2019

Classification of human ovarian cancer using functional, spectral, and imaging features obtained from in vivo photoacoustic imaging

Eghbal Amidi
Washington University in St. Louis

Atahar Mostafa
Washington University in St. Louis

Sreyankar Nandy
Washington University in St. Louis

Guang Yang
Washington University in St. Louis

William Middleton
Washington University School of Medicine in St. Louis

See next page for additional authors

Research Article | Vol. 10, No. 5 | 1 May 2019 | BIOMEDICAL OPTICS EXPRESS 2203
Biomedical Optics EXPRESS

Classification of human ovarian cancer using functional, spectral, and imaging features obtained from in vivo photoacoustic imaging

EGHBAL AMIDI,[1] ATAHAR MOSTAFA,[1] SREYANKAR NANDY,[1] GUANG YANG,[1] WILLIAM MIDDLETON,[2] CARY SIEGEL,[2] AND QUING ZHU[1,2,*]

[1]*Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA*
[2]*Department of Radiology, Washington University School of Medicine, St. Louis, MO 63110, USA*
**Zhu.q@wustl.edu*

Abstract: We report in this pilot study the diagnostic results of in vivo imaging of patients with ovarian lesions, using a co-registered photoacoustic and ultrasound (PAT/US) system. A total of 39 ovaries from 24 patients were imaged in vivo. PAT functional features, i.e., blood oxygen saturation (sO2) and relative total hemoglobin (rHbT), PAT image features, and PAT spectral features within a region of interest (ROI) in each ovarian tissue were extracted. To select the significant features, a t-test on each feature was performed, and the independent predictors were determined by evaluating correlation between each pair of predictors. To classify the ovarian lesions, we employed a generalized linear model (GLM) and a support vector machine (SVM). We used 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. We developed classifiers once by inclusion of PAT functional features to assess the best diagnostic performance of the classifiers when multiple wavelengths data are available. Second time, we excluded 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 for distinguishing normal/benign ovarian lesions from all types of tumors. In this case, an area under ROC curve (AUC) of 0.92, 0.93 of testing data was achieved using a GLM and SVM classifier when functional features were included in the feature set while excluding these features resulted in an AUC of 0.89, 0.92, respectively.

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 will be diagnosed in the United States, and about 13,980 of these women will die 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

#356171 https://doi.org/10.1364/BOE.10.002303
Journal © 2019 Received 3 Jan 2019; revised 17 Mar 2019; accepted 20 Mar 2019; published 8 Apr 2019

Biomedical Optics EXPRESS

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 relative total hemoglobin (rHbT) and blood oxygen saturation (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 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.

2. Materials and methods

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.

Biomedical Optics EXPRESS

2.2 PAT functional features

The relative total hemoglobin concentration ($rHbT(r,\theta)$) at each pixel in the region of interest (ROI) was calculated as the summation of the relative oxy hemoglobin concentration ($rHbO_2(r,\theta)$) and relative deoxy hemoglobin concentration ($rHb(r,\theta)$) at that pixel. The relative oxy and deoxy hemoglobin concentration values can be approximated as [19]

$$rHbO_2(r,\theta) = \bar{C}(r,\theta)HbO_2(r,\theta), \text{and} \tag{1}$$

$$rHb(r,\theta) = \bar{C}(r,\theta)Hb(r,\theta), \tag{2}$$

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 oxygen saturation (*sO2*) at each pixel was calculated by dividing the oxy hemoglobin by the sum of the oxy and deoxy hemoglobin:

$$sO2(r,\theta) = \frac{HbO_2(r,\theta)}{HbO_2(r,\theta) + Hb(r,\theta)} \times 100\%. \tag{3}$$

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.

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.

Biomedical Optics EXPRESS

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 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.

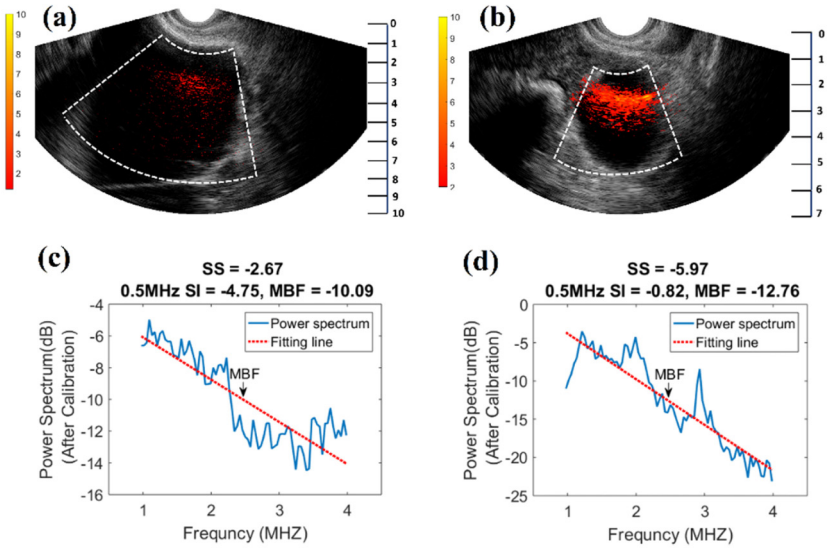

Fig. 1. Top row: co-registered rHbT and US images of a benign mucinous cystadenoma (a) and a high-grade serous carcinoma (b). The vascular distribution of the benign lesion is more scattered, but more localized and intense for the malignant ovary. Bottom row: the calibrated PAT power spectra and their fitted lines in the regions associated with the angular dashed lines in each image. Note the different Y-axis depth ranges.

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 (Fig. 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 [25]. 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

Biomedical Optics EXPRESS

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:

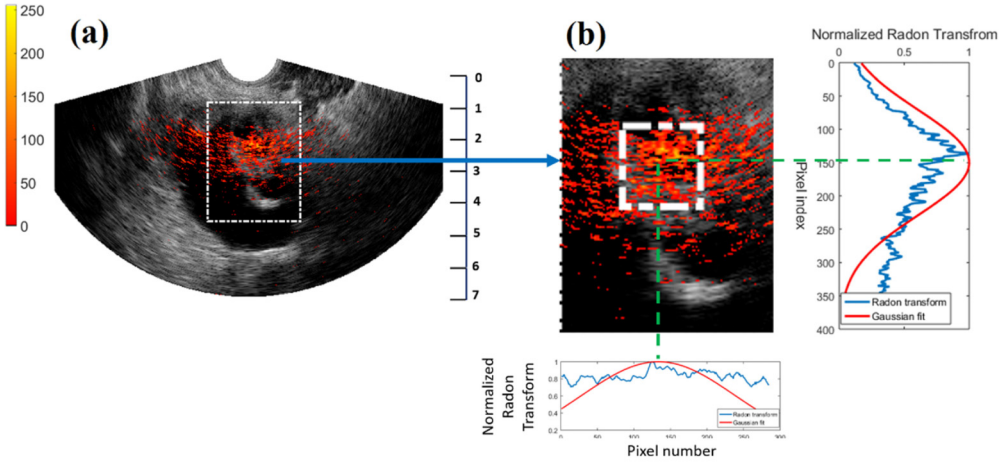

Fig. 2. ROI selection for image analysis. A larger rectangular region associated with the ovarian tissue is first selected (a). After that, the Radon transforms of the image at angles of 0 and 90 degrees in the selected area are calculated (b). These Radon transforms are then normalized, and a Gaussian 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.

$$Contrast = \sum_{|i-j|=0}^{N-1} |i-j|^2 \sum_{i=1}^{N}\sum_{j=1}^{N} c(i,j), \tag{4}$$

$$Correlation = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(i-\mu_i)(j-\mu_j)c(i,j)}{\sigma_i\sigma_j}, \tag{5}$$

$$Energy = \sum_{i=1}^{N}\sum_{j=1}^{N} c(i,j)^2, \text{and} \tag{6}$$

$$Homogeneity = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{c(i,j)^2}{1+|i-j|}, \tag{7}$$

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 [25].

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].

Biomedical Optics EXPRESS

2.5 Classification

We imaged 24 patients (39 ovaries) using our hybrid PAT/US system (see Table 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 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 performed a cross-validation of 200 times. 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

Biomedical Optics EXPRESS

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). 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. Results and discussion

3.1 PAT spectral features

In Fig. 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 ($P<0.05$).

3.2 PAT image features

Co-registered PAT and US images of a set of benign fibrothecoma and high-grade serous carcinoma are shown in Fig. 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 Classification by inclusion of functional features in the features set

3.3.1. 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 Fig. 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.

Biomedical Optics EXPRESS

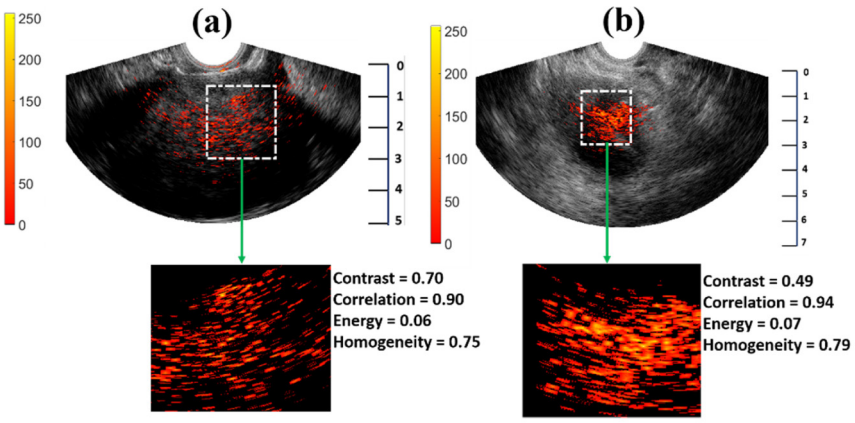

Fig. 3. Co-registered PAT and US images and magnification of the PAT images in the areas indicated by the dashed rectangle for a benign fibrothecoma (a) and an ovary with epithelial cancer (b). The values of the textural features for each image are also shown.

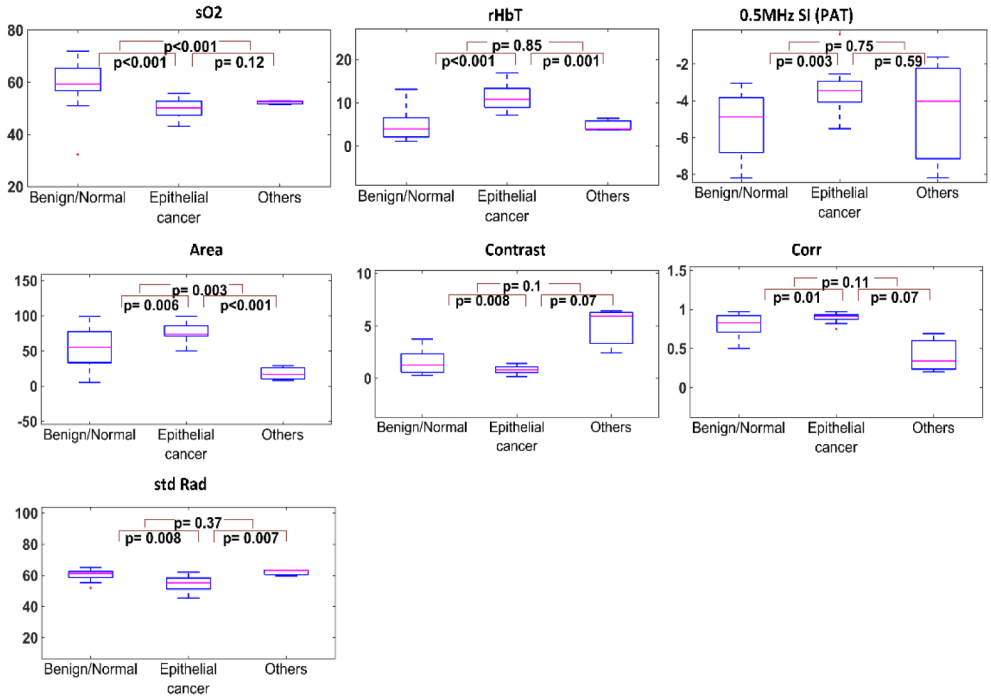

Fig. 4. Box plots of the significant features for the three groups of ovaries. For each feature, the p-value between each pair of the three groups is shown in the plots.

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 2. As can be seen, rHbT is at the top of the ranking, based on

Biomedical Optics EXPRESS

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. 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 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. 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 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 Fig. 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.

Biomedical Optics EXPRESS

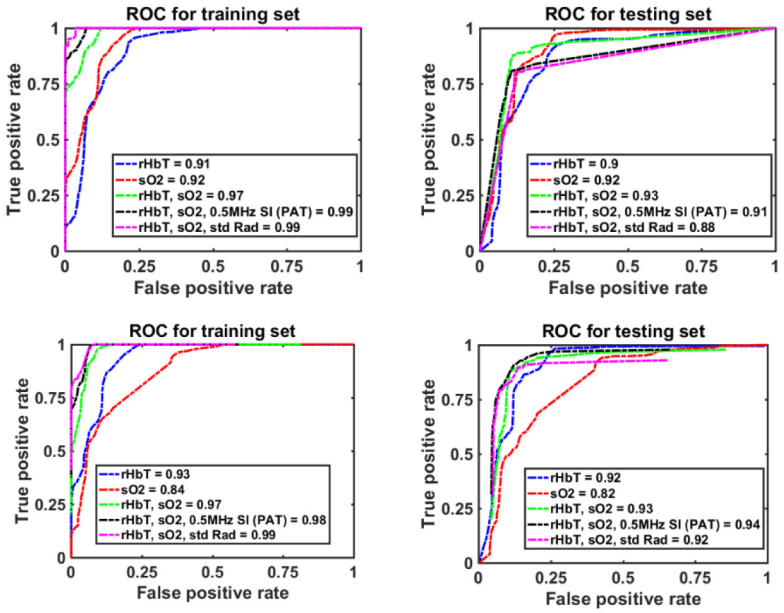

Fig. 5. ROCs for the training (left) and testing (right) data sets and the associated AUCs for different feature sets, using the GLM (upper row) and SVM (lower row) classifiers. The classifiers distinguish benign/normal from epithelial cancers.

3.3.2. 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 Fig. 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 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 5.

Figure 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.

Biomedical Optics EXPRESS

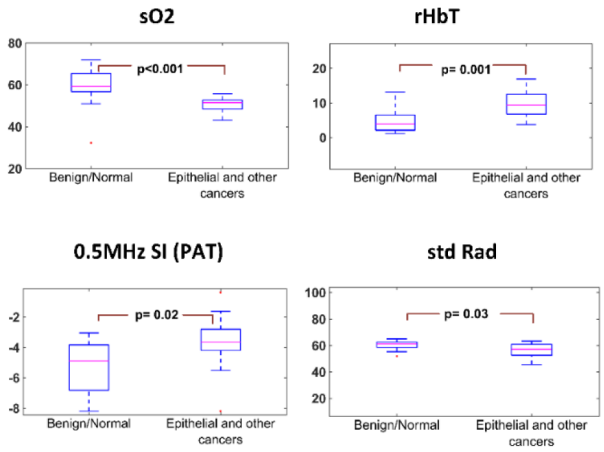

Fig. 6. Box plots of the significant features for two groups of ovaries. For each feature, the p-value between the two groups is shown.

Table 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 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 Fig. 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.

3.4 Classification by exclusion of functional features from 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.

Biomedical Optics EXPRESS

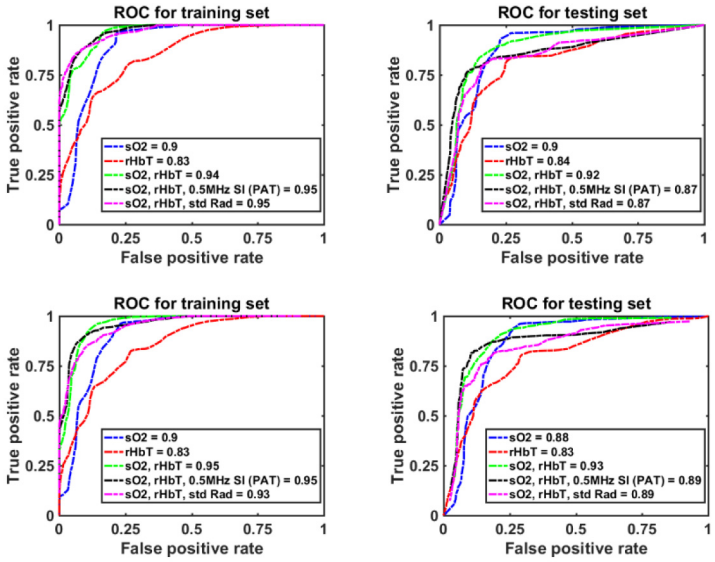

Fig. 7. ROC for the training (left) and testing (right) data sets and the associated AUCs for different feature sets, using the GLM (upper row) and SVM (lower row) classifiers. The classifiers distinguish benign/normal ovarian masses from epithelial cancer and other neoplasms.

3.4.1. 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 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.

3.4.2. 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 Fig. 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.

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.

Biomedical Optics EXPRESS

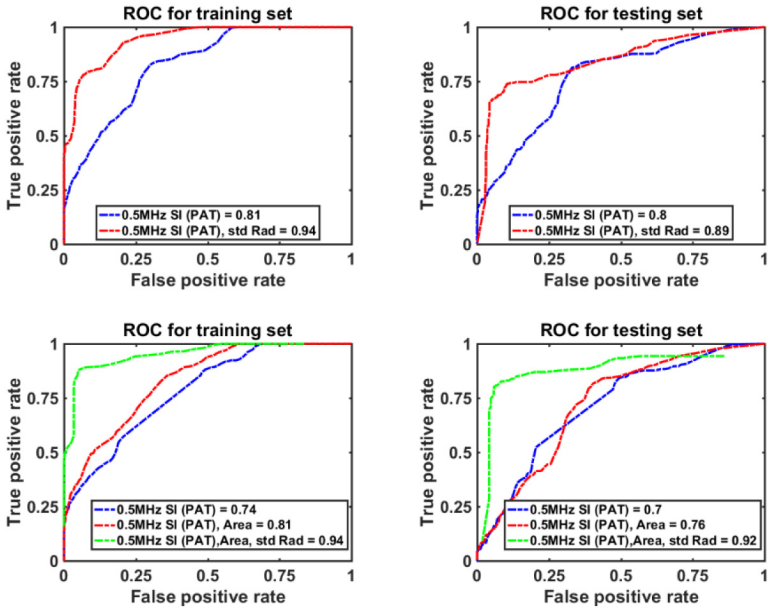

Fig. 8. ROC for the training (left) and testing (right) data sets and the associated AUCs for different feature sets, using the GLM (upper row) and SVM (lower row) classifiers. The classifiers distinguish benign/normal ovarian masses from epithelial cancers. Functional features are not included in the features set.

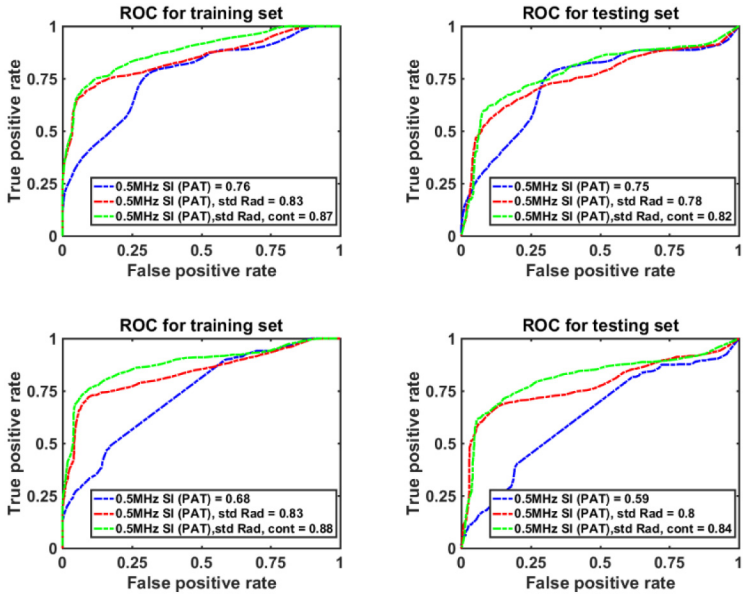

Fig. 9. ROC for the training (left) and testing (right) data sets and the associated AUCs for different feature sets, using the GLM (upper row) and SVM (lower row) classifiers. The classifiers distinguish benign/normal ovarian masses from epithelial cancers and other neoplasms. Functional features are not included in the feature set.

The training and testing data sets are small (26 for training and 13 for testing), and overfitting can occur when the training data set is limited [26]. We have selected the minimal number of independent predictors for each prediction model, performed cross-validation, and used fairly high amounts (33%) of the patient data for testing. The performances of the

Biomedical Optics EXPRESS

prediction models based on respective training and testing data sets are similar with no obvious pattern of higher AUC values for training data and much lower AUCs for testing data, which would be expected if there were the problem of overfitting. With more patients recruited to the study, we will be able to establish a large database to validate the prediction models with more input predictors.

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 tHb, 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.

Funding

NCI (R01CA151570); partial support from NIBIB (R01EB002136).

Acknowledgements

We thank study coordinators Ruth Holdener and Lynne Lippmann for coordinating study schedules, identifying and consenting patients to the study. We thank Stacey Reime, Radiology Technical Supervisor for ultrasound, for her help on ultrasound imaging of patients. We thank the entire GYN oncology group for helping identify patients, and the Radiology US group for helping with US scans. The authors also thank James Ballard for editing the manuscript.

Disclosures

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

References

1. American Cancer Society, Cancer facts and figures, 2017, American Cancer Society, Atlanta, GA (2017)
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**(2), 170–177 (2009).
4. K. R. Lee, F. A. Tavassoli, J. Prat, M. Dietel, D. J. Gersell, and A. I. Karseladze, Surface Epithelial Stromal Tumours: Tumours of the Ovary and Peritoneum. FA. Tavassoli, P. Devilee, IARC Press: Lyon; 117–145 (2003).
5. N. J. Finkler, B. Benacerraf, P. T. Lavin, C. Wojciechowski, and R. C. Knapp, "Comparison of serum CA 125, clinical impression, and ultrasound in the preoperative evaluation of ovarian masses," Obstet. Gynecol. **72**(4), 659–664 (1988).

Biomedical Optics EXPRESS

6. I. Jacobs, D. Oram, J. Fairbanks, J. Turner, C. Frost, and J. G. Grudzinskas, "A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer," Br. J. Obstet. Gynaecol. **97**(10), 922–929 (1990).
7. B. Van Calster, D. Timmerman, T. Bourne, A. C. Testa, C. Van Holsbeke, E. Domali, D. Jurkovic, P. Neven, S. Van Huffel, and L. Valentin, "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, M. Amneus, F. Su, J. Lang, J. M. Janco, S. T. Reddy, and R. Farias-Eisner, "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. D. X. Chen, P. E. Schwartz, X. G. Li, and Z. Yang, "Evaluation of CA 125 levels in differentiating malignant from benign tumors in patients with pelvic masses," Obstet. Gynecol. **72**(1), 23–27 (1988).
10. A. Shaaban and M. Rezvani, "Ovarian cancer: detection and radiologic staging," Clin. Obstet. Gynecol. **52**(1), 73–93 (2009).
11. R. E. Bristow, R. L. Giuntoli 2nd, H. K. Pannu, R. D. Schulick, E. K. Fishman, and R. L. Wahl, "Combined PET/CT for detecting recurrent ovarian cancer limited to retroperitoneal lymph nodes," Gynecol. Oncol. **99**(2), 294–300 (2005).
12. Y. Nakamoto, T. Saga, and S. Fujii, "Positron emission tomography application for gynecologic tumors," Int. J. Gynecol. Cancer **15**(5), 701–709 (2005).
13. L. V. Wang and S. Hu, "Photoacoustic tomography: in vivo imaging from organelles to organs," Science **335**(6075), 1458–1462 (2012).
14. K. S. Valluru, K. E. Wilson, and J. K. Willmann, "Photoacoustic imaging in oncology: translational preclinical and early clinical experience," Radiology **280**(2), 332–349 (2016).
15. A. A. Oraevsky and A. A. Karabutov, Optoacoustic tomography. In: Vo-Dinh T, ed. Biomedical photonics handbook. Boca Raton, Fla: CRC, (2003).
16. H. Li, P. Kumavor, U. Salman Alqasemi, and Q. Zhu, "Utilizing spatial and spectral features of photoacoustic imaging for ovarian cancer detection and diagnosis," J. Biomed. Opt. **20**(1), 016002 (2015).
17. H. S. Salehi, H. Li, A. Merkulov, P. D. Kumavor, H. Vavadi, M. Sanders, A. Kueck, M. A. Brewer, and Q. Zhu, "Coregistered photoacoustic and ultrasound imaging and classification of ovarian cancer: ex vivo and in vivo studies," J. Biomed. Opt. **21**(4), 046006 (2016).
18. S. Nandy, A. Mostafa, I. S. Hagemann, M. Powell, E. Amidi, K. Robinson, D. Mutch, C. Siegel, and Q. Zhu, "Role of Co-Registered Photoacoustic and Ultrasound Tomography: Initial Application for Evaluation of Ovarian Cancer," Radiology **289**(3), 740–747 (2018).
19. B. Cox, J. G. Laufer, S. R. Arridge, and P. C. Beard, "Quantitative spectroscopic photoacoustic imaging: a review," J. Biomed. Opt. **17**(6), 061202 (2012).
20. T. Feng, J. E. Perosky, K. M. Kozloff, G. Xu, Q. Cheng, S. Du, J. Yuan, C. X. Deng, and X. Wang, "Characterization of bone microstructure using photoacoustic spectrum analysis," Opt. Express **23**(19), 25217–25224 (2015).
21. G. Xu, Z. X. Meng, J. D. Lin, J. Yuan, P. L. Carson, B. Joshi, and X. Wang, "The functional pitch of an organ: quantification of tissue texture with photoacoustic spectrum analysis," Radiology **271**(1), 248–254 (2014).
22. X. Leng, W. Chapman, Jr., B. Rao, S. Nandy, R. Chen, R. Rais, I. Gonzalez, Q. Zhou, D. Chatterjee, M. Mutch, and Q. Zhu, "Feasibility of co-registered ultrasound and acoustic-resolution photoacoustic imaging of human colorectal cancer," Biomed. Opt. Express **9**(11), 5159–5172 (2018).
23. R. E. Kumon, C. X. Deng, and X. Wang, "Frequency-domain analysis of photoacoustic imaging data from prostate adenocarcinoma tumors in a murine model," Ultrasound Med. Biol. **37**(5), 834–839 (2011).
24. E. Hysi, L. A. Wirtzfeld, J. P. May, E. Undzys, S. D. Li, and M. C. Kolios, "Photoacoustic signal characterization of cancer treatment response: Correlation with changes in tumor oxygenation," Photoacoustics **5**(5), 25–35 (2017).
25. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Trans. Syst. Man Cybern. **3**(6), 610–621 (1973).
26. J. Concato, A. R. Feinstein, and T. R. Holford, "The risk of determining risk with multivariable models," Ann. Intern. Med. **118**(3), 201–210 (1993).