



OPEN

## Artificial intelligence-based locoregional markers of brain peritumoral microenvironment

Zahra Riahi Samani<sup>1,2,3</sup>✉, Drew Parker<sup>1,2,3</sup>, Hamed Akbari<sup>2,3</sup>, Ronald L. Wolf<sup>3</sup>, Steven Brem<sup>4</sup>, Spyridon Bakas<sup>2,3,5</sup> & Ragini Verma<sup>1,2,3</sup>

In malignant primary brain tumors, cancer cells infiltrate into the peritumoral brain structures which results in inevitable recurrence. Quantitative assessment of infiltrative heterogeneity in the peritumoral region, the area where biopsy or resection can be hazardous, is important for clinical decision making. Here, we derive a novel set of Artificial intelligence (AI)-based markers capturing the heterogeneity of tumor infiltration, by characterizing free water movement restriction in the peritumoral region using Diffusion Tensor Imaging (DTI)-based free water volume fraction maps. We leverage the differences in the peritumoral region of metastasis and glioblastomas, the former consisting of vasogenic versus the latter containing infiltrative edema, to extract a voxel-wise deep learning-based peritumoral microenvironment index (PMI). Descriptive characteristics of locoregional hubs of uniformly high PMI values are then extracted as AI-based markers to capture distinct aspects of infiltrative heterogeneity. The proposed markers are utilized to stratify patients' survival and *IDH1* mutation status on a population of 275 adult-type diffuse gliomas (CNS WHO grade 4). Our results show significant differences in the proposed markers between patients with different overall survival and *IDH1* mutation status (*t* test, Wilcoxon rank sum test, linear regression;  $p < 0.01$ ). Clustering of patients using the proposed markers reveals distinct survival groups (logrank;  $p < 10^{-5}$ , Cox hazard ratio = 1.82;  $p < 0.005$ ). Our findings provide a panel of markers as surrogates of infiltration that might capture novel insight about underlying biology of peritumoral microstructural heterogeneity, providing potential biomarkers of prognosis pertaining to survival and molecular stratification, with applicability in clinical decision making.

The tumor burden of diffuse gliomas extends beyond the radiographically visible border of tumor margin<sup>1-3</sup>. However, current clinical practice considers the tumor core (T1 contrast enhancing boundary) as the primary target of treatment<sup>4,5</sup>; therefore, infiltrative cancer in the peritumoral region (peritumoral T2 hyperintense tissue) remains un-resected which may result in tumor recurrence. Obtaining tissue biopsies can be hazardous in the peritumoral region due to the possibility of functional deficit<sup>4,5</sup>. As a result, characterization of the infiltrative heterogeneity in the peritumoral region is a critical need to inform clinical decision making. Previous attempts to characterize tumor infiltration applied various imaging modalities using manually delineated or heuristic-based labeling of infiltrative tissue<sup>6-12</sup>, but approaches using Diffusion Tensor Imaging (DTI) have been limited to clinically used measures<sup>7-10</sup>.

DTI is the Magnetic Resonance Imaging (MRI) modality that provides insight into tissue microstructure by measuring water diffusivity. In particular, the free water volume fraction, a measure of the amount of extracellular water<sup>13</sup>, is able to capture differences between infiltrative and vasogenic peritumoral regions by exploiting the variation in water movement restriction<sup>14-16</sup>. The overarching goal of this paper is to leverage this unique information of DTI and derive artificial intelligence (AI)-based markers to capture infiltrative heterogeneity using the extracellular water-based voxel-wise characterization of tissue in the peritumoral region.

By leveraging the differences in water diffusivity properties in the peritumoral region of brain metastases and glioblastomas (Central Nervous System (CNS) World Health Organization (WHO) grade 4,

<sup>1</sup>Diffusion & Connectomics In Precision Healthcare Research (DiCIPHR) Lab, University of Pennsylvania, Philadelphia, PA 19104, USA. <sup>2</sup>Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, PA 19104, USA. <sup>3</sup>Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. <sup>4</sup>Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. <sup>5</sup>Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. ✉email: zari@upenn.edu

Isocitrate-Dehydrogenase 1 (*IDH1*)-wildtype); consisting of purely vasogenic versus infiltrative edema respectively, we train a deep learning model to derive a novel voxel-wise peritumoral microenvironment index (PMI) without using any manual delineation of infiltrative regions. The PMI exploits characterization of water movement restriction in the voxels with and without infiltration and hence captures the infiltrative heterogeneity in the peritumoral region. We later apply connected component analysis to extract locoregional hubs of uniformly high PMI values, hypothesizing them as regions with high infiltration. Descriptive characteristics of locoregional hubs are calculated as AI-based markers of infiltrative heterogeneity, including number and size of hubs and the differences in their shape, direction, and spatial location.

The proposed AI-based markers are utilized for two clinical applications on a population of adult-type diffuse gliomas (CNS WHO grade 4) to demonstrate their potential in capturing distinct locoregional aspects of infiltrative heterogeneity beyond standard diffusion measures: (1) analysis of the duration of survival among *IDH1*-wildtype glioblastomas, to investigate whether higher infiltration of cancer cells, as measured by the proposed markers, is associated with shorter survival; and (2) differences across patients with varying *IDH1* mutation status (astrocytoma *IDH1*-mutant versus glioblastoma *IDH1*-wildtype), notably, the *IDH1*-wildtype tumors, when compared with *IDH1*-mutants, have poorer prognosis, and higher peritumoral infiltration<sup>17,18</sup>.

## Materials and methods

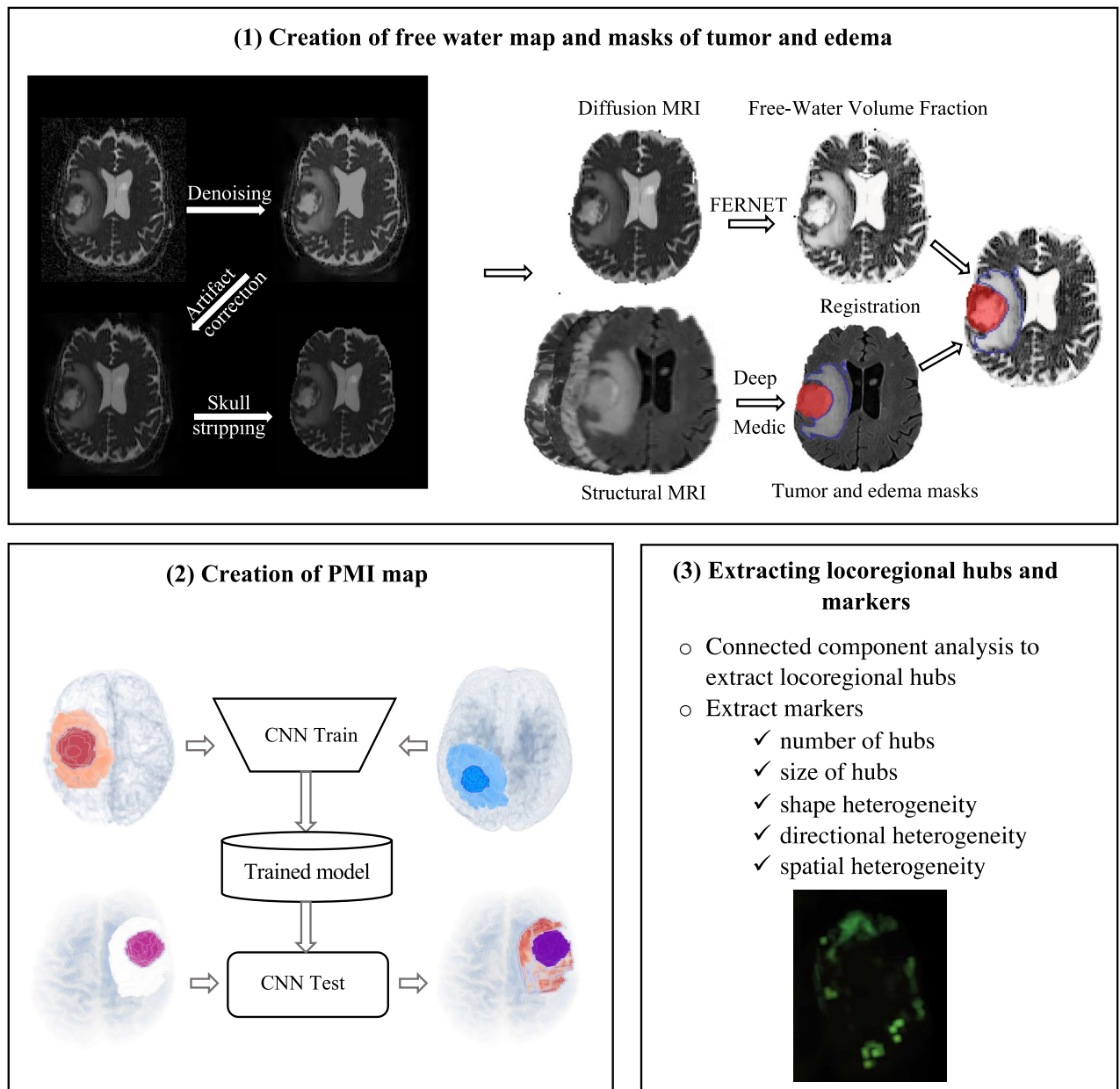
In this section, we describe the details of the datasets and steps of the pipeline for extracting the proposed AI-based markers. The pipeline consists of three steps. (i) creation of free water map and masks of tumor and edema, (ii) creation of the PMI map, (iii) extracting locoregional hubs and AI-based markers. Figure 1 demonstrates the overview of the pipeline. The proposed markers are evaluated on two clinical applications to investigate whether the differences in infiltrative heterogeneity are able to capture differences in overall survival and *IDH1* mutation status in adult-type diffuse glioma patients.

**Datasets.** This study was approved by the institutional review board of the University of Pennsylvania. Informed consent was obtained from all participants or their legally authorized representative. All methods were carried out in accordance with relevant guidelines and regulations. The population was identified based on retrospective review of the electronic medical record of patients diagnosed with adult-type diffuse glioma (CNS WHO grade 4). Study inclusion criteria were (i) histopathological tissue diagnosis of adult-type diffuse glioma (CNS WHO grade 4) or brain metastasis and preoperative; (ii) availability of structural sequences (pre (T1) and post-contrast T1 weighted (T1CE), T2 weighted (T2), and T2 weighted fluid attenuated inversion recovery (T2-FLAIR)) and diffusion MRI at time of diagnosis. For the *IDH1* mutation study, sufficient tumor tissue collected at time of surgery was required. We identified 381 patients with adult-type diffuse gliomas (CNS WHO grade 4) and 50 patients with brain metastases, and they were randomly divided into three datasets. Figure 2 demonstrates an overall view of the datasets. **The training dataset** included 106 patients with brain tumors, 66 *IDH1*-wildtype glioblastomas, aged 23–83 years (mean: 60.5, standard deviation (SD): 11.8), and in a male: female proportion of 37:29; and 40 metastases, aged 29–87 years (mean: 62.12, SD: 12.6), and in a male: female proportion of 18:22. **The validation dataset** was used to make locoregional hubs. This dataset was independent from the training dataset and consisted of 30 patients with brain tumors, 20 *IDH1*-wildtype glioblastomas and 10 brain metastases, aged 42–84 years (mean: 64.3, SD: 10.5), and in a male: female proportion of 16:14. **The test dataset** was used for two different applications, comprising survival among *IDH1*-wildtype glioblastomas and differences among *IDH1* mutation status. **The overall survival dataset** included 264 *IDH1*-wildtype glioblastomas in the survival range of 0.43 to 76.9 month, aged 21–88 years (mean: 63.6, SD: 11.4), and in a male: female proportion of 160:104. **The mutation dataset** consisted of 275 CNS WHO grade 4 adult-type diffuse gliomas, 264 patients with *IDH1*-wildtype glioblastoma as in the survival data set and 11 *IDH1*-mutant astrocytomas, aged 21–88 years (mean: 62.5, SD: 12.5), and in a male: female proportion of 165:110. The acquisition parameters for all datasets were based on single-shell diffusion data on two Siemens 3 T scanners, Verio or TrioTim, with TR/TE in range of 4200–7400 ms/83–88 ms, with 1 unweighted volume and either 30 or 12-direction diffusion-weighted volumes, at a *b* value of 1000 s/mm<sup>2</sup>. The acquisition was repeated between 1 and 6 times for improved signal to noise ratio (SNR). The spatial resolution was 1.7 × 1.7 × 3 mm.

**Creation of free water map and masks of tumor and edema.** All diffusion data was pre-processed using local Principal Component Analysis (PCA) denoising<sup>19</sup>, eddy current and motion correction using FSL EDDY<sup>20</sup>, and skull-stripping with Brain Extraction Tool (BET)<sup>21</sup>. Fractional anisotropy (FA) and mean diffusivity (MD) maps were calculated after DTI fitting with DIPY using weighted least squares<sup>22</sup>.

Structural scans (T1CE, T2, and T2-FLAIR) were registered to the pre-contrast T1 with rigid registration in Advanced Normalization Tools (ANTs)<sup>23,24</sup>, and then registered to the FA image with a nonlinear registration to account for Echo Planar Imaging (EPI) distortions in the diffusion data in ANTs. Masks of the tumor and edema for each patient were created from the registered structural images using Deep-Medic<sup>25</sup>. We used Freewater Estimator using Interpolated Initialization (FERNET)<sup>13</sup>, a free water elimination paradigm designed for single-shell diffusion MRI data using a novel interpolated initialization approach, to estimate the free water compartment in single-shell diffusion MRI data. FERNET provides the user with a free water volume fraction map that were resampled to 2 × 2 × 2 mm spatial resolution.

**Creation of the PMI map.** We used the training dataset to train a convolutional neural network (CNN)<sup>26</sup> model using the differences in water diffusivity properties of metastasis and glioblastomas. Figure 3 demonstrates a schematic view for the generation of the PMI map. We automatically extracted a set of (16 × 16) voxel patches from the peritumoral edema of metastases and *IDH1*-wildtype glioblastomas placed at random locations

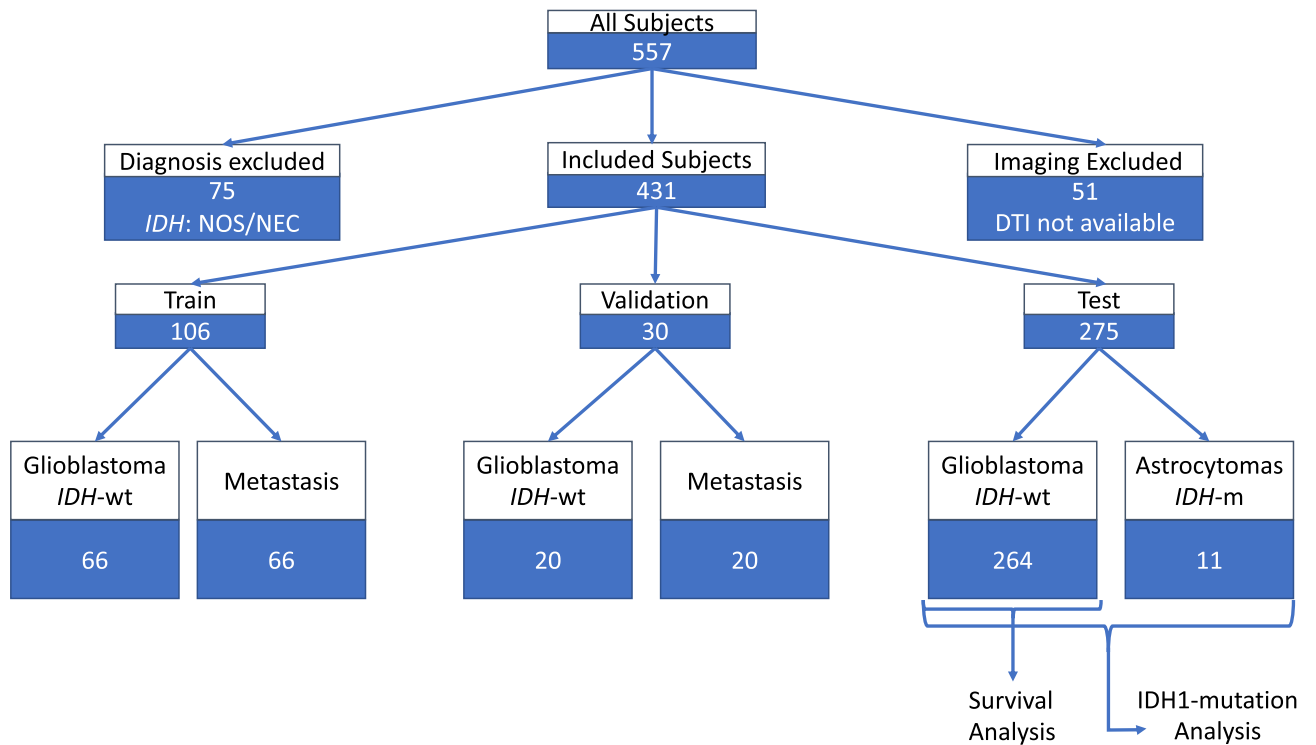


**Figure 1.** Different steps of the processing pipeline. The pipeline consists of three steps. (1) creation of free water map and masks of tumor and edema, (2) creation of the PMI map, (3) extracting locoregional hubs and AI-based markers. PMI: Peritumoral microenvironment index, CNN: Convolutional neural networks, FERNET: Freewater estimator using Interpolated Initialization, MRI: Magnetic resonance imaging.

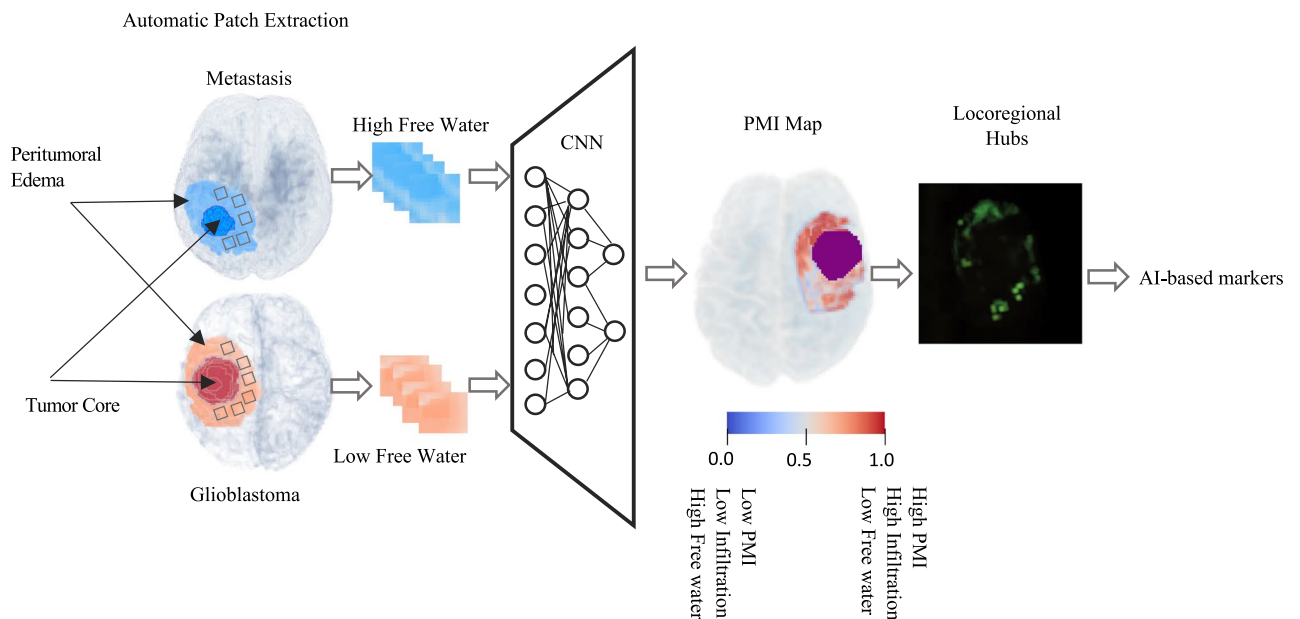
and random directions (within axial, sagittal and coronal planes) using random seed generators. This was the largest patch that could be fit into edema without overlapping into the main tumor mass. Patches in the peritumoral edema of metastases and glioblastomas were labeled as high free water and low free water, respectively. The convolutional neural network consisted of 6 convolutional layers followed by a max-pooling and a fully connected layer. A softmax layer at the end produced a probability value for every input patch that indicated its membership to each class, either high free water or low free water<sup>27</sup>. Data augmentation was done on the patches by shifting them in random directions, letting a maximum of 20% overlap with the healthy brain. The hyperparameters of CNN were weight decay  $5 \times 10^{-5}$ , momentum 0.9, initial learning rate  $10^{-4}$ . The details of the CNN architecture can be found in supplementary material S.1.

To obtain the PMI map for new unseen patients, we placed patches centered at each voxel within the sagittal, axial, and coronal directions. The PMI value for each voxel was then calculated by averaging among the result of the CNN for the patches.

**Extracting locoregional hubs and AI-based markers.** For extracting locoregional hubs, we used the validation dataset. Connected components (CCs) of high PMI values<sup>28</sup> were extracted and defined as locore-

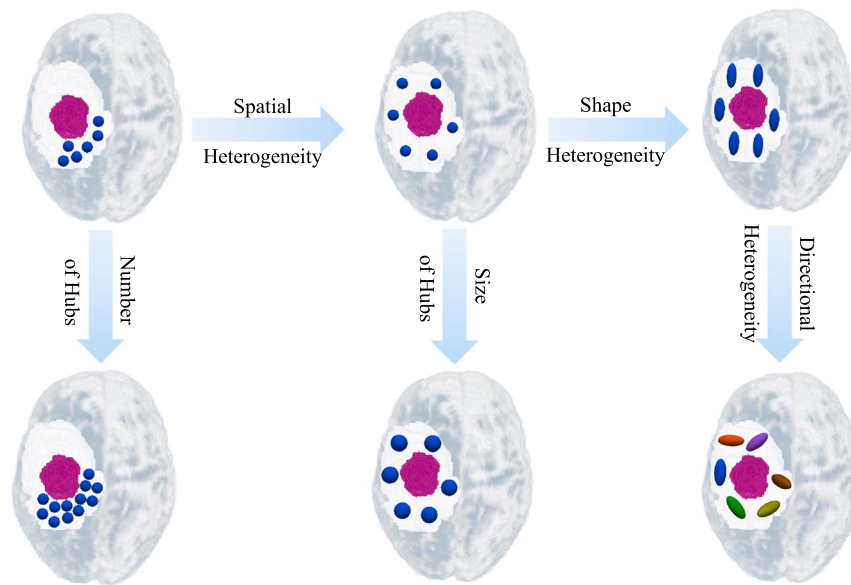


**Figure 2.** Datasets: train, validation, test (survival and *IDH1*-mutation analysis). *IDH1*: Isocitrate-Dehydrogenase. NOS/NEC: Not elsewhere classified/Not otherwise specified.



**Figure 3.** Generation of the PMI map and locoregional hubs. The inputs to the CNN are patches (boxes) extracted from the free water volume fraction map in the peritumoral region from both glioblastomas (red) and metastases (blue) labeled as low free water and high free water which are used to train the CNN. Locoregional hubs are extracted from PMI. Descriptive characteristics of the locoregional hubs are extracted as AI-based markers. PMI: Peritumoral microenvironment index, CNN: Convolutional neural networks.

gional hubs. Descriptive characteristics of locoregional hubs were extracted as AI-based markers for each patient. These characteristics comprised the number and size of the hubs, as well as the differences in their (i) shape (quantified by their individual anisotropic property), (ii) directionality, and (iii) spatial location (see Fig. 4 for a schematic view).



**Figure 4.** A schematic view of locoregional hubs descriptive characteristics: number of hubs, size of hubs, shape heterogeneity, directional heterogeneity, and spatial heterogeneity. The purple circle displays the tumor core and blue hubs are located in the peritumoral edema. Arrows are directed toward increasing the values of the descriptive characteristics. Figure in the top-left is the reference figure for the hubs. (i) Moving from left to right, to the top-middle: the spatial heterogeneity of hubs increases; to the top-right: the shape heterogeneity increases. (ii) Moving from top to the bottom, left: the number of hubs increases; center: size of hubs increases; right: directional heterogeneity increases.

CCs were created for patients in different thresholds in range of [0 1] by steps of 0.1. Connected components with diameters less than 2 voxels were not taken into consideration. The threshold which provided the most significant differences in the descriptive characteristics between patients of the validation data set (30 samples of metastases and glioblastomas) was chosen (threshold = 0.9) and used for the mutation and the survival analysis.

For obtaining descriptive characteristics as AI-based markers, the *number* and *size* of the locoregional hubs were extracted and normalized based on the number of voxels in edema. The size was averaged among the hubs of each patient. Shape heterogeneity and directional heterogeneity were computed as follows: For each locoregional hub, coordinates for all its voxels along three (axial, sagittal, coronal) directions were extracted. Standard deviation of all voxels coordinate values along those three directions were calculated to make a three directional feature vector called *sdvec*. *Shape heterogeneity* was defined as the normalized difference between the maximum and second largest *sdvec*, which was later averaged among hubs of each patient. For *directional heterogeneity*, Hausdorff distance<sup>29</sup> was applied on the pair-wise cosine distance among *sdvec* of different hubs. For computing *spatial heterogeneity*, the center of gravity for each hub and the average Euclidean distance among them was calculated. This distance was divided by the diameter, or maximum possible distance between any two points to make a normalized value between 0 and 1. A prototype code for extracting AI-based markers is provided in supplementary material S.2.

**Analysis of survival and *IDH1* mutation.** The PMI maps were generated for the patients in the survival and mutation datasets. 8 patients were removed from both datasets as they had small edema size (less than two voxels around tumor). Next, the locoregional hubs of high PMI values were generated. 9 patients from the survival dataset and 10 patients from the mutation dataset were removed as they had less than 2 hubs which was the minimum number needed to compute directional and shape heterogeneity.

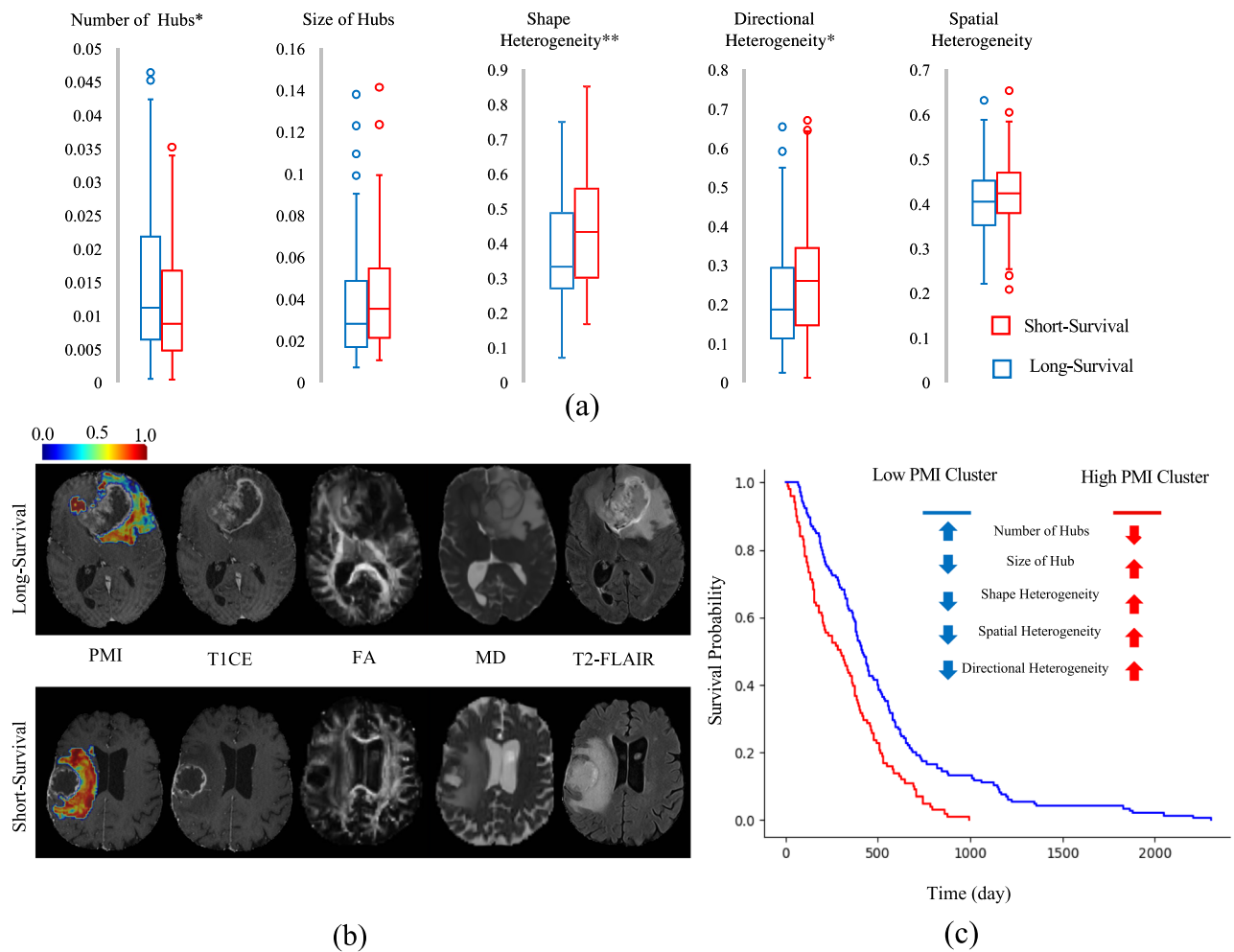
Survival analysis was done in two parts. First, we divided the population into short- and long-survival groups at the median and analyzed their statistical differences. This was done by doing: (i) two sample two tailed *t* test for each of the markers, followed by Bonferroni correction for multiple comparison; and (ii) linear regression on each of the markers with sex and age as covariates. Second, clusters of patients were generated downstream of the integration of all markers using K-means. We tried different number of clusters in the range of 1 to 9 and the one that maximized the Calinski-Harabasz index<sup>30</sup> was chosen. Two clusters were derived, namely, low-PMI and high-PMI patients. Kaplan-Meier (KM)<sup>6</sup> curves were fit for low-PMI and high-PMI clusters using nonparametric Turnbull estimator<sup>31</sup>. Logrank test and Cox hazard ratio were applied to compare KM-curves. The correlation between different markers were also calculated to explore their similarities.

*IDH1* mutation analysis was done by statistical group difference in two different ways: (i) to consider differences in the number of samples in each group, Wilcoxon rank sum<sup>32</sup> test was done for each of the markers, followed by Bonferroni correction for multiple comparison; (ii) linear regression was performed on each of the markers with sex and age as covariates.

## Results

The proposed AI-based markers were evaluated on two clinical applications to investigate whether the differences in infiltrative heterogeneity, as captured by the proposed descriptive characteristics, were able to capture differences overall survival and *IDH1* mutation status of adult-type diffuse glioma patients.

**Survival analysis in *IDH1*-wildtype grade 4 gliomas.** The survival analysis was done on a population of 264 *IDH1*-wildtype grade 4 glioma patients in the survival range of 0.43 to 76.9 months. The goal was to determine whether the proposed markers contain information pertaining to patient survival. Statistical analysis between short and long-survival groups demonstrated significant differences in the proposed markers, using *t* test and Bonferroni correction for multiple comparison. The short-survival group had significantly lower number of locoregional hubs ( $t_{\text{number}} = 2.54$ ,  $p_{\text{number}} = 0.01$ ). Shape and directional heterogeneity were also significantly higher in the short-survival group ( $t_{\text{shape}} = 3.89$ ,  $p_{\text{shape}} < 0.001$ ,  $t_{\text{directional}} = 2.79$ ,  $p_{\text{directional}} = 0.005$ ). Linear regression analysis of the markers using sex and age as covariates revealed significant differences between short and long-survival groups (Fig. 5a). The short-survival group had significantly lower number of locoregional hubs ( $t_{\text{number}} = 2.65$ ,  $p_{\text{number}} = 0.008$ ). Shape and directional heterogeneity were significantly higher in the short-survival group ( $t_{\text{shape}} = 3.74$ ,  $p_{\text{shape}} < 0.001$ ,  $t_{\text{directional}} = 2.44$ ,  $p_{\text{directional}} = 0.015$ ). Size and spatial heterogeneity were not found to be significantly different. Examples of PMI maps (overlaid with structural MRI), along with T1CE, FA, MD, and T2-FLAIR images, for a short and a long-survival patient are presented in Fig. 5b. KM estimates<sup>6</sup> of the low- and high-PMI groups of patients (generated using K-means clustering on integration of all markers) are provided in Fig. 5c. A logrank test showed a significant difference between KM curves of low- and high-PMI clusters ( $t = 19.9$ ,  $p < 10^{-3}$ ), and the Cox hazard ratio was 1.82 (95% confidence interval: 1.39, 2.37;  $p < 0.005$ ).



**Figure 5.** The PMI map and AI-based markers for *IDH1*-wildtype grade 4 glioma patients with different duration of survival. (a) AI-based markers (Descriptive characteristics of PMI locoregional hubs) for long and short survival groups,  $p$  value  $< 0.05$  (\*),  $p$  value  $< 0.005$  (\*\*), Linear regression was used with age and sex as covariates. (b) representative samples of the PMI map with T1CE, FA, MD, and T2-FLAIR images for long and short survival patients. (c) Kaplan–Meier estimates of the two clusters generated by K-means clustering. PMI: Peritumoral Microenvironment Index, *IDH1*: Isocitrate-Dehydrogenase 1, T1CE: T1 post-contrast, FA: Fractional anisotropy, MD: Mean diffusivity, T2-FLAIR: T2 weighted fluid attenuated inversion recovery.

Group differences among high- and low-PMI clusters, were similar to short and long-survival groups, respectively (Fig. 5c for details and further information in S.3). The correlations between the proposed markers showed the highest correlations between shape heterogeneity and directional heterogeneity. However, they still provide complementary information as coefficient of determination ( $R^2$ ) showed that only 43% of information in directional heterogeneity was covered by shape heterogeneity (details in S.4).

**IDH1 mutation analysis of grade 4 gliomas.** *IDH1* mutation analysis was done on a population of 275 CNS WHO grade 4 adult-type diffuse gliomas with different *IDH1* mutation status (i.e., 264 *IDH1*-wildtype glioblastomas vs. 11 *IDH1*-mutant astrocytomas) to investigate whether the proposed markers can characterize mutation status. Statistical analysis across the population, using Wilcoxon rank sum test and Bonferroni correction for multiple comparison, demonstrated significant differences in the proposed markers between *IDH1*-mutants and *IDH1*-wildtypes. The *IDH1*-wildtype group had significantly lower number of locoregional hubs ( $z_{\text{number}} = 4.11$ ,  $p_{\text{number}} < 10^{-4}$ ). Shape heterogeneity and directional heterogeneity were also significantly higher in *IDH1*-wildtypes ( $z_{\text{shape}} = 3.15$ ,  $p_{\text{shape}} = 0.001$ ,  $z_{\text{directional}} = 3.5$ ,  $p_{\text{directional}} < 10^{-3}$ ). Linear regression analysis of the proposed markers using sex and age as covariates revealed significant differences between the two groups (Fig. 6a). *IDH1*-wildtypes had significantly lower number of locoregional hubs, when compared with *IDH1*-mutants ( $t_{\text{number}} = 2.861$ ,  $p_{\text{number}} = 0.005$ ). Shape heterogeneity and directional heterogeneity were significantly higher in *IDH1*-wildtypes comparing to *IDH1*-mutants ( $t_{\text{shape}} = 2.407$ ,  $p_{\text{shape}} = 0.01$ ,  $t_{\text{directional}} = 2.380$ ,  $p_{\text{directional}} = 0.01$ ). Size and spatial heterogeneity were not found to be significantly different. Examples of PMI maps (overlaid with structural MRI), along with the corresponding T1CE, FA, MD, and T2-FLAIR images for two patients (*IDH1*-mutant astrocytoma and *IDH1*-wildtype glioblastoma) are presented in Fig. 6(b). AI-based markers of the complete population of *IDH1*-mutants versus *IDH1*-wildtype long-survival group versus *IDH1*-wildtype short-survival group are provided in S.5.

## Discussion

We introduced fully automated, novel AI-based markers of the peritumoral microenvironment (PME) of adult-type diffuse gliomas using unexplored information of water restriction extracted from DTI. The markers were based on descriptive characteristics derived from locoregional hubs of the PMI map that captured a unique aspect of the infiltrative heterogeneity from the extracellular water properties of the peritumoral regions of metastases and glioblastomas. Our findings identified a panel of markers as surrogates of infiltration, that could be considered as potential biomarkers of prognosis based on survival and *IDH1* mutation status.

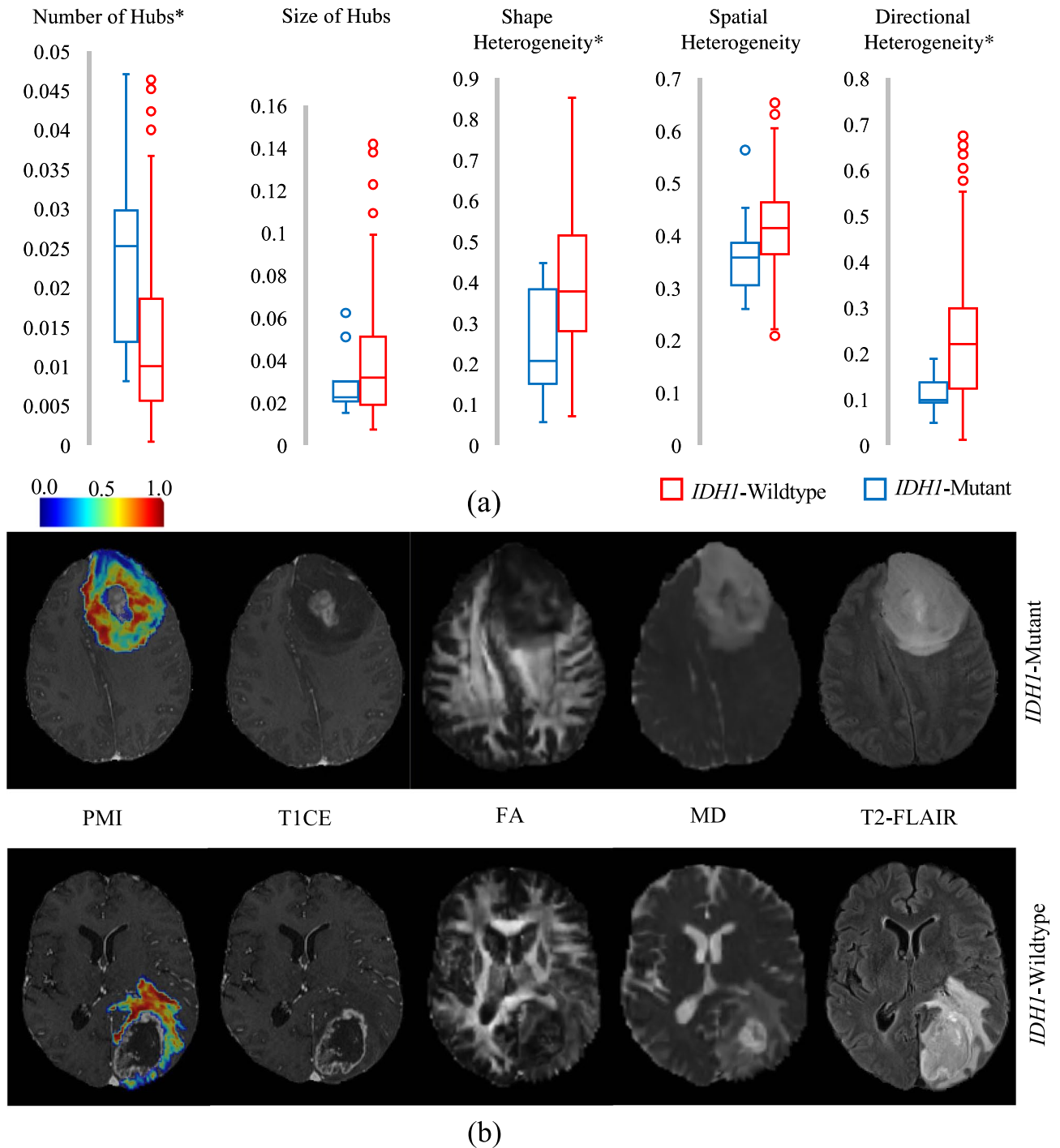
Previous methods on characterizing the infiltrative heterogeneity utilized various imaging modalities using manual or heuristic-based labeling of infiltrative tissue<sup>6–12</sup>. DTI, the modality to characterize water movement, has been used for defining infiltrative heterogeneity in the peritumoral region<sup>33</sup>. However, the extracellular free water movement differences have not been explored for peritumoral region. Therefore, the proposed AI-based markers presented here quantified the heterogeneity in the PME which could complement and further strengthen other diffusion-based measures and MRI parameters to aid in personalized treatment planning toward optimizing patient outcome.

Peritumoral edema, happens due to excess accumulation of fluid in the brain parenchyma which results from infiltrating tumor cells as well as biological responses to the permeability of the spatially adjacent tumor cells<sup>6–8</sup>. The spatial distribution and level of water dynamics pattern in edema has not been well defined yet<sup>6–8</sup>. Using a CNN model trained by characteristics of water diffusivity in glioblastomas versus brain tumor metastases, we were able to capture peritumoral heterogeneity in adult-type diffuse gliomas, without using any manual labeling, to capture surrogates of infiltration and learn how brain parenchyma and immune system responded to different malignancies.

Brain peritumoral tissue is a heterogeneous microenvironment and accurately quantifying this heterogeneity can be crucial for understanding tumor progression. The proposed markers described regionally distinct water movement centers that contained voxels sharing the same brain tissue microstructures and we demonstrated their utility in two clinical applications.

Previous studies showed that higher infiltration of tumor cells is associated with poorer prognosis<sup>6,7,34,35</sup>. The PMI maps also demonstrated higher PMI values in the voxels of the short-survival patients. Previous literature reporting differences in the levels of FA and MD in altered infiltrated tissues which might be related to higher cellularity and lower water content<sup>7,8,36,37</sup>. Our result provides complementary information in regard to extracellular water movement restriction differences. Analysis of the proposed markers between poor prognosis (short-survivors and high-PMI cluster) and good prognosis (long-survivors and low-PMI cluster), showed a significantly lower number of locoregional hubs in the poor prognosis group which implies larger hubs corresponding to a more infiltrative and aggressive type of tumor. A significantly higher value of shape heterogeneity in the poor prognosis group could reflect the fact that in adult-type diffuse gliomas, cancer cells mainly infiltrate along the white matter<sup>38–40</sup>. Likewise, higher value of directional heterogeneity for poor prognosis patients could reflect the fact that cancer spreads in different directions comparing to good prognosis group. This is consistent with previous literature showing that glioma cells infiltrate brain tissue by at least two topographic paths, including perivascular invasion along the vascular system or infiltration along the extracellular matrix, nerve, and astrocytic tracts<sup>41</sup>.

*IDH1*-wildtype glioblastoma has a poorer prognosis than the *IDH1*-mutant astrocytoma<sup>17,18,34</sup> which shows higher PMI values. The statistical differences of the proposed markers between *IDH1*-mutants and *IDH1*-wildtypes suggests them as powerful markers to discriminate mutation status, even prior to biopsy. *IDH1*-wildtypes had a lower number of locoregional hubs, which could reflect bigger size of adjacent tumoral cells compared to *IDH1*-mutants. This might represent extracellular vesicles which are involved in the rich network



**Figure 6.** The PMI map and AI-based markers for grade 4 glioma patients with different *IDH1* mutation status. **(a)** AI-based markers (Descriptive characteristics of PMI locoregional hubs) for *IDH1*-mutant and *IDH1*-wildtype groups,  $p$  value  $< 0.05$  (\*),  $p$  value  $< 0.005$  (\*\*), Linear regression was used with age and sex as covariates. **(b)** representative samples of the PMI map with T1CE, FA, MD, and T2-FLAIR images for *IDH1*-mutant and *IDH1*-wildtype patients. *IDH1*: Isocitrate-Dehydrogenase 1, T1CE: T1- and post-contrast, FA: Fractional anisotropy, MD: Mean diffusivity, T2-FLAIR: T2 weighted fluid attenuated inversion recovery.

of intercellular connections and develop pathologic cascade leading to neurological diseases<sup>42</sup>. Likewise, higher shape and directional heterogeneity in *IDH1*-wildtypes, could demonstrate that in *IDH1*-wildtypes, cancer spreads more throughout the surrounding brain tissue, reflecting poorer prognosis which might result in higher tumor cellularity and infiltration, causing FA and MD changes<sup>43,44</sup>.

The locoregional hubs-based markers introduced here capture the heterogeneity in masses of adjacent voxels with high infiltration which could correspond to a connected set of glioma cells in the peritumoral region.



Multicellular networks with filamentous microtubes are found to connect tumoral cells to a network in the peritumoral space in animal models<sup>41,45</sup>. These cellular networks allow rapid progression and are likely related to treatment resistance<sup>41,45</sup>. Therefore, biologically applicable markers can be identified in highly connected tumor cells that provide insights into the tumor microenvironment and offer translation to the clinic<sup>41,45</sup>.

The proposed markers could be interpreted as signatures of connected tumoral cells. Size and number of hubs quantitated modularity which is a measure of the structure of networks. Lower number of hubs in the poor prognosis group demonstrated their higher modularity. Higher values of shape and directional heterogeneity in poor prognosis group suggest connected tumoral cells with higher shape and directional heterogeneity. This might provide potentially important biomarkers that could be further explored as key targets for treatment planning and patient selection for clinical trials, including novel immunotherapy and anti-invasion therapy<sup>41,45,46</sup>.

The data in this study was collected from a single institution. This limitation was addressed by utilizing different acquisition settings and use of an independent cohort for testing, which provides strong sign of a generalizability of our results. The number of *IDH1* mutant subjects were limited in our study since only approximately 5.6% of primary glioblastomas are *IDH1* mutant<sup>47</sup>. The statistical significance of our findings can be improved by higher number of subjects. Our markers can be extracted fully automatic and do not require any manual intervention or reference labels and are easily translatable to the clinic. Although the proposed markers without a corresponding resected specimen cannot pathologically prove the infiltrative heterogeneity of the peritumoral region, they can be used as surrogate markers for prognosis.

In the future, the proposed markers could be integrated with other imaging modalities to provide biologically relevant characterization of the PME to quantitate invasion and microenvironmental heterogeneity for effective cancer therapy<sup>48–50</sup>. Further studies using the markers described here could be useful to elucidate biological processes linked to the *IDH1* mutation status and better understand multiple sources of heterogeneity in adult-type diffuse gliomas which has major clinical implications<sup>51–53</sup>.

### Data availability

The datasets generated during the current study are not publicly available due to the IRB requirements of the Hospital of University of Pennsylvania. The datasets can be made available on request to the corresponding author, after required data transfer and IRB paperwork is completed. The code is available here: <https://github.com/zrsamani/AI-Based-LoCoregional-Markers>.

Received: 23 March 2022; Accepted: 14 December 2022

Published online: 18 January 2023

### References

- Sebök, M. *et al.* Hypermetabolism and impaired cerebrovascular reactivity beyond the standard MRI-identified tumor border indicate diffuse glioma extended tissue infiltration. *Neuro Oncol. Adv.* **3**, 1–9 (2021).
- Claes, A., Idema, A. J. & Wesseling, P. Diffuse glioma growth: A guerilla war. *Acta Neuropathol.* **114**, 443–458 (2007).
- Wang, J. *et al.* Glioma invasion along white matter tracts: A dilemma for neurosurgeons. *Cancer Lett.* **526**, 103–111 (2022).
- Cuddapah, V. A., Robel, S., Watkins, S. & Sontheimer, H. A neurocentric perspective on glioma invasion. *Nat. Rev. Neurosci.* **15**, 455–465 (2014).
- Kelly, P. J. *et al.* Imaging-based stereotaxic serial biopsies in untreated intracranial glial neoplasms. *J. Neurosurg.* **66**, 865–874 (1987).
- Akbari, H. *et al.* Pattern analysis of dynamic susceptibility contrast-enhanced MR imaging demonstrates peritumoral tissue heterogeneity. *Radiology* **273**, 502–510 (2014).
- Akbari, H. *et al.* Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma. *Neurosurgery* **78**, 572–580 (2016).
- Rathore, S. *et al.* Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: Implications for personalized radiotherapy planning. *J. Med. Imaging* **5**(021219), 021211–21210 (2018).
- Yan, J.-L. *et al.* Multimodal MRI characteristics of the glioblastoma infiltration beyond contrast enhancement. *Therap. Adv. Neurol. Disord.* **12**, 1–10 (2019).
- d'Este, S. H., Nielsen, M. B. & Hansen, A. E. Visualizing glioma infiltration by the combination of multimodality imaging and artificial intelligence, a systematic review of the literature. *Diagnostics* **11**, 592 (2021).
- Yan, J.-L. *et al.* A neural network approach to identify the peritumoral invasive areas in glioblastoma patients by using MR Radiomics. *Sci. Rep.* **10**, 1–10 (2020).
- Dasgupta, A. *et al.* Quantitative mapping of individual voxels in the peritumoral region of IDH-wildtype glioblastoma to distinguish between tumor infiltration and edema. *J. Neuro Oncol.* **153**(2), 1–11 (2021).
- Parker, D. *et al.* Freewater estimator using interpolated initialization (FERNET): Characterizing peritumoral edema using clinically feasible diffusion MRI data. *PLoS ONE* **15**, e0233645 (2020).
- Samani, Z. R. *et al.* Distinct tumor signatures using deep learning-based characterization of the peritumoral microenvironment in glioblastomas and brain metastases. *Sci. Rep.* **11**, 14469 (2021).
- Samani, Z. R., Parker, D., Alappatt, J. A., Brem, S. & Verma, R. Deep learning-based peritumoral microstructure mapping in glioblastomas using free water volume fraction. *Neuro Oncol.* **22**, 157–158 (2020).
- Samani, Z. R., Parker, D., Alappatt, J. A., Brem, S. & Verma, R. Differentiating tumor types based on the peritumoral microenvironment using convolutional neural networks. *Neuro Oncol.* **22**, ii151 (2020).
- Aliotta, E. *et al.* Increased intratumoral infiltration in IDH wild-type lower-grade gliomas observed with diffusion tensor imaging. *J. Neurooncol.* **145**, 257–263 (2019).
- Chang, Y.-Z. *et al.* Transcriptional characteristics of IDH-wild type glioma subgroups highlight the biological processes underlying heterogeneity of IDH-wild type WHO grade IV gliomas. *Front. Cell Dev. Biol.* **8**, 1132 (2020).
- Manjón, J. V. *et al.* Diffusion weighted image denoising using overcomplete local PCA. *PLoS ONE* **8**, e73021 (2013).
- Andersson, J. L. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).
- Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155 (2002).
- Garyfallidis, E. *et al.* Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinformatics* **8**, 1–17 (2014).
- Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).

24. Tustison, N. J. Explicit B-spline regularization in diffeomorphic image registration. *Front. Neuroinform.* **7**, 39 (2013).
25. Kamnitsas, K. *et al.* Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017).
26. Samani, Z. R., Alappatt, J. A., Parker, D., Ismail, A. A. O. & Verma, R. QC-Automator: Deep learning-based automated quality control for diffusion mr images. *Front. Neurosci.* **13**, 1456 (2020).
27. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **5**, 1106–1114 (2012).
28. He, L. *et al.* The connected-component labeling problem: A review of state-of-the-art algorithms. *Pattern Recogn.* **70**, 25–43 (2017).
29. Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 850–863 (1993).
30. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat. Theory Methods* **3**, 1–27 (1974).
31. Hudgens, M. G. On nonparametric maximum likelihood estimation with interval censoring and left truncation. *J. Royal Stat. Soc. Ser. B Stat. Methodol.* **67**, 573–587 (2005).
32. Fay, M. P. & Proschan, M. A. Wilcoxon-Mann-Whitney or *t*-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Stat. Surveys* **4**, 1 (2010).
33. Hoefnagels, F. W. *et al.* Differentiation of edema and glioma infiltration: Proposal of a DTI-based probability map. *J. Neurooncol.* **120**, 187–198 (2014).
34. Di Jia, S. L., Li, D., Xue, H., Yang, D. & Liu, Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging Albany N. Y.* **10**, 592 (2018).
35. Brancato, V. *et al.* Predicting survival in glioblastoma patients using diffusion MR imaging metrics—A systematic review. *Cancers* **12**, 2858 (2020).
36. Akbari, H. *et al.* Histopathology-validated machine learning radiographic biomarker for noninvasive discrimination between true progression and pseudo-progression in glioblastoma. *Cancer* **126**, 2625–2636 (2020).
37. Metz, M.-C. *et al.* Predicting glioblastoma recurrence from preoperative MR scans using fractional-anisotropy maps with free-water suppression. *Cancers* **12**, 728 (2020).
38. Sharifi, G., Pajavand, A. M., Nateghinia, S., Meybodi, T. E. & Hasooni, H. Glioma migration through the corpus callosum and the brainstem detected by diffusion and magnetic resonance imaging: Initial findings. *Front. Hum. Neurosci.* **13**, 472 (2020).
39. Hara, A. *et al.* Treatment strategies based on histological targets against invasive and resistant glioblastoma. *J. Oncol.* **2019**, 1–10 (2019).
40. Ferrer, V. P., Moura Neto, V. & Mentlein, R. Glioma infiltration and extracellular matrix: Key players and modulators. *Glia* **66**, 1542–1565 (2018).
41. Gritsenko, P. G. *et al.* p120-catenin-dependent collective brain infiltration by glioma cell networks. *Nat. Cell Biol.* **22**, 97–107 (2020).
42. Ciccocioppo, F., Lanuti, P., Marchisio, M. & Miscia, S. Extracellular vesicles involvement in the modulation of the glioblastoma environment. *J. Oncol.* **2020**, 1–8 (2020).
43. Rathore, S. *et al.* Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci. Rep.* **8**, 1–12 (2018).
44. Lee, S. *et al.* Evaluation of the microenvironmental heterogeneity in high-grade gliomas with IDH1/2 gene mutation using histogram analysis of diffusion-weighted imaging and dynamic-susceptibility contrast perfusion imaging. *J. Neurooncol.* **121**, 141–150 (2015).
45. Hai, L. *et al.* A connectivity signature for glioblastoma. *bioRxiv. Preprint.* **2208**, 14445. <https://doi.org/10.1101/20211107465791> (2021).
46. Verma, R. *et al.* Prediction of response to combination of nivolumab and bevacizumab in patients with recurrent glioblastoma via radiomic analysis on clinical MRI scans. *Neuro Oncol. Adv.* **3**, iv6–iv6 (2021).
47. Liu, A., Hou, C., Chen, H., Zong, X. & Zong, P. Genetics and epigenetics of glioblastoma: Applications and overall incidence of IDH1 mutation. *Front. Oncol.* **6**, 16 (2016).
48. Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor heterogeneity: The rosetta stone of therapy resistance. *Cancer Cell* **37**, 471–484 (2020).
49. Khalafallah, A. M. *et al.* “Zooming in” on glioblastoma: Understanding tumor heterogeneity and its clinical implications in the era of single-cell ribonucleic acid sequencing. *Neurosurgery* **88**, 477–486 (2021).
50. Majumder, B. *et al.* Predicting clinical response to anticancer drugs using an ex vivo platform that captures tumour heterogeneity. *Nat. Commun.* **6**, 1–14 (2015).
51. Neftel, C. *et al.* An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849 (2019).
52. Qazi, M. *et al.* Intratumoral heterogeneity: pathways to treatment resistance and relapse in human glioblastoma. *Ann. Oncol.* **28**, 1448–1456 (2017).
53. Gill, B. J. *et al.* MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. *Proc. Natl. Acad. Sci.* **111**, 12550–12555 (2014).

## Acknowledgements

Research reported in this publication was supported by the National Institutes of Health (NIH) Grant R01NS096606 (PI: Ragini Verma), National Cancer Institute (NCI) of the National Institutes of Health (NIH) under award numbers: NCI: U01CA242871, NCI: U24CA189523, NINDS: R01NS042645. The content of this publication is solely the responsibility of the authors and does not represent the official views of the NIH.

## Author contributions

Conceptualization: Z.R.S., R.V.; Methodology: Z.R.S.; Experiments: Z.R.S.; Data processing: D.P.; Analysis: Z.R.S., D.P., H.A.; Interpretation: Z.R.S., D.P., H.A., R.L.W., S.B., Sp.B., R.V.; Writing and editing: Z.R.S., D.P., H.A., R.L.W., S.B., Sp.B., R.V.; Supervision: S.B., Sp.B., R.V.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-26448-9>.

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

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

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



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

© The Author(s) 2023