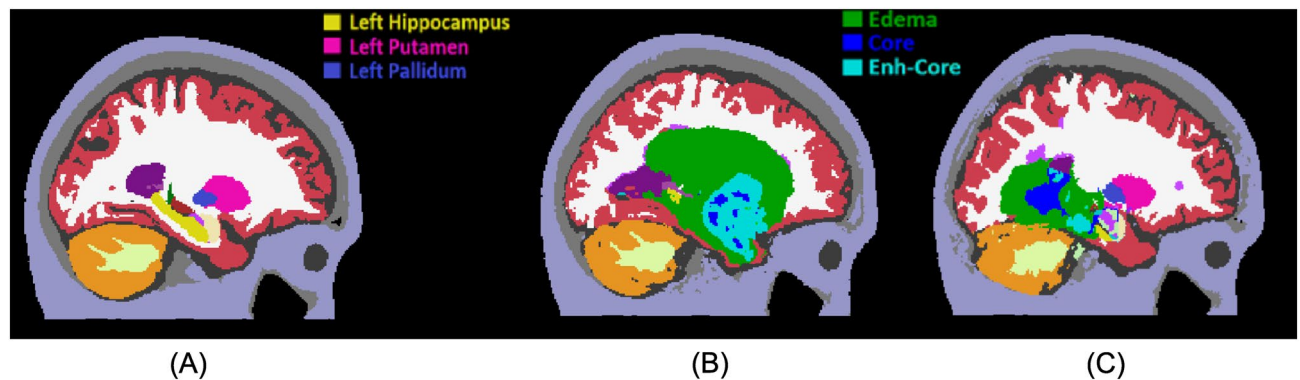


**Figure 1.** From MR images to survival prediction in three steps: segmentation, feature extraction, and survival prediction.



**Figure 2.** An illustration of how the Hausdorff 95% distance (Hd95) is computed between two example shapes.

In order to quantitatively compare the two segmentations, we compute a robust version of the Hausdorff distance<sup>39</sup> for each of 26 relevant structures: Accumbens area (L &R), amygdala (L &R), brain stem, caudate (L &R), cerebellum cortex (L &R), cerebral cortex (L &R), hippocampus (L &R), lateral ventricle (L &R), optic chiasm, pallidum (L &R), putamen (L &R), thalamus (L &R), ventral diencephalon (L &R), 3rd- and 4th-ventricles. The Hausdorff distance measures the distance between the outer borders of a pair of segmentation masks; its robust version is an often-used metric to quantify the performance of automatic segmentation methods with respect to manual “ground truth” delineations performed by human experts<sup>40</sup>. Let *A* and *B* denote the outer border of the segmentation masks of a particular brain structure, obtained from the atlas and warped segmentation, respectively. The Hausdorff distance computes, for all voxels on the border *A*, the shortest Euclidean distance to voxels on the border *B*, and vice versa, and returns the maximum value over all the computed distances. Because the maximum distance is highly sensitive to outliers, the robust version instead returns the 95th percentile of the distances (Fig. 2). The robust version is often called the 95% Hausdorff distance but for short, will be referred to as Hd95 throughout the rest of the paper.



**Figure 3.** The atlas segmentation reflecting average anatomy (A) and automatic brain segmentations of two subjects with glioblastoma (B,C). The subject in (B) has a highly deformed left hippocampus, pallidum, and putamen, which is reflected in high Hd95 values for these structures. While the hippocampus in (C) is also deformed, the pallidum and putamen are largely unaffected.

In cases where no voxel is assigned to a structure when obtaining the automatic segmentation, the Hd95 is not defined. In such cases, we instead use a single voxel located at the center of mass of the corresponding atlas segmentation.

For an example of how Hd95 captures the deformation of brain structures, Fig. 3 shows two subjects with glioblastoma (Fig. 3B,C) and the corresponding atlas segmentation (Fig. 3A) for comparison. The tumor in Fig. 3B has a clear effect on the shape of the left hippocampus, putamen, and pallidum, with an estimated Hd95 of 21.7, 28.5, and 49.3 [mm], respectively. While also showing a clear deformation of the left hippocampus, the left pallidum and putamen in Fig. 3C seems largely unaffected, with Hd95 of 24.6, 2.5, and 2.2 [mm], respectively.

The proposed Hd95 features contain some information about where the tumor is located in the brain and its size, both of which have been studied before and shown to carry prognostic value<sup>6,9,28,41–44</sup>. To verify that any prognostic value of our features is not solely based on tumor size and location, in our experiments we also evaluate the performance of our survival prediction models when they are trained directly on the estimated tumor size and the center-of-mass (CoM) coordinates of the whole tumor (defined as the set of voxels assigned to any tumor component). The contrast-enhancing tumor volume (CEV) is the tumor size definition most widely used clinically, but we will also consider the volume of each tumor component (TCV), including resection cavity in case of post-operative images, the extent of which has previously been shown to be an important independent prognostic factor<sup>45</sup>.

**Survival prediction.** Survival predictions models were trained following a standard machine learning workflow. The workflow consists of feature selection to remove uninformative features, and subsequent fitting of a survival prediction model to the remaining features (see Fig. 1 (Step 3)).

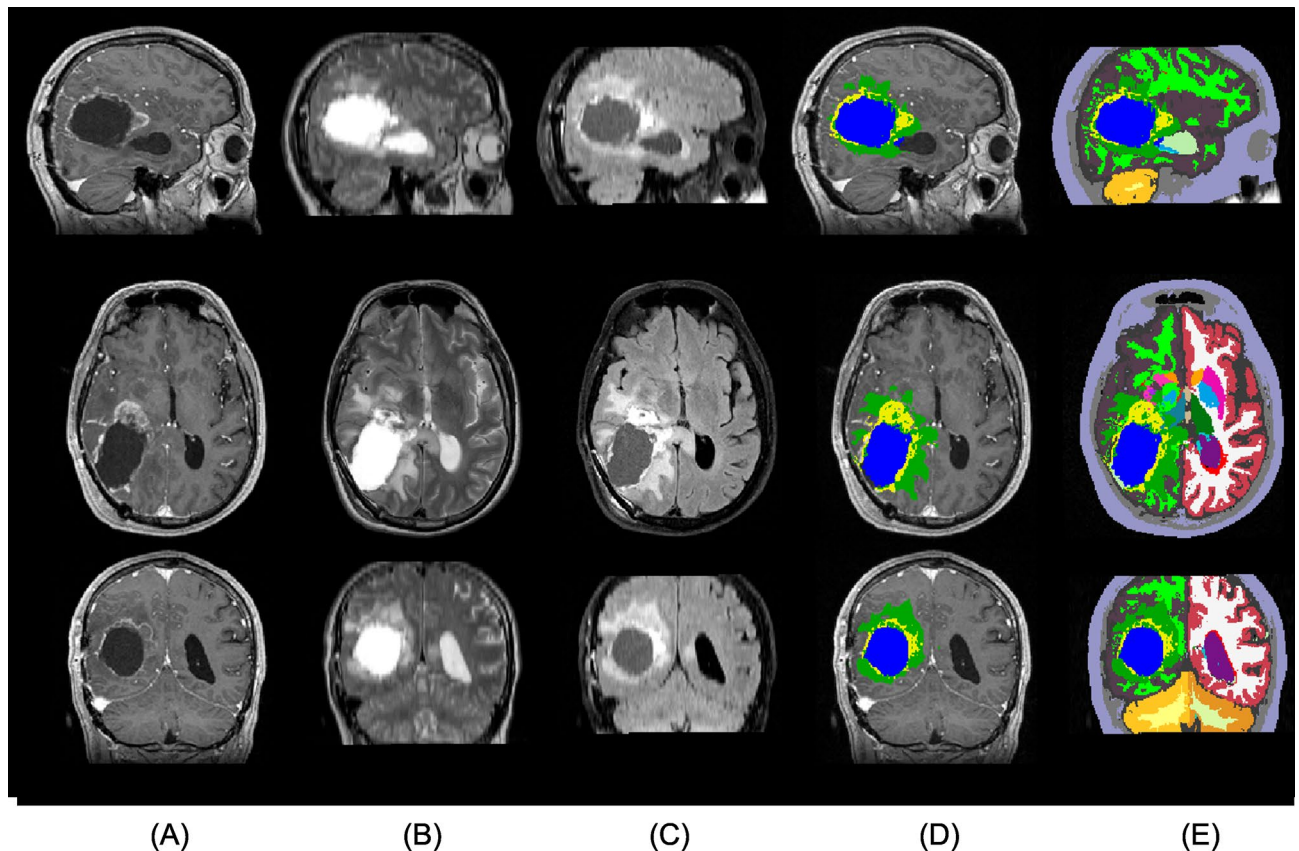
For feature selection, we used the univariate Cox proportional hazards (Cox PH) model<sup>46</sup> (implemented in python<sup>47</sup>), considering one feature at a time and retaining it if its coefficient is significantly nonzero. We used two-sided P values and considered  $P < 0.05$  statistically significant.

A random survival forest (RSF)<sup>48</sup> was used as the prediction model (implemented in python<sup>49</sup>). RSF extends the random forest model<sup>50</sup> to handle right-censored data, i.e., subjects who had not yet died by the end of the study—knowing that these subjects survived at least until their recorded time still contributes to fitting the RSF parameters. The RSF is an ensemble of trees whose leaf nodes estimate the subject's survival function from training data seen by the node. The survival prediction for a subject is taken as the expected survival of the average survival function across all leaf nodes the subject visits. Due to the small number of subjects in our datasets, we did not optimize over the RSF hyperparameters but left them at the default setting in the survival analysis software: 100 trees, no maximum depth, 6 subjects minimum to split a node, and minimum 3 subjects in leaf nodes. Models were trained via K-fold cross-validation where K was chosen such that in each fold, 5 subjects were left out while the model was trained on the remaining  $N-5$  subjects ( $K = N/5$ ); the model was then used to predict survival of the 5 left-out subjects. We repeated this procedure 100 times for more accurate estimation of model performance.

## Experiments and results

To demonstrate the versatility and reproducibility of the proposed method across data acquisitions, we performed experiments on two fundamentally different datasets: an in-house dataset of post-operative scans, and a publicly accessible dataset of pre-operative scans. Here we first describe these datasets, and subsequently present results for each.

**Ethical approval.** All patients included in our in-house post-operative dataset gave consent for imaging to be performed as part of their treatment. Permission for data collection was given by the Danish Data Protection Agency (2006-41-6979). The Danish National Committee of Health Research Ethics waived the need for explicit



**Figure 4.** A sample from the Copenhagen (post-operative) dataset. From top to bottom: sagittal, axial, and coronal view. The columns show (A) T1c, (B) T2, (C) FLAIR, and (D,E) the automatic segmentation output. (D) shows the tumor components only, while (E) shows the full segmentation output. The tumor components in (D,E) are edema (green), enhancing core (yellow), and non-enhancing core (blue). Resection cavity is shown in light-green color in the sagittal view of (E).

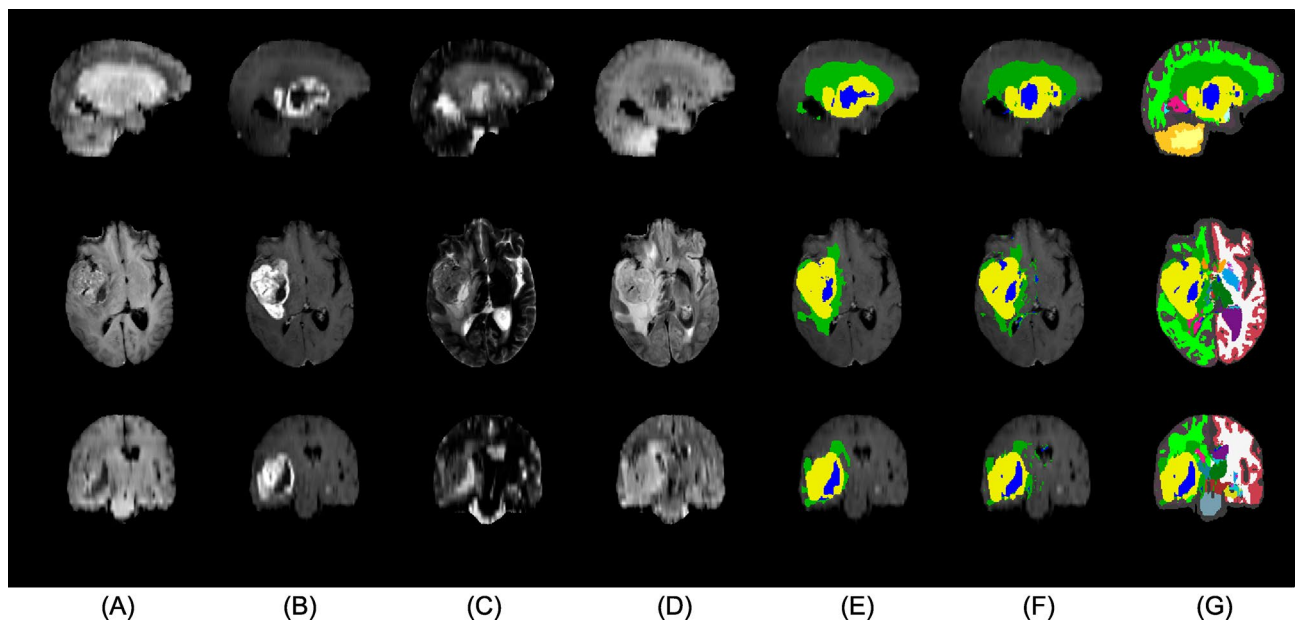
written informed consent since the study operated exclusively on anonymized retrospective data. This study was performed according to the Declaration of Helsinki and Danish legislation.

**Datasets.** *Copenhagen dataset (post-operative).* Our primary focus is on a set of post-operative scans acquired at Rigshospitalet, Copenhagen. It contains MR scans of 146 histologically verified glioblastoma patients, diagnosed in the period September 2011–April 2014 using the WHO standard at the time<sup>2</sup>. Using the 2021 WHO standard<sup>1</sup> retrospectively, the distribution of IDH mutation types in these subjects is as follows: GBM, IDH wild type, WHO grade IV (n = 119); Astrocytoma, IDH-mutant, WHO grade IV (previously Glioblastoma, IDH-mutant) (n = 9); and GBM, NOS (Not Otherwise Specified) (n = 19). Each patient received radiation therapy with concomitant and adjuvant temozolomide (see <sup>6</sup> for details about treatment). OS and PFS were recorded in months for all subjects with 14 and 6 censored subjects (i.e. still alive/non-progression at the end of the study), respectively.

MR scans were acquired for radiation planning 2–3 weeks post-operatively. The acquired MR modalities included 3D T1 (MPRAGE) post-administration of gadolinium contrast (T1c), T2, and FLAIR (Fig. 4A–C), using a 1.5T Siemens Espree scanner. The T1c scans were acquired using a voxel size of  $0.5 \times 0.5 \times 1.0 \text{ mm}^3$  (matrix size  $384 \times 512 \times 176$ ); the FLAIR scans with a voxel size of  $0.45 \times 0.45 \times 3.3 \text{ mm}^3$  (matrix size  $448 \times 512 \times 40$ ); and the T2 scans using a voxel size of  $0.3 \times 0.3 \times 3.3 \text{ mm}^3$  (matrix size  $672 \times 768 \times 39$ ). As the only form of pre-processing, intra-subject registration and resampling to  $1 \text{ mm}^3$  resolution was performed using FLIRT<sup>51</sup>. Three out of the 146 subjects were excluded as their post-operative MR data was unavailable. Out of the remaining 143 subjects, 11 were missing FLAIR scans and 3 were missing T2. However, our segmentation algorithm is robust with respect to missing modalities, allowing all 143 subjects to be included in the study.

Additional features recorded in the clinic were the patient’s age, performance status, and MGMT protein status. As mentioned in the introduction, these are features that have been previously shown to have prognostic value and are thus commonly considered for radiotherapy planning. We will refer to this set of variables as the “clinical features”.

*BraTS20 dataset (pre-operative).* To test the reproducibility of the methods we propose, we also applied them to a fundamentally different (namely, pre-operative) dataset, obtained with other acquisition settings and pre-



**Figure 5.** A sample from the BraTS20 (pre-operative) dataset. From top to bottom: sagittal, axial, and coronal view. The columns show (A) T1, (B) T1c, (C) T2, (D) FLAIR, (E) manual segmentation of tumor, and (F,G) the automatic segmentation output. (F) Shows the tumor components only, while (G) shows the full segmentation output. Some major differences to the Copenhagen (post-operative) dataset (see Fig. 4) can be seen in this figure.

processing steps. The Multi-modal Brain Tumor Segmentation Challenge 2020 (BraTS20)<sup>52–55</sup> released a publicly available set of 235 high-grade glioma subjects with overall-survival times. This dataset contains both glioblastoma and anaplastic astrocytoma<sup>40</sup>, although more detailed information on the subjects' sub-classification is not provided. For each subject, information on their age and OS is provided, but PFS or other clinical features are not available. None of the 235 subjects are censored.

The MR scans originate from multiple clinics and were acquired on different scanners, with magnetic field strengths of 1.5T and 3T. For each subject, the dataset contains a T1 pre- and post-administration of gadolinium contrast, a T2, and a T2 FLAIR scan (Fig. 5A–D). In a pre-processing step, the images were aligned to a brain template, interpolated to 1 mm<sup>3</sup> isotropic resolution, and skull-stripped by the challenge organizers<sup>40,52</sup>. Despite differences in available MR contrasts and in pre-processing compared to the Copenhagen dataset, our segmentation method did not need adjustment to handle the BraTS20 data (see in Fig. 5E–G, the manual and automatic tumor segmentation along with the whole-brain segmentation).

**Results on the Copenhagen dataset.** We present several different aspects of the proposed prediction method. First, we look at which Hd95 features were automatically selected for inclusion in the survival models. We then make a comparison between models trained on different feature sets, and we test the proposed method's ability to stratify patients into high- and low-risk groups based on their predictions. Finally, we evaluate the discriminative power of individual features for predicting short and long survival. In order to demonstrate the robustness of the proposed method to potential inaccuracies in the automatic segmentations, we also include simulation experiments in which the Hd95 distances are artificially enlarged as supplementary material.

*Feature selection.* Rather than reporting on Hd95 features selected within each fold during cross-validation, for conciseness here we present results of the Cox PH feature selection method on the entire cohort. Although this introduces information leakage between training and test sets, in our experiments we found that the selected features across folds were highly consistent, thus having minimal impact on the overall prediction performance (details provided in supplementary file). As shown in Table 1, our feature selection resulted in 10 retained Hd95 features for OS prediction, and 4 for PFS.

*Subject-level prediction performance.* To evaluate the prognostic value of the Hd95 features, in this section we investigate the performance of RSF prediction models trained on different sets of input features. In particular, we are interested in the comparison of models trained with the clinical features alone; the Hd95 features alone; and the combination of both. In addition, we compare with models that include tumor size (either TCV or CEV) and center-of-mass (CoM) as input features, as well as with models that only use age as the clinical variable. Note that feature selection was only performed on the Hd95 features as the clinical, size and location features have all been previously shown to carry prognostic value<sup>6,9,28,41–44</sup>.

The first two columns of Table 2 show performance of the RSF models, computed from the cross-validated predictions on the Copenhagen dataset. To quantify the performance of any given model, its predictions were

Feature	OS (Copenhagen)	PFS (Copenhagen)	OS (BraTS20)
Amygdala	✓ (L)		✓ (L)
3rd-Ventricle	✓		
Hippocampus			✓ (L)
Lateral ventricle	✓ (L)		
Pallidum	✓ (L, R)	✓ (L)	✓ (L)
Putamen	✓ (L)	✓ (L)	✓ (L)
Thalamus	✓ (L, R)	✓ (L)	
Ventral diencephalon	✓ (L, R)	✓ (L)	

**Table 1.** Brain structures whose Hd95 feature was selected by the feature selection method are marked with a checkmark, accompanied by L and R denoting left and right-sided structures. Shown for both the Copenhagen and BraTS20 datasets.

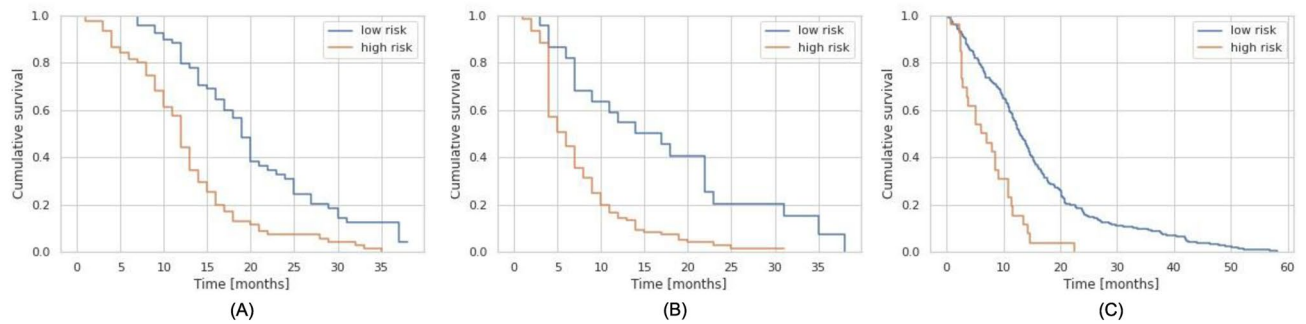
Features	Copenhagen		BraTS20
	OS	PFS	OS
Hd95 + Clinical + CoM + CEV	0.670 (0.668–0.671)	0.637 (0.634–0.641)	0.619 (0.618–0.620)
Hd95 + Clinical + CoM + TCV	0.657 (0.655–0.659)	0.629 (0.624–0.634)	0.612 (0.611–0.614)
Hd95 + Clinical + CEV	0.666 (0.665–0.668)	0.597 (0.595–0.600)	0.631 (0.630–0.632)
Hd95 + Clinical	0.669 (0.668–0.671)	0.614 (0.612–0.616)	0.612 (0.611–0.613)
Hd95 + CoM + CEV	0.635 (0.633–0.637)	0.629 (0.625–0.634)	0.564 (0.562–0.565)
Hd95 + CEV	0.643 (0.641–0.644)	0.580 (0.575–0.584)	0.594 (0.593–0.595)
Clinical + CoM + CEV	0.624 (0.623–0.625)	0.622 (0.618–0.626)	0.599 (0.598–0.600)
Clinical + CoM + TCV	0.595 (0.592–0.598)	0.613 (0.608–0.618)	0.600 (0.598–0.602)
Clinical + CEV	0.591 (0.589–0.593)	0.567 (0.563–0.572)	0.616 (0.614–0.617)
CoM + CEV	0.548 (0.545–0.551)	0.605 (0.599–0.611)	0.517 (0.516–0.518)
CoM + TCV	0.540 (0.537–0.543)	0.617 (0.612–0.622)	0.536 (0.533–0.539)
Hd95	0.644 (0.643–0.646)	0.552 (0.548–0.556)	0.571 (0.570–0.572)
Clinical	0.574 (0.572–0.576)	0.524 (0.522–0.527)	0.581 (0.579–0.583)
CoM	0.550 (0.548–0.551)	0.591 (0.588–0.594)	0.504 (0.503–0.506)
CEV	0.479 (0.476–0.482)	0.551 (0.547–0.554)	0.534 (0.533–0.535)
TCV	0.525 (0.522–0.528)	0.574 (0.570–0.577)	0.553 (0.551–0.555)
Age	0.509 (0.506–0.513)	0.519 (0.516–0.522)	0.581 (0.579–0.583)

**Table 2.** Prediction performance measured with the C-index for models trained on several different sets of features. The first two columns show results for OS and PFS prediction on the Copenhagen post-operative dataset, whereas the last column contains OS prediction performance on the BraTS20 pre-operative dataset. Note that the clinical features for the Copenhagen dataset include age, performance status, and MGMT methylation, while the available clinical features for the BraTS dataset consist only of age. Including the proposed Hausdorff (Hd95) features in the survival model provides an improvement in prediction performance over models that only consider conventional clinical features, tumor size and location.

compared with the ground truth survival times using Harrell's concordance index (C-index)<sup>56</sup>. The C-index computes the probability that for a pair of randomly selected subjects, their predicted survival is correctly ordered with respect to their true survival times. A C-index value of 1 means perfect prediction performance while 0.5 is the expected result of blindly guessing. The reported C-index is the average over the 100 repetitions of cross-validation, accompanied by the 95% confidence interval of the mean in brackets.

The best model for OS was achieved by combining the proposed features with the previously known prognostic clinical features: further addition of CoM, TCV, and CEV did not provide significant improvement. Individually, the clinical, size and location features all showed lower performance than the Hd95 features for OS prediction, and when combined they achieved only 0.624 C-index, compared to the 0.670 C-index when Hd95 was also included. The Hd95 features thus seem to bring prognostic value that is not contained in simple size and location based features.

For PFS, the best model was achieved by combination of Hd95, clinical, CoM and CEV, achieving a C-index of 0.637. Individually, the CoM was the best predictor of PFS, and combining it with clinical and size features achieved a C-index of 0.622. The benefit of including the Hd95 features is clear for PFS but is considerably lower than for OS.



**Figure 6.** Kaplan–Meier curves showing the cumulative survival (fraction of the population alive/without progression at a given time), for (A) OS and (B) PFS in the Copenhagen dataset and (C) for OS in the BraTS20 dataset. The two survival groups in each figure are obtained by splitting the cohort based on the predicted survival at the threshold that best separates the cohort.

Hd95 features	OS (Copenhagen)		PFS (Copenhagen)		OS (BraTS20)	
	% short	<i>P</i>	% short	<i>P</i>	% short	<i>P</i>
Left lateral ventricle	92	$1 \times 10^{-3}$	71	$4 \times 10^{-2}$	–	–
Left putamen	85	$2 \times 10^{-3}$	71	$2 \times 10^{-3}$	71	$2 \times 10^{-3}$
Left pallidum	83	$4 \times 10^{-3}$	71	$7 \times 10^{-3}$	67	$5 \times 10^{-3}$
Left thalamus	82	$7 \times 10^{-3}$	91	$2 \times 10^{-2}$	65	$2 \times 10^{-2}$
Left ventral diencephalon	77	$6 \times 10^{-3}$	69	$2 \times 10^{-2}$	81	$2 \times 10^{-4}$
4th ventricle	58	$4 \times 10^{-2}$	–	–	–	–
Left amygdala	–	–	–	–	79	$1 \times 10^{-2}$
Left hippocampus	–	–	–	–	75	$7 \times 10^{-4}$

**Table 3.** Percentage of short survivors among the subjects in the highest 10% range of individual Hd95 feature values. The table also shows the *P*-value of a log-rank test between the survival times of subjects within and outside the highest 10% range. Brain structures where the log-rank *P*-value  $> 0.05$  are omitted.

**Risk group stratification.** Here we demonstrate that the proposed survival models can be used to stratify patients into low- and high-risk groups. For this purpose, a threshold was selected by searching, among the predictions for all Copenhagen patients, for the value that best separates the dataset in terms of the recorded survival<sup>19,57</sup>. Separation quality was measured with the log-rank test<sup>58</sup>, which tests the hypothesis that two groups have the same survival distribution. The prediction value yielding the lowest *P*-value (of the log-rank test) was chosen as the threshold separating the low- from the high-risk patients. Visualization of the resulting groups, using the RSF models trained on the combination of clinical and selected Hd95 features as prediction models, is shown with Kaplan–Meier survival curves<sup>59</sup> for OS and PFS in Fig. 6A,B, respectively.

We further computed the corresponding hazard ratio for the obtained splits (ratio of hazard rates between the two groups under the proportional hazards assumption<sup>60</sup>), using the univariate Cox proportional hazards model where the input covariate was the group membership. In addition to the hazard ratio, its 95% confidence interval and log-rank *P*-value were also computed. For OS, the hazard ratio was 2.65 (1.85–3.79),  $P = 10^{-8}$  and for PFS the hazard ratio was 1.85 (1.7–4.78),  $P = 10^{-5}$ . These results show that our survival models can stratify patients into significantly different survival groups for both OS and PFS.

**Prognostic potential of individual features.** The Hd95 features we propose have a clear anatomical-functional interpretation: higher values reflect more severe deformation in the corresponding brain structures. To test the intuition that highly deformed individual structures are associated with poor outcomes, we concentrated on subjects with very high deformations and tested to what degree their survival differs from that of the remaining subjects. Specifically, for each of the 26 brain structures for which we computed Hd95 features, we split the subjects into two groups according to whether or not they are in the highest 10% range of feature values. We then computed (1) the percentage of short survivors (below the median survival of the cohort) among the subjects in the highest 10% range, and (2) the log-rank test between the two groups.

The results of this experiment are listed in Table 3 for structures where the log-rank *P*-value was significant. The results show that for several brain structures, high Hd95 value is a strong predictor of short survival. The best predictor of OS was deformation of the left lateral ventricle, where 92% of the subjects with the most deformation were short survivors. For PFS, the best predictor was the deformation of the left thalamus, with 91% of the subjects with the most deformation of that structure being short survivors.



**Results on the BraTS20 dataset.** The same methods were applied to the BraTS20 (pre-operative) dataset, for which our goal was to predict the OS only.

*Feature selection.* Feature selection on the full BraTS20 cohort resulted in selection of 4 Hd95 features: left putamen, left pallidum, left hippocampus and the left amygdala (listed in Table 1 for comparison with Copenhagen dataset). Similar to the Copenhagen dataset results, selecting features within each fold of cross-validation resulted in mostly the same features being chosen (see details in Supplementary File). Note that three of the selected features on the BraTS20 data were also selected for OS prediction on the Copenhagen dataset (vs. two for PFS, cf. Table 1).

*Subject-level prediction performance.* Model comparison to test whether the Hd95 features contain prognostic information not included in the clinical, size or location data was done in the same manner as with the Copenhagen dataset. An important difference is that the only clinical data available here is the subject's age, while the Copenhagen data also included MGMT and performance status.

As shown in the last column of Table 2, the best OS prediction model was obtained with a combination of Hd95, CEV, and age. This is largely in line with the results we obtained for OS prediction on the Copenhagen data (cf. first column of Table 2), where the best models were the ones combining Hd95 with other features. The results for size and location features are similar between the datasets: neither are good OS predictors individually. However, individually, here the age was the best feature, achieving a C-index of 0.581, which is substantially higher than in the Copenhagen dataset where age alone only achieved 0.509. Although the performance of the proposed Hd95 features and CEV individually was quite low (0.571 and 0.534, respectively), combining them both with the age achieved the best C-index of 0.631. While one of the best models for OS prediction on the Copenhagen dataset was the model combining Hd95 with clinical features, that specific combination only achieved a C-index of 0.612 on the BraTS20 dataset. Nevertheless, this is still an improvement over considering either of the two feature sets individually. It is further worth noting that age is the only clinical feature provided in the BraTS20 dataset—addition of MGMT and performance status information could improve the performance and possibly outperform the model using Hd95, CEV, and age also here.

*Risk group stratification.* As in the Copenhagen dataset, the prediction model trained on the combination of clinical and selected Hd95 features was used to stratify the BraTS20 cohort. The Kaplan–Meier curves in Fig. 6C show the proportion of subjects alive at any given time point for the two resulting groups. The corresponding hazard ratio was 2.81 (1.84–4.29) and log-rank P-value  $5 \times 10^{-7}$ , indicating that the two resulting groups have significantly different OS.

*Prognostic potential of individual features.* To explore to what degree individual Hd95 features can predict short survival in the BraTS20 dataset, we repeated the experiment of exploring the percentage of short survivors among the subjects with the most deformed brain structures. As shown in Table 3, the highest 10% range of feature values is predictive of short survival for several structures. Compared to our results on the Copenhagen dataset, two new structures show high predictive power in the BraTS20 data.

## Discussion

In this paper we have proposed a new set of imaging features for glioblastoma survival prediction. Our main goal was to introduce imaging features that are more interpretable and that can be replicated across different MR contrasts, scanning equipment, or preprocessing. The proposed Hd95 features can be interpreted as measuring the deviation from normal brain morphology due to glioblastoma, and are computed by comparing an automatic whole-brain segmentation with its expected equivalent in healthy subjects. To achieve robustness to missing MR modalities and variations in scanners or acquisition protocols, the automatic segmentations were obtained with a method that was designed to have these properties.

Using experiments on two different datasets—one post-operative and one pre-operative—we showed that the proposed features carry prognostic information and can improve survival models that use conventional clinical features such as age, MGMT, and performance status. Group analysis based on the output of our models showed that they could clearly stratify the datasets into low- and high-risk groups with significantly different survival characteristics. Furthermore, individual feature predictiveness was explored, indicating that for some brain structures, very high deformation is a reliable indicator of short survival.

Through feature selection, we discovered several brain structures whose Hd95 value correlates with survival and were therefore retained for training our prediction models. Although the same set of structures was not selected in each case (OS vs. PFS and pre- vs. post-operative), two structures were selected in all three cases: the left pallidum and left putamen. Interestingly, we found that right-sided structures were overall less associated with survival. Two recent studies have explored the association of OS with left-sided glioblastoma (having higher volume in the left hemisphere than the right side) but with contrary results<sup>43,44</sup>. One study<sup>43</sup>, which showed the association of left-sidedness and worse prognosis, proposed a possible explanation could be that the left hemisphere's functions may be more essential for survival.

In glioma surgery the term “minimal common brain” refers to a probabilistic atlas of primarily centrally located areas of the brain that house the core of connectivity of the main associative pathways, where functional deficits cannot be compensated through functional plasticity and are, thus, considered non-resectable<sup>61</sup>. In our analysis, the most salient brain structures (Table 3) are all located centrally in the brain and are overlapping with “minimal common brain”. The biological mechanism that couples structural deformation to OS can be speculated

to arise from functional disruption in these critically important structures and/or their pathways, that cannot be compensated for by functional reorganization leading to an early decrease in clinical performance status and death. The functional disruption could arise from mechanically induced pressure caused by mass effects, disruption of blood supply, and direct tumor infiltration.

By automatically computing detailed segmentations of both the tumor and the surrounding brain, the method presented here makes it easy to analyze the relative role of both aspects on survival prediction, as we have done in our experiments (Table 2). Nevertheless, answering long-standing questions such as whether patients with tumors presenting more infiltration than mass-effect have a better prognosis, remains difficult. This is because FLAIR MRI signal changes are unspecific and cannot be interpreted as identical with infiltration: Pre-treatment these changes can represent tumor infiltration or edema, and post-treatment additionally ischemia, demyelination and gliosis. Furthermore, prognosis needs to be interpreted within specific tumor groups, which is challenged by continuous revisions of pathological classifications, last in 2021. For GBM, classified by 2007 standards, a progression pattern *without* contrast enhancement had a better OS than progression *with*, likely because of the hypoxic and toxic environment associated with the contrast enhancing area<sup>62</sup>. In terms of recurrence pattern after radiation therapy, only 6% of GBM patients had distant recurrences, indicating poor local control close to the original contrast enhancement to be the primarily cause, and not tumor distant infiltration<sup>63</sup>. GBM patients with a longer progression free survival, e.g., MGMT-methylated patients, are more prone to non-central failures. Thus, treatment failure in contrast enhancing areas is the primary reason for a poor GBM prognosis while infiltration mainly influences OS in long survivor patients.

As demonstrated in our experiments, the features proposed in this paper readily generalize across datasets: They are independent of scanner and imaging parameters, and they can be computed from both pre- or post-operative images; from data that is skull-stripped or not; and from subjects with missing modalities. We did, however, see worse prognostic performance of the Hd95 features for OS prediction on the pre-operative cohort (BraTS20) compared to the post-operative one (Copenhagen). One possible reason for this discrepancy may be that the BraTS20 dataset contains both glioblastoma and anaplastic astrocytomas, which have different survival characteristics. The fact that BraTS20 is pre-operative may play an important role as well, as the effects of surgery can not be taken into account.

Although the extent of resection (EOR) has been shown to be an important independent prognostic factor<sup>45</sup>, EOR was not recorded in our post-operative dataset and was therefore not used in our experiments. However, we did consider the volume of each tumor component, including both the contrast-enhancing tumor volume and resection cavity in case of post-operative images (cf. Table 2). It has been argued by the response assessment in neuro-oncology (RANO) working group<sup>64</sup> that EOR does not represent the biological reality of the tumor burden, and the use of post-op residual tumor volumes is recommended as the measurement of cytoreductive treatment effect, as used e.g., in the fluorescence-guided surgery trials with 5-aminolevulinic acid (5-ALA)<sup>65</sup>. Since a 90% resection of a 1 mL and a 100 mL GBM is not the same thing, we have placed emphasis on the *absolute* and not *relative* measure of tumor remnants in our experiments.

The proposed Hd95 features measure how much each brain structure is deformed compared to its expected shape in the absence of pathology, and therefore they contain information about the location of the tumor, which has been shown previously to be a prognostic factor for OS<sup>9,28,41–44</sup>. Nevertheless, our results show that the proposed Hd95 features carry richer prognostic information for predicting OS than tumor location alone. For predicting PFS, tumor location was found to be a stronger predictor than Hd95 when considering these feature sets individually; however, substantially higher model performance was achieved with a combination of the two, together with clinical and size features. To the best of our knowledge, considerations of tumor location has not been a parameter used in the stratification of patients to treatment in clinical glioblastoma trials, although the poor prognostic feature is recognized in clinical management. Based on our results, the application of survival models exploiting advanced imaging features, such as the ones proposed here, could potentially help minimize bias in stratification in future clinical trials. High quality prognostic information could also potentially guide clinicians in adjusting the intensity of interventions, based on expected outcome and quality-of-life considerations.

Since the submission of this paper, independent research that also uses peri-tumoral deformation for survival prediction has been published elsewhere<sup>32</sup>. In that work, three tumor compartments are manually delineated from pre-operative T1c, T2 and FLAIR scans; the T1c scan with the tumor masked out<sup>66</sup> is nonlinearly deformed to a template; and features summarizing the properties of the resulting deformation field are then computed (five first-order statistics of the deformation field magnitude in each of twelve equidistant bands drawn around the tumor boundary). In contrast to this *registration-based* approach centered around manual tumor delineations, the method proposed here is *segmentation-based* and fully automatic: Not only does it perform detailed segmentations of both the tumor region and the surrounding brain directly from multi-contrast MRI scans (including post-operative ones), the method also computes its predictions from the *shape* of individual neuroanatomical structures instead. Such segmentation-based features are arguably easier to interpret than the features used in<sup>32</sup>, which closely resemble classical histogram radiomics features (*in casu* mean, median, standard deviation, skewness, and kurtosis of deformation field amplitudes in 5mm thick bands around the tumor).

While radiomics studies focus on patterns within the tumor region, in this study we have focused on the rest of the brain and ignored the tumor region itself entirely. Using such an approach, we demonstrated that considering out-of-region deformation features together with conventional clinical prognostic factors significantly improves survival models. A recent study<sup>19</sup> showed how 18 radiomic features could similarly improve RSF model accuracy when combined with clinical features. Future work may therefore explore combining both within-tumor radiomic features and our Hd95 features to further improve model accuracy, following the example of<sup>32</sup>.

Received: 24 September 2021; Accepted: 25 August 2022

Published online: 17 November 2022

## References

- Louis, D. N. *et al.* The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro Oncol.* **23**, 1231–1251 (2021).
- Louis, D. N. *et al.* The 2007 who classification of tumours of the central nervous system. *Acta Neuropathol.* **114**, 97–109 (2007).
- Louis, D. N. *et al.* The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* **131**, 803–820 (2016).
- Gutman, D. A. *et al.* MR imaging predictors of molecular profile and survival: Multi-institutional study of the TCGA glioblastoma data set. *Radiology* **267**, 560–569 (2013).
- Stupp, R. *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.* **10**, 459–466 (2009).
- Poulsen, S. H. *et al.* The prognostic value of fet pet at radiotherapy planning in newly diagnosed glioblastoma. *Eur. J. Nucl. Med. Mol. Imaging* **44**, 373–381 (2017).
- Michaelsen, S. R. *et al.* Clinical variables serve as prognostic factors in a model for survival from glioblastoma multiforme: an observational study of a cohort of consecutive non-selected patients from a single institution. *BMC Cancer* **13**, 402 (2013).
- Hegi, M. E. *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* **352**, 997–1003 (2005).
- Gorlia, T. *et al.* Nomograms for predicting survival of patients with newly diagnosed glioblastoma: prognostic factor analysis of EORTC and NCIC trial 26981–22981/ce. 3. *Lancet Oncol.* **9**, 29–38 (2008).
- Lambin, P. *et al.* Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–446 (2012).
- Booth, T. C. *et al.* Machine learning and glioma imaging biomarkers. *Clin. Radiol.* **75**, 20–32 (2020).
- Isensee, F., Kickingereder, P., Wick, W., Bendszus, M. & Maier-Hein, K. H. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop*, 287–297 (Springer, 2017).
- Weninger, L., Haarbuerger, C. & Merhof, D. Robustness of radiomics for survival prediction of brain tumor patients depending on resection status. *Front. Comput. Neurosci.* **13**, 73 (2019).
- Agravat, R. R. & Raval, M. S. Brain tumor segmentation and survival prediction. In *International MICCAI Brainlesion Workshop*, 338–348 (Springer, 2019).
- Sun, L., Zhang, S., Chen, H. & Luo, L. Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning. *Front. Neurosci.* **13**, 810 (2019).
- Baid, U. *et al.* Deep learning radiomics algorithm for gliomas (drag) model: a novel approach using 3d unet based deep convolutional neural network for predicting survival in gliomas. In *International MICCAI Brainlesion Workshop*, 369–379 (Springer, 2018).
- Baid, U. *et al.* Overall survival prediction in glioblastoma with radiomic features using machine learning. *Front. Comput. Neurosci.* **14**, 61 (2020).
- Ingrisch, M. *et al.* Radiomic analysis reveals prognostic information in t1-weighted baseline magnetic resonance imaging in patients with glioblastoma. *Investig. Radiol.* **52**, 360–366 (2017).
- Bae, S. *et al.* Radiomic MRI phenotyping of glioblastoma: Improving survival prediction. *Radiology* **289**, 797–806 (2018).
- Parekh, V. S. & Jacobs, M. A. Deep learning and radiomics in precision medicine. *Expert Rev. Precis. Med. Drug Dev.* **4**, 59–72 (2019).
- Shortliffe, E. H. & Sepúlveda, M. J. Clinical decision support in the era of artificial intelligence. *JAMA* **320**, 2199–2200 (2018).
- Traverso, A., Wee, L., Dekker, A. & Gillies, R. Repeatability and reproducibility of radiomic features: A systematic review. *Int. J. Radiat. Oncol. Biol. Phys.* **102**, 1143–1158 (2018).
- Zwanenburg, A. *et al.* The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328–338 (2020).
- Welch, M. L. *et al.* Vulnerabilities of radiomic signature development: The need for safeguards. *Radiother. Oncol.* **130**, 2–9 (2019).
- Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **278**, 563–577 (2016).
- Orlhac, F. *et al.* Tumor texture analysis in 18f-fdg pet: Relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J. Nucl. Med.* **55**, 414–422 (2014).
- Ellingson, B. M. *et al.* Validation of postoperative residual contrast-enhancing tumor volume as an independent prognostic factor for overall survival in newly diagnosed glioblastoma. *Neuro Oncol.* **20**, 1240–1250 (2018).
- Awad, A.-W. *et al.* Impact of removed tumor volume and location on patient outcome in glioblastoma. *J. Neurooncol.* **135**, 161–171 (2017).
- Agn, M. *et al.* A modality-adaptive method for segmenting brain tumors and organs-at-risk in radiation therapy planning. *Med. Image Anal.* **54**, 220–237 (2019).
- Prasanna, P. *et al.* Radiographic-deformation and textural heterogeneity (r-depth): an integrated descriptor for brain tumor prognosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 459–467 (Springer, 2017).
- Prasanna, P. *et al.* Mass effect deformation heterogeneity (MEDH) on gadolinium-contrast t1-weighted MRI is associated with decreased survival in patients with right cerebral hemisphere glioblastoma: A feasibility study. *Sci. Rep.* **9**, 1–13 (2019).
- Ismail, M. *et al.* Radiomic deformation and textural heterogeneity (r-depth) descriptor to characterize tumor field effect: Application to survival prediction in glioblastoma. *IEEE Trans. Med. Imaging* (2022) (accepted).
- Fischl, B. Freesurfer. *Neuroimage* **62**, 774–781 (2012).
- Puonti, O., Iglesias, J. E. & Van Leemput, K. Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling. *Neuroimage* **143**, 235–249 (2016).
- Van Leemput, K. Encoding probabilistic brain atlases using Bayesian inference. *IEEE Trans. Med. Imaging* **28**, 822–837 (2008).
- Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y. Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Commun. ACM* **54**, 95–103 (2011).
- Kingma, D. P. & Welling, M. Auto-encoding variational bayes. arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114) (2013).
- Cerri, S. *et al.* A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. *Neuroimage* **225**, 117471. <https://doi.org/10.1016/j.neuroimage.2020.117471> (2021).
- Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 850–863 (1993).
- Menze, B. H. *et al.* The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2014).
- Gorlia, T. *et al.* New prognostic factors and calculators for outcome prediction in patients with recurrent glioblastoma: A pooled analysis of eortc brain tumour group phase i and ii clinical trials. *Eur. J. Cancer* **48**, 1176–1184 (2012).
- Chaichana, K. L. *et al.* Relationship of glioblastoma multiforme to the lateral ventricles predicts survival following tumor resection. *J. Neurooncol.* **89**, 219–224 (2008).
- Abou Jaoude, D. *et al.* Glioblastoma and increased survival with longer chemotherapy duration. *Kansas J. Med.* **12**, 65 (2019).

44. Yersal, Ö. Clinical outcome of patients with glioblastoma multiforme: Single center experience. *J. Oncol. Sci.* **3**, 123–126 (2017).
45. Lacroix, M. *et al.* A multivariate analysis of 416 patients with glioblastoma multiforme: Prognosis, extent of resection, and survival. *J. Neurosurg.* **95**, 190–198 (2001).
46. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B (Methodol.)* **34**, 187–202 (1972).
47. Davidson-Pilon, C. lifelines: Survival analysis in python. *J. Open Source Softw.* **4**, 1317. <https://doi.org/10.21105/joss.01317> (2019).
48. Ishwaran, H. *et al.* Random survival forests. *Ann. Appl. Stat.* **2**, 841–860 (2008).
49. Pölsterl, S. scikit-survival: A library for time-to-event analysis built on top of scikit-learn. *J. Mach. Learn. Res.* **21**, 1–6 (2020).
50. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
51. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
52. Bakas, S. *et al.* Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint [arXiv:1811.02629](https://arxiv.org/abs/1811.02629) (2018).
53. Bakas, S. *et al.* Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **4**, 1–13 (2017).
54. Bakas, S. *et al.* Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The cancer imaging archive. *Nat. Sci. Data* **4**, 170117 (2017).
55. Bakas, S. *et al.* Segmentation labels and radiomic features for the pre-operative scans of the tcga-igg collection. *Cancer Imaging Arch.* **286** (2017).
56. Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *JAMA* **247**, 2543–2546 (1982).
57. Contal, C. & O’Quigley, J. An application of changepoint methods in studying the effect of age on survival in breast cancer. *Comput. Stat. Data Analysis* **30**, 253–270 (1999).
58. Mantel, N. *et al.* Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* **50**, 163–170 (1966).
59. Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958).
60. Sashegyi, A. & Ferry, D. On the interpretation of the hazard ratio and communication of survival benefit. *Oncologist* **22**, 484 (2017).
61. Ius, T., Angelini, E., de Schotten, M. T., Mandonnet, E. & Duffau, H. Evidence for potentials and limitations of brain plasticity using an atlas of functional resectability of WHO grade II gliomas: Towards a “minimal common brain”. *Neuroimage* **56**, 992–1000 (2011).
62. Nowosielski, M. *et al.* Radiologic progression of glioblastoma under therapy—An exploratory analysis of AVAglio. *Neuro Oncol.* **20**, 557–566 (2018).
63. Lundemann, M. *et al.* Patterns of failure for patients with glioblastoma following O-(2-[18F] fluoroethyl)-L-tyrosine PET-and MRI-guided radiotherapy. *Radiother. Oncol.* **122**, 380–386 (2017).
64. Vogelbaum, M. A. *et al.* Application of novel response/progression measures for surgically delivered therapies for gliomas: Response Assessment in Neuro-Oncology (RANO) Working Group. *Neurosurgery* **70**, 234–244 (2012).
65. Stummer, W. *et al.* Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: A randomised controlled multicentre phase III trial. *Lancet Oncol.* **7**, 392–401 (2006).
66. Dawant, B. M., Hartmann, S. & Gadamsetty, S. Brain atlas deformation in the presence of large space-occupying tumors. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 589–596 (Springer, 1999).

## Author contributions

S.P., K.V.L. and I.L. conceived and planned the experiments; S.P. carried out the experiments; S.P. took the lead in writing the manuscript, closely collaborating with I.L. and K.V.L., who supervised the project; S.C. contributed to the design and implementation of the methods; H.P. and T.U. carried out data acquisition and preparation; all authors reviewed the manuscript and provided feedback.

## Funding

This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant agreement No. 765148, as well as from the National Institute Of Neurological Disorders and Stroke under project number R01NS112161.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-19223-3>.

**Correspondence** and requests for materials should be addressed to S.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022