Summer 8-15-2021

Photoacoustic Imaging, Feature Extraction, and Machine Learning Implementation for Ovarian and Colorectal Cancer Diagnosis

Eghbal Amidi
Washington University in St. Louis

Recommended Citation
Amidi, Eghbal, "Photoacoustic Imaging, Feature Extraction, and Machine Learning Implementation for Ovarian and Colorectal Cancer Diagnosis" (2021). *Engineering and Applied Science Theses & Dissertations*. 642.
https://openscholarship.wustl.edu/eng_etds/642

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Biomedical Engineering

Dissertation Examination Committee:
Quing Zhu, Chair
Hong Chen
Abhinav Jha
Joseph O'Sullivan
Umberto Villa

Photoacoustic Imaging, Feature Extraction, and Machine Learning Implementation for Ovarian and Colorectal Cancer Diagnosis

by
Eghbal Amidi

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2021
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Dr. Quing Zhu, my Ph.D. advisor, is the first person who truly deserves to be well acknowledged. She has been an amazing supervisor, an experienced leader, and a super kind-hearted person. She guided me when I was stuck in challenging problems. There is no word to describe how grateful I was to have such a caring and knowledgeable supervisor.

I acknowledge the value of the critical feedback I received from my thesis committee members, Dr. O'Sullivan, Dr. Chen, Dr. Jha, and Dr. Villa, which guided my thesis in the right direction.

I would like to thank all the physicians, coordinators, and pathologists who helped us recruit patients and ex-vivo tissue samples.

I am grateful to all my labmates and friends for their help and support. Especially, Guang, Shihab, Sreyankar, Atahar, Yun, and Hongbo, with whom I have directly worked on different projects.

I also want to thank Lin Chen for his amazing help with software development and code debugging, and Prof. James Ballard for editing my manuscripts.

I would also like to acknowledge our generous NCI funding sources (R01CA151570 and R01CA237664).

I offer special thanks to the Washington University School of Engineering for allowing me to use their dissertation template as a starting point for the development of this document.

Eghbal Amidi

Washington University in St. Louis

August 2021

Dedicated to my caring parents, Hassan Amidi and Setareh Moradpour, and my loving siblings, Jamal and Sorayya Amidi

ABSTRACT OF THE DISSERTATION

Photoacoustic Imaging, Feature Extraction, and Machine Learning Implementation for Ovarian and Colorectal Cancer Diagnosis

by

Eghbal Amidi

Doctor of Philosophy in Biomedical Engineering

Washington University in St. Louis, 2021

Professor Quing Zhu, Chair

Among all cancers related to women’s reproductive systems, ovarian cancer has the highest mortality rate. Pelvic examination, transvaginal ultrasound (TVUS), and blood testing for cancer antigen 125 (CA-125), are the conventional screening tools for ovarian cancer, but they offer very low specificity. Other tools, such as magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET), also have limitations in detecting small lesions.

In the USA, considering men and women separately, colorectal cancer is the third most common cause of death related to cancer; for men and women combined, it is the second leading cause of cancer deaths. It is estimated that in 2021, 52,980 deaths due to this cancer will be recorded. The common screening tools for colorectal cancer diagnosis include colonoscopy, biopsy, endoscopic ultrasound (EUS), optical imaging, pelvic MRI, CT, and PET, which all have specific limitations.

In this dissertation, we first discuss in-vivo ovarian cancer diagnosis using our coregistered photoacoustic tomography and ultrasound (PAT/US) system. The application of this system is also explored in colorectal cancer diagnosis ex-vivo. Finally, we discuss the capability of our photoacoustic microscopy (PAM) system, complemented by machine learning algorithms, in distinguishing cancerous rectums from normal ones.

The dissertation starts with discussing our low-cost phantom construction procedure for pre-clinical experiments and quantitative PAT. This phantom has ultrasound and photoacoustic properties similar to those of human tissue, making it a good candidate for photoacoustic imaging experiments. In-vivo ovarian cancer diagnosis using our PAT/US system is then discussed. We demonstrate extraction of spectral, image, and functional features from our PAT data. These features are then used to distinguish malignant (n=12) from benign ovaries (n=27). An AUC of 0.93 is achieved using our developed SVM classifier. We then explain a sliding multi-pixel method to mitigate the effect of noise on the estimation of functional features from PAT data. This method is tested on 13 malignant and 36 benign ovaries. After that, we demonstrate our two-step optimization method for unmixing the optical absorption (μa) of the tissue from the system response (C) and Grüneisen parameter (Γ) in quantitative PAT (QPAT). Using this method, we calculate the absorption coefficient and functional parameters of five blood tubes, with sO2 values ranging from 24.9% to 97.6%.

We then demonstrate the capability of our PAT/US system in monitoring colorectal cancer treatment as well as classifying 13 malignant and 17 normal colon samples. Using PAT features to distinguish these two types of samples (malignant and normal colons), our classifier can achieve an AUC of 0.93. After that, we demonstrate the capability of our coregistered photoacoustic microscopy and ultrasound (PAM/US) system in distinguishing normal from malignant colorectal tissue. It is shown that a convolutional neural network (CNN) significantly outperforms the generalized regression model (GLM) in distinguishing these two types of lesions.

Chapter 1: Introduction

Ovarian cancer has the highest mortality rate among all cancers related to the reproductive system of women. It is estimated that in 2021 in the USA, 21,410 new cases of ovarian cancer will be diagnosed and 13,770 women will die of this cancer [1]. In the USA, colorectal cancer is the third most common cause of deaths related to cancer in men and women considered separately, and the second leading cause of cancer deaths when males and females are combined [2].

Photoacoustic imaging (PAI) is an imaging modality that advantageously combines the high resolution of ultrasound (US) imaging with the high contrast of optical modalities [3-4]. Photoacoustic (PA) contrast is related to the tumor’s optical absorption, which is dependent on tumor hypoxia and angiogenesis. PAI has been widely applied in diagnosing different cancers, such as breast [5-7], thyroid [8-9], cervical [10], colorectal [11], and prostate cancers [12-13]), and diseases like peripheral vascular diseases (PVDs) [14], joint inflammations [15-17], and skin diseases [18-20]. By collecting PA data at four wavelengths, our lab has been able to use photoacoustic tomography (PAT) to extract functional information from ovarian and colorectal tissue. We have used blood oxygen saturation (sO2) and relative total hemoglobin (rHbT) to develop classifiers that can successfully distinguish malignant from benign lesions.

This thesis begins by introducing ovarian and colorectal cancers. In the first chapter, we clarify the importance of detecting these malignancies in their early stages, review statistics that are concerning, and describe the limitations of current screening tools. We then introduce photoacoustic imaging and discuss how this modality can help achieve early diagnosis of ovarian and colorectal cancer.

In the second chapter of the dissertation, we introduce our low-cost phantom construction procedure. Phantoms made using this procedure can be used in preclinical applications, e.g., ultrasound system calibration and quantitative photoacoustic tomography. Our gelatin-based phantom includes evaporated milk to generate ultrasound attenuation, and it possesses ultrasound and optical properties close to those of soft tissues. Two sets of phantoms are constructed to explore the effects of the gelatin and evaporated milk concentrations on the phantom's ultrasound properties. The first set of phantoms contains different amounts of gelatin mixed with deionized water (no evaporated milk in this set), while in the second set, the evaporated milk concentration is varied (constant gelatin concentration). We measure the ultrasound attenuation of these phantoms in low and high frequency ranges and show that when gelatin concentration is fixed at 5 %, the ultrasound attenuation varies from 0.4 to 0.6 dB/MHz/cm as the evaporated milk concentration increases from 20% to 50%. After finding the concentration of gelatin and evaporated milk that most closely approximates ultrasound attenuation of tissue, we add n-propanol alcohol, glass microspheres, and Germall Plus preservative to our recipe. We then measure the optical properties of the resulting phantom using a diffuse optical tomography (DOT) system.

Next, we report the diagnostic results of in-vivo imaging of patients with ovarian lesions, using our co-registered photoacoustic and ultrasound (PAT/US) system. A total of 39 ovaries from 24 patients are imaged in-vivo. We extract PAT functional features (sO2 and rHbT), PAT image features, and PAT spectral features within a region of interest (ROI) in each ovary. To select the significant features, a t-test is performed, and the independent predictors are determined by evaluating the correlation between each pair of predictors. To classify the ovarian lesions, we employ a generalized linear model (GLM) and a support vector machine (SVM). We use these classifiers first to distinguish benign/normal lesions from ovaries with invasive epithelial tumors and then to separate normal/benign lesions from all types of ovarian tumors. To assess the best diagnostic performance of the classifiers when multiple wavelength data are available, we first include PAT functional features. Next, we exclude the PAT functional features from the features set to evaluate the best diagnostic performance if only a single wavelength is available. Our results show that using functional features improves the classification performance, especially in distinguishing normal/benign ovarian lesions from all types of tumors. In this case, the GLM classifier yields an area under the ROC curve (AUC) of 0.92; the SVM classifier has an AUC of 0.93. When these features are excluded, the AUCs of the GLM and SVM are 0.89 and 0.92, respectively.

Next, using our PAT data, we explain our multi-pixel procedure to reduce the effect of noise on sO2 estimation in ovarian lesions. In PAT, a tunable laser typically illuminates the tissue at multiple wavelengths, and the received photoacoustic waves are used to form functional images of rHbT and sO2. Due to measurement errors, the estimation of these parameters can be challenging, especially in clinical studies. We use a multi-pixel method to smooth the measurements before calculating rHbT and sO2. We first perform phantom studies using blood tubes of calibrated sO2 to evaluate the accuracy of our sO2 estimation. We then present diagnostic results from PAT of 33 patients with 51 ovarian masses imaged by our PAT/US system. The ovarian masses are divided into malignant and benign/normal groups. Using the PAT data from all ovaries in these two groups, we construct functional maps of rHbT and sO2 and plot their histograms. We also calculate the spectral features of the PAT data: their slope, intercept, and midband fit. Support vector machine (SVM) models are trained on different combinations of the significant features. An AUC of 0.93 (0.95%CI: 0.90 – 0.96) on the testing data set is achieved by combining the mean sO2, the spectral intercept, and the score of the study radiologist.

We then discuss one of the challenges of quantitative photoacoustic (PA) imaging: unmixing the optical absorption (μa) of the tissue from the system response (C) and Grüneisen parameter (Γ). In this study, we calculate the absorption coefficient and functional parameters, i.e., the total hemoglobin (HbT) and sO2 of five blood tubes with sO2 values ranging from 24.9% to 97.6% at different depths in intralipid solution. Beer's law is used to calculate the optical fluence in the target area. Initial values for μa and C×Γ are found by fitting a line to the log of the PA beamform envelope data. These initial values are iteratively updated using a conjugate gradient method, and this process is repeated for all 11 wavelengths. The absorption coefficient spectrum follows the molar extinction coefficient spectrum of deoxyhemoglobin for lower sO2 percentages, and it more closely resembles the spectrum of oxyhemoglobin when the sO2 percentage increases. The calculated absorption coefficients at 11 wavelengths are used to estimate the absolute value of the HbT and sO2 of each blood sample at different depths. The mean errors of the estimated HbT values for blood tubes at all depths, with respect to the real values, are less than 13%. Moreover, the largest sO2 estimation error is 7.5%, for a blood sample with an sO2 of 24.9%. Our quantitative PA method performs well for the data collected from blood samples, and we are investigating this method further on our clinical data.

We then study the application of PAT in colorectal cancer diagnosis. This cancer is the second most common malignancy diagnosed globally. Critical gaps exist in diagnostic and surveillance imaging modalities for colorectal neoplasia. While prior studies have demonstrated the capability of photoacoustic imaging techniques to differentiate normal from neoplastic tissue in the gastrointestinal tract, the capability of evaluating tissue deeper than 1 cm with high speed and a large field of view remains limited. To investigate the ability of photoacoustic technology to image deeper tissue, we conduct a pilot study using a real-time co-registered PAT and US system. A total of 20 ex-vivo human colorectal tissue samples are imaged immediately after surgical resection. Co-registered photoacoustic images of malignancies show significantly increased PAT signal amplitude compared to normal regions of the same sample. Between untreated colorectal tumors and normal tissue, there are statistically significant differences in the rHbT, in spectral features (such as the mean spectral slope and the intercept extracted from PAT and US data), and in image features (such as the first and second order statistics along with the standard deviation of the mean radon transform of PAT images). In predicting histologically confirmed invasive carcinoma using either a logistic regression model or a support vector machine, the combination of the rHbT and PAT intercept parameters achieved the highest AUC value, 0.93, for the testing data set.

We then demonstrate the application of our novel photoacoustic microscopy/ultrasound (PAM/US) endoscope in imaging post-treatment rectal cancer for surgical management of residual tumor after radiation and chemotherapy. In this work, we compare the performance of the CNN models to that of GLM models across 24 ex-vivo samples and 10 in-vivo patient examinations. First order statistical features are extracted from histograms of PAM and US images to train, validate, and test the GLM models, while PAM and US images are directly used to train, validate, and test CNN models. The PAM-CNN model performs best, with an AUC of 0.96 (95% CI: 0.95-0.98), compared to the best PAM-GLM model using kurtosis, with an AUC of 0.82 (95% CI: 0.82-0.83). We also show that both CNN and GLMs derived from photoacoustic data outperform those utilizing ultrasound alone. We conclude that pairing deep-learning neural networks with photoacoustic images is the optimal analysis framework for determining the presence of residual cancer in the treated human rectum.

In the last chapter, we summarize the dissertation and outline future work in quantitative photoacoustic tomography (QPAT) and the detection of nonvalid sO2 maps created from noisy data in clinical applications.

1.1 Cancer

Throughout humans' lifetimes, cells in different organs constantly die and are replaced with new cells. Cancer starts when a group of cells start to grow in an uncontrollable manner. This abnormal condition can appear as small or large tumors in an organ. The cancerous cells can also spread to other organs and affect different parts of human's body. American cancer society has estimated that one in each three people are diagnosed with some type of cancer in their lifetime [21]. Therefore, this disease is a huge concern throughout the world, and in spite of all the advances in medicine and related field, its treatment is still very difficult or, in many cases, impossible, especially when it is diagnosed at later stages.

1.2 Ovarian cancer

1.2.1 Statistics

Among women, ovarian cancer is fifth most common cause of death due to cancer, and it is the deadliest of all the gynecological cancers. According to the American Cancer Society, 21410 women will be diagnosed with ovarian cancer in the US in 2021, and 13770 deaths will be reported [1]. If ovarian cancer is diagnosed at early stages, the 5-year survival rate can be higher than 90%. Unfortunately, as ovaries are located deep in the abdominal cavity, the chance of noticing an abnormality in this area is low when the tumor is at its early stages. Therefore, about 80% of women are diagnosed with ovarian cancers at later stages (stage 3 or 4).

1.2.2 Screening tools

A recent update from the U.S. Preventive Services Task Force has concluded that major trials of promising ovarian cancer screening tools have null findings to date among healthy average-risk women, and there are considerable harms associated with screening, which have included major surgical complications in women found to have no cancer [22]. Women with a screening abnormality will generally undergo prophylactic bilateral salpingo-oophorectomy. Prophylactic salpingo-oophorectomy results in morbidity and mortality of premature menopause, including accelerated bone loss and cardiovascular death. Current screening tools for ovarian cancer are pelvic examinations, transvaginal ultrasound (US), and assessment of cancer antigen 125 (CA125) serum marker levels with all provide high false-positive rate. Conventional MRI and diffusion-weighted MRI can be also employed for ovarian cancer diagnosis, but they are expensive and are useful just as follow up modalities in the investigation of sonographically indeterminate adnexal masses. Moreover, computed tomography (CT) is non-specific for small lesions. Furthermore, FDG-PET can measure the residual or recurrent disease and also help select the surgical treatment. However, this modality has a limited value in lesion localization in the early stages of ovarian cancer because of the difficulty in distinguishing between the signal from early-stage cancers and the background uptake signals from the normal tissue. Thus, there is an urgent need to develop better and more sensitive tools to effectively evaluate ovary.

1.3 Colorectal cancer

1.3.1 Statistics

In the USA, colorectal cancer is the third most common cause of deaths related to cancer in men and women, and the second leading cause of cancer deaths in combined men and women group

[2]. It's estimated that in 2021, 52980 deaths due to this cancer will be recorded. This cancer can be in either of these 3 stages: localize, regional, and distal. The localize stage is the first stage of colorectal cancer development, while the distal stage is its last stage. The survival rates of colon cancer at these three stages are 91%, 72%, and 14%, respectively [2]. These numbers are 89%, 72%, and 16% for rectal cancer. Obviously, it is crucial for these cancers to be diagnosed as early as possible.

1.3.2 Screening tools

Common screening tools for colorectal cancer diagnosis include: White light endoscopy which only detects macroscopic morphology and provides no functional assessment of the imaged tissue. EUS is another modality which is highly user-dependent and unable to resolve small islands of the tumor. MRI has limited between-slice resolution and is often unable to differentiate early tumors from benign neoplasia, committing patients to potentially more invasive treatment regimens than needed. Monitoring of tumors after chemotherapy and radiation with MRI is often confounded by fibrotic reaction and edema, which can appear similar to residual tumor. CT has poorer resolution of the bowel wall layers in comparison to MRI, subsequently limiting its ability to describe circumferential resection margin status or serosal invasion in locally advanced cases. Additionally, CT also cannot distinguish induration or peritumoral fibrosis from frank malignant disease with a high degree of specificity, further limiting its application in local tumor staging. Finally, PET is plagued by poor resolution

1.4 Photoacoustic Imaging

Photoacoustic imaging is based on the generation of ultrasound waves using short pulse lasers. When a tissue is shined by short pulsed lasers, it absorbs the optical energy of the laser, leading to its thermal expansion. This process generates ultrasound wave. The ultrasound waves that are generated using the light energy are called photoacoustic (PA) waves. These waves are propagated in the tissue and are detected by ultrasound detectors (transducers). If PA waves are detected by multiple transducer elements, which is the case in photoacoustic tomography (PAT), an image reconstruction technique can be used to form images from the detected PA waves. Unlike pure optical modalities, the detected signals are actually ultrasound waves which are scattered in tissue much less than optical waves. Moreover, unlike ultrasound modalities which contrast between different areas is dependent on difference in structural/mechanical properties of those areas, in photoacoustic modalities, the difference in optical properties is responsible for contrast between different areas. Therefore, photoacoustic imaging modalities offer the advantages of both ultrasound imaging (high resolution) and optical imaging (high contrast) modalities.

Photoacoustic imaging modalities are divided into two main categories: photoacoustic tomography (PAT), and photoacoustic microscopy (PAM). In PAT, usually an array of low frequency ultrasound detectors is employed to collect PA data. As multiple detectors collect data at the same time, data acquisition is fast in PAT. Moreover, as a wide laser beam is used to shine the tissue as well as array transducers can easily be programmed to steer at various angles (by applying different delays to transducer elements), a large field of view can be imaged using PAT. Finally, PAT can image up to 5 to 7 cm (depending on the properties of tissue) at the cost of lower resolution than PAM because it uses a low frequency transducer to detect PA signals.

PAM on the other hand, uses highly focused laser beams or ultrasound detectors to image tissue. Optical-resolution PAM (OR-PAM) and acoustic-resolution PAM (AR-PAM) are two categories of PAM. In OR PAM, the light beam is focused much tighter than acoustic beams while acoustic beams are much more tightly focused than optical beams in AR-PAM. This modality is employed in this dissertation for colorectal cancer diagnosis as it provides better spatial resolution than OR-PAM beyond the diffusion limit. Generally, PAM offers a better image resolution than PAT, but it can only image a very limited depth in the tissue.

PAT uniquely provides functional imaging at high resolution using hemoglobin as an endogenous contrast agent. The sO2 of the blood vessels in the lesion area can be calculated using PAT data. PAT is a non-expensive, high resolution, high contrast imaging modality, which is able to identify malignant or residual tumors, otherwise, undetectable by current clinical imaging.

References

1. https://www.cancer.org/cancer/ovarian-cancer/about/key-statistics.html
2. https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html
3. L. V. Wang, & S. Hu, “Photoacoustic tomography: in vivo imaging from organelles to organs,” Science, 335(6075), 1458-1462, (2012).
4. L. V. Wang, & J. Yao, “A practical guide to photoacoustic tomography in the life sciences,” Nature methods, 13(8), 627, (2016).
5. A. A. Oraevsky et al., “Clinical optoacoustic imaging combined with ultrasound for coregistered functional and anatomical mapping of breast tumors,” Photoacoustics, 12, 30-45, (2018).
6. L. Lin et al., “Single-breath-hold photoacoustic computed tomography of the breast,” Nature communications, 9(1), 1-9, (2018).
7. L. Lin et al. "High-speed three-dimensional photoacoustic computed tomography for preclinical research and clinical translation," *Nature communications* 12.1: 1-10 (2021).

8. W. Roll et al., “Multispectral Optoacoustic Tomography of Benign and Malignant Thyroid Disorders: A Pilot Study,” Journal of Nuclear Medicine, 60(10), 1461-1466, (2019).
9. M. Yang et al., “Photoacoustic/ultrasound dual imaging of human thyroid cancers: an initial clinical study,” Biomedical optics express, 8(7), 3449-3457, (2017).
10. K. Peng et al., “Detection of cervical cancer based on photoacoustic imaging—the in-vitro results,” Biomedical optics express, 6(1), 135-143, (2015).
11. G. Yang et al., “Co-registered photoacoustic and ultrasound imaging of human colorectal cancer,” Journal of biomedical optics, 24(12), 121913, (2019).
12. C. Liu et al., “In vivo transrectal imaging of canine prostate with a sensitive and compact handheld transrectal array photoacoustic probe for early diagnosis of prostate cancer,” Biomedical optics express, 10(4), 1707-1717, (2019).
13. V. S. Dogra et al., “Multispectral photoacoustic imaging of prostate cancer: preliminary ex-vivo results,” Journal of clinical imaging science, 3, (2013).
14. Y. Wang et al., “A portable three-dimensional photoacoustic tomography system for imaging of chronic foot ulcers,” Quantitative imaging in medicine and surgery, 9(5), 799, (2019).
15. J. Jo et al., “A functional study of human inflammatory arthritis using photoacoustic imaging,” Scientific reports, 7(1), 1-9, (2017).
16. P. J. van den Berg et al., “Feasibility of photoacoustic/ultrasound imaging of synovitis in finger joints using a point-of-care system,” Photoacoustics, 8, 8-14, (2017).
17. J. Jo et al., “Photoacoustic tomography for human musculoskeletal imaging and inflammatory arthritis detection,” Photoacoustics, 12, 82-89, (2018).

18. J. Aguirre et al., “Precision assessment of label-free psoriasis biomarkers with ultra-broadband optoacoustic mesoscopy,” Nature Biomedical Engineering, 1(5), 1-8, (2017).
19. J. Kim et al., “Multispectral ex vivo photoacoustic imaging of cutaneous melanoma for better selection of the excision margin,” British Journal of Dermatology, 179(3), 780-782, (2018).
20. A. B. E. Attia et al., “Noninvasive real-time characterization of non-melanoma skin cancers with handheld optoacoustic probes,” Photoacoustics, 7, 20-26 (2017).
21. https://www.cancer.org/
22. J. T. Henderson et al., “Screening for ovarian cancer: updated evidence report and systematic review for the US preventive services task force,” JAMA. 13;319(6):595-606. doi: 10.1001/jama.2017.21421. Review, (2018).

Chapter 2: Low-cost phantoms for photoacoustic imaging

2.1 Introduction

Photoacoustic (PA) imaging has recently been widely used in clinical studies for cancer diagnosis [1-6]. This modality is based on the generation of ultrasound waves using a short pulsed laser. This means that the generated photoacoustic signals depend on both optical and ultrasound properties of the medium. Thus, it is important to make a phantom which resembles both optical and ultrasound properties of soft tissues. Such a phantom not only could be useful in photoacoustic imaging for calibration purposes to cancel the system effect on the received PA signals [7, 8], but also can be helpful in other optical modalities, such as optical coherence tomography [9-11] and spectroscopy [12].

Tissue-mimicking ultrasound phantoms can be either in liquid or gelatin/agar-based form. Liquid phantoms are not the subject of this study and are reviewed elsewhere [13]. Homogeneous gelatin/ agar based phantoms have been widely used, especially in the context of quantitative ultrasound to cancel the ultrasound system effects. One of the earliest work on gelatin-based ultrasound phantom construction was by Madsen et al. [14]. They made ultrasound phantoms with different properties by changing the concentration of graphite powder and 1-propanol alcohol in their phantom mixture. They were able to construct ultrasound phantoms with a speed of sound between 1520 and 1650 m/s and an ultrasound attenuation which ranged from 0.2 to 1.5 dB/cm/MHz. As the graphite powder highly scattered ultrasound waves, this material was replaced with evaporated milk in another study [15]. Agar-based phantoms made using this method mimicked ultrasound properties of soft tissue. However, the optical properties of the phantoms were not measured in that study. A procedure to make gelatin-based phantoms which mimic both ultrasound and optical properties of soft tissue is reported in [16]. In that study, Intralipid is used to enhance both optical scattering and ultrasound attenuation. Moreover, to induce optical absorption, any one of India Ink or Direct Red 81 or Evans Blue are added to the mixture.

Besides the ultrasonic and optical properties of phantoms, their durability is another important factor to consider when making phantoms. For this reason, phantom recipes usually include adding a preservative, such as Formaldehyde, Thimerosal, and Germall plus to the phantom mixture. Furthermore, gelatin/agar based phantoms usually have to be kept in the fridge to protect them from being spoiled.

In this study, we have made phantoms with different concentrations of gelatin and evaporated milk. Then the ultrasound properties of these phantoms have been measured to see which phantom provides the best ultrasound properties. Then n-propanol alcohol, glass spheres, and Germall plus are added to the recipe and the optical properties of the resulted phantom are found using a diffuse optical system (DOT) developed in our lab [17,18] to make sure that the phantom has optical properties close to soft tissues.

2.2 Methods and Materials

2.2.1 Phantom construction procedure

First, to get some idea about the effect of evaporated milk and gelatin concentration on the ultrasound properties of our phantom, we made two sets of phantoms. In the first set, different phantoms were made by adding different amounts of gelatin to deionized water. No evaporated milk was added to these phantoms. In the second set, we kept the concentration of gelatin at 5%

and then added different amounts of evaporated milk. After realizing how the ultrasound properties are affected by different concentrations of evaporated milk and gelatin, we added n-propanol, glass spheres and Germall plus to our phantom recipe.

In this study, a procedure similar to that described in [15] was followed to make a gelatin-based phantom with ultrasound and optical properties close to those of soft tissues. In table 2.1, the materials used to make the phantom and the amount of each material can be found. To increase the ultrasound attenuation, evaporated milk was employed in the phantom instead of the graphite powder which was conventionally a component of ultrasound phantoms. The reason for this is that besides high ultrasound scattering, graphite powder has a very high optical absorption, whereas employing the evaporated milk results in a yellowish phantom with optical absorption and scattering coefficients close to those of soft tissues. To control the speed of sound in the phantom, n-propanol alcohol was added to the mixture. The higher amount of this material leads to a higher speed of sound of the phantom. Moreover, 70 gr (100 beads/cm3) of 3000E glass microsphere beads with a mean diameter of around 35 µm (Potters Industries Inc., Valley Forge, PA) were used to enhance ultrasound scattering in the phantom. Finally, to protect the phantom from bacterial invasion, Germall plus is added to the mixture.

The following is a summary of the step by step procedure of the phantom construction:

- The ultra-purified deionized water is mixed with n-propanol alcohol. The mixture is placed on top of a magnetic stirrer and gelatin (G2500, Sigma-Aldrich Corp., St. Louis, MO) is gradually added to the mixture. The magnetic stirrer heats up the mixture besides mixing its components. This process of simultaneous heating and mixing is continued until a transparent mixture is generated. Then the mixture is allowed to be cooled down.

- The preservative is added to the evaporated milk and the mixture is heated up to the temperature of 60 °C.

- The two mixtures generated in step 1 and 2 are then combined when they both reach a temperature of 55°C. After that, the microspheres, which have been heated up to the same temperature, are added to the mixture.

- The mixture is then poured into a flask for degassing to ensure that the phantom will not contain any air bubbles. After that, a small portion of the mixture is poured into a cup for ultrasound properties measurements, and the rest of it is transferred from the flask to a larger container to measure the optical properties. Both the cup and container are kept in the fridge for 24 hours to allow its solidification.

Table 2.1: The materials employed to make the phantom and the amount of each of them.

Material	**Amount**
Ultra-purified DI water	1 L
Evaporated milk (unfiltered)	1 L
Gelatin	100 gr
n-propanol alcohol	50 mL
microspheres	70 gr
Germall plus	25 gr

2.2.2 Ultrasound properties

The set up for ultrasound attenuation measurements is shown in Figure 2.1. The ultrasound signals generated by the ultrasound (US) transmitter are received by a second transducer, once without phantom between the two transducers and a second time, in the presence of the phantom. The phantom ultrasound attenuation is then calculated as

$$\alpha_s(f) = \frac{20}{z_2}\log_{10}\left[\frac{< V_r(f,z) >}{< V_s(f,z) >}10^{-\frac{z_2}{10}\alpha_w(f)}\right] \qquad \text{Eq. (2.1)}$$

where αs(f) is the phantom length, αw(f) is the ultrasound attenuation of water as a function of frequency, and <Vr(f,z)> and <Vs(f,z)> are the Fourier transforms of the recorded ultrasound signals before and after placing the phantom between the transducer and planar reflector, respectively.

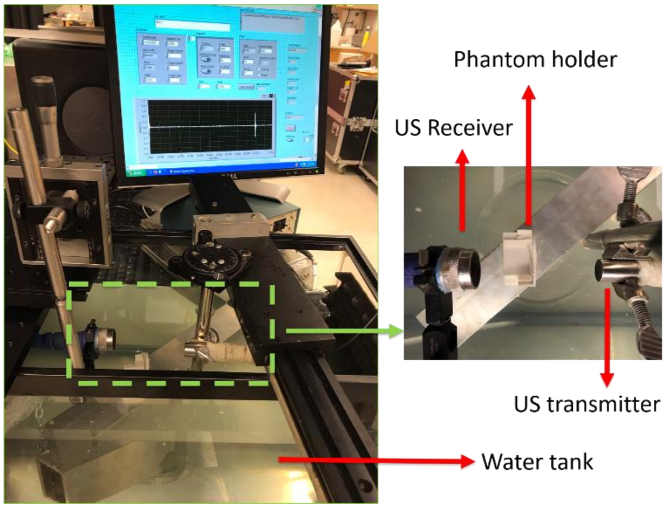


Figure 2.1: The setup used to measure the ultrasound attenuation coefficient

The speed of sound of our phantom was measured by a pulse-echo method. To calculate the speed of sound, the phantom was placed inside the water tank filled with deionized water. A single element transducer (with a 5.4MHz center frequency) generated an ultrasound pulse from the transducer towards the phantom located at the focal distance of the transducer (8cm). This pulse hits both surfaces of the phantom and reflects back towards the transducer. The speed of sound (c) in the phantom is a function of the temporal distance between the two reflections from proximal and distal surfaces of the phantom (Δt) and the phantom length (z) and is given by

$$c = \frac{2z}{\Delta t}. \qquad \text{Eq. (2.2)}$$

2.2.3 Optical properties

The reduced scattering and absorption coefficients of our phantom was measured using a diffuse optical tomography (DOT) system developed in our lab [18]. This system consists of 9 sources and 14 detectors. The optical sources transmit diffused lights sequentially, and every time the light is detected by all detectors. For each source-detector pair, the amplitude and phase of the detected signals are computed using a Hilbert transform. The log of amplitude and phase measurements at different wavelengths can be plotted with respect to the source-detector distance (Figure 2.3) and the optical scattering and absorption can then be calculated from these plots. A detailed description about how to find the absorption and scattering coefficient using the amplitude and phase plots is found in [19].

2.3 Results

2.3.1 Ultrasound and optical properties

Figure 2.2 shows the calculated ultrasound attenuations of the two sets of phantoms using (1) and their linear fit within the -6dB bandwidth of the transducer. The top plots show the ultrasound attenuation increases from 0.07 to 1.26 dB/cm/MHz when the gelatin concentration increases form 4% to 15%. The lower plots illustrate the effect of evaporated milk on the ultrasound attenuation.

These phantoms have a fixed concentration of gelatin (5%). As can be seen, increasing the evaporated milk concentration from 20 to 50 % results in increase of 0.2 dB/cm/MHz in the ultrasound attenuation.

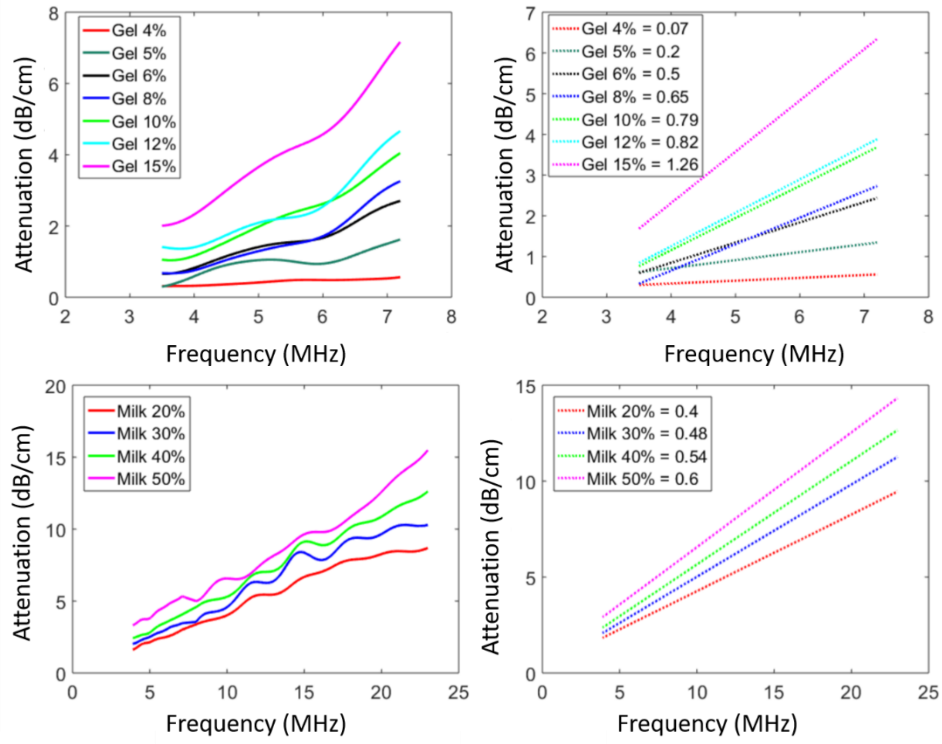


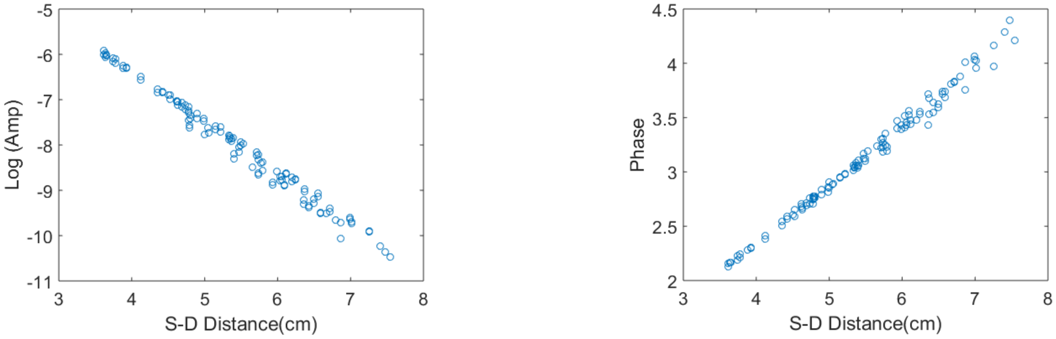
Figure 2.2: Top: The effect of gelatin concentration on ultrasound attenuation (no evaporated milk in the mixture). Bottom: The effect of evaporated milk concentration on ultrasound attenuation when the gelatin concentration is fixed at 5%. The plots on the left show the ultrasound attenuation as a function of frequency, and those on the right are their associated fitted lines.

Using the DOT system, we acquired the log of amplitude (left column) and the phase (right column) measurements of the phantom introduced in Table 1 at 4 different wavelengths of 740, 780, 808, and 830 nm (Figure 2.3). These measurements were then used to calculate the absorption and reduced scattering coefficients of the phantom at these wavelengths. The results have been summarized in Table 2.2. As can be seen, our phantom has optical properties very similar to those of soft tissue.

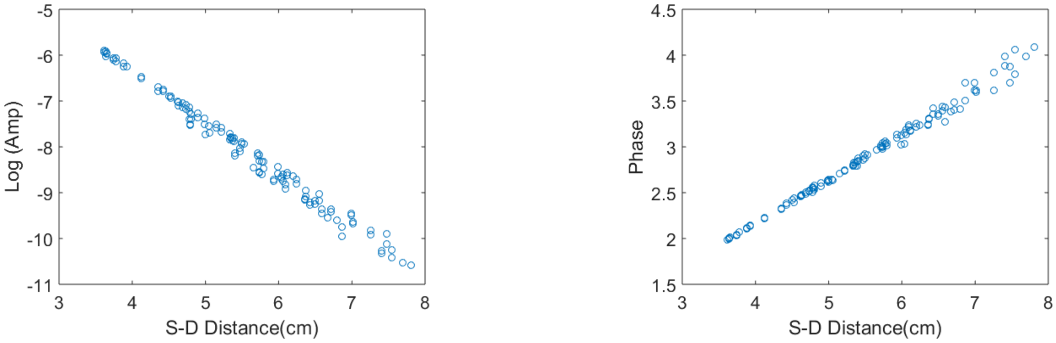
Table 2.2: Absorption and reduced scattering coefficient of our phantom at 740, 780, 808, and 830 nm wavelengths

	740 nm	**780 nm**	**808 nm**	**830 nm**
Absorption coefficient (cm^{-1})	0.029	0.034	0.027	0.032
Reduced scattering coefficient (cm^{-1})	10.41	9.21	9.21	8.74

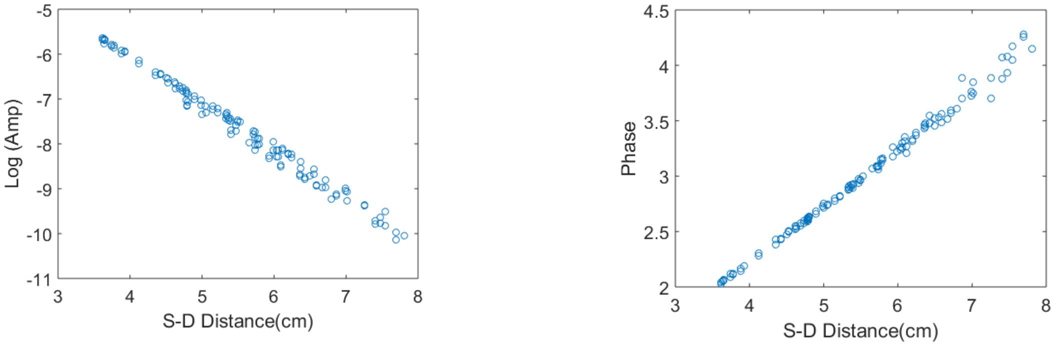
740 nm



780 nm



808 nm



830 nm

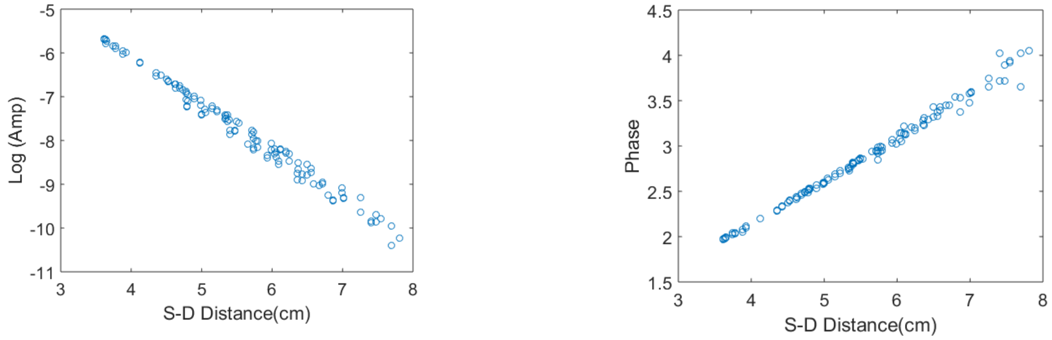


Figure 2.3: The log of amplitude (left column) and the phase (right column) measurements of our phantom at 740, 780, 808, and 830 nm wavelengths as a function of source-detector distance.

2.4 Discussion and Summary

In this study, a phantom was made which has acoustic and optical properties similar to soft tissues. At first, we made phantoms with different concentration of gelatin and evaporated milk, and the effect of each material on the ultrasound attenuation was investigated. The ultrasound attenuation was measured using two transducers, one of which was used to transmit the ultrasound waves and the other one to receive the ultrasound signals. To calculate the optical properties of our phantom, we used a DOT system developed in our lab. The results show that our phantom has ultrasound and optical properties close to soft tissues. However, our study has two main limitations. First, although evaporated milk is low cost, different brands of it with different fat concentrations will affect acoustic and optical properties. Second, using DOT to calculate the optical properties is suitable for measuring lower background optical absorption and scattering properties. Inverse adding doubling sphere method is investigated to measure a range of optical properties of the phantoms.

References

1. L. V. Wang, & S. Hu, “Photoacoustic tomography: in vivo imaging from organelles to organs,” science, 335(6075), 1458-1462 (2012).
2. K. S. Valluru et al., “Photoacoustic imaging in oncology: translational preclinical and early clinical experience,” Radiology, 280(2), 332-349 (2016).
3. G. Yang et al., “Optimized light delivery probe using ball lenses for co-registered photoacoustic and ultrasound endo-cavity subsurface imaging,” Photoacoustics, 13, 66-75 (2019).
4. G. Yang et al., “Optimizing light delivery through ball-shaped multimode fiber tips in co-registered photoacoustic and ultrasound endo-cavity imaging: simulation and experimental validation,” Proc. SPIE 10878, 108785H (2019).
5. G. Yang et al., “Co-registered photoacoustic and ultrasound real-time imaging of colorectal cancer: ex-vivo studies,” Proc. SPIE 10878, 108784K (2019).

6. X. Leng et al., “Feasibility of co-registered ultrasound and acoustic-resolution photoacoustic imaging of human colorectal cancer,” Biomedical optics express, 9(11), 5159-5172 (2018).
7. S. Nandy et al., “Evaluation of ovarian cancer: initial application of coregistered photoacoustic tomography and US,” Radiology, 289(3), 740-747 (2018).
8. A. Mostafa et al., “Dual-mode photoacoustic and ultrasound system for real-time in-vivo ovarian cancer imaging,” Proc. SPIE 10494 (2018).
9. Y. Zeng et al. “3D visualization of the ovarian tissue scattering coefficient with swept-source optical coherence tomography,” Proc. SPIE 10483, 1048338 (2018).
10. Y. Zeng et al., “The Angular Spectrum of the Scattering Coefficient Map Reveals Subsurface Colorectal Cancer,” Scientific Reports, 9(1), 2998 (2019).
11. T. Wang et al., “Label-free biomolecular imaging using scanning spectral interferometry,” Chinese Optics Letters, 11(11), 111102 (2013).
12. R. Wang et al. “Rapid, on-site detection of residual explosives based on a lab-in-a-capillary and UV fiber sensor,” Analytical Methods, 6(24), 9628-9633 (2014).
13. M. O. Culjat et al., “A review of tissue substitutes for ultrasound imaging,” Ultrasound in medicine & biology, 36(6), 861-873 (2010).
14. E. L. Madsen et al., “Tissue mimicking materials for ultrasound phantoms,” Medical physics, 5(5), 391-394 (1978).
15. E. L. Madsen et al., “Liquid or solid ultrasonically tissue-mimicking materials with very low scatter,” Ultrasound in medicine & biology, 24(4), 535-542 (1998).
16. J. R. Cook et al., “Tissue-mimicking phantoms for photoacoustic and ultrasonic imaging,” Biomedical optics express, 2(11), 3193-3206 (2011).

17. K. M. S. Uddin et al., "Two step imaging reconstruction using truncated pseudoinverse as a preliminary estimate in ultrasound guided diffuse optical tomography," Biomedical optics express, 8(12), 5437-5449 (2017).
18. H. Vavadi et al., "Compact ultrasound-guided diffuse optical tomography system for breast cancer imaging," Journal of Biomedical Optics, 24(2), 021203 (2018).
19. N.G. Chen et al., "Simultaneous near-infrared diffusive light and ultrasound imaging," Applied optics, 40(34), 6367-6380 (2001).

Chapter 3: In-vivo ovarian cancer diagnosis using coregistered photoacoustic tomography and ultrasound system

3.1 Introduction

Among women, ovarian cancer is fifth most common cause of death due to cancer, and it is the deadliest of all the gynecological cancers [1]. In 2019, an estimated 22,530 women were diagnosed in the United States, and about 13,980 of these women died from this disease [2]. Due to the lack of early screening and diagnostic techniques, many women are diagnosed with ovarian cancer when it is already at stages III or IV, where the mortality rates are high (70 to 75%) [3,4]. Pelvic examination [5,6], transvaginal ultrasound [7,8], and blood testing for CA-125 [8-9] are the conventional screening tests, but they all lack enough specificity for early ovarian cancer diagnosis [9]. Moreover, imaging modalities such as computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI) have been used for surgical guidance. However, all of these modalities have limitations: for example, they are non-specific for small lesions (CT), costly (MRI), or need specific tracers and have difficulty in separating tumors from background (MRI, PET) [10-12]. Clearly, improved diagnostic methods and more effective detection tools are needed to diagnose ovarian cancer.

Recently, photoacoustic tomography (PAT) has been explored in medical diagnosis because it provides functional information of biological tissue at ultrasound resolution. This modality is based on the photoacoustic effect [13-15]. A pulsed laser light absorbed by tissue causes a local temperature increase which creates thermoelastic expansion and generates photoacoustic waves. The propagated waves are then detected by an ultrasound (US) transducer and used to reconstruct optical absorption distribution. If photoacoustic imaging is performed at two or more optical wavelengths, information about functional parameters of a tissue, such as its rHbT and sO2, can be obtained.

Previously, our group performed a study on classification of excised ovarian samples [16]. In that study, an average area under Receiver Operating Characteristic (ROC) curve (AUC) of 0.92 ± 0. 05 for 50 testing samples was achieved using PAT spectral and beam envelope features along with several PAT image features. In a subsequent study, US spectral features were added to the PAT features to evaluate the classification performance on another set of excised ovaries [17]. It was demonstrated that the generalized linear model (GLM) and support vector machine (SVM) classifiers could respectively achieve sensitivities of 70.4 and 87.7%, and specificities of 95.6 and 97.9%, for the testing data. Moreover, two patients with malignant and benign ovaries were imaged in-vivo in that study, but this number of patients was too low to evaluate the performances of the classifiers. In our most recent study, we reported the imaging results of co-registered US and PAT in a pilot group of 26 ovaries from16 patients. We demonstrated in this study that the difference in rHbT was statistically significant between invasive epithelial ovarian cancers and benign/normal ovarian masses (p=0.01), and the sO2 was statistically significant between invasive epithelial ovarian cancers/other neoplasms and benign/normal ovarian masses (p=0.03) [18]. However, no classification results and ROC analysis were reported in this first pilot group of patients.

In this study, we extracted the PAT functional, spectral, and image features from 24 patients of 39 ovaries (mean age, 54 years; range, 34-76) and performed ROC analysis using the GLM and SVM classifiers. The first 16 patients reported in [18] were included in this study for feature extraction and ROC analysis. We categorized the ovarian lesions into three groups of benign/normal ovaries (n=27), invasive epithelial cancers (n=9), and other types of neoplasms (n=3) (see Table 3.1). First, we developed GLM and SVM classifiers to distinguish benign/normal ovaries from epithelial cancers only. Subsequently, we differentiated benign/normal ovaries from all types of cancers (epithelial and other neoplasms) using new

GLM and SVM classifiers. To evaluate the performances of the classifiers when the data from just one wavelength is available, we repeated the same procedure to design GLM and SVM classifiers without inclusion of PAT functional features in the features set. The performance of each classifier was evaluated by computing its ROC curve for both training and testing data sets and calculating the area under these curves. To the best of our knowledge, the reported patient diagnostic results using the GLM and SVM classifiers with ROC analysis based on PAT features are the first of its kind and may improve current practice on ovarian tissue diagnosis once the results are validated with a large patient pool.

3.2 Methods and Materials

3.2.1 Co-registered PAT/US System

The co-registered PAT/US system was described in detail in [18]. Briefly, this system consists of a Ti-sapphire laser which can be tuned from 690 nm to 900 nm, a light delivery system that includes four optical fibers coupled with a transvaginal transducer (6 MHz, 80% bandwidth), and a commercial ultrasound system (EC-12R, Alpinion Medical Systems, Republic of Korea). The system was programmed to image patients at four wavelengths (730, 780, 800, and 830 nm), and at each wavelength, several (3-10) PAT and US frames were collected.

3.2.2 PAT functional features

The $rHbT\ (r,\theta)$ at each pixel in the region of interest (ROI) was calculated as the summation of the relative oxyhemoglobin concentration $rHbO\,(r,\theta)$ and relative deoxyhemoglobin concentration $rHb\ (r,\theta)$ at that pixel. The relative oxy and deoxy hemoglobin concentrations can be approximated as [19]

$$rHbO(r,\theta) = \bar{C}(r,\theta)HbO(r,\theta), \qquad \text{Eq. (3.1)}$$

$$rHb(r,\theta) = \bar{C}(r,\theta)Hb(r,\theta), \qquad \text{Eq. (3.2)}$$

where $\bar{C}(r,\theta) = \Gamma C_0(r,\theta)\varphi(r,\theta)$, Γ is the tissue's Grüneisen parameter, $C_0(r,\theta)$ is the system acoustic operator, and $\varphi(r,\theta)$ is the local fluence, which can be approximated as wavelength independent at the narrow wavelength window we used. As can be seen in (1) and (2), calculating the absolute values of oxy, deoxy, and total hemoglobin requires knowledge of the local fluence in the tissue, which is difficult to estimate in clinical studies. For this reason, the relative values of these parameters have been computed in this study. To calculate the mean rHbT in the ROI, the maximum of this parameter in this region was found, and the average of the rHbT values higher than half of this maximum value was then computed.

The blood sO2 at each pixel was calculated by dividing the oxyhemoglobin by the sum of the oxy and deoxyhemoglobin:

$$sO2(r,\theta) = \frac{HbO(r,\theta)}{HbO(r,\theta) + Hb(r,\theta)} \times 100\%. \qquad \text{Eq. (3.3)}$$

Note that in *sO2* calculation, the unknown $\bar{C}(r,\theta)$ is cancelled out in the numerator and denominator of the division. We calculated the mean *sO2* in the ROI by taking an average of *sO2* values over pixels with a *sO2* value higher than a noise threshold. Based on our system noise level, this threshold was defined as 5% of the maximum *SO2* in the ROI.

3.2.3 PAT spectral feature extraction

Photoacoustic spectral features have been shown to be valuable tools in clinical applications, such as characterization of bone microstructure [20], quantification of normal and fatty livers [21], cancer diagnosis [22-23], and monitoring cancer treatment response [24]. In earlier studies we evaluated the feasibility of these features for distinguishing malignant from benign ex-vivo ovaries [16-17]. Here, we extract these features from the data collected from in-vivo studies of patients.

To calculate the spectral features, co-registered ultrasound images were used to select an ROI corresponding to the examined ovary. The angular beam segments in the ROI with a maximum value greater than the noise level of our US system (60 mV for both PAT and US) were then selected for spectral analysis. Each of these beams was gated using a Hamming window before its spectrum was computed within a -10dB frequency range (0.5 to 4 MHz for PAT, and 3.5 to 7 MHz for US) using a fast Fourier transform (FFT).

The transducer response was calibrated using several sets of calibration data. To generate each set of data, we recorded the photoacoustic signals from a black string with a diameter of 250 µm at a particular depth. Then the distance from the transducer to the string was varied from 0.5 cm to 7 cm in steps of 0.25 cm, and for each transducer-string distance a set of photoacoustic signals was recorded and averaged. After that, the power spectra of these data were found using an FFT algorithm. Finally, the power spectra of all PAT beam lines in the ROI of the examined ovary were divided by this calibration spectrum, depending on the depth of the ROI center.

After calibration, a line was fitted to each of the calibrated PAT spectra, and the mean spectral slope (SS), midband fit (MBF), and 0.5 MHz spectral intercept (0.5 MHz SI (PAT)) were obtained from all fitted lines within the ROI. Although the 0 MHz intercept has been widely used as a feature in the literature, we chose the 0.5 MHz spectral intercept instead because our transducer's lower band is at ~ 0.5 MHz.

Figure 3.1 shows the mean spectra of PAT signals in the ROI (dashed rectangular area), along with their fitted lines, for a benign mucinous cystadenoma (left) and an ovary of a high-grade serous carcinoma (right). The SS, 0.5 MHz SI (PAT), and MBF of the mean PAT spectrum are also shown in each image. Note that the spectral features were obtained at four different wavelengths (730, 780, 800, and 830 nm), but each feature was highly correlated for different wavelengths. Thus, the spectral features at one wavelength (730 nm) were used for our classifiers.

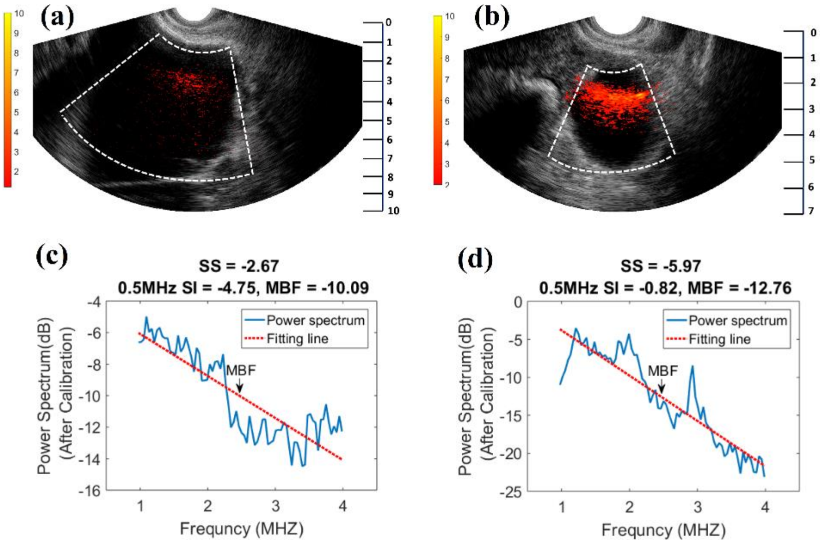


Figure 3.1: Top row: co-registered rHbT and US images of a benign mucinous cystadenoma (a) and a high-grade serous carcinoma (b). The vascular distribution of the benign lesion is more scattered, but more localized and intense for the malignant ovary. Bottom row: the calibrated PAT power spectra and their fitted lines in the regions associated with the angular dashed lines in each image. Note the different Y-axis depth ranges.

3.2.4 PAT image features

In addition to functional and spectral features, we observed that the textures of PAT images for benign and malignant ovarian tissue look different in the patient data. This observation led us to quantify this difference by calculating more imaging features from the PAT images. The first step was to choose an ROI. To find the ROI in each frame, we first selected a larger rectangular region associated with the ovarian tissue. Then the Radon transforms of the image at angles of 0 and 90 degrees in the ovary area were calculated. A Gaussian curve was fitted to each of them after they were normalized to their peak values. The means of the Gaussian curves determined the center of a 2 cm by 2 cm square where the image analysis was performed (Figure 3.2).

Second order statistics of the normalized PAT images were computed within the ROI. These features provide information about the relation between pixel connections. To calculate the textural features of the PAT images, a gray-level co-occurrence matrix was created [27]. This matrix had dimensions of N*N, where N is the number of gray levels in the PAT image. In this study, we used N=16. The value c(i,j) of the (i, j) element of the GLCM represents the number of times that gray levels i and j are adjacent to each other in the PAT image. In the present study, we considered two gray levels g1 and g2 as adjacent if g2 was located at the immediate right of g1. After forming the GLCM matrix, the following four textural features were computed for each PAT image:

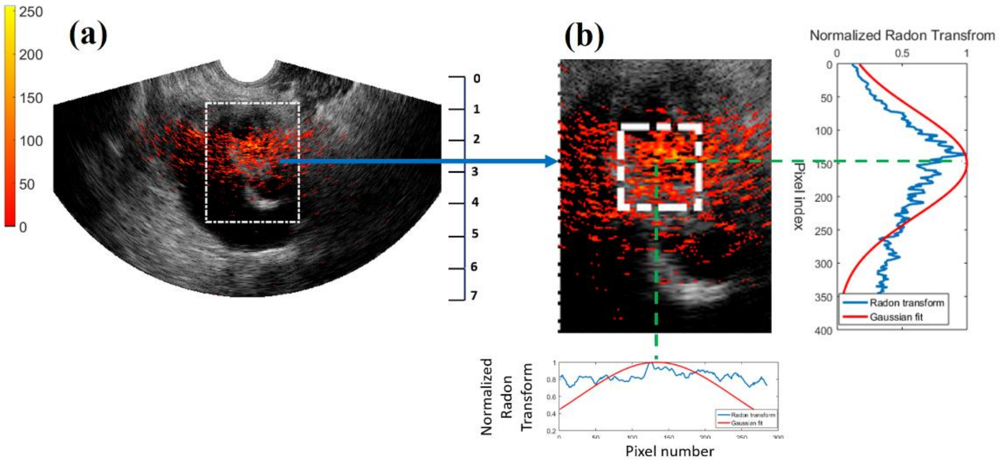


Figure 3.2: ROI selection for image analysis. A larger rectangular region associated with the ovarian tissue is first selected (a). After that, the Radon transforms of the image at angles of 0 and 90 degrees in the selected area are calculated (b). These Radon transforms are then normalized, and a Gaussian curve is fitted to each of them. The means of the Gaussian curves determine the center of a square with a side of 2 cm where the image analysis is performed.

$$Contrast = \sum_{|i-j|=0}^{N-1} |i-j|^2 \sum_{i=1}^{N}\sum_{j=1}^{N} c(i,j), \qquad \text{Eq. (3.4)}$$

$$Correlation = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(i-\mu_i)(j-\mu_j)c(i,j)}{\sigma_i\sigma_j}, \qquad \text{Eq. (3.5)}$$

$$Energy = \sum_{i=1}^{N}\sum_{j=1}^{N} c(i,j)^2, \text{ and} \qquad \text{Eq. (3.6)}$$

$$Homogeneity = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{c(i,j)^2}{1+|i-j|}, \qquad \text{Eq. (3.7)}$$

where N is the dimension of this matrix, and μ and σ are the mean and standard deviation for row i or column j of the GLCM [26].

Besides the second order statistics of the normalized PAT images, we calculated other features related to the non-normalized PAT envelope data. Two features of this type include the standard deviation and fitting error of the Gaussian function that is fitted to the mean Radon transform of the non-normalized PAT images (std Rad). To calculate the mean Radon transform, we computed the Radon transform of the non-normalized PAT images in the ROI for angles from 0 to 90 degrees, with a step size of one degree, and took the average of them. Another feature that we extracted from the non-normalized images was the area of the PAT image in the ROI. This parameter was defined as the percentage of the pixels in the ROI with a value higher than the noise level. Other PAT image features, including malignant and benign spatial filters and Rayleigh fit parameters, were also calculated, but no significant differences in these features were observed between benign and malignant masses. A detailed explanation of how to calculate these features can be found in [16].

3.2.5 Classification

We imaged 24 patients (39 ovaries) using our hybrid PAT/US system (see Table 3.1). Among these, 17 patients (26 ovaries) had benign/normal ovaries, 4 patients (8 ovaries) had invasive epithelial cancers, 1 patient had one normal ovary and one epithelial cancer, and the rest (2 patients, 3 ovaries) had other types of neoplasm. In this study, each ovary was considered as an independent sample.

Table 3.1: Lesion Characteristics (24 patients, 39 ovaries; average age 54 years, range 34-76 years)

Invasive epithelial ovarian cancer	high grade serous carcinoma (n=5), endometrioid carcinoma (n=4) (average size 10 cm, range 2.8 - 20 cm)
Other neoplasm	serous borderline (n=2) tumor, sertoli-Leydig cell tumor (n=1) (average size 11.1 cm, range 4.5-19.2 cm)
Benign ovaries	fibrothecoma (n=1, size 14 cm), mature teratoma (n=1, size 6 cm), serous or mucinous cystadenoma or cystic endometriosis (n=11, average

	10.1 cm, range 1.8-37 cm), complex or simple cysts (n=10, average 4.5 cm, range 2.5-7.6 cm)
Normal ovaries	no histopathological abnormalities (n=4, average 2.4 cm, range 2.1 -2.8 cm)

Two types of linear classifiers were investigated in this study, i.e., GLM and SVM. GLM has proven to be an excellent classifier for binary classification problems. To train this classifier, the true response (0 or 1) is plotted as a function of the input parameters for each sample in the training data set. Then a logistic function is fitted to the data points. Therefore, the output of this classifier is a value between 0 and 1 which represents the probability of an observation belonging to an output category. To test this classifier, a threshold value is defined to put each testing sample in one of the two classes (0 or 1). If the probability is less than this threshold, the new sample is classified as label 0 and otherwise, as label 1. The ROC curve can be derived by changing the threshold value and the AUC value can be computed to evaluate the performance of the classifier.

In a binary SVM classification problem, the training data sets are plotted on an n-dimensional space, where n indicates the number of input features. Then the support vectors, i.e., the points which are located nearest to the margin of each of the two classes are found. These vectors are employed to find the hyperplane which best separates the two classes. If the data sets are not linearly separable, a kernel function can be defined to map the data to a higher dimension space. An SVM classifier which is developed using this method is called a non-linear SVM with a kernel function. In this study, we used a linear SVM which has proved to perform better than the non-linear SVM classifiers.

We first developed models to differentiate benign/normal masses from epithelial cancers. This was performed by constructing GLM and SVM classifiers. We were also interested to find out how distinguishable the benign/normal group was from the epithelial and other neoplasm. Therefore, another set of GLM and SVM classifiers were developed to achieve this goal.

To train each classifier, we used a cross-validation method with 200 iterations. In each iteration, two-third of the data were randomly selected for training the classifier, and the rest were used for testing. For both training and testing data sets, the ROC curves were calculated at each iteration. The mean ROCs were then determined by taking an average over all individual ROCs, and the AUCs were then employed to evaluate the performances of classifiers.

As previously discussed, the PAT functional features in our study were calculated by extracting PAT data at four different wavelengths. However, we were also interested to see the performances of the classifiers in the case when the PAT signals are acquired at one wavelength. This performance is very important to evaluate the co-registered PAT and US technique if a laser system only has one wavelength available. For this evaluation, the PAT functional features were excluded from the feature set and new classifiers were constructed using other PAT features. These features were divided into three groups based on the correlation value between each two features (Table 3.3). The highly correlated features were in one group. Based on this criterion, the first group includes the PAT spectral feature only, i.e., 0.5 MHZ SI (PAT). The second group includes area, contrast and correlation. Finally, the third group consists of the remaining image feature, i.e., std Rad, which has low correlation values with all the other image features. To develop the classifiers, we first used just 0.5 MHz SI (PAT). Then one feature in the second group whose combination with 0.5 MHz SI (PAT) resulted in the highest testing AUC was added to the feature set. Finally, std Rad was also included in the features set, and its effects on the performance of our classifiers were investigated by evaluating the training and testing AUCs.

3.3 Results

3.3.1 PAT spectral features

In Figure 3.1, the co-registered rHbT and US images of a benign (a) and an epithelial cancer (b) are shown. The PAT spectral analysis for each of these ovaries was performed in the ROI (specified by the dashed angular area). The plots of the average PAT spectra after calibration for these ovaries along with their fitted lines are shown in the bottom row of the figure. As can be seen in this figure, the value of the 0.5 MHz SI (PAT) of the ovary with an epithelial tumor is higher than that for a benign/normal ovary. This is the only spectral feature which is significantly different between benign/normal ovaries and the ovaries with epithelial cancers ($P<0.05$).

3.3.2 PAT image features

Co-registered PAT and US images of a set of benign fibrothecoma and high-grade serous carcinoma are shown in Figure 3.3. The rectangular dashed area in each image indicates the region where the PAT image analysis is performed. The center of this area was found using the Radon transforms of the PAT signals as explained in the previous section. The images on the lower panel are a zoom-in view of the PAT image in the selected rectangular area. As can be seen, the benign ovary seems to have higher local variations in pixel values of the PAT image in comparison with the malignant ovary which leads to a larger contrast and lower correlation value for the benign ovary. The pixel values in the malignant ovary, on the other hand, seem to be more correlated than those in the benign ovary, so a higher value of correlation is expected in the images of malignant ovaries.

3.3.3 Classification by inclusion of functional features in the features set

Classification of benign/normal ovaries and epithelial tumors

To determine the difference between the three groups of ovarian masses in terms of the extracted features, a student t-tests was performed on each feature for each pair of the three groups. Then the features with a p-value of less than 0.05 between benign/normal and epithelial cancer groups were selected as the significant features. The box plots of these features are shown in Figure 3.4. The p-value between each pair of groups can also be seen on each plot. As can be seen, except for sO2, there is a higher similarity between “others” and “benign/normal” than “others” and “epithelial cancer”. This suggests that “others” group may not be separable from “benign/normal” group if no sO2 is used.

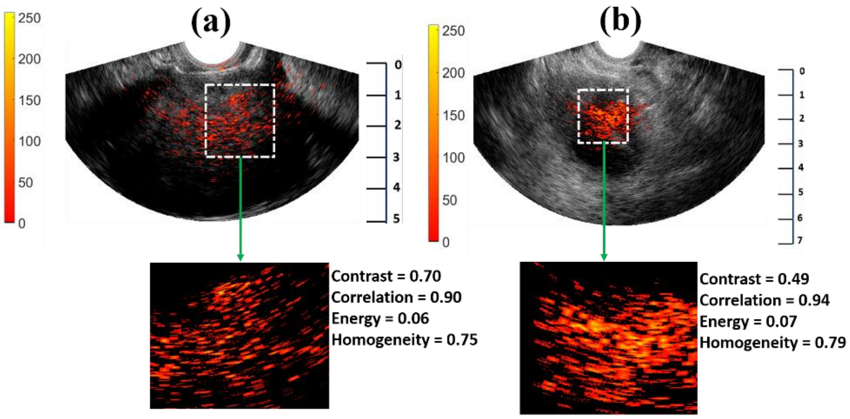


Figure 3.3: Co-registered PAT and US images and magnification of the PAT images in the areas indicated by the dashed rectangle for a benign fibrothecoma (a) and an ovary with epithelial cancer (b). The values of the textural features for each image are also shown.

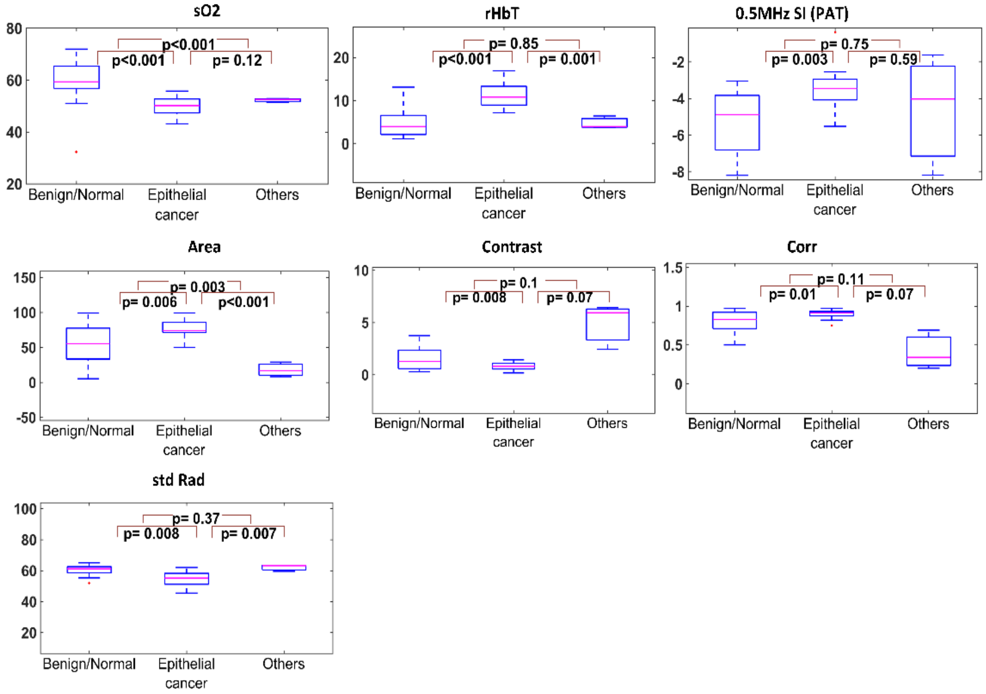


Figure 3.4: Box plots of the significant features for the three groups of ovaries. For each feature, the p-value between each pair of the three groups is shown in the plots.

First, we have constructed GLM and SVM models to distinguish the benign/normal group from the epithelial cancer group. Note that a key step in any classification algorithm is model construction using a proper set of representative predictors. A proper set includes a combination of features which are highly correlated with the class labels (0 or 1), but poorly correlated with each other. To evaluate the correlation between each feature and the class label and between each two features, Spearman's rho value was calculated. After that, we ranked the features based on the value of Spearman's correlation between each feature and the labels. This ranking for the seven significant features, along with the ranking based on the p-values, is shown in Table 3.2. As can be seen, rHbT is at the top of the ranking, based on Spearman's correlation as well as p values. Also note that, although 0.5 MHz SI (PAT) has a lower p-value than std Rad, it is ranked lower than std Rad when the ranking criterion is the correlation value between the feature value and the ovarian mass label.

The correlation between each two significant features can be seen in Table 3.3. The values in this table show that rHbT has a rather high correlation with all the image features, the lowest one being the std Rad, with a correlation value of 0.46. Moreover, all the image features except for std Rad are highly correlated. Among the image features, contrast has the highest correlation with std Rad, with a Spearman’s rho of 0.13.

Table 3.2: Ordering the significant features for distinguishing benign/normal ovarian masses from epithelial cancer, based on their p-values (left) and Spearman’s rho between each feature and the class label.

Ordering features based on p-value		Ordering features based on correlation	
Feature	p-value	Feature	|Spearman’s rho|
rHbT	<0.001	rHbT	0.64
sO2	<0.001	sO2	0.61
0.5 MHz SI (PAT)	0.003	std Rad	0.52
Area	0.006	0.5 MHz SI (PAT)	0.46
Contrast	0.008	Area	0.36
std Rad	0.008	Correlation	0.28
Correlation	0.01	Contrast	0.26

Table 3.3: Spearman’s cross correlation between each two features in the set of significant features for distinguishing benign/normal ovarian masses from epithelial cancer.

	rHbT	0.5 MHz SI (PAT)	Area	Contrast	Correlation	std Rad
sO2	0.39	0.15	0.35	0.19	0.19	0.28
rHbT		0.36	0.8	0.6	0.63	0.46
0.5 MHz SI			0.51	0.46	0.47	0.05
Area				0.61	0.68	0.11
Contrast					0.97	0.13
Correlation						0.09

To train classifiers, we started constructing our model by using the feature with the highest correlation with the class label. Then we added the other significant features one by one until we found that adding features does not improve the classification performance. Figure 3.5 (top row) shows the ROCs for the training and testing data sets using a GLM model. The top left shows that the AUC for the training data set is 0.99 when rHbT, sO2 and 0.5 MHz SI (PAT) are used for the classifier construction. For the testing data set (top right), however, the highest value of AUC is achieved when rHbT and sO2 features are employed for the classification (AUC=0.93). Adding 0.5 MHz SI (PAT) or std Rad to the feature set decreases this value by 0.02 and 0.05, respectively, and further inclusion of any of the other significant features results in an additional decrease in the AUC values (not shown in the plot).

The SVM classifier performance in differentiating benign/normal lesions from epithelial cancers for the training and testing data sets is shown in Figure 3.5 (bottom row). As can be seen, in this case, unlike for GLM, adding 0.5 MHz SI (PAT) to the features set of "rHbT, sO2" increases the value of the AUC from 0.93 to 0.94. However, similar to the GLM classifier, inclusion of std Rad in the features set decreases the AUC value for the testing data set.

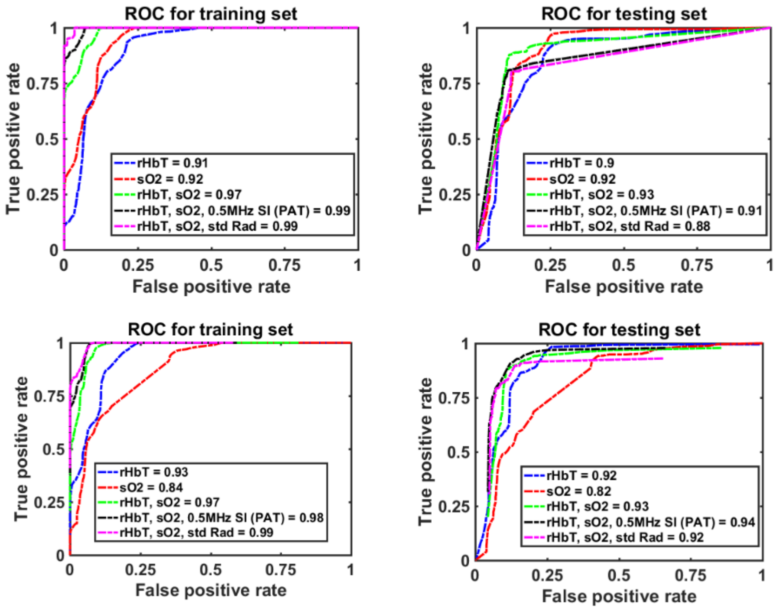


Figure 3.5: ROCs for the training (left) and testing (right) data sets and the associated AUCs for different feature sets, using the GLM (upper row) and SVM (lower row) classifiers. The classifiers distinguish benign/normal from epithelial cancers.

Classification of benign/normal ovaries and all types of tumors

We then had epithelial cancer and other neoplasm types in one group and compared the calculated features of this combined group with those of benign/normal group. Only four features, i.e., sO2, rHbT, 0.5 MHz SI (PAT), and sdt Rad, among all the calculated features were statistically significant ($p<0.05$). Box-plots of these features are shown in Figure 3.6. The sO2 and rHbT values for the “epithelial and other cancers” group are respectively lower and higher than those of the benign/normal group. The 0.5 MHz SI (PAT) value is generally higher for cancerous ovaries than for of benign/normal ones, which indicates more low frequency components for cancers. Moreover, the std Rad is lower for cancer than for normal ovaries, because the vascular distribution of the benign mass is more scattered than in tumors (higher standard deviation).

In Table 3.4, these four significant features are ordered based on their p-values (left), as well as on the value of Spearman’s rho (right). In this case, both ranking criteria determine sO2 to be the most valuable parameter. Also note that among the three significant features selected for model construction, std Rad and rHbT are the most well correlated pair (Spearman’s rho = 0.47), while Spearman’s rho between std Rad and 0.5 MHz SI (PAT) is the lowest (0.02), as shown in Table 3.5.

Figure 3.7 (top) shows the ROC plots for training and testing data sets when a GLM classifier was employed to distinguish the benign/normal group from all types of ovarian cancers. As can be seen, sO2 alone results in training and testing AUC values of 0.90, while considering rHbT as the only feature in the features set generates lower training and testing AUCs (0.83 and 0.84, respectively). The highest testing AUC value is obtained when the feature set includes just sO2 and rHbT (0.92). Note that although adding either 0.5 MHz SI (PAT) or std Rad improves the classifiers performance for the training data set, it reduces the testing AUC from 0.92 to 0.87.

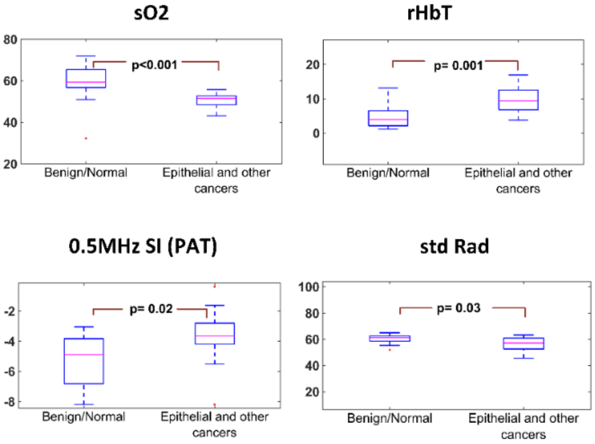


Figure 3.6: Box plots of the significant features for two groups of ovaries. For each feature, the p-value between the two groups is shown.

Table 3.4: Ordering the significant features for distinguishing benign/normal ovarian masses from epithelial and other ovarian cancers, based on their p-values (left) and Spearman’s rho between each feature and the class label.

Ordering features based on p-value		Ordering features based on correlation	
Feature	p-value	Feature	|Spearman’s rho|
sO2	<0.001	sO2	0.64
rHbT Mean	0.001	rHbT Mean	0.54
0.5 MHz SI (PAT)	0.02	0.5 MHz SI (PAT)	0.41
std Rad	0.03	std Rad	0.36

Table 3.5. Spearman’s cross correlation between each two features in the significant features set for distinguishing benign/normal ovarian masses from epithelial and other ovarian cancers.

	rHbT Mean	0.5 MHz SI (PAT)	std Rad
sO2	0.36	0.16	0.2
rHbT Mean		0.30	0.47
0.5 MHz SI			0.02

The ROC plots for the training and testing data sets using the SVM classifier to differentiate the benign/normal ovary group for all types of cancers are shown in Figure 3.7 (bottom) and compared for different feature sets. As with GLM, the best performance of the SVM model is achieved when sO2 and rHbT are the only features employed in the features set (training and testing AUCs of 0.95 and 0.93, respectively). Moreover, adding 0.5 MHz SI (PAT) feature decreases the testing AUC of the SVM classifier by 0.04, but it does not have much influence on the AUC of the training data set. Finally, inclusion of std Rad decreases the AUC values for both the training and testing data sets.

3.3.3 Classification by exclusion of functional features in the features set

Next, we evaluated the performance of the classifiers assuming that PAT functional features were not available. As discussed before, based on the correlation values between different features, we divided the non-functional PAT features into three groups. The first group contains the single significant spectral feature of 0.5 MHz SI (PAT). PAT image area, contrast, and correlation are included in the second group. The third group is made of std Rad feature only.

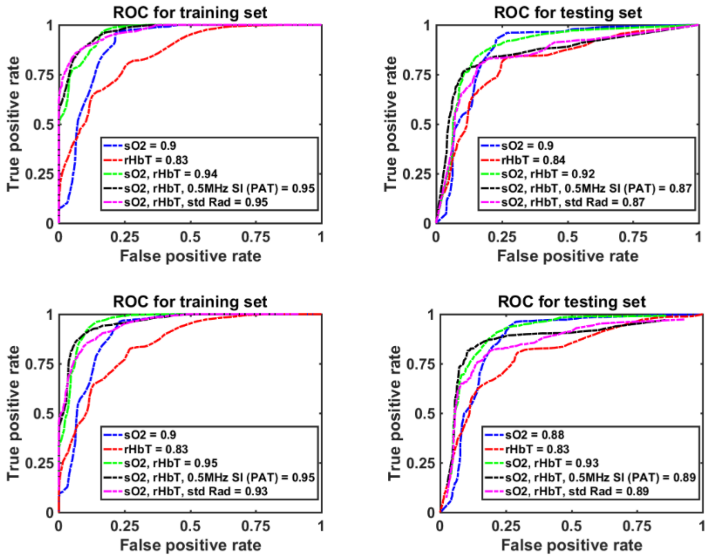


Figure 3.7: ROC for the training (left) and testing (right) data sets and the associated AUCs for different feature sets, using the GLM (upper row) and SVM (lower row) classifiers. The classifiers distinguish benign/normal ovarian masses from epithelial cancer and other neoplasms.

The ROC plots for the training and testing data sets using the SVM classifier to differentiate the benign/normal ovary group for all types of cancers are shown in Figure 3.7 (bottom) and compared for different feature sets. As with GLM, the best performance of the SVM model is achieved when sO2 and rHbT are the only features employed in the features set (training and testing AUCs of 0.95 and 0.93, respectively). Moreover, adding 0.5 MHz SI (PAT) feature decreases the testing AUC of the SVM classifier by 0.04, but it does not have much influence on the AUC of the training data set. Finally, inclusion of std Rad decreases the AUC values for both the training and testing data sets.

Classification of benign/normal ovaries and epithelial tumors

First, we developed classifiers to distinguish benign/normal masses from the epithelial cancer group using PAT spectral and image features. Figure 3.8 shows the training and testing ROCs for these new GLM (top) and SVM (bottom) classifiers. The SVM classifier shows better performance (training and testing AUC of 0.94 and 0.92) when the feature set includes 0.5 MHz SI (PAT), area, and std Rad. Note that we have not used any features from the second group (including area, contrast, and correlation) to develop the GLM classifier, since no feature in this group improved the testing AUC of the GLM classifier.

Classification of benign/normal ovaries and all types of tumors

Finally, we distinguished the benign/normal group from epithelial and other types of neoplasms by using GLM and SVM classifiers without PAT functional features. The training and testing ROC plots for the GLM (top) and SVM (bottom) are shown in Figure 3.9. The performances of the classifiers are moderate in this case, and the highest testing AUC is achieved when 0.5 MHz SI (PAT), std Rad, and contrast are included in the feature set and an SVM model is used to classify the data. Even so, the testing AUC value in this case is just 0.84.

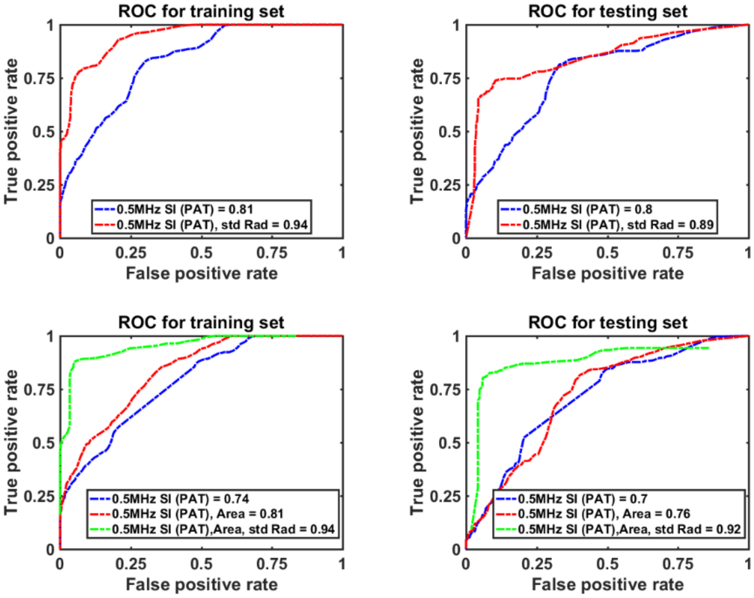


Figure 3.8: ROC for the training (left) and testing (right) data sets and the associated AUCs for different feature sets, using the GLM (upper row) and SVM (lower row) classifiers. The classifiers distinguish benign/normal ovarian masses from epithelial cancers. Functional features are not included in the features set.

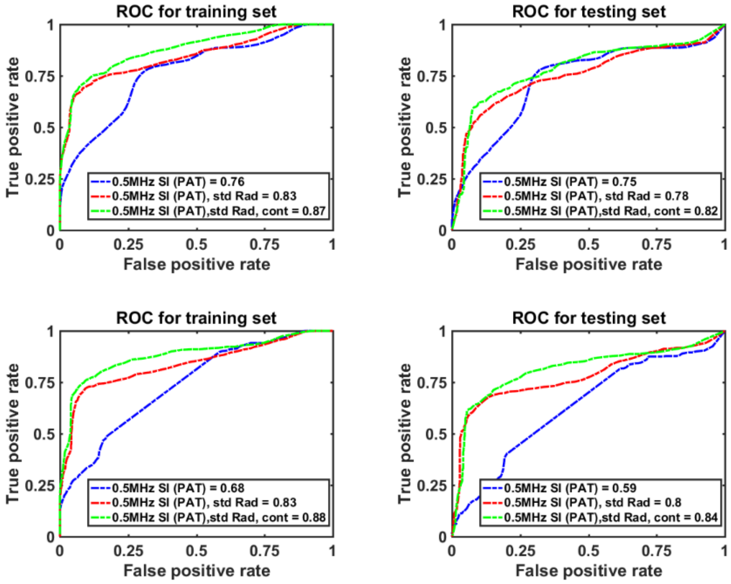


Figure 3.9. ROC for the training (left) and testing (right) data sets and the associated AUCs for different feature sets, using the GLM (upper row) and SVM (lower row) classifiers. The classifiers distinguish benign/normal ovarian masses from epithelial cancers and other neoplasms. Functional features are not included in the feature set.

3.4 Discussion and Summary

In this study, 22 out of the 24 patients have CA-125. Thus, it could be useful to analyze the effect of this parameter on our classification performance. However, because the value of CA-125 is patient-dependent (one number for each patient) rather than ovary-dependent, each patient should be considered as one independent sample. This would decrease the total number of available independent samples from 39 to 22, not adequate to train and test our classifiers. The effect of this parameter on the performances of the classifiers will be evaluated in future studies when more patients' data are available.

Besides the PAT spectral and functional features, US spectral features were also evaluated in this study. As demonstrated by Lizzi et al [27], the SS of the US spectrum is dependent on the sizes of the scatterers and the attenuation of the medium, while its SI is a function of the size, concentration, and acoustic impedance of the scatterers relative to the background medium. Finally, the MBF of the US spectrum is related to the attenuation of the medium, and the size, concentration, and impedance of the scatterers relative to the medium. The US spectral features were not reported in this study because for this feature set, a statistical difference was not found between the benign/normal masses and epithelial cancers ($p>0.05$). We will continue to evaluate the diagnostic values of US spectral features in future studies wen more patients' data are available.

The textural features of the PAT image proved useful in some of the classifiers in this study. Among these features, contrast helpfully quantifies the local variations in the PAT images. When intensity varies significantly from one pixel to another, the contrast values are very large. Correlation is a measure of the dependence among neighboring pixels, and energy represents the local uniformity of the pixel values. A higher similarity in pixel values results in larger energy values. Finally, homogeneity determines the local homogeneity in the image. When there are small local variations in the image, the homogeneity is large.

We were interested in distinguishing normal/benign ovarian masses from the ovaries with epithelial cancers. GLM and SVM classifiers were designed to specifically distinguish this difference. The best performance of an SVM classifier, a testing AUC of 0.94, was achieved when rHbT, sO2, and 0.5 MHz SI (PAT) were used in the features set. For the GLM classifier, the highest testing AUC was 0.93, when rHbT and sO2 were employed as the input features. When functional features were excluded from the classification procedure, the best testing AUC (0.92) was associated with an SVM classifier which was trained using the 0.5 MHz SI (PAT), area, and std Rad features.

Photoacoustic features were also employed to differentiate patients with normal/benign ovaries from patients with ovarian cancers, either epithelial or other neoplasms. In that case, the SVM classifier also performed slightly better than GLM. The testing AUC for these classifiers were 0.93 and 0.92, respectively. For both classifiers, using sO2 and rHbT features in the feature set resulted in the highest testing AUC, and including other features decreased their performance. When functional features were excluded from analysis, the performances of the classifiers were moderate (best testing AUC = 0.84 using an SVM classifier). In summary, these results demonstrate that functional, spectral and imaging features obtained from photoacoustic imaging can comprehensively classify human ovarian cancer.

References

1. American Cancer Society. Cancer facts and figures, 2019, American Cancer Society, Atlanta, GA (2019)
2. https://www.cancer.org/cancer/ovarian-cancer/about/key-statistics.html
3. D.L. Clarke-Pearson, “Screening for Ovarian Cancer,” N Engl J Med 361, 170–177, (2009).
4. K. R. Le, et al., “Surface Epithelial Stromal Tumours: Tumours of the Ovary and Peritoneum,” IARC Press: Lyon; 117–145 (2003).
5. N. J. Finkler et al., “Comparison of serum CA 125, clinical impression, and ultrasound in the preoperative evaluation of ovarian masses,” Obstet Gynecol;72:659-64 (1988).

6. I. Jacobs et al., “A risk of malignancy index incorporating CA125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer,” Br J Obstet Gynarcol;97:922-9 (1990).
7. V. Calster et al., “Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125,” J. Natl. Cancer Inst. 99(22), 1706–1714 (2007).
8. V. Nossov et al., “The early detection of ovarian cancer: from traditional methods to proteomics. Can we really do better than serum CA-125?,” Am. J. Obstet. Gynecol. 199(3), 215–223 (2008).
9. C. Di-Xia et al., “Evaluation of CA 125 levels in differentiating malignant from benign tumors in patients with pelvic masses,” Obstet Gynecol;72:23-7 (1988).
10. A. Shaaban & M. Rezvani. “Ovarian cancer: detection and radiologic staging,” Clin Obstet Gynecol 52(1), 73-93 (2009).
11. R. Bristow et al. “Combined PET/CT for detecting recurrent ovarian cancer limited to retroperitoneal lymph nodes,” Gynecologic Oncology, 99 (2005).
12. Y. Nakamoto et al., “Positron emission tomography application for gynecologic tumors,” Int J Gynecol Cancer (2005).
13. L. V. Wang & S. Hu, “Photoacoustic tomography: in vivo imaging from organelles to organs,” Science; 335(6075):1458–1462, (2012).
14. K. S. Valluru et al., “Photoacoustic imaging in oncology: translational preclinical and early clinical experience,” Radiology; 280(2):332–349, (2016).
15. A. A. Oraevsky & A. A. Karabutov, “Optoacoustic tomography. In: Vo-Dinh T, ed. Biomedical photonics handbook,” Boca Raton, Fla: CRC, (2003).

16. H. Li et al., “Utilizing spatial and spectral features of photoacoustic imaging for ovarian cancer detection and diagnosis,” J Biomed Opt. (2015).
17. H. S. Salehi et al., “Coregistered photoacoustic and ultrasound imaging and classification of ovarian cancer: ex vivo and in vivo studies,” J Biomed Opt (2016).
18. S. Nandy et al., “Role of Co-Registered Photoacoustic and Ultrasound Tomography: Initial Application for Evaluation of Ovarian Cancer,” Radiology (2018).
19. B. Cox et al., “Quantitative spectroscopic photoacoustic imaging: a review,” J Biomed Opt.;17(6):061202, doi: 10.1117/1.JBO.17.6.061202. Review, (2012).
20. T. Feng et al. “Characterization of bone microstructure using photoacoustic spectrum analysis,” Optics express vol. 23,19: 25217-24 (2015).
21. G. Xu et al. “The functional pitch of an organ: quantification of tissue texture with photoacoustic spectrum analysis,” Radiology ;271(1):248-54, (2014).
22. X. Leng et al., "Feasibility of co-registered ultrasound and acoustic-resolution photoacoustic imaging of human colorectal cancer," Biomedical Optics Express Vol. 9, Issue 11, (2018).
23. R. E. Kumon et al., “Frequency-domain analysis of photoacoustic imaging data from prostate adenocarcinoma tumors in a murine model,” Ultrasound Med Biol.;37(5):834-9, (2011).
24. E. Hysi et al., “Photoacoustic signal characterization of cancer treatment response: Correlation with changes in tumor oxygenation,” Photoacoustics. 2017;5:25-35, (2017).
25. U. Alqasemi et al., “Recognition algorithm for assisting ovarian cancer diagnosis from coregistered ultrasound and photoacoustic images: ex vivo study,” Journal of biomedical optics 17 (12), (2012).

26. R. M. Haralick et al., “Textural Features for Image Classification,” IEEE Trans. Syst. Man. Cybern.;3:610–621 (1973).
27. F. L. Lizzi et al, “Relationship of ultrasonic spectral parameters to features of tissue microstructure,” IEEE Trans Ultrason Ferroelectr Freq Contr 34:319–329 (1987).

Chapter 4: Sliding multi-pixel method to improve oxygen saturation estimation using photoacoustic tomography

4.1 Introduction

Ovarian cancer remains the deadliest of all the gynecological malignancies. According to the American Cancer Society, 21,750 women in the US will be diagnosed with ovarian cancer in 2020, and 13,940 deaths will be reported [1]. Due to a lack of effective screening tools, only 20-25% of ovarian cancers are diagnosed early. A recent update from the U.S. Preventive Services Task Force concluded that major trials of promising ovarian cancer screening tools have null findings to date among healthy average-risk women, and there are considerable harms associated with screening, which include major surgical complications in women found to have no cancer [2]. Women with a screening abnormality generally undergo prophylactic bilateral salpingo-oophorectomy [3-5], which can result in morbidity and mortality from premature menopause, including accelerated bone loss and cardiovascular death [6]. Thus, there is an urgent need to develop better and more sensitive tools to effectively evaluate the ovary.

Current screening modalities include bimanual pelvic examination and transvaginal ultrasound (US), as well as assessment of cancer antigen 125 (CA125) serum marker levels, which are associated with a high false-positive rate [2,8]. Conventional MRI and diffusion-weighted MRI are useful follow-up modalities for investigating sonographically indeterminate adnexal masses [9]. Using MRI to measure sO2 for ovarian tumor assessment has not been well studied and is potentially valuable for future research. CT has been historically preferred for the pretreatment evaluation of ovarian cancer to determine both the extent of the disease and the optimal surgical treatment [10]. FDG-PET/CT can measure residual or recurrent disease and also help select the surgical treatment [11]. However, FDG-PET has limited value in localizing lesions in the early stages of ovarian cancer [12].

In the past decade, with advances in lasers, ultrasound transducers, and tomographic reconstruction techniques, photoacoustic imaging or photoacoustic tomography (PAT) has seen an immense growth, providing high spatial resolution same as ultrasound and functional information at depths ranging from several millimeters up to several centimeters [13-19]. PAT is a hybrid imaging technology that uses a short-pulsed laser to excite the tissue. Acoustic (or photoacoustic) waves are then generated from thermoelastic expansion due to a transient temperature rise and are measured by US transducers. PAT image contrast is related to tissue optical absorption properties and therefore tumor vasculature or tumor angiogenesis. Oncologic targets of PAT to date include breast cancer [20-28], prostate cancer [29, 30], skin cancer [31], thyroid cancer [32], colon cancer [33], and ovarian cancers [34-35]. Our limited data have indicated that high grade stage I invasive ovarian cancers show a higher hemoglobin or vasculature content and a lower sO2 [34].

sO2, an indicator of tumor metabolism and therapeutic response, is one important diagnostic parameter measured by PAT. To compute sO2, PAT data are acquired at multiple wavelengths, then used to solve linear equations that relate the PAT signals to the oxy and deoxyhemoglobin concentrations at each pixel in the region of interest (ROI), simultaneously obtaining the concentrations of these chromophores at that pixel. This method of calculating sO2, named “linear unmixing”, has been utilized in many studies [36-45]. In this study, for 49 ovarian masses from 33 patients, we present sO2 maps computed using this method. However, to mitigate the effect of the measurement errors (e.g. spatial and temporal variations in light fluence, system noise, and motion), we replace the value of each pixel with the average of the values of 100 pixels of 1.0 mm by 1.0 mm in size around it before solving the linear equations. The estimated sO2 maps calculated using this method, along with rHbT mean value and PAT spectral features, are used to develop support vector machine (SVM) models. We demonstrate with a considerable patient population that a combination of PAT features with the radiologists' diagnostic score, can significantly improve diagnostic accuracy, achieving an area under the receiver operating characteristic curve (ROC) of 0.93 on the testing data set.

4.2 Materials and methods

4.2.1 Phantoms

To evaluate the accuracy of our sO2 estimation, we prepared five blood tube phantoms, with sO2 values of 24.9%, 44.2%, 64.9%, 83.9%, and 97.6%. The red blood cells (RBCs) were collected from healthy volunteers and were mixed with 60% saline water inside a hypoxia chamber with a temperature/humidity control. Each blood sample with the desired sO2 value was produced by controlling the amount of oxygen and nitrogen in the chamber. The sO2 value of each blood sample was calibrated using an ABL90 FLEX Radiometer. This value was considered as the calibrated sO2 value in the phantom study.

After each blood sample was prepared, it was injected into a 3 mm diameter tube, which was then tightly capped in the chamber in order to prevent the room oxygen from affecting the sO2 of the blood sample. Each tube was sequentially positioned at a progression of depths from 1 to 5 cm, with a step size of 0.5 cm, in a homogeneous Intralipid® solution with calibrated absorption and reduced scattering coefficients of 0.02 and 4 cm^{-1}, similar to the optical properties of soft tissue. The PAT/US system was used to collect data from each blood tube at each depth and at wavelengths of 730, 780, 800, and 830 nm. To prevent motion during the data acquisition, the transducer was fixed to a post. The acquired PAT data were employed to calculate the sO2 maps inside the blood tubes, using linear unmixing (described in 4.2.4). Finally, the mean sO2 value at each depth was calculated by taking an average over all the pixels with sO2 values greater than zero.

4.2.2 Patients

The study protocol was approved by the institutional review board and was compliant with the Health Insurance Portability and Accountability Act. A total of 40 patients with ovarian masses gave informed consent and participated in this study from February 2017 to November 2018. Out of these patients, 7 were excluded from the data analysis: 3 patients due to PAT/US system problems, 2 patients without any ovarian masses found at the time of imaging, and 2 patients with deep ovarian masses in the range of 6 to 7 cm from the probe. In 18 of the remaining 33 patients, both ovaries were imaged. Among the other 15 patients, only one suspicious ovary (with pathological evaluation available) was imaged and surgically resected. Thus, the total number of the imaged ovaries was 51, out of which two ovaries were not considered in our analysis because pathology data was not available for them (Figure 4.1).

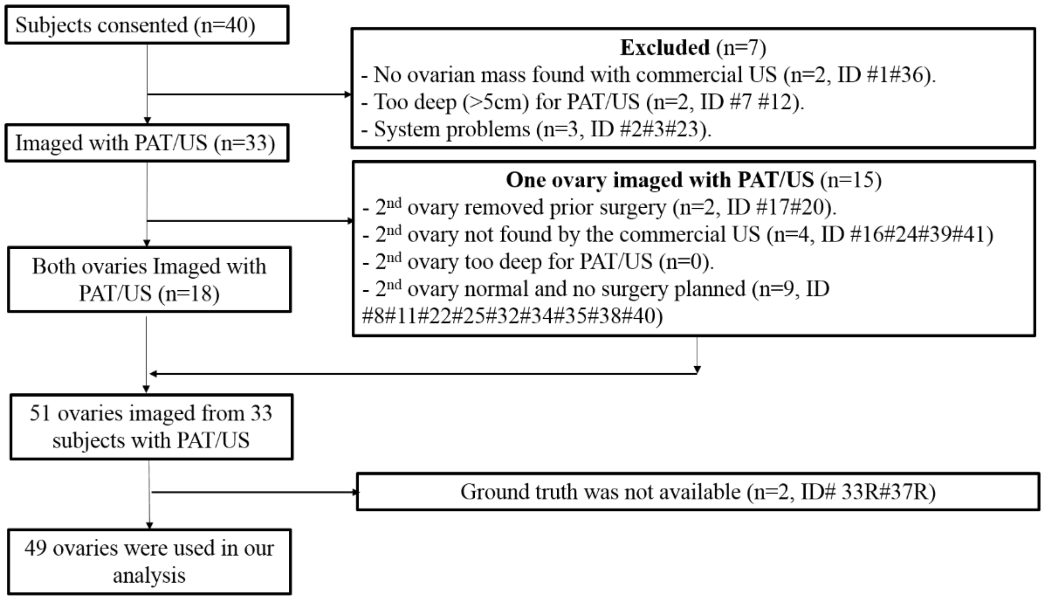


Figure 4.1: Flowchart for inclusion and exclusion of study participants, including reasons for exclusion.

In Table 4.1, the imaged ovaries in this study have been divided into four main categories of invasive epithelial ovarian cancer, other neoplasm, benign ovaries, and normal ovaries. On the right side of the table, each mass type has been divided into more subgroups. The number of ovaries belonging to each subgroup as well as the information about the mass size and depth is also provided in the table.

Table 4.1: Lesion Characteristics (33 patients with 49 ovaries: average age of 56 years, range 33-87 years)

Invasive epithelial ovarian cancer	high grade serous carcinoma (n=7), endometrioid carcinoma (n=3) (average size, 9 cm; size range, 2.8 - 20 cm; average depth, 3 cm; depth range, 2.2 - 4.3 cm)
Other neoplasm	serous borderline (n=2) tumor, Sertoli-Leydig cell tumor (n=1) (average size, 11.1 cm; size range, 4.5 - 19.2 cm, average depth, 3.6 cm; depth range, 2.6 - 5.5 cm)
Benign ovaries	fibrothecoma (n=1; size, 14 cm; depth, 1.9 cm), mature teratoma (n=1; size, 6 cm; depth, 3.1 cm), serous or mucinous cystadenoma or cystic endometriosis (n=17;

	average size, 10 cm; size range, 1.8 – 3.7 cm; average depth, 2.7 cm; depth range, 0.7 - 7.4 cm), complex or simple cysts (n=10; average size, 4.3 cm; size range, 3 - 7.6 cm; average depth, 2.7 cm; depth range, 0.6 – 5.5 cm), benign leiomyoma (n=1; size, 5.5 cm; depth, 1.8 cm), fibrosis and cyst (n=1; size, 4 cm; depth, 1.9 cm)
Normal ovaries	no histopathological abnormalities (n=5; average size, 2.5 cm; size range, 2.1 -2.8 cm; average depth, 1.6 cm; depth range, 0.6 - 2.6 cm)

4.2.3 Ovarian mass ranking

For each patient, one attending radiologist evaluated all available images of x-ray CT, MRI, prior US from the patient's medical record as well as the real-time US images. Based on this information, the radiologist provided a score in a range of 1 to 5 for each ovarian mass before the co-registered photoacoustic and ultrasound exam. Scores 1 and 2 are normal, 3 is likely benign, 4 is suspicious for malignancy, score 5 is highly suspicious for malignancy. For the data analysis, we categorized scores 1 and 2 as normal, score 3 as benign, and 4 and 5 as malignant. This mapped feature is referred to as the radiology ranking (Rad rank) in this manuscript.

4.2.4 PAT functional features

To calculate the PAT functional features, i.e. rHbT and sO2, we need to estimate relative oxy and deoxy hemoglobin concentrations, $\hat{C}_{oxy}(r,\theta)$ and $\hat{C}_{deoxy}(r,\theta)$, which can be obtained using the linear unmixing approach [36-45]. Briefly, we find a solution at each pixel by solving a linear least square problem with the constraint of non-negative solutions as:

$$\text{Object F} = \text{argmin}\ \|g - Hf\|^2\ ; f \geq 0\ , \qquad \text{Eq. (4.1)}$$

where g represents measurements at 4 wavelengths used, H is the known matrix consists of extinction coefficients of the four wavelengths, and f represents the $\hat{C}_{oxy}(r,\theta)$ and $\hat{C}_{deoxy}(r,\theta)$ that need to be estimated.

$$g=\begin{bmatrix} P_{0n-730} \\ P_{0n-780} \\ P_{0n-800} \\ P_{0n-830} \end{bmatrix},\ H=\begin{bmatrix} \varepsilon_{oxy-730} & \varepsilon_{deoxy-730} \\ \varepsilon_{oxy-780} & \varepsilon_{deoxy-780} \\ \varepsilon_{oxy-800} & \varepsilon_{deoxy-800} \\ \varepsilon_{oxy-830} & \varepsilon_{deoxy-830} \end{bmatrix}, and$$

$$f=\left[\hat{C}_{oxy}(r,\theta) \quad \hat{C}_{deoxy}(r,\theta)\right].$$

The subscripts 730, 780, 800, and 830 in the above equations denote the wavelengths used in this study. The rHbT is then simply calculated by summing them. Furthermore, sO2 is estimated by dividing the relative oxyhemoglobin by rHbT.

In our previous studies [34,35], we calculated the functional features at each single pixel by solving Eq. (4.1). We then applied a spatial filter to improve the sO2 image quality. However, due to the measurement errors arise from spatial and temporal variations in light fluence caused by tissue optical absorption and scattering heterogeneities, system noise, and motion, solving a linear equation at each single pixel might not be a robust method for finding the functional features. To mitigate this problem, we replaced the value of the envelope data in each pixel by the mean of the envelope at an area around it with a size of 10×10 pixels. We named it multi-pixel method. Thus, for each pixel, the area of 1.0 mm × 1.0 mm is chosen so that the smaller pixel is in the center of the 10×10 pixels.

We also investigated using Gaussian and Lorentzian algorithms as spatial filters to smooth the PAT data before calculating the sO2 maps. The size of each filter was fixed at 1.0 mm × 1.0 mm (10 pixels by 10 pixels), the same as the size in the multi-pixel method. To take advantage of the spatial filter's ability to mitigate measurement errors without significantly impairing the spatial resolution, we had to assign appropriate values to the parameters of the filters. The filter parameters that we adjusted were the standard deviation of the Gaussian filter and the full width at half maximum (FWHM) of the Lorentzian filter, both indicated by “sigma” in this manuscript. Small standard deviations of the Gaussian filters or small FWHMs of the Lorentzian filters give too much weight to the pixels in the center of each kernel, neglecting the contributions of the pixels around it. The resulting sO2 maps are undesirably similar to maps without any spatial filters. On the other hand, if a very large standard deviation or Lorentzian parameter is chosen, all the pixels in the kernel will have almost the same weight. In this case, there will be no difference between the generated sO2 maps calculated using these filters and those estimated using sliding multi-pixels. Because our multi-pixel kernel size is 10 pixels, we decided to assign half of this value (5) to the standard deviation of the Gaussian filter and Lorentzian filter parameter.

After finding the sO2 map, we constructed its histogram and calculated the histogram features (mean, standard deviation, skewness, kurtosis, energy, and entropy). Among these features, three (mean, skewness, and energy) showed significant differences between the benign and malignant ovarian groups.

To compare the robustness of the different smoothing methods, we calculated the value of the *squared norm of the normalized residuals* (for simplicity, called the *normalized residual* throughout this manuscript) at that pixel. To find this parameter at each pixel, after solving the linear optimization of Eq. (4.1), we normalized each element of $(g - Hf)$ by the summation of all four wavelengths of g at that pixel, thereby removing the dependency of the residuals on the magnitude of g. Then we calculated the normalized residual at that pixel as

$$Normalized\ Residual = (g - Hf)_n^T \times (g - Hf)_n, \qquad \text{Eq. (4.2)}$$

where superscript T indicates the transpose of the matrix, and subscript n means the matrix has been normalized pixel by pixel after optimization. The better the data fits the model, the lower the value of the normalized residual will be. Thus, we use this parameter to evaluate sO2 robustness.

4.2.5 Statistical analysis, feature selection, and classification

Each investigated ovarian mass in this study was included in either the benign/normal group (36 ovaries) or the malignant group (13 ovaries). To evaluate the difference between these two groups, a two-tailed t-test with the assumption of an unequal variance between the two groups was performed on each available feature. A feature with a p-value equal to or less than 0.05 was considered significant. Moreover, a random forest model was employed to rank the importance of each feature. The importance of each feature is calculated by averaging the Gini impurity at each node of the forest where that feature is used. Scikit-Learn, a module in Python, was used to calculate the features importance. Further information about feature selection using random forest models is found in Ref [49].

Support vector machine (SVM) models were trained to distinguish benign/normal ovaries from those with malignant tumors. Models were developed for three subsets of uncorrelated significant features extracted from all the ovaries: radiology ranking (Rad rank) alone; uncorrelated PAT features (sO2 mean, and rHbT mean, spectral intercept (SI)) alone; and the combination of uncorrelated PAT features and the radiologist's ranking (sO2 mean, rHbT mean, SI, and Rad rank). Each model was trained 100 times, each time using a random selection of 2/3 of the available dataset for training, and the rest of the data was reserved for testing. At each iteration the ROC of the training and testing data sets were computed. The mean receiver operating characteristic (ROC) curve and the area under the curve (AUC) over all iterations were then calculated for the training and testing data sets. The model which provided the highest AUC value for the testing data set was considered as the best model. Note that, the AUC of the training data set was found just to monitor the classifiers for overfitting (overfitting occurs when training AUC increases, but the testing AUC does not).

The mean AUC values were also used to calculate the 95% confidence interval (95% CI) for each constructed model using a binomial formula.

4.3 Results

4.3.1 Ranking by the radiologists

A boxplot of the ovarian mass ranking is shown in Figure 4.5 (a). Note that our expert radiologist successfully detected all malignant ovaries, but a few benign ovaries were misclassified as malignant.

4.3.2 sO2 calculation for blood tube phantoms

In Figure 4.2, we demonstrate the accuracy of our sO2 calculations when the medium was filled with homogeneous Intralipid® solution. The multi-pixel method was used before linear unmixing. Each column in this figure shows the result for a different blood tube, with its associated calibrated sO2 value indicated at the top of the column. The images are from experiments in which the tube was located at a depth of 2.5 cm. Figure 4.3 summarizes the calculated mean sO2 values at different depths. Each box plot in this figure shows the calculated mean sO2 value at depths from 1 to 5 cm, with a step size of 0.5 cm. The mean absolute estimation errors at different depths for the blood tubes with the calibrated sO2 of 24.9, 44.2, 64.9, 83.9, and 97.6 are 5.6%, 3.8%, 9.8%, 8.7%, and 17.7%, respectively.

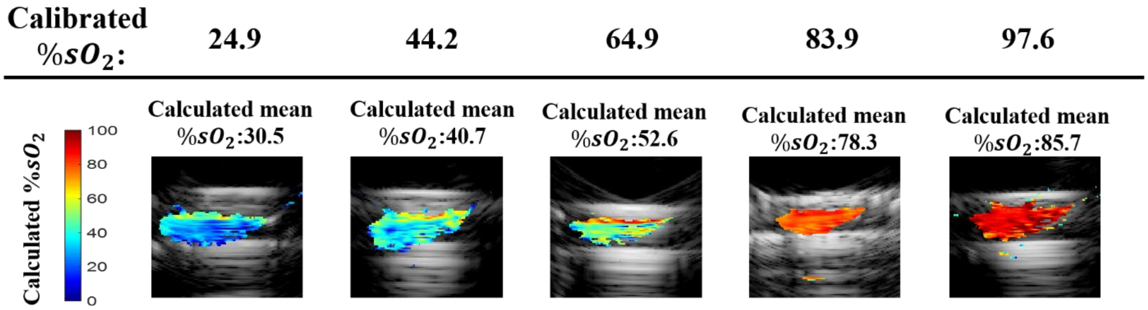


Figure 4.2: The co-registered US and sO2 maps for different blood tubes located at the depth of 2.5 cm in intralipid. Each column indicates the sO2 maps for a blood tube with the calibrated sO2 value specified at the top of the column. The mean of the calculated sO2 is above each sub image.

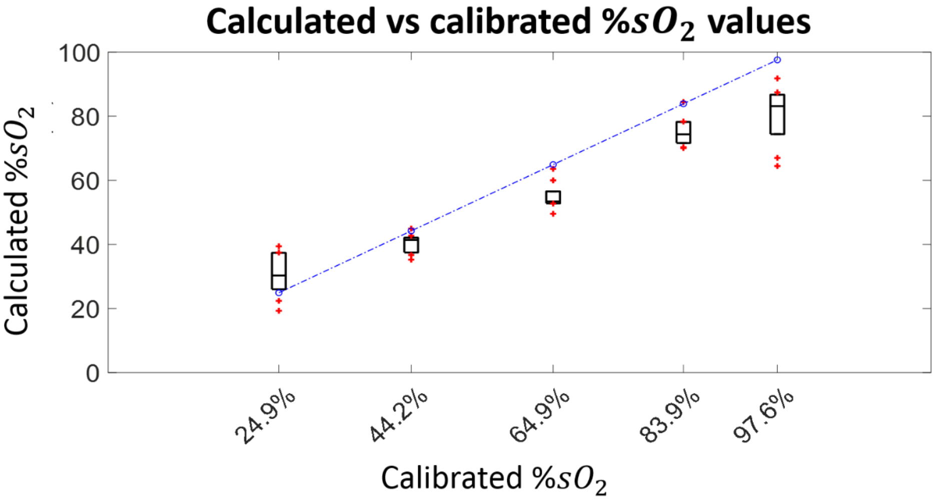


Figure 4.3: Calculated sO2 values vs calibrated values. Each blood vessel was placed in Intralipid at depths from 1 to 5 cm below the probe surface, in nine successive steps of 0.5 cm

each. At each depth, the mean sO2 value was calculated. Each box plot summarizes the mean calculated sO2 at these 9 different depths.

4.3.3 PAT features for patients

In Figure 4.4, the rHbT and sO2 maps and PAT spectral features are compared for a malignant (a-d) and a benign (e-h) ovarian mass. Multi-pixel method is used before linear unmixing. In these figures, the lesion regions are indicated by the dashed rectangles in the overlaid rHbT (color-scale image) and background US (gray-scale) images. The malignant ovary has a stronger and more concentrated rHbT map than the benign mass. This difference is also shown in Figure 4.5 (h), where the box plot of the mean rHbT for all the benign/normal masses is compared with that of the malignant cases. A significant difference was not observed between these two groups, in terms of the mean of rHbT maps ($p=0.06$).

The histogram for each sO2 map is shown below in Figure 4.4 (c) and (g). As can be seen, the mean sO2 is lower for the malignant group. Also, the histogram of the malignant ovary is skewed toward the lower values, while the benign ovary has a sO2 histogram skewed towards the higher values. Furthermore, Figure 4.5 (b) to Figure 4.5 (d) verify the expected conclusion that in terms of skewness, mean, and energy, the sO2 histograms (e and f) show a significant difference between the benign/normal and malignant groups. Because the sO2 calculation is very sensitive to the PA signal to noise ratio (SNR), we have excluded four ovaries with a poor SNR in the final analysis. For these ovaries, the maximum signal levels at all wavelengths (elements of matrix g in Eq. (4.1)) in the ROI were around the noise level of our system (60 mV). Therefore, performing linear unmixing on these noisy data led to a non-robust estimation of sO2 in the ROI.

Finally, in Figure 4.4 (d) and (h), the mean spectra of the PAT beamlines in the ROI along with their fitted lines are presented for a malignant and a benign ovary, respectively. The malignant ovaries show a smaller value of SS (more negative) and a larger value of SI (less negative). The box plots in Figure 4.5 (e) to (g) verify these observations. As demonstrated in Ref [50], the lower SS in the malignant ovaries is related to the larger size of absorbers that are present in this type of masses, and the higher SI is associated to the larger size of the absorbers as well as a higher concentration of them in malignant ovaries.

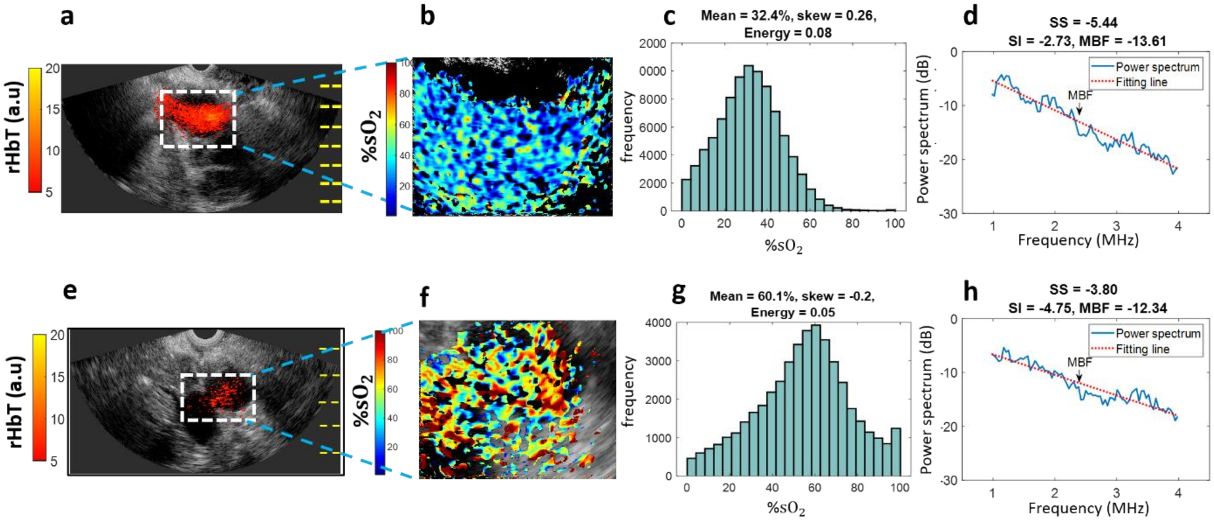


Figure 4.4: Comparison of PAT functional and spectral features of a malignant ovary (a-d) with a benign case (e-h). “a” and “e” are the coregistered US and rHbT maps for the two types of ovarian masses. “b” and “f” show the coregistered US and sO2 map calculated in the ROI indicated by the rectangles in “a” and “e”, respectively. “The histogram of the sO2 maps are shown in “c” and “g”. The mean spectra of the beamlines in the ROIs and their fitted lines are shown in “d” and “h”.

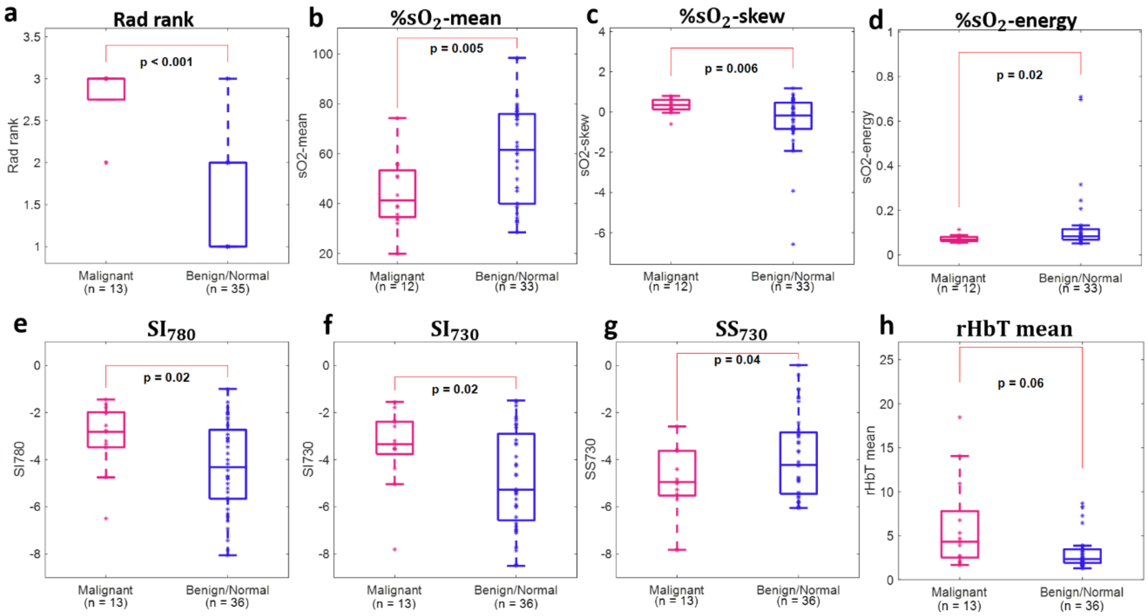


Figure 4.5: Box plots of the most significant features and rHbT. The p-value from a t-test on each feature is shown in the associated plot. The number of samples in each group is also shown below the x-axis of each plot. The sO2 maps were calculated in 10×10 multi-pixels. The three-digit numbers that follow SI or SS in the plots indicate the optical wavelength at which the data was acquired.

4.3.4 Spatial filtering methods

We investigated the effect of the three smoothing filters on the calculated sO2 and residual norm maps. Multi-pixel smoothing used a 10×10 pixel window. The standard deviation of the Gaussian filter and the Lorentzian parameter of the Lorentzian filter were fixed at sigma=5 (see details in Appendix A). An example of the calculated sO2 and normalized residual maps using different smoothing methods is shown in Figure 6 for the malignant ovary in Figure 4. As can be seen, the mean values of the calculated sO2 maps and mean values of the normalized residuals do not change much when different spatial filter methods are used.

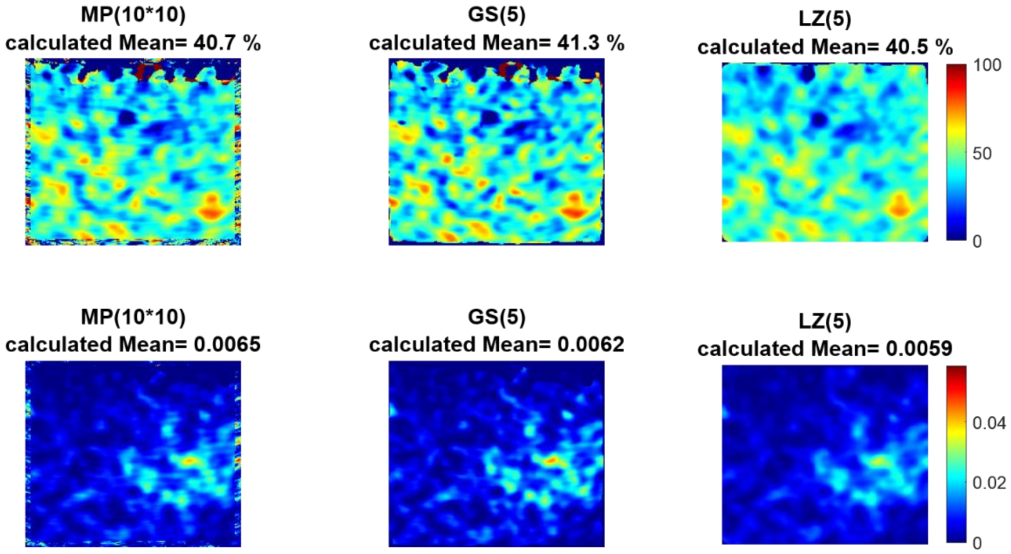


Figure 4.6: The sO2 (a-c) and the corresponding normalized residual (d-f) maps calculated using different smoothing methods for the malignant ovary in Figure 4. The multi-pixel size or sigma for each smoothing method is in the parenthesis in the image title. MP, GS, and LZ in these plots represent multi-pixel, Gaussian, and Lorentzian smoothing methods, respectively.

We calculated the histogram features of the sO2 maps using different smoothing filters (Figure 7). The box plots of the malignant and benign/normal ovaries are located separately in these plots, and the p-values calculated by two-tailed t-tests between the benign/normal and malignant ovaries are shown in each plot. In terms of p-values, the different smoothing methods generate similar results, except for the energy of the sO2 maps. For this parameter, unlike the multi-pixel and Gaussian filters, the Lorentzian filter does not show a statistically significant difference between the benign/normal and malignant groups ($p>0.05$). Moreover, the box plots of the benign ovaries in the sO2-skew and sO2-energy plots calculated using Gaussian method are more spread out than the associated boxplots for the features calculated using the multi-pixel method. Therefore, for our data, the multi-pixel method is a better option than using the other two spatial filters, and it is used in data analysis throughout this manuscript.

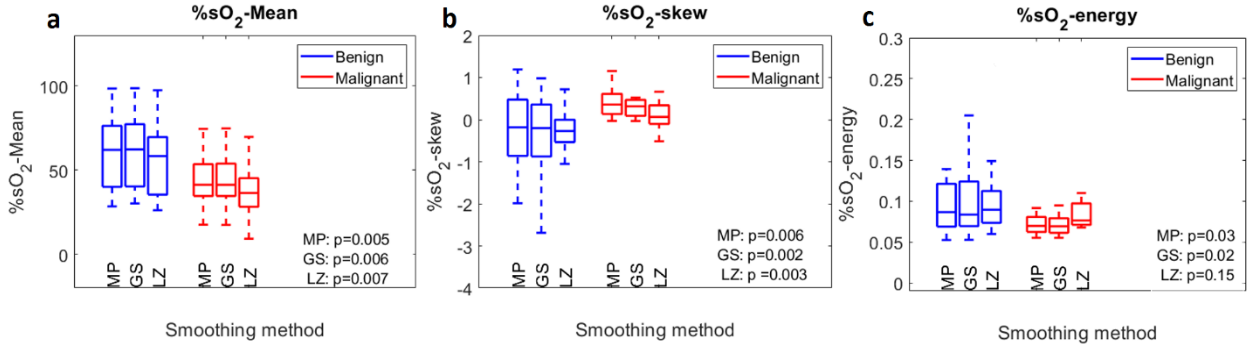


Figure 4.7: Box plots of histogram features calculated by using different smoothing methods. “a” shows the mean of the sO2 maps, and “b” and “c” are the skewness and energy of these maps, respectively. MP, GS, and LZ indicate the multi-pixel, Gaussian, and Lorentzian smoothing methods. The p-values for each method are in the lower right of each plot.

We also calculated the mean residual norm for all the ovarian masses, shown in Figure 4.8. The average and standard deviation of the mean normalized residual maps using each method are shown on the right of each plot. Overall, sO2 maps with a slightly lower normalized residual error are achieved by using the multi-pixel method. Therefore, this method was used for estimation of sO2 maps in this manuscript.

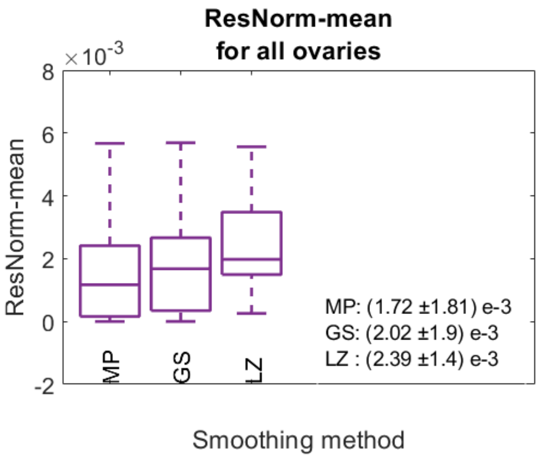


Figure 4.8: Box plots of the means of the normalized residual error maps of all ovaries, calculated using MP, GS, and LZ smoothing. The decimal numbers on the right are the mean ± standard deviation of all the samples in each box.

4.3.5 Features ranking

As mentioned before, the significance of each extracted feature was evaluated using the Student's t-test. These features were then ranked in ascending order, meaning that features with lower p-values (shorter bar length in Figure 4.9 (a)) were located higher in the ranking. Figure 4.9 (a) shows the first eight most important features based on the p-value ranking. These eight features were ranked using a random forest model (explained section 2.7) as well and presented in Figure 4.9 (b). In this plot, unlike the p-value ranking, the longer bars represent more important features.

Note that Rad rank is the most important feature in both rankings. Moreover, although rHbT is located on the bottom of the ranking in Figure 4.9 (a), the random forest model has ranked it as the second most important feature. Another interesting point about these plots is that three of the first four features in Figure 4.9 (a) were highly correlated (sO2-mean, sO2-skew, sO2 -energy), while in the random forest ranking (Figure 4.9 (b)), no correlation was found between each pair of the first 4 features (The correlation between each two features was evaluated using Spearman's method). This is an important consideration when developing classifiers as highly correlated features lead to overfitting our model. A second point to keep in mind when designing classifiers to decrease the chance of overfitting is that the number of features that is used for training should be less than one-tenth of the training size [51]. The training size in this study was $(\frac{2}{3} \times 49 \approx 33)$, so we did not use more than four features. As the first four features in the random forest ranking are not correlated, designing the classifiers based on this ranking seems to reduce the chance of overfitting.

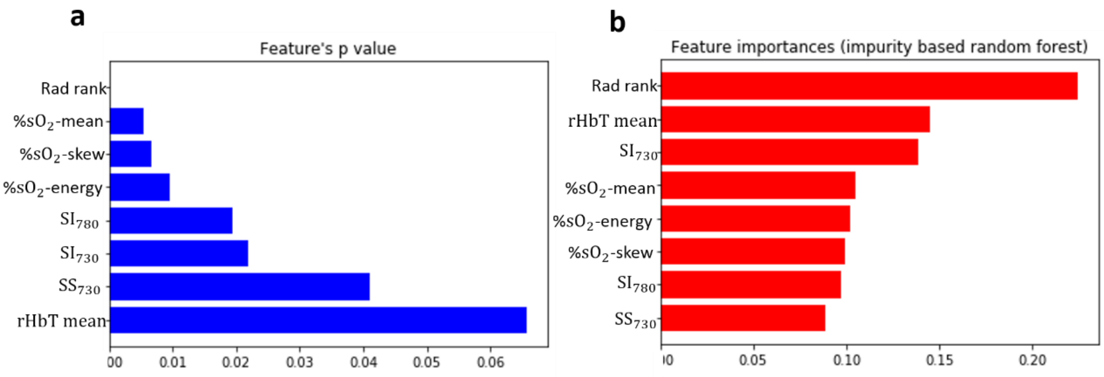


Figure 4.9: Feature ranking based on (a) p-value. (b) random forest importance. In (a), more significant features have shorter bar length and located higher in the ranking. In (b), more significant features have longer bars and located higher in the ranking.

4.3.6 Classification

We developed SVM classifiers to distinguish between "malignant" and "benign/normal" groups. In this case, the data set was randomly divided into 2 groups. The first group included two-thirds of the data used for training, and the rest of the data was employed for testing the classifiers. To lower the chance of overfitting, this process was repeated 100 times.

Four SVM models were trained in this study. The first model was trained using the single radiology feature (Rad rank). The second one was trained using 3 most important PAT features based on the random forest ranking (rHbT, sO2 mean, and SI 730). The third model was trained using the combination of the features used in the last two models. Finally, in the last model, rHbT was removed from the features set. For each model, the 95% confidence of interval (CI) was computed. Moreover, the mean ROC of each of the training and testing data sets for these four models were calculated and are shown in Figure 4.10. In the legend of each plot, the left-hand side of each equation represents the features set employed to train the classifiers and the decimal numbers on the right-hand side are the mean AUC values and the 95% CI associated with the ROC of that classifier. As can be seen on the plots, when just radiology ranking was used to train the SVM classifier, a mean AUC value of 0.85 (95% CI: 0.81 - 0.89) was achieved. On the other hand, the performance of the classifier was not impressive when just PAT features were used to train the classifiers (mean testing AUC = 0.77 (95%CI: 0.73– 0.81). By combining the radiology and PAT features, a superior AUC value of 0.92 (95%CI: 0.89 – 0.95) on the testing data set was achieved. However, when rHbT was removed from the features set, the performance of the SVM classifier on the testing data set improved to 0.93 (95% CI: 0.90 – 0.96).

Two of the thirteen malignant ovaries imaged in this study were serous borderlines, and another one was sertoli-Leydig cell tumor. These types of malignancies are considered low-grade cancers. The estimated sO2 maps for these masses were similar to those for high-grade cancers. However, their rHbT maps and spectral features were close to the ovaries in the benign group. Without these outliers, performing the t-test on rHbT resulted in a p-value of 0.03. However, including these outliers, p-value between all malignant lesions and the benign group is marginal significant ($p=0.06$). As these ovaries have some features similar to the malignant group and some others close to the benign group, performing a multi-class classification to divide the ovaries into 3 groups of high-grade cancers, low grade cancers, and benign ovaries might be a better choice. However, as we currently have a very limited number of the low-grade cancers ($n=3$), it is not feasible to train such a model at this time.

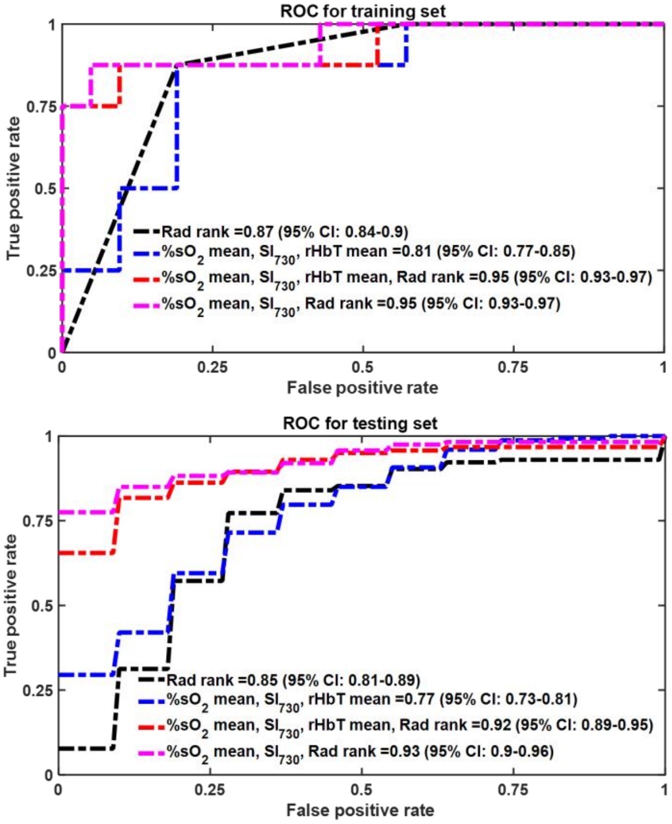


Figure 4.10: The mean ROCs, AUCs, and 95% CI of the four SVM models developed to classify normal/benign ovaries and malignant ovaries for training (top) and testing (bottom) data sets.

To evaluate the contribution of other significant sO2 features (skewness and energy), we combined these features with other types of extracted features and constructed new models. Each model was developed once by including rHbT in, and a second time, by excluding this feature from the features set. The performances of different classifiers are summarized in Table 4.2. By comparing the testing AUC values, we concluded that rHbT did not improve the performance of any of the constructed models which could be due to the inclusion of the three low grade tumors in the cancer group.

Table 4.2: AUC of models constructed using different sO2 features, with and without rHbT

Features (predictors)	**Training $\overline{AUC}$ (0.95% CI)**	**Testing $\overline{AUC}$ (0.95% CI)**
sO2 mean, SI_{730}, Rad rank	0.95 (0.93 – 0.97)	0.93 (0.90 – 0.96)
sO2 mean, SI_{730}, Rad rank, rHbT	0.95 (0.93 – 0.97)	0.92 (0.89 – 0.95)
sO2 skew, SI_{730}, Rad rank	0.95 (0.93 – 0.97)	0.92 (0.89 – 0.95)
sO2 skew, SI_{730}, Rad rank, rHbT	0.95(0.92 – 0.97)	0.92 (0.89 – 0.95)
sO2 energy, SI_{730}, Rad rank	0.93 (0.90 – 0.96)	0.92 (0.89 – 0.95)
sO2 energy, SI_{730}, Rad rank, rHbT	0.93 (0.91 – 0.96)	0.90(0.87 – 0.93)

4.4 Discussion and Summary

In this work, we used the linear unmixing method to calculate sO2 values from phantom and clinical PAT data. In our earlier publication [35], the sO2 maps were calculated by solving linear equations independently for every single pixel. A spatial filter was then applied to smooth the sO2 maps. Here, before the linear equations are solved, the value of the envelope data at each pixel is replaced with the average value of the envelope data for 100 surrounding pixels, which makes the results more robust to measurement errors. Additionally, whereas our earlier study computed only the mean sO2 from each sO2 map, here we calculated six histogram features of the sO2 maps (mean, skewness, energy, variance, kurtosis, entropy), and the first three features were statistically significant. Furthermore, our earlier study ranked features based on their p-values estimated from t-tests. Here, a random forest model has been utilized for ranking the features, reducing the chance of overfitting. Finally, the radiology score has been used in this study, which is a significant step toward the clinical translation of the technology.

We demonstrated the accuracy of estimating the sO2 for different calibrated blood tube phantoms embedded inside a homogeneous Intralipid medium**.** Whole blood consists of a substantial number of impurities, including enzymes that cause blood coagulation. Therefore, we followed the standard process of centrifuging whole blood in order to collect red blood cells (RBCs). The RBCs were diluted with saline water and placed inside a hypoxia chamber, where oxygen and nitrogen concentrations were controlled to produce the desired sO2 condition. A 60% saline solution was used to dilute the RBCs and avoid saturation of the oximeter used to calibrate each blood sample. Since the goal was to estimate sO2 of the blood sample using PAT, diluting the RBCs with 60% saline provided strong PA signals without affecting the results.

We also calculated sO2 using clinical data. The calculated sO2 and rHbT maps, along with SI and Rad rank, were then used to construct different SVM models. The highest AUC value on the testing data set was found for the model constructed from sO2 mean, SI730, and Rad rank. In our earlier publication [35], we developed two classifiers, models A and B. Model A distinguished “benign/normal” masses from those with “epithelial ovarian cancers”. For the testing data, the best AUC of 0.94 was achieved using an SVM classifier. Model B distinguished “benign/normal” lesions from those with “epithelial ovarian cancers and other types of tumors”. With this model, we achieved a best testing AUC of 0.93. In the present manuscript with a larger sample size, we have developed one classifier, model C, which differentiates “benign/normal” masses from "epithelial cancers and other types of tumors". With the addition of the radiology scores, an AUC of 0.93, closely similar to that of model B, was also achieved. In practice, model C has more translational potential because it integrates conventional imaging with PAT functional imaging features.

Depleted oxygen and glucose in the microenvironment of ovarian cancer limits the potential for cellular metabolic plasticity, which has been appreciated as playing a key role in cancer progression and chemoresistance [52]. In a recent study [53], Qiu et al. reported the first evidence that levels of a natural antisense transcript of hypoxia-inducible factor 1 (aHIF) were increased in epithelial ovarian cancer (EOC) tissues and were upregulated by hypoxia in EOC cells. Functional data revealed that aHIF knockdown accelerated cell apoptosis under hypoxia and inhibited EOC tumorigenesis and tumor growth in-vivo. Photoacoustic tomography provides blood sO2, which is related to the tumor oxygen microenvironment and may have a greater role in ovarian cancer diagnosis and assessing chemotherapy response.

Our ovarian mass reader evaluation is a single institution experience, and the ovarian mass ranking was performed by two highly experienced radiologists with over 55 years of combined expertise in pelvic ultrasound. Their experience and breadth in pelvic imaging does not reflect the wider radiology community and will not be the case as PAT gets integrated in the clinical imaging armamentarium. In fact, we anticipate that the addition of PAT to ovarian mass imaging may bring the confidence and accuracy of a general radiologist into the realm of a specialist.

Currently, there are limitations of the photoacoustic technique. First, the typical penetration depth of PAT is about 5 cm. In cystic ovarian tissue, however, the PAT penetration depth may reach 6 to 7 cm. We were able to image more than 95% of the ovaries at these imaging depths. A second limitation is the assumption of wavelength-independent fluence in the linear unmixing method.

Quantitative sO2 calculation without this assumption has been a hot topic in the field of photoacoustic tomography in recent years, and several methods, including deep learning models and model-based algorithms, have been proposed [54-59]. Nevertheless, although interesting results have been observed for simulation and phantom data, applying these methods to clinical data remains a challenge.

In summary, our study with a considerable patient population has shown for the first time that sO2 and its distribution can play a significant role in ovarian cancer diagnosis. Combined with the diagnosis of radiologists, the overall accuracy of identifying both high grade epithelial ovarian cancer and low-grade ovarian tumor can achieve an AUC of 0.93. A large clinical study is currently underway to validate these initial findings.

References

1. American Cancer Society. Cancer facts and figures, 2019, American Cancer Society, Atlanta, GA (2019).
2. J. T. Henderson et al., “Screening for ovarian cancer: updated evidence report and systematic review for the US preventive services task force,” JAMA. 13;319(6):595-606. doi: 10.1001/jama.2017.21421. Review, (2018).
3. S. M. Domchek, et al., “Mortality after bilateral salpingo-oophorectomy in BRCA1 and BRCA2 mutation carriers: a prospective cohort study,” Lancet Oncol;7:223–229, (2006)
4. A. Finch et al., “Salpingo-oophorectomy and the risk of ovarian, fallopian tube, and peritoneal cancers in women with a BRCA1 or BRCA2 Mutation,” JAMA;296:185–192, (2006).

5. J. S. Kwon et al., “Prophylactic salpingectomy and delayed oophorectomy as an alternative for BRCA mutation carriers,” Obstet Gynecol;121:14–24, (2013).
6. L. T. Shuster et al., “Prophylactic oophorectomy in premenopausal women and long-term health,” Menopause Int;14:111–116, (2008).
7. L. Maehle et al., “High risk for ovarian cancer in a prospective series is restricted to BRCA1/2 mutation carriers,”, Clin Cancer Res.;14(22):7569-73, (2008).
8. K. B. Mathieu et al., “Screening for ovarian cancer: imaging challenges and opportunities for improvement,” Ultrasound Obstet Gynecol.;51(3):293-303. doi: 10.1002/uog.17557, (2018).
9. H. Addley et al., “Diffusion-weighted imaging in gynaecological malignancy,” Clin Radiol.;72(11):981-990. doi: 10.1016/j.crad.2017.07.014, (2017).
10. V. R. Iyer & S. I. Lee, “MRI, CT, and PET/CT for ovarian cancer detection and adnexal lesion characterization,” AJR;194(2):311-321,(2010).
11. P. Mapelli et al., “Imaging biomarkers in ovarian cancer: the role of ^{18}F-FDG PET/CT,” Q J Nucl Med Mol Imaging.;60(2):93-102, (2016).
12. R. Kumar et al., “Positron emission tomography in gynecological malignancies,” Expert Review of Anticancer Therapy;6(7):1033-1044, (2006).
13. A. A. Oraevsky & A. A. Karabutov, “Optoacoustic tomography,” In: T. Vo-Dinh, ed., Biomedical photonics handbook. Boca Raton, Fla: CRC, (2003).
14. R. Bouchard et al., “Ultrasound-guided photoacoustic imaging: current state and future development,” IEEE Trans Ultrason Ferroelectr Freq Control.;61(3):450-66, (2014).
15. K. S. Valluru & J. K. Willmann, “Clinical photoacoustic imaging of cancer,” Ultrasonography;35(4): 267-280, (2016).

16. S. Zackrisson et al., “Light in and sound out: emerging translational strategies for photoacoustic imaging,” Cancer Res.;74(4):979-1004, (2014).
17. R. Su et al., “Laser OptoAcoustic Tomography: Towards new technology for biomedical diagnostics,” Nucl Instrum Methods Phys Res A. Author manuscript; available in PMC. Published in final edited form as: Nucl Instrum Methods Phys Res A. 2013 Aug 21;720: 58–61, (2014).
18. G. Yang et al., “Optimized light delivery probe using ball lenses for co-registered photoacoustic and ultrasound endo-cavity subsurface imaging,” Photoacoustics; 13:66-75, (2019).
19. H. Luo et al., “Fiber endface illumination diffuser for endo-cavity photoacoustic imaging,”, Optics Letters; 45(3):632-635, (2020).
20. N. Weidner et al. “Tumor angiogenesis and metastasis – correlation in invasive breast carcinoma,” N Engl J Med.;324(1):1-8, (1991).
21. P. Vaupel et al., “Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review,” Cancer Research; 49:6449-6465, (1989).
22. R. A. Kruger et al., “Dedicated 3D photoacoustic breast imaging,” Med Phys.;40(11):113301, (2013).
23. M. Heijblom et al., “The state of the art in breast imaging using the Twente Photoacoustic Mammoscope: results from 31 measurements on malignancies,” Eur Radiol;26(11):3874–3887, (2016).
24. J. Menke et al., “Photoacoustic breast tomography prototypes with reported human applications,” Eur Radiol.;25(8):2205-13. Review, (2015).

25. T. T. W. Wong et al., “Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy,” Sci Adv.;3(5), (2017).
26. A. Garcia-Uribe et al., “Dual-modality photoacoustic and ultrasound imaging system for noninvasive sentinel lymph node detection in patients with breast cancer,” Sci Rep.;5:15748, (2015).
27. R. A. Kruger et al., “Photoacoustic angiography of the breast,” Med Phys.;37(11):6096-100, (2010).
28. A. Becker et al., “Multispectral optoacoustic tomography of the human breast: characterisation of healthy tissue and malignant lesions using a hybrid ultrasound-optoacoustic approach,” Eur Radiol.;28(2):602-609, (2018).
29. V. S. Dogra et al. “Multispectral photoacoustic imaging of prostate cancer: preliminary ex-vivo results,” J Clin Imaging Sci.;3:41, (2013).
30. X. Wang et al., “Photoacoustic tomography: a potential new tool for prostate cancer,” Biomed Opt Express;1(4):1117–1126, (2010).
31. Y. Zhou et al., “Noninvasive determination of melanoma depth using a handheld photoacoustic probe,” J Invest Dermatol.;137(6):1370-1372, (2017).
32. M. Yang et al., “Photoacoustic/ultrasound dual imaging of human thyroid cancers: an initial clinical study,” Biomed Opt Express.;8(7):3449-3457, (2017).
33. G. Yang et al., “Co-registered photoacoustic and ultrasound imaging of human colorectal cancer,” J Biomed Opt.;24(12):1-13, (2019).
34. S. Nandy et al. “Evaluation of ovarian cancer: initial application of coregistered photoacoustic tomography and US,” Radiology.;289(3):740-747, (2018).

35. E. Amidi et al., “Classification of human ovarian cancer using functional, spectral, and imaging features obtained from in vivo photoacoustic imaging,” Biomed Opt Express.;10(5): 2303-2317, (2019).

36. M. Li et al., “Simultaneous molecular and hypoxia imaging of brain tumors in vivo using spectroscopic photoacoustic tomography,” Proceedings of the IEEE;96(3):481-489 (2008).

37. X. Wang et al., “Noninvasive imaging of hemoglobin concentration and oxygenation in the rat brain using high-resolution photoacoustic tomography,” J Biomed Opt.;11(2):024015 (2006).

38. A. Needles et al., “Development and initial application of a fully integrated photoacoustic micro-ultrasound system,” IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control; 60(5):888-897, (2013).

39. L. V. Wang et al., “High-resolution spectroscopic photoacoustic tomography for noninvasive functional imaging of small-animal brains in vivo,” The Second Asian and Pacific Rim Symposium on Biophotonics,. APBP 2004.,246-247, (2004).

40. S. N. Hennen et al., “Photoacoustic tomography imaging and estimation of oxygen saturation of hemoglobin in ocular tissue of rabbits,” Exp Eye Res.;138:153-8, (2015).

41. R. Hochuli et al., “Estimating blood oxygenation from photoacoustic images: can a simple linear spectroscopic inversion ever work?,” J. of Biomedical Optics;24(12),121914, (2019).

42. J. Lavaud et al., “Noninvasive monitoring of liver metastasis development via combined multispectral photoacoustic imaging and fluorescence diffuse optical tomography,” Int J Biol Sci.;16(9):1616-1628, (2020).

43. X. Li et al., “Multispectral interlaced sparse sampling photoacoustic tomography,” IEEE Transactions on Medical Imaging; doi: 10.1109/TMI.2020.2996240, (2020).
44. J. Kim et al., “Multispectral photoacoustic assessment of thyroid cancer nodules in vivo,” Photons Plus Ultrasound: Imaging and Sensing; 1124004, (2020).
45. D. A. Durairaj et al., “Unsupervised deep learning approach for photoacoustic spectral unmixing,” Photons Plus Ultrasound: Imaging and Sensing; 112403H, (2020).
46. T. Feng et al., “Characterization of bone microstructure using photoacoustic spectrum analysis,” Optics express;23(19):25217-24, (2020).
47. X. Leng et al., “Feasibility of co-registered ultrasound and acoustic-resolution photoacoustic imaging of human colorectal cancer,” Biomed Opt Express.;9(11):5159-5172, (2018).
48. E. Hysi et al., “Photoacoustic signal characterization of cancer treatment response: Correlation with changes in tumor oxygenation,” Photoacoustics;2017(5):25-35, (2017).
49. B. H. Menze et al., “A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data,” BMC Bioinformatics. 10(213). doi: https://doi.org/10.1186/1471-2105-10-213, (2009).
50. G. Xu et al, “Photoacoustic spectrum analysis for microstructure characterization in biological tissue: A feasibility study,” Appl Phys Lett.;101(22):221102, (2012).
51. F. Harrell, “The PHGLM procedure,” In: SUGI Supple- mental Library User’s Guide, SAS Institute, Cary, North Carolina; 267-294, (1983).
52. E. Emmings et al., “Targeting mitochondria for treatment of chemoresistant ovarian cancer,” Int J Mol Sci.;20(1):229. doi: 10.3390/ijms20010229, (2019).

53. J. J. Qiu et al., “Natural antisense transcript of hypoxia-inducible factor 1 regulates hypoxic cell apoptosis inepithelial ovarian cancer,” Onco Targets Ther.;14(11):9101-9110. doi: 10.2147/OTT.S173816, (2018).
54. B. T. Cox et al., “The challenges for quantitative photoacoustic imaging,” Proc. SPIE 7177, Photons Plus Ultrasound: Imaging and Sensing;717713, (2009).
55. A. Pulkkinen et al., “A Bayesian approach to spectral quantitative photoacoustic tomography,” Inverse Problems;30(6), (2014).
56. T. Kirchner et al., “Context encoding enables machine learning-based quantitative photoacoustics,” J. Biomed. Opt.;23(5):056008, (2018).
57. M. Li et al., “Photoacoustic tomography of blood oxygenation: A mini review,” Photoacoustics.;10:65-73. doi: 10.1016/j.pacs.2018.05.001, (2018).
58. C. Cai et al., “End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging,” Opt. Lett.;43(12):2752-2755, (2018).
59. C. Bench, “Towards accurate quantitative photoacoustic imaging: learning vascular blood oxygen saturation in 3D,” arXiv:2005.01089 [Preprint]., Available from: https://arxiv.org/abs/2005.01089, (2020).

Chapter 5: Quantitative photoacoustic tomography using two-step optimization method for estimation of properties of blood samples

5.1 Introduction

Photoacoustic tomography (PAT) is based on the generation of ultrasound waves by a short pulse laser. This modality can simultaneously provide decent imaging contrast as well as resolution [1-5]. It also has the potential of obtaining functional information about the tissue. However, accurate estimation of the functional parameters using photoacoustic imaging is challenging. This is because of the fact that the detected signals by the transducers are not just a function of the absorption coefficient of the target, but rather a multiplication of this parameter with some other tissue and system related parameters.

Recently, our group has demonstrated the potential of photoacoustic tomography in ovarian and colon cancer diagnosis [6-10]. In those studies, the rHbT was calculated using the PAT beam data at four wavelengths. To calculate this parameter, the value of optical fluence on the surface of the tissue was employed as this value was not available at the other locations in the tissue. Clearly, this method of approximating the total hemoglobin is crude and cannot be considered as being related to just the chromophores in the target area. In fact, it is also related to the optical properties of the background tissue.

One of the barriers in quantitate evaluation of the tissue using photoacoustic tomography is related to the optical fluence estimation in the tissue [11-14]. One of the most known methods to estimate light fluence in the tissue is using radiative transfer equation (RTE). This is an integro-differential equation, relating the light radiance to the source and the optical properties of the medium and may be solved using the finite element method (FEM) [15]. When the reduced optical scattering coefficient of the medium is much higher than its optical absorption coefficient, RTE can be approximated as the diffusion equation (DE). If the medium is heterogenous, the DE equation can be solved using a numerical method, such as FEM [16, 17]. In the case of homogenous medium, a simpler way using green function exists to solve DE. For 2-D case, the fluence can be found using Beer's law: $\Phi(z) = \Phi_0 \exp\left(-\mu_{eff} r\right)$, where Φ_0 is the fluence on the surface, μ_{eff} is the effective attenuation coefficient and r is the distance from the surface of the tissue. Besides these analytical methods, Monte-Carlo, which is a numerical method, has also been used by some researchers to deal with the fluence estimation problem in the medium.

In this study, we have started with the simplest case of using Beer's law to estimate the fluence in the target area. We have verified this method in the case when blood tube phantoms are located at different depths in an intralipid medium (with optical properties like human tissue). Then, the optical absorption coefficients of the blood samples have been found by solving an unmixing problem using a conjugate-gradient method. These coefficients have been used to estimate the oxy and deoxyhemoglobin concentrations in the tissue. These concentration values are then employed to calculate the total hemoglobin (HbT) and sO2 of each blood sample.

5.2 Methods and Materials

Experiment: We have conducted five sets of blood tube experiments with calibrated blood sO2 values of 24.9%, 44.2%, 64.9%, 83.9%, and 97.6% using the co-registered PAT/US system. Blood from healthy volunteers was processed to collect red blood cells (RBCs) using a protocol of Dr.

Allan Doctor, WashU Med School. RBCs diluted by adding 60% saline water, referred as “blood” below, were then prepared inside a hypoxia chamber with temperature/humidity control and a “glove box” for handling samples in a controlled gas environment (Figure 5.1). Oxygen and nitrogen gas were controlled to produce the desired sO2 conditions inside the chamber, as measured by a Radiometer ABL90 Series.

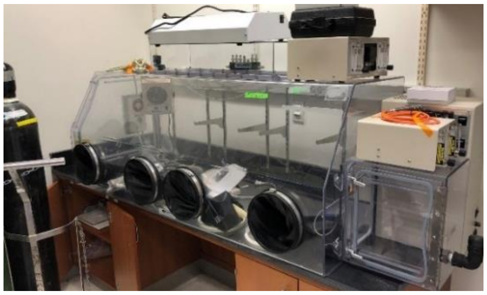


Figure 5.1: The chamber used to control the sO2 values of the blood samples

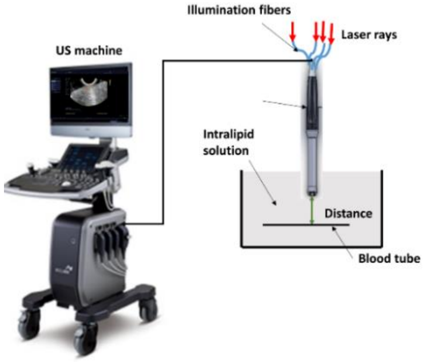


Figure 5.2. The PAT/US set up used to acquire the PAT signals

Each 3-mm tube was filled with blood of a certain sO2 and tightly capped inside the chamber. PAT/US experiment was performed immediately after each tube was prepared. sO2 was measured again with the Radiometer and sO2 change after each PAT/US experiment was less than 2% of the initially calibrated value. During PAT/US experiment, each tube was imbedded in an intralipid scattering medium (μ_a =0.02-0.03 cm^{-1} and μ_s'=4-5 cm^{-1}, typical soft tissue properties) and imaged at depths from 1 cm to 4 cm with 0.5cm spacing controlled by a stage (Figure 5.2). For each sO2 image obtained at each depth, pixel values greater than zero were averaged to obtain the average sO2 value.

Data analysis: The initial photoacoustic pressure P_0 (λ, r) at any point r =(x,y,z) within the tissue is proportional to the wavelength-dependent absorption distribution, μ_a, the fluence Ø (energy/unit area) distribution, and the Grüneisen parameter, Γ, and can be written as

$$P_0\ (\lambda,r)=\Gamma Ø(\lambda,r)\mu_a\ (\lambda,r). \qquad \text{Eq. (5.1)}$$

Also, the reconstructed pressure is related to the initial pressure as

$$P(\lambda,r)=C_0(r)P_0(\lambda,r), \qquad \text{Eq. (5.2)}$$

where C_0(r) is the acoustic operator [18]. Due to the very small change in wavelength-dependent tissue scattering, as well as the constant absorption of water in the narrow wavelength window of 730 nm to 830 nm, the system response and fluence distribution can be approximated as wavelength-independent. Thus

$$P\ (\lambda,r) = \overline{C(r)}\mu_a(\lambda,r), \text{ where } \overline{C(r)}=\Gamma C_0(r)Ø(r). \qquad \text{Eq. (5.3)}$$

Considering oxygenated and deoxygenated hemoglobin, HbO and Hb, to be the only dominant absorbing chromophores in this wavelength range, μ_a at each wavelength λ_i can be written as

$$\mu_a(\lambda_i,r)= \varepsilon_{HBO}(\lambda_i\ )HbO(r)+\varepsilon HB(\lambda_i\ )Hb(r), \qquad \text{Eq. (5.4)}$$

where ε_{HbO} and ε_{Hb} are the wavelength-dependent molar extinction coefficients of HbO and Hb. From equations (3) and (4), we have

$$P(\lambda_i, r) = \varepsilon_{HBO}(\lambda_i)\,\overline{C(r)}HbO(r) + \varepsilon_{HB}(\lambda_i)\overline{C(r)}Hb(r). \qquad \text{Eq. (5.5)}$$

The relative total hemoglobin concentration is given by

$$rHbT(r) = \overline{C(r)}[HbO(r) + Hb(r)]. \qquad \text{Eq. (5.6)}$$

rHbT can be computed from multi-wavelength $P(\lambda_i,r)$ data. To assess quantitative HbT(r), we need to estimate $\overline{C(r)}$. Furthermore, estimating the local fluence, Ø(r), is challenging because Ø(r) is also a function of the tissue's μ_a and μ_s'. Another challenge is that the US transducer receives PAT signals only within a limited field of view. There are several approaches to estimate the fluence and then quantitatively compute the distribution from $P(\lambda,r)$ measurements [18]. We have employed a revised layered-tissue approach similar to the 1-D method reported in Ref. 19 which approximates the light fluence at depth $\rho = \sqrt{z^2 + x^2}$ as

$$\varnothing_0 exp^{-\mu_{eff}\times\rho} \text{, where } \mu_{eff=}\sqrt{3\,\mu_a\left(\mu_a + \mu'_s\right)} \qquad \text{Eq. (5.7)}$$

and $\varnothing_0$ is the fluence on the tissue surface. For our particular application, with the vagina muscle wall between the examined ovary and the PAT/US probe and under wide-field light illumination, this simple 2-D model is more suitable. We formulate the quantitative reconstruction of $\mu_a(\lambda,\rho,\Theta)$ as an optimization problem:

$$f(x) = \underset{\hat{C}(\rho,\Theta),\mu_a(\lambda,\rho,\Theta)}{\arg\min}\left(\left\|P(\lambda_i,\rho,\Theta) - \hat{C}(\rho,\Theta)\times \exp^{-\mu_{eff}\times\rho}\mu_a(\lambda_i,\rho,\Theta)\right\|^2\right) \qquad \text{Eq. (5.8)}$$

where $\hat{C}(\rho,\theta) = \Gamma C_0(\rho,\theta) \times \varnothing_0$ and Θ is the angle from the central field of view of the US transducer. We first fit the PAT beam lines along the depth ρ and Θ to estimate $\mu_{eff}(\lambda, \Theta)$ and then use the known μ_s' to estimate $\mu_a(\lambda, \theta)$ within the US identified angular section of the ovary. μ_a

(λ, θ) is used as an initial guess of μ_a (λ,ρ,Θ) to estimate initial guess of $\hat{C}(\rho,\theta)$. From the estimated initial $\hat{C}(\rho,\theta)$, we iteratively reconstruct $\hat{C}(\rho,\theta)$ and μ_a (λ,ρ,Θ). With these two initial estimates of $\mu_a(\lambda, \theta)$ and $\hat{C}(\rho,\theta)$, the optimization converges.

However, this method requires that the tissue μ'_s value is known and its change from background is small. From our initial experience on blood tube phantoms, the reconstructed $\mu_a(\lambda,\rho,\Theta)$ is not very sensitive to μ'_s, thus one wavelength measurement of μ'_s should offer a good estimation.

5.3 Results

We have demonstrated this approach in blood tube experiments. The optical absorptions of the five tested blood tubes at different wavelengths are shown in Figure 5.3. By comparing the plots in this figure to the molar extinction coefficient of oxy and deoxy hemoglobin shown in Figure 5.4, it is clear that the absorption coefficient spectrum follows the molar extinction coefficient spectrum of deoxyhemoglobin for lower sO2 percentages and it becomes closer to the spectrum of oxyhemoglobin when the sO2 percentage increases.

Figure 5.5 shows a scatter plot of quantitative HbTs in uM (a) and a scatter plot of the calculated sO2 values (b) calculated from μ_a (λ,ρ,Θ) obtained from depths of 1 cm to 4 cm for each of five sets of blood tube experiments described before. The calibrated HbT shown in black circle was measured from a Microplate Spectrometer (BIoTek) at Dr. Doctor's lab and a whole blood $\mu'_s=10\text{cm}^{-1}$ was used as an estimate for computing $\mu_a(\lambda, \theta)$. As shown in Figure 5.5 a, HbT is quite accurate for the sO2 values in the middle but has slight larger errors at the two ends of the sO2 values of 24.9% and 97.6%. This is related to the relatively lower laser power at 730nm and

830 nm of our system, and we will improve the system by replacing the pump laser which is quite old.

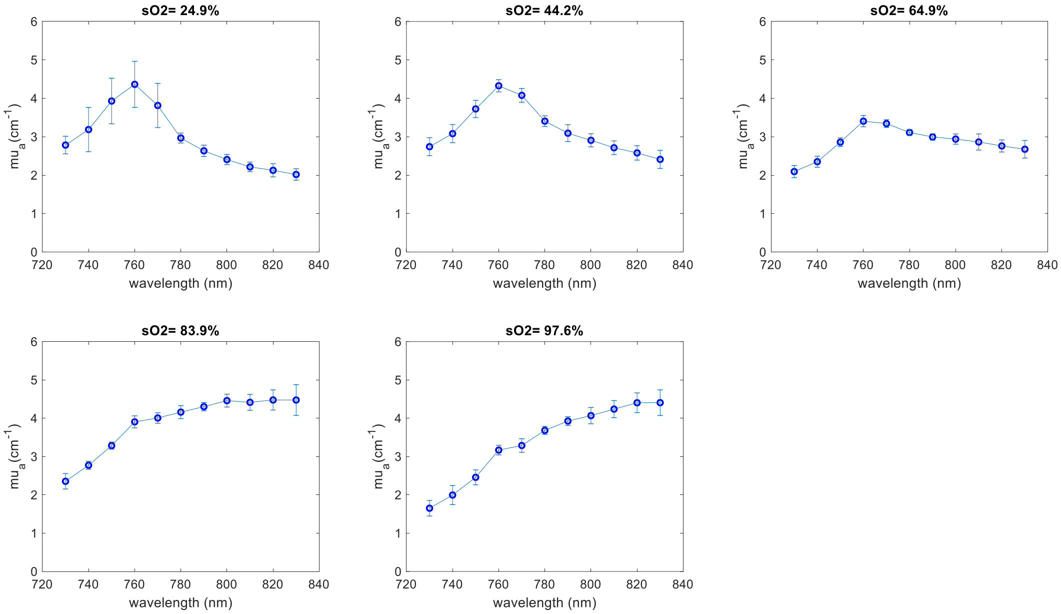


Figure 5.3: The calculated optical absorption for blood samples with different oxygen concentrations.

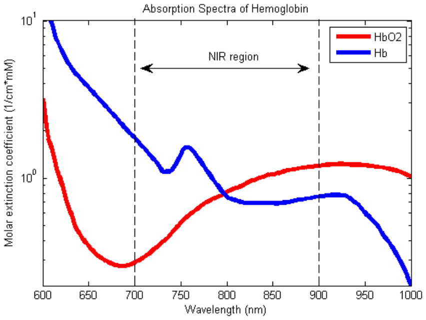
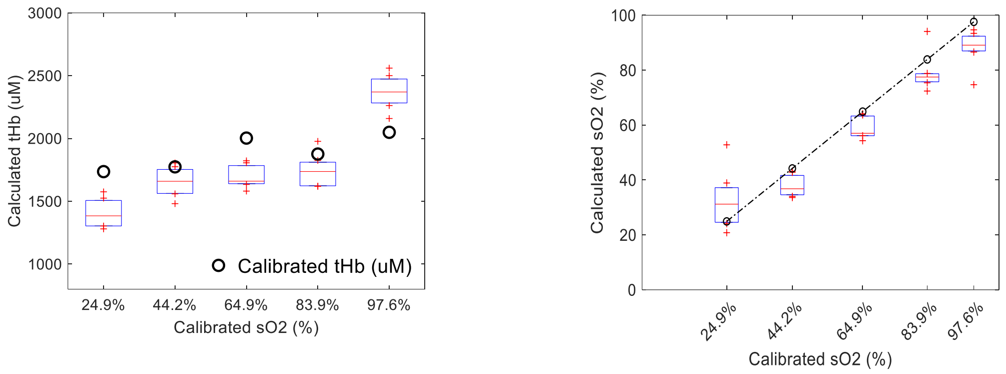


Figure 5.4: the molar extinction coefficient of oxy and deoxyhemoglobin.



a

b

Figure 5.5: a: The calculated vs the actual total hemoglobin values of the tested blood samples. b: The calculated vs the actual oxygen saturation percentages of the tested blood samples.

About the sO2 scatter plot (Figure 5.5 b), the standard deviations (STD) for each calculated sO2 at all depths were 10.8%, 3.8%, 4.1%, 6.9% and 6.5%, respectively with mean STD 6.4%. The sO2 curve from PAT data showed a typical bell shape, which slightly reduces the calculated sO2 dynamitic range by providing a higher sO2 estimate at extremely low blood sO2 levels and a lower sO2 estimate at extremely high sO2 levels. However, this dynamic range reduction affects the calculated sO2 of both malignant and benign/normal ovarian tissues equally, they are all distributed in a similar depth range of different sO2 values. To sum up, based on the results in Figure 5.5, the mean error of the estimated HbT values for blood tubes at all depths with respect to the real values are less than 13%. Furthermore, the largest sO2 estimation error is 7.5% for the blood sample with sO2 of 24. 9%

In the future, using tissue phantoms, like the one introduced in [20], we will validate this approach for sensitivity to μ'_s, sensitivity to the initial estimate of $\mu_a(\lambda, \theta)$and $\hat{C}(\rho, \theta)$, sensitivity to lower HbT concentrations, and robustness of the approach to different types of ovarian tissue (solid vs.

cystic). More ex-vivo measurements of average ovarian tissue μ_a and μ'_s will be made at 730nm, 780nm, 800 and 830nm using our DOT system [21]. The improvement in diagnostic accuracy for malignant vs. benign/normal ovaries will be evaluated using our pilot PAT/US patient data.

Moreover, in the next step, we will try this method on our in-vivo ovarian data. In this case, as the ovarian tissue, unlike intralipid solution, is not homogenous, 2-D model may not be a suitable approximation for the problem and more complicated methods can be used to find the local fluence in the tissue. If the above simple 2-D model is not robust to patient data, we will investigate the full-field 3-D diffusion equation to iteratively compute the $\emptyset(r, \mu_a, \mu'_s)$ distributions [19, 22]. Other computationally intensive inverse methods include non-linear iterative updating of the absorption distribution [23], and optimization using model-based [24] and Monte Carlo based approaches [25-27]. These methods have been demonstrated in simulations and phantom studies, and their robustness in clinical applications will be evaluated.

5.4 Discussion and Summary

In this study, we conducted an experiment using blood tube phantoms to quantitatively evaluate the absorption coefficient and functional features of different blood samples using PAT. The blood tube was located at different depths in an intralipid solution, and PAT signals were acquired at each depth. As the medium was homogenous, the optical fluence at each depth was simply calculated using Beer’s law. To find the absorption coefficient of the blood, an optimization problem with two unknowns was solved using conjugate gradient method. One of the unknown variables in this problem was the multiplication of system response and Grüneisen parameter, and the other one was the absorption coefficient of the blood. Our estimated absorption coefficient, HbT, and sO2 were close to the real values. In the future we will evaluate the performance of this method on the data we have acquired from patients with ovarian masses.

References

1. L. V. Wang et al., “Photoacoustic tomography: in vivo imaging from organelles to organs,” science, 335(6075), 1458-1462 (2012).
2. K. S. Valluru et al., “Photoacoustic imaging in oncology: translational preclinical and early clinical experience,” Radiology, 280(2), 332-349 (2016).
3. G. Yang et al., “Optimized light delivery probe using ball lenses for coregistered photoacoustic and ultrasound endo-cavity subsurface imaging,” Photoacoustics, 13, 66–75 (2019).
4. G. Yang et al., "Optimizing light delivery through ball-shaped multimode fiber tips in co-registered photoacoustic and ultrasound endo-cavity imaging: simulation and experimental validation," Proc. SPIE 10878, Photons Plus Ultrasound: Imaging and Sensing 2019, 108785H (2019);
5. H. Luo et al., “Fiber endface illumination diffuser for endo-cavity photoacoustic imaging,” Optics Letters, 45(3), 632-635 (2020).
6. A. Mostafa et al., "Dual-mode photoacoustic and ultrasound system for real-time in-vivo ovarian cancer imaging," Proc. SPIE 10494, Photons Plus Ultrasound: Imaging and Sensing 2018, 104944K (2018).
7. S. Nandy et al., “Evaluation of ovarian cancer: initial application of coregistered photoacoustic tomography and US,” Radiology, 289(3), 740-747 (2018).
8. E. Amidi et al., "Classification of human ovarian cancer using functional, spectral, and imaging features obtained from in vivo photoacoustic imaging," Biomed. Opt. Express 10, 2303-2317 (2019).

9. G. Yang et al., "Co-registered photoacoustic and ultrasound imaging of human colorectal cancer," J. Biomed. Opt. 24(12) 121913 (2019).
10. G. Yang et al., "Co-registered photoacoustic and ultrasound real-time imaging of colorectal cancer: ex-vivo studies," Proc. SPIE 10878, Photons Plus Ultrasound: Imaging and Sensing 2019, 108784K (2019);
11. A. Ishimaru, “Wave Propagation and Scattering in Random Media,” Vol 1: Single Scattering and Transport Theory, Academic Press (1978).
12. A. J. Welch & M. J. C. van Gemert, eds., “Optical-Thermal Response of Laser-Irradiated Tissue,” Plenum Press (1995).
13. F. Martelli et al., “Light Propagation through Biological Tissue and Other Diffusive Media,” SPIE (2010).
14. L. V. Wang & H.-I. Wu, “Biomedical Optics: Principles and Imaging,” Wiley (2007).
15. T. Tarvainen et al., “Finite element model for the coupled radiative transfer equation and diffusion approximation,” Int. J. Numer. Methods Eng. 65(3), 383–405 (2006).
16. S. Arridge & M. Schweiger, “A finite element approach for modeling photon transport in tissue,” Med. Phys. 20(2), 299–309 (1993).
17. M. Schweiger et al., “The finite element method for the propagation of light in scattering media: boundary and source conditions,” Med. Phys. 22(11), 1779–1792 (1995).
18. B. Cox et al., “Quantitative spectroscopic photoacoustic imaging: a review,” J Biomed Opt. 17(6):061202. doi: 10.1117/1.JBO.17.6.061202. Review, (2012).
19. J. Ripoll & V Ntziachristos., “quantitative point source photoacoustic inversion formulas for scattering and absorbing media,” Phys. Rev. E 71(3),1-9 (2005).

20. E. Amidi et al., "Low-cost ultrasound and optical gelatin-based phantoms," Proc. SPIE 10878, Photons Plus Ultrasound: Imaging and Sensing 2019, 108784A (2019).
21. H. Vavadi et al., "Compact ultrasound-guided diffuse optical tomography system for breast cancer imaging," Journal of biomedical optics 24, no. 2 (2018).
22. R. J. Zemp. “Quantitative photoacoustic tomography with multiple optical sources,”. Appl Opt.;49(18):3566-72 (2010).
23. BT. Cox et al., “Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method,” Appl Opt.;45(8):1866-75 (2006).
24. Z. Yuan & H. Jiang., “A calibration-free, one-step method for quantitative photoacoustic tomography,” Med Phys.; 39(11): 6895–6899 (2012).
25. R. Hochuli et al., “Quantitative photoacoustic tomography using forward and adjoint Monte Carlo models of radiance,” J Biomed Opt;21(12):126004, (2016).
26. Y. Liu et al., “Two schemes for quantitative photoacoustic tomography based on Monte Carlo simulation,” Med Phys. 43(7):3987, (2016).
27. T. Shan et al., “GPU-based acceleration and mesh optimization of finite-element-method-based quantitative photoacoustic tomography: a step towards clinical applications,” Appl Opt.;56(15):4426-4432, (2017)

Chapter 6: Colorectal cancer diagnosis using coregistered photoacoustic tomography and ultrasound system

6.1 Introduction

Photoacoustic imaging (PAI) is an emerging technique which can provide high optical absorption contrast images at reasonable microscale resolution and clinically relevant depths [1]. Several studies have established that optical absorption parameters are important biomarkers directly related to the tissue microvasculature, tumor angiogenesis or tumor hypoxia [2-4]. In general, PAI is classified into photoacoustic microscopy (PAM) and photoacoustic tomography (PAT) [1]. Previously, PAM or Photoacoustic endoscopy (PAE) have demonstrated the capability of detecting human colorectal cancer [5,6]. However, the low imaging speed (limited by the laser repetition rate and scanning scheme), small imaging area, and shallow penetration depth (<7 mm) created obstacles for clinical applications.

Compared with PAM, PAT is able to penetrate deeper with a faster data acquisition speed and a larger field of view due to the use of ultrasonic arrays and a wide optical beam. Several studies have demonstrated that PAT/US dual-modality imaging system can provide anatomical and functional information in tumors [7-13], but no prior applications in the human distal GI tract have been reported using PAT/US dual-modality imaging.

Adenocarcinoma of the colon and rectum is the second most common malignancy diagnosed globally and the fourth leading cause of cancer mortality, with more than 100,000 new cases diagnosed annually in the US [14-15]. Accurate staging and post-treatment surveillance of this prevalent disease are critical because treatment strategies are predicated upon the stage at presentation and response to therapy – in some instances, detailed imaging allows certain patients to avoid surgery altogether. While colonoscopy and biopsy are the gold-standard diagnostic tests for colorectal cancers [16], multiple imaging modalities including optical imaging [17-18], endoscopic ultrasound (EUS), pelvic magnetic resonance imaging (MRI), computed tomography (CT) and positron emission tomography (PET) are also utilized.

Unfortunately, each of these modalities have critical weaknesses when evaluating colorectal tumors. White light endoscopy (WLE) only detects macroscopic morphology and provides no functional assessment of the imaged tissue. MRI has limited between-slice resolution and is often unable to differentiate early tumors from benign neoplasia, committing patients to potentially more invasive treatment regimens than needed [19,20]. Monitoring of tumors after chemotherapy and radiation with MRI is often confounded by fibrotic reaction and edema, which can appear similar to residual tumor [21]. CT has poorer resolution of the bowel wall layers in comparison to MRI, subsequently limiting its ability to describe circumferential resection margin (CRM) status or serosal invasion in locally advanced cases. Additionally, CT also cannot distinguish induration or peritumoral fibrosis from frank malignant disease with a high degree of specificity, further limiting its application in local tumor staging [19]. PET imaging is also plagued by poor resolution, and EUS remains highly user-dependent and unable to resolve small islands of tumor [19]. Therefore, critical need exists for precise imaging modalities of colorectal tumors for both staging and therapeutic response evaluations.

In this study, we have imaged colorectal masses using a real-time co-registered PAT/US system to delineate differences between benign and malignant tissue. To the best of our knowledge, this study is the first utilizing co-registered PAT/US to evaluate human colon samples. We investigated the potential qualitative and quantitative capability of PAT functional and spectral parameters as well as PAT/US images to identify malignant tissue.

6.2 Methods

6.2.1 Human Sample Preparation

Freshly resected colon and rectum samples obtained from patients undergoing surgery at Washington University School of Medicine were imaged immediately after surgery. Patients with known benign neoplasia (polyps) as well as malignancies (adenocarcinoma) were eligible for imaging. Cancer patients who had received preoperative treatment with chemotherapy and /or radiation were also included. The study was approved by the Institutional Review Board (IRB) at Washington University (#201707066). Informed consent was obtained from all patients. Specimens were obtained from the operating room as previously described [6].

A total of 20 tissue samples were imaged in the pilot study using the PAT/US system. This included untreated colorectal adenocarcinomas (n=10), precancerous polyps (n=5), colorectal cancer following chemotherapy or radiation and chemotherapy (n=4), and post polypectomy (n=1). Two treated patients have achieved complete pathological response and two partial response. The majority of patients underwent hemicolectomy for cancer and were found to have malignancy on histologic analysis (Table 6.1).

Table 6.1: Summary of Specimens

Patient ID	Surgery	Pathology
1	Total colectomy	Moderately differentiated adenocarcinoma (T3)
2	Right Hemicolectomy	Tubular adenoma (precancerous polyp)
3	Right Hemicolectomy	Moderately differentiated adenocarcinoma (T2)

4	Sigmoid colectomy	Treated moderately differentiated adenocarcinoma (T3); post chemotherapy
5	Right Hemicolectomy	Tubular adenoma (precancerous polyp)
6	Low anterior resection	Complete pathologic response-no residual tumor after radiation and chemotherapy
7	Left colectomy	Moderately differentiated adenocarcinoma (T2)
8	Low anterior resection	Complete pathologic response-no residual tumor following radiation and chemotherapy
9	Low anterior resection	Tubulovillous adenoma (precancerous polyp)
10	Right hemicolectomy	Moderately differentiated adenocarcinoma (T3)
11	Right hemicolectomy	Moderately differentiated adenocarcinoma (T2)
12	Right hemicolectomy	Moderately differentiated adenocarcinoma (T4)
13	Right hemicolectomy	Tubular adenoma (precancerous polyp)
14	Left colectomy	Moderately differentiated adenocarcinoma (T4)
15	Left Hemicolectomy	Tubulovillous adenoma (precancerous polyp)
16	Low anterior resection	No residual tumor following prior polypectomy*
		Moderately differentiated adenocarcinoma (T3)
17	Sigmoidectomy	Moderately differentiated adenocarcinoma (T2)
18	Transverse colectomy	Treated moderately differentiated adenocarcinoma
19	Low anterior resection	(T3); post radiation and chemotherapy
		Moderately differentiated adenocarcinoma (T3)
		Moderately to poorly differentiated adenocarcinoma
20	Low anterior resection	(T3)
21	Total colectomy	Moderately differentiated adenocarcinoma (T3)
		Differentiated adenocarcinoma (mixed polyp and
22	Right hemicolectomy	adenocarcinoma)
23	Right hemicolectomy	

Note: T is the primary tumor depth of invasion, per TNM guidelines.

* No residual tumor was found after polypectomy. We have grouped this case with the complete responders.

6.2.2 Extraction of Functional, Spectral, and Textural features

Several functional, spectral, and textural features were extracted from the PAT and US data and images as given in Table 6.2. These features were calculated as given in previous chapters in this dissertation. Examples of spectral feature calculation for a malignant and a benign colon sample are given in Figure 6.1.

Table 6.2: Abbreviations

Abbreviation	Description
rHbT	relative total hemoglobin

SS (PAT)	mean PAT spectral slope
0.5MHz SI (PAT)	0.5 MHz spectral intercept from PAT spectra
MBF (PAT)	mid-band fit from PAT spectra
SS (US)	mean US spectral slope
0.5MHz SI (US)	0.5 MHz spectral intercept from US spectra
MBF (US)	mid-band fit from US spectra
Sig_rad	standard deviation of the mean radon transform
Homogeneity	the homogeneity of image textures
Energy	the grayscale distribution homogeneity of images and texture crudeness
Contrast	the sharpness of images and the depth of texture grooves
Correlation	the consistency of image texture

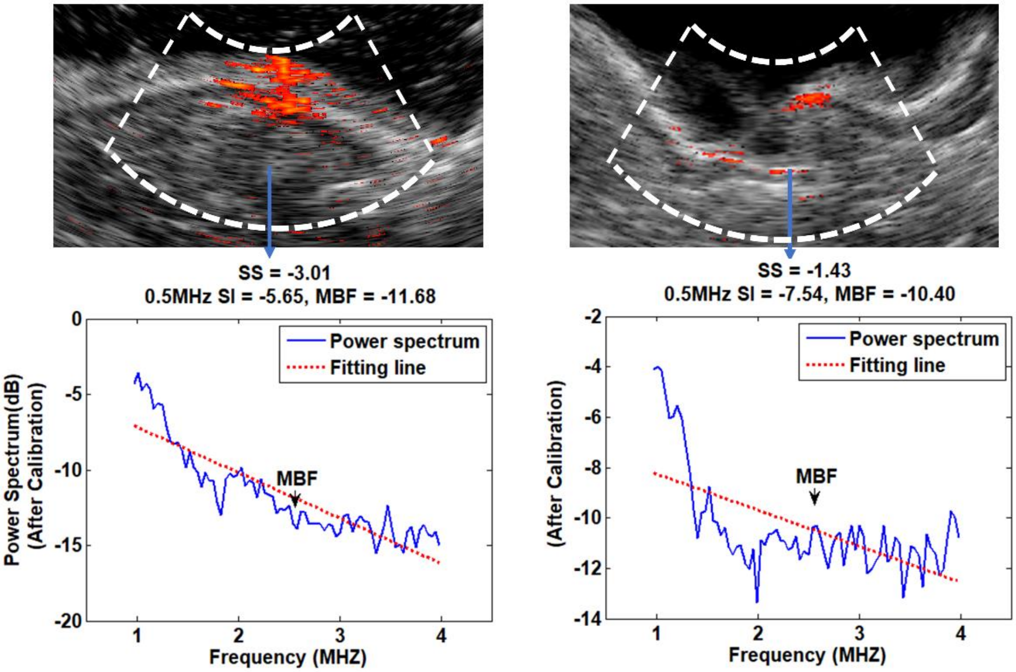


Figure 6.1: Top row: co-registered rHbT and US images of a cancerous (left) and a normal (right) colon sample. Bottom row: the calibrated PAT power spectra along with their fitted lines in the regions marked with the angular dashed lines in each image.

6.2.3 Feature Selection and Classification

A two-step approach was used to select features most likely to differentiate normal from untreated malignant tissue. In the first step, all previously discussed PAT/US features were tested in univariate analysis between untreated cancer and normal regions, and *p* values were generated by two-sample two-sided Student's *t* tests. The features where $p > 0.05$ – which we concluded *a priori* not to be significantly associated with malignancy – were excluded from the classification model (Table 6.3).

Table 6.3: Significance testing of individual covariates as related to tissue diagnosis

Feature	p-value
rHbT	<0.001
SS (PAT)	<0.001
0.5 MHz SI (PAT)	0.002
0.5 MHz SI (US)	0.01
Homogeneity	0.01
Energy	0.02
Sig_rad	0.03
MBF (PAT)	0.12
MBF (US)	0.23
SS (US)	0.55

Next, logistic a general logistic mode (GLM) and a support vector machine (SVM) were used to evaluate the strength of association of each feature with the ultimate tissue diagnosis, and a prediction model was then constructed with significant covariates. In total, 18 areas selected from 18 specimens and 12 malignant areas from 12 untreated cancer specimens were used to construct and evaluate the prediction models. Out of these, 12 normal and 8 malignant areas were used for prediction model derivation and the rest (6 normal and 4 malignant areas) for internal model validation. The receiver operating curve (ROC) and the area under the curve (AUC) were used to evaluate the accuracy of the model. Finally, a second prediction model was constructed without rHbT to determine how limiting the PAT/US device to a single wavelength would affect identification of malignancies.

6.3 Results

6.3.1 Qualitative Analysis: Baseline Characteristics of US and PAT Images

Colorectal tissues are composed primarily of fluid, lipid, collagen, and muscle. The general architecture (from superficial to deep) in a normal specimen is mucosa (fluid-filled cells surrounded by lipid bilayers), submucosa (largely composed of extracellular collagen matrix and some muscle fibers), muscularis propria (muscle), and adipose tissue (lipid). In malignancy, the individual cell types are similar but the architecture is distorted because cancerous cells of mucosal origin penetrate into the deeper layers of the organ. As these cells invade, the organized structure of the tissue is lost.

Figure 6.2 shows specimen photographs, US images, and co-registered PAT/US rHbT maps as well as histologic images from a representative region of normal colon samples (6.4 a - 6.4 d), and a colorectal malignancy (6.4 e - 6.4 h). The white arrows indicate the scanning direction along which B-scans were recorded at four different wavelengths (the imaging plane is perpendicular to the scanning direction). In the standalone ultrasound images, the normal layered structure of the colorectal wall is clearly delineated (6.4 b). In the presence of malignancy, however, this organized structure is distorted by the tumor and loses the clear delineation of mucosal, submucosal, and muscular layers. These findings mirror the differences in histology among the specimens; in contrast to the ordered layering of the normal colonic wall (images 6.4 d), the tumors appear disorganized with destruction of the underlying colonic architecture (images 6.4 h).

Additionally, the rHbT maps computed from coregistered PAT/US images of benign regions show significantly lower rHbT signals (6.4 c) than maps from the malignant lesions (6.4 g). As demonstrated in these representative images, normal tissue was found to have almost no detectable rHbT signal. In contrast, malignant tissue showed much higher concentrations of hemoglobin around the tumor bed. Again, these findings appear corroborated by histologic examination. In comparison to the relative paucity of large blood vessels in normal tissue, the malignancies were more vascular and contained large blood vessels (red arrows in images 6.4 h).

It is interesting to note that fatty tissues have limited PAT signals in the outer portions of the specimens. This is not surprising since we are specifically targeting hemoglobin – which is not concentrated in fatty tissue – as our chromophore of interest, and therefore we image within the 730-830 nm wavelength range. Additionally, all PAT images are displayed with the same dynamic range of -10 dB, so anything below this level is not displayed. The fatty tissue, due to its lack of vascular structures, falls below this range.

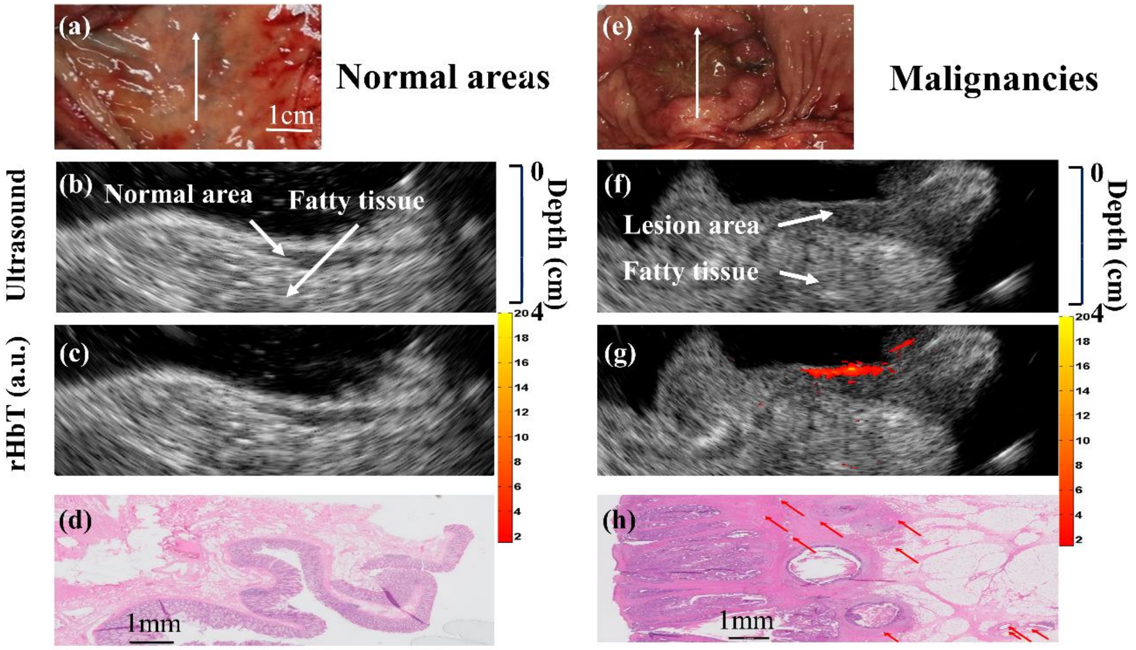


Figure 6.2: Color photograph, US image, rHbT map, and H&E image from representative areas of (a)-(d) a normal region and (e)-(h) a malignant region of pretreatment colorectal cancer tissue. Red arrows identify blood vessels within the histologic images.

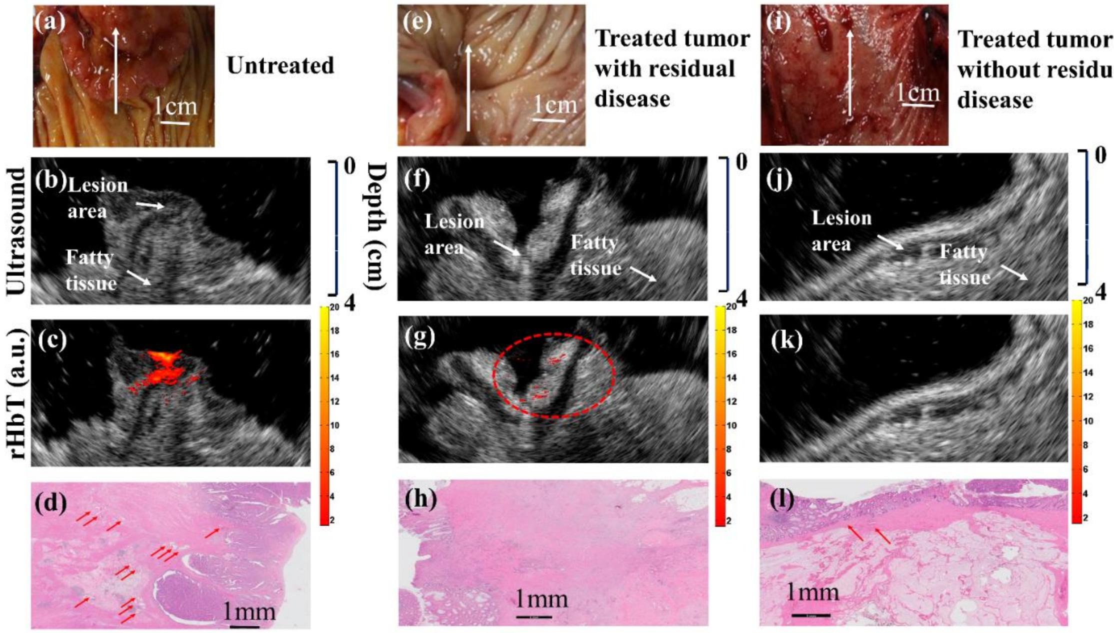


Figure 6.3: Color photograph, US image, rHbT map, and H&E image from representative areas of (a)-(d), a pretreatment colorectal cancer, (e)-(h) a post-treatment colorectal cancer tissue with residual disease, and (i)-(l) a post-treatment colorectal cancer tissue without residual disease.

6.3.2 Evaluation of Treated Tumors

Figure 6.3 shows corresponding images from a representative untreated colon cancer sample (6.5 a - 6.5 d), a colon tumor treated with preoperative chemotherapy (6.5 e – 6.5 h), and a rectal malignancy that received radiation and chemotherapy prior to surgical resection (6.5 i - 6.5 l). The untreated tumor displays findings consistent with other untreated colorectal cancers: loss of layered wall structure, increase in rHbT signal, and increased vascularity throughout the tumor bed. However, treatment appears to reverse these changes. For example, the PAT signal appears to diminish with chemotherapy (6.5 g) and even disappear altogether with complete destruction of the tumor (6.5 k). Additionally, ultrasound imaging demonstrates a return to the normal wall structure with complete tumor destruction (6.5 j). Histologic comparison among specimens also correlates with these findings; a reduction in vasculature along with return to a semi-organized mucosal structure is noted throughout the treated specimens.

6.3.3 Quantitative Analysis

In addition to the above qualitative comparisons, we extracted features from 23 non-overlapping areas obtained from 12 untreated malignant tumors, 6 polyps, 3 post-treatment complete responders or patients with no residual tumor following prior polypectomy, and 2 post-treatment non-responders. We also extracted features from 18 normal non-overlapping areas from specimens of normal regions. Thus, a total of 41 non-overlapping regions were used. Note that each non-overlapping area was obtained from each separate sample. These regions were identified by the attending pathologist.

Figure 6.4 (a-g) shows the boxplots of the seven features calculated from the functional, spectral and image differences between the different types of colorectal tissue. The n number given in the plots corresponds to the total number of areas. The malignant regions demonstrated elevated rHbT, 0.5MHz SI (PAT) and 0.5MHz SI (US) score smaller (less negative) compared to normal and precancerous regions. For SS (PAT), the malignant regions score below normal and precancerous polyps. Treated tumors with complete response were found to have similar scores to normal tissue, while treated regions with residual cancer have scores similar to untreated cancers. Due to the limited number of treated cancers, statistics were not performed for these two treated categories.

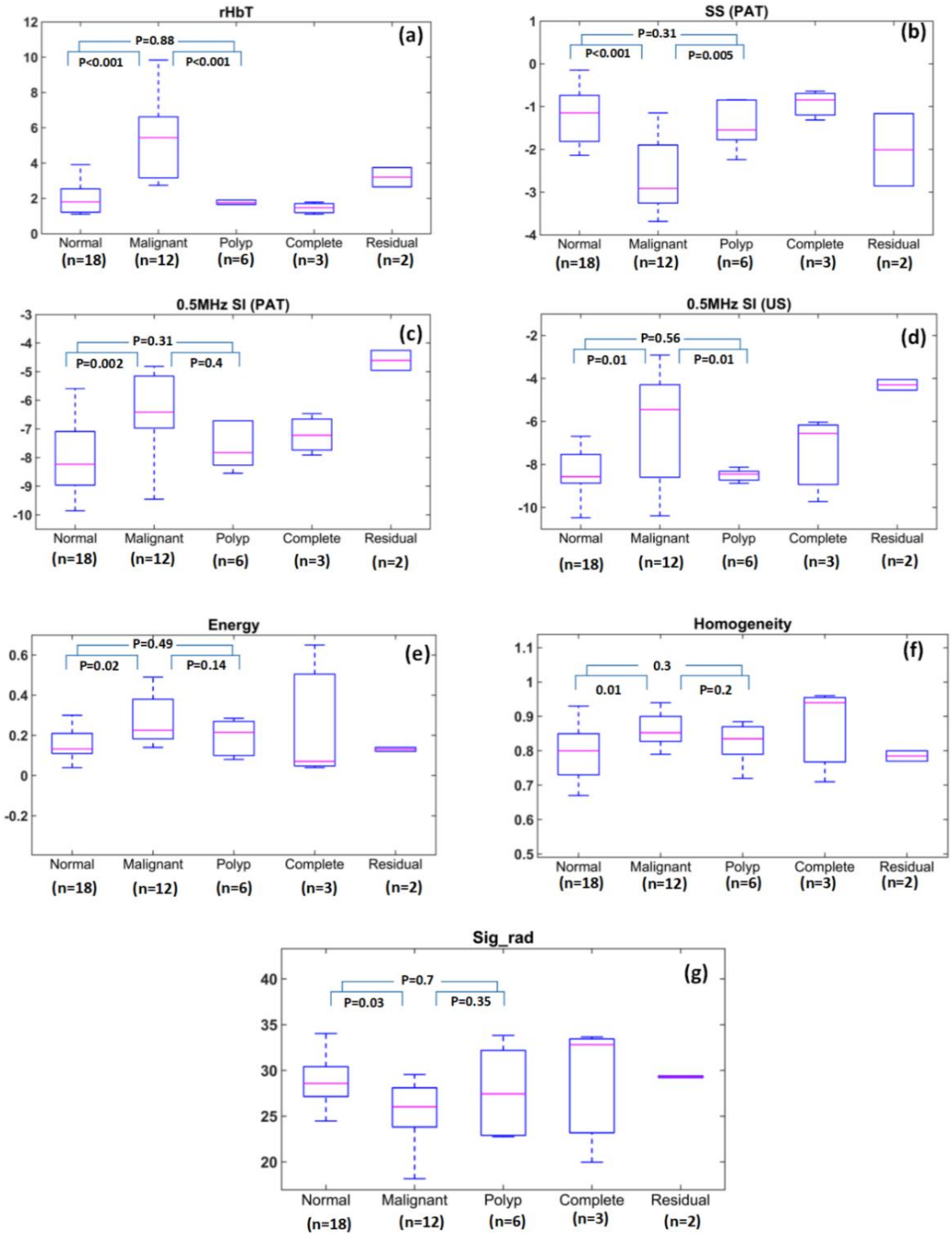


Figure 6.4: Boxplots of (a) total hemoglobin; (b) the mean spectral slope from PAT spectra; (c) 0.5 MHz spectral intercept from PAT spectra; (d) 0.5 MHz spectral intercept from US spectra; (e) energy from the second order statistics of PAT images; (f) homogeneity from the second order statistics of PAT images; (g) standard deviation of the mean radon transform.

To distinguish untreated malignant from normal colon tissues, GLM and SVM classifiers were established. These classifiers were developed using the independent features with a p-value less than 0.05 between malignant and normal colon tissues. To determine if two features are independent, Spearman's correlation was calculated between each pair of features (Table 6.4). To train each classifier, we first used the feature with the lowest p-value and then added other features to the feature set one by one. We continued to include features to the feature set until no increase in the AUC value for the testing data set was observed. We found that when rHbT was included in the feature set, the best performance of both the GLM and SVM classifiers (the highest AUC value for the testing data set) was achieved when rHbT and 0.5MHz SI (PAT) were employed to train the classifier, although SS (PAT) has a lower p-value than 0.5MHz SI (PAT). Adding other features did not improve the AUC for the testing data set.

Table 6.4: The correlation between significant features used in this study

	SS (PAT)	**0.5MHz SI (PAT)**	**0.5MHz SI (US)**	**Homogeneity**	**Energy**	**Sig_rad**
rHbT	0.65	0.45	0.27	0.34	0.41	0.37
SS (PAT)		0.67	0.23	0.42	0.46	0.29
0.5MHz SI (PAT)			0.21	0.49	0.41	0.4
0.5MHz SI (US)				0.41	0.37	0.31
Homogeneity					0.79	0.91
Energy						0.82

Figure 6.5 shows the ROC curves and AUC values of the training (left) and testing (right) data sets using GLM (top row) and SVM (bottom row) classifiers. As shown in this figure, when the features set include just rHbT, the AUC values for the training and testing data sets are 0.95 and 0.93 for both classifiers, respectively. Adding 0.5MHz SI (PAT) to the features set results in a significant improvement in the AUC values for both the training and testing data sets (0.97 and 0.95 for the training and testing data sets for both classifiers, respectively). The three image features (Sig_rad,

Homogeneity, and Energy) did not improve the AUC values for either the training or testing data sets.

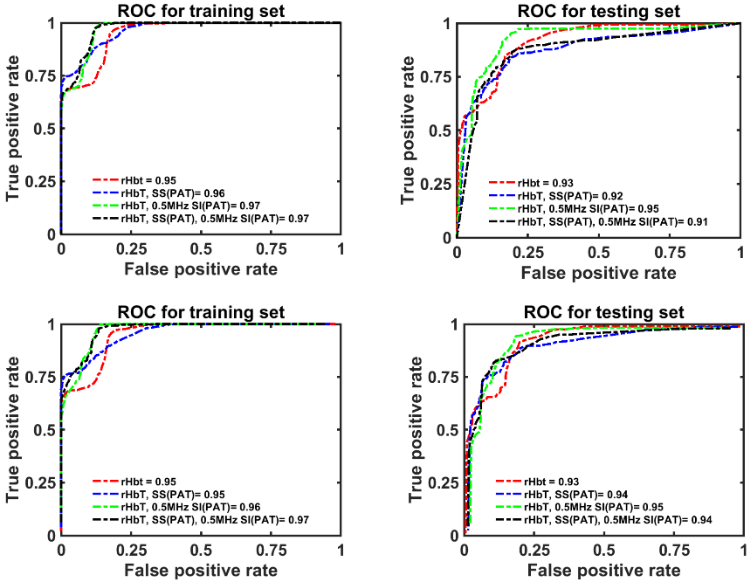


Figure 6.5: ROC curves and their associated AUC values for the training and testing data sets in the presence of rHbT in the feature set. (a), (b) GLM classifier performance. (c), (d) SVM classifier performance.

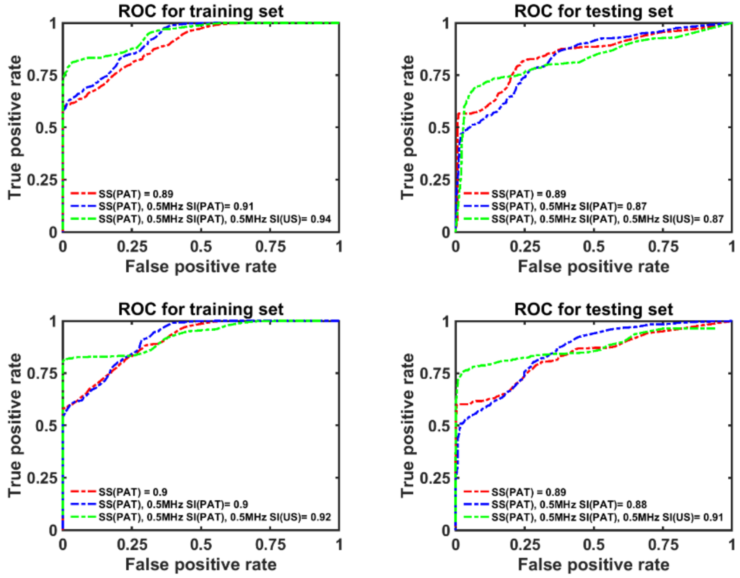


Figure 6.6: ROC curves and their associated AUC values for the training and testing data sets in the absence of rHbT in the feature set. (a), (b) GLM classifier performance. (c), (d) SVM classifier performance.

Finally, the performance of the GLM (top row) and SVM (bottom row) classifiers without rHbT (the single-wavelength model) are presented in Figure 6.6. Note that although the difference between some of the PAT image features in malignant and normal samples is statistically significant, none of these features improve the AUC for the testing data sets. The best performance of GLM classifier is achieved when the only spectral feature SS(PAT) is included in the feature set. The best performance of SVM classifier is achieved when the spectral features SS(PAT), 0.5MHz SI(PAT), and 0.5 MHz SI(US) are included in the feature set. The testing AUC in this case is 0.89 for the GLM classifier and 0.91 for the SVM classifier.

6.4 Discussion and Summary

In this pilot study of co-registered ultrasound and photoacoustic tomography, we found significant qualitative and quantitative differences between malignant tumors and normal tissue within human colorectal specimens. Specifically, the parameters rHbT, 0.5MHz SI (PAT), 0.5MHz SI (US), and SS (PAT) differ between the two tissue types imaged, suggesting that PAT may be able to differentiate malignant from normal tissue in the colon and rectum. Combined with the PAT system's tissue penetration depth of over 4 to 5 cm (depending on the background tissue optical properties), these findings suggest that PAT may be able to augment extant radiographic technology in the diagnosis, management, and surveillance of colorectal cancer.

As demonstrated by Guan et al [22] and Ronald [23], PAT spectral features are related to the size and concentration of the optical absorbers. The slope decreases (more negative) as PA absorber sizes increase and the intercept increases (less negative) as the sizes and concentrations of the absorbers increase. We believe that malignant lesions have larger absorber sizes and higher concentrations compared with normal colorectal tissues due to their increased microvessel networks. As introduced by Lizzi and co-workers [24], US spectral slope depends on acoustic scatter size while spectral intercept depends on scatter sizes, concentrations and acoustic impedances of tissue scatter matrix. These parameters have been found valuable to characterize liver, eye [24], prostate [25], and breast lesions [26]. We believe that the distorted tissue architecture and abundance of cancerous cells are the source of the US spectral contrast between malignant and normal colorectal tissues. However, the findings of the PAT and US spectral features of colorectal diseases may or may not be applicable to diseases of other organs.

Several technical limitations must be considered with our data. First, we imaged colorectal specimens obtained from routine surgeries and these tissues were typically with large pathologic components that often appeared malignant by visual inspection after specimens were open. These lesions may or may not need advanced PAT and US features for diagnosis. However, these lesions are excellent examples for identifying PAT and US feature characteristics that differ between cancerous and normal tissue. With this information known, we can target less obvious lesions as we look to test the utility of the device in identifying cancer margins and residual tumors after chemo-radiation treatment in patient.

The second limitation of this study is the low image resolution of our prototype. The image resolution is only ~250 μm due to the commercial endo-cavity ultrasound transducer array (6 MHz central frequency, 80% bandwidth). Because this resolution will impact future clinical applications of the device, we plan to upgrade the ultrasound system with a transducer array of more than 15 MHz to address this problem in future studies. Third, sO2 was not calculated in this study since all specimens were imaged after resection, resulting in significantly altered sO2 compared to normal living tissue. sO2 is a significant biomarker for characterization of cancer [11] and assessment of treatment response.

Thirdly, the limited sample size could lead to overfitting of the classifiers if enough care is not taken to develop the classifiers. As a rule of thumb, overfitting is least possible to occur if the number of samples is 10 times or higher than the number of independent predictors [27]. Based on this rule, as we have a total of 30 samples (18 normal colorectal tissues and 12 untreated malignant colorectal tissues) for ROC analysis, the maximum number of the predictors that should be used to avoid overfitting would be three. Figure 6.5 shows that when rHbT is present in the feature set, the best performances of both GLM and SVM are achieved when rHbT and 0.5MHz SI (PAT) are the only features used to train the classifiers. Adding SS (PAT) to these feature set neither changes the value of the AUC for training data sets, nor increases the AUC for the testing data set. Moreover, when rHbT is not included in the feature set, employing the combination of SS (PAT), 0.5MHz SI (PAT), and 0.5MHz SI (US) features for developing the classifiers would result in the best performance of SVM classifier and SS (PAT) only would result in the best performance of GLM classifier (Figure 6.6). While adding Sig_rad increases the AUC value for the training data set, it decreases the AUC values for the testing data set in both classifiers. This would mean that our classifiers have most probably been overfitted when four features have been used. In this study, to further protect our classifiers from overfitting, repeated rounds (100 times) of cross validation were applied by randomly selecting 2/3 of the samples for training and 1/3 of the samples for testing. The average ROC and AUC values were reported as the results.

In summary, a real-time co-registered PAT/US system was used to image and characterize colorectal masses ex-vivo in this pilot study. 23 colon and rectum samples (19 colon and 4 rectum) were imaged, rHbT was computed from 4 wavelength data, and seven quantitative features were extracted from PAT and US power spectra and images. In pre-treated malignant colorectal tumors, we found the cross-section structure to be highly disorganized with a significantly higher rHbT concentration compared to normal and precancerous regions. We performed classifications on the malignant and normal colon regions using GLM and SVM classifiers both with and without HbT in the feature set. When rHbT was employed to construct the classifiers with 0.5 MHz SI (PAT), GLM and SVM classifiers achieved optimal AUC values for the training and testing data sets (0.97 and 0.95, respectively). The small number of treated tumors included in this dataset limits the statistical power of the analysis, but the functional, spectral and image parameters do appear more similar to normal colorectal tissue in tumors that have experienced complete responses compared to partial responders. These results indicate potential of using PAT/US for future cancer screening and post-treatment surveillance of the colon and rectum. Moving forward, we plan to increase the resolution of our system by using a high frequency US array and then adapt the technology to an endo-rectal probe, which will allow us to test the functional and spectral feature differences in in-vivo human tissue.

References

1. L. V. Wang et al., “Photoacoustic tomography: in vivo imaging from organelles to organs,” science, 335(6075), 1458-1462 (2012).
2. N. Weidner et al., “Tumor angiogenesis and metastasis—correlation in invasive breast carcinoma,” New England Journal of Medicine, 324(1), 1-8 (1991).

3. P. Vaupel et al., “Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review,” Cancer research, 49(23), 6449-6465 (1989).
4. M. Hockel et al., “Tumor hypoxia: definitions and current clinical, biologic, and molecular aspects,” Journal of the National Cancer Institute, 93(4), 266-276 (2001).
5. Y. Yuan et al., “Preclinical photoacoustic imaging endoscope based on acousto-optic coaxial system using ring transducer array. Optics letters, 35(13), 2266-2268 (2010).
6. X. Leng et al., “Feasibility of co-registered ultrasound and acoustic-resolution photoacoustic imaging of human colorectal cancer,” Biomedical optics express, 9(11), 5159-5172 (2018).
7. H. S. Salehi et al., “Coregistered photoacoustic and ultrasound imaging and classification of ovarian cancer: ex vivo and in vivo studies,” Journal of biomedical optics, 21(4), 046006 (2016).
8. T. Wang et al., “Characterization of ovarian tissue based on quantitative analysis of photoacoustic microscopy images,” Biomedical optics express, 4(12), 2763-2768 (2013).
9. H. Li et al., “Utilizing spatial and spectral features of photoacoustic imaging for ovarian cancer detection and diagnosis,” Journal of biomedical optics, 20(1), 016002 (2015).
10. T. Wang et al., “A low-cost photoacoustic microscopy system with a laser diode excitation,” Biomedical optics express, 5(9), 3053-3058 (2014).
11. S. Nandy et al., “Evaluation of ovarian cancer: initial application of coregistered photoacoustic tomography and US,” Radiology, 289(3), 740-747 (2018).

12. G. Yang et al., “Optimized light delivery probe using ball lenses for co-registered photoacoustic and ultrasound endo-cavity subsurface imaging,” Photoacoustics, 13, 66-75 (2019).
13. G. Yang, et al., “Co-registered photoacoustic and ultrasound real-time imaging of colorectal cancer: ex-vivo studies,” in Photons Plus Ultrasound: Imaging and Sensing 2019, Proc. SPIE 108784K, (2019) [doi:10.1117/12.2510914].
14. R. L. Siegel et al., “Colorectal cancer statistics, 2017,” CA: a cancer journal for clinicians, 67(3), 177-193 (2017).
15. American Cancer Society, “Cancer Facts & Figures,” American Cancer Society, Atlanta, Ga, USA (2014-2016).
16. J. Pan et al., “Colonoscopy reduces colorectal cancer incidence and mortality in patients with non-malignant findings: a meta-analysis,” The American journal of gastroenterology, 111(3), 355 (2016).
17. Y. Zeng et al., “The Angular Spectrum of the Scattering Coefficient Map Reveals Subsurface Colorectal Cancer,” Scientific reports, 9(1), 2998 (2019).
18. T. Wang et al., “Label-free biomolecular imaging using scanning spectral interferometry,” Chinese Optics Letters, 11(11), 111102 (2013).
19. S. P. Raman et al., “Evolution of imaging in rectal cancer: multimodality imaging with MDCT, MRI, and PET,” Journal of gastrointestinal oncology, 6(2), 172 (2015).
20. E. Al-Sukhni et al., “Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis,” Annals of surgical oncology, 19(7), 2212-2223 (2012).

21. S. P. Hiotis et al., “Assessing the predictive value of clinical complete response to neoadjuvant therapy for rectal cancer: an analysis of 488 patients,” Journal of the American College of Surgeons, 194(2), 131-135 (2002).
22. G. Xu et al., “Photoacoustic spectrum analysis for microstructure characterization in biological tissue: A feasibility study,” Applied Physics Letters, 101(22), 221102 (2012).
23. R. Kumon, “Frequency-domain analysis of photoacoustic imaging data from prostate adenocarcinoma tumors in a murine model,” Ultrasound in Medicine & Biology, 37(5), 834-839 (2011).
24. F. L. Lizzi et al., “Relationship of ultrasonic spectral parameters to features of tissue microstructure,” IEEE Transactions on ultrasonics, ferroelectrics, and frequency control, 34(3), 319-329 (1987).
25. E. Feleppa et al., “Ultrasonic spectral-parameter imaging of the prostate,” imaging systems and technology, 8(1), 1098 (1997).
26. H. Tadayyon et al., “Quantitative ultrasound characterization of locally advanced breast cancer by estimation of its scatterer properties,” Med Phys, 41(1), 012903 (2014).
27. J. Concato et al., “The risk of determining risk with multivariable models,” Ann Intern Med, 118 (3), 201-210 (1993).

Chapter 7: Colorectal cancer diagnosis using coregistered photoacoustic microscopy and ultrasound system – comparison of CNN and GLM classifiers

7.1 Introduction

Colorectal cancer is the third most common cancer diagnosed in both men and women in the United States [1]. While treatment often involves chemotherapy, radiation, and surgical resection, recent advances in neoadjuvant (preoperative) treatment of locally advanced rectal cancers (LARC) have enabled 20-30% of patients to safely avoid surgery altogether [2-5]. However, this "watch and wait" depends on accurate assessments of tumor regression and high-resolution and high-sensitivity surveillance imaging for tumor recurrence.

Standard surveillance modalities include physical exam, endoscopy with biopsy, and MRI; however, each of these modalities have distinct weaknesses in the post-treatment setting. [6-12]. The poor performance of current technology makes it extremely difficult to identify patients who can safely avoid surgery (pCRs) from those who need resection (non-responders).

To overcome these challenges, we developed a co-registered endorectal photoacoustic microscopy and ultrasound (PAM/US) system to assess rectal cancer treatment response [13-14]. Photoacoustic imaging (PAI) is a hybrid imaging technology that uses a short laser pulses to excite hemoglobin molecules endogenous to the human body. The resulting acoustic waves are then acquired by US transducers and analyzed for vascular bed quantification. This process has been utilized in many different areas such as breast cancer [17-18], lung cancer [19-20], ovarian cancer [21], skin cancer [22], and colorectal cancer [13-14].

A convolutional neural network (CNN) is an artificial intelligence algorithm with remarkable capabilities for automated image analysis [23]. To quantitatively interpret the large volumes of data acquired by the PAM/US system, we designed and incorporated deep-learning CNN models in the PAM system (PAM-CNN) [14]. While our deep-learning PAM-CNN model can accurately assess rectal cancer treatment response, it requires a large training and validation data set. The key question remains if the PAM-CNN outperforms traditional histogram-feature based models. In this study, using 24 ex-vivo and 10 in-vivo data sets, we compare the performances of the PAM-CNN and the traditional histogram-parameter-based classifiers in rectal cancer treatment evaluation. Unlike CNN models, a generalized logistic regression (GLM) classifier does not require a large dataset for training and validation, however, imaging features must be extracted and evaluated on their diagnostic accuracy. We have computed five PAM image histogram features and used them to train, validate and test GLM classifiers. The performance of deep learning based CNN models is compared with GLM classifiers. To the best of our knowledge, this study is the first to establish the role of deep-learning PAM-CNN in rectal cancer evaluation.

7.2 Methods

7.2.1 Patients, specimens, and PAM imaging

Briefly, 10 participants (mean age, 58 years; range 42 – 68 years; 2 women and 8 men) completed radiation and chemotherapy from September 2019 to September 2020 and were imaged with the PAM/US system prior to surgery. Colorectal specimens from another group of 24 patients who had undergone surgery were studied ex-vivo. All studies were approved by the institutional review board of the Washington University School of Medicine, and all patients provided written informed consent. In the ex-vivo study, each specimen was evaluated within one hour of surgical resection and prior to formalin fixation. In the in-vivo study, patients who had previously undergone preoperative treatment with radiation and chemotherapy were imaged in-vivo before resection.

7.2.2 PAM/US endoscope

The PAM endoscope consists three parts: a handle, a water channel (the main body), and an imaging head, as shown in Figure 7.1 A [14]. Briefly, the water inlet which allows water injected from a syringe to inflate a water balloon covering the image head for ultrasound coupling. A stepper motor in the handle turns a hollow shaft in the water channel to rotate the image head 360° for full circle imaging. An optical fiber inside the hollow shaft delivers laser pulses to the imaging head. An ultrasonic transducer (20 MHz, 75% bandwidth) fixed on the imaging head both transmits and receives ultrasound signals, and also receives PA signals. An Nd: YAG laser working at 1064 nm with a 1000 Hz pulse repetition rate is the light source. A 0.15 cm^2 tissue area is illuminated by 3.6 mJ laser pulses from the probe tip, resulting in a surface optical fluence of 24 mJ/cm^2, which is well within the ANSI safety threshold (100 mJ/cm^2) at 1064 nm [24]. This fluence is further reduced by energy diffusion caused by the balloon.

During imaging, the PAM endoscope is inserted transanally through a proctoscope, (Figure 7.1 C). Ruled scales on the water channel (Figure 7.1 B) show how deeply the endoscope is inserted into the rectum where the images are obtained.

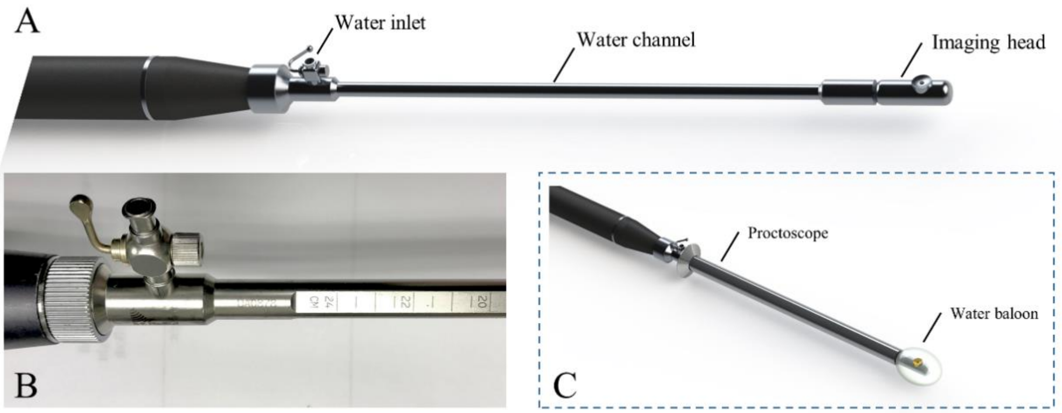


Figure 7.1: PAM endoscope (A), scales on water channel (B) and endoscope in a proctoscope, with a balloon on the tip (C).

7.2.3 PAM and US data selection for training/validation and testing of models

For training and validation, three to five regions of interest (ROIs) were selected at uniformly spaced locations on each PAM or US B-scan image acquired from normal regions or a tumor bed (Figure 7.2). For example, the red ring in Figure 7.2 represents mucosa vasculature, which is continuous in the normal image of Figure 7.2 B. The blue rectangles indicating ROIs are uniformly spaced along the perimeter of the image. In the cancer image, the dark zones and discontinuities in the red ring from approximately 9:00 to 1:00 o’clock indicate tumor, so the ROIs are uniformly spaced in that segment. A total of 2600 US ROIs (1262 normal and 1496 cancerous) and 2004 PA ROIs (1207 normal and 797 cancerous) were compiled from 24 patients’ ex-vivo images and 10 patients’ in-vivo images. Two ex-vivo samples showed a low signal-to-noise ratio (SNR) on PAM images due to a laser energy problem. We excluded those two samples in training PAM-CNN and PAM-GLM models. For the US-CNN and US-GLM models, we used all 24 ex-vivo and all 10 in-vivo patient data.

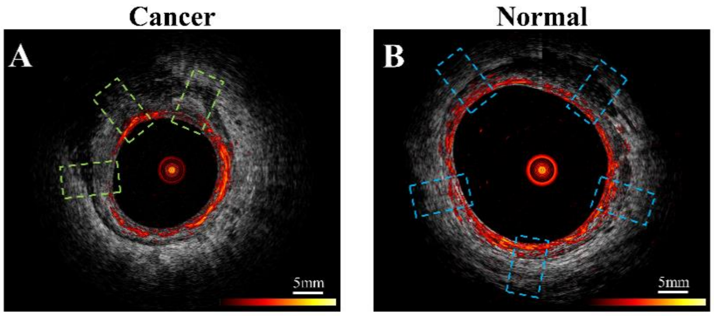


Figure 7.2: Example co-registered PAM and US images showing ROIs of (A) residual cancer tissue, area in green dashed line boxes, and (B) normal tissue, area in blue boxes. PAM ROIs are cropped from PAM images, and US ROIs are cropped from US images.

We divided the total of 2004 PA ROIs and total of 2600 US ROIs into two discrete data sets – one for model training and validation and another for testing, respectively. The training set included all ex-vivo cases (including normal, pCR responders, and cancers) and half of the in-vivo patient data consists of normal regions, including the normal area and pCR treated tumor bed, and 3 cancer beds from non-responders. Of the training set ROIs, 80% were used for training with the remainder for internal validation. The testing set contained the other half of the in-vivo patient data consisting of 5 normal regions, one tumor bed of a pCR, and 3 cancer samples from non-responders.

7.2.4 GLM models

We used selected image features of ROIs to develop PAM-GLM and US-GLM models. To calculate the histogram of each ROI, we divided the ROI into 32 bins. The bar height of each bin was then computed by dividing the number of pixels with a given value in an associated range by the size of the image. From the histogram of each ROI, we then extracted five features: mean, standard deviation, skewness, kurtosis, and energy.

All the PAM and US features showed significant differences between malignant and normal colorectal tissues ($p<0.05$) (Appendix Figure 7.1S and Figure 7.2S). Therefore, all these features were considered as potential candidates when building PAM-GLM and US-GLM models. To prevent model overfitting, the Spearman's correlation coefficient between each of the histogram features were calculated (Appendix, Table 7.1S). We developed PAM-GLM classifiers using each histogram feature separately, as well as using combinations of features with low correlation values. The mean AUCs of the training/validation and testing data sets as well as their 95% confidence of interval were computed for each classifier. The same process was followed to construct US-GLM classifiers.

To remove bias in selecting in-vivo data for training and validation, we trained the classifiers 10 times. The training/validation and testing data sets are the same as those used for CNN models described in next section.

Figure 7.3 (PAM-GLM) and Figure 7.4 (US-GLM) show examples of the first order statistical features calculated from malignant rectal tissue ROIs (shown in Figure 7.2 A) and normal rectal tissue ROIs (shown in Figure 7.2 B). As shown in Figure 7.3, in PAM ROIs, the malignant tissue has a lower mean and standard deviation, while the other three features are higher. In Figure 7.4, malignant US ROIs show a lower mean and standard deviation than that of the normal US ROIs.

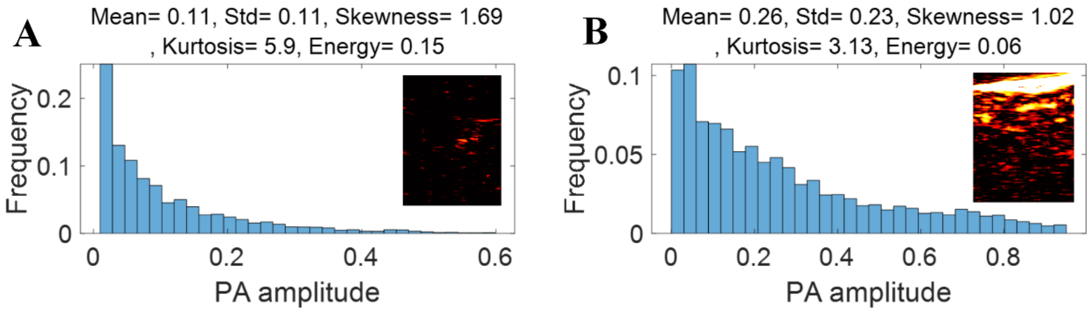


Figure 7.3: First order statistical features calculated from malignant rectal tissue PAM ROIs (A) and normal rectal tissue PAM ROIs (B)

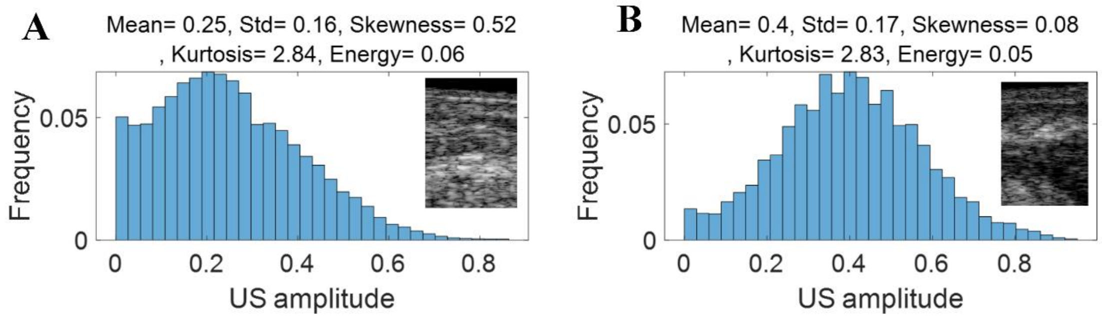


Figure 7.4: First order statistical features calculated from malignant rectal tissue US ROIs (A) and normal rectal tissue US ROIS.

Figures 7.1S in Appendix show the boxplots of the histogram features of the PAM ROIs. The p-value for each feature, calculated from a two-sided statistical t-test, is indicated on each plot. All features are statistically significant ($p<0.05$), however, they are not equally important. To assess the importance of each feature, we first fit a regression model to each feature separately, using all the available data (ex-vivo and in-vivo patients), and then we found the AUC of the fitted model. As shown in Table 7.1, Std, Mean, and Kurtosis respectively provide the highest AUC values among all the features of PAM images. While Std and Mean are highly correlated, the correlation value between Mean and Kurtosis is less than 0.5 (Table 7.S1). Therefore, these two features are used together to develop PAM-GLM classifiers.

Similarly, boxplots of the five features from US ROIs are given Figure 7.2S, and the AUC feature values of the fitted model are shown in Table 7.1. Based on this table, Std, Energy, and Mean are respectively the most important features of the US images. However, they all are highly correlated with each other (Table 7.S2).

Table 7.1: AUCs of the fitted regression model developed using features of PAM and US images.

Feature	AUC (PAM)	AUC (US)
Mean	0.76	0.81
Std	0.79	0.86
Skewness	0.71	0.57
Kurtosis	0.73	0.62
Energy	0.70	0.85

7.2.5 CNN models

The PAM-CNN (or US-CNN) architecture contained two sequential feature extraction layers and two fully connected layers [14]. Briefly, each extraction layer had a convolutional layer followed by a pooling layer. To optimize the validation results, the convolution kernel and max-pooling kernel sizes were set to 3 × 3 and 4 × 4, respectively. The first fully connected layer was a 512-node hidden layer, and the second fully connected layer (output layer) generated two output classifications – normal or cancerous. “Normal” described a layer-like vascular distribution in a PAM image or a layer structure in a US image, and “cancerous” described an absence of the normal vasculature pattern in PAM images, or an absence of the layer structure in US images. A softmax activation function in the output layer generated the probabilities of each of the two possible classifications (cancer or normal) for an input image; for each input ROI of a PAM or US image, the CNN model outputted the probability of a normal classification compared to the threshold (e.g.

>50% is normal). In all the other layers, a “ReLU” activation function immediately set all negative values to 0 and left positive values unchanged, avoiding exploding or vanishing gradient problems.

To avoid biased selection, we trained and validated 10 PAM-CNN and US-CNN models each using all the ex-vivo data and a randomly selected half of the in-vivo patient data, while reserving the other half for testing. The maximum number of epochs was 20, with early stopping (a tolerance of 2 epochs) monitored by validation accuracy. If there was no increase in validation for two successive epochs, training was stopped. Stochastic gradient descent was used with a batch size of 20, and the RMSprop optimizer function was used to optimize the neural net weights. The learning rate was set to 10^{-3} with a decay of 10^{-5}. In each model, 80% of the ROIs from the training & validation set were used to train the model, the remaining 20% were used for validation, and 20× cross validation was performed.

The ROIs of each in-vivo normal or tumor bed patient images were either all used in training or all used in testing. Each of the 10 CNN models was tested on a randomly selected half of the in-vivo data and generated an ROC. The overall performance of the classifier was measured by the mean AUC of the 10 models.

The method for calculating PAM-CNN’s AUC is different from that of our previous report [14], which leads to a slightly different AUC value. In previous work, the training and validation data set was fixed, which was consisted of 24 ex-vivo and five in-vivo data set. The PAM-CNN’s AUC obtained from another five in-vivo data set unseen by PAM-CNN for testing is 0.98. In this study, we have done a more thorough investigation. The ex-vivo data set is still fixed for training and validation, but the five *in vivo* data set for training and validation and the five in-vivo data set for testing were interchanged randomly for 10 times, and the 10 AUC was used to generate the mean value of AUC.

7.3 Results

7.3.1 GLM models

Table 7.2 shows the mean AUCs and 95% confident of interval for PAM-GLM classifiers developed using single features, as well as feature pairs that are weakly correlated (based on Table S1). As can be seen, the "Mean-Kurtosis" combination results in a better testing performance than "Mean" alone, and a better training performance than "Kurtosis" alone. In the case of US-GLM (Table 7.3), the classifier which is built using "Std" alone performs best on both training and testing data sets (mean AUCs of 0.86 and 0.66 for training and testing data sets, respectively).

Table 7.2: Training and testing mean AUC values for PAM-GLM classifiers developed using different combinations of weakly correlated features. The 95% confidence of interval values are also shown in front of each mean AUC value.

Feature combinations	**Training AUC (95% CI)**	**Testing AUC (95% CI)**
Mean	0.77 (0.767-0.777)	0.80 (0.793-0.807)
Std	0.79 (0.788-0.793)	0.76 (0.746-0.770)
Skewness	0.71 (0.708-0.719)	0.82 (0.815-0.825)
Kurtosis	0.73 (0.724-0.734)	0.82 (0.817-0.827)
Energy	0.72 (0.712-0.727)	0.74 (0.724-0.758)
Mean, Kurtosis	0.74 (0.732-0.743)	0.82 (0.808-0.820)
Std, Energy	0.80 (0.799-0.807)	0.76 (0.750-0.773)
Kurtosis, Energy	0.75 (0.744-0.750)	0.81 (0.805-0.817)

Table 7.3: Training and testing AUC values for US-GLM classifiers developed using different combinations of weakly correlated features. The 95% confidence of interval values are also shown in front of each mean AUC value.

Feature combinations	**Training AUC (95% CI)**	**Testing AUC (95% CI)**
Mean	0.82 (0.818-0.820)	0.64 (0.629-0.657)
std	0.86 (0.860-0.862)	0.66 (0.650-0.674)
skewness	0.59 (0.587-0.591)	0.42 (0.405-0.443)
Kurtosis	0.64 (0.635-0.639)	0.34 (0.326-0.344)
energy	0.85 (0.851-0.854)	0.61 (0.600-0.621)
Mean, kurtosis	0.82 (0.819-0.822)	0.60 (0.581-0.618)
Std, skew	0.86 (0.860-0.862)	0.65 (0.643-0.664)
Std, kurtosis	0.86 (0.858-0.860)	0.65 (0.642-0.666)
Kurtosis, energy	0.86 (0.856-0.858)	0.63 (0.617-0.638)

Figure 7.5 (A) and (B) respectively show the mean training and testing ROCs of three of the best performing (based on both training and testing AUCs) classifiers developed using PAM histogram features. As shown in these plots, “Kurtosis” alone results in a slightly better performance on the testing data set than the other feature combinations (see the 95% CI values in the table). It is worth noting that although adding “Mean” to the features set negligibly lowers the AUC of the testing data set, it increases the AUC of the training data set by 0.01. Finally, the reason for the slightly poor training performance than testing for different combinations of features is that the training data set includes both in-vivo and ex-vivo ROIs while the testing data set contains only in-vivo ROIs. Overall, our in-vivo data have demonstrated slightly better classification between malignant and normal colorectal tissue than the ex-vivo data.

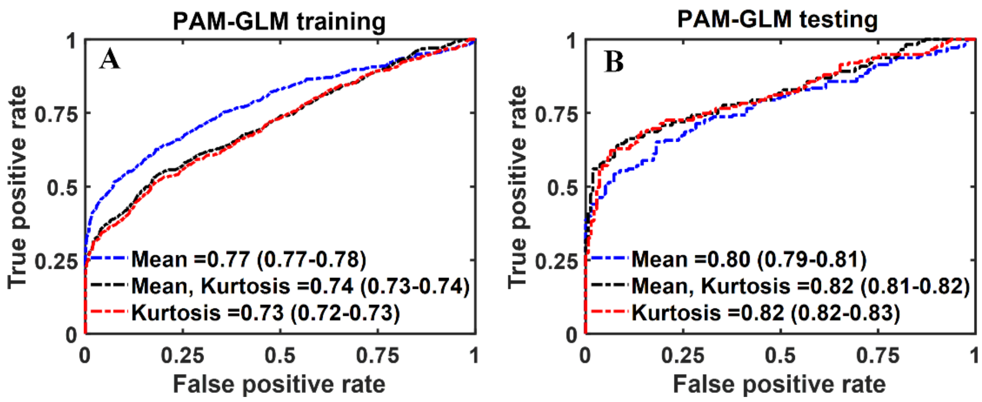


Figure 7.5: The average ROC of the training (A) and testing (B) data sets for different combinations of features set. The features were extracted from PAM images. The 95% CIs are indicated in parentheses.

In the case of US-GLM, using the "Std" histogram feature demonstrates the best prediction AUC of 0.68, as seen in Figure 7.6 (B). Adding any other uncorrelated features does not improve the AUCs of the training or testing data sets as shown in Figure 7.6 (A) and (B).

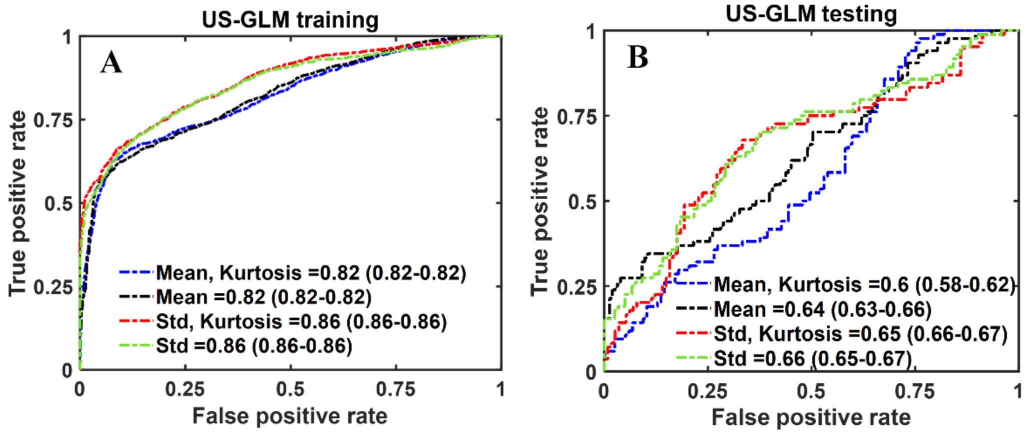


Figure 7.6: The average ROC of the training (A) and testing (B) data sets for different combinations of features set. The features were extracted from US images. The 95% CIs are indicated in parentheses.

7.3.2 CNN models

The mean ROC and AUC of the CNN models were computed from 10 CNN models, using the same shuffle method as in GLM. PAM-CNN demonstrated high performance in training and testing, with a 0.96 AUC for both (Figure 7.7). For US-CNN (Figure 7.8), the average AUC was 0.71 in testing.

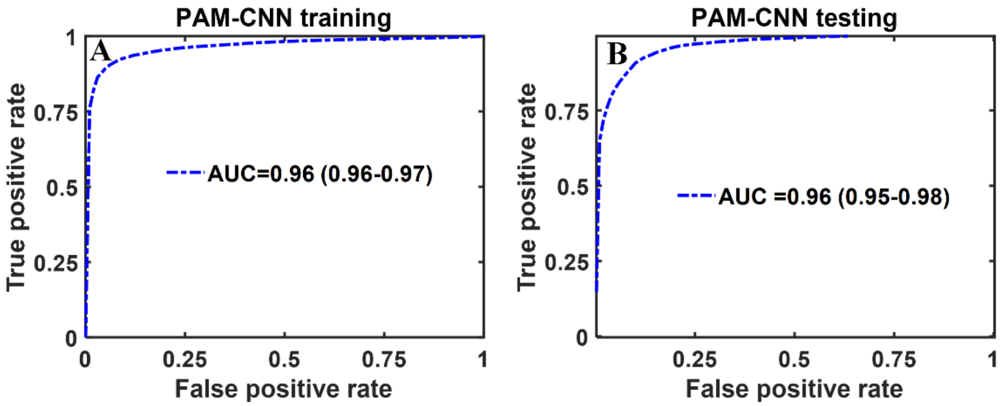


Figure 7.7: Average ROCs of PAM-CNN model. (A) training and validation results, (B) testing results. The 95% CIs are indicated in parentheses.

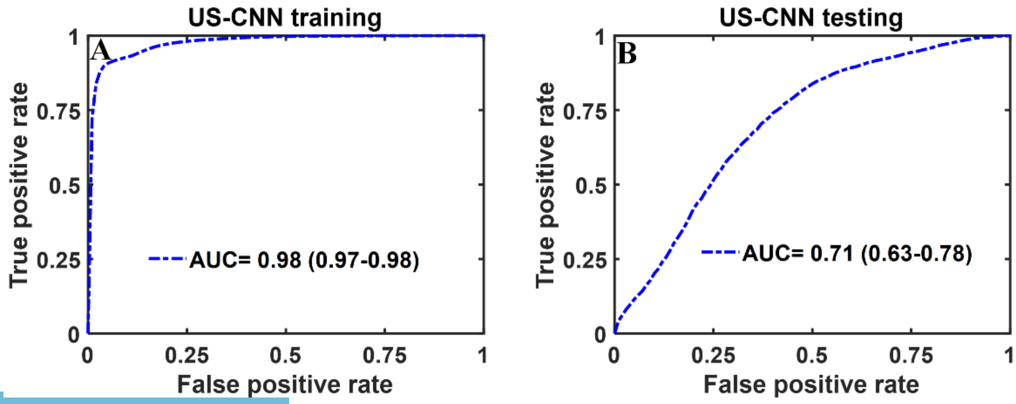


Figure 7.8: Average ROCs of US-CNN model. (A) training and validation, (B) testing results. The 95% CIs are indicated in parentheses.

7.4 Discussion

The general architecture of the normal colon and rectal tissue consists of the mucosa (a thin layer of epithelial cells, a layer of connective tissue, a thin layer of muscle), submucosa (mucous glands, blood vessels, lymph vessels), muscularis propria (a thick layer of muscle), and serosa (an outer layer of the colon). In malignancy, the individual cell types are similar, but the architecture is distorted because cancerous cells of mucosal origin penetrate into the deeper layers of the organ. As these cells invade, the organized structure and vascular network are lost. We have observed uniform, layer-like vasculature with intense photoacoustic signals within normal rectal submucosa and in the tumor beds where complete tumor destruction has occurred. In contrast, heterogeneous and often microvascular-deficient regions have been found consistently in tumor beds with residual cancer at treatment completion [13-14]. The return of a “normal” vascular pattern to the tumor bed appears to signal complete tumor destruction, though this mechanism is not well-understood. As demonstrated, PAM-CNN captures this unique pattern and predicts pCR with a high diagnostic accuracy. PAM-GLM uses first order statistical features extracted from PAM histograms and these features do not contain spatial micro-features that can be leaned by deep-learning neural networks. Thus, the performance of PAM-GLM is significantly poorer than PAM-CNNs.

In summary, we have shown that the performance of deep-learning based PAM-CNN models was significantly better than that of the PAM-GLM classifier with AUC of 0.96 (95% CI: 0.95 - 0.98) vs. 0.82 (95% CI: 0.81-0.83) using PAM Kurtosis. Both ultrasound-derived models (US-CNN and

US-GLM) performed poorly with AUCs of 0.71 (95% CI: 0.63 – 0.78) and 0.66 (95% CI: 0.65 – 0.67), respectively. While easier to train and validate and requiring smaller data sets, GLM diagnostic performance is inferior to CNN models.

Our study has a significant impact in rectal cancer treatment management. The PAM/US endoscopy paired with CNNs has a great potential to improve curent standard of care imaging in accurately predicting complete pathological response (pCR) of rectal cancer post-treatment. For those who have achieved a pCR, unnecessary surgery can be avoided without compromising cancer-related outcomes, and thereby lowering morbidity and health care cost.

Our study has limitations. First, the patient data is limited. With more patient data available, the diagnostic performance of PAM-CNN models can be further improved. For example, in our current study, 1-D ROIs from PAM and US B-scans were used as input images to CNNs. Misclassifications can occur in ROIs' when SNRs are low. 2-D ROIs from a small number of sequential B-scans can be trained together to reduce the dependence of CNNs on the SNR of individual 1-D ROIs and further improve the performance of CNNs. Second, the quality of ex-vivo data was not as good as in-vivo data which can be seen from slightly lower training/validation PAM-GLM data compared with testing results of PAM-GLM. Future studies will be focused on recruiting more patients to the study to further validate the initial results reported in this manuscript.

References

1. R. L. Siegel et al., "Colorectal cancer statistics, 2020," CA: A cancer journal for clinicians," 70(3), 145-164, (2020).

2. A. G. Renehan et al., “Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis,” The Lancet Oncology, 17(2), 174-183, (2016)..
3. F. Dossa et al., “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 Gastroenterology & Hepatology, 2(7), 501-513, (2017).
4. J. C. Kong et al., “Outcome and salvage surgery following watch and wait for rectal cancer after neoadjuvant therapy: a systematic review,” Diseases of the Colon & Rectum, 60(3), 335-345, (2017).
5. J. Yahya et al, “Survey results of US radiation oncology providers’ contextual engagement of watch-and-wait beliefs after a complete clinical response to chemoradiation in patients with local rectal cancer,” Journal of gastrointestinal oncology, 9(6), 1127, (2018).
6. A. Habr-Gama et al., “Complete clinical response after neoadjuvant chemoradiation therapy for distal rectal cancer: characterization of clinical and endoscopic findings for standardization,” Dis Colon Rectum 2010 53(12):1692-1698, (2010)
7. I. M. Blazic et al., “MRI for evaluation of treatment response in rectal cancer,” The British journal of radiology, 89(1064), 20150964, (2016).
8. R. G. Beets-Tan et al., “Magnetic resonance imaging for clinical management of rectal cancer: updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting,” European radiology, 28(4), 1465-1475, (2018).

9. M. J. Gollub et al., "Use of magnetic resonance imaging in rectal cancer patients: Society of Abdominal Radiology (SAR) rectal cancer disease-focused panel (DFP) recommendations 2017," Abdominal Radiology, 43(11), 2893-2902, (2018).
10. D. M. Lambregts et al., "Diffusion-weighted MRI to assess response to chemoradiotherapy in rectal cancer: main interpretation pitfalls and their use for teaching," European radiology, 27(10), 4445-4454, (2017).
11. P. Marone et al., "Role of endoscopic ultrasonography in the loco-regional staging of patients with rectal cancer," World journal of gastrointestinal endoscopy, 7(7), 688, (2015).
12. S. Liu et al., "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 & Rectum, 61(8), 903-910, (2018).
13. X. Leng et al, "Feasibility of co-registered ultrasound and acoustic-resolution photoacoustic imaging of human colorectal cancer," Biomedical optics express, 9(11), 5159-5172, (2018).
14. X. Leng et al., "Assessing Rectal Cancer Treatment Response Using Coregistered Endorectal Photoacoustic and US Imaging Paired with Deep Learning," Radiology, 202208 (2021).
15. J. G. Laufer et al., "In vivo preclinical photoacoustic imaging of tumor vasculature development and therapy," Journal of biomedical optics, 17(5), 056016, (2012).
16. S. Hu & L. V. Wang, "Photoacoustic imaging and characterization of the microvasculature,". Journal of biomedical optics, 15(1), 011101, (2010).

17. M. Heijblom et al., “Visualizing breast cancer using the Twente photoacoustic mammoscope: what do we learn from twelve new patient measurements?,” Optics express, 20(11), 11582-11597, (2012).
18. M. Heijblom, et al., “Clinical photoacoustic breast imaging: the Twente experience,” IEEE pulse, 6(3), 42-46, (2015).
19. F. Raes et al., “High resolution ultrasound and photoacoustic imaging of orthotopic lung cancer in mice: new perspectives for onco-pharmacology,” PloS one, 11(4), e0153532, (2016).
20. D. K. Apriyanto & M. Satriawan, “CO2 Laser Photoacoustic Spectrometer for Measuring Acetone in the Breath of Lung Cancer Patients,”. Biosensors, 10(6), 55, (2020).
21. S. Nandy et al., “Evaluation of ovarian cancer: initial application of coregistered photoacoustic tomography and US,” Radiology, 289(3), 740-747, (2018).
22. C. P. Favazza et al., “In vivo photoacoustic microscopy of human cutaneous microvasculature and a nevus,” Journal of biomedical optics, 16(1), 016015, (2011).
23. D. Shen et al., “Deep learning in medical image analysis,” Annual review of biomedical engineering, 19, 221-248, (2017).
24. ANSI. "American National Standard for Safe Use of Lasers ANSI Z136. 1–2014.", (2014).

Appendix

Using highly correlated features to develop a classifier increases the chance of overfitting. To prevent this issue, we calculated the Spearman’s correlation between each two histogram features. Table S1 and S2 show the Spearman's correlation between each pair of histogram features of the PAM and US images, respectively. These tables are used to select an optimized feature set before developing GLM classifiers.

Table 7.S1: Spearman's correlation between histogram features of the PAM images

	Mean	**Std**	**Skewness**	**Kurtosis**	**Energy**
Mean	1	0.79	0.72	0.42	0.79
Std	-	1	0.75	0.57	0.52
Skewness	-	-	1	0.9	0.77
Kurtosis	-	-	-	1	0.5
Energy	-	-	-	-	1

Table 7.S2: Spearman's correlation between histogram features of the US images

	Mean	**Std**	**Skewness**	**Kurtosis**	**Energy**
Mean	1	0.7	0.63	0.38	0.78
Std	-	1	0.15	0.35	0.9
Skewness	-	-	1	0.69	0.41
Kurtosis	-	-	-	1	0.52
Energy	-	-	-	-	1

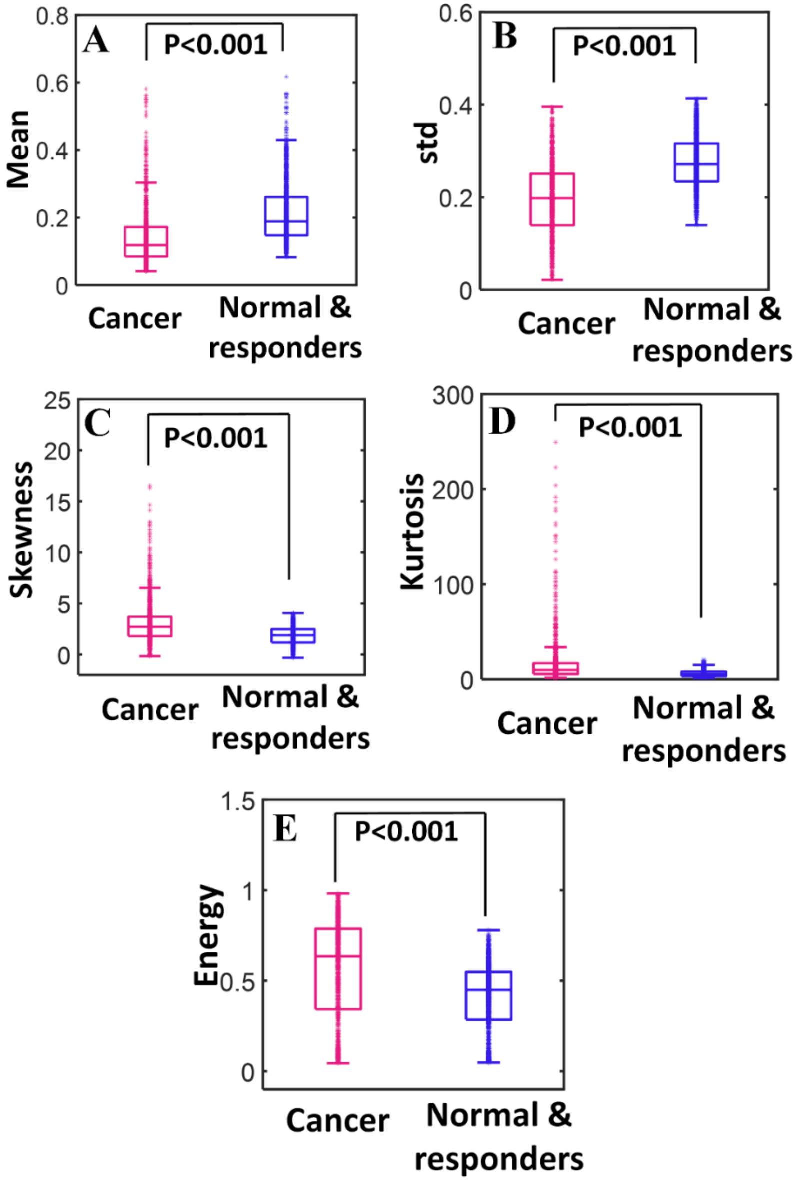


Figure 7.1S. Boxplots of histogram features (Y axes) of PAM images. Each plotted point represents the histogram feature in one ROI. The p-value for each feature is shown on the plot.

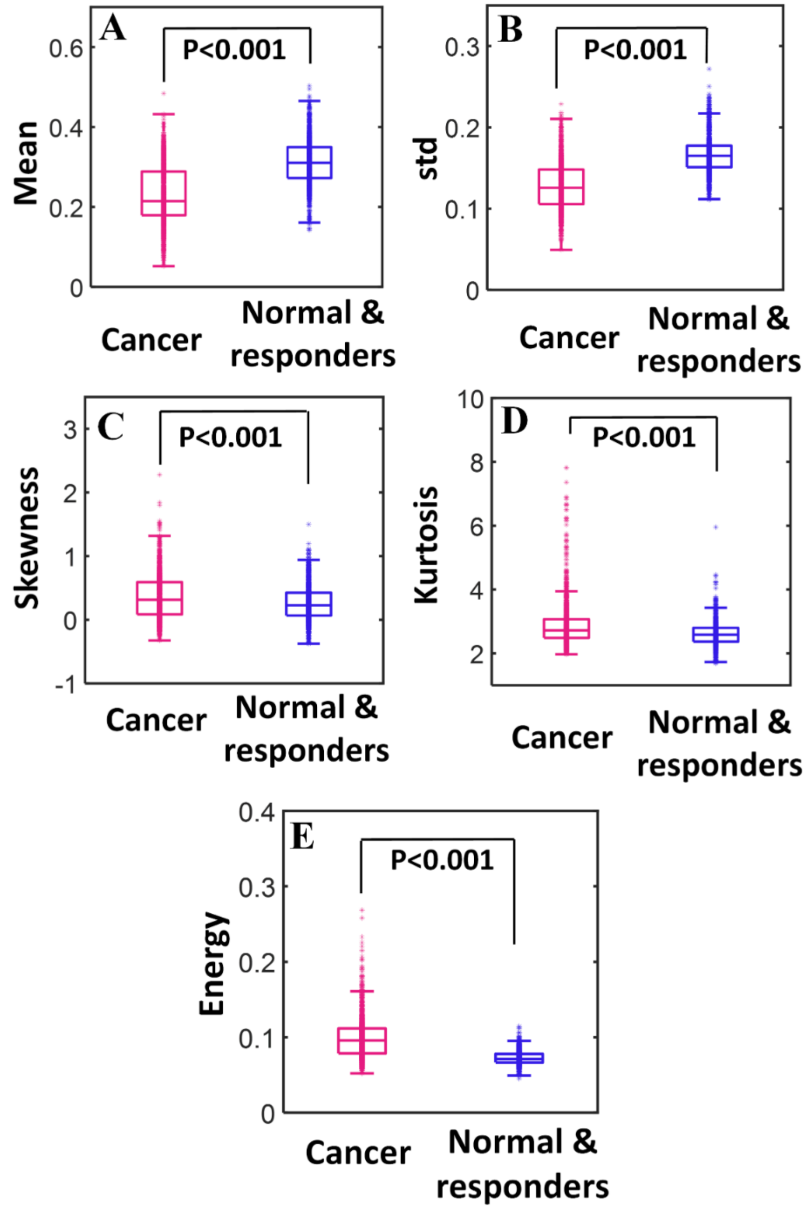


Figure 7.2S. Boxplots of histogram features (Y axes) of US images. Each plotted point represents the histogram feature in one ROI. The p-value for each feature is shown on the plot

Chapter 8: Summary and Future Work

8.1 Summary

In this dissertation, we focused on ovarian and colorectal cancer diagnosis using coregistered ultrasound and photoacoustic tomography (PAT/US) system. We also showed the capability of our PAM/US system in colorectal cancer diagnosis. The dissertation started by introducing some basic knowledge and statistics about ovarian and colorectal cancer, as well as the most common tools for detecting these cancers. We then introduced photoacoustic effect and different modalities which are based on this effect (PAT and PAM).

Next, we introduced our phantom construction procedure which can be used for pre-clinical studies of ultrasound and photoacoustic tomography. In this gelatin-based phantom, we use different concentrations of evaporated milk to generate desired ultrasound attenuation and optical scattering in the phantom. Moreover, alcohol is used to achieve a speed of sound close to real tissue.

We then discussed our in-vivo study on ovarian cancer diagnosis using our PAT/US system. In that study, we successfully classified 12 malignant and 27 ovarian ovaries using our developed GLM and SVM classifiers. These classifiers were trained using spectral, image, and functional features extracted from our PAT data. To extract functional features, the well-known linear unmixing method at each single pixel was employed. This method has three problems: first, when noise and motion artifact exist in our data, the computed functional maps are not reliable; second, the computed maps are not just a function of the concentrations of chromophores, but also a function of PAT/US system, Grüneisen coefficient, and fluence; third, it assumes that optical fluence is wavelength independent. We mitigated the effect of the first problem using sliding multi-pixel method and the second problem using a two-step optimization algorithm. Solving the third problem is very challenging and is an ongoing project.

We then sifted our focus to colorectal cancer diagnosis. Two studies were conducted relating to this subject. First, our PAT/US system was employed to collect data from ex-vivo colorectal cancer samples. The effects of different treatments on colorectal lesions were quantitatively presented, and the capability of features extracted from PAT data in distinguishing colon samples was studied (AUC of testing data set = 0.93). Finally, in another study on colorectal cancer diagnosis, the data from PAM/US system were used to classify rectal lesions in-vivo. We showed that a CNN classifier would significantly outperform the performance of a GLM classifier (AUC_CNN= 0.96 vs AUC_GLM=0.82) in distinguish rectal lesions.

8.2 Future Work

8.2.1 end-to-end deep learning model to estimate functional features from PAT data

As mentioned before, one of the limitations of our current method for evaluating functional parameters using PAT is that we assume that optical fluence is a wavelength independent parameter. To remove this assumption from our analysis, we are working on an end-to-end deep learning model (U-net) to estimate optical absorption at each wavelength from the reconstructed PA images at that wavelength. The input of this model is the reconstructed initial pressure, and the output is absorption coefficient map. We have used digital phantoms with targets of different shapes, sizes and locations to train the deep learning model (Figure 8.1). Monte Carlo [1] and k-wave [2] toolboxes are used to estimate the optical fluence and reconstructed PA images from the phantoms. We are also making several real phantoms (using the procedure mentioned in chapter 2) to test the performance of our classifiers. Once we assure that our model works well on our phantom data, we apply this model on our clinical data to estimate functional feature more accurately.

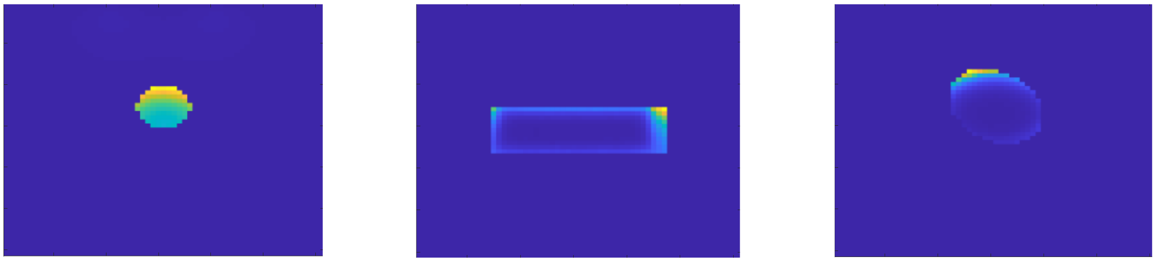


Figure 8.1: Examples of created digital phantoms used to train our deep learning model.

8.2.2 A generalized linear model to detect invalid sO2 maps of the ovarian areas calculated via photoacoustic tomography

To generate an sO2 map, we independently solve a linear equation at each pixel of that map. PAT data at at least three wavelengths are required to get a robust sO2 at each pixel. These PAT data should be higher than the noise level of the system to get a reliable estimate of sO2 at each pixel. Some ovarian areas, such as cystic areas, have very low optical absorptions. This result in low PA signals (less than the noise level) in most of the pixels of the map, so sO2 is not calculated for these pixels. The remaining pixels appear as few random scattered points that although they satisfy our condition of sO2 calculation (PAT signal>noise level), their signal levels are still not high enough to convince us to rely on the estimated sO2 values. Detecting these maps from imaged ovarian areas using a GLM will the aim of this study. First row of Figure 8.2 shows the maps of some examples of accepted sO2 maps, and the second row shows examples of those that should not be considered for further data analysis.

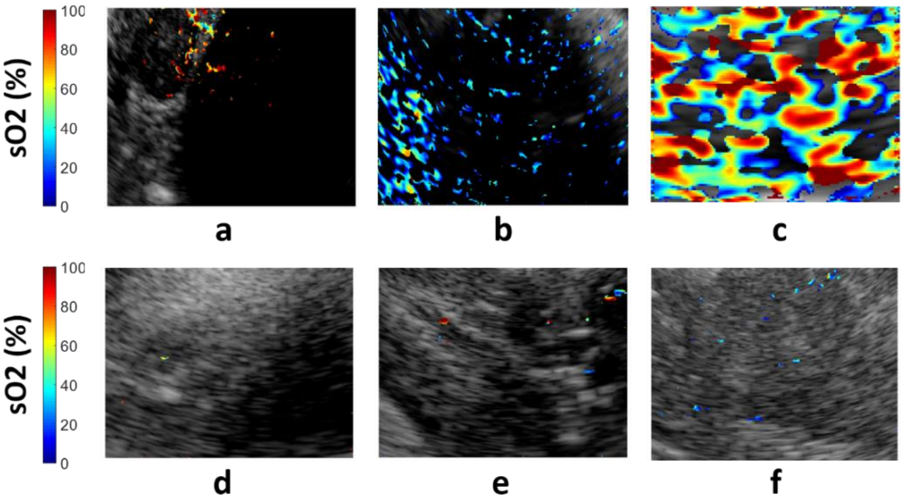


Figure 8.2: Coregistered US and sO2 maps of 3 valid (a-c) and 3 invalid maps (d-f).

References

1. B. E. Treeby & B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave-fields," J. Biomed. Opt., vol. 15, no. 2, p. 021314, (2010).

2. Q. Fang & D. Boas, "Monte Carlo Simulation of Photon Migration in 3D Turbid Media Accelerated by Graphics Processing Units," Opt. Express, vol. 17, issue 22, pp. 20178-20190 (2009).