5-1-2023

Deep learning based on co-registered ultrasound and photoacoustic imaging improves the assessment of rectal cancer treatment response

Yixiao Lin
Washington University in St. Louis

Sitai Kou
Washington University in St. Louis

Haolin Nie
Washington University in St. Louis

Hongbo Luo
Washington University in St. Louis

Ahmed Eltahir
Washington University School of Medicine in St. Louis

See next page for additional authors

Recommended Citation

Lin, Yixiao; Kou, Sitai; Nie, Haolin; Luo, Hongbo; Eltahir, Ahmed; Chapman, Will Jr.; Hunt, Steven; Mutch, Matthew; and Zhu, Quing, "Deep learning based on co-registered ultrasound and photoacoustic imaging improves the assessment of rectal cancer treatment response." Biomedical Optical Express. 14, 5. 2015 - 2027. (2023).
https://digitalcommons.wustl.edu/oa_4/2047

Authors
Yixiao Lin, Sitai Kou, Haolin Nie, Hongbo Luo, Ahmed Eltahir, Will Chapman Jr., Steven Hunt, Matthew Mutch, and Quing Zhu

Deep learning based on co-registered ultrasound and photoacoustic imaging improves the assessment of rectal cancer treatment response

YIXIAO LIN,[1,†] SITAI KOU,[1,†] HAOLIN NIE,[1,†] HONGBO LUO,[2] AHMED ELTAHIR,[3] WILL CHAPMAN JR.,[3] STEVEN HUNT,[3] MATTHEW MUTCH,[3] AND QUING ZHU[1,4,*] (iD)

[1]*Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63105, USA*
[2]*Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63105, USA*
[3]*Department of Surgery, Washington University School of Medicine, St. Louis, MO 63110, USA*
[4]*Department of Radiology, Washington University School of Medicine, St. Louis, MO 63110, USA*
[†]Contributed equally.
[*]*zhu.q@wustl.edu*

Abstract: Identifying complete response (CR) after rectal cancer preoperative treatment is critical to deciding subsequent management. Imaging techniques, including endorectal ultrasound and MRI, have been investigated but have low negative predictive values. By imaging post-treatment vascular normalization using photoacoustic microscopy, we hypothesize that co-registered ultrasound and photoacoustic imaging will better identify complete responders. In this study, we used *in vivo* data from 21 patients to develop a robust deep learning model (US-PAM DenseNet) based on co-registered dual-modality ultrasound (US) and photoacoustic microscopy (PAM) images and individualized normal reference images. We tested the model's accuracy in differentiating malignant from non-cancer tissue. Compared to models based on US alone (classification accuracy $82.9 \pm 1.3\%$, AUC 0.917(95%CI: 0.897-0.937)), the addition of PAM and normal reference images improved the model performance significantly (accuracy $92.4 \pm 0.6\%$, AUC 0.968(95%CI: 0.960-0.976)) without increasing model complexity. Additionally, while US models could not reliably differentiate images of cancer from those of normalized tissue with complete treatment response, US-PAM DenseNet made accurate predictions from these images. For use in the clinical settings, US-PAM DenseNet was extended to classify entire US-PAM B-scans through sequential ROI classification. Finally, to help focus surgical evaluation in real time, we computed attention heat maps from the model predictions to highlight suspicious cancer regions. We conclude that US-PAM DenseNet could improve the clinical care of rectal cancer patients by identifying complete responders with higher accuracy than current imaging techniques.

1. Introduction

In 2020, there were over 45,000 cases of rectal cancer in the US. Worldwide, the rectal cancer incidence was over 732,000, with a mortality over 339,000 [1]. The current standard-of-care treatment for Stage II and III tumors (locally advanced rectal cancer, or LARC) involves radiation, chemotherapy, and rectal resection. However, rectal surgery is a procedure with high morbidity, and it may lead to lifelong bowel dysfunction, incontinence, sexual dysfunction, and ostomy [2,3]. Studies of non-operative strategies – collectively known as "watch and wait" – have shown that patients who achieve complete tumor destruction with no residual cancer cells, called complete responders (CR), can safely avoid surgery while achieving satisfactory long-term oncologic and

Corrected 22 May 2023

functional outcomes [4,5,6,7]. However, the main barrier to widespread adoption of "watch and wait" is that existing imaging techniques, including MRI and endorectal ultrasound, have very low sensitivity in identifying patients who have been completely cured by preoperative chemo- and radiotherapy. The problem is mainly that current imaging techniques cannot accurately differentiate post-treatment scar from viable cancer in the patient rectum [8,9,10]. If this problem could be resolved by improving existing imaging techniques, the "watch and wait" management would be adopted more widely, and hundreds of thousands of LARC patients could benefit from this non-operative alternative. One promising imaging modality, photoacoustic imaging, uses hemoglobin as an endogenous contrast agent to map tissue vascular networks. Photoacoustic imaging has been shown to be effective in monitoring cancer treatment response for various cancer types, including breast cancer, brain cancer, and head-and-neck cancer, in both animal [11,12,13,14] and clinical studies [15]. In our previous work, we developed a co-registered US and PAM (US-PAM) probe and conducted *in vivo* patient imaging. In that pilot study, we showed that, compared with US imaging alone, PAM helped better differentiate residual cancer from tissue with CR [16,17].

Several studies have employed deep learning to classify colorectal cancer treatment response using data from routine imaging modalities, including endoscopy, CT, MRI, and histology. They demonstrated that compared to the standard clinical procedure for determining CR, their deep learning models could successfully provide a complementary "clinical decision support system" for selecting patients who do not need surgeries [18,19,20,21,22]. In this study, using *in vivo* images from 21 patients, we developed a robust deep-learning model based on co-registered dual-modality US and PAM images and individualized normal reference images to differentiate residual cancer from normal and normalized tissue, or CR. Additionally, based on the model predictions, we generated attention heat maps with hot spots that indicate suspicious cancer regions to facilitate surgeons' decision-making. The objective of this study is to more accurately and robustly identify CRs with co-registered US and PAM imaging. To the best of our knowledge, this study is the first to use co-registered US and PAM images as well as normal reference images paired with machine learning to assess rectal cancer treatment response.

2. Methods

2.1. Co-registered ultrasound and photoacoustic (US-PAM) endorectal system

Imaging was performed with an in-house co-registered US-PAM system, described previously [17]. The endorectal probe has a side-viewing design, where a stepper motor rotates the probe head at 1 revolution per second, producing 360° co-registered US and PAM B scans at 1 frame per second. US is transmitted and received by a single-element ring transducer with a 20 MHz center frequency, a 75% bandwidth, and a 12.7 mm focal length (Capistrano Laboratories, Inc.). PA signals are generated from a 1064 nm Nd:YAG laser with a pulse repetition rate of 1 kHz, providing an angular resolution of 0.36°. The light is delivered through the 2-mm hole at the center of the ring transducer. The system has a US axial resolution of 0.116 mm and a PA axial resolution of 0.231 mm. Co-registered B scans were continuously captured throughout each imaging session. Because normal rectal wall is less than 1 cm thick [13], our system excites and detects PA signals from the entire thickness of the rectal wall. Each displayed B scan showed an imaging depth of 24.6 mm. Due to this short imaging depth, no time-gain compensation was applied to either US or PAM A-lines.

2.2. Patient imaging

This study was approved by the Institutional Review Board of the Washington University School of Medicine. All patients provided informed consent. Between May 2019 and April 2022, a total of 21 patients were imaged with the endorectal probe prior to surgery. The imaging protocol was

as follows. After the administration of anesthesia, a study surgeon used a rigid proctoscope and followed standard procedure to identify the location of the treated tumor bed. The surgeon then inserted the US-PAM endoscope and acquired co-registered US and PAM B scans at 1 frame per second across the tumor bed. The scanning process was then repeated. The entire procedure took about ten minutes.

2.3. DenseNet model training

2.3.1. Dataset preparation and ground truth labeling

US images of the normal rectum show alternating hyperechoic and hypoechoic layers, corresponding anatomically to the mucosa, submucosa, muscle, and fat/ligament layers of the rectal tissue. To highlight this layered structure, a custom surface detection algorithm automatically identified the rectal tissue surface in polar coordinates on each US B scan, as illustrated in Fig. 2. The algorithm first searches for the position of the first positive gradient for each A line on a B scan. Using these positions as the initial trial for the rectal tissue surface, the algorithm then optimizes a balance between the depth-directional gradient magnitude, which is a measure of edge strength, and the overall surface smoothness. Using the found surface, the tissue was then digitally flattened on both the US scan and its corresponding co-registered PAM scan by translating each A line to the average surface depth across the entire B scan. This flattening procedure corrected for off-center probe positions inside the rectum during scanning, and made the rectal tissue layers appear approximately horizontal on the B scans when displayed in the polar coordinates. As a result, the layered structure of the normal rectum was characterized by alternating horizontal bands in all images or sections of images, regardless of the endoscopic position inside the lumen when the scans were captured.

To prepare the model inputs, a set of co-registered US and PAM B scans were first divided into subsections, each covering 60° of the circumference and measuring 11 mm in radius. Because the normal colon wall is less than 1 cm thick, 11 mm was chosen to ensure the entire layered structure of the patient's rectal tissue was included in the subsection [23]. In polar coordinates, these subsections corresponded to rectangular regions of interest (ROIs) measuring 167 (W) $\times$ 302 (H) pixels on the raw images. To automate ROI generation for model training, a horizontal sliding window uniformly sampled ROIs, with a 50% overlap between adjacent ROIs. The selected ROIs were subsequently downsampled to 64×128 pixels to reduce the effect of interpatient variation in microscopic rectal features, while maintaining the mesoscopic layered structure. Additionally, downsampling the ROIs effectively reduced the number of model parameters and helped avoid overfitting the training data.

On each US scan, a study surgeon identified the tumor bed region. The ground truth label of an ROI extracted from this scan was determined by its overlap with the tumor bed region, quantified by $\frac{ROI \cap tumor\ bed}{ROI}$. We observed that if <5 mm of the tumor bed was included at the edge of an ROI, it was challenging to distinguish it from the normal morphological variation of rectal tissue. At the focal depth of the probe, a lateral length of 5 mm corresponded to 25% of the ROI width, so an empirical threshold of 0.25 was selected, such that if the overlap exceeded 0.25, the ROI would receive a preliminary label of suspected cancer. Based on the final diagnosis in the surgical pathology report, if there were no cancer cells remaining in the patient's tumor bed (CR), the tissue was considered fully normalized and the ROI was relabeled as normal. Otherwise, if the pathologic diagnosis was cancer, the ROI was confirmed as cancer.

Ground truth artifact ROIs were labeled through a similar process. Two frequent sources of image artifacts are poor signal coupling between tissue and the water balloon surface, and air bubbles trapped inside the water balloon. Both types of artifacts were clearly identifiable on US scans. A researcher experienced in reading ultrasound determined if any artifacts were on US scans. If the overlap between an ROI with artifact regions on the scan exceeded 0.15, the ROI

was labeled as an artifact. The threshold for an artifact was smaller than for a cancer because some air bubbles could be smaller than a typical tumor bed.

2.3.2. Model design and dataset preparation

Our classification model adopted a densely connected neural network (DenseNet) architecture. This architecture was chosen because it requires considerably fewer parameters to achieve comparable performance with competing neural network architectures such as VGG and ResNet [24]. To select the appropriate model architecture for our study, we conducted transfer learning using pretrained DenseNet121, VGG11, and ResNet18 models for differentiating cancer from normal tissue, and compared their classification performance, as summarized in Table 1. Two types of inputs were used: US images alone and co-registered US and PAM images arranged in two channels. DenseNet121 yielded the highest accuracy and area under the ROC curve (AUC) for both input types. This finding confirmed that the DenseNet architecture was the most suitable for our dataset, which was from a small patient population and had limited available data. In this model architecture, a significantly more complex data connectivity is established, where data in each layer are connected with data in every other layer. Consequently, complex and hierarchical information can be learned with relatively few parameters. Additionally, in conventional neural networks with a sequential model architecture, there is a vanishing gradient risk, where the errors during model training become vanishingly small in the shallowest layers because the errors are attenuated as they are propagated back from the last layer to previous layers sequentially. However, DenseNet does not have a vanishing gradient risk, because during DenseNet training, errors in deeper layers are directly propagated to all previous layers

Table 1. Comparison of model performance on normal vs cancer classification using transfer learning based on pretrained models. The performance results were averaged from 10 independently trained models using the same architecture with different initializations.

Model Architecture	Model Parameters	US Accuracy (%)	US AUC	US-PAM Accuracy (%)	US-PAM AUC
VGG11	128.8M	76.8	0.880	79.8	0.899
ResNet18	11.2M	77.4	0.877	81.6	0.914
DenseNet121	6.9M	80.2	0.903	83.7	0.918

To further avoid overfitting during model training, we monitored the model performance on the validation dataset while gradually reducing the model complexity of the original DenseNet 121 architecture, which has 6.9 M parameters (Fig. S1). The final model architecture was a simplified 3-layer DenseNet with 200 k parameters (US-PAM DenseNet), as illustrated in Fig. 1. This simplification significantly improved the model accuracy, from 83.7% to 92.0%, and the AUC, from 0.918 to 0.974.

The final model input contained five channels, where each input channel is an image providing different or complementary information to guide the model's learning. The first two channels were the coregistered US and PAM ROIs, respectively, of the region to be classified. One limitation of PA B scans is that they do not provide information about the anatomical depths of PA signals, which could have clinical significance. While normal vasculature lies in the submucosal and, to a lesser extent, mucosal layers, strong PA signals from deeper tissue could suggest tissue abnormality. To provide structural information for a PAM ROI, the vertical image gradient of the corresponding US ROI was computed and overlaid on the PAM ROI. This superimposed image was used as the model's third input channel. The fourth and fifth channels were a US ROI and its co-registered PAM ROI from the same patient's normal rectal tissue, respectively, acquired either proximal or distal to the tumor bed. These two channels established an individualized reference for the patient's normal rectal tissue to reduce the impact of interpatient variability and to make
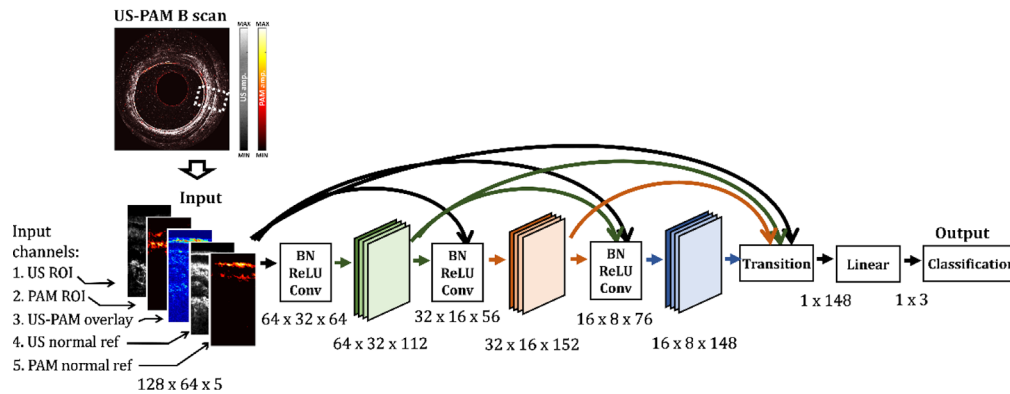
Fig. 1. US-PAM DenseNet architecture. The white dotted box shows an example ROI selected from a co-registered US-PAM B scan. Five channels are generated from the selected ROI as the model input, which has dimensions of $128 \times 64 \times 5$. Solid arrows indicate data flows and connections inside the model: different colors correspond to different data origins. Connections are made between every pair of layers in the DenseNet architecture. The model has three layers, with 64 initial kernels in the first layer, a kernel growth rate of 12 from one layer to the next, and block repetition numbers of 4, 8, and 6 respectively for the three layers. The size of each model layer is marked under the layer icons. BN: batch normalization, ReLU: rectified linear unit. Conv: convolution.

the model more generalizable when applied to patients not encountered during model training. The five channels were then stacked along a third dimension to augment the original US-PAM ROI and fed into the model as a single input.

Image data from 21 patients were used for training and testing the models. The dataset included rectal cancers of all stages with four 4 T0 (CR), four T1, three T2, eight T3, and two T4 cancers. It was partitioned into a training cohort of 15 patients (two T0, three T1, two T2, seven T3, and one T4), a validation cohort of 3 patients (one T0, one T1, one T4b), and a testing cohort of 3 patients (one T0, one T2, and one T3). From the training cohort, a total of 1758 normal ROIs, 1651 cancer ROIs, and 735 artifact ROIs were selected. Artifact ROIs were given twice the sampling weight during model training to ensure the number of ROIs in each classification outcome was balanced. The training dataset was then augmented using random adjustment of image contrast and random horizontal flip. To determine the stopping point during model training, 219 normal ROIs, 176 cancer ROIs, and 88 artifact ROIs from the validation cohort were used. Finally, the testing dataset consisted of 187 normal ROIs, 168 cancer ROIs, and 87 artifact ROIs from the three patients in the testing cohort. The models used the cross entropy loss as their objective loss function, and were trained using Adam optimization with an initial learning rate (LR) of 10^{-4}. To avoid overfitting, the models were trained for only 8 epochs with a multistep LR decrease, namely, LR was reduced to 10% of its previous value after epoch number 2, 5 and 7. The models were constructed in the PyTorch framework and training one model took less than 90 minutes on an NVIDIA GeForce RTX 2080 Ti Graphics Processing Unit.

2.3.3. Two-class models and three-class models

Because image artifacts can be easily identified by an experienced US reader or a surgeon, two-class models were first trained to differentiate residual cancer from normal or normalized tissue. Excluding image artifacts as a possible model outcome by manually removing them from the training dataset helped to isolate the task of making clinical diagnosis for the model. To investigate the added benefit of including co-registered photoacoustic and normal reference images to improve diagnosis, DenseNet models with the same architecture but four different types of

inputs were constructed and trained separately: First, models trained on only individual ultrasound ROIs established the baseline reference for comparing model performances. Additionally, models that accepted co-registered US and PAM ROIs, US ROIs with a normal US reference, and co-registered US and PAM ROIs with normal US and PAM references were trained. To evaluate and compare the performance of models for each input type, the average performance of 10 independently trained models with different initializations was computed.

To achieve automatic real-time diagnosis in the clinical setting, the models need to identify image artifacts on a scan and exclude those regions from further analysis. Therefore, a third category was included during model training to distinguish image artifacts from normal or cancerous tissue. Similar to training two-class models, three-class DenseNet models accepting four different types of inputs were trained and compared.

2.4. US-PAM DenseNet interpretation and clinical application

Three-class US-PAM DenseNet classifies ROIs independently. To make predictions on an entire B scan, 24 ROIs with 75% horizontal overlap were densely sampled from the B scan and classified sequentially. Model predictions at 24 angles along the rectal luminal circumference were obtained. The B scan was predicted to contain cancer if the model classified at least one ROI as cancer, and the cancer location could be further resolved based on the angles corresponding to all the ROIs classified as cancer.

To facilitate the model's clinical application in real time, suspicious cancer regions on the scan were highlighted through guided backpropagation based on the model predictions [25,26]. By back propagating the model outputs of each ROI, the regions in the ROIs that contributed to the model prediction of either normal or cancer were identified and assigned the weights of the model's prediction scores. Color-coded heat maps were reconstructed to differentiate regions that the model predicted as cancer from normal. Image artifacts were not color-coded.
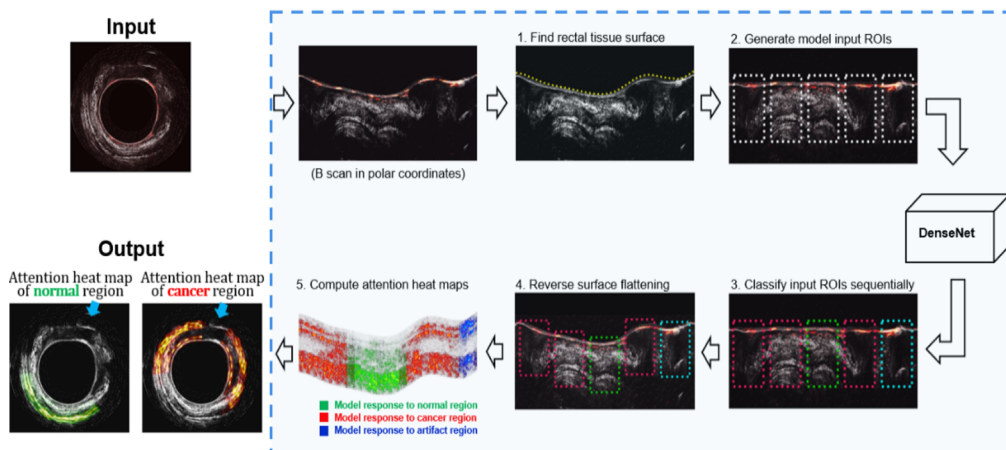

Fig. 2. Pipeline for applying US-PAM DenseNet to diagnose a whole US-PAM B scan and generate an attention heat map of suspicious cancer regions to facilitate surgeons' decision making. The processing pipeline is illustrated in the blue box. In steps 3, 4, and 5, green dotted boxes show the ROIs that US-PAM DenseNet classifies as normal, red shows the cancer ROIs, and cyan shows artifacts. In step 5, guided backpropagation is computed for all three potential classification outcomes, i.e., normal, cancer and artifact, weighted with their respective prediction scores. When displaying the attention heat maps, normal regions and cancer regions are plotted separately, while suspected artifacts, indicated by the blue arrows in the output images, are not color coded.

The pipeline for applying the US-PAM DenseNet to an entire B scan and displaying interpreted model predictions is summarized in Fig. 2 (detailed description of image results for each ROI see Fig. S2).

3. Results

Figure 3 shows representative co-registered US and PAM B scans of normal rectal tissue, tissue with residual rectal cancer, and a tumor bed with no residual cancer (CR). A clear layered structure was observed on ultrasound scans of normal rectal tissue, as shown in Fig. 3(a). There is typically one thick or two closely-spaced hyperechoic layers indicating the mucosa and submucosa. The microvasculature distributed in the submucosa layer generates most PA signals from a normal rectum. The muscular layer beneath the submucosa is hypoechoic on US and does not generate significant PA signals. Cancer typically appears hypoechoic and disrupts the layered structure seen on US, as seen in Fig. 3(b). PA signals from cancer are discontinuous from the surrounding normal tissue. Depending on the patterns of cancer-related blood vessel growth and the friability of the tumor's surface, cancers generate significantly more varied PA signals than normal tissue. Figure 3(c) shows a B scan of the tumor bed from a complete responder. While abnormal morphology was clearly visible on the US scan, PA signals in the tumor bed region appeared consistent with the PA signals on the opposite side of the rectum where tissue was normal, suggesting vascular normalization in the tumor bed region. These observations implied that although US was sufficiently sensitive to cancer, its specificity was not satisfactory because it could not accurately differentiate the tumor beds from complete responders. On the other hand, PA imaging provides functional information of a tumor bed that reflects the extent of the patient's treatment response.
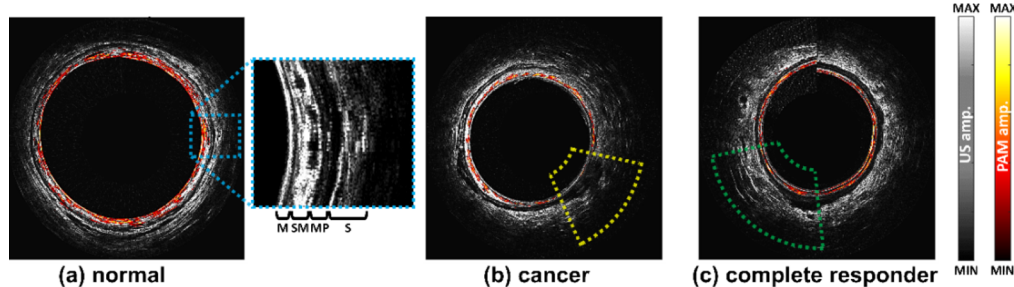

Fig. 3. Representative co-registered US-PAM scans. PAM images are overlaid on co-registered grayscale US images in hot color maps. (a) normal rectal tissue. The region in the blue dashed box is zoomed in to illustrate the layered rectal tissue structure visible on ultrasound scans. M = mucosa, SM = submucosa, MP = muscularis propria, S = serosa. (b) cancer (yellow dashed box) persisted after rectal cancer therapy. (c) tumor bed (green dashed box) with no residual cancer from a complete treatment responder.

3.1. Model classification performance

As shown in Table 2 and Fig. 4, when differentiating cancer from normal tissue, the addition of co-registered PA images significantly improved the classification performance in both accuracy and area under the ROC curve (AUC). This finding suggested that the local structural information from US ROIs was not sufficient for accurate diagnosis, and that the functional information from the co-registered PA ROIs provided a useful diagnostic marker to complement the US ROIs. Compared with models based on only ultrasound ROIs, model performance also improved significantly with the inclusion of an individualized normal reference input channel. The normal

reference effectively reduced the interpatient variability by providing each ROI with a global context for the patient's tissue structure. However, the addition of a normal reference to co-registered US and PAM ROIs did not result in a significant improvement in model performance. Overall, US-PAM DenseNet achieved an accuracy of $92.4 \pm 0.6\%$ and an AUC of 0.968 (95%CI: 0.960-0.976) when differentiating residual cancer from normal ROIs. On the patient level, US-PAM DenseNet accurately classified all three patients in the testing cohort.
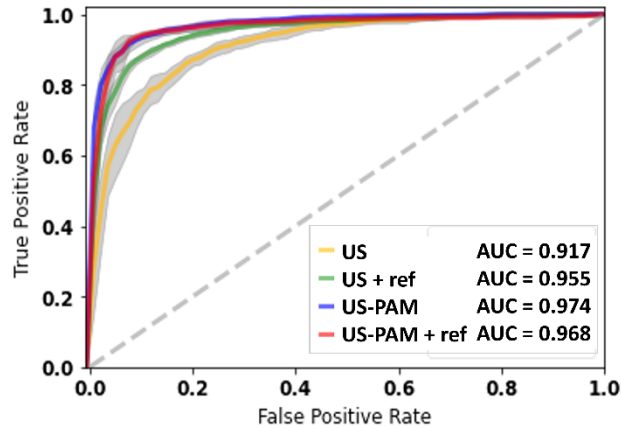

Fig. 4. Average ROC curves of models for differentiating normal rectal tissue and cancer based on different model inputs. Models based solely on ultrasound image inputs are compared to models based on ultrasound images complemented with co-registered PAM and/or a normal reference. The standard deviation of each ROC curve is shown as a gray shaded region.

Table 2. Effect of incorporating co-registered PAM and/or normal reference on model performance in differentiating normal rectal tissue and residual cancer

Model input	Accuracy (%)	AUC
US	82.9 ± 1.3	0.917 (95% CI: 0.897-0.937)
US + normal reference	89.3 ± 0.5	0.955 (95% CI: 0.945-0.965)
US-PAM	92.0 ± 1.6	0.974 (95% CI: 0.966-0.982)
US-PAM + normal reference	92.4 ± 0.6	0.968 (95% CI: 0.960-0.976)

The capability to identify completely normalized tissue with no residual cancer (CR) is of ultimate clinical interest. Accordingly, the model was tested on only tumor bed ROIs from CRs and those with residual cancers. As shown in Table 3 and Fig. 5, the addition of either co-registered PA images or individualized normal references significantly improved the model's performance.

Table 3. Effect of incorporating co-registered PAM and/or a normal reference on model performance in differentiating a normalized tumor bed in a complete treatment responder from residual cancer

Model input	Accuracy (%)	AUC
US	78.6 ± 3.8	0.873 (95% CI: 0.816-0.930)
US + normal reference	81.6 ± 3.2	0.922 (95% CI: 0.853-0.985)
US-PAM	84.4 ± 2.0	0.914 (95% CI: 0.875-0.951)
US-PAM + normal reference	94.7 ± 1.9	0.984 (95% CI: 0.955-1)

When both were incorporated, US-PAM DenseNet achieved an accuracy of $94.7 \pm 1.9\%$ and an AUC of 0.984 (95%CI: 0.955-1), compared to $78.6 \pm 3.8\%$ and 0.873 (95%CI: 0.816-0.930) from models trained on individual US ROIs alone. It should be noted that interpatient variability was not accounted for in this evaluation because there was only one complete responder in the testing cohort, and all normalized tumor bed ROIs were selected from the same patient.
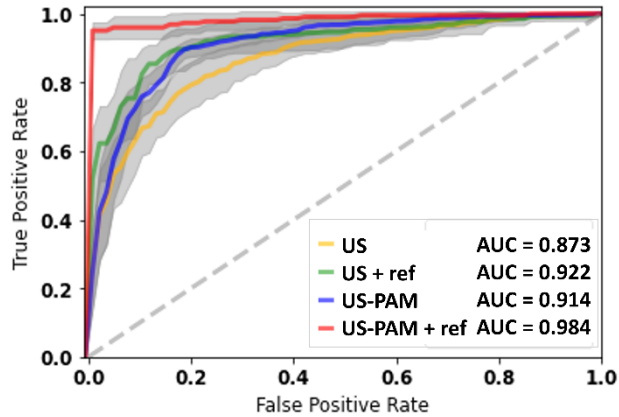

Fig. 5. Average ROC curves of models for differentiating a normalized tumor bed in a complete treatment responder and residual cancer, based on different model inputs. Models based solely on ultrasound image inputs are compared to models based on ultrasound images complemented with co-registered PAM and/or a normal reference. The standard deviation of each ROC curve is shown as a gray shaded region.
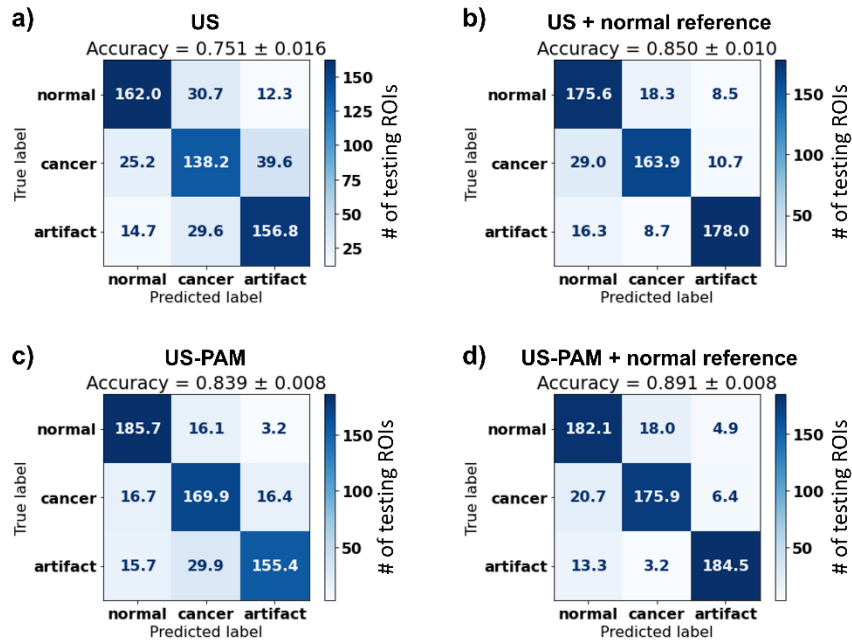

Fig. 6. Comparison of three-class classification model performance with and without the addition of co-registered PA images and/or a normal reference of the patient's rectal tissue.
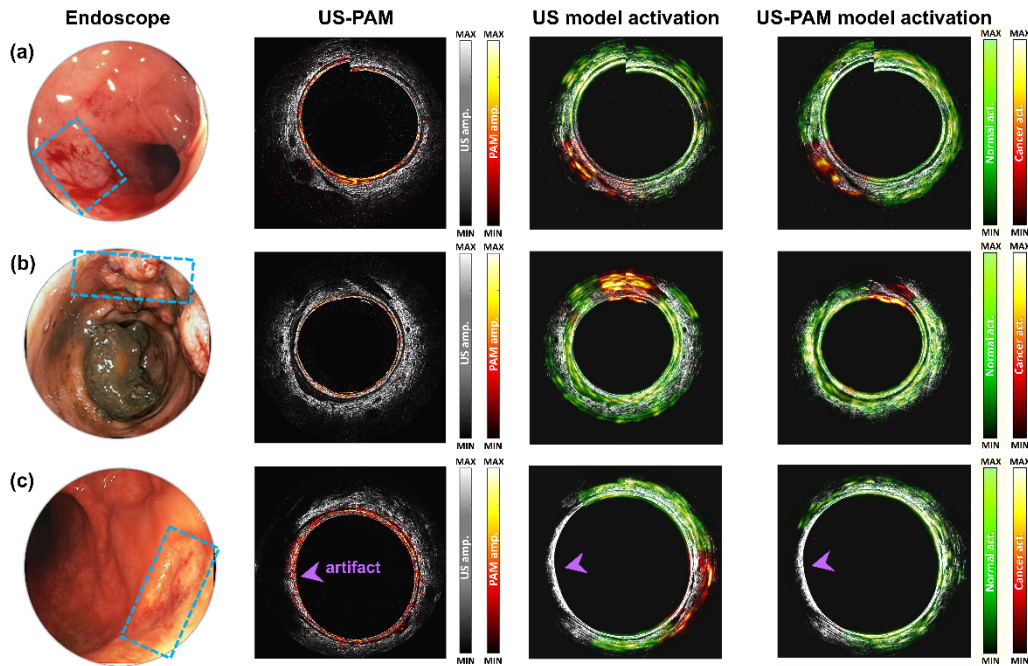
Fig. 7. Representative images from patients in the testing cohort and their US-PAM DenseNet predictions. The blue boxes in the endoscopic photos indicate the cancer sites. The US model activation and the US-PAM activation columns show the attention heat maps computed from the US model and US-PAM DenseNet, respectively. On the attention heat maps, green highlights the normal regions based on the model predictions, and red highlights the cancer regions. (a) Patient with T3 cancer with little or no treatment response. (b) Patient with T2 cancer with extensive scarring. (c) Complete treatment responder (pCR) with no residual cancer. Purple arrowhead indicates an image artifact due to poor ultrasound signal coupling. It is correctly classified in both the US model and US-PAM DenseNet.

The presence of image artifacts, such as air bubbles or fecal debris, can lead to misclassification. Therefore, to facilitate real time clinical application without manual ROI selection, an extension of US-PAM DenseNet included a third classification category to identify image artifacts. Figure 6 illustrates the three-class classification performance of US-PAM DenseNet compared to that of models trained on only US images or images without a normal reference. The confusion matrices suggest that incorporating co-registered PA images better differentiates normal from either cancer or artifact images, whereas adding a normal reference allows the model to more accurately differentiate cancer images from image artifacts. Overall, US-PAM DenseNet achieved an average accuracy of $89.1 \pm 0.8\%$, compared to $75.1 \pm 1.6\%$ from models trained on only US images.

3.2.   US-PAM DenseNet-guided attention heat map

Three-class US-PAM DenseNet was applied to whole B scans by making predictions on overlapping ROIs sequentially. Guided backpropagation was used for more straightforward interpretation of model outputs and visualization. Figure 7 illustrates the model performance on patients in the testing cohort with different degrees of treatment response and cancer stages after treatment. When the post-treatment cancer boundary was clear, the attention heat maps reconstructed from US-PAM DenseNet predictions were consistent with those from the US model, as shown in Fig. 7(a). However, when there was extensive scarring around the treated

tumor, as shown in Fig. 7(b), US-PAM DenseNet predicted a more localized cancer region than the US model, because the PA signals appeared more continuous with and similar to the surrounding normal tissue at the scar boundary. Finally, the models were tested on a CR with no residual cancer. As shown in Fig. 7(c), the US model incorrectly highlighted the tumor bed as a suspicious cancer region. But according to US-PAM DenseNet, while the prediction score for cancer was slightly elevated at the scar region, the model correctly classified the region as cancer-free. This result suggested that US-PAM DenseNet better distinguished scar from residual tumor than models based on US ROIs alone.

4. Discussion

In our previous work, we found that PA imaging could accurately differentiate disrupted cancer vasculature from normal vasculature using a model trained from PA images only. In this expanded study population with a total of 21 *in vivo* patients, we have increased the model's robustness by using co-registered dual-modality images and incorporating an individualized normal reference in a custom DenseNet model. Compared to models trained on only US images, models including either co-registered PA images or normal references as additional inputs showed that both additions significantly improved the model's performance in different aspects. Importantly, we demonstrated that US-PAM DenseNet had higher accuracy in identifying normalized tissue after treatment (CR) compared to models for which only US images were available. Further, to visually interpret the model and potentially facilitate the surgeons' clinical decision-making, we implemented guided backpropagation to highlight hot spots of normal and cancer regions based on model predictions. We observed from the model activation maps that signals from the submucosa layer were important to the model's final classifications. This finding is expected because normal submucosa is hyperechoic to US and rich in blood vessels, generating both strong US and PA signals. Interestingly, the outermost serosa layer, where there are typically very weak US and PAM signals if any, was found to be equally important to the model's final classifications. One reason could be that because LARCs are cancers that invade at least as deep as the muscularis, irregular or strong US or PAM signals in the serosa may suggest presence of cancer. The algorithm was validated on the testing cohort, which included two patients with residual cancer and one complete treatment responder. US-PAM DenseNet classified all three patients correctly. Moreover, the generated attention heat maps accurately corresponded to the tumor bed regions indicated by the study surgeon in the two cancer patients.

   The size of the study population is a main limitation. The appearance of normal rectal tissue varies slightly among patients, and the morphology of rectal cancer is more highly heterogeneous, especially when rectal cancers of all stages are included for model training. Therefore, the model's generalizability could be affected because the training dataset consisted of only ROIs selected from a limited pool of 15 patients. We mitigated this limitation with the following two strategies: first, we downsampled the input images and simplified the model architecture. We noticed that a normal diagnosis is mainly based on a clearly layered tissue structure, which is a mesoscopic feature not affected by downsampling, and that the microscopic differences in rectal tissue on US and PAM images are mainly due to interpatient variability, whose effect is reduced by image downsampling. Second, we included individualized normal references as additional model inputs because we hypothesized that by comparing US-PAM ROIs and their corresponding normal references, the model could also learn the variation in the tissue morphology of an individual patient, regardless of differences in normal rectal tissue among different patients. Our results suggested that normal references made US-PAM DenseNet more generalizable in both two-class and three-class models. Also, due to the limited study population, only three patients were used for testing the model's performance. Data from more patients, especially complete treatment responders, will be collected to further improve the model's robustness and obtain a more reliable evaluation of the model's performance. Additionally, to enhance the clinical utility

of our model, we will implement the US-PAM DenseNet classification and attention heat map generation in real-time as we scan the patients in the next phase of our imaging studies. Based on the evaluation and feedback from the study surgeons, we will further improve the model to better support clinical decision making.

US-PAM DenseNet has demonstrated satisfactory accuracy in differentiating residual cancer from normal or normalized tissue after treatment. However, due to high signal variation and limited imaging depth, it remains quite challenging to extract quantifiable features, which can directly affect diagnosis based on PA images alone. As a result, the structural and depth information from a co-registered US image is a necessary complement for us to draw conclusions from PA images. Additionally, vascular normalization is not only manifested in the volume of blood vessels, which can be inferred from the PA signal density and intensity, but in the vascular morphology as well. To resolve both limitations, an ongoing effort is to implement axial scanning along the length of the probe, in addition to radial scanning. Instead of acquiring B scans at different anatomical positions by hand, we will acquire 3D data and reconstruct the vascular morphology by computing its en face projection.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data Availability. Associated code has been uploaded to Code Ocean [27]. Data is available from the corresponding author upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA: A Cancer J. Clin. **71**(3), 209–249 (2021).
2. P. Luna-Pérez, S. Rodríguez-Ramírez, J. Vega, E. Sandoval, and S. Labastida, "Morbidity and mortality following abdominoperineal resection for low rectal adenocarcinoma," Rev. Invest Clin. **53**(5), 388–395 (2001).
3. A. Nesbakken, K. Nygaard, T. Bull-Njaa, E. Carlsen, and L. M. Eri, "Bladder and sexual dysfunction after mesorectal excision for rectal cancer," British J. Surg. **87**(2), 206–210 (2002).
4. A. Habr-Gama, R. O. Perez, W. Nadalin, R. Sabbaga, U. Jr, S. e. Sousa, A. H. Jr, F. G. Campos, D. R. Kiss, and J. Gama-Rodrigues, "Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: Long-term results," Ann. Surg. **240**(4), 711–718 (2004).
5. F. Dossa, T. R. Chesney, S. A. Acuna, and N. N. Baxter, "A watch-and-wait approach for locally advanced rectal cancer after a clinical complete response following neoadjuvant chemoradiation: a systematic review and meta-analysis," The Lancet Gastroenterol. Hepatol. **2**(7), 501–513 (2017).
6. J. J. Smith, P. Strombom, and O. S. Chow, *et al.*, "Assessment of a watch-and-wait strategy for rectal cancer in patients with a complete response after neoadjuvant therapy," JAMA Oncol. **5**(4), e185896 (2019).
7. F. López-Campos, M. Martín-Martín, R. Fornell-Pérez, J. C. Garcia-Perez, J. Die-Trill, R. Fuentes-Mateos, S. Lopez-Duran, J. Dominguez-Rullan, A. Ferreiro, A. Riquelme-Oliveira, A. Hervas-Moron, and F. Couñago, "Watch and wait approach in rectal cancer: Current controversies and future directions," World J. Gastroenterol. **26**(29), 4218–4239 (2020).
8. J. G. Guillem, D. B. Chessin, J. Shia, H. G. Moore, M. Mazumdar, B. Bernard, P. B. Paty, L. Saltz, B. D. Minsky, M. R. Weiser, L. K. F. Temple, A. M. Cohen, and W. Douglas Wong, "Clinical examination following preoperative chemoradiation for rectal cancer is not a reliable surrogate end point," J. Clin. Oncol. **23**(15), 3475–3479 (2005).
9. R. G. H. Beets-Tan and G. L. Beets, "MRI for assessing and predicting response to neoadjuvant treatment in rectal cancer," Nat. Rev. Gastroenterol. Hepatol. **11**(8), 480–488 (2014).
10. S. Liu, G. X. Zhong, W. Zhou, H. Xue, W. Pan, L. Xu, J. Lu, B. Wu, G. Lin, H. Qiu, and Y. Xiao, "Can endorectal ultrasound, MRI, and mucosa integrity accurately predict the complete response for mid-low rectal cancer after preoperative chemoradiation? A prospective observational study from a single medical center," Diseases of the Colon and Rectum **61**(8), 903–910 (2018).
11. E. Hysi, L. A. Wirtzfeld, J. P. May, E. Undzys, S. Li, and M. C. Kolios, "Photoacoustic signal characterization of cancer treatment response: Correlation with changes in tumor oxygenation," Photoacoustics **5**, 25–35 (2017).
12. S. Mallidi, K. Watanabe, D. T. Timerman, D. Schoenfeld, and T. Hasan, "Prediction of Tumor Recurrence and Therapy Monitoring Using Ultrasound-Guided Photoacoustic Imaging," Theranostics **5**(3), 289–301 (2015).

13. L. J. Rich, A. Miller, A. K. Singh, and M. Seshadri, "Photoacoustic imaging as an early biomarker of radio therapeutic efficacy in head and neck cancer," Theranostics **8**(8), 2064–2078 (2018).

14. I. Quiros-Gonzalez, M. R. Tomaszewski, M. A. Golinska, E. Brown, L. Ansel-Bollepalli, L. Hacker, D. Couturier, R. M. Sainz, and S. E. Bohndiek, "Photoacoustic tomography detects response and resistance to bevacizumab in breast cancer models," Cancer Res. **82**(8), 1658–1668 (2022).

15. L. Lin, X. Tong, P. Hu, M. Invernizzi, L. Lai, and L. V. Wang, "Photoacoustic computed tomography of breast cancer in response to neoadjuvant chemotherapy," Adv. Sci. **8**(7), 2003396 (2021).

16. X. Leng, W. Chapman, B. Rao, S. Nandy, R. Chen, R. Rais, I. Gonzalez, Q. Zhou, D. Chatterjee, M. Mutch, and Q. Zhu, "Feasibility of co-registered ultrasound and acoustic-resolution photoacoustic imaging of human colorectal cancer," Biomed. Opt. Express **9**(11), 5159 (2018).

17. X. Leng, K.M.S. Uddin, W. Chapman Jr., H. Luo, S. Kou, E. Amidi, G. Yang, D. Chatterjee, A. Shetty, S. Hunt, M. Mutch, and Q. Zhu, "Assessing rectal cancer treatment response using coregistered endorectal photoacoustic and us imaging paired with deep learning," Radiology **299**(2), 349–358 (2021).

18. H. E. Haak, X. Gao, M. Maas, S. Waktola, S. Benson, R. G. H. Beets-Tan, G. L. Beets, M. van Leerdam, and J. Melenhorst, "The use of deep learning on endoscopic image to assess the response of rectal cancer after chemoradiation," Surg. Endosc. **36**(5), 3592–3600 (2022).

19. D. Zhang, Y. Duan, J. Guo, Y. Wang, Y. Yang, Z. Li, K. Wang, L. Wu, and M. Yu, "Using multi-scale convolutional neural network based on multi-instance learning to predict the efficacy of neoadjuvant chemoradiotherapy for rectal cancer," IEEE J. Transl. Eng. Health Med. **10**, 1–8 (2022).

20. L. Shi, Y. Zhang, K. E. Nie, X. Sun, T. Niu, N. Yue, T. Kwong, P. Chang, D. Chow, J. H. Chen, and M. Y. Su, "Machine learning for prediction of chemoradiation therapy response in rectal cancer using pre-treatment and mid-radiation multi-parametric MRI," Magn. Reson Imaging **61**, 33–40 (2019).

21. A. Kleppe, O. J. Skrede, S. De Raedt, T. S. Hveem, H. A. Askautrud, J. E. Jacobsen, D. N. Church, A. Nesbakken, N. A. Shepherd, M. Novelli, and R. Kerr, "A clinical decision support system optimising adjuvant chemotherapy for colorectal cancers by integrating deep learning and pathological staging markers: a development and validation study," Lancet Oncol. **23**(9), 1221–1232 (2022).

22. F. Wang, B. F. Tan, S. S. Poh, T. R. Siow, F. Lynette, W. T. Lim, C. Siew, P. Yip, M. Lian, C. Wang, W. Nei, and H. Q. Tan, "Predicting outcomes for locally advanced rectal cancer treated with neoadjuvant chemoradiation with CT-based radiomics," Sci. Rep. **12**(1), 6167 (2022).

23. W. Wiesner, K. J. Mortele, H. Ji, and P. R. Ros, "Normal colonic wall thickness at CT and its relation to colonic distension," J. Comput. Assist. Tomogr. **26**(1), 102–106 (2002).

24. G. Huang, Z. Liu, L. Van der Maaten, and K. Q Weinberger., "Densely connected convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708 (2017).

25. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *Proceedings of the IEEE international conference on computer vision*, pp. 618–626 (2017).

26. A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," *International conference on machine learning*, pp. 3145–3153 (2017).

27. Y. Lin, S. Kou, and H. Nie, *et al.*, "ARPAM-DenseNet-classification," Zenodo (2022), https://github.com/OpticalUltrasoundImaging/ARPAM-DenseNet-classification.