Observing Deep Radiomics for the Classification of Glioma Grades

Kazuma Kobayashi (kazumkob@ncc.go.jp)
National Cancer Center Research Institute
Mototaka Miyake
National Cancer Center Hospital
Masamichi Takahashi
National Cancer Center Hospital
Ryuji Hamamoto
National Cancer Center Research Institute

Research Article

Keywords: Deep Radiomics, Glioma Grades, Deep learning, magnetic resonance imaging (MRI)

Posted Date: February 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-182617/v1

Observing Deep Radiomics for the Classification of Glioma Grades

Kazuma Kobayashi[1,2,*], Mototaka Miyake[3], Masamichi Takahashi[4], and Ryuji Hamamoto[1,2]

[1]Division of Molecular Modification and Cancer Biology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
[2]Cancer Translational Research Team, RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo
[3]Department of Diagnostic Radiology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
[4]Department of Neurosurgery and Neuro-Oncology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
[*]kazumkob@ncc.go.jp

ABSTRACT

Deep learning is promising for medical image analysis because it can automatically acquire meaningful representations from raw data. However, a technical challenge lies in the difficulty of determining which types of internal representation are associated with a specific task, because feature vectors, which collectively constitute the feature maps, can vary dynamically according to individual inputs. Therefore, based on the magnetic resonance imaging (MRI) of gliomas, we propose a novel technique to extract a shareable set of feature vectors that encode various parts in tumor imaging phenotypes. Because the set of feature vectors is shared across a population, it can be used in other downstream tasks as if these feature vectors are imaging markers. Then, based on the feature vectors, a classifier is established using logistic regression to predict the glioma grade, and an accuracy of 90% is achieved. Besides, we develop an algorithm to visualize the image region encoded by each feature vector, and demonstrate that the classification model preferentially relies on feature vectors associated with the presence or absence of contrast enhancement in tumor regions. The proposed method provides a data-driven approach to enhance the understanding of physicians on the imaging appearance of gliomas.

Introduction

The scientific community has become interested not only in harnessing the predictive performance of machine learning models, but also in dissecting such models to distill useful knowledge that can potentially advance scientific understanding[1]. When a model achieves high prediction performance in a particular task, one would expect it to have acquired an expressive internal representation that approximates the explanatory patterns underlying the phenomena of interest. Therefore, one can interpret the internal representations of trained models to obtain meaningful insights and scientific knowledge without directly observing the phenomena. Based on this concept of acquiring medical knowledge in a data-driven manner, the objective of this study is to discover common features in medical imaging associated with specific clinical information across a patient population.

Particularly, this study focuses on the imaging phenotypes of gliomas, which are the most common central nervous system tumors[2,3]. According to the grading system of the World Health Organization (WHO), gliomas are classified into grades I to IV, based on histopathological findings obtained from surgical biopsies or specimens[4]. Because the degree of aggressiveness and infiltrative characteristics significantly affect the prognostics of patients, the differential diagnosis between lower-grade gliomas (LGG, WHO grades II and III) and high-grade gliomas (HGG, WHO grade IV) is an important issue with regard to treatment options and prognosis prediction[5].

Currently, the standard procedure for classifying tumors into the WHO grades is based on pathological study. However, there are still limitations for tumor classification, including the requirement for invasive procedures such as surgical resections or biopsies, inherent sampling errors caused by the heterogeneity of tumors, and the time-consuming process of histopathological analysis. There are also cases wherein it may be dangerous to perform surgical procedures on tumors located in critical brain sites. To address these issues, the computational analysis of magnetic resonance imaging (MRI) for tumor grading has attracted significant attention[6,7]. Because MRI can non-invasively observe an entire tumor in vivo, it is free from sampling errors. Therefore, the management of gliomas based on multi-parametric MRI analysis can play a complementary role in pathology-based diagnosis.

Radiomics and deep learning are two mainstays for computationally analyzing the appearance of tumors in imaging. Many intensive studies have attempted to analyze the imaging phenotypes of glioma, and each of these approaches has certain advantages and disadvantages with regard to gaining meaningful insights from trained models.

Radiomics is a research field focusing on decoding tumor phenotypes based on quantitative imaging features[8]. Typically, in the analysis pipeline, suitable sets of handcrafted imaging features are sought to be extracted from the region of interest (ROI). Subsequently, a prediction model based on a machine learning algorithm is trained for a particular prediction task that is relevant to clinical decision-making. For glioma grading, many previous studies have demonstrated that the tumor characteristics can be quantified using radiomics, and have reported that good discriminative performance was achieved[9–12]. Because the radiomics approach uses handcrafted imaging features that have been defined in advance, it has the advantage of high interpretability with regard to which features contribute to the prediction. However, to implement problem-specific handcrafted features, domain knowledge is often required. Because the optimal representative features for a given task are not always obvious, a data-driven approach should be considered to represent the data distribution.

Deep learning has emerged as an innovative technology that enables end-to-end learning between the input data and ground-truth labels[13]. By using backpropagation to tune the parameters of multilayered nonlinear operations during the training process, deep neural networks can automatically abstract useful representations from data. In other words, deep neural networks are capable of feature extraction in a data-driven manner. Therefore, a deep learning model can learn internal representations that are meaningful for distinguishing the attributes of samples without relying on feature engineering based on domain knowledge. For example, with regard to glioma grade classification, deep-learning-based algorithms have achieved remarkable prediction performance[14–16]. However, in such complex models, a tradeoff between accuracy and explainability has traditionally existed[17]. Hence, complex models, such as deep learning models, are occasionally referred to as black-box models[18], which implies that it is difficult to interpret how such a model reaches a particular outcome.

In this paper, we propose a new approach that combines the advantages of radiomics and deep learning, which we call *deep radiomics*. When a convolutional neural network (CNN) is trained to predict the imaging characteristics of gliomas, the internal representation can be acquired as low-dimensional feature vectors, which collectively constitute the feature maps. These feature vectors can then be used as imaging markers in downstream tasks because they are expected to adequately represent the appearance of tumors. However, only a few studies have closely investigated which types of imaging characteristics that can be exploited by deep learning models are predictable in clinical tasks in the field of glioma imaging. Among existing studies, Banerjee et al.[15] investigated the properties of convolutional kernels in different layers using visualization. However, in typical CNNs, the feature vectors contained in each feature map change dynamically depending on the inputs, and this makes it difficult to determine which internal representations are shared across a population. Because the objective of many medical studies is to find specific factors that are significantly common in a diseased population, we develop a method for obtaining a shareable set of feature vectors associated with the imaging characteristics of tumors that are obtained by deep learning models. Subsequently, we attempt to identify specific types of internal representations associated with particular clinical information. Thus, the proposed deep radiomics approach combines the flexible representative capacities of deep learning and the highly interpretable aspects of radiomics to acquire meaningful knowledge in a data-driven manner.

This paper proposes a novel technique for identifying the internal representation types of deep learning models that are associated with the binary classification of glioma grades, specifically LGG or HGG. As a pre-task, we first train a segmentation network, which can store a shareable set of feature vectors as its latent representation to predict the pixel-wise classification of glioma imaging characteristics from a two-dimensional (2D) axial slice of multi-parametric MRI (Figure 1a). To classify the glioma grades, the latent representation of each slice of an MRI volume is concatenated into a histogram, which represents the information of how many times each feature vector appears in the MRI volume (Figure 1b). By applying simple logistic regression to classify different glioma grades using the histogram representation, a set of feature vectors that are significantly associated with the prediction is identified. Furthermore, we conduct a feature ablation study to visualize which types of imaging characteristics are associated with glioma grades in the image space to provide interpretable feedback to physicians for the evaluation of grade-specific radiological findings (Figure 2). The results reveal that our classification model emphasizes the imaging appearance with or without contrast enhancement to distinguish between LGG and HGG (Figure 6). This observation is consistent with the findings reported in the literature.

Methods

In this section, we describe a scheme for using a shareable set of feature vectors obtained by a segmentation network for the classification of glioma grades using simple logistic regression. This task was formulated as a binary classification task whereby an input magnetic resonance (MR) volume is diagnosed either as LGG or HGG. Additionally, the types of imaging characteristics that enable the prediction were investigated by conducting a feature ablation study.

Proposed Image Processing Pipeline

Here, a technical description of the proposed method is provided.

Notation

Let us consider a multi-parametric three-dimensional (3D) MRI volume $\boldsymbol{X} \in \mathbb{R}^{C \times W \times H \times I}$, where C is the number of channels, W and H represent the height and width of the axial slices, respectively, and I is the number of axial slices. We define $\boldsymbol{x} \in \mathbb{R}^{C \times W \times H}$ as a slice in the axial view. The segmentation network encodes a slice-wise input $\boldsymbol{x}$ into the low-dimensional latent representation $\boldsymbol{z} \in \mathbb{R}^{C' \times W' \times H'}$ and decodes the segmentation output $\hat{\boldsymbol{y}} \in \mathbb{R}^{S \times W \times H}$, where S is the number of segmentation labels. The ground-truth segmentation label $\boldsymbol{y} \in \mathbb{R}^{S \times W \times H}$ is used to train the segmentation network. The series of latent representations $\boldsymbol{z}$ for each slice of the MRI volume can be concatenated into a summarized representation $\boldsymbol{Z} \in \mathbb{R}^{C' \times W' \times H' \times I'}$, which is considered as a volume-based representation. The glioma grades are classified on a volume basis, because grading is carried out clinically by referring to images of the entire brain.

Pre-training of Segmentation Networks with Shareable Set of Feature Vectors

As a pre-task, a segmentation network was trained to extract a shareable set of feature vectors. As shown in Figure 1a, the network consisted of an encoder and decoder pair connected to a discrete latent space containing a set of feature vectors in a codebook. Through the encoder, a 2D MRI slice $\boldsymbol{x}$ used as input is mapped to a latent representation $\boldsymbol{z}_e$. In the latent space, vector quantization is performed based on a codebook $\boldsymbol{e} = \{e_k | k = 1, \ldots, K\} \in \mathbb{R}^{K \times D}$, which stores a shareable set of K feature vectors as codewords $e_k \in \mathbb{R}^D$, by replacing each feature vector in $\boldsymbol{z}_e$ with the nearest codeword to produce the quantized latent representation $\boldsymbol{z}_q$. This vector quantization process is analogous to that of a vector-quantized variational autoencoder (VQ-VAE)[19,20]. The feature vectors corresponding to each voxel of $\boldsymbol{z}_e$ are quantized by executing a nearest-neighbor lookup on the codebook, as follows:

$$z_i = \underset{k \in [K]}{\arg\min} \| \boldsymbol{z}_{e_i} - e_k \|_2. \tag{1}$$

Thereafter, the codewords in the codebook are collected as a quantized latent representation $\boldsymbol{z}_q$, as follows:

$$\boldsymbol{z}_{q_i} = e_{z_i}. \tag{2}$$

To optimize this process, the codebook and encoder are trained to minimize the objective, which is referred to as *latent loss*, as follows:

$$L_{\text{latent}} = \| \text{sg}[\boldsymbol{z}_e(x)] - \boldsymbol{e} \|_2^2 + \beta \| \boldsymbol{z}_e(x) - \text{sg}[\boldsymbol{e}] \|_2^2, \tag{3}$$

where sg represents the stop-gradient operator, which serves as an identity function at the forward computation time and has zero partial derivatives. During training, the *codebook loss*, which is the first term in the above equation, updates the codebook variables by delivering the codewords to the encoder's output. Simultaneously, the *commitment loss*, which is the second term, encourages the output of the encoder to move closer to the target codewords. The hyperparameter β controls the reluctance of changing the encoder output to match the corresponding codewords. Backpropagation or the exponential moving average can be used to train the codebook[21]. Notably, the size of the codebook can be arbitrarily tuned, which ensures that a certain amount of information is preserved in a compressed manner within the latent space[20].

Then, the decoder takes $\boldsymbol{z}_q$ as an input and generates the segmentation map $\hat{\boldsymbol{y}}$, which is encouraged to be similar to the ground-truth labels $\boldsymbol{y}$. The segmentation loss function consists of the soft Dice[22] and focal losses[23]. In summary, the overall training objectives for the segmentation network are as follows:

$$L_{\text{total}} = L_{\text{latent}} + L_{\text{segmentation}}. \tag{4}$$

After the tumor segmentation training, we can consider the codewords as a shareable set of feature vectors that represent the imaging phenotypes of gliomas.

Histogram Representation of brain MRI based on Deep Radiomics

The set of feature vectors obtained inside the codebook can capture the semantic composition of the imaging phenotypes of gliomas. Thus, we hypothesize that these feature vectors are important for distinguish between LGG and HGG. Hereinafter, the encoder in the segmentation network is used as the feature extractor f that produces the slice-wise quantized latent representation $\boldsymbol{z}_q$ (Figure 1b). Further, it is necessary to summarize the slice-wise deep radiomics into a volume-wise representation because the diagnosis of the glioma grades is based on the entire images obtained by brain MRI. Therefore, we concatenate the I

quantized latent representations $\boldsymbol{z}_q$ into a volume-wise representation $\boldsymbol{Z}_q$. Subsequently, we consider a histogram representation to approximate the imaging appearance as a count of each feature vector on a volume basis, as follows:

$$\boldsymbol{Z}_q = \sum_{i \in I} f(\boldsymbol{x}) = \sum_{i \in I} \boldsymbol{z}_q \approx \sum_{i \in I} \mathrm{hist}_{k \in K}(c_{k_i}, e_k) = \mathrm{hist}_{k \in K}(c_k, e_k), \tag{5}$$

where K is the number of discrete feature vectors in the codebook, c_{k_i} is the frequency of occurrence of the k-th feature vector in the i-th axial slice, and c_k is the summed occurrence of the k-th feature vector appearing in the MRI volume.

Feature Extraction for Classification of Glioma Grades

Because the codebook is shared among a population, it is possible to identify the feature vectors used in classification tasks if simple classifiers, such as logistic regression, are adopted. By considering the frequency of occurrence c_i of each feature vector as an explanatory variable, the logistic regression model can be formulated as follows:

$$\mathrm{logit}(p) = \beta_0 + \sum_{k \in K^*} \beta_k c_k, \tag{6}$$

where p indicates the probability of a particular class, β is a regression coefficient, and K^* denotes a set of significant coefficients in the classifier. The set of feature vectors with coefficients that are statistically significant, based on an effect likelihood ratio test, in the logistic regression model are referred to as *responsible vectors*. We analyzed the frequency of each feature vector with statistical significance according to the glioma grade (LGG or HGG) using the Wilcoxon signed-rank test with a significance level of 0.05. If a feature vector is significantly frequent in LGG patients, it is called an *LGG responsible vector*. Similarly, *HGG responsible vectors* are defined as common feature vectors in HGG patients.

Feature Ablation Study to Visualize Responsible Regions

A feature ablation study was conducted to assess which types of imaging characteristics are encoded by the responsible vectors. The overall scheme is presented in Figure 2. The encoder and decoder trained in the segmentation network were used.

The objective of the feature ablation study is to visualize the type of imaging characteristics that are encoded by a specific feature vector. First, an input image is projected onto a corresponding latent representation by the encoder (Figure 2a). The quantized latent representation $\boldsymbol{z}_q$ is then fed into the decoder to generate the logit map $\tilde{\boldsymbol{y}}$, which is subsequently converted into the segmentation output $\hat{\boldsymbol{y}}$ through the argmax function. Here, the logit map $\tilde{\boldsymbol{y}}$ is retained for further processing. Next, the feature vector of interest in the initial latent representation $\boldsymbol{z}_q$ is replaced with a background vector, which is defined as the most common vector in the background of the images (for example, the black region outside the body in MRI). The replaced latent representation $\boldsymbol{z}'_q$ is subsequently input into the decoder and the corresponding logit map $\tilde{\boldsymbol{y}}'$ is retained. Finally, the per-pixel L1 difference between the two logit maps, $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{y}}'$, is evaluated. Because the difference map reflects the changed segmentation output caused by the replacement, we can assess the imaging characteristics encoded by each feature vector in the image space. Therefore, we call this difference map the *responsible region* (Figure 2b). The responsible regions from all LGG responsible and HGG responsible vectors are collectively denoted as *LGG responsible region* and *HGG responsible region*, respectively.

Dataset

We considered brain MRIs containing gliomas from the 2019 BraTS Challenge[24–27]. This dataset contains T1, Gd-enhanced T1, T2, and FLAIR sequences for patients diagnosed with LGG or HGG. In this study, all four sequences were used and three types of datasets were obtained: a training dataset (MICCAI_BraTS_Training) containing 355 patients, a validation dataset (MICCAI_BraTS_Validation) containing 125 patients, and a test dataset (MICCAI_BraTS_Testing) containing 167 patients. Only MICCAI_BraTS_Training contains a patient-basis diagnosis of LGG (76 patients) and HGG (259 patients). With regard to segmentation labels, MICCAI_BraTS_Training originally contained three ground-truth segmentation labels for abnormalities: Gd-enhanced tumor (ET), peritumoral edema (ED), and necrotic and non-enhancing tumor core (NET). Under the supervision of expert radiologists, we segmented MICCAI_BraTS_Validation and MICCAI_BraTS_Testing into the three above-mentioned abnormal categories (ET, ED, and NET). Note that the names of the datasets given in the 2019 BraTS Challenge and the purpose of using each dataset in this study are different. To train the segmentation network, a dataset obtained by concatenating MICCAI_BraTS_Validation and MICCAI_BraTS_Testing was used. After training the segmentation network, a classification model (Eq. 6) was constructed based on MICCAI_BraTS_Training, which is the only dataset containing information on the glioma grades. The classification performance of the logistic regression model was assessed through five-fold cross-validation.

Experimental Settings

Here, the implementation details of the segmentation network and training configuration are described.

Encoder Implementation

The encoder consists of residual blocks[28], wherein two [*convolution* + *group normalization*[29] + *LeakyReLU*] sequences are processed with residual connection. The kernel size, stride, and padding size of the convolution function in the residual blocks are set to 3, 1, and 1, respectively. From the first to the last residual blocks, the encoder uses $32-64-128-128-128-128$ filter kernels. Each residual block is followed by a downsampling block to halve the feature map size, except for the bottom of the network. The downsampling block consists of a sequence of [*convolution* + *group normalization* + *LeakyReLU*], whose kernel size, stride, and padding size are set to 3, 2, and 1, respectively. The input image is required to have a size of $4 \times 256 \times 256$ ($=$ channel $\times$ height $\times$ width). The encoder output, which is denoted as $\boldsymbol{z}_e$, has a size of $64 \times 8 \times 8$.

Decoder Implementation

The decoder architecture is approximately symmetrical to that of the encoder. From the first to the last residual block, the decoder uses $128-128-128-128-64-32$ filter kernels. The residual blocks consist of two [*convolution* + *group normalization* + *LeakyReLU*] sequences that follow an upsampling layer using an interpolation function coupled with a convolutional function to double the size of the feature map. Latent variables sampled from $p(\boldsymbol{z})$ with a size of $64 \times 8 \times 8$ pass through the decoder to yield reconstructed 2D images with a size of $4 \times 256 \times 256$.

Training Setups for Segmentation Model

All neural networks were implemented using Python 3.7 with the PyTorch library 1.6.0[30] on an NVIDIA Tesla V100 GPU with CUDA 10.0. He initialization[31] was applied to all networks. Adam optimization[32] with a learning rate of 1×10^{-4} was used for the segmentation network. The other hyperparameters were empirically determined as follows: batch size=72, maximum number of epochs=600. The size of the latent codebook is 512×64 ($= K \times D$). During training, the data augmentation included horizontal flipping, random rotation, and random-intensity shifting and scaling.

Results

Segmentation Performance of Segmentation Network

The segmentation performance of the segmentation network based on the Dice score is as follows: 0.42 for NET, 0.66 for ED, and 0.68 for ET. These intermediate Dice scores are expected, because the segmentation network has a bottleneck where the imaging features are compressed according to the limited size of the codebook. Notably, the primary objective of the segmentation network is not segmentation, but rather to provide a shareable set of feature vectors that sufficiently cover the imaging phenotypes of gliomas and are discriminative in downstream tasks.

Histogram Representation

Figure 3 shows the average histogram representations of HGG patients and LGG patients. These histograms indicate the average value of how many times each feature vector appears per MRI volume according to the glioma grading. A slight difference can be observed between these two histograms, particularly with regard to low-frequency feature vectors.

Classification Accuracy

According to five-fold cross-validation, the classification results (mean $\pm$ standard deviation) of the glioma-grading model are as follows: $0.90 \pm 0.02\%$ of accuracy, 0.79 ± 0.05 of precision, 0.72 ± 0.09 of recall (sensitivity), and 0.92 ± 0.02 of negative predictive value. With regard to the negative predictive value, HGG is considered to be positive and LGG is considered to be negative. After evaluating the classification performance based on five-fold cross-validation, we trained the classification model again on all samples, without leaving any samples out for further analysis. Additionally, the classification model identified two HGG responsible vectors and three LGG responsible vectors, which were significant covariates in the logistic regression models (effect likelihood ratio test: $p < 0.05$) and had significantly uneven distribution according to the glioma grading (Wilcoxon signed-rank test: $p < 0.05$).

Qualitative evaluation of responsible regions

As demonstrated by the classification performance, the feature vectors in the codebook appear to represent the imaging characteristics of gliomas and may convey meaningful information to identify the glioma grade. Therefore, we investigated which types of imaging characteristics are encoded by each feature vector by conducting a feature ablation study. To this end, we visualized both the HGG responsible regions and LGG responsible regions to evaluate the overlap with the segmented tumor regions that were provided as ground-truth labels.

Figure 4 shows the distribution of the HGG responsible regions and LGG responsible regions in patients with HGG. Notably, the HGG responsible regions are strongly correlated with the tumor regions of the HGG patients. The large difference values (indicated in red color) were preferentially gathered in the central region of the tumor corresponding to the ET label. In contrast, although a small overlap with the LGG responsible regions was observed in the peripheral regions of the tumor, the values were

relatively low as indicated by the color map. Therefore, it is concluded that the discriminative imaging characteristics of HGG are characterized by the preferential localization of ET at the central regions of the disease.

Figure 5 presents the distribution of HGG responsible regions and LGG responsible regions in patients with LGG. In contrast to the above-mentioned results, the LGG responsible regions significantly overlapped with the central region of the tumor, and particularly the region labelled as NET. The signals of the HGG responsible regions were not significant, as indicated by their low values. Thus, it is concluded that the discriminative imaging characteristics of LGG are associated with the NET regions in the central areas.

Quantitative evaluation of responsible regions

Finally, we quantitatively evaluated the preferences of each responsible region according to the ET, ED, and NET segmentation labels. The difference values in each segmented area were summed and statistically compared, as shown in Figure 6. For the HGG responsible regions, the mean and standard deviation values for the NET, ED, and ET labels were 5.48 ± 4.69, 3.78 ± 2.79, and 7.66 ± 5.37, respectively. The Kruskal-Wallis test and the non-parametric comparisons carried out for all pairs using the Dunn method for joint ranking revealed that the highest values appeared in the ET region ($p < 0.0001$). For the LGG responsible regions, the values for the NET, ED, and ET labels were 1.22 ± 1.26, 1.02 ± 1.10, and 0.92 ± 1.02, respectively. The same statistical tests revealed that the highest values appeared in the NET region ($p < 0.0001$). Because these quantitative observations are consistent with the qualitative results, it is concluded that the imaging characteristics associated with the discrimination of HGG and LGG are indicated by their localization in the ET and NET regions, respectively. In other words, the discriminative information can be considered in terms of whether contrast enhancement is present or not in the central area of the tumor.

Discussion

Multi-parametric MRI can reveal the morphological heterogeneity of gliomas, which contain various sub-regions (edematous regions, enhancing and non-enhancing tumor cores) with various histological and genomic phenotypes. This intrinsic heterogeneity can also be observed in imaging phenotypes because their sub-regions exhibit different intensity patterns across different MR sequences. In this study, three different regions were considered. The ET is defined by areas exhibiting hyper-intensity in the Gd-enhanced T1 sequences compared with T1 signals[25]. Such regions generally correspond to areas of contrast enhancement, where contrast leakage caused by blood-brain barrier damage may exist[33,34]. The ED is defined by areas with high T2/FLAIR signal intensity[25], which represent either low cellularity or edema[35]. The NET indicates non-enhancing tumor regions and pre-necrotic and/or necrotic regions located in the non-enhancing part of the tumor core[25]. The imaging appearance of NET typically exhibits hypo-intensity in the Gd-enhanced T1 sequences compared with T1 signals[25].

The imaging differences between LGG and HGG have attracted a substantial amount of attention with regard to early differential diagnosis. However, these differences are still debated. Typically, LGG appear as an area of focal signal abnormality with minimal or no contrast enhancement[36], and do not cause significant blood-brain barrier disruption, which results in less contrast leakage around the lesions. In contrast, most HGG in Gd-enhanced T1 sequences exhibit moderate to strong contrast enhancement, which reflects the degree of microvascularity and the presence of a disrupted blood-brain barrier[37]. Occasionally, necrosis can be observed inside a tumor, and is an important diagnostic feature for HGG[38]. Furthermore, HGG commonly cause significant damage to the blood-brain barrier, which appears as a large ED area surrounding the tumor core. Therefore, based on the segmentation categories adopted in this study, the presence of NETs in the central region of a tumor surrounded by a small ED region can be considered as a typical LGG characteristic. For HGG, a tumorous lesion represented by ET with or without NET and extensively surrounded by ED areas can be considered as a typical representation.

Based on these considerations, our results are in good agreement with the known imaging characteristics of LGG and HGG. Moreover, the feature ablation study revealed that NET is the most discriminative component of LGG, while ET is the most discriminative component of HGG. The presence of contrast enhancement (ET) is often considered as a sign of HGG[39]. Therefore, the fact that the classification model captured the presence (ET) or absence (NET) of contrast enhancement in the tumor core is compelling.

Several studies have investigated the classification of glioma grades using deep learning. For example, Yang et al. demonstrated that ImageNet-pretrained deep learning models, such as AlexNet[40] and GooleNet[41] can outperform a comparative model trained from scratch, and achieve a maximum test accuracy above 90%[14]. However, their method requires the manual segmentation of the ROIs before the classification. Recently, Banerjee et al. proposed a deep-learning-based algorithm that incorporates volumetric tumor information and achieves a maximum accuracy of 97%[15]. Similarly, Zhuge et al. proposed a two-step approach to automatically segment brain tumor regions and carry out classification according to the bounded image regions that contain tumors[16]. They also achieved a maximum classification accuracy of 97%. To achieve superior performance, an important aspect of deep-learning-based models is the size and extent of the input images. Banerjee et al.[15] compared several neural networks using patch-wise, slice-wise, and volume-wise inputs, and achieved glioma grading accuracy of 82%, 86%,

and 95%, respectively. Particularly, when considering the input as a 3D volume, these deep-learning-based approaches can outperform machine-learning-based approaches that use logistic regression based on brain tumor radiomics features (accuracy of 88%)[42].

Compared with previous studies, the classification accuracy of the proposed model is ranked between the accuracy achieved when using slice-wise inputs and the accuracy achieved when using volume-wise inputs[15]. Even though the proposed feature extraction process was performed using slice-wise inputs, the classification model is as simple as using logistic regression. Therefore, the proposed classification model's performance is remarkable compared with that of end-to-end deep learning models that take slice-wise inputs. Notably, Rudin[43] insisted that the belief whereby more complex models are more accurate is not always true, particularly when a good representation in terms of meaningful features is constructed for a target task. She also argued that there is often no significant difference between the prediction accuracy achieved by more complex models, such as deep neural networks, and much simpler models, such as logistic regression, when the representative data features are given. Accordingly, we confirmed that the feature vectors obtained from the pre-task of tumor segmentation are sufficiently informative for the discrimination of glioma grading.

To the best of our knowledge, this is the first study that uses vector quantization to obtain a shareable set of feature vectors across a population for the purpose of identifying specific factors associated with clinical information. Notably, logistic regression, which was adopted as the classifier in this study, can be considered as *transparent* in terms of interpretability[44]. Therefore, the variables that significantly contribute to the prediction can be identified. Based on the identified set of feature vectors associated with glioma grade classification, we conducted a feature ablation study to visualize the imaging region encoded by each feature vector. The results are consistent with those reported in the literature and can equip physicians with an enhanced understanding of the inner reasoning process of classification models. Our simple deep radiomics approach leverages the internal representations obtained by deep neural networks in downstream tasks using transparent models, such as logistic regression, and is versatile and easily applicable to other research fields.

References

1. Liu, S. *et al.* Actionable attribution maps for scientific machine learning. In *International Conference on Machine Learning (ICML) Workshop on ML Interpretability for Scientific Discovery* (2020).
2. Wesseling, P. & Capper, D. WHO 2016 Classification of gliomas. *Neuropathol. Appl. Neurobiol.* **44**, 139–150 (2018).
3. DeAngelis, L. M. Brain tumors. *New Engl. J. Medicine* **344**, 114–123 (2001).
4. Louis, D. N. *et al.* The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* **131**, 803–820 (2016).
5. The Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *New Engl. J. Medicine* **372**, 2481–2498 (2015).
6. Sotoudeh, H. *et al.* Artificial intelligence in the management of glioma: Era of personalized medicine. *Front. Oncol.* **9**, 768 (2019).
7. Shaver, M. M. *et al.* Optimizing neuro-oncology imaging: A review of deep learning approaches for glioma imaging. *Cancers* **11** (2019).
8. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 1–9 (2014).
9. Xiao, T., Hua, W., Li, C. & Wang, S. Glioma grading prediction by exploring radiomics and deep learning features. In *ACM International Conference Proceeding Series*, 208–213 (2019).
10. Banerjee, S., Mitra, S., Masulli, F. & Rovetta, S. Glioma classification using deep radiomics. *SN Comput. Sci.* **1**, 209 (2020).
11. Chen, W., Liu, B., Peng, S., Sun, J. & Qiao, X. Computer-aided grading of gliomas combining automatic segmentation and radiomics. *Int. J. Biomed. Imaging* **2018** (2018).
12. Cho, H., Lee, S., Kim, J. & Park, H. Classification of the glioma grading using radiomics analysis. *PeerJ* **2018** (2018).
13. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
14. Yang, Y. *et al.* Glioma grading on conventional MR images: A deep learning study with transfer learning. *Front. Neurosci.* **12**, 804 (2018).
15. Banerjee, S., Mitra, S., Masulli, F. & Rovetta, S. Deep radiomics for brain tumor detection and classification from multi-sequence MRI. *arXiv preprint arXiv:1903.09240* (2019).

16. Zhuge, Y. *et al.* Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Med. Phys.* **47**, 3044–3053 (2020).

17. Nanayakkara, S. *et al.* Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study. *PLOS Medicine* **15**, e1002709 (2018).

18. Holm, E. A. In defense of the black box. *Science* **364**, 26–27 (2019).

19. van den Oord, A., Vinyals, O. & Kavukcuoglu, K. Neural discrete representation learning. In *Advances in Neural Information Processing Systems 30 (NeurIPS)*, 6306–6315 (2017).

20. Razavi, A., van den Oord, A. & Vinyals, O. Generating diverse high-fidelity images with VQ-VAE-2. In *Advances in Neural Information Processing Systems 32 (NeurIPS)*, 14866–14876 (2019).

21. Łukasz Kaiser *et al.* Fast decoding in sequence models using discrete latent variables. In *Proceedings of the 35st International Conference on International Conference on Machine Learning (ICML)* (2018).

22. Sudre, C., Li, W., Vercauteren, T. K. M., Ourselin, S. & Cardoso, M. J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *arXiv preprint arXiv:1707.03237* (2017).

23. Lin, T., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2999–3007 (2017).

24. Menze, B. H. *et al.* The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Med. Imaging* **34**, 1993–2024 (2015).

25. Bakas, S. *et al.* Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **4** (2017).

26. Bakas S *et al.* Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *The Cancer Imaging Arch.* DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q (2017).

27. Bakas S *et al.* Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *The Cancer Imaging Arch.* DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF (2017).

28. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778 (2016).

29. Wu, Y. & He, K. Group normalization. *arXiv preprint arXiv:1803.08494* (2018).

30. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32 (NeurIPS)*, 8024–8035 (2019).

31. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026–1034 (2015).

32. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. In *The 3rd International Conference on Learning Representations (ICLR)* (2015).

33. Stadlbauer, A. *et al.* Preoperative grading of gliomas by using metabolite quantification with high-spatial-resolution proton MR spectroscopic imaging. *Radiology* **238**, 958–969 (2006).

34. Dowling, C. *et al.* Preoperative proton MR spectroscopic imaging of brain tumors: Correlation with histopathologic analysis of resection specimens. *Am. J. Neuroradiol.* **22**, 604–612 (2001).

35. Kono, K. *et al.* The role of diffusion-weighted imaging in patients with brain tumors. *Am. J. Neuroradiol.* **22**, 1081–1088 (2001).

36. Sawlani, V. *et al.* Multiparametric MRI: Practical approach and pictorial review of a useful tool in the evaluation of brain tumours and tumour-like lesions. *Insights into Imaging* **11** (2020).

37. Burger, P. C. Malignant astrocytic neoplasms: Classification, pathologic anatomy, and response to treatment. *Semin. oncology* **13**, 16–26 (1986).

38. Raza, S. M. *et al.* Necrosis and glioblastoma: A friend or a foe? A review and a hypothesis. *Neurosurgery* **51**, 2–13 (2002).

39. Scott, J. N., Brasher, P. M., Sevick, R. J., Rewcastle, N. B. & Forsyth, P. A. How often are nonenhancing supratentorial gliomas malignant? A population study. *Neurology* **59**, 947–949 (2002).

40. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS)*, 1097–1105 (2012).

41. Szegedy, C. *et al.* Going deeper with convolutions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1–9 (2015).

42. Hsieh, K. L. C., Lo, C. M. & Hsiao, C. J. Computer-aided grading of gliomas based on local and global MRI features. *Comput. Methods Programs Biomed.* **139**, 31–38 (2017).

43. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).

44. Barredo Arrieta, A. *et al.* Explainable explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020).

Acknowledgements

This study was supported by JST CREST (Grant Number JPMJCR1689), JST AIP-PRISM (Grant Number JPMJCR18Y4), and JSPS Grant-in-Aid for Scientific Research on Innovative Areas (Grant Number JP18H04908). The RIKEN AIP Deep Learning Environment (RAIDEN) supercomputer system was used to perform the calculations.

Author contributions

K.K. conceived the experiments; K.K. conducted the experiments; K.K., M.M., and M.T. analyzed the results. All the authors discussed the results and reviewed the manuscript.

Additional information

Competing interests: K.K. and R.H. have received research funding from Fujifilm Corporation.

Data availability

Data analyzed during the current study are available from the corresponding author upon reasonable request.

Code availability

All source code described in this project can be available from the corresponding author upon reasonable request.

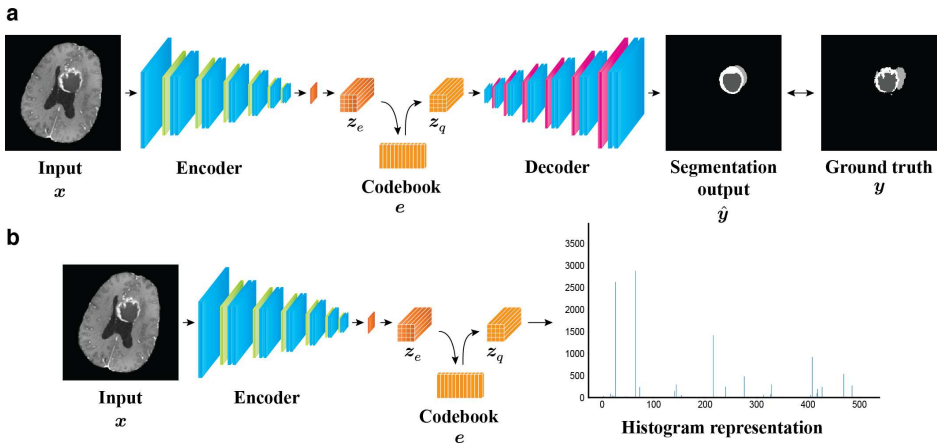

Figure 1. Obtaining a shareable set of feature vectors from a segmentation network. (**a**) A segmentation network consists of an encoder and decoder pair and stores a shareable set of feature vectors in a codebook. At the training stage of a tumor segmentation pre-task, an input image $\boldsymbol{x}$ is mapped onto a latent representation $\boldsymbol{z}_e$ through the encoder. Vector quantization is performed based on the codebook $\boldsymbol{e}$ by replacing each feature vector in $\boldsymbol{z}_e$ with the nearest codeword to produce a quantized latent representation $\boldsymbol{z}_q$. Then, the decoder produces a segmentation output by taking $\boldsymbol{z}_q$ as the input. The error between the segmentation output and a ground-truth label is evaluated to train the network. (**b**) When using the shareable set of feature vectors in a downstream task, the encoder is employed as a feature extractor. The latent representation of an input image is mapped onto the quantized latent representation $\boldsymbol{z}_q$, and then a histogram representation is constructed. This histogram representation contains information on the frequency with which each feature vector appears in the input image.

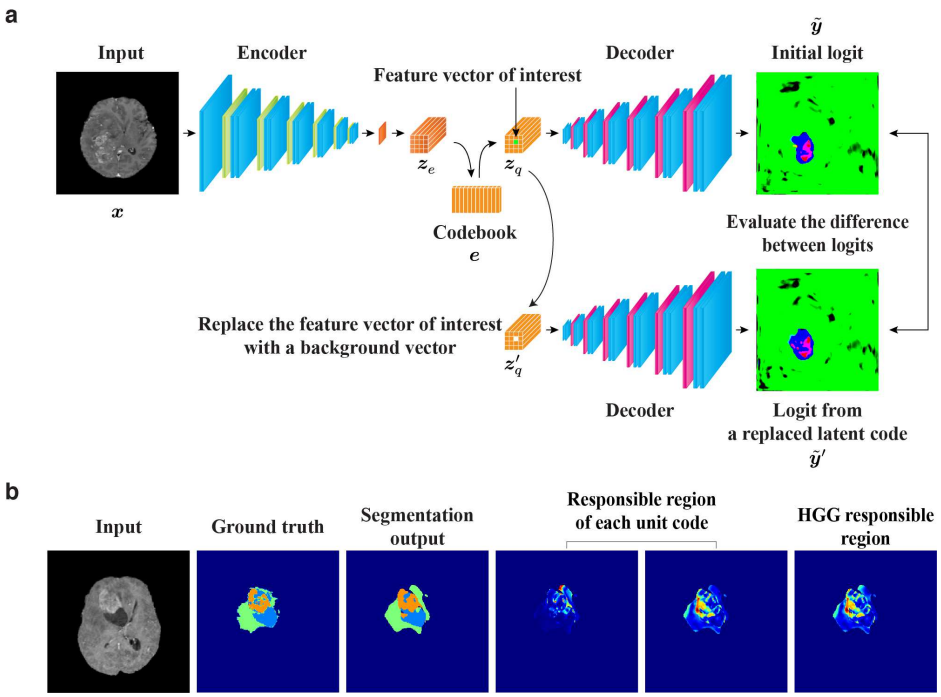

Figure 2. Overview of feature ablation study conducted to visualize the image region encoded by each feature vector. (**a**) The input image is initially mapped onto the quantized latent representation $\boldsymbol{z}_q$ through the encoder, which functions as a feature extractor. This initial latent representation is subsequently fed into the decoder to generate the segmentation output $\hat{\boldsymbol{y}}$, and the logit map $\tilde{\boldsymbol{y}}$ obtained before the final argmax operation is retained in the subsequent procedure. Then, the feature vector of interest in $\boldsymbol{z}_q$ is replaced with a background vector to generate the replaced latent representation $\boldsymbol{z}'_q$. The background vector is identified as the most common feature vector in the background of the images (for example, the region outside the body). Next, the decoder outputs the logit map $\tilde{\boldsymbol{y}}'$ again by taking $\boldsymbol{z}'_q$ as the input. Because the difference between $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{y}}'$ reflects the image region affected by the replacement, the difference map is referred to as the *responsible region* of the feature vector of interest. (**b**) The two responsible regions corresponding to the HGG responsible vectors are shown along with examples of an input image, ground-truth label, and segmentation output. By gathering these responsible regions from all responsible vectors for a particular glioma grade, we can observe which type of imaging characteristics are related to that glioma grade.

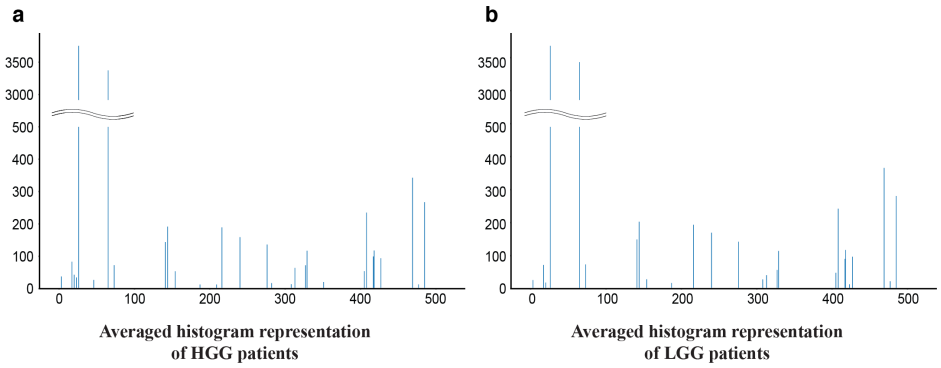

Figure 3. Average histogram representation for patients with (**a**) HGG and (**b**) LGG.

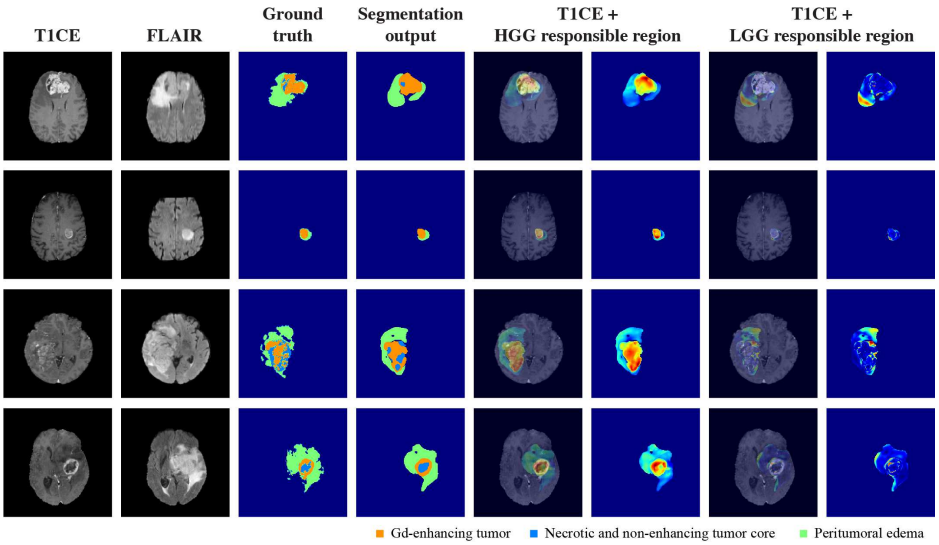

Figure 4. Example results for responsible regions in HGG patients. For patients with HGG, the Gd-enhanced T1 (T1CE) and FLAIR sequences, ground-truth labels, segmentation outputs, HGG responsible regions, and LGG responsible regions are shown. The tumor regions are adequately correlated with the HGG responsible regions, but overlap with the LGG responsible regions is scarce. The color map indicates the high-difference values in red and the lower-difference values in blue; the values are standardized for each patient.

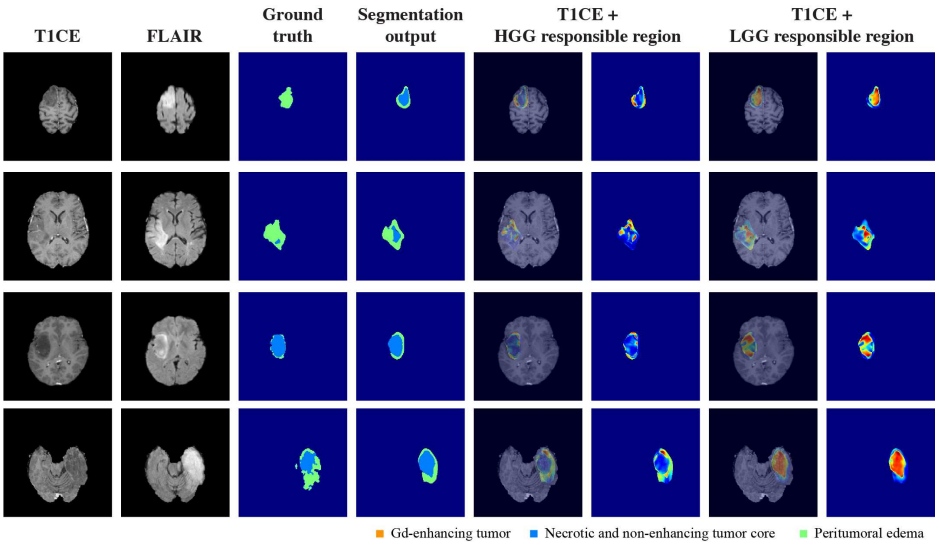

Figure 5. Example results for responsible regions in LGG patients. For patients with LGG, the Gd-enhanced T1 (T1CE) and FLAIR sequences, ground-truth labels, segmentation outputs, HGG responsible regions, and LGG responsible regions are shown. The tumor regions are strongly correlated with the LGG responsible regions, particularly in the central area of the tumor. The overlap with the HGG responsible regions is relatively insignificant and peripherally distributed at best. The color map indicates the high-difference values in red and the low-difference values in blue; the values are standardized for each patient.

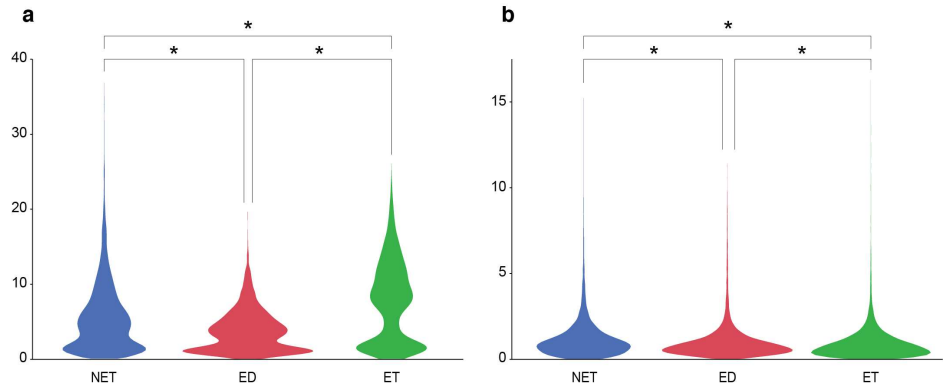

Figure 6. Quantitative evaluation of overlap between responsible regions and segmentation labels. (**a**) Difference values of HGG responsible regions in each segmentation label: Gd-enhanced tumor (ET), peritumoral edema (ED), and necrotic and non-enhancing tumor core (NET). The values in the ET region are the highest among the three segmentation categories. (**b**) Difference values of LGG responsible regions for the same segmentation labels. The NET regions have the highest values; * indicates a statistical significance < 0.0001.

Figures

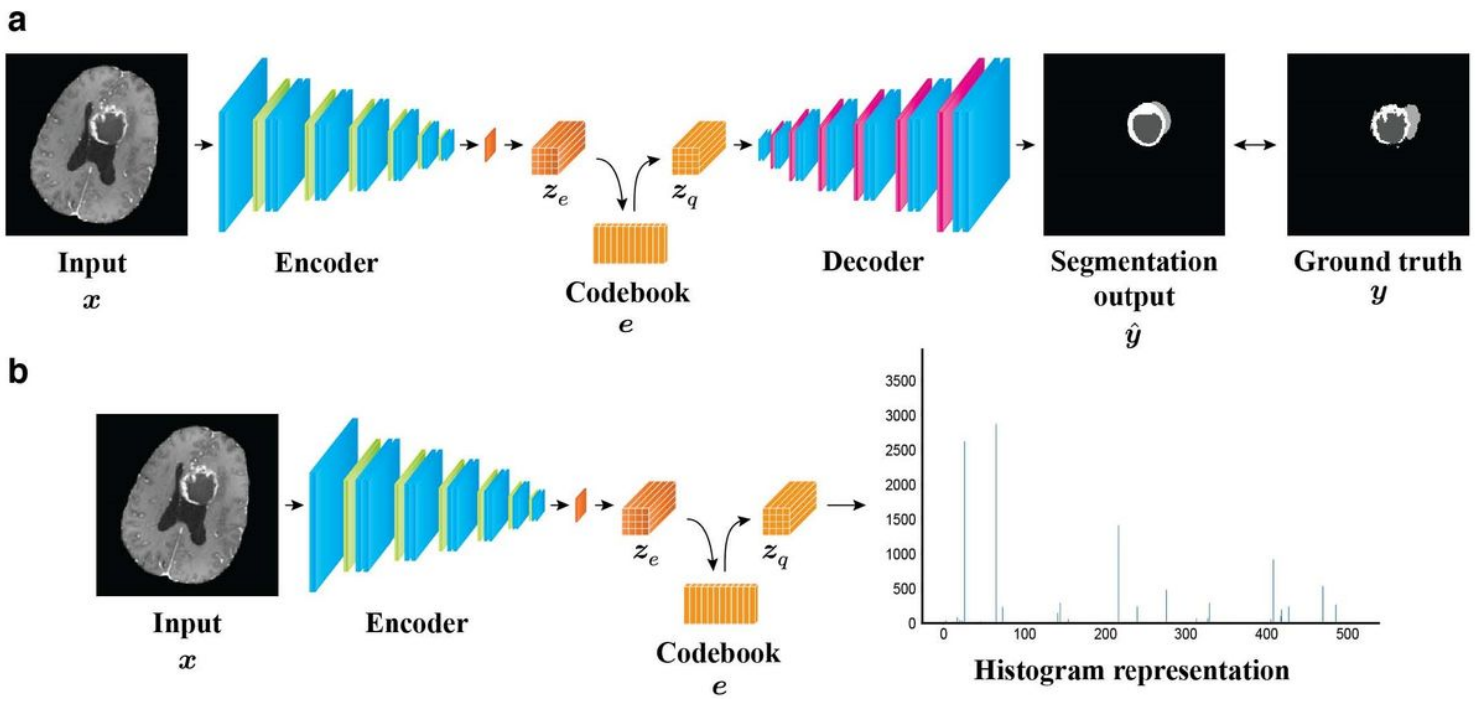

Figure 1

Obtaining a shareable set of feature vectors from a segmentation network. (a) A segmentation network consists of an encoder and decoder pair and stores a shareable set of feature vectors in a codebook. At the training stage of a tumor segmentation pre-task, an input image x is mapped onto a latent representation ze through the encoder. Vector quantization is performed based on the codebook e by replacing each feature vector in ze with the nearest codeword to produce a quantized latent representation zq. Then, the decoder produces a segmentation output by taking zq as the input. The error between the segmentation output and a ground-truth label is evaluated to train the network. (b) When using the shareable set of feature vectors in a downstream task, the encoder is employed as a feature extractor. The latent representation of an input image is mapped onto the quantized latent representation zq, and then a histogram representation is constructed. This histogram representation contains information on the frequency with which each feature vector appears in the input image.

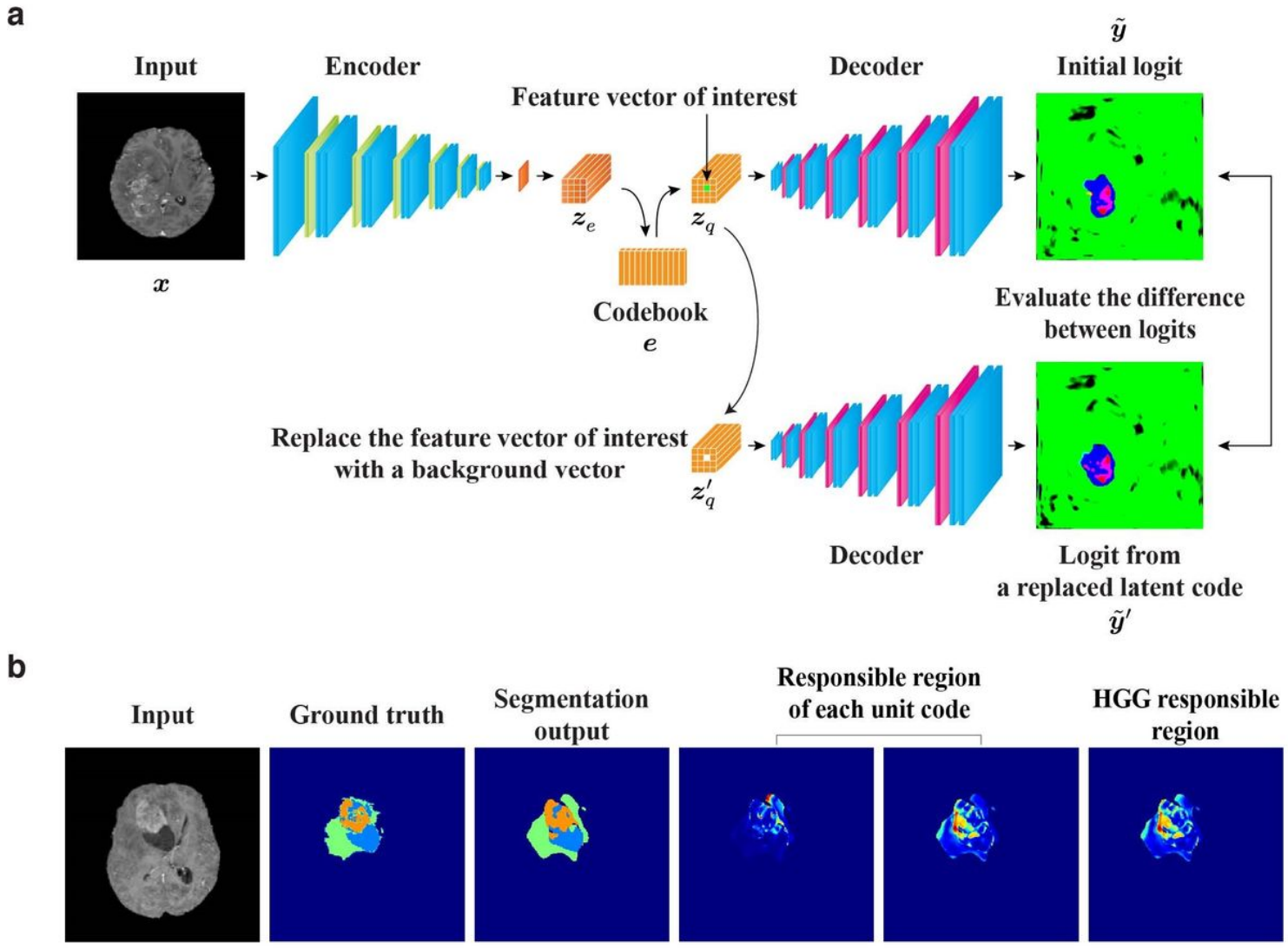

Figure 2

Please see the Manuscript PDF file for the complete figure caption.

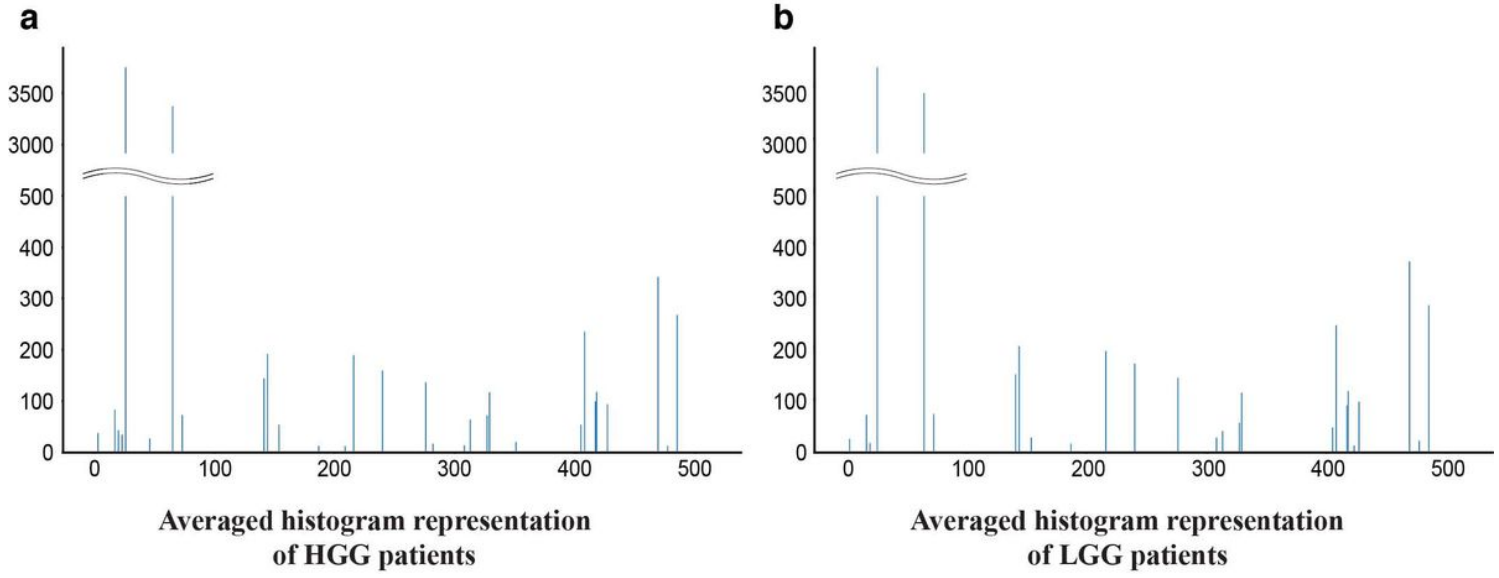

Figure 3

Average histogram representation for patients with (a) HGG and (b) LGG.

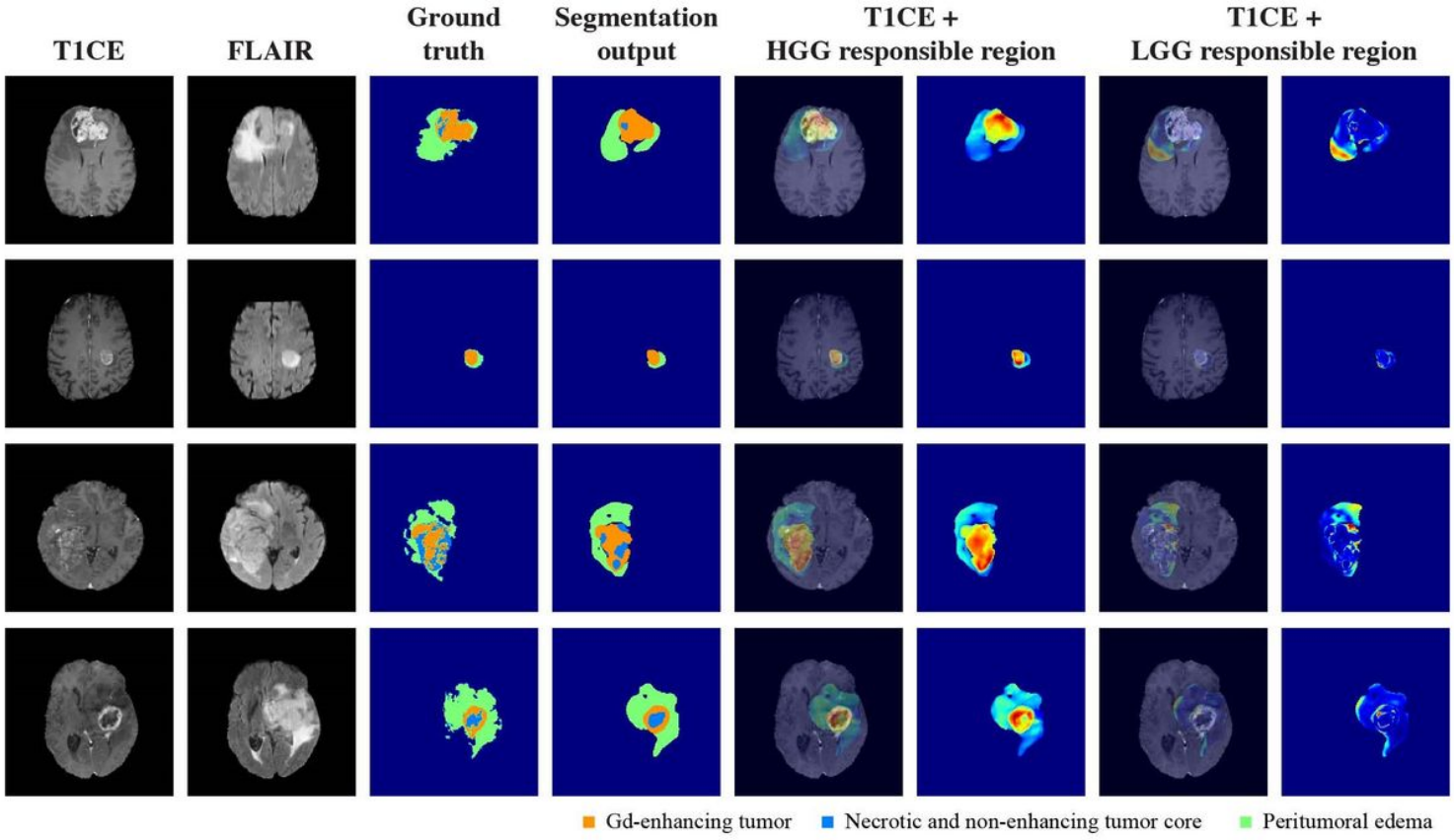

Figure 4

Example results for responsible regions in HGG patients. For patients with HGG, the Gd-enhanced T1 (T1CE) and FLAIR sequences, ground-truth labels, segmentation outputs, HGG responsible regions, and LGG responsible regions are shown. The tumor regions are adequately correlated with the HGG responsible regions, but overlap with the LGG responsible regions is scarce. The color map indicates the high-difference values in red and the lower-difference values in blue; the values are standardized for each patient.

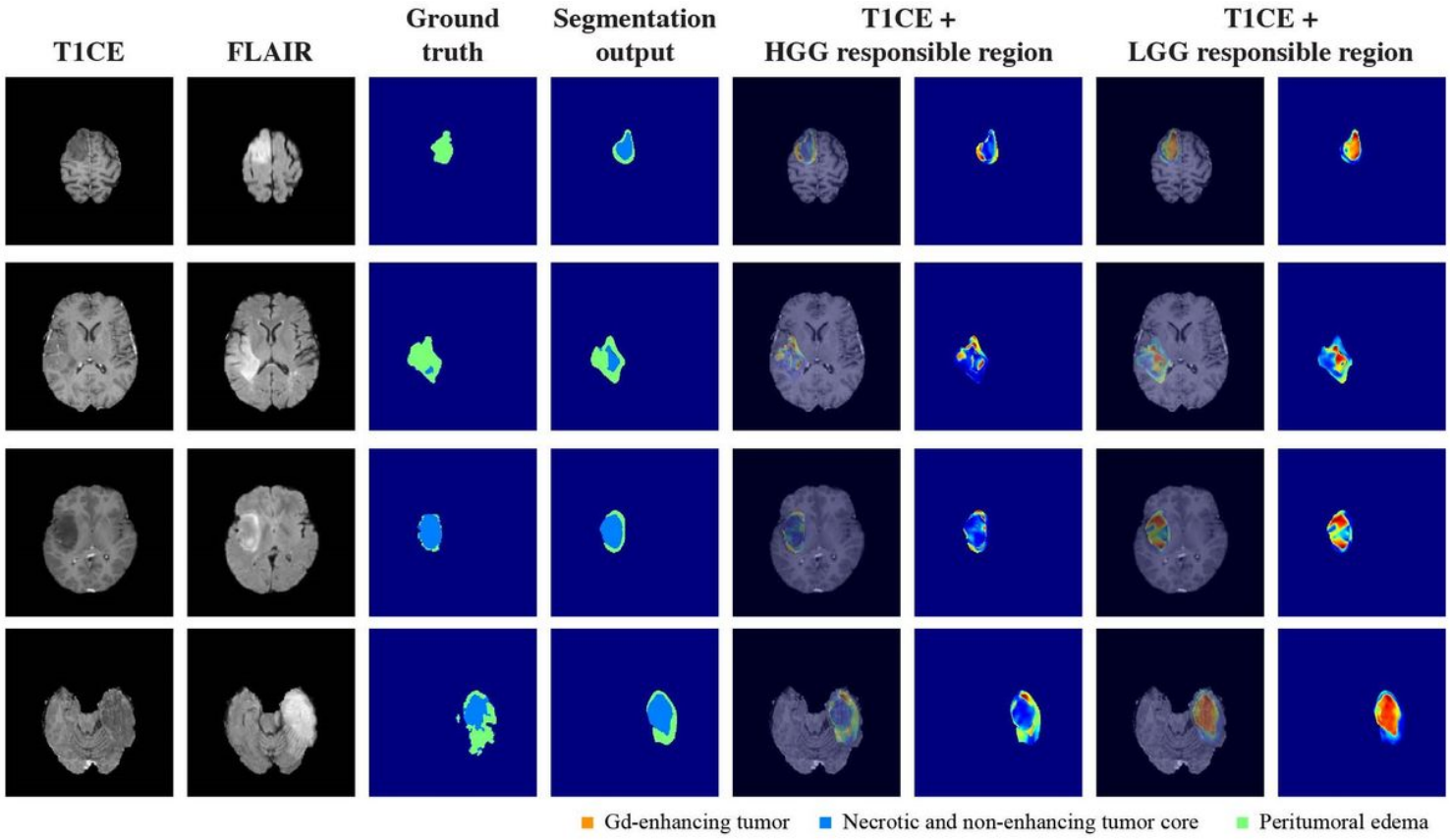

Figure 5

Example results for responsible regions in LGG patients. For patients with LGG, the Gd-enhanced T1 (T1CE) and FLAIR sequences, ground-truth labels, segmentation outputs, HGG responsible regions, and LGG responsible regions are shown. The tumor regions are strongly correlated with the LGG responsible regions, particularly in the central area of the tumor. The overlap with the HGG responsible regions is relatively insignificant and peripherally distributed at best. The color map indicates the high-difference values in red and the low-difference values in blue; the values are standardized for each patient.

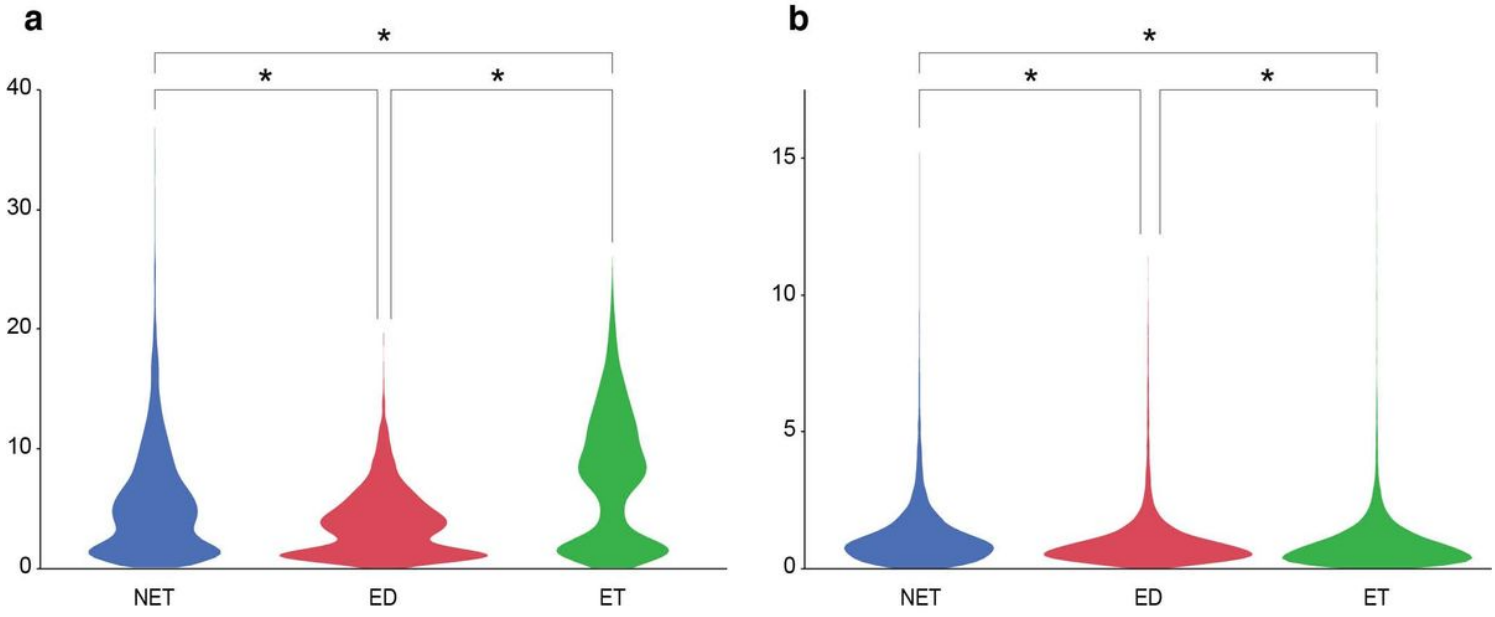

Figure 6

Quantitative evaluation of overlap between responsible regions and segmentation labels. (a) Difference values of HGG responsible regions in each segmentation label: Gd-enhanced tumor (ET), peritumoral edema (ED), and necrotic and non-enhancing tumor core (NET). The values in the ET region are the highest among the three segmentation categories. (b) Difference values of LGG responsible regions for the same segmentation labels. The NET regions have the highest values; * indicates a statistical significance < 0.0001.