Shape and Intensity Analysis of *Glioblastoma Multiforme* Tumors

Yi Tang Chen, Sebastian Kurtek
Department of Statistics
The Ohio State University
Columbus, OH

chen.8414@osu.edu, kurtek.1@stat.osu.edu

Abstract

We use a geometric approach to characterize tumor shape and intensity along the tumor contour in the context of Glioblastoma Multiforme. Properties of the proposed shape+intensity representation include invariance to translation, scale, rotation and reparameterization, which allow for objective comparison of tumor features. Controlling for the weight of intensity information in the shape+intensity representation results in improved comparisons between tumor features of different patients who have been diagnosed with Glioblastoma Multiforme; further, it allows for identification of different partitions of the data associated with different median survival among such patients. Our findings suggest that integrating and appropriately balancing information regarding GBM tumor shape and intensity can be beneficial for disease prognosis. We evaluate the proposed statistical framework using simulated examples as well as a real dataset of Glioblastoma Multiforme tumors.

1. Introduction

Glioblastoma Multiforme (GBM) is a morphologically heterogeneous disease and is a very common form of brain tumor found in adults [5]; it is a severe disease with median survival time of approximately 12 months [9]. Notably, fewer than 10% of patients survive 5 years after diagnosis [15]. In previous studies, it has been shown that using quantitative imaging features for survival analysis in the context of GBM is beneficial for diagnostic as well as treatment purposes in the context of this severe disease [4]. Magnetic resonance imaging (MRI) is a common tool used to detect and examine brain abnormalities, e.g., brain tumors such as GBM. In particular, tumor shape and intensity, as captured by MRI, have been recognized as important prognostic factors for GBM [3, 12]. GBM tumor intensity captures regions of infiltrating tumor and edema, and other tissue properties. GBM tumor shape, on the other hand, captures protrusions of the tumor into surrounding

brain tissues. Both are thus important for characterization of GBM severity. However, it is often challenging to obtain a reliable characterization of, and to quantify differences in, tumor shape since its geometric features are described in a non-objective and (statistically) informal manner by domain experts, e.g., circularity, irregularity, etc. [1, 6]. Such descriptions are only crude summaries of GBM tumor shape and do not fully reflect the complexities of tumor growth and shape. Furthermore, most studies consider tumor shape and intensity separately, thus not capturing the potential dependence between these MRI-derived GBM tumor signatures [4]. These motivate our work to develop an approach that can integrate tumor shape and intensity via a unified mathematical representation that allows objective comparison and further joint statistical analysis of these two features of GBM tumors.

GBM tumor shape was analyzed in [3] using a geometric approach. We build on their framework by additionally incorporating information about intensity (texture) along tumor contours as captured in MRIs. Our approach leads to a distance between shape and intensity features of GBM tumors that is invariant to all shape-preserving transformations, and can be used for further statistical analysis, e.g., clustering. Using the proposed framework, we discovered clusters that are associated with large absolute differences in survival.

1.1. Data description

MRI imaging and associated survival data for 63 patients diagnosed with GBM, who consented under the Cancer Genome Atlas protocols[1], were obtained from the Cancer Imaging Archive[2]. In our study, we consider pre-surgical T1-weighted post contrast (henceforth referred to simply as T1) and T2-FLAIR (henceforth referred to simply as FLAIR) MRIs; data pre-processing is described in detail in [12] and we omit it here for brevity. In particular, our focus is on statistical analysis of GBM tumor shape and intensity

---

[1] http://cancergenome.nih.gov/
[2] http://www.cancerimagingarchive.net/

(a)   (b)
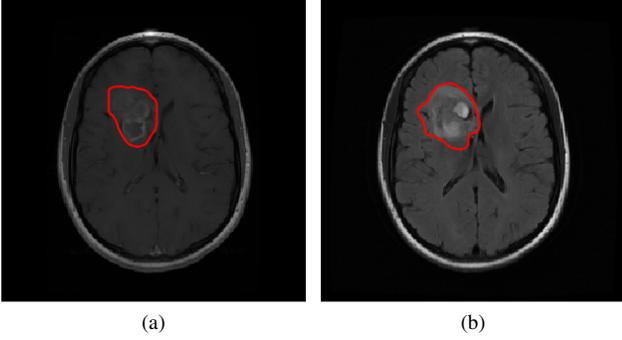
Figure 1. (a) T1 and (b) FLAIR MRI slices for one GBM patient in the data. The tumor outline is highlighted as a red curve.

as captured by the axial image slice with largest tumor area in T1, and the corresponding slice in FLAIR [3,12]. An example of a T1 image, and the corresponding FLAIR image, for one GBM patient are shown in Figure 1. Because of different properties of tissues highlighted by different types of MRI modalities, tumor shape and associated intensity differ in their appearance in the T1 and FLAIR images. The GBM tumor contours, which were segmented semi-automatically using the Medical Image Interaction Toolkit MITK3M3 Image Analysis (v 1.1.0) and the NIFTI toolbox in MATLAB, are illustrated as red closed curves overlaid on the MRI images. More details on the semi-automated segmentation, as well as other image processing steps related to this dataset of GBM tumors, can be found in Saha et al. (2016) [12].

2. Mathematical framework

The proposed representation of shape and intensity of GBM tumors, which we refer to as shape+intensity for simplicity of presentation, and the associated proposed distance, build on the elastic shape analysis framework [8,13].

2.1. Representation and comparison of GBM tumor shape+intensity

Let $\beta_c : \mathbb{S}^1 \to \mathbb{R}^2$ denote the 2D closed curve representing the tumor outline, and $\beta_I : \mathbb{S}^1 \to \mathbb{R}$ be the intensity function along the tumor outline; here, $\mathbb{S}^1$ acts as the natural domain for a closed curve and denotes the unit circle. As seen in Figure 1, segmented GBM tumor contours (as well as the associated intensity functions) form closed curves making this a natural representation. We model the tumor outline, and intensity along the outline, jointly via a 3D, parameterized closed curve $\beta(t) = [\beta_c(t), \; \beta_I(t)]^T \in \mathbb{R}^3 \; \forall \; t \in \mathbb{S}^1$. When constructing this representation based on segmented MRI images of GBM tumors, we (1) smooth each tumor outline using a moving average smoother and scale it to unit length (i.e., $\int_{\mathbb{S}^1} |\dot{\beta}_c(t)| dt = 1$) to ensure that the resulting shape representation is scale-invariant, and (2) smooth each intensity function using cubic splines and stan-

dardize it by subtracting the average and dividing by the standard deviation. Smoothing of the tumor contour and associated intensity values is motivated by the discrete nature of MRI imaging data. Standardization of intensity values for each observation is motivated by a well-known issue with MRI where intensity values in different images, corresponding to the same or similar tissues, can vary, making them not directly comparable [11]. We denote the standardized tumor outline and intensity function by β_c^* and β_I^*, respectively, and define $\beta_\lambda^*(t) = \begin{bmatrix} \beta_c^*(t) \\ \lambda\beta_I^*(t) \end{bmatrix} \in \mathbb{R}^3 \; \forall \; t \in \mathbb{S}^1$ as a composite coordinate function, where β_I^* is additionally scaled by a parameter λ, which controls the emphasis of intensity information when comparing MRI-derived GBM tumors.

A nuisance source of variation in the presented representation of tumor shape and intensity is the choice of parameterization. Different parameterizations of a closed curve, corresponding to different starting points or seeds on the domain $\mathbb{S}^1$ and different speeds of traversal along the curve, do not alter its shape. Let $\Gamma = \{\gamma : \mathbb{S}^1 \to \mathbb{S}^1 \, | \, \gamma$ is an orientation preserving diffeomorphism$\}$ denote the group of reparameterization functions. Then, a reparameterization of a closed curve β_λ^* is given by the group action of composition: $(\beta_\lambda^*, \gamma) = \beta_\lambda^* \circ \gamma$, where $\gamma \in \Gamma$. A parameterization-invariant distance between two closed curves $\beta_{\lambda,1}^*$ and $\beta_{\lambda,2}^*$ must satisfy the isometry property $d(\beta_{\lambda,1}^*, \beta_{\lambda,2}^*) = d(\beta_{\lambda,1}^* \circ \gamma, \beta_{\lambda,2}^* \circ \gamma)$, and it is well-known that the convenient $\mathbb{L}^2$ distance does not satisfy this property.

In this work, we use an elastic Riemannian metric [10] that is known to be parameterization invariant to define a distance between shape+intensity representations of GBM tumors. The resulting elastic distance is difficult to compute in practice, but under a convenient transformation called the square-root velocity function (SRVF) [7, 14], it simplifies to the $\mathbb{L}^2$ distance; this enables efficient quantification of shape+intensity differences between GBM tumors in a parameterization-invariant manner. First, we transform composite coordinate functions β_λ^* into their SRVF representation given by $q_\lambda^*(t) = \frac{\dot{\beta}_\lambda^*(t)}{\sqrt{|\dot{\beta}_\lambda^*(t)|}} \; \forall \; t \in \mathbb{S}^1$, where $\dot{\beta}_\lambda^*$ denotes the (coordinatewise) derivative of β_λ^* and $|\cdot|$ is the Euclidean norm in $\mathbb{R}^3$; one can reconstruct the original representation β_λ^* from its SRVF q_λ^*, up to a translation, by $\beta_\lambda^*(t) = \int_0^t q_\lambda^*(u)|q_\lambda^*(u)|du$. Note that the SRVF representation is automatically invariant to translation of the contour and intensity components since it depends on the derivative of β_λ^* only. Further, as described earlier, each GBM tumor contour is scaled to unit length via $\int_{\mathbb{S}^1} |\dot{\beta}_c^*(t)| dt = 1$ ensuring scale-invariance of this component. This results in the pre-shape+intensity space of GBM tumors, which is given by $\mathcal{C} = \{q_\lambda^* : \mathbb{S}^1 \to \mathbb{R}^3 \, | \, \int_{\mathbb{S}^1} q_\lambda^*(t)|q_\lambda^*(t)|dt = 0\}$, i.e., the

space of SRVFs of closed 3D curves.

We refer to $\mathcal{C}$ as the pre-shape+intensity space, because while translation and scale variabilities have now been removed from the representation, we have not yet accounted for variation in orientation and parameterization. The reparameterization of a curve β_λ^* via $\gamma \in \Gamma$, given by $\beta_\lambda^* \circ \gamma$, becomes $(q_\lambda^*, \gamma) = (q_\lambda^* \circ \gamma)\sqrt{\dot{\gamma}}$ for SRVFs, where $\dot{\gamma}$ is the derivative of γ. When considering rotations, it is important to ensure that only the GBM tumor contour, and not the intensity component, is transformed. For rotations, our focus is thus on a subgroup of $SO(3)$, $R = \begin{bmatrix} SO(2) & 0 \\ 0 & 1 \end{bmatrix} \subset SO(3)$, that rotates the first two coordinates of β_λ^* (or equivalently q_λ^*) related to the tumor contour, but not the third coordinate related to intensity. Then, for a rotation $O \in R$, $O\beta_\lambda^*$ and Oq_λ^* represent rotations of the contour+intensity curve and its SRVF, respectively. Importantly, the $\mathbb{L}^2$ distance between two SRVFs $q_{\lambda,1}^*, q_{\lambda,2}^* \in \mathcal{C}$ is isometric under the actions of Γ and R: $d_\mathcal{C}(q_{\lambda,1}^*, q_{\lambda,2}^*) = \|q_{\lambda,1}^* - q_{\lambda,2}^*\| = \|O(q_{\lambda,1}^*, \gamma) - O(q_{\lambda,2}^*, \gamma)\| = d_\mathcal{C}(O(q_{\lambda,1}^*, \gamma), O(q_{\lambda,2}^*, \gamma))$ for a $\gamma \in \Gamma$ and $O \in R$. This is a crucial property that allows us to use the $\mathbb{L}^2$ distance on the pre-shape+intensity space $\mathcal{C}$ to define a distance on the resulting shape+intensity space as described next. To unify all contour+intensity SRVFs that are within a reparameterization and rotation of each other, we define equivalence classes of the form $[q_\lambda^*] = \{O(q_\lambda^* \circ \gamma)\sqrt{\dot{\gamma}} \,|\, O \in R, \gamma \in \Gamma\}$, which provide a unique representation of shape+intensity for GBM tumors. In other words, different orientations and reparamaterizations of a GBM tumor's shape+intensity representation are contained in one equivalence class. The set of all such equivalence classes is denoted by $\mathcal{S} = \{[q_\lambda^*] \,:\, q_\lambda^* \in \mathcal{C}\}$ and is the shape+intensity space.

The distance between two shape+intensity equivalence classes, $[q_{\lambda,1}^*], [q_{\lambda,2}^*] \in \mathcal{S}$, is defined via the $\mathbb{L}^2$ distance on $\mathcal{C}$ as follows:

$$d_\mathcal{S}([q_{\lambda,1}^*], [q_{\lambda,2}^*]) = \inf_{O \in R, \gamma \in \Gamma} d_\mathcal{C}(q_{\lambda,1}^*, O(q_{\lambda,2}^* \circ \gamma)\sqrt{\dot{\gamma}})$$
$$= \inf_{O \in R, \gamma \in \Gamma} \|q_{\lambda,1}^* - O(q_{\lambda,2}^* \circ \gamma)\sqrt{\dot{\gamma}}\|. \tag{1}$$

The optimization problems over R (optimal rotation) and Γ (optimal reparameterization) are solved using singular value decomposition and the dynamic programming algorithm [2] (with an additional seed search), respectively. This process aligns or registers the contour+intensity representations of GBM tumors such that the $\mathbb{L}^2$ distance between them is minimized, which results in improved comparisons. An important aspect of this procedure is that the contribution of geometric contour information versus intensity information depends on the chosen value of λ that is used in the representation. When λ is small, the registration is driven by geometric features of the tumor contours; on the other hand,

when λ is large, it is driven by intensity information. We can visualize the optimal path of deformation (geodesic) between two GBM tumor shape+intensity objects, the length of which is given by the distance $d_\mathcal{S}$, via a linear interpolation between $q_{\lambda,1}^*$ and the optimally rotated and reparameterized $q_{\lambda,2}^*$. Such deformation paths are useful as they provide insight into the contribution of the shape and intensity components in the comparison of GBM tumors.

2.2. Assessment of GBM tumor shape+intensity heterogeneity

The proposed representation and distance are useful in assessing inter-patient GBM tumor heterogeneity with respect to shape and intensity. In particular, the shape+intensity distance between GBM tumors enables implementation of distance-based clustering approaches such as hierarchical clustering. Subsequently, we are able to assess differences in survival across the generated clusters. Importantly, depending on the choice of the tuning parameter λ, the clustering can be driven either by geometric shape differences (for small values of λ) or intensity differences (for large values of λ).

We apply hierarchical clustering based on the proposed distance, with complete linkage, to partition GBM tumor shape+intensity data into two groups. We use an implementation of hierarchical clustering as given by the MATLAB (version R2022a) functions *linkage()* and *cluster()*. We consider different values of λ to compute the shape+intensity distances between GBM tumors to try to understand the importance of geometric shape versus intensity in the clustering results and subsequent survival analysis. Then, within each cluster and for each considered value of λ, we estimate median survival times based on the Kaplan-Meier estimate of the survival function, which accounts for censoring in the patients' survival times. One can also use other summaries of survival times such as mean survival time, which is the area under the Kaplan-Meier curve. In this work, we use median survival rather than mean survival, as the latter can be underestimated if the longest observation time is censored, which is the case in our data. Finally, we compare the estimates between the two groups generated via our clustering approach. To evaluate the reliability of the presented clustering results, and the subsequent cluster-wise survival analysis, we provide multi-dimensional scaling (MDS) plots to visualize/assess separation across clusters where each GBM tumor shape+intensity observation is colored by its cluster membership, and further conduct three-fold cross-validation (over λ), respectively. While the cross-validation assesses differences in survival times across clusters for different values of λ, the MDS is based only on distances and does not incorporate any survival information.

3. Results

We present results of applying the proposed framework to analyze GBM tumor shape and intensity information. We begin with illustrations of registration and geodesic computation using a simulated example and multiple real GBM tumors. Then, we describe clustering results, based on the entire dataset of 63 GBM patients, for different values of λ, and relate them to survival. The smallest value of λ that we consider is $\lambda = 0.01$, where very little intensity information is used in the registration and comparison of MRI-derived GBM tumors; the majority of contribution comes from the difference in shape of the tumor contours. The largest value of λ that we consider is $\lambda = 0.5$, where the intensity information dominates comparisons of GBM tumors. Due to the relative scales of the GBM contour coordinates (after rescaling the contours to unit length) and the intensity values (after standardization), we found empirically that registration and comparison results for $\lambda > 0.5$ are very similar to those for $\lambda = 0.5$.

3.1. Registration and deformation of GBM tumors

Figure 2 illustrates registration results for two simulated contour+intensity objects. The simulated contours have a round shape with one region of protrusion (peak); the size of the peak differs between the two simulated contours. The corresponding simulated intensity functions both have a single peak of high intensity with different magnitude. In each panel, we show the two tumor contours along with the corresponding intensity functions in red and blue. For improved display, we plot the intensity functions as if they were defined on the domain $[0, 1]$. In Figure 2a, we show the standardized tumor contour+intensity objects. Recall that standardization refers to (1) rescaling of the contours, and (2) translation of the intensity functions via subtraction of the mean and rescaling of the intensity functions via division by the standard deviation. In Figure 2b, we show the registration results for a small value of $\lambda = 0.01$. In this case, as expected, the registration process is driven by geometric information of the contours: the peaks on the two contours are well-aligned. On the other hand, the peaks on the two intensity functions are misaligned since this information plays a small role in registration. When λ is increased to 0.5 in Figure 2c, intensity information starts to dominate the registration procedure. Thus, the peaks on the intensity functions become well-aligned while the peaks on the contours are not.

Figure 3 presents registration results for two GBM tumors, which have similar shape and were segmented from the T1 MRI modality. The blue and red tumor contour+intensity representations in this figure came from patients with 12.20 months and 2.83 months survival times, respectively. In Fig. 4, we also present registration results for two T1 MRI GBM tumors from patients with the shortest

(0.72 months, red) and longest (57.8 months, blue) survival times in our dataset. As illustrated in Fig. 4a, the patient with the shortest survival has a more irregularly shaped tumor contour, i.e., more protrusions, as compared to the patient with the longest survival. Similar to the registration results based on simulated contour+intensity objects (Fig. 2), Figures 3b and 4b show that when $\lambda = 0.01$, corresponding to less emphasis on intensity information during registration, geometric features of the tumor contours are better aligned as compared to the intensity functions along the tumor contours. As λ is increased to 0.5, the registration process focuses more on intensity information as demonstrated by improved alignment of the intensity functions in Figs. 3c and 4c.

The rows in Figs. 5 to 7 show geodesic deformations between the shape+intensity representations for the simulated example and real GBM tumors considered in Figs. 2 to 4, respectively, for seven values of $\lambda = 0.01, 0.09, 0.17, 0.26, 0.34, 0.42, 0.5$ (from top to bottom); each geodesic starts from the blue GBM tumor and ends at the red one, and is sampled with six intermediate points. In each geodesic, shape components are drawn as 2D tumor contours with overlaid colors representing intensity. As seen in the top row of each figure, when λ is small, the deformation is primarily driven by the shape component. Here, geometric features of the contours, e.g., protrusions, are in correspondence across the two shapes and this is reflected in the resulting geodesic. As λ increases, the influence of the intensity component on the geodesic deformation becomes apparent. For instance, in the bottom row of each figure, we see that the deformation is mostly driven by the correspondence of high/low intensity values.

3.2. Hierarchical clustering and survival analysis

Next, we consider the entire dataset consisting of 63 subjects; for each subject, we have GBM tumor contour+intensity data from T1 and FLAIR MRIs. We first computed ten 63×63 pairwise shape+intensity distance matrices using Eq. (1), corresponding to ten equally spaced values of λ between 0.01 and 0.5. We then applied hierarchical clustering, with complete linkage, to each of the ten distance matrices and split the data into two clusters, with the aim of discovering a value of λ that provides large separation between the clusters in terms of median survival.

First, in Fig. 8a, we evaluated the absolute difference in median survival between the two clusters for each value of λ. For the T1 modality, the largest absolute difference of approximately 6.5 months occurs for a relatively small value of $\lambda = 0.12$, i.e., where shape plays the dominant role in the comparison of GBM tumors. For the FLAIR modality, the largest absolute difference of approximately 8 months occurs at a relatively large value of $\lambda = 0.5$, i.e., where intensity plays the dominant role in the comparison of GBM
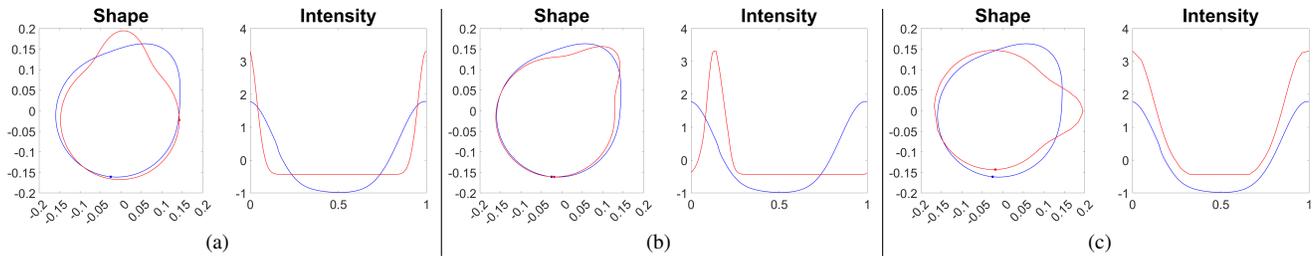
Figure 2. Registration of two simulated contour+intensity objects. (a) Standardized contour+intensity for two simulated examples. Registration results: (b) $\lambda = 0.01$, (c) $\lambda = 0.5$. Shape plots: the x- and y-axes are used to visualize the x- and y-coordinates of the shape component. Intensity plots: the x-axis shows the domain of the intensity function, while the y-axis captures the magnitude of intensity values.
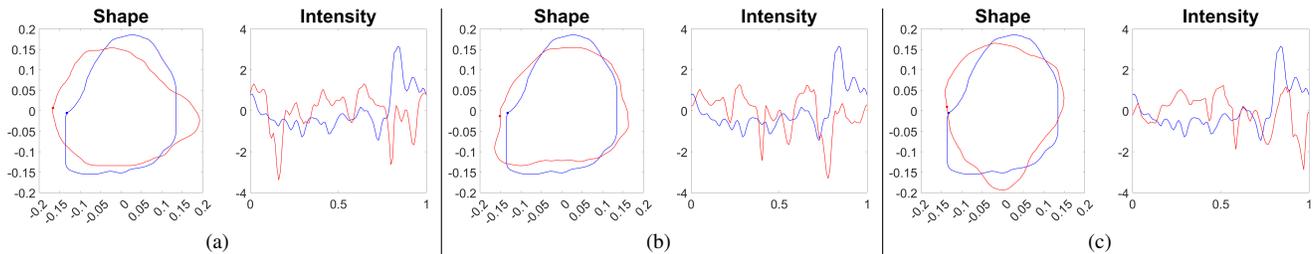
Figure 3. Registration of two GBM tumor contour+intensity representations from T1 (left: contours; right: intensity functions). (a) Standardized contour+intensity for two GBM tumors. Registration results: (b) $\lambda = 0.01$, (c) $\lambda = 0.5$. Shape plots: the x- and y-axes are used to visualize the x- and y-coordinates of the shape component. Intensity plots: the x-axis shows the domain of the intensity function, while the y-axis captures the magnitude of intensity values.

tumors. Next, we assessed balance of sample sizes in the estimated clusters by computing the absolute difference between the number of patients in cluster 1 and the number of patients in cluster 2. As shown in Fig. 8b, for T1 MRI data, the two estimated clusters are fairly balanced when $\lambda = 0.06, 0.12, 0.17$; for FLAIR, the clusters are balanced when $\lambda = 0.06, 0.5$. Note that the clusters are balanced in terms of sample size for $\lambda = 0.12$ in T1 and $\lambda = 0.5$ in FLAIR, suggesting that the corresponding largest differences in median survival discovered based on the proposed shape+intensity representation of GBM tumors are reliable statistically; whether or not the balanced sample sizes between groups of patients with distinct survival are clinically relevant, however, needs to be further investigated. In addition, the survival differences based on the estimated clusters are quite large, since median survival time in GBM is only 12 months [9].

In Fig. 9, we show clusterwise Kaplan-Meier estimates of survival probabilities for clusters estimated based on (a) T1 data and $\lambda = 0.12$, and (b) FLAIR data and $\lambda = 0.5$, i.e., λ values that yield the largest absolute difference in median survival. In both cases, patients placed in cluster 2 had lower survival probabilities across most of the study period. Particularly, when clustering is based on the FLAIR data, the difference in survival probabilities between the two clusters is larger across the study period, as compared to

clustering based on T1 data. This suggests that the FLAIR MRI-derived tumor shape and intensity can be more useful in identifying two groups of patients with distinct survival patterns.

Figure 10 shows MDS plots of the shape+intensity distance matrices based on (a) T1 data and $\lambda = 0.12$, and (b) FLAIR data and $\lambda = 0.5$. Each observation is colored according to its cluster membership. In both panels, we see that there is reasonable separation between the two estimated clusters based on the proposed distance, although this is not as clear based on the FLAIR data. These plots provide some evidence that the clustering results based on the proposed representation and distance are reliable. However, the MDS coordinates are computed using distances only and do not incorporate survival information.

Finally, Table 1 reports three-fold cross-validation results for the ten equally spaced values of λ between 0.01 and 0.5. In this cross-validation setup, the subjects' cluster memberships for different values of λ remain the same as a result of hierarchical clustering based on the entire dataset. We randomly split the data into three folds, and use two training folds to find the value of λ that yields the largest difference in median survival between the two clusters. Then, we evaluate the difference in median survival and (clusterwise) sample sizes in the remaining testing fold for that λ value. We repeat this procedure three times by considering
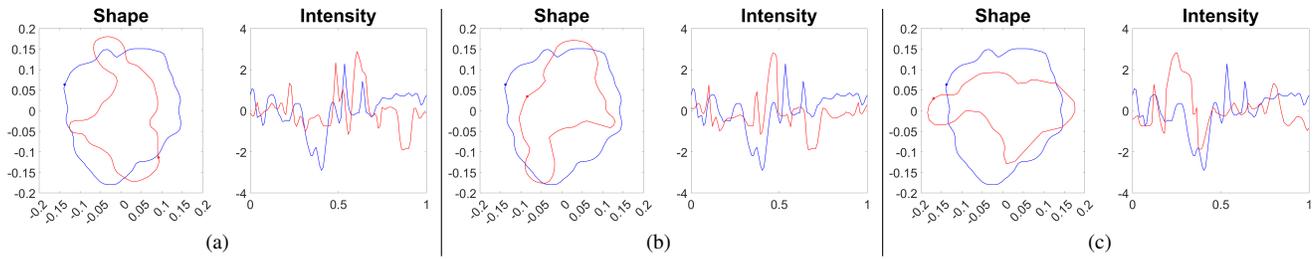
Figure 4. Registration of two T1 GBM tumor contour+intensity representations for patients with shortest (less than 1 month; red) and longest survival (more than 50 months; blue). (a) Standardized contour+intensity for two GBM tumors. Registration results: (b) $\lambda = 0.01$, (c) $\lambda = 0.5$. Shape plots: the x- and y-axes are used to visualize the x- and y-coordinates of the shape component. Intensity plots: the x-axis shows the domain of the intensity function, while the y-axis captures the magnitude of intensity values.
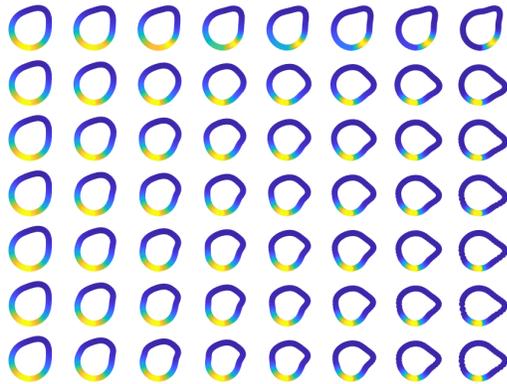
Figure 5. Shape+intensity geodesics (along rows) between the two simulated contour+intensity objects considered in Figure 2 for different values of λ. Top row to bottom row: $\lambda = 0.01, 0.09, 0.17, 0.26, 0.34, 0.42, 0.5$.
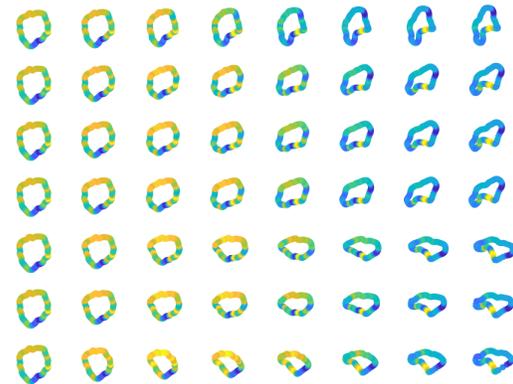
Figure 7. Shape+intensity geodesics (along rows) between the two GBM tumors considered in Figure 4 for different values of λ. Top row to bottom row: $\lambda = 0.01, 0.09, 0.17, 0.26, 0.34, 0.42, 0.5$.

Figure 8. Absolute difference (y-axis) in (a) median survival, and (b) sample size between estimated clusters for different values of λ (x-axis). Results based on T1 data are shown in blue while results based on FLAIR data are shown in red.

Figure 6. Shape+intensity geodesics (along rows) between the two GBM tumors considered in Figure 3 for different values of λ. Top row to bottom row: $\lambda = 0.01, 0.09, 0.17, 0.26, 0.34, 0.42, 0.5$.

each of the folds as a testing fold. As evident in the table, these results corroborate our findings in the earlier analysis based on the entire dataset. For the FLAIR data, the optimal value of λ estimated using the training folds is 0.5

in each of the three cases. The resulting median survival differences across clusters evaluated using the left out testing data are large and comparable (5.55, 9.00 and 6.05); the sample sizes across clusters in the testing fold are also fairly balanced in each case. For the T1 data, the optimal value of λ is estimated to be 0.12 based on the second and third training folds yielding median survival differences across clusters evaluated using the left out testing data of 9.10 and 0.60; the sample sizes across clusters in the corresponding
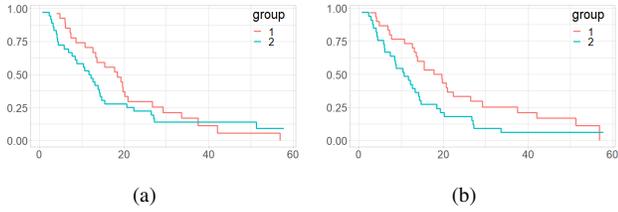
Figure 9. Clusterwise Kaplan-Meier survival estimates, for λ values that yield the largest difference in median survival, based on (a) T1 data, and (b) FLAIR data. The study period time (in months) is shown on the x-axis, while the Kaplan-Meier estimates are on the y-axis.

Figure 10. MDS plots, for λ values that yield the largest absolute difference in median survival, for (a) T1 data, and (b) FLAIR data. Each patient is colored according to cluster membership.

Modality	Optimal λ	Sample size difference across test clusters	Difference in median survival
T1	0.5	20	3.50
	0.12	0	9.10
	0.12	5	0.60
FLAIR	0.5	2	5.55
	0.5	4	9.00
	0.5	5	6.05

Table 1. Three-fold cross-validation results for ten equally spaced values of λ between 0.01 and 0.5.

Figure 11. Level curves for one T1 MRI-derived GBM tumor allowing the study of shape and intensity on and inside the tumour boundary.

testing folds are fairly balanced. Note, however, the small difference in median survival (0.60) in the third testing fold (bottom row of the T1 results in Table 1). This indicates that the choice of training and testing folds is critical as we have small sample sizes in each of the folds. Based on the first training fold, the estimated optimal value of λ is 0.5, but results in very unbalanced clustering on the corresponding testing fold. This again shows the limitation of the small sample size in our dataset. Nevertheless, these results provide some evidence that the clusterwise survival analysis results reported in this section are reliable. The results reported here can be strengthened further by repeating this analysis on a larger cohort of GBM patients.

The discrepancy between the clustering results based on T1 and FLAIR is expected since the tumor outlines, and associated intensities, are captured differently in each modality; they highlight properties of tissues differently. Tumors captured by FLAIR are generally more irregularly shaped, having more complex protrusions, and also have higher intensity compared to T1. Therefore, the same weight λ for the intensity component can lead to different clustering results.

4. Discussion

We proposed a framework that provides objective characterization of tumor shape and intensity and allows assessment of tumor heterogeneity among patients. The proposed representation combines and balances information of tumor shape and intensity along the tumor outline, estimated from T1 and FLAIR MRIs, which is found to be useful in identifying patients with different disease prognosis. The balance of information is achieved by scaling the intensity along the tumor contours to control its influence in tumor shape and intensity comparisons. The proposed representation and associated distance are invariant to translation, scale, rotation and reparameterization, which are desirable mathematical properties. The definition of distance then enables statistical analysis of GBM tumor shape+intensity, e.g., clustering. Besides the context of GBM, the proposed approach is general and can be applied to data from other medical imaging modalities.

We discovered that a certain emphasis on the intensity component is helpful in distinguishing patients with poor disease prognosis, in terms of survival, from those with good prognosis, as shown in the clustering and survival analysis results. Such information can be beneficial both for

patients and doctors to make medical decisions after GBM diagnosis, e.g., whether to undergo invasive surgery to remove the tumor.

The proposed method uses information about the shape of the tumor outline and the intensity along the outline. A natural direction for future work is to extend the proposed approach to additionally use information about the (shape of) level curves inside the tumor and the associated intensity values along these curves. Such a representation of a GBM tumor for a single patient is illustrated in Fig. 11. This extension would capture information regarding the spatial organization of intensity values inside the tumor, thus providing more information regarding GBM tumor heterogeneity. We additionally plan to develop further statistical tools for analyzing GBM tumor shape+intensity objects including summarization via the mean and exploration of variability through principal component analysis.

References

[1] Chaddad A, Desrosiers C, Hassan L, and Tanougast C. A quantitative study of shape descriptors from glioblastoma multiforme phenotypes for predicting survival outcome. *The British Journal of Radiology*, 89:1068, 2016. 1

[2] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, 1995. 3

[3] K. Bharath, S. Kurtek, A.U.K. Rao, and V. Baladandayuthapani. Radiologic image-based statistical shape analysis of brain tumours. *Journal of the Royal Statstical Society, Series C*, 67(5):1357–1378, 2018. 1, 2

[4] O. Gevaert, L.A. Mitchell, A.S. Achrol, J. Xu, S. Echegaray, G.K. Steinberg, S.H. Cheshier, S. Napel, G. Zaharchuk, and S.K. Plevritis. Glioblastoma multiforme: Exploratory radiogenomic analysis by using quantitative image features. *Radiology*, 273(1):168–174, 2014. 1

[5] E.C. Holland. Glioblastoma multiforme: The terminator. *Proceedings of the National Academy of Sciences*, 97(12):6242 – 6244, 2000. 1

[6] K. Krabbe, P. Gideon, P. Wagn, U. Hansen, C. Thomsen, and F. Madsen. MR diffusion imaging of human intracranial tumors. *Neuroradiology*, 39:483–489, 1997. 1

[7] S. Kurtek, A. Srivastava, E. Klassen, and Z. Ding. Statistical modeling of curves using shapes and related features. *Journal of the American Statistical Association*, 107:1152–1165, 2012. 2

[8] W. Liu, A. Srivastava, and E. Klassen. Joint shape and texture analysis of objects boundaries in images using a Riemannian approach. In *Asilomar Conference on Signals, Systems and Computers*, pages 833–837, 2008. 2

[9] R. McLendon, A. Friedman, D. Bigner, E.G. Van Meir, D.J. Brat, G.M. Mastrogianakis, J.J. Olson, T. Mikkelsen, N. Lehman, K. Aldape, and et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–1068, 2008. 1, 5

[10] W. Mio, A. Srivastava, and S.H. Joshi. On shape of plane elastic curves. *International Journal of Computer Vision*, 73(3):307–324, 2007. 2

[11] L.G. Nyúl and J.K. Udupa. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine*, 42(6):1072–1081, 1999. 2

[12] A. Saha, S. Banerjee, S. Kurtek, S. Narang, J. Lee, G. Rao, J. Martinez, K. Bharath, A.U.K. Rao, and V. Baladandayuthapani. DEMARCATE: Density-based magnetic resonance image clustering for assessing tumor heterogeneity in cancer. *Neuroimage Clinical*, 12:132–43, 2016. 1, 2

[13] A. Srivastava and E. Klassen. *Functional and Shape Data Analysis*. Springer, 2016. 2

[14] A. Srivastava, E. Klassen, S.H. Joshi, and I.H. Jermyn. Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:1415–1428, 2011. 2

[15] B. Tutt. Glioblastoma cure remains elusive despite treatment advances. *OncoLog*, 56(3):1 – 8, 2011. 1