Exploring the predictive value of additional peritumoral regions based on deep learning and radiomics: A multicenter study

Xiangjun Wu and Di Dong
School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Lu Zhang
Department of Radiology, The First Affiliated Hospital, Jinan University, Guangzhou, China

Mengjie Fang, Yongbei Zhu and Bingxi He
School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Zhaoxiang Ye[a)]
Department of Radiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China

Minming Zhang[a)]
Department of Radiology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

Shuixing Zhang[a)]
Department of Radiology, The First Affiliated Hospital, Jinan University, Guangzhou, China

Jie Tian[a)]
CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China
Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing, China

(Received 10 May 2020; revised 6 January 2021; accepted for publication 4 February 2021; published xx xxxx xxxx)

Purpose: The present study assessed the predictive value of peritumoral regions on three tumor tasks, and further explored the influence of peritumors with different sizes.
Methods: We retrospectively collected 333 samples of gastrointestinal stromal tumors from the Second Affiliated Hospital of Zhejiang University School of Medicine, and 183 samples of gastrointestinal stromal tumors from Tianjin Medical University Cancer Hospital. We also collected 211 samples of laryngeal carcinoma and 233 samples of nasopharyngeal carcinoma from the First Affiliated Hospital of Jinan University. The tasks of three tumor datasets were risk assessment (gastrointestinal stromal tumor), T3/T4 staging prediction (laryngeal carcinoma), and distant metastasis prediction (nasopharyngeal carcinoma), respectively. First, deep learning and radiomics were respectively used to construct peritumoral models, to study whether the peritumor had predictive value on three tumor datasets. Furthermore, we defined different sizes peritumors including fixed size (not considering tumor size) and adaptive size (according to average tumor radius) to explore the influence of peritumor of different sizes and types of tumors. Finally, we visualized the deep learning and radiomic models to observe the influence of the peritumor in three datasets.
Results: The performance of intra-peritumors are better than intratumors alone in three datasets. Specifically, the comparisons of area under receiver operating characteristic curve in the testing set between intra-peritumoral and intratumoral models are: 0.908 vs 0.873 (P value: 0.037) in gastrointestinal stromal tumor datasets, 0.796 vs 0.756 (P value: 0.188) in laryngeal carcinoma datasets and 0.660 vs 0.579 (P value: 0.431) in nasopharyngeal carcinoma datasets. Furthermore, for gastrointestinal stromal tumor datasets, deep learning is more stable to learn peritumors with both fixed and adaptive size than radiomics. For laryngeal carcinoma datasets, the intra-peritumoral radiomic model could make model performance more balanced. For nasopharyngeal carcinoma datasets, radiomics is also more suitable for modeling peritumors than deep learning. The size of the peritumor is critical in this task, and only the performance of 1.5 mm–4.5 mm peritumors is stable.
Conclusions: Our results indicate that peritumors have additional predictive value in three tumor datasets through deep learning or radiomics. The definitions of the peritumoral region and artificial intelligence method also have great influence on the performance of the peritumor. [https://doi.org/10.1002/mp.14767]

Key words: deep learning, peritumor, radiomics

Abbreviations

GIST	gastrointestinal stromal tumor
LC	laryngeal carcinoma
NPC	nasopharyngeal carcinoma
IPTR	intra- and peritumoral region
IPDL	intra- and peritumoral deep learning model
LASSO	least absolute shrinkage and selection operator
MRF	multiregion radiomic features fused model
IPRD	intra- and peritumoral radiomic model
IPCM	intra- and peritumoral combination model
CAM	class active map
AUC	area under receiver operating characteristic curve

1. INTRODUCTION

The peritumoral region or tumor microenvironment consists of various cell types including endothelial cells, fibroblasts, immune cells, and other types, as well as extracellular components.[1,2] The microenvironment determines many aspects of tumor behavior, including tumor progression, therapeutic response, and metastasis.[3–5] For example, Evans et al. reported that the peritumoral microenvironment was significantly associated with progression and metastasis in head and neck squamous cell carcinoma.[6] Carraro et al. also demonstrated that the peritumoral microenvironment played an important role in tumor resistance to neoadjuvant therapy.[7] To further improve therapeutic strategy precision, however, a range of peritumoral regions impacting on different tasks is required.

Radiomics has been used to mine the intratumor information based on many medical images for assisting the tasks of diagnosis or prognostics.[8–11] This strategy involves the extraction of quantitative manual features from regions of interest and correlates these features with specific tasks via machine learning algorithms.[12,13] Previous studies have explored the effect of peritumoral regions using radiomic methods, and peritumoral radiomic features were found to be associated with tumor prognosis.[14,15] Beig et al. reported that combining intramodular and perinodular regions improved model performance beyond that based on intramodular regions alone in the determining nonsmall cell lung cancer type.[16] Moreover, Braman et al. reported that the peritumoral environment, determined via radiomic methods, was associated with treatment response, and that a combination of intratumoral and peritumoral radiomic features could distinguish between the intrinsic molecular subtypes of HER2 + breast cancers.[17]

Although the peritumoral region is helpful for many diagnostic and prognostic tasks, the extent of influence of peritumoral regions of different sizes or types of tumors has rarely been studied. Furthermore, for small tumors, if the 30 mm peritumor is considered as the previous work, the peritumor is far larger than the intratumor, and even some other tissues have been introduced, this will cause additional pressure on the artificial intelligence models. Given these constraints, it is important to explore the influence of peritumoral regions of different sizes or types of tumors on methods of artificial intelligence.

In previous study, the methods for studying the peritumoral region almost are radiomics. Furthermore, the combined use of deep learning and radiomics has additional benefits, as has been reported previously. For instance, Ning et al. fused radiomic features and deep learning features to predict malignant potential for gastrointestinal stromal tumors, and achieved better performance than radiomics or deep learning alone.[18] Given this, it appears that deep learning is necessary for the analysis of peritumoral regions.

In the present study, we systematically analyzed the predictive value of peritumoral regions across different tumors and different tasks (Risk assessment in gastrointestinal stromal tumor datasets; T3/T4 staging prediction in laryngeal carcinoma datasets; Distant metastasis prediction in nasopharyngeal carcinoma datasets) through deep learning and radiomics. We further explored the influence of peritumoral regions by constructing a series of different sizes peritumoral models, the purpose of which is also to study which method is more suitable for learning peritumor in three tumor datasets. The study design is illustrated in Fig. 1.

2. MATERIALS AND METHODS

The present study used a retrospective design and was approved by the institutional review boards of all participating hospitals. All research was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

We collected imaging datasets of three types of tumors, namely gastrointestinal stromal tumor (GIST), laryngeal carcinoma (LC), and nasopharyngeal carcinoma (NPC) (See Table I for details). The inclusion criteria are shown in Methods S1. For GIST datasets, they were collected from two centers, so the data from hospital1 were used as the training set and the data from hospital2 were used as the testing set. For the nasopharyngeal cancer and laryngeal cancer datasets, they were collected from one hospital, so we divided the datasets according to the patient diagnostic time, taking 80% of the data as the training set and the remaining 20% of the data as the testing set.

We retrospectively collected 333 contrast-enhanced CT data from the Second Affiliated Hospital of Zhejiang University School of Medicine and 183 GIST from Tianjin Medical University Cancer Hospital, respectively, between 2009 and 2017. All patients were pathologically confirmed to have GIST. The task of GIST datasets is risk evaluation. According to NIH standards, GISTs were classified into four risk levels: very low, low, medium, and high.[19] We further defined very low, low, and medium as low risk and defined high as high risk.[20] By predicting the preoperative risk of GIST, it can provide valuable clues for predicting prognosis and assisting personalized clinical decision-making.

A total of 211 retrospective cases of LC contrast-enhanced CT data from the First Affiliated Hospital of Jinan University were collected between 2007 and 2017. All patients were

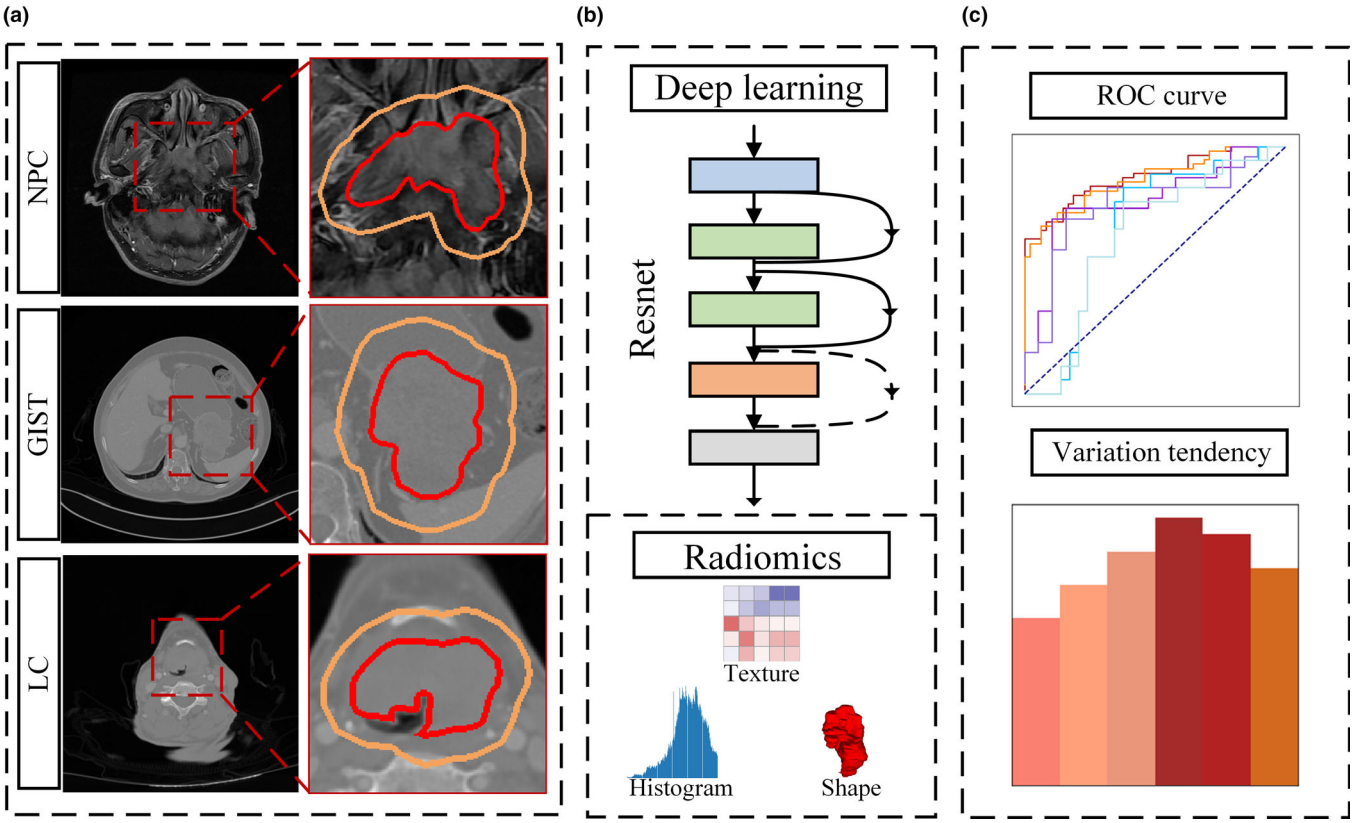

FIG. 1. A diagram depicting the experimental protocols used in this study. (a) is the step of segmentation of intratumoral regions and peritumoral regions in three tumors. (b) is the step of model construction based deep learning and radiomics. (c) is the step of model evaluation.

TABLE I. The information and datasets split of three tumor datasets.

	Training set	Time	Testing set	Time	Image modality	Task	Hospital
GIST	333	2009–2017	/	/	CT	Risk evaluation	The Second Affiliated Hospital of Zhejiang University School of Medicine
GIST	/	/	183	2011–2017	CT	Risk evaluation	Tianjin Medical University Cancer Hospital
LC	168	2007–2016	43	2016–2017	CT	T3/T4 stage classification	The First Affiliated Hospital of Jinan University
NPC	186	2007–2014	47	2014–2016	MR	Distant metastasis	The First Affiliated Hospital of Jinan University

Abbreviations: GIST = gastrointestinal stromal tumor; LC = laryngeal carcinoma; NPC = nasopharyngeal carcinoma.

pathologically confirmed to have T3 or T4 staging LC after surgery. The task of LC datasets is T3/T4 stage classification. Accurate prediction of preoperative T staging of LC can be beneficial to the decision of total laryngectomy or larynx-preserving treatment.

We retrospectively collected 233 NPC cases of MRI data from the First Affiliated Hospital of Jinan University between 2007 and 2016. All patients were followed for at least 3 years. The task of NPC datasets is to predict whether distant metastasis would occur. Distant metastasis was defined as in our prior work,[21] and treatment decisions can be improved by accurately stratifying the risk of distant metastasis of NPC.

2.A. Segmentation

Tumor regions for the three kinds of tumors included in the present study were manually segmented by radiologists. Tumor boundaries were first delineated by a radiologist with 5 yr of experience and then confirmed by a radiologist with 10 yr of experience. Boundaries were delineated using ITK-snap (version, 3.6) software. The maximum slice of the tumor was selected. Peritumoral regions are automatically obtained according to the delineation result for the intratumor region. We first determined the coordinates of the boundary points for the peritumoral region, as delineated by the radiologists, and then expanded these outward according to different requirements to determine the peritumoral region.

2.B. Definition of peritumor size

In previous studies, to extract peritumoral radiomic features, peritumoral regions were defined according to fixed sizes [e.g., 3–15 mm or 5–30 mm].[16,17] We considered the fact that tumor size as well as morphology vary greatly from

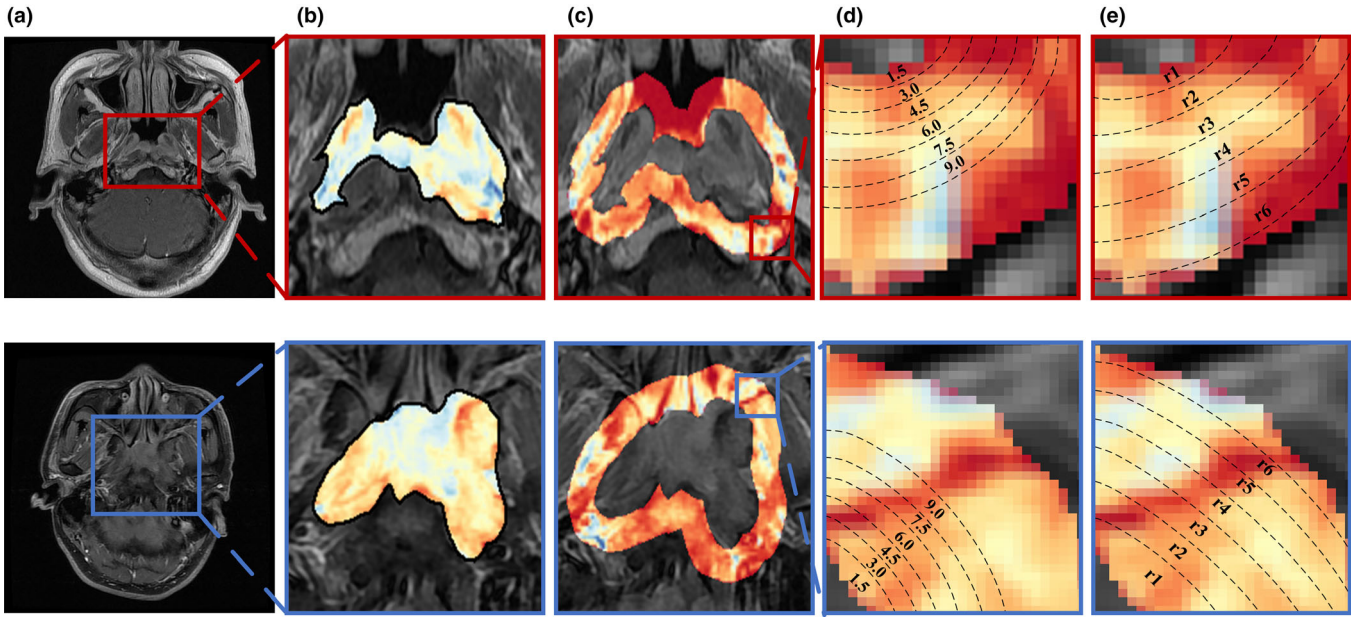

FIG. 2. A schematic diagram representing the definition of the peritumoral region. (a), the slices with the maximum tumor area were selected. (b), the intratumoral regions were segmented by expert radiologists. (c), the peritumoral regions were auto extended using an algorithm. (d), the peritumors with fixed sizes. (e), the peritumors with adapted sizes.

patient to patient and across tumors. So, to study the influence of different sizes of peritumors in the present experiment, we defined the adaptive peritumor according to the average tumor radius, and compared it with the fixed peritumor that did not vary with tumor size. The fixed peritumors were defined by millimeters [Fig. 2(d)], from 1.5 to 9 mm for LC and NPC, and from 2 to 12 mm for GISTs. The adaptive peritumors were defined by the average tumor radius, from 1/6 radius (r1) to the radius (r6) [Fig. 2(e)], and the average tumor radius was calculated according to the area of the tumor.

2.C. Peritumors analyzed through deep learning

After the performance comparison of different networks (more detail is shown in Tables S10, S12), the classification network we used is based on the structure of Resnet, that is, the structure of residual connection, which is a general learning structure that can fuse the features of shallow layer and deep layer in the deep neural network.[22] And Resnet has already been widely used in both natural images and medical images. For medical images, Resnet has shown good performance in different tasks, such as, Reddy et al used Resnet to classify malarial infected cells, and obtained a great performance on microscopic cell images,[23] and Hu et al used Resnet as features extraction to extract deep learning features to construct the model for distinguishing benign and malignant lesions of breast cancer.[24] Therefore, we chose Resnet as the classification network here. For the three data sets, we used the same classification network in order to fairly compare the influence range of the peritumor region. Specifically, we used the classification network of Resnet18, that is, the network contains 18 convolution and fully connected layers (more detail is shown in Table S11). We used CT or MR images to train the deep learning network, and they are grayscale images. So, the input size of the classification network is $512 \times 512 \times 1$. After a series of convolution operations, pooling operations and activation layer, a probability value was obtained at the end of the network, which was called the deep learning signature. Its distribution is from 0 to 1. The closer this value is to 1, the more likely it was predicted as a positive sample. Similarly, the closer this value is to 0, the more likely it was predicted as a negative sample.

As tumor region only occupies a very small part of the whole image, and the background outside the tumor and peritumor are noises for the deep neural network, which is not conducive to the network's full learning of the information of intratumor and peritumor. So, we used intratumor or intraperitumor as input by multiplying intratumor mask or intraperitumor mask with the original image to remove the background. After that, we performed data augmentation in the training set. The specific operation was to rotate the original image from different angles to get some new samples. The function of data enhancement is to increase the amount of data used to train the neural network, and to improve the generalization performance of the model. When training the network, in order to fairly compare the performance of intratumor, peritumor, and intra-peritumor, we used the same hyperparameters to train these three networks. The network optimizer is root mean square prop (RMSprop); Learning rate is 10^{-6}; Learning rate attenuation is 10^{-7}; And batchsize is 4.

2.D. Peritumors analyzed through radiomics

In order to ensure the fairness of comparison, features extraction and features selection methods of all radiomic

models were consistent. For each region, 107 radiomic features were extracted, each of which had three kinds of features: histogram, shape, and texture. The feature selection methods used were least absolute shrinkage and selection operator (LASSO) and multivariate analysis. Logistics was further used to fit models with significant radiomic features. Similar to the deep learning model, the radiomic features are fitted through the Logistics to obtain a probability value of 0–1 distribution range, which was called radiomic signature. We used the radiomic signature to calculate the performance of the model.

2.E. Combination of deep learning signature and radiomic signature

Deep learning is an end-to-end learning approach for images, and radiomics is the method used to analyze manually defined features extracted from region of interest. We fused the results of two methods to study if there were further improvements. Specifically, deep learning signature was fused with radiomic signature, both of which are probability values distributed in a range from 0 to 1, so they could be regarded as two features. We used a Logistics model to fit these two features, to obtain the final combination model.

2.F. Model evaluation and visualization

For the three tasks, the area under receiver operating characteristic curve (AUC), sensitivity (Sen) and specificity (Spe) were used to evaluate the model performance. To measure the robustness of the model, we also calculated the 95% confidence intervals for each evaluation metrics. We used the Delong Test to compare the performance differences between the two models, and $P < 0.05$ was considered to have a significant difference.

To demonstrate how the model learns peritumor, we visualized intratumoral and intra-peritumoral models. The class active maps were used to visualize the deep learning model, which was obtained by calculating the average gradient change of the input image to the label.[25,26] The greater the active value of the region, the greater the gradient of the model in the region. For the radiomic model, different features have different meanings, so they were visualized in different ways. Specifically, for shape features, we used the white arrow to signal; For histogram features, we used the specific value or statistical histogram to display; And for texture features, we calculated the neighborhood features of each point in the region of interest as the eigenvalues of that point, then the color map of texture features were obtained through color mapping (more detail is shown in Methods S2).

2.G. Statistical analysis

The deep learning network used in this study was created using the Keras framework (version, 2.1.6) based on TensorFlow (version, 1.10.1). Radiomic features were extracted using the pyradiomics library (version, 2.1.2) in Python (version, 3.6). Radiomic models were constructed using R language (version, 1.10.1). Experimental results were visualized using the matplotlib library (version, 3.1.1) in Python. The evaluate matrices of AUC, sensitivity, specificity, and confidence intervals were calculated using pROC (1.12.1) and reportROC (3.2) packages in R language.

3. RESULTS

3.A. Performance of intratumors and peritumors

For GIST datasets, the results are shown in Table II. The performance of the intratumoral deep learning model in the testing set are: AUC, 0.873, Sen, 0.635, and Spe, 0.877. The performance of the intra-peritumoral deep learning model in the testing set are: AUC, 0.908, Sen, 0.730, and Spe, 0.895. The statistical difference of the AUC between these two models is 0.037. The AUC of the intra-peritumoral model in the testing set is significantly better than intratumor. And the Sen improved by nearly 10%. For the radiomic model, although there was no significant difference between the performance of the intratumoral model and the intra-peritumoral model (AUC: 0.892 vs 0.890, P: 0.957), the Sen and Spe of the

TABLE II. The performance of the peritumor in the GIST datasets.

	Testing set			
	AUC	Sensitivity	Specificity	*P* value
Intratumor				
Deep learning	0.873 [0.820–0.926]	0.635 [0.551–0.719]	0.877 [0.792–0.962]	\
Radiomics	0.890 [0.841–0.934]	0.952 [0.915–0.990]	0.421 [0.293–0.549]	\
Peritumor				
Deep learning	0.840 [0.781–0.898]	0.571 [0.485–0.658]	0.895 [0.815–0.974]	\
Radiomics	0.893 [0.845–0.941]	0.651 [0.568–0.734]	0.965 [0.917–1.000]	\
Intra-peritumor				
Deep learning	0.908 [0.863–0.953]	0.730 [0.653–0.808]	0.895 [0.815–0.974]	0.037
Radiomic	0.892 [0.844–0.940]	0.754 [0.679–0.829]	0.930 [0.864–0.996]	0.957
Deep learning + Radiomics	0.891 [0.844–0.938]	0.730 [0.653–0.808]	0.930 [0.864–0.996]	

TABLE III. The performance of the peritumor in the NPC datasets.

	Testing set			
	AUC	Sensitivity	Specificity	*P* value
Intratumor				
Deep learning	0.579 [0.397–0.761]	0.312 [0.152–0.473]	0.800 [0.598–1.000]	\
Radiomics	0.608 [0.443–0.773]	0.688 [0.527–0.848]	0.400 [0.152–0.648]	\
Peritumor				
Deep learning	0.581 [0.406–0.757]	0.688 [0.527–0.848]	0.400 [0.152–0.648]	\
Radiomics	0.657 [0.490–0.825]	0.656 [0.492–0.821]	0.667 [0.428–0.905]	\
Intra-peritumor				
Deep learning	0.660 [0.484–0.837]	0.344 [0.179–0.508]	0.800 [0.598–1.000]	0.431
Radiomics	0.648 [0.481–0.815]	0.625 [0.457–0.793]	0.667 [0.428–0.905]	0.540
Deep learning + Radiomics	0.631 [0.463–0.799]	0.625 [0.457–0.793]	0.667 [0.428–0.905]	

intra-peritumoral model (Sen: 0.754, Spe: 0.930) were more balanced than intratumoral model (Sen: 0.952, Spe: 0.421).

For NPC datasets, the results are shown in Table III. The performance of intra-peritumoral deep learning and radiomic models are both better than intratumor in the testing set. In this task, the radiomic method is more suitable to learn peritumor. The Sen of intratumoral deep learning model is 0.312, and the Spe of intratumor radiomic model is 0.400. Such results indicate that the models cannot effectively distinguish between two classes of samples, and the results of model tend to be one class. Although intra-peritumoral deep learning model has some promotion in the AUC performance, the elevation to Sen is too faint. However, the intra-peritumor radiomic model can greatly improve Spe performance (Sen, 0.625; Spe, 0.667), thus effectively improving the generalization performance.

For LC datasets, the results are shown in Table IV. As for the deep learning method, the performance of intra-peritumor (AUC, 0.796; Sen, 0.667; Spe, 0.750) is slightly better than intratumor (AUC, 0.756; Sen, 0.593; Spe, 0.938) in the testing set. As for the performance of intra-peritumoral radiomic model and intratumoral radiomic model, although the AUC did not show any improvement, the results of Sen and Spe became more consistent (Sen, 0.741; Spe, 0.750). The Sen and Spe of intratumoral radiomic model are 0.630 and 0.875, respectively. This means that most of the samples were predicted to be negative, which is not good for the generalization performance.

3.B. Performance of different sizes peritumors

We compared the effects of a series of peritumors with different sizes on the two methods of deep learning and radiomics. The purpose of this experiment was to explore influence of peritumors for different tumors and different tasks, and to study the strengths and weaknesses of deep learning and radiomics in learning different size peritumors.

For the GIST datasets, Fig. 3 compares the performance of different peritumor models, and the p values of each pair models. The specific results are in Tables S1, S2. It can be seen from Fig. 3 that for deep learning, the performance of the intra-peritumor models is relatively stable with the expansion of the peritumor, and most of the intra-peritumor models are better than the intratumor. However, for radiomics, with the change of peritumors, the performance of the intra-

TABLE IV. The performance of the peritumor in the LC datasets.

	Testing set			
	AUC	Sensitivity	Specificity	*P* value
Intratumor				
Deep learning	0.756 [0.606–0.903]	0.593 [0.407–0.778]	0.938 [0.819–1.000]	\
Radiomics	0.796 [0.663–0.930]	0.630 [0.447–0.812]	0.875 [0.713–1.000]	\
Peritumor				
Deep learning	0.745 [0.588–0.903]	0.556 [0.368–0.743]	0.875 [0.713–1.000]	\
Radiomics	0.674 [0.501–0.846]	0.481 [0.293–0.670]	0.812 [0.621–1.000]	\
Intra-peritumor				
Deep learning	0.796 [0.657–0.936]	0.667 [0.489–0.844]	0.750 [0.538–0.962]	0.188
Radiomics	0.764 [0.614–0.914]	0.741 [0.575–0.906]	0.750 [0.538–0.962]	0.236
Deep learning + Radiomics	0.771 [0.622–0.920]	0.741 [0.575–0.906]	0.750 [0.538–0.962]	

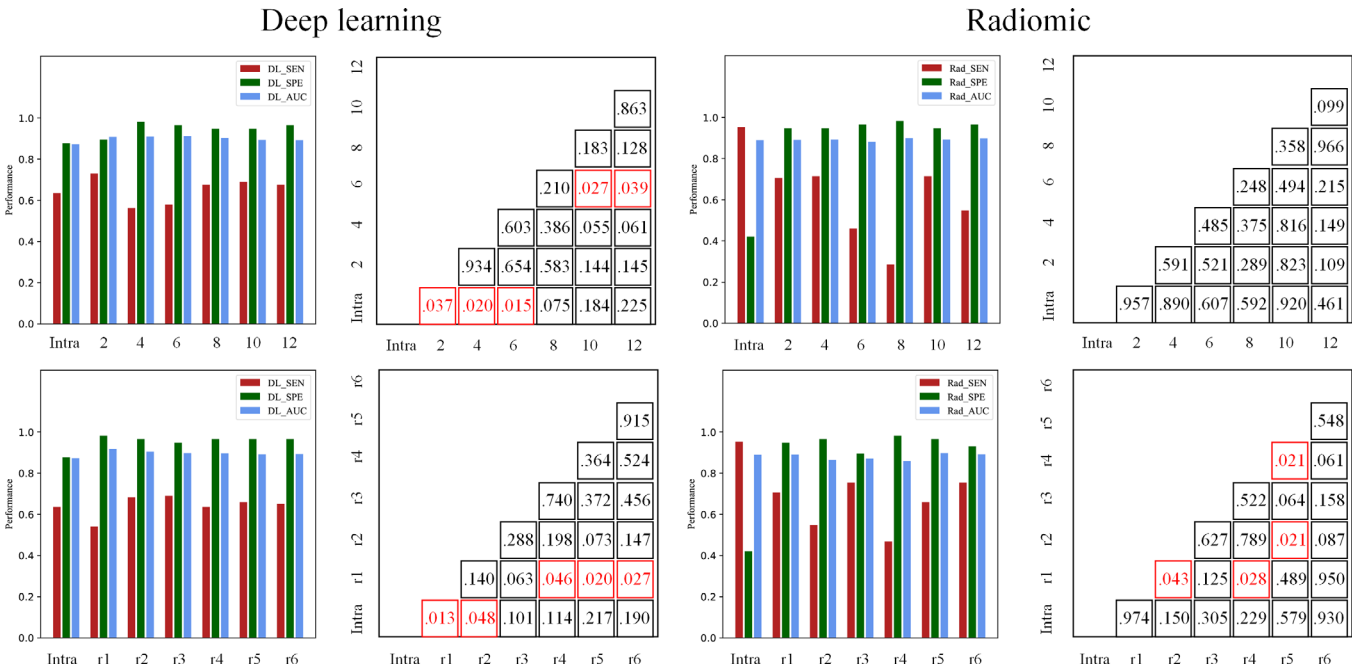

FIG. 3. The performance of peritumors with different sizes in GIST datasets.

peritumor models varies greatly, and there is a large gap between the Sen and Spe of the models. This indicates that for the GIST datasets, deep learning is a better way to learn the peritumor region, and has relatively stable performance in different sizes of peritumors.

For NPC datasets, Fig. 4 compares the performance of different peritumor models, and specific results are in Tables S3, S4. According to Fig. 4, with the change of peritumor, the Spe value of many radiomic models become 0, which means that the models predict all samples in the testing set into one class. And the model has no discrimination ability at all, or the Spe values are much higher than Sen (deep learning model with fixed size peritumor). This means that for the NPC datasets, the definition of peritumor size is very important for peritumor performance, and the smaller peritumor is more stable. In addition, only the radiomic models based on fixed size peritumors are stable, which can reduce the gap between Spe and Sen. This indicates that for NPC datasets, radiomics is more suitable for modeling peritumors than deep learning.

For LC datasets, Fig. 5 compares the performance of different peritumor models, and specific results are in Tables S5, S6. In the LC datasets, the fixed and adaptive size peritumors both show stable performance in deep learning and radiomic models. However, for the deep learning method, the Spe of the intratumor model is much higher than the Sen, and the intra-peritumoral model cannot effectively improve this gap. For the radiomic method, most of the peritumor models can make the SEN, SPE, and AUC of the model tend to be at the same level, so that the performance of the model is more balanced. This indicates that for the LC datasets, radiomics is more suitable for modeling of peritumors in this task than deep learning.

The radiomic features used in each radiomic model are displayed in Tables S7, S8, S9. Many intra-peritumoral models used the same features as the intratumoral model, but their performance are superior. These results indicated that the peritumors have additional valuable information and can improve the stability of the radiomic features. For GIST datasets, "Maximum_2D_Diameter_Row" is the most stable feature, which represents the size of tumor. For NPC datasets, "Shape_Maximum_2D_Diameter_Row" and "Texture_Gldm_Large_Dependence_Low_Gray_Level_Emphasis" are the most stable features. Although these two features were used in many intra-peritumoral models, the model performance collapses with the expansion of peritumor. This indicates that the excessively large peritumor would introduce extra noise, which makes the radiomic features lose the ability to characterize the tumor. For LC datasets, "Shape_Elongation" and "Texture_Glszm_Zone_Percentage" were almost used in all intra-peritumoral models. In this task, the performance of radiomic features is stable with the expansion of peritumor.

3.C. Model visualization

For GIST datasets, the deep learning and radiomic models are visualized in Fig. 6. First, the red arrows indicate the regions of interest (highlighted regions) of the deep learning model, that is, the regions that contributes a lot to the model result. From the regions indicated by the arrows, a large part of the weights of the deep learning model falls on the margins of the tumor. This indicates that the tumor marginal region has a greater impact on the results of deep learning model. Second, the intratumoral radiomic model was constructed by two shape features: "Shape_Maximum_2D_Diameter_Row" and "Shape_Sphericity." They

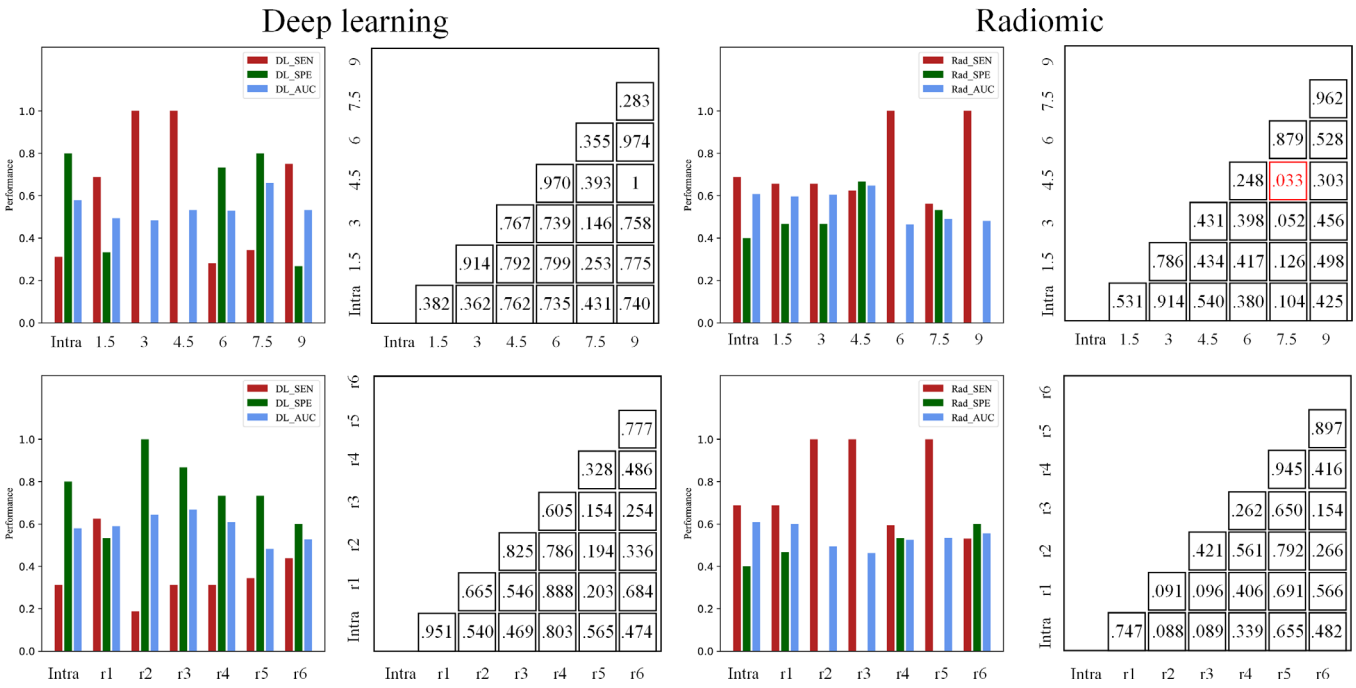

FIG. 4. The performance of peritumors with different sizes in NPC datasets.

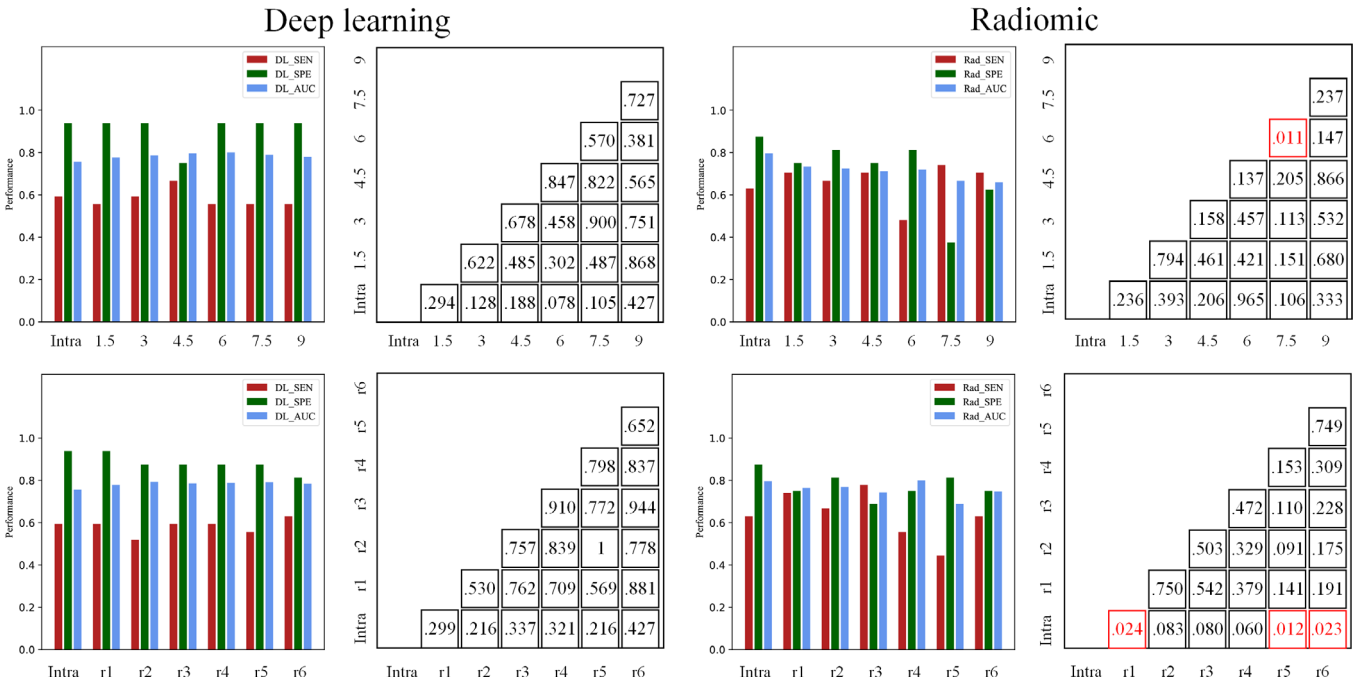

FIG. 5. The performance of peritumors with different sizes in LC datasets.

respectively represent the size and sphericity of the tumor. The larger and more irregular of the tumor may bring the greater risk. As for intra-peritumoral model, three texture features were used, including "Texture_Glszm_GrayLevel_Non_Uniformity," "Texture_Glszm_Zone_Entropy," and "Texture_Ngtdm_Contrast." It can be seen from Fig. 6 that, unlike the intratumoral shape features, the features used in intra-peritumoral model represent the gray intensity and regional texture changes of the intra-peritumor, especially at the marginal region of the tumor.

For NPC datasets, the deep learning and radiomic models are visualized in Fig. 7. Similar to GIST datasets, peritumor occupies the main weights of the deep learning model. However, due to the limited amount of data, the intra-peritumor deep learning model did not show obvious advantages than

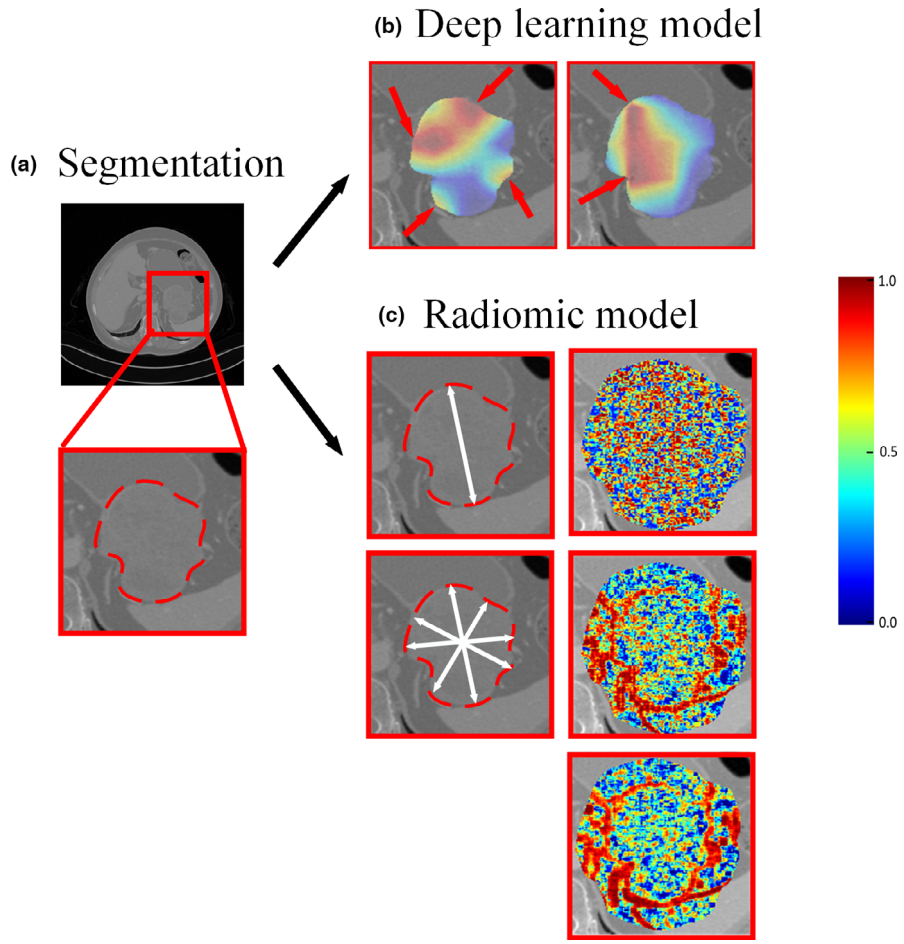

FIG. 6. The class active map of the deep learning model and features color map of the radiomic model in GIST datasets. (a), represents the segmentation of the tumor. In (b) and (c), the left column represents the intratumoral models, and the right column represents the intra-peritumoral models.

intratumor alone. Both the intratumoral and intra-peritumoral radiomic model only contained a shape feature: "Shape_Maximum_2D_Diameter_Row." This shows that in our datasets, for the radiomic model, the maximum diameter of the tumor is the largest factor that affects the distant metastasis of NPC.

For LC datasets, the deep learning and radiomic models are visualized in Fig. 8. There are obvious differences between the regions of high weights in the intratumoral and intra-peritumoral models, which may be the reason causing the imbalanced performance of deep learning model. The significant radiomic features are stable in intratumors and intra-peritumors, including a shape feature ("Shape_Elongation") and a texture feature ("Texture_Glszm_Zone_Percentage"). The shape feature represents the elongation rate of the tumor, and the texture feature represents the regional texture roughness of the tumor. This indicates that irregularly shaped and rough-textured laryngeal cancer may bring a higher risk of T4 staging.

4. DISCUSSION

In the present study, we investigated the performance of different peritumors through deep learning and radiomic methods. Our experimental results demonstrated that the peritumors have additional predictive value relative to intratumors in three tumor datasets. Furthermore, the definition of peritumor and artificial intelligence methods (deep learning or radiomics) also affect the performance of peritumor.

Previous studies explored how different definitions of the peritumoral region might influence the performance of different models. For instance, Beig et al. combined intranodular and different perinodular region radiomic features to distinguish between lung nonsmall cell lung cancer adenocarcinomas and benign granulomas. They included a 30 mm perinodular region, divided into 5 mm rings, and found that the best features were those extracted from a perinodular of 5 mm beyond the tumor.[16] Braman et al. explored the performance of different sizes of peritumors ranging from 3 to 15 mm for predicting the response to treatment in breast cancer cases, and found that the radiomic features from 3 mm peritumor were significantly associated with the density of tumor-infiltrating lymphocytes.[17] The results of our experiments are consistent with them. The peritumors have additional predictive value in three tumor datasets, especially for GIST datasets, the AUC of the intra-peritumor deep learning model is significantly better than intratumor alone. And the size of peritumor is also important for effectively characterizing the intra-peritumor with tumor task. As for LC datasets,

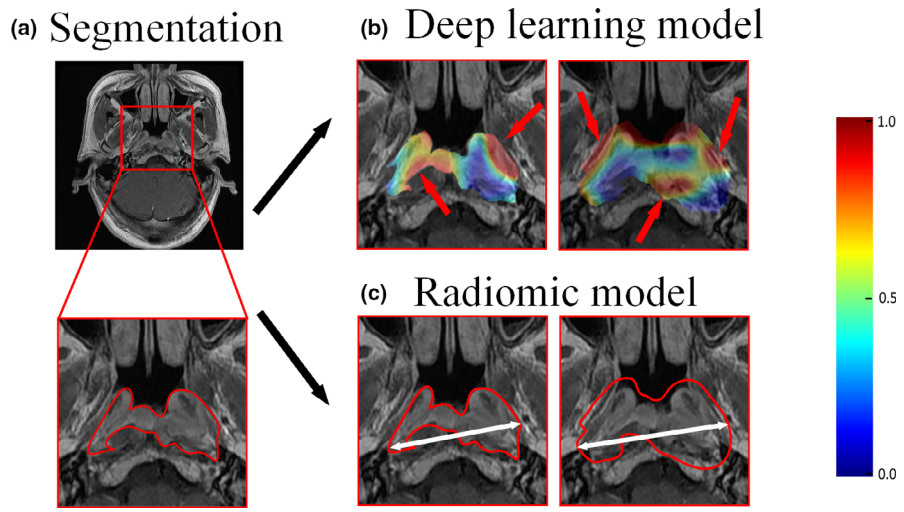

FIG. 7. The class active map of the deep learning model and features color map of the radiomic model in NPC datasets. (a), represents the segmentation of the tumor. In (b) and (c), the left column represents the intratumoral models, and the right column represents the intra-peritumoral models.

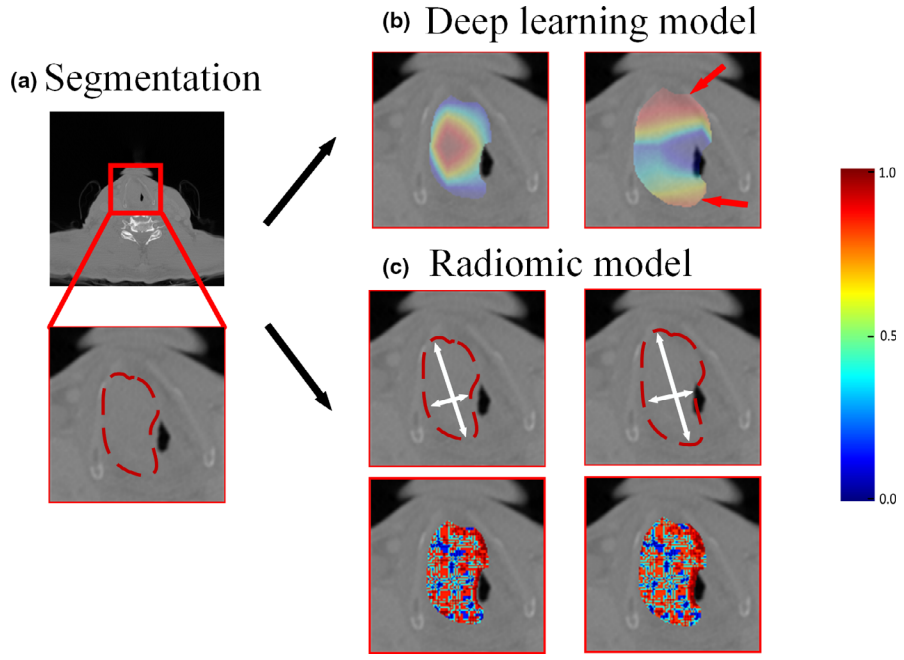

FIG. 8. The class active map of the deep learning model and features color map of the radiomic model in LC datasets. (a), represents the segmentation of the tumor. In (b) and (c), the left column represents the intratumoral models, and the right column represents the intra-peritumoral models.

when the peritumor is too large, the model performance would collapse, only the peritumor ranging from 1.5 to 4.5 mm could get stable performance through radiomic method. This indicated that, although the peritumor indeed has a positive influence in the three tumor tasks, the peritumor with a suitable size would make the model more stable and balanced.

At present, radiomics and deep learning are two commonly used methods for the quantitative analysis of tumors. For the radiomic, from the concept to a wide range of applications is used for the quantitative analysis of the tumor.[8,9,11,12] Specifically, high-dimensional artificially defined features, such as shape, histogram, and texture features, are extracted from regions of interest to represent tumor information, many of which are difficult to be observed by the naked eye. As for deep learning, after the success in natural images, it was widely applied to various fields, including medical imaging.[18,22–24] Deep neural network is an end-to-end structure. For tumor diagnosis tasks, images or regions of interest are directly taken as the input of the network, and the correlation between images and target tasks is automatically learned by optimizing the loss function. In our results, radiomics and deep learning have their strengths and weakness in the three datasets. For the GIST dataset, there are a total of 561 cases of data from two hospitals, and deep learning is more stable according to Fig. 3. As

can be seen from the Table S7, with the change of the peritumor range, the significant features after features selection used in radiomic model changed greatly, resulting in the instability of the radiomic model. There may be some deviations in the larger datasets from different centers. Compared with radiomics, deep learning seems more resistant to such deviations. However, for LC and NPC datasets, radiomic method has better performance. These two datasets are relatively small, with 211 and 233 cases, respectively. The features used in the radiomic models with different peritumors are highly consistent, which indicates that peritumor radiomic features are very stable in small datasets. As far as deep learning is concerned, it is known that it has a huge demand for data volume, which usually leads to a large performance loss in a small dataset. Therefore, for small data, it may be more appropriate to use radiomics to analyze the peritumor.

The performance of Deep learning + Radiomics models on intra-peritumor has not improved compared to deep learning alone or radiomics alone. We believe that this may be caused by insufficient performance of one of the two models. Such as, for the GIST dataset, the performance of radiomic models in different intra-peritumors is instability; For NPC and LC datasets, the deep learning model has not achieved good performance due to the limited training data. When the performance of the two models is stable, and the results are complementary, model fusion may bring further performance gains. But insufficient performance of one of the models cannot improve the performance of the fusion model.

The present study also has some limitations which might be addressed by future work. First, although three tasks were analyzed, the datasets were small. In future work, more data (including other countries) should be collected to further verify the performance of peritumor. Second, the maximum slice of the tumor was directly used for analysis, not including the multislices of the tumor. In future work, we intend to study the strengths and weaknesses of maximum slice and multislices of the tumor for analyzing the peritumor.

In conclusion, the present study demonstrated that peritumoral regions have additional predictive value in three tumor datasets. The definition of the peritumor and artificial intelligence methods also have great influence on the performance of peritumor.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2017YFC1309100, 2017YFC1308700, 2017YFA0205200, 2017YFA0700401), National Natural Science Foundation of China (82022036, 91959130, 81971776, 81771924, 6202790004, 81930053, 81527805), The Beijing Natural Science Foundation (L182061), Strategic Priority Research Program of Chinese Academy of Sciences (XDB 38040200), the Instrument Developing Project of the Chinese Academy of Sciences (YZ201502), Chinese Academy of Sciences under Grant No. GJJSTD20170004 and QYZDJ-SSW-JSC005, and the Youth Innovation Promotion Association CAS (2017175).

CONFLICT OF INTEREST

The authors have no conflict to disclose.

DATA AVAILABILITY STATEMENT

The datasets will not be made public because of privacy concerns.

Xiangjun Wu, Di Dong, and Lu Zhang contributed equally as co-first authors.
[a)]Authors to whom correspondence should be addressed. Electronic mails: yezhaoxiang@163.com, zhangminming@zju.edu.cn, shui7515@126.com, jie.tian@ia.ac.cn; Telephones: 86-18622221316, 86-571-87315255, 86-20-83870125, 86-010-82618465; Fax: 86-22-23537796, 86-571-87315255, 86-20-83870125, 86-10-62527995.

REFERENCES

1. Wu T, Dai Y. Tumor microenvironment and therapeutic response. *Cancer Lett*. 2017;387:61–68.
2. Hirata E, Sahai E. Tumor microenvironment and differential responses to therapy. *Cold Spring Harb Perspect Med*. 2017;7:a026781.
3. Semenza GL, Ruvolo PP. Introduction to tumor microenvironment regulation of cancer cell survival, metastasis, inflammation, and immune surveillance. *Biochimica et Biophysica Acta (BBA) - Mol Cell Res*. 2016;1863:379–381.
4. Zhou Z, Lu Z-R. Molecular imaging of the tumor microenvironment. *Adv Drug Deliv Rev*. 2017;113:24–48.
5. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013;19:1423.
6. Evans M, Baddour Jr HM, Magliocca KR, et al. Prognostic implications of peritumoral vasculature in head and neck cancer. *Cancer Med*. 2019;8:147–154.
7. Carraro A, Trevellin E, Fassan M, et al. Esophageal adenocarcinoma microenvironment: peritumoral adipose tissue effects associated with chemoresistance. *Cancer Sci*. 2017;108:2393–2404.
8. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
9. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–446.
10. Dong DI, Zhang F, Zhong L-Z, et al. Development and validation of a novel MR imaging predictor of response to induction chemotherapy in locoregionally advanced nasopharyngeal cancer: a randomized controlled trial substudy (NCT01245959). *BMC Med*. 2019;17:190.
11. Dong D, Fang M-J, Tang L, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multi-center study. *Ann Oncol*. 2020;31:912–920.
12. Dong D, Tang L, Li Z-Y, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol*. 2019;30:431–438.
13. Peng H, Dong DI, Fang M-J, et al. Prognostic value of deep learning PET/CT-based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. *Clin Cancer Res*. 2019;25:4271–4279.
14. D'Antonoli TA, Farchione A, Lenkowicz J, et al. CT radiomics signature of tumor and peritumoral lung parenchyma to predict nonsmall cell lung cancer postsurgical recurrence risk. *Acad Radiol*. 2019;27:497–507.
15. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-

parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *Eur Radiol*. 2017;27: 4188–4197.
16. Beig N, Khorrami M, Alilou M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. *Radiology*. 2018;290:783–792.
17. Braman N, Prasanna P, Whitney J, et al. Association of peritumoral radiomics with tumor biology and pathologic response to preoperative targeted therapy for HER2 (ERBB2)–positive breast cancer. *JAMA Netw Open*. 2019;2:e192561.
18. Ning Z, Luo J, Li Y, et al. Pattern classification for gastrointestinal stromal tumors by integration of radiomics and deep convolutional features. *IEEE J Biomed Heal informatics*. 2018;23:1181–1191.
19. Fletcher CDM, Berman JJ, Corless C, et al. Diagnosis of gastrointestinal stromal tumors: a consensus approach. *Int J Surg Pathol*. 2002;10: 81–89.
20. Wang C, Li H, Jiaerken Y, et al. Building CT radiomics-based models for preoperatively predicting malignant potential and mitotic count of gastrointestinal stromal tumors. *Transl Oncol*. 2019;12:1229–1236.
21. Zhang L, Dong D, Li H, et al. Development and validation of a magnetic resonance imaging-based model for the prediction of distant metastasis before initial treatment of nasopharyngeal carcinoma: a retrospective cohort study. *EBioMedicine*. 2019;40:327–335.
22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016:770–778.
23. Reddy ASB, Juliet DS. Transfer Learning with ResNet-50 for Malaria Cell-Image Classification. In: *2019 International Conference on Communication and Signal Processing (ICCSP)*. IEEE; 2019:945–949.
24. Hu Q, Whitney HM, Giger ML. Using ResNet feature extraction in computer-aided diagnosis of breast cancer on 927 lesions imaged with multiparametric MRI. In: *Medical Imaging 2020: Computer-Aided Diagnosis*. Vol 11314. International Society for Optics and Photonics; 2020:1131411.
25. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016:2921–2929.
26. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Grad-cam BD. Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*; 2017:618–626.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. The performance of the intra-peritumoral deep learning model in GIST datasets.
Table S2. The performance of the intra-peritumoral radiomic model in GIST datasets.
Table S3. The performance of the intra-peritumoral deep learning model in NPC datasets.
Table S4. The performance of the intra-peritumoral radiomic model in NPC datasets.
Table S5. The performance of the intra-peritumoral deep learning model in LC datasets.
Table S6. The performance of the intra-peritumoral radiomic model in LC datasets.
Table S7. The features used in the radiomic model in GIST datasets.
Table S8. The features used in the radiomic model in NPC datasets
Table S9. The features used in the radiomic model in LC datasets
Table S10. The performance of different depth Resnet in GIST testing set.
Table S11. The structure of Resnet18
Table S12. The performance of different networks trained with/without the pretrained weights in GIST testing set.
Method S1. Inclusion criteria.
Method S2. Texture feature visualization.