Radiomic Deformation and Textural Heterogeneity (R-DepTH) Descriptor to Characterize Tumor Field Effect: Application to Survival Prediction in Glioblastoma

Marwa Ismail[ID], Prateek Prasanna, Kaustav Bera[ID], Volodymyr Statsevych, Virginia Hill, Gagandeep Singh, Sasan Partovi[ID], Niha Beig, Sean McGarry, Peter Laviolette, Manmeet Ahluwalia, Anant Madabhushi[ID], *Fellow, IEEE*, and Pallavi Tiwari[ID], *Member, IEEE*

Abstract—The concept of tumor field effect implies that cancer is a systemic disease with its impact way beyond the visible tumor confines. For instance, in Glioblastoma (GBM), an aggressive brain tumor, the increase in intracranial pressure due to tumor burden often leads to brain herniation and poor outcomes. Our work is based on the rationale that highly aggressive tumors tend to grow uncontrollably, leading to pronounced biomechanical tissue deformations in the normal parenchyma, which when combined with local morphological differences in the tumor confines on MRI scans, will comprehensively capture tumor field effect. Specifically, we present an integrated MRI-based descriptor, radiomic-Deformation and Textural Heterogeneity (r-DepTH). This descriptor comprises measurements of the subtle perturbations in tissue deformations throughout the surrounding normal parenchyma due to mass effect. This involves non-rigidly aligning the patients' MRI scans to a healthy atlas via diffeomorphic registration. The resulting inverse mapping is used to obtain the deformation field magnitudes in the normal parenchyma. These measurements are then combined with a 3D texture descriptor, Co-occurrence of Local Anisotropic Gradient Orientations (COLLAGE), which captures the morphological heterogeneity and infiltration within the tumor confines, on MRI scans. In this work, we extensively evaluated r-DepTH for survival risk-stratification on a total of 207 GBM cases from 3 different cohorts (Cohort 1 ($n1 = 53$), Cohort 2 ($n2 = 75$), and Cohort 3 ($n3 = 79$)), where each of these three cohorts was used as a training set for our model separately, and the other two cohorts were used for testing, independently, for each training experiment. When employing Cohort 1 for training, r-DepTH yielded Concordance indices (C-indices) of 0.7 and 0.65, hazard ratios (HR) and Confidence Intervals (CI) of 10 ($6 - 19$) and 5 ($3 - 8$) on Cohorts 2 and 3, respectively. Similarly, training on Cohort 2 yielded C-indices of 0.6 and 0.7, HR and CI of 1 ($0.7 - 2$) and 3 ($2 - 5$) on Cohorts 1 and 3, respectively. Finally, training on Cohort 3 yielded C-indices of 0.75 and 0.63, HR and CI of 24 ($10 - 57$) and 12 ($6 - 21$) on Cohorts 1 and 2, respectively. Our results show that r-DepTH descriptor may serve as a comprehensive and a robust MRI-based prognostic marker of disease aggressiveness and survival in solid tumors.

Index Terms—Glioblastoma, survival, field-effect, biomechanical deformations, LASSO.

I. INTRODUCTION

IT IS well recognized that cancer is not a bounded, self-organized system [1], but a systemic disease. Most malignant tumors have heterogeneous growth patterns, leading to disorderly proliferation well beyond the visible surgical margins. In fact, in solid tumors, depending on the malignant phenotype, the impact of the tumor is observed not just within the visible tumor, but also in the immediate peri-tumoral, as well as in seemingly normal-appearing adjacent regions [1], [2], a phenomenon known as 'tumor field effect' [3]. For instance, Glioblastoma (GBM), one of the most aggressive brain tumors, is known to extend several millimeters distal to the tumor margins, which ultimately leads to recurrence in

GBM patients [4]. Similarly, the herniation or gross distortion of the brainstem, remote to the tumor location, is the proximal cause of deaths in 61% of GBM studies [5]. The infiltrating brain tumor mass pushes and displaces the surrounding tissue structures (known as mass effect), leading to a mid-line shift and an increase in the intracranial pressure [6], [7], which ultimately results in destroying white matter tracts and alterations of consciousness and other chronic conditions such as seizures and headaches in GBM patients. While tumor infiltration [8], [9] and mass effect [10], [11] have both, to different extents, been shown to be associated with more aggressive tumor behavior and poor prognosis, it may be reasonable to assume that there might be latent disease-specific information to quantify both phenomena on routine imaging. Specifically, mass effect may be captured via the subtle tissue deformations in the seemingly-normal brain regions adjacent to tumor (also known as "brain around tumor (BAT)"). The rationale being that more aggressive tumors may exert increased intra-cranial pressure on the surrounding BAT regions, resulting in pronounced structural deformations and thus worse prognosis, as compared to less aggressive tumors. Additionally, the pronounced deformations, when combined with features from within the tumor confines that may quantify intra-tumoral heterogeneity and tumor infiltration, may serve as image-based prognostic markers of overall survival in GBM tumors.

"Recently, 'radiomics' (i.e. extraction of quantitative image features such as co-occurrence, gray-level dependence, directional gradients, and shape-based) has provided a surrogate mechanism to non-invasively capture sub-visual cues of intra-tumoral morphological heterogeneity on routine imaging, for different prognostic and predictive applications. Many studies in the literature have involved machine and deep-learning approaches using routine magnetic resonance imaging (MRI) sequences and have shown potential in survival prediction and response assessment for brain tumors [12]–[20]. Specifically, the recent works in [19], [20] attempted to segment the tumor sub-compartments using pre-operative scans, then extracted radiomic features from these sub-compartments and combined them with clinical information such as age and resection status, into machine learning algorithms to predict patient overall survival. These studies have largely focused on capturing local textural changes within the tumor and the peri-tumoral regions [21], and their associations with patient survival [22]. However, a missing gap in previous work has been to leverage the subtle tumor-induced deformations in the BAT regions as measured on routine MRI scans, as a complementary radiomic marker for prognosticating patient survival. In this context, our group developed an integrated MRI-based descriptor, that captures radiomic-Deformation and Textural Heterogeneity (r-DepTH) [23], which accounts for both the tumor-induced deformations in the BAT regions as well as the intra-tumoral heterogeneity from within its visible confines. Overall, the r-DepTH descriptor involves capturing phenotypic attributes of tumor infiltration as well as mass effect, as both of these aspects, to varying degrees, have been shown to be associated with worse outcomes in GBM tumors. This is achieved by computing Co-occurrence of Local Anisotropic Gradient

Orientations (COLLAGE) descriptor [24] from the tumor and the surrounding peritumoral regions to quantify intra-tumoral heterogeneity and tumor infiltration, as well as computing the local biomechanical deformations to quantify mass effect and its impact on the rest of the brain. This work expands on the original implementation of r-DepTH [23] in several aspects, including rigorous robustness analysis and registration strategies across different data cohorts and extensive comparisons with existing radiomic and deep-learning strategies, as well as clinical parameters, as shown throughout the paper.

This paper is organized as follows. In Section 2, previous work on characterizing field effect in survival prediction of GBM tumors using routine MRI scans is discussed. In Section 3, we describe the methodology for computation of the r-DepTH descriptor. The experimental design and implementation details for the risk assessment model are provided in Section 4, followed by results in Section 5. Section 6 provides discussion and concluding remarks.

II. PREVIOUS WORK AND NOVEL CONTRIBUTIONS

The concept of interrogating the tumor field using routine MRI scans has gained significant interest over the years, both to study its impact on tumor growth as well as correlating its impact on overall patient survival [25]–[27]. For instance, works have previously developed deterministic mathematical models that model cancerous growths from aggressive cellular proliferation in GBM tumors [25], [28], [29]. Through these models, studies investigated the induced significant mechanical stress on the surrounding tissue that results in mass effect in GBM tumors [25], [26]. These mathematical models consider the cellular motility factor in GBMs, to account for its invasiveness as well as the ability of its cells to migrate and proliferate [25]. A mechanically coupled model was also suggested in [27], to address the mechanical stress caused by tumor expansion, while also incorporating a diffusion coefficient that accounts for local tissue stress due to the field effect. In addition to these deterministic mathematical models, multiple studies have explored the utility of 'data driven' approaches, such as radiomic features extracted from GBM patients in survival prediction. For instance, [14] showed that radiomic features outperformed clinical and radiologic risk models in predicting overall survival in GBM tumors. Similarly, the studies conducted by [15], [30]–[32] assessed the utility of texture features extracted from the tumor and peritumoral regions for survival prediction in GBM. Our own group has developed a textural radiomic descriptor, COLLAGE [24], that has demonstrated success in capturing subtle differences between similar appearing disease phenotypes on routine imaging, across different types of malignant tumors [33], [34]. COLLAGE captures local anisotropic differences in intra-tumoral heterogeneity by calculating per-voxel gradient orientations, followed by obtaining statistics of Gray-Level Co-Occurrence Matrix (GLCM) heterogeneity to quantify patterns of local gradient alignment. However, those existing prognostic studies in GBM have been limited to interrogating texture representations from the enhancing lesion, inner necrotic core, and peritumoral area, and have not
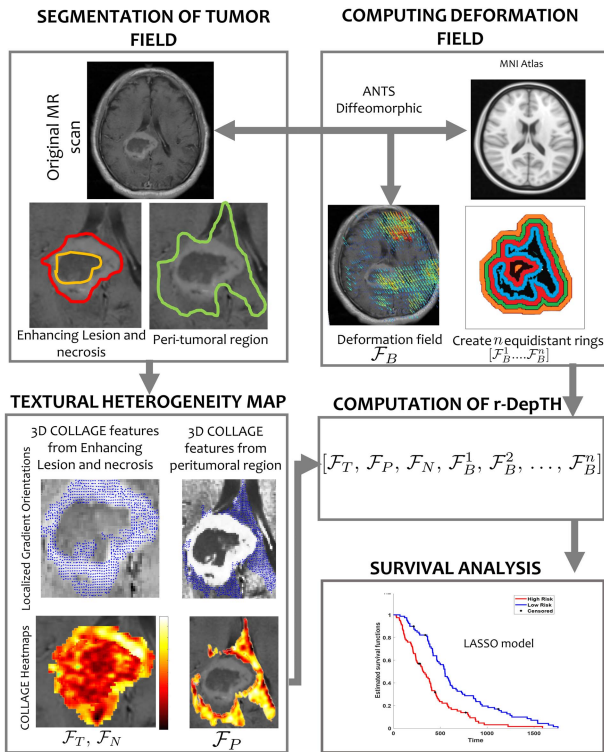
Fig. 1. Overview of r-DepTH framework. First, segmentation of tumor compartments of interest (enhancing lesion (outlined in red), peri-tumoral area (outlined in green), and necrotic core (outlined in orange)) is performed. Following pre-processing, feature extraction is performed via COLLAGE features to characterize the intra-tumoral textural heterogeneity, and deformation heterogeneity features to characterize the tumor impact on BAT region. Next, the sets of COLLAGE and deformation features are concatenated to compute the r-DepTH descriptor. The r-DepTH descriptor could then be employed for classification/survival analysis (in our case using a LASSO model for stratifying GBM patients into low- and high-risk groups based on their overall survival).

TABLE I
LIST OF THE NOTATIONS AND ACRONYMS USED IN THIS PAPER

Notation	Description	Notation	Description
I_T	intra-tumoral subvolume	I_P	peri-tumoral subvolume
I_N	necrotic subvolume	I_B	normal parenchyma subvolume
C_T	intra-tumoral region voxels	C_P	peri-tumoral region voxels
C_N	necrotic region voxels	C_{B_j}	normal parenchyma voxels from band j
I_{Atlas}	Healthy brain atlas	I_{mask}	tumor mask region
$\mathcal{F}_T$	COLLAGE features from intra-tumoral region	$\mathcal{F}_P$	COLLAGE features from peri-tumoral region
$\mathcal{F}_N$	COLLAGE features from necrotic region	$\mathcal{F}_{T,P,N}$	COLLAGE features from tumor compartments
$\mathcal{F}_B$	Deformation features	$\mathcal{F}_{rDepTH}$	r-DepTH descriptor features
$\theta(c), \phi(c)$	Principal orientations for COLLAGE	M^θ, M^ϕ	Co-occurrence matrices for COLLAGE
$\mathcal{F}_{rad}$	Comparative radiomic approach features	$\mathcal{F}_{DL}$	Deep features
S_t^{CCF}	CCF cohort as training set	S_v^{CCF}	CCF cohort as test set
S_t^{TCIA}	TCIA cohort as training set	S_v^{TCIA}	TCIA cohort as test set
S_t^{MCW}	MCW cohort as training set	S_v^{MCW}	MCW cohort as test set
Acronym	**Description**	**Acronym**	**Description**
TCIA	The Cancer Imaging Archive	CCF	Cleveland Clinic
MCW	Medical College of Wisconsin	S_t	Training set
S_v	Test set	$Risc$	Survival risk score
KM	Kaplan Meier	C-index	Concordance Index
HR	Hazard Ratio	CI	Confidence Interval
EOR	Extent of resection	GTR	Gross total resection
NTR	Near total resection	STR	Subtotal resection

explicitly accounted for any possible biomechanical deformational changes in the BAT region.

The key contribution of this work is the development of r-DepTH descriptor that combines measurements of biomechanical and morphological features of the tumor regions, for predicting patient survival. Briefly, we capture the biomechanical deformations in the BAT regions by using diffeomorphic registration between the GBM brain scans and a healthy atlas. We then utilize the inverse mapping during registration to calculate the deformation magnitudes within uniformly sized annular sub-volumes that belong to the surrounding BAT regions, both adjacent and distal to the tumor boundaries (as close as $5mm$ and up to $60mm$). In addition, we use 3D COLLAGE features [24] to capture the textural heterogeneity from within the tumor visible confines. The deformation features from the normal parenchyma and the COLLAGE features from the tumor confines, are finally combined to obtain the integrated r-DepTH descriptor. Figure 1 provides an overview of the r-DepTH framework.

III. METHODOLOGY

A. Notation

We define an image scene I as $I = (C, f)$, where I is a spatial grid of voxels $c \in C$, in a 3-dimensional

space, R^3. Each voxel, $c \in C$, is associated with an intensity value $f(c)$. I_T, I_P, I_N, and I_B correspond to the intra-tumoral, peri-tumoral, necrotic, and surrounding normal parenchymal sub-volumes within every I respectively, such that $[I_T, I_P, I_N, I_B] \subset I$. We further divide the sub-volume I_B into uniformly sized annular sub-volumes I_{B_j}, where j is the number of uniformly-sized annular bands, such that $j \in \{1, \ldots, m\}$, and m is a user-defined proximity parameter that is dependent on the distance from the tumor margin. We extract each feature set $\mathcal{F}$ from each training S_t and test S_v set, across the different cohorts employed in this study. The common notations and acronyms employed in this paper are listed in Table I.

B. r-DepTH Descriptor

1) Deformation Heterogeneity Features From the Normal Parenchyma: Healthy T1-w MNI atlas (I_{Atlas}) is used to measure the tissue deformation in the normal appearing brain regions of every patient volume I. I_{Atlas} is first non-rigidly aligned to I using mutual-information-based similarity measure provided in ANTs (Advanced Normalization Tools) SyN (Symmetric Normalization) toolbox [35]. This toolbox is specifically used due to its proved efficiency in mapping brain images containing lesions into healthy templates [36].

It also efficiently handles the constrained cost-function masking approach, where the mapping within a tumor exclusive region is determined by the solution of the negative tumor mask region I_{mask}. We employ Lagrangian diffeomorphic-based registration [37], as it possesses symmetry properties required for a geodesic connecting two images in the space of diffeomorphic transformation that guarantees symmetry regardless of the chosen similarity measure [35]. We also wanted to ensure that only the intensity differences due to structural deformations are accounted for, during registration, while excluding the intensity differences within the tumor area. I_{mask} was hence removed from I during registration to I_{Atlas}. Given the reference (I) and floating (I_{Atlas}), the non-rigid alignment can be formulated as: $(I, I_{mask}) = Tr(I_{Atlas})$, where $Tr(.)$ is the forward transformation of the composite voxel-wise deformation field (including affine components) that maps the displacements of the voxels between the reference and floating volumes. This transformation also propagates the atlas brain mask (I_{atlas}) to the subject space, thereby skull stripping the subjects. As ANTs SyN satisfies the conditions of a diffeomorphic registration, an inverse $Tr^{-1}(.)$ exists, that successfully maps I to the I_{Atlas} space. This inverse mapping yields the tissue deformation of I with respect to I_{Atlas}, representing the deformations exerted on every $c \in C_B$, due to tumor mass effect. Considering (c'_t, c'_u, c'_v) as new voxel positions of I when mapped to I_{Atlas}, the displacement vector is given as $[\delta t, \delta u, \delta v]$ where vector $(c'_t, c'_u, c'_v) = (c_t, c_u, c_v) + (\delta t, \delta u, \delta v)$, and the magnitude of deformation is given by calculating the Euclidean norm of the scalar values of the deformation orientations as:

$$D(c) = \sqrt{(\delta t)^2 + (\delta u)^2 + (\delta v)^2}, \qquad (1)$$

for every $c \in C_{B_j}$, and $j \in \{1, \ldots, m\}$. First order statistics (i.e. mean, median, standard deviation, skewness, and kurtosis) are then calculated by aggregating $D(c)$ for every c within every sub-volume I_{B_j} yielding a feature descriptor $\mathcal{F}_{B_j}$ for every annular sub-region C_{B_j}, where $C_{B_j} \subset C_B$, $j \in \{1, \ldots, m\}$.

2) 3D COLLAGE Features From Within the Tumor Confines: COLLAGE, a 3D gradient-based texture descriptor, captures intra-tumoral heterogeneity by calculating local per-voxel gradient orientations [24]. Briefly, for every voxel c, intensity gradients in X, Y, Z directions are calculated, followed by centering a $3D$ window around every c to compute the vector gradient matrix F. Then, two principal orientations, $\theta(c)$ and $\phi(c)$, can be obtained from F for every c, followed by computing two separate co-occurrence matrices, M^θ and M^ϕ, that capture orientation pairs between voxels in a local neighborhood. From every co-occurrence matrix, a total of 13 Haralick statistics are calculated for every c [38]. We finally obtain first order statistics (mean, median, standard deviation, skewness, and kurtosis) for every $c \in \{C_P, C_T, C_N\}$, which yields a feature descriptor $\mathcal{F}_T$ for the enhancing lesion, $\mathcal{F}_P$ for the T2/FLAIR hyperintensities, and $\mathcal{F}_N$ for the necrotic areas. Additional details regarding COLLAGE computation can be found in [24].

3) Construction of r-DepTH Descriptor: The r-DepTH descriptor is obtained for every patient, by concatenating the deformation descriptor, $\mathcal{F}_B$, along with COLLAGE texture descriptors, $\mathcal{F}_T$, $\mathcal{F}_P$, and $\mathcal{F}_N$, as $\mathcal{F}_{rDepTH} = [\mathcal{F}_B, \mathcal{F}_T, \mathcal{F}_P, \mathcal{F}_N]$. $\mathcal{F}_{rDepTH}$ descriptor can then be employed within a supervised or an unsupervised approach for disease characterization. The algorithm for computing r-DepTH is provided in Algorithm 1.

Algorithm 1 Computation of r-DepTH Descriptor

Data: I, I_{Atlas}, ROI Volume C
Result: $\mathcal{F}_{rDepTH}$

begin
  1- Obtain Deformation Features $\mathcal{F}_B$
  for *each patient volume I* **do**
    Remove I_{mask} from I, align I_{Atlas} to I to get $(I, I_{mask}) = Tr(I_{Atlas})$
  end
  for *each $c \in C_B$* **do**
    Get the deformation of I w.r.t. I_{Atlas}, $[\delta t, \delta u, \delta v]$, using $Tr^{-1}(.)$.
    Get deformation magnitude $D(c) = \sqrt{(\delta t)^2 + (\delta u)^2 + (\delta v)^2}$.
  end
  for *each annular sub-region $C_{B_j} \subset C_B$* **do**
    Aggregate $D(c)$ for every c within sub-volume I_{B_j}.
    Calculate first order statistics for I_{B_j} to get $\mathcal{F}_{B_j}$.
  end
  2- Obtain 3D COLLAGE Features $\mathcal{F}_T, \mathcal{F}_P$
  for *each voxel $c \in C$* **do**
    Obtain gradients $\partial f_i(c)$ along i-axes, $\partial f_i(c) = \frac{\partial f(c)}{\partial i}$, $i \in x, y, z$.
  end
  Define $\mathcal{N} \times \mathcal{N} \times \mathcal{N}$ neighborhood centered around $c \in C$.
  for *each voxel $c \in C$* **do**
    Calculate gradient vectors $\overrightarrow{\partial f_i}(c_k)$ in $\mathcal{N} \times \mathcal{N} \times \mathcal{N}$, $i \in (x, y, z)$, $k \in \{1, \ldots, \mathcal{N}^3\}$, where
    $\partial f_i(c) = [\partial f_i(c_1) \ \partial f_i(c_2) \ . \ \partial f_i(c_{N^3})]^T$
    Obtain localized gradient vector matrix
    $\overrightarrow{\mathcal{F}} = [\overrightarrow{\partial f_X}(c_k) \ \overrightarrow{\partial f_Y}(c_k) \ \overrightarrow{\partial f_Z}(c_k)]$
    Calculate dominant components $\psi(c_{k11})$, $\psi(c_{k12})$, and $\psi(c_{k13})$, $k \in \{1, \ldots, \mathcal{N}^3\}$, by singular value decomposition of $\overrightarrow{\mathcal{F}}$
    Obtain dominant directions $\theta(c_k)$ and $\phi(c_k)$, using
    $\theta^{3D}(c_k) = \tan^{-1} \frac{\psi_Y(c_k)}{\psi_X(c_k)}$ and
    $\phi^{3D}(c_k) = \tan^{-1} \frac{\psi_Z(c_k)}{\sqrt{\psi_Y^2(c_k) + \psi_X^2(c_k)}}$
  end
  Compute co-occurrence matrices M^θ and M^ϕ from $\theta(c_k)$ and $\phi(c_k)$
  for *each voxel $c \in \{C_P, C_T\}$* **do**
    Get 13 Haralick statistics $[S_{\theta_b}, S_{\phi_b}]$, $b \in [1, 13]$ from M^θ and M^ϕ
    Calculate first order statistics to get $\mathcal{F}_T, \mathcal{F}_P, \mathcal{F}_N$
  end
  3- Compute r-DepTH $\mathcal{F}_{rDepTH}$
  Concatenate $\mathcal{F}_B$, $\mathcal{F}_T$, $\mathcal{F}_P$, and $\mathcal{F}_N$ to get $\mathcal{F}_{rDepTH}$
end

IV. EXPERIMENTAL DESIGN

While the r-DepTH descriptor is generalizable, in this work, we calculated the $\mathcal{F}_{rDepTH}$ from routine pre-treatment MRI scans as a prognostic marker for overall survival in GBM subjects. Details on experimental design are provided below.

A. Data

A total of 207 cases were collected from 3 different cohorts for this study, including the publicly available Cancer Imaging Archive (TCIA) ([39]), Cleveland Clinic (CCF), and Medical College of Wisconsin (MCW). TCIA is an

TABLE II

DESCRIPTION OF PATIENT DEMOGRAPHICS FOR THE THREE COHORTS USED IN THIS STUDY

Group	TCIA	CCF	MCW
Number	75	53	79
Mean of age (Years)	59.6	59.5	61.5
Gender	45 males 30 females	36 males 17 females	39 males 40 females
OS (mean $\pm$ STD) (days)	467.4 $\pm$ 392.7	520.2 $\pm$376.6	468.7 $\pm$430.8
Censored Subjects	2	3	0
Extent of Resection	n=67 Gross total resection: 64 Biopsy: 3	n=53 Gross total resection: 16 Near total resection: 9 Neuroblate:7 Biopsy: 6 Subtotal Resection: 15	N/A
MGMT	n=38 Methylated: 23 Non-meth: 15	n=46 Methylated: 16 Non-meth: 30	N/A
IDH	n = 75 Wild Type: 73 Mutant: 2	n = 53 Wild Type: 47 Mutant: 6	N/A

open archive of radiological scans for different cancer types including GBM consisting of MRIs and its associated clinical metadata, with regulations and policies for the protection of human subjects and approvals by institutional review boards in place. Data analysis from CCF and MCW was approved by the institutional Ethics Committee. Our inclusion criteria included the availability of (1) 1.5 Tesla (T) routine MRI sequences (Gadolinium (Gd)-contrast T_{1w}, T_{2w}, T_{2w}-FLAIR) for treatment-naive patients with diagnostic image quality, and (2) Overall Survival (OS) information. This inclusion criteria yielded a total of (1) 75 GBM studies from TCIA, (2) 53 studies from CCF, and (3) 79 studies from MCW. In our experiments, we interchangeably used one of the three cohorts for training (S_t) and used the other two cohorts independently for testing (S_{v_1},S_{v_2}). Table II details the associated patient demographics for each of the 3 cohorts.

B. Preprocessing

A total of three experts were asked to perform the manual annotations on every MRI slice, via a hand-annotation tool in 3D Slicer. The senior-most expert (V.H, $>$10-years of experience in neuro-radiology) independently annotated the studies obtained from CCF, while expert 2 (V.S) with 7 years of experience in neuro-radiology supervised expert 3 (K.B, with 3 years of radiology experience), to manually annotate the TCIA and MCW cases. In rare cases with disagreement across the two readers (expert 2 and expert 3), the senior-most radiologist (V.H, expert 1) was consulted to obtain the final segmentations. Every tumor was annotated into three regions: enhancing lesion (I_T), T_{2w}/FLAIR hyperintense peri-lesional component (I_P), and necrotic core (I_N). T_{1w} MRI scans were used to delineate I_T and I_N, while both T_{2w} and FLAIR scans were used to annotate I_P. Following segmentation, for every

patient study, the 3 MRI sequences, Gd-T_{1w} MRI, T_{2w}, and FLAIR were co-registered to a brain atlas (MNI152; Montreal Neurological Institute) using ANTs (Advanced Normalization Tools) SyN (Symmetric Normalization) toolbox [35]. Skull stripping was performed simultaneously during registration of I with I_{Atlas}, as detailed in Section 3.2.1. Finally, bias field correction was conducted using a non-parametric non-uniform intensity normalization technique [40].

C. Implementation Details

We calculated the deformation features $\mathcal{F}_{B_j}$ for $j \in \{1, 2, \ldots, 12\}$ annular regions that are equidistant to each other at a distance of $5mm$. Specifically, we created concentrated annular rings around the tumor mask boundaries, using morphological operations, by dilating the tumor mask at several distances that were 5 mm apart for a total distance of 60 mm from the tumor mask boundaries, generating a total of 12 annular rings. The choice for the size of the annular rings was based on retrospective studies that have recommended $5mm$ as safe clinical target volume margin for GBM tumors [41]. This resulted in a 60×1 deformation vector that included 5 statistics (mean, median, standard deviation, skewness, and kurtosis) calculated for each of the 12 bands. This resulted in $12 \times 5 = 60$ features corresponding to $\mathcal{F}_B$. In addition, the 5 statistics were similarly obtained for each of the 13 Haralick statistics across each of the two co-occurrence matrices, resulting in $13 \times 5 \times 2 = 130$ COLLAGE features that are extracted from each of the three compartments; $\mathcal{F}_T$, $\mathcal{F}_P$, and $\mathcal{F}_N$. The descriptor $\mathcal{F}_{rDepTH}$ was finally obtained by aggregating $\mathcal{F}_B$, $\mathcal{F}_T$, $\mathcal{F}_P$, and $\mathcal{F}_N$. Hence the $\mathcal{F}_{rDepTH}$ descriptor for every tumor included a total of 450 features.

D. Survival Risk Assessment

Following computation of $\mathcal{F}_\alpha$, where $\alpha = \{T, P, N, B, rDepTH\}$, feature selection (reduction) was conducted using least absolute shrinkage and selection operator (LASSO), along with a cox regression model [42] on S_t. We used LASSO to utilize its capability in reducing variance when shrinking features, while simultaneously not increasing the bias. The top features selected by LASSO model were then used to create a continuous survival risk score ($Risc$), calculated as:

$$Risc(\mathcal{F}_\alpha) = \sum_{g=1}^{A} q_g \mathcal{F}_\alpha^g \quad (2)$$

where A represents the number of selected imaging features from LASSO, $\mathcal{F}_\alpha^g$ is the gth feature for $\alpha = \{T, P, N, B, rDepTH\}$ and q_g is the respective coefficient. An observation was deemed censored if the patient withdrew from the study or there was no follow up available. All $Risc$ scores were obtained based on the threshold value provided by the fitted cox model, to stratify patients into high-risk and low-risk groups obtained from S_t. Log-rank test along with Kaplan Meier (KM) survival analysis were then performed to see how the survival rate varies between the two identified risk

groups. Additionally, performance measures such as hazard ratios (HR), 95% Confidence Interval (CI), and Concordance index (C-index) were obtained to assess the performance of our survival models. HR is defined as the risk of experiencing the event of interest at a time point [43], whereas CI measures the level of uncertainty about the point estimates [44]. C-index was calculated using R (v4.0), and is commonly used to validate the predictive ability of a survival model by calculating the probability of concordance between the predicted and the observed survival [45]. Finally, on S_v, the top features obtained from S_t were used to calculate $Risc$ for every patient, followed by the log-rank test to obtain the level of significance between the two identified risk groups (low-risk and high-risk).

E. Comparative Strategies

In order to evaluate the efficacy of r-DepTH for GBM survival prediction, we performed the following comparisons: (1) Employing clinical features (age, gender, tumor volume, molecular information), in uni-variate and multi-variate settings, (2) Evaluating $Risc(\mathcal{F}_{T,P,N})$ using $\mathcal{F}_T$, $\mathcal{F}_P$, and $\mathcal{F}_N$, (3) Evaluating $Risc(\mathcal{F}_B)$ using $\mathcal{F}_B$, (4) State-of-the-art radiomics [46], and convolutional neural network (CNN) [16] approaches previously used in the literature for GBM prognosis, and (5) hybrid approaches where (a) age and gender information were combined with $\mathcal{F}_{rDepTH}$ and (b) deep features obtained via a CNN approach were combined with $\mathcal{F}_{rDepTH}$.

Table III details the feature families for all the comparative approaches as well as the total number of extracted features from each feature family and the number of features that were selected by the LASSO survival model for survival prognostication. When employing the comparative radiomic approach on C_T, C_P, C_N regions using the CapTk software as described in [46], this resulted in a total of 4376 features ($\mathcal{F}_{Rad}$) for every tumor region in S_t. This was followed by feeding ($\mathcal{F}_{Rad}$) to our LASSO model to calculate $Risc(\mathcal{F}_{rad})$ for every patient.

Additionally, we compared the performance of the r-DepTH descriptor to a CNN model previously utilized in the literature [16], to predict survival in GBM. This model is similar to most of the models found in literature in the context of survival prediction in GBM [17], [18], where transfer learning has been exploited to stratify patients into risk groups in GBM. Specifically, we extracted deep features from the GBM patients using a pre-trained CNN model via transfer learning using the strategy presented in [16]. We extracted deep features from one single tumor image that included the largest tumor region for each subject across all cohorts. The CNN acted as a feature extractor only and the obtained deep features from the CNN were then fed into the LASSO model for risk stratification. The CNN contained 5 convolutional layers and 3 fully-connected layers. The model was trained on ImageNet database with predetermined weights that are summarized in Table III. The input to the model was the cropped tumor sub-regions from the MR scans, obtained from the slice that had the largest tumor area for every patient, followed by resizing the sub-regions to 224 × 224 with bicubic interpolation.

TABLE III
COMPARATIVE STRATEGIES TO R-DEPTH

Approach	Extracted Features/ Parameters	Final Features		
		S_t^{CCF}	S_t^{TCIA}	S_t^{MCW}
Radiomic-based	4376 features: maximum, minimum, variance, standard deviation, skewness, kurtosis, mean, and median of: Intensity histogram distributions, Texture, Shape, and Spatial features	18	20	15
CNN-based	4096 features: Weight decay = 5×10^{-4} Momentum = 0.9 Initial learning rate = 10^{-2}	24	30	26
Clinical	6 features: Uni-, Multi-variate settings of: Age, Gender, Tumor Volume, EOR, MGMT, IDH	-	-	-
COLLAGE	390 features: Mean, Median, Skewness, Standard deviation, Kurtosis of the 13 Haralick statistics for each tumor region	11	11	10
Deformation	60 features: Mean, Median, Skewness, Standard deviation, Kurtosis of the 12 annular bands of the BAT region	10	10	10

Deep features were computed by forward propagation, and extracted from the second-last fully-connected layer (similar to the implementation in [16]). A total of 4096 features ($\mathcal{F}_{DL}$) were extracted from each patient, which were finally fed to our LASSO model for survival prognostication, to compute $Risc(\mathcal{F}_{DL})$ for every patient. We also conducted a hybrid approach, where we combined $\mathcal{F}_{DL}$ with $\mathcal{F}_{rDepTH}$ into the survival model to assess if this combination will improve the performance of $\mathcal{F}_{rDepTH}$ in survival prognostication.

V. RESULTS

In order to assess the robustness of r-DepTH features with respect to intra-site variability, our experiments were set up such that the survival models were once trained on each data cohort and tested on the two other cohorts independently. Hence, we report the survival results for 1) training with CCF cohort (S_t^{CCF}) and testing on TCIA (S_v^{TCIA}) and MCW cohorts (S_v^{MCW}), 2) training with TCIA cohort (S_t^{TCIA}) and testing on CCF (S_v^{CCF}) and MCW cohorts (S_v^{MCW}), and 3) training with MCW cohort (S_t^{MCW}) and testing on TCIA (S_v^{TCIA}) and CCF cohorts (S_v^{CCF}).

A. Survival Risk Assessment Using r-DepTH

LASSO survival analysis while employing $Risc(\mathcal{F}_{rDepTH})$ yielded 10 features, listed in supplementary material, when employing each of the 3 cohorts for training (S_t^{TCIA}, S_t^{CCF}, and S_t^{MCW}). Interestingly, there were 5 COLLAGE features that were consistently picked by the regression model across the three training cohorts (e.g., skewness of correlation in enhancing lesion and median of correlation for necrotic core).
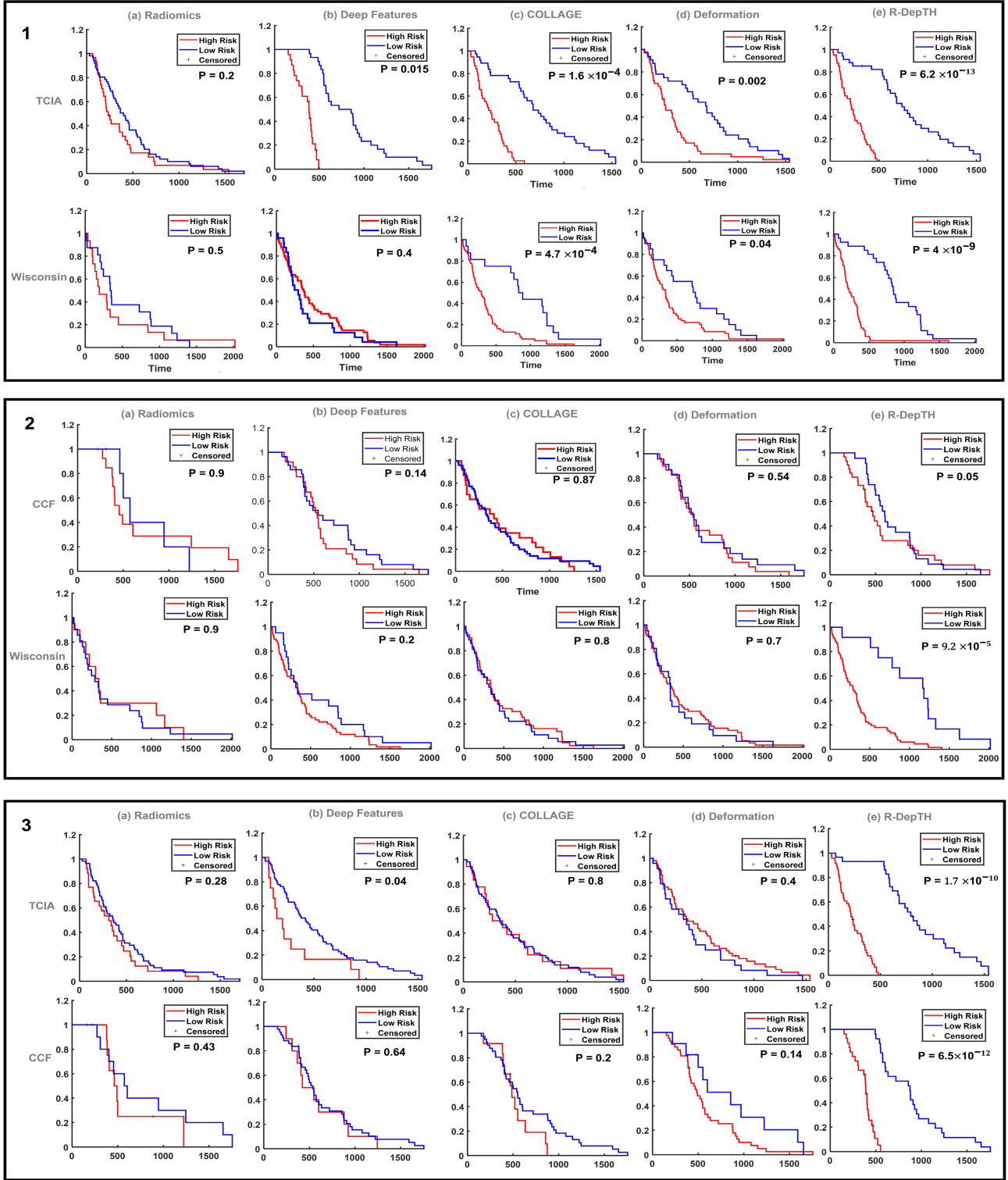
Fig. 2. Kaplan Meier curves for estimating overall survival for 1) TCIA (S_v^{TCIA}) and MCW (S_v^{MCW}) cohorts as independent test cohorts when using CCF as the training cohort (S_t^{CCF}), 2) CCF (S_t^{CCF}) and MCW (S_v^{MCW}) as independent test cohorts when using TCIA as the training cohort (S_t^{TCIA}), and 3) TCIA (S_v^{TCIA}) and CCF (S_v^{CCF}) as independent test cohorts when using MCW as the training cohort (S_t^{MCW}). Boxes 1, 2, 3 show the Kaplan Meier curves for estimating survival using (a) the comparative radiomic approach, (b) the DL approach, (c) COLLAGE features, (d) deformation features, and (e) r-DepTH features. X-axis represents the overall survival in days and Y-axis represents the estimated survival function.

The KM curves obtained for S_v^{TCIA} and S_v^{MCW} demonstrated significant differences between the two risk groups (Figure 2.1(e)), with p-value $= 6.2 \times 10^{-13}$ for S_v^{TCIA}

and p-value $= 4 \times 10^{-9}$ for S_v^{MCW}. Similarly, training the model with TCIA cohort (S_t^{TCIA}) resulted in significant differences between the two risk groups in S_v^{CCF} and S_v^{MCW}
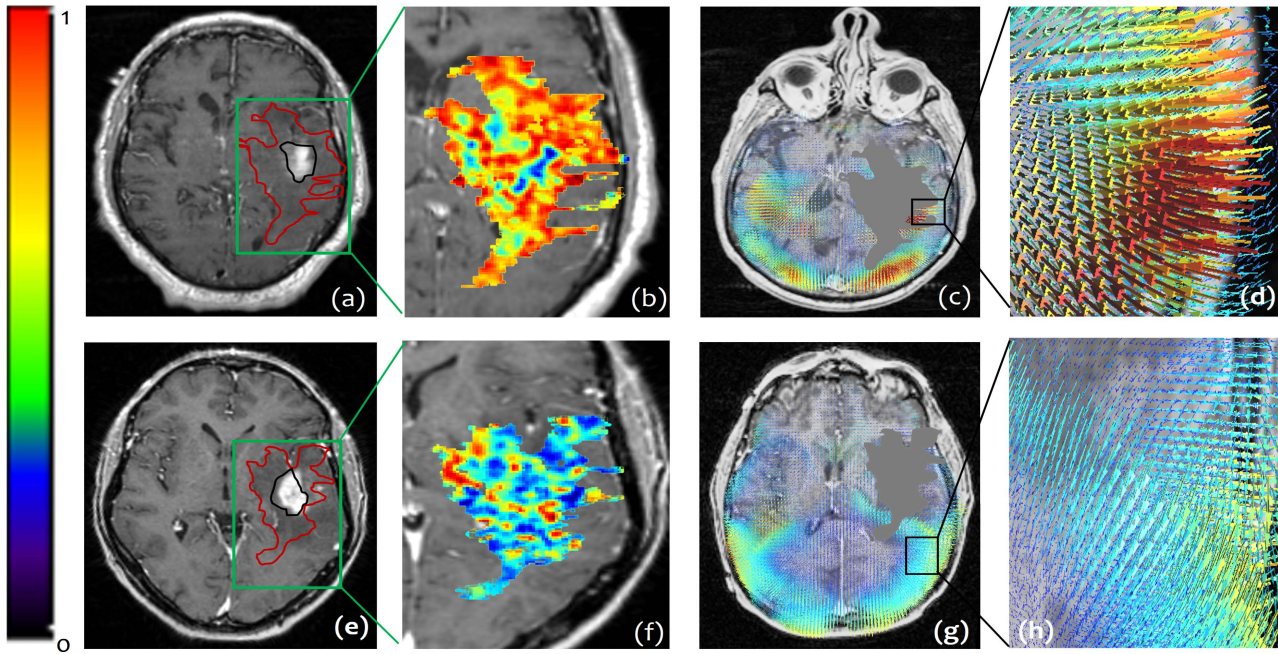
Fig. 3. Two subjects from S_t^{CCF}; a patient with poor survival (top row), OS = 30 days, and a patient with prolonged survival (bottom row), OS = 691 days. (a), (e) show the corresponding Gd-T1w MR scans of the two patients with their tumors segmented into 2 compartments: enhancing lesion (outlined in black) and peri-tumoral area (outlined in red). (b), (f) demonstrate the COLLAGE heatmaps generated for each of the two subjects, with higher values (red) being more prevalent in the patient with poor survival, compared to the patient with prolonged survival. (c), (g) illustrate the extracted deformation field magnitudes respectively for each of the two patients. For the patient with poor survival (d), higher magnitude values (represented in red) were observed in close proximity of the tumor, whereas lower values were observed (blue) for the patient with prolonged survival (h).

(Figure 2.2(e)), with p-values of 0.05 and 9.2×10^{-5} for S_v^{CCF} and S_v^{MCW}, respectively. Finally, training the model with MCW cohort (S_t^{MCW}) resulted in significant differences between the two risk groups in S_v^{TCIA} and S_v^{CCF} (Figure 2.3(e)), with p-values of 1.7×10^{-10} and 6.5×10^{-12} for S_v^{TCIA} and S_v^{CCF}, respectively. Additional measures for all these experiments are provided in Table IV.

Qualitative differences for both COLLAGE and deformation fields for two GBM subjects, one with poor survival (OS = 30 days) (top row) and one with prolonged survival (OS = 691 days) (bottom row), are presented in Figure 3. The patient with poor survival seemed to exhibit higher values of the COLLAGE feature (Kurtosis of Energy) (Figure 3(b)), compared to the patient with prolonged survival (Figure 3(f)). Similarly, the deformation field magnitudes for Skewness (measure of data asymmetry) at $10mm$ are visualized, and seem to reflect higher values for the patient with worse survival (Figure 3(d)), both in close proximity as well as distal to the tumor, compared to that for the patient with improved survival (Figure 3(h)).

B. Comparative Approaches

1) Risk-Scores Using Clinical Information: In a uni-variate setting, each of age, gender, and tumor volume did not demonstrate significant differences in the survival groups when training the survival model using S_t^{CCF}, S_t^{TCIA}, and S_t^{MCW}, Table IV. When combining age and gender in a multi-variate setting, significant differences were not observed between the

two risk groups across all the cohorts as well, Table IV. Additionally, when combining age and gender information with $\mathcal{F}_{rDepTH}$ into the survival model to assess if it is going to add value to our descriptor, results did not change when training the model using S_t^{CCF}. Interestingly, employing this experiment using S_t^{TCIA} and S_t^{MCW} increased the statistical significance between the two risk groups, Table IV. Additionally, we evaluated molecular markers including MGMT (available for 84 subjects), and IDH (available for 128 subjects), as well as extent of resection (EOR) (available for 120 subjects) for prognosis of GBM survival (the status of the subjects for which the EOR and MGMT were available are provided in the supplementary document). These clinical and molecular parameters unfortunately were not available for MCW and hence could not be evaluated. EOR and MGMT did not demonstrate significant differences between the two risk groups when training the survival model with subjects that had the available information (p-value = 0.13, C-index = 0.6 for EOR, p-value = 0.3, C-index = 0.52 for MGMT). Similarly, IDH did not demonstrate significant differences in the survival groups when training the model using S_t^{CCF} and S_t^{TCIA}, Table IV. Additionally, in order to control for EOR parameter during the survival analysis, we conducted an experiment where we trained the LASSO survival model with $\mathcal{F}_{rDepTH}$ using TCIA cohort cases that underwent GTR only (n = 64). The test set for this experiment was the CCF cases that underwent both NTR and GTR (n = 25). We excluded all the other cases with different EOR information (e.g., subtotal resection (STR)) from the analysis to ensure data balance, across the

TABLE IV

p-VALUE, CONCORDANCE INDEX, HAZARD RATIO, AND 95% CONFIDENCE INTERVAL FOR THE EXPERIMENTS CONDUCTED, ON THE INDEPENDENT TEST SETS FOR EACH OF THE 3 TRAINING COHORTS

	Feature	S_v^{CCF}, S_v^{MCW}			S_v^{TCIA}, S_v^{MCW}			S_v^{TCIA}, S_v^{CCF}		
		p-value	C-index	HR (95% CI)	p-value	C-index	HR (95% CI)	p-value	C-index	HR (95% CI)
Univariate	Age	0.9, 0.2	0.5, 0.59	1 (0.6 − 2), 1 (0.9 − 2)	0.2, 0.2	0.6, 0.53	1 (0.9 − 2), 1 (0.9 − 2)	0.2, 0.4	0.6, 0.5	1 (0.9 − 2), 1 (0.7 − 2)
	Gender	0.4, 0.42	0.54,0.54	0.7(0.5 − 1), 1 (0.5 − 1)	0.63, 0.7	0.53, 0.5	1 (0.7 − 2),2(0.6 − 0.9)	0.3, 0.25	0.56,0.59	1 (1 − 2), 1 (1 − 2)
	IDH	0.7, −	0.5, −	0.9 (0.9 − 1), −	0.8, −	0.5, −	1 (1 − 3), −	−, −	−, −	−, −
	Tumor Volume	0.12, 0.2	0.61,0.58	2 (2 − 2.4), 1 (1.5 − 2)	0.07, 0.7	0.52,0.46	2 (2 − 3), 0.9 (0.6 − 1)	0.3, 0.32	0.56,0.55	2 (1 − 2), 2 (1.5 − 2)
Multivariate	Age, Gender	0.75, 0.2	0.5, 0.59	1 (0.8 − 2), 1 (0.9 − 2)	0.3, 0.23	0.6, 0.55	1 (0.8 − 2), 1 (0.9 − 2)	0.2, 0.3	0.61,0.57	1 (0.9 − 2),2 (1 − 2)
	COLLAGE	0.87, 0.8	0.5, 0.5	1 (0.5 − 2), 1 (0.6 − 2)	1.6×10^{-4},4.7×10^{-4}	0.68,0.64	8 (4 − 14), 3 (2 − 4)	0.8, 0.2	0.52,0.59	0.9 (0.5 − 2),2 (0.8 − 4)
	Deformation	0.54, 0.7	0.6, 0.6	1 (0.7 − 2), 1 (0.5 − 2)	0.002, 0.04	0.63,0.6	2 (1 − 4), 2 (1 − 3)	0.4, 0.14	0.6, 0.57	0.8 (0.5 − 1),2(0.9 − 3)
	Radiomics	0.9, 0.9	0.5, 0.5	1 (0.4 − 3),0.5 (0.4 − 2)	0.2, 0.5	0.6, 0.5	1.4 (0.9 − 2),1(0.7 − 3)	0.28, 0.43	0.65,0.54	1 (0.8 − 2), 2 (0.6 − 6)
	DL	0.14, 0.2	0.6, 0.5	2 (0.6 − 8), 2 (0.8 − 3)	0.015, 0.4	0.63, 0.58	1(0.6 − 1.6),0.8(0.5 − 1	0.04, 0.64	0.65, 0.6	3 (1 − 6),1 (0.6 − 3)
	rDepTH	0.05, 9.2×10^{-5}	0.6, 0.7	1 (0.7 − 2), 3 (2 − 5)	6.2×10^{-13},4×10^{-9}	0.7, 0.65	10 (6 − 19), 5 (3 − 8)	1.7×10^{-10},6.5×10^{-12}	0.63,0.75	12 (6 − 21),24(10 − 57)
	r-DepTH, Age, Gender	9×10^{-4},7×10^{-6}	0.75, 0.7	3 (2 − 6), 3 (2 − 5)	6.2×10^{-13},4×10^{-9}	0.7, 0.65	10 (6 − 19), 5 (3 − 8)	2×10^{-11}, 2×10^{-11}	0.64,0.78	8 (4 − 14), 23 (9 − 55)
	rDepTH + DL	0.67, 0.56	0.6, 0.6	1 (0.6 − 3), 1 (0.5 − 1)	0.38, 0.62	0.61, 0.5	1(0.8 − 2), 0.8(0.4 − 2)	0.75, 0.86	0.6, 0.5	0.9(0.6 − 2), 1(0.6 − 2)

training and test sets. Results showed statistical significance on S_t^{TCIA} (p-value $= 1.5 \times 10^{-6}$, C-index $= 0.62$), but not on S_v^{CCF} (p-value $= 0.53$, C-index $= 0.56$). Similarly, to control for the methylation status while predicting survival, we conducted an additional experiment, where we trained the survival model with $\mathcal{F}_{rDepTH}$ using the cases from CCF cohort that were non-methylated (n $= 30$) and tested on non-methylated TCIA cases (n $= 15$). Results showed statistical significance on S_t^{CCF} (p-value $= 0.05$, C-index $= 0.64$), but not on S_v^{TCIA} (p-value $= 0.12$, C-index $= 0.54$).

2) Survival Risk-Assessment Using COLLAGE Features From the Tumor and Peri-Tumoral Regions: When training the LASSO model to obtain $Risc(\mathcal{F}_{T,P,N})$ using each of the 3 training cohorts, significant differences were observed when employing S_t^{CCF}, but not when employing S_t^{TCIA} or S_t^{MCW}. Specifically, the KM curves (Figure 2.1 (c)) obtained from $Risc(\mathcal{F}_{T,P,N})$ for S_t^{CCF} demonstrated statistically significant differences between the two risk groups for S_v^{TCIA} (p-value $= 1.6 \times 10^{-4}$) and S_v^{MCW} (p-value $= 4.7 \times 10^{-4}$). When employing S_t^{TCIA} to obtain $Risc(\mathcal{F}_{T,P,N})$, significant differences were not observed between the two risk groups for S_v^{CCF} (p-value $= 0.87$) and S_v^{MCW} (p-value $= 0.8$), Figure 2.2(c). Similarly, employing S_t^{MCW} to obtain $Risc(\mathcal{F}_{T,P,N})$ did not result in significant differences between the two risk groups for S_v^{TCIA} (p-value $= 0.4$) and S_v^{CCF} (p-value $= 0.14$), Figure 2.3(c). Additional measures for all these experiments are listed in Table IV.

3) Survival Risk Assessment Using Deformation Features From the Normal Parenchymal Regions: When training the LASSO model to obtain $Risc(\mathcal{F}_B)$ using each of the 3 training cohorts, significant differences were observed when employing S_t^{CCF}, but not when employing S_t^{TCIA} or S_t^{MCW}. The KM curves obtained for S_v^{TCIA} and S_v^{MCW} when employing S_t^{CCF} demonstrated significant differences between the two risk groups (Figure 2.1 (d)), with p-values of 0.002 and 0.04 for S_v^{TCIA} and S_v^{MCW}, respectively. Additionally, when training the LASSO model to obtain $Risc(\mathcal{F}_B)$ using S_t^{TCIA}, significant differences were not observed between the two risk groups for S_v^{CCF} (p-value $= 0.54$) and S_v^{MCW} (p-value $= 0.7$) (Figure 2.2(d)). Similarly, when training the LASSO model to obtain $Risc(\mathcal{F}_B)$ using S_t^{MCW}, significant differences were not observed between the two risk groups for S_v^{TCIA} (p-value $= 0.4$) and S_v^{CCF} (p-value $= 0.14$) (Figure 2.3(d)). Additional measures for these experiments are listed in Table IV.

a) Sensitivity analysis on the number of annular bands for deformation features extraction: As mentioned in Section IV.C, the choice for the size of the annular rings in our experiments (5mm) was based on retrospective studies that have recommended 5mm as a safe clinical target volume margin for GBM tumors [41]. However, for the sake of completeness, in the supplementary document, we provide a comparative analysis that shows the prognostication results using 20 annular bands that were 3mm apart from each other. Results of this experiment suggested that the measures derived from the 5mm band seem to lead to improved metrics (i.e., p-values, C-indices), as compared to the 3mm band measurements.

b) Extracting deformation features using another registration approach: We have also compared the performance of the deformation features obtained from ANTs to those obtained using another registration approach, called Greedy [47], to assess whether results obtained from the r-DepTH descriptor would be affected as a function of the registration approach. Our results demonstrated that both registration approaches generated similar measures (p-values and C-indices) suggesting that the r-DepTH descriptor may not specifically rely on a certain registration approach to obtain the deformation heterogeneity features. Results of this experiment are provided in the supplementary document.

Additionally, box-plots for the 2 most discriminative COLLAGE and deformation features on S_t^{CCF}, S_v^{TCIA}, and S_v^{MCW} are shown in Figure 4. The top deformation features included skewness at 10 mm as well as kurtosis at 15 mm. Similarly, the top 2 COLLAGE features were median of sum average and standard deviation of correlation.

4) Survival Risk Assessment Using Comparative Radiomic and CNN Approaches: $Risc(\mathcal{F}_{Rad})$ obtained from the LASSO model when employing S_t^{CCF} did not result in significant differences on S_v^{TCIA} (p-value $= 0.9$), or on S_v^{MCW} (p-value $= 0.5$), Figure 2.1 (a). Similarly, when training the LASSO model using S_t^{TCIA} (Figure 2.2 (a)) and using S_t^{MCW} (Figure 2.3(a)), significant differences were not observed between the two risk groups on any of the test cohorts. Table IV details all measures for these experiments.

For the adopted comparative CNN approach, the LASSO analysis to obtain $Risc(\mathcal{F}_{DL})$ using S_t^{CCF} demonstrated significant differences between the two risk groups for S_v^{TCIA}, (Figure 2.1 (b)), with p-value $= 0.015$. However, on S_v^{MCW}, significant differences were not observed between the two groups (Figure 2.1(b)), with p-value $= 0.4$. Similarly, when training the LASSO model using S_t^{TCIA}, significant differences were not observed between the two risk groups for S_v^{CCF} (p-value $= 0.14$) and for S_v^{MCW} (p-value $= 0.2$) (Figure 2.2 (b)). Interestingly, training the LASSO model using S_t^{MCW} demonstrated significant differences between the two risk groups for S_v^{TCIA} (p-value $= 0.04$), but not for S_v^{CCF} (p-value $= 0.64$) (Figure 2.3 (b)). Additional measures for these experiments are listed in Table IV.

Additionally, combining $\mathcal{F}_{DL}$ with $\mathcal{F}_{rDepTH}$ for survival prognostication did not result in statistically significant differences between the two risk groups across all experiments. Specifically, using S_t^{CCF}, a p-value of 0.38 was obtained on S_v^{TCIA} and a p-value of 0.62 was obtained on S_v^{MCW}. Similarly, when using S_t^{TCIA}, a p-value of 0.67 was obtained on S_v^{CCF} and a p-value of 0.56 was obtained on S_v^{MCW}. Similarly, p-values of 0.75 and 0.86 were obtained for S_v^{TCIA} and S_v^{CCF} when training the LASSO model using S_t^{MCW}. The poor results of this hybrid approach could be on account of the poor results obtained from some of the experiments that employed $\mathcal{F}_{DL}$ alone. Additional measures for these experiments are reported in Table IV.

VI. Discussion

Highly aggressive tumors such as Glioblastoma (GBM) tend to proliferate way beyond their visible tumor confines on
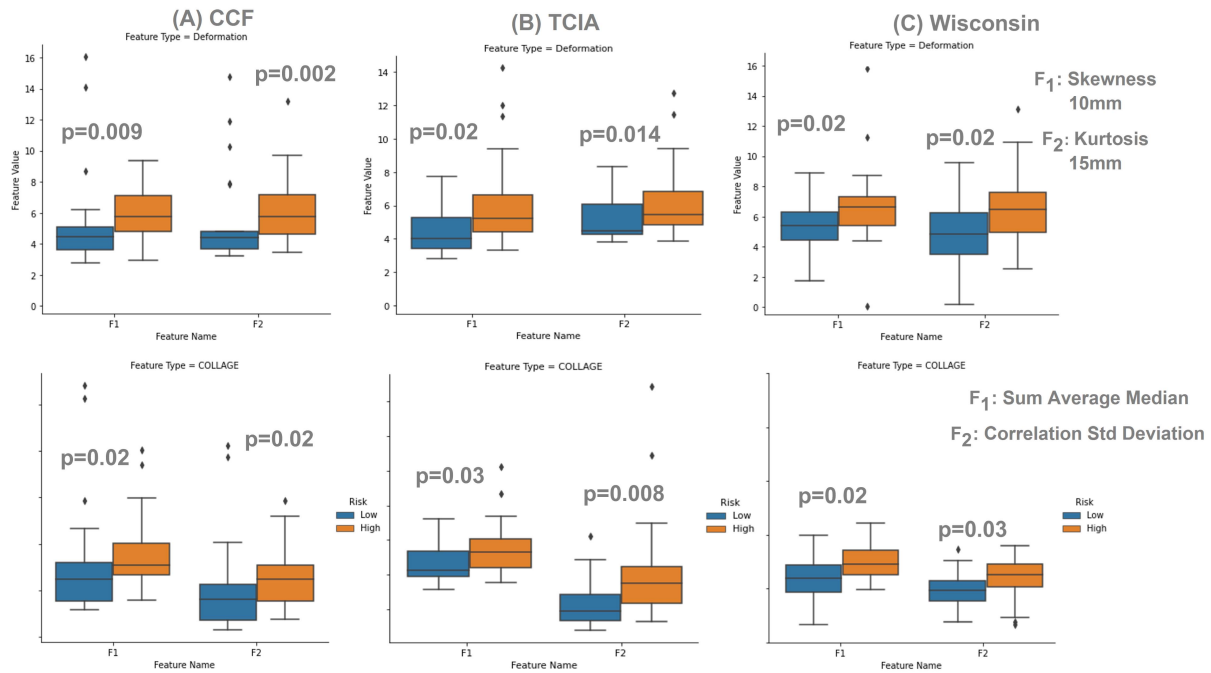
Fig. 4. Box plots of four statistically significantly different features between the high-risk and low-risk patients for (A) CCF cohort used for training (S_t^{CCF}), (B) TCIA cohort used for testing (S_v^{TCIA}), and (C) MCW cohort used for testing (S_v^{MCW}). The top row shows 2 deformation features and their p-values for S_t^{CCF} (A), S_v^{TCIA} (B), and S_v^{MCW} (C). The first feature is skewness, a measure of data symmetry, at 10 mm and the second one is kurtosis, a measure of the extreme values of a dataset, at 15 mm. The bottom row shows 2 COLLAGE features and their p-values for S_t^{CCF} (A), S_v^{TCIA} (B), and S_v^{MCW} (C). The first feature is median of sum average, a measure of the mean of the gray level sum distribution of the image, and the second one is standard deviation of correlation, a measure of the linear dependency of gray levels on those of neighboring voxels. The high-risk group is in orange, whereas the low-risk group is in blue.

routine MRI scans. For instance, GBM tumors are known to displace the surrounding tissue structures (phenomenon known as mass effect), which often causes herniation in the normal brain around tumor (BAT) parenchymal regions. Additionally, extensive brainstem infiltration in GBM patients which displaces the surrounding structures, is known to lead to worse prognosis in end-stage patients [1], [3], [6]–[8]. In guided-surgery procedures, the additional use of 5-Aminolevulinic acid (5-ALA) has been explored in many studies to assess its impact on overall survival and if it can provide prognostic cues in GBM. For instance, the study in [48] was conducted to assess the impact of additional use of 5-ALA in intraoperative MRI (iMRI) assisted surgery of GBMs on overall survival, and a significant increase of extent of resection (EOR) was found when combining 5-ALA and iMRI compared to use of iMRI alone but could not find any correlations between increase of EOR and progression free survival or overall survival. Also, the study in [49] showed a small but significant increase in survival measures associated with the use of 5-ALA-guided surgical resection with high grade gliomas. Although promising, the use of 5-ALA is still being explored. In the context of routine MR imaging, while previous studies have employed radiomic (textural and shape representations) and deep-learned features obtained from within the visible peri-tumor confines alone [13], [14], [34], no work to our knowledge has explicitly attempted to exploit the quantifiable biomechanical deformation attributes from the BAT regions, as a complementary radiomic feature,

in conjunction with features from the tumor and peri-tumoral confines.

In this work, we presented r-DepTH, a radiomic descriptor that leverages both morphological and biomechanical attributes of the tumor regions, and employed it in the context of survival prognostication in GBM. This was achieved by combining measurements that capture subtle tissue deformation features occurring in the surrounding healthy BAT regions due to mass effect, with 3D COLLAGE descriptor, which measures local heterogeneity arising from tumor infiltration, via higher order statistics of local gradient tensors on a voxel-wise basis, from the tumor and peri-tumoral confines. This integrated feature set was then employed to predict overall survival in GBM tumors. The current work builds on the original implementation of r-DepTH [23] and provides substantial improvements in its performance and validation. Our previous work employed a linear discriminant analysis (LDA) classifier to obtain the top features to construct r-DepTH and classify every GBM patient as belonging to poor or improved outcome. In this work, the r-DepTH descriptor is evaluated for prognosticating overall patient survival in GBM tumors using the top features that were obtained using least absolute shrinkage and selection operator (LASSO), along with a cox regression model to stratify patients into low- or high-risk, based on their overall survival (as a continuous variable). We also included textural features from the necrotic core regions existing within the visible tumor confines, in addition to the other two compartments (enhancing lesion and peri-tumoral edema),

into r-DepTH descriptor features, for survival prognostication. Further, we conducted extensive comparative strategies to demonstrate improved performance of r-DepTH compared to other approaches in the context of survival prognostication in GBM (radiomic-feature based [46] as well as deep-feature based [16]). Additionally, we performed extensive comparative experiments to assess if the performance of r-DepTH would further improve, by combining r-DepTH features with deep features [16] into the survival model as well as with clinical features such as age and gender. We further evaluated the robustness of the deformation features within r-DepTH across two different registration approaches. Specifically, aside from ANTs registration approach [35] that we used in our experimental design, we evaluated the use of another registration approach, namely, Greedy [47], to obtain the deformations employed within r-DepTH to predict survival. Finally, our evaluation of the r-DepTH descriptor in [23] was limited to a total of $n = 79$ cases; 68 cases for training, and 11 cases for testing. In this work, we evaluated the efficacy of the r-DepTH descriptor on a total of $n = 207$ 1.5 Tesla (T) MRI studies (Gadolinium (Gd)-contrast T_{1w}, T_{2w}, T_{2w}-FLAIR), obtained from a multi-institutional cohort (with each cohort used once for training and the other two being used independently for testing ($n_1 = 53$, $n_2 = 75$, $n_3 = 79$)).

Previous radiomic studies have investigated extracting features from the intra-tumor and peri-tumoral boundaries, for survival prediction or improving disease diagnosis. For instance, the study in [21] exploited radiomic shape and texture features extracted from both intra- and perinodular regions (where annular rings were generated around the nodule), to differentiate between cancerous lung nodules and benign masses. Similarly, another study, [22], attempted to prognosticate survival in GBM patients using radiomic texture features extracted from the peritumoral brain parenchyma. The COLLAGE features employed within r-DepTH have previously demonstrated success in characterizing tumor heterogeneity from tumor and peritumoral regions to distinguish similar-appearing pathologies, as well as prognosticate survival, across different applications such as brain, lung, and prostate cancer [34], [50], [51]. In line with our findings, most of these works have reported improvement in their diagnostic/prognostic models with inclusion of both intra and peri-lesion textural features to characterize disease heterogeneity. Our work builds on these previous findings by combining deformation features from brain around tumor region with the textural features from the lesion and peri-lesional compartments, into our descriptor, to further improve the prognostic model for GBM survival. Results of our work from employing COLLAGE features or deformation features in a siloed manner, to prognosticate survival across different cohorts, demonstrated significant differences between the two risk groups when training the model with one cohort (CCF), but not when training with TCIA or MCW cohorts. Interestingly, combining both feature sets into r-DepTH descriptor yielded statistically significant differences across all training and test cohorts, which demonstrates the impact of the proposed descriptor in predicting survival as well as its generalizability across different sites, through the feature combination that

allows for capturing mass effect (via deformation features) as well as tumor infiltration (via COLLAGE features), with both phenomena being associated with more aggressive tumor behavior as well as poor prognosis.

Apart from radiomic features, a few DL approaches in literature have investigated survival prediction in GBM by interrogating features from the visible tumor confines [16]–[18]. When replicating one such model in [16] for GBM survival prognosis on our cohort, the results did not yield significant differences between the high-risk and low-risk groups on some of our experiments (e.g. p-value $= 0.4$, C-index $= 0.58$ when training with CCF set and testing on MCW cohort). This, we posit, could be on account of training the model on a single slice as compared to the entire 3D volume. Further, the scanner-specific variations in our multi-institutional cohort are likely impacting the results of the adopted CNN approach, similar to previous works demonstrating the same drawback when applying CNN approaches on data collected from multiple institutions [52]. Additionally, when the CNN model was combined with r-DepTH to prognosticate survival, some of our results were not significantly different between the two risk groups, which could be on account of the average performance of the DL approach alone. Also, when combining clinical information, e.g., age and gender, into our descriptor for survival analysis, results showed that this information might aid in improving survival prognostication.

The closest work to our work on exploring structural brain deformations was performed by [53], where structural deformations were obtained from different parcellations in the brain, which were then associated with overall survival in GBM patients. Decreased survival time was found to be associated with increased deformations in certain cognitive and motor-control brain areas. Uniquely, our study found statistically significant structural deformation changes around the tumor boundaries up to $55mm$, both in the training and the test sets, that contributed to the prognostic signature for distinguishing between high- and low-risk GBM patients. As shown in Figure 4, some of our top deformation features included skewness, an indicator of lack of data symmetry, at $10mm$. The higher skewness values exhibited by the high-risk group with poor survival (Figure 4) could be linked to the way such aggressive tumors proliferate and exert pressure on BAT, and hence may lead to more lopsidedness in the frequency distribution of the deformation magnitude values at these regions, compared to the group with prolonged survival. Kurtosis, a measure of the extreme values in a dataset, at $15mm$ also turned out to be a top prognostic feature of the two risk groups, where it showed higher values for the high-risk group (Figure 4). This could be on account of the higher heterogeneity of BAT in patients with poor survival, due to active proliferation and herniation beyond tumor confines, leading to higher deformation magnitudes with extreme values.

Our study did have some limitations. One limitation is that we did not explicitly account for direction (i.e phase) attributes of tissue deformations obtained in the BAT region. Further, while r-DepTH in this study was hypothesized to serve as a surrogate measure for the pressure/force exerted by the tumor, this can only be affirmed via controlled in-vivo

experiments in a pre-clinical setting. Lastly, molecular and clinical information (i.e IDH, MGMT, EOR) was not available for one of the three cohorts used, and hence could not be used to control for clinical parameters, molecular status, and EOR, while building our prognostic risk assessment models.

VII. CONCLUSION

In this work, we presented r-DepTH, an integrated radiomic descriptor that aimed at comprehensively characterizing the field effect from tumor, peri-tumor, and brain around tumor regions, towards predicting overall survival in glioblastoma patients. Our results suggest that combining measurements that quantify subtle biomechanical deformations from the brain around tumor, along with morphological features within the tumor and peri-tumor confines allowed for improved prognostic models for predicting overall survival in GBM, as compared to using clinical variables, as well as using radiomic and deep-learning features from the tumor confines alone. Future work will involve integrating the direction (i.e. phase) attributes of the tissue deformation along with deformation magnitude features to build an integrated prognostic signature of the tumor regions. Additionally, we plan to extend our analysis to a large multi-site retrospective cohort, and eventually to prospectively collected scans, for validation of r-DepTH as a prognostic marker for GBM and other solid tumors.

REFERENCES

[1] H. Chai and R. E. Brown, "Field effect in cancer—An update," *Ann. Clin. Lab. Sci.*, vol. 39, no. 4, pp. 331–337, 2009.

[2] J. S. K. Chan *et al.*, "Cancer-associated fibroblasts enact field cancerization by promoting extratumoral oxidative stress," *Cell Death Disease*, vol. 8, no. 1, p. e2562, Jan. 2018.

[3] P. Lochhead *et al.*, "Etiologic field effect: Reappraisal of the field effect concept in cancer predisposition and progression," *Mod. Pathol.*, vol. 28, no. 1, pp. 14–29, Jan. 2015.

[4] O. M. Salazar and P. Rubin, "The spread of glioblastoma multiforme as a determining factor in the radiation treated volume," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 1, nos. 7–8, pp. 627–637, Jul. 1976.

[5] D. L. Silbergeld, R. C. Rostomily, and E. C. Alvord, "The cause of death in patients with glioblastoma is multifactorial," *J. Neuro-Oncol.*, vol. 10, no. 2, pp. 179–185, 1991.

[6] J.-C. Tonn, M. Westphal, J. T. Rutka, and S. A. Grossman, *Neuro-Oncology of CNS Tumors*. Germany: Springer, 2006.

[7] E. Pan and M. Prados, "Spinal cord tumors," in *Holland-Frei Cancer Medicine*, 6th ed. Hamilton, ON, USA: BC Decker, 2003.

[8] M. R. Drumm *et al.*, "Extensive brainstem infiltration, not mass effect, is a common feature of end-stage cerebral glioblastomas," *Neuro-Oncol.*, vol. 22, no. 4, pp. 470–479, Apr. 2020.

[9] J. S. Cordova *et al.*, "Whole-brain spectroscopic MRI biomarkers identify infiltrating margins in glioblastoma patients," *Neuro-Oncol.*, vol. 18, no. 8, pp. 1180–1189, Aug. 2016.

[10] T. C. Steed *et al.*, "Quantification of glioblastoma mass effect by lateral ventricle displacement," *Sci. Rep.*, vol. 8, no. 1, pp. 1–8, Dec. 2018.

[11] M. Dallabona *et al.*, "Impact of mass effect, tumor location, age, and surgery on the cognitive outcome of patients with high-grade gliomas: A longitudinal study," *Neuro-Oncol. Pract.*, vol. 4, no. 4, pp. 229–240, Dec. 2017.

[12] K. Bera, N. Braman, A. Gupta, V. Velcheti, and A. Madabhushi, "Predicting cancer outcomes with radiomics and artificial intelligence in radiology," *Nature Rev. Clin. Oncol.*, vol. 19, pp. 1–15, Oct. 2021.

[13] R. Lambin *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[14] P. Kickingereder *et al.*, "Radiomic profiling of glioblastoma: Identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models," *Radiology*, vol. 280, no. 3, pp. 880–889, 2016.

[15] D. Yang, G. Rao, J. Martinez, A. Veeraraghavan, and A. Rao, "Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma," *Med. Phys.*, vol. 42, no. 11, pp. 6725–6735, Nov. 2015.

[16] J. Lao *et al.*, "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, Dec. 2017.

[17] L. Chato and S. Latifi, "Machine learning and deep learning techniques to predict overall survival of brain tumor patients using MRI images," in *Proc. IEEE 17th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2017, pp. 9–14.

[18] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 212–220.

[19] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," 2018, *arXiv:1811.02629*.

[20] S. Bakas *et al.*, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Sci. Data*, vol. 4, no. 1, pp. 1–13, Dec. 2017.

[21] N. Beig *et al.*, "Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas," *Radiology*, vol. 290, no. 3, pp. 783–792, Mar. 2019.

[22] P. Prasanna, J. Patel, S. Partovi, A. Madabhushi, and P. Tiwari, "Radiomic features from the peritumoral brain parenchyma on treatment-naïve multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings," *Eur. Radiol.*, vol. 27, no. 10, pp. 4188–4197, Oct. 2017.

[23] P. Prasanna *et al.*, "Radiographic-deformation and textural heterogeneity (r-DepTH): An integrated descriptor for brain tumor prognosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 459–467.

[24] P. Prasanna, P. Tiwari, and A. Madabhushi, "Co-occurrence of local anisotropic gradient orientations (CoLlAGe): A new radiomics descriptor," *Sci. Rep.*, vol. 6, no. 1, p. 37241, Dec. 2016.

[25] K. R. Swanson, C. Bridge, J. D. Murray, and E. C. Alvord, "Virtual and real brain tumors: Using mathematical modeling to quantify glioma growth and invasion," *J. Neurol. Sci.*, vol. 216, no. 1, pp. 1–10, Dec. 2003.

[26] H. L. Harpold, E. C. Alvord, Jr., and K. R. Swanson, "The evolution of mathematical modeling of glioma proliferation and invasion," *J. Neuropathol. Exp. Neurol.*, vol. 66, no. 1, pp. 1–9, 2007.

[27] D. A. Hormuth *et al.*, "A mechanically coupled reaction–diffusion model that incorporates intra-tumoral heterogeneity to predict *in vivo* glioma growth," *J. Roy. Soc. Interface*, vol. 14, no. 128, Mar. 2017, Art. no. 20161010.

[28] N. Gaw *et al.*, "Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019.

[29] S. Stein, R. Zhao, H. Haeno, I. Vivanco, and F. Michor, "Mathematical modeling identifies optimum lapatinib dosing schedules for the treatment of glioblastoma patients," *PLoS Comput. Biol.*, vol. 14, no. 1, Jan. 2018, Art. no. e1005924.

[30] O. Gevaert *et al.*, "Glioblastoma multiforme: Exploratory radiogenomic analysis by using quantitative image features," *Radiology*, vol. 273, no. 1, pp. 168–174, Oct. 2014.

[31] Y. Cui *et al.*, "Prognostic imaging biomarkers in glioblastoma: Development and independent validation on the basis of multiregion and quantitative analysis of MR images," *Radiology*, vol. 278, no. 2, pp. 546–553, Feb. 2016.

[32] S. D. McGarry *et al.*, "Magnetic resonance imaging-based radiomic profiles predict patient prognosis in newly diagnosed glioblastoma before therapy," *Tomography*, vol. 2, no. 3, p. 223, 2016.

[33] N. M. Braman *et al.*, "Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI," *Breast Cancer Res.*, vol. 19, no. 1, p. 57, Dec. 2017.

[34] R. Shiradkar *et al.*, "Radiomic features from pretreatment biparametric MRI predict prostate cancer biochemical recurrence: Preliminary findings," *J. Magn. Reson. Imag.*, vol. 48, no. 6, pp. 1626–1636, Dec. 2018.

[35] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008.

[36] A. Eloyan *et al.*, "Health effects of lesion localization in multiple sclerosis: Spatial registration and confounding adjustment," *PLoS ONE*, vol. 9, no. 9, Sep. 2014, Art. no. e107263.

[37] B. Avants, P. Schoenemann, and J. Gee, "Landmark and intensity-driven Lagrangian frame diffeomorphic image registration: Application to structurally and functionally based inter-species comparison," *Med. Image Anal.*, vol. 10, no. 3, pp. 397–412, 2006.

[38] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[39] K. Clark *et al.*, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

[40] N. J. Tustison *et al.*, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, p. 1310, Jun. 2010.

[41] A. R. Cabrera *et al.*, "Radiation therapy for glioblastoma: Executive summary of an American society for radiation oncology evidence-based clinical practice guideline," *Practical Radiat. Oncol.*, vol. 6, no. 4, pp. 217–225, Jul. 2016.

[42] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[43] P. C. Austin, P. Wagner, and J. Merlo, "The median hazard ratio: A useful measure of variance and general contextual effects in multilevel survival analysis," *Statist. Med.*, vol. 36, no. 6, pp. 928–938, Mar. 2017.

[44] M. J. Pencina and R. B. D'Agostino, "Overall c as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation," *Statist. Med.*, vol. 23, no. 13, pp. 2109–2123, 2004.

[45] H. Steck, B. Krishnapuram, C. Dehing-Oberije, P. Lambin, and V. C. Raykar, "On ranking in survival analysis: Bounds on the concordance index," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 1209–1216.

[46] C. Davatzikos *et al.*, "Cancer imaging phenomics toolkit: Quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome," *J. Med. Imag.*, vol. 5, no. 1, 2018, Art. no. 011018.

[47] P. A. Yushkevich, J. Pluta, H. Wang, L. E. Wisse, S. Das, and D. Wolk, "Fast automatic segmentation of hippocampal subfields and medial temporal lobe subregions in 3 Tesla and 7 Tesla T2-weighted MRI," *Alzheimer's Dementia*, vol. 7, no. 12, pp. P126–P127, 2016.

[48] J. Coburger, V. Hagel, C. R. Wirtz, and R. König, "Surgery for glioblastoma: Impact of the combined use of 5-aminolevulinic acid and intraoperative MRI on extent of resection and survival," *PLoS ONE*, vol. 10, no. 6, Jun. 2015, Art. no. e0131872.

[49] S. Gandhi *et al.*, "Survival outcomes among patients with high-grade glioma treated with 5-aminolevulinic acid–guided surgery: A systematic review and meta-analysis," *Frontiers Oncol.*, vol. 9, p. 620, Jul. 2019.

[50] P. Prasanna, A. Karnawat, M. Ismail, A. Madabhushi, and P. Tiwari, "Radiomics-based convolutional neural network for brain tumor segmentation on multiparametric magnetic resonance imaging," *J. Med. Imag.*, vol. 6, no. 2, 2019, Art. no. 024005.

[51] M. Khorrami *et al.*, "A combination of intra- and peritumoral features on baseline CT scans is associated with overall survival in non-small cell lung cancer patients treated with immune checkpoint inhibitors: A multi-agent multi-site study," *Proc. SPIE*, vol. 10950, Mar. 2019, Art. no. 109500R.

[52] W. Yan *et al.*, "MRI manufacturer shift and adaptation: Increasing the generalizability of deep learning segmentation for MR images acquired with different scanners," *Radiol., Artif. Intell.*, vol. 2, no. 4, Jul. 2020, Art. no. e190195.

[53] P. Prasanna *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.*, vol. 9, no. 1, pp. 1–13, Dec. 2019.