Biomedical Optics EXPRESS

A three-parameter logistic model to characterize ovarian tissue using polarization-sensitive optical coherence tomography

Tianheng Wang, Yi Yang, and Quing Zhu*

University of Connecticut, Dept. of Electrical and Computer Engineering, Storrs, CT 06269, USA
**zhu@engr.uconn.edu*

Abstract: In this paper, a logistic prediction model is introduced to characterize the ovarian tissue. A new parameter, the phase retardation rate, was extracted from phase images of polarization-sensitive optical coherence tomography (PS-OCT). Statistical significance of this parameter between normal and malignant ovarian tissues was demonstrated ($p<0.0001$). Linear regression analysis showed that this parameter was positively correlated ($R = 0.74$) with collagen content, which was associated with the development of ovarian tissue malignancy. When this parameter and the optical scattering coefficient and the phase retardation estimated from the 33 ovaries were used as input predictors to the logistic model, 100% sensitivity and specificity in classifying malignant and normal ovaries were achieved. Ten additional ovaries were imaged and used to validate the prediction model and 100% sensitivity and 83.3% specificity were achieved. These results showed that the three-parameter prediction model based on quantitative parameters estimated from PS-OCT images could be a powerful tool to detect and diagnose ovarian cancer.

OCIS codes: (110.4500) Optical coherence tomography; (170.4500) Optical coherence tomography; (170.3880) Medical and biological imaging.

References and links

1. T. R. Rebbeck, H. T. Lynch, S. L. Neuhausen, S. A. Narod, L. Van't Veer, J. E. Garber, G. Evans, C. Isaacs, M. B. Daly, E. Matloff, O. I. Olopade, and B. L. Weber; Prevention and Observation of Surgical End Points Study Group, "Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations," N. Engl. J. Med. **346**(21), 1616–1622 (2002).
2. N. D. Kauff, J. M. Satagopan, M. E. Robson, L. Scheuer, M. Hensley, C. A. Hudis, N. A. Ellis, J. Boyd, P. I. Borgen, R. R. Barakat, L. Norton, M. Castiel, K. Nafa, and K. Offit, "Risk-reducing salpingo-oophorectomy in women with a BRCA1 or BRCA2 mutation," N. Engl. J. Med. **346**(21), 1609–1615 (2002).
3. W. A. Rocca, B. R. Grossardt, M. de Andrade, G. D. Malkasian, and L. J. Melton 3rd, "Survival patterns after oophorectomy in premenopausal women: a population-based cohort study," Lancet Oncol. **7**(10), 821–828 (2006).
4. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," Science **254**(5035), 1178–1181 (1991).
5. M. R. Hee, D. Huang, E. A. Swanson, and J. G. Fujimoto, "Polarization-sensitive low-coherence reflectometer for birefringence characterization and ranging," J. Opt. Soc. Am. B **9**(6), 903–908 (1992).
6. J. F. de Boer, T. E. Milner, M. J. C. van Gemert, and J. S. Nelson, "Two-dimensional birefringence imaging in biological tissue by polarization-sensitive optical coherence tomography," Opt. Lett. **22**(12), 934–936 (1997).
7. M. A. Brewer, U. Utzinger, J. K. Barton, J. B. Hoying, N. D. Kirkpatrick, W. R. Brands, J. R. Davis, K. Hunt, S. J. Stevens, and A. F. Gmitro, "Imaging of the ovary," Technol. Cancer Res. Treat. **3**(6), 617–627 (2004).
8. Y. Yang, T. Wang, N. C. Biswal, X. Wang, M. Sanders, M. Brewer, and Q. Zhu, "Optical scattering coefficient estimated by optical coherence tomography correlates with collagen content in ovarian tissue," J. Biomed. Opt. **16**(9), 090504 (2011).
9. Y. Yang, T. Wang, X. Wang, M. Sanders, M. Brewer, and Q. Zhu, "Quantitative analysis of estimated scattering coefficient and phase retardation for ovarian tissue characterization," Biomed. Opt. Express **3**(7), 1548–1556 (2012).
10. M. C. Pierce, R. L. Sheridan, B. Hyle Park, B. Cense, and J. F. de Boer, "Collagen denaturation can be quantified in burned human skin using polarization-sensitive optical coherence tomography," Burns **30**(6), 511–517 (2004).
11. Y. Chen, L. Otis, D. Piao, and Q. Zhu, "Characterization of dentin, enamel, and carious lesions by a polarization-sensitive optical coherence tomography system," Appl. Opt. **44**(11), 2041–2048 (2005).

#187788 - $15.00 USD Received 26 Mar 2013; revised 18 Apr 2013; accepted 18 Apr 2013; published 29 Apr 2013

12. D. Kleinbaum, L. Kupper, K. Muller, and A. Nizam, *Applied Regression Analysis and Other Multivariable Methods*, 3rd ed. (Duxbury Press, 1998).
13. X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez, and M. Müller, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," BMC Bioinformatics **12**(1), 77 (2011).

1. Introduction

Ovarian cancer has the highest mortality rate among all the gynecologic cancers because it is predominantly diagnosed at late stages due to the unreliable early symptoms and the poor screening techniques. Prophylactic oophorectomy (PO) has been accepted as the standard of care for high risk women and it reduces the risk of ovarian cancer by more than 50% [1,2]. However, there appears to be a higher mortality for premenopausal oophorectomy [3]. Therefore, there is an urgent need to develop effective tools to inspect ovaries, so that the mortality rate of ovarian cancer can be reduced and the quality of patients' life can be improved.

Optical coherence tomography (OCT) is an emerging high resolution imaging technique [4], which measures backscattered light generated from an infrared light source directed to the tissues. OCT typically achieves a resolution of several microns and a penetration depth of several millimeters, and has been used to image tissues in the body that can be accessed by endoscope or catheter. Polarization-sensitive OCT (PS-OCT) is a functional extension of OCT [5,6] and capable of detecting birefringence changes caused by collagen, and collagen changes in human ovary are indicators of malignancy [7,8]. Therefore, PS-OCT could be an effective tool to detect ovarian cancer. In our initial study [9], optical scattering coefficient and phase retardation of 33 *ex vivo* ovaries obtained from 18 patients were extracted from time domain (TD) PS-OCT intensity and phase images, respectively. While the scattering coefficient was significant in predicting malignancy, the phase retardation achieved low sensitivity of 43%. In this study, a more sensitive parameter, the phase retardation rate, was extracted from PS-OCT phase images and used together with the scattering coefficient and phase retardation to characterize ovarian tissue. In the literature, the PS-OCT phase retardation rate was introduced by M. C. Pierce et al. to quantify collagen denaturation in burned human skin [10]. In our study, these three parameters extracted from 33 ovaries were used as inputs to a logistic model to predict or classify the malignant and benign ovaries. In addition, 10 more ovaries from 5 patients were imaged with our upgraded Fourier domain (FD) PS-OCT system and used to test the model. To the best of our knowledge, this is the first study of using multiple parameters extracted from PS-OCT images as predictors for ovarian tissue characterization.

2. Materials and methods

2.1 Ovary sample and histopathology

A total of 43 ovaries were extracted from 23 patients undergoing PO at the University of Connecticut Health Center (UCHC). 33 ovaries from 18 patients were imaged using TD-PS-OCT while 10 ovaries from 5 patients were imaged using FD-PS-OCT. These patients were at risk for ovarian cancer or they had ovarian mass or pelvic mass suggesting malignancy. This study was approved by the Institutional Review Board of UCHC, and informed consent was obtained from all patients. The details of imaging procedures and histological processing were described in [9]. Sirius Red staining protocol was applied to the sectioned slides to analyze the collagen content. The amount of collagen was quantitatively analyzed using ImageJ software (National Institute of Health). The average collagen area fraction (CAF) was measured as "Stained collagen area/tissue area".

2.2 PS-OCT systems

The TD-PS-OCT and upgraded FD-PS-OCT systems are shown in Fig. 1(a). The essential optical configurations of the TD-PS-OCT and upgraded FD-PS-OCT systems are the same. The technical details of the TD-PS-OCT system were described in our earlier study [9]. The main differences between the upgraded FD system and the TD system are: (1) the super

#187788 - $15.00 USD Received 26 Mar 2013; revised 18 Apr 2013; accepted 18 Apr 2013; published 29 Apr 2013

luminescent diode source was replaced with an 110 nm bandwidth swept source (HSL-2000, Santec Corp., Japan) with center wavelength of 1310 nm and scan rate of 20 kHz; (2) the detectors were replaced with 75MHz bandwidth photodetectors (Thorlabs PDB120C); (3) the reference mirror was fixed instead of moving back and forth by a stepper motor. The conventional OCT intensity images were obtained from calculating the summation of squares of two orthogonally polarized signals, and the phase retardation images were obtained by calculating arctangent between vertical and horizontal polarization components [11].

2.3 Phase retardation rate

During imaging, similar conditions for all ovary samples were obtained by mounting the ovary on a three-dimensional stage and adjusting the ovarian tissue surface to the same depth position. The phase retardation rate was obtained by linearly fitting phase retardation depth profile. The region of interest (ROI) selection was consistent with that in our earlier study [9] when calculating scattering coefficient and phase retardation. Overall, each image was evenly divided into several ROIs with 1mm width. Values in all ROIs from all images of one ovary were averaged to obtain the phase retardation rate of this ovary. The same procedures were followed for all ovaries. An example of fitting phase retardation rate of a normal ovary is shown in Figs. 1(b) and 1(c). Figure 1(b) is the phase retardation image, where the dark blue represents phase retardation value of zero degree and the dark red represents 90 degrees. The white dashed rectangular area was selected for fitting. The depth profile of the averaged A-lines in the selected area was shown as blue curve in Fig. 1(c), and the numerical fitting curve was plotted as red. The slope of the red curve was calculated as the phase retardation rate. The phase retardation decreases with depth after about 1.5mm. This is because the ratio of vertical and horizontal signals reduces as light penetrates deeper in the tissue. The fitting error of the phase retardation rate is estimated as the norm of the fitting residue divided by the norm of the original curve.

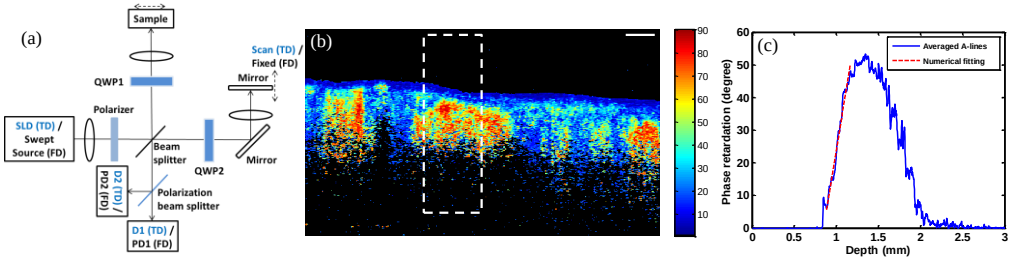

Fig. 1. (a) TD/FD-PS-OCT systems configuration. QWP: quarter-wave plate; PD: photodetector. (b) Phase retardation image; white dashed rectangular: selected area for fitting; scale bar: 0.5mm; (c) averaged A-lines and numerical fitting curve.

2.4 Optical scattering coefficient and phase retardation

The quantification of scattering coefficient and phase retardation were described in our earlier publication [9]. Scattering coefficient was estimated by numerically fitting compounded conventional OCT depth profile to a single scattering model based on Beer's law. 1mm tissue was averaged to minimize the speckle noise. The phase retardation was obtained by calculating the average phase values from PS-OCT phase images of the same area.

2.5 Logistic model and receiver operating characteristic curve

Logistic regression belongs to the class of generalized linear model (GLM) based on the exponential distribution family. It is a statistical model that can describe the relationship of several predictor variables X_1, X_2, …, X_k to a dichotomous response variable Y (0 or 1) [12]. The probability of occurrence of one of the two possible outcomes of Y can be described by the following equation:

#187788 - $15.00 USD Received 26 Mar 2013; revised 18 Apr 2013; accepted 18 Apr 2013; published 29 Apr 2013

$$pr(Y=1|X_1,X_2,...,X_K)=\frac{1}{1+\exp\left[-\left(\beta_0+\sum_{n=1}^{k}\beta_n X_n\right)\right]}$$

Given the data Y, X_1, X_2, …, X_k, the unknown coefficients $\beta_n, n=0,1,...,k$ can be estimated using the maximum likelihood method. In this paper, we use three predictors (scattering coefficient, phase retardation, and phase retardation rate) to classify normal and malignant ovarian tissue. The MATLAB GLMFIT function was used to fit the logistic model using the predictors and the response (0 represents normal and 1 represents malignant). The coefficients $\beta_n, n=0,1,...,k$ of the model that best follow the actual diagnosis were estimated and used to calculate the estimated responses (the numbers between 0 and 1) using GLMVAL function. The GLMFIT function also computed the deviance, which is a generalization of the residual sum of squares (comparison of log-likelihood function of actual fitted values with perfectly fitted values). The deviance was used to compare different prediction models, in which different parameter-combinations were used as predictors to classify normal and malignant ovaries. The deviance value decreases as the model fit improves.

The quality of the logistic prediction model was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC). The estimated responses from different prediction models were used to compute the ROC curves and AUCs using R package pROC [13]. We also estimated the 95% confidence interval (CI) using bootstrap method with 10,000 stratified bootstrap replicates. The optimal threshold provided by pROC was used to calculate the sensitivity and specificity, positive and negative predictive values (PPV, NPV). To further evaluate the logistic prediction model and testing results, we have also investigated the correlation coefficients R_{train} and R_{test} between calculated responses and the actual diagnosis.

3. Results and discussion

3.1 Statistical results of 33 ovaries imaged by TD-PS-OCT

A total of 33 *ex vivo* ovaries from 18 patients were imaged using TD-PS-OCT system. 26 ovaries were diagnosed as normal and 7 ovaries were diagnosed as malignant. Normal ovaries show higher average values of scattering coefficient and phase retardation than malignant ones, with the normal/malignant ratio of 1.36, 1.11, respectively [9]. For phase retardation rate, the average fitting range of normal group is 36.7-329.8 μm from the tissue surface, and the malignant group is 38.4-347.3 μm. The range of average value of normal group is 28.8-154.8 degree/mm, and malignant group is 8.4-121.6 degree/mm. The normal group has mean value of 79.5 degree/mm (± 19.0), which is higher than that of the malignant group with mean value of 45.0 degree/mm (± 19.6). The normal/malignant ratio of phase retardation rate is 1.77. Phase retardation rate of normal and malignant ovaries shows larger difference ($p < 0.0001$) than the other two parameters. The fitting error of the phase retardation rate of the normal and the malignant group is 5.13% (± 0.82%) and 4.69% (± 0.96%), respectively.

The scatter plot in Fig. 2(a) shows the average phase retardation rate of each ovary for normal and malignant groups. By setting a threshold of phase retardation rate at the value of 55 degree/mm, we could achieve 85.7% sensitivity and 92.3% specificity. However, by using phase retardation as a classifier, we could only achieve 42.9% sensitivity [9]. Linear regression analysis is shown in Fig. 2(b). A positive correlation was found between phase retardation rate and collagen content, with Pearson's correlation coefficient $R = 0.74$ ($p<0.0001$), which is higher than those from scattering coefficient ($R = 0.57$, $p<0.0001$) and phase retardation ($R = 0.47$, $p<0.01$) [9]. A multiple linear regression shows that those three parameters together positively correlate with collagen content with $R = 0.76$, which is higher than that using each parameter alone. Collagen is associated with the development of ovarian cancers; the collagen amount and structure are quite different between normal and malignant ovaries. The normal and malignant groups have CAF values of 46.0% (± 9.1%), and 28.4% (± 8.3%), respectively [9]. Since CAF, measured from Sirius Red staining on ovary samples,

#187788 - $15.00 USD Received 26 Mar 2013; revised 18 Apr 2013; accepted 18 Apr 2013; published 29 Apr 2013

directly assesses collagen, the highly positive correlation indicates that the phase retardation rate may measure the complicated process of collagen development of ovarian cancer.

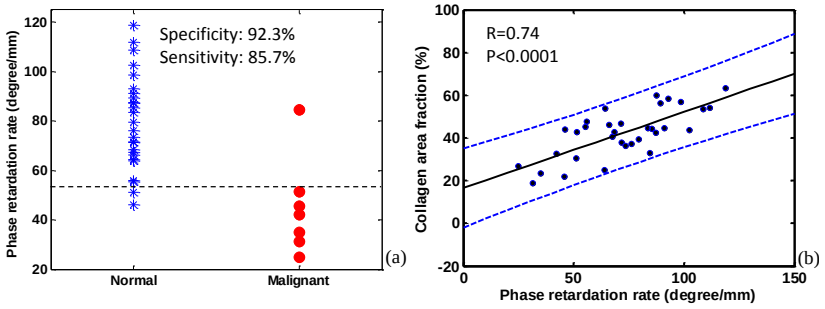

Fig. 2. (a) Scatter plot of phase retardation rate of normal and malignant ovary groups. (b) Positive correlation demonstration between phase retardation rate and collagen content; the blue dashed lines show 95% confidence interval.

3.2 Training results based on logistic model using the 33-ovary data

The three parameters extracted from 33-ovary images were used to train the logistic classifier. As shown by the ROC curves in Fig. 3(a), the use of all three parameters shows much better performance than each parameter alone. The more specific prediction results, including sensitivity, specificity, PPV, NPV, AUC (95% CI), correlation coefficient R_{train} between estimated responses and actual responses (p value), and deviance, of different parameter-combinations are summarized in Table 1. By using only one parameter as a predictor, none of the models could achieve perfect sensitivity and specificity. By using combinations of any two parameters except one set using phase retardation and phase retardation rate, or using three parameters as predictors, 100% sensitivity and specificity are achieved. The deviance of using three parameters together is smaller than that of using two parameters, which indicates that the three-parameter model is more reliable.

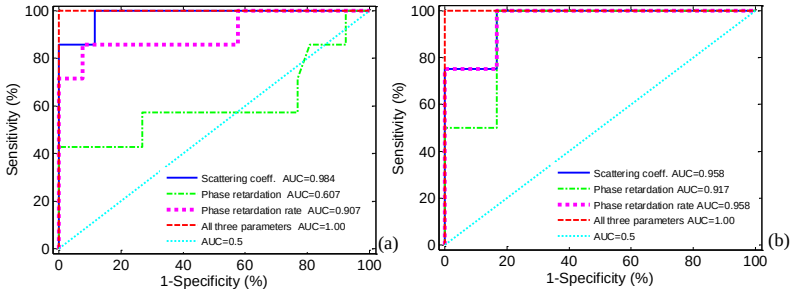

Fig. 3. ROC and AUC of different prediction models. (a) Training and (b) testing results.

3.3 Testing results of 10 ovaries imaged by FD-PS-OCT

10 ovaries (6 normal and 4 malignant) from 5 patients were imaged using the upgraded FD system and were tested using logistic prediction model based on different parameters described above. The testing results are summarized in Table 2. The same threshold of the training group was used for this testing group to calculate the sensitivity, specificity, PPV, and NPV. The R_{test} and AUC are highest (R_{test} = 0.893, $p<0.001$, AUC = 1.0) when using the three-parameter prediction model. Note that the three-parameter model achieved AUC = 1, but the sensitivity (100%) and specificity (83.3%) are not perfect. This is because we set a threshold of 0.5 for the training and testing groups for classifying normal and malignant ovaries. If we set a threshold of 0.7, we could achieve 100% sensitivity and specificity. However, in the three-parameter model training, all estimated responses of normal cases are

#187788 - $15.00 USD Received 26 Mar 2013; revised 18 Apr 2013; accepted 18 Apr 2013; published 29 Apr 2013

very close to normal response 0, and all estimated responses of malignant cases are very close to malignant response 1, so it makes more sense to set the middle point 0.5 as a threshold based on the training results.

Table 1. Summary of Logistic Model Results by Using Different Parameters[a]

Prediction model (Th)	Sensitivity	Specificity	PPV	NPV	AUC (95% CI)	R_{train} (p value)	Deviance
SC (0.65)	85.7%	100%	100%	96.3%	0.984 (0.934:1.0)	0.911 (<0.0001)	5.94
PR (0.30)	42.9%	100%	100%	86.7%	0.607 (0.302:0.885)	0.212 (0.237)	33.14
PRR (0.38)	85.7%	92.3%	75.0%	96.0%	0.907 (0.720:1.0)	0.747 (<0.0001)	18.92
SC + PR (0.50)	100%	100%	100%	100%	1.000 (1.0:1.0)	1.000 (<0.0001)	6.22e-13
SC + PRR (0.50)	100%	100%	100%	100%	1.000 (1.0:1.0)	1.000 (<0.0001)	1.94e-14
PR + PRR (0.18)	100%	92.3%	77.0%	100%	0.973 (0.912:1.0)	0.774 (<0.0001)	13.30
SC + PR + PRR (0.50)	100%	100%	100%	100%	1.000 (1.0:1.0)	1.000 (<0.0001)	1.60e-14

[a]SC: scattering coefficient, PR: phase retardation, PRR: phase retardation rate, Th: threshold.

Table 2. Summary of Testing Results

Prediction model (Th)	Sensitivity	Specificity	PPV	NPV	AUC (95% CI)	R_{test} (p value)
SC (0.65)	75.0%	83.3%	75.0%	83.3%	0.958 (0.833:1.0)	0.764 (0.010)
PR (0.30)	0%	100%	0%	60.0%	0.917 (0.667:1.0)	0.624 (0.054)
PRR (0.38)	50.0%	100%	100%	75.0%	0.958 (0.833:1.0)	0.751 (0.012)
SC + PR (0.50)	75.0%	100%	100%	85.7%	0.875 (0.625:1.0)	0.802 (0.005)
SC + PRR (0.50)	100%	83.3%	80.0%	100%	0.958 (0.750:1.0)	0.789 (0.007)
PR + PRR (0.18)	75.0%	83.3%	75.0%	83.3%	0.833 (0.500:1.0)	0.495 (0.146)
SC + PR + PRR (0.50)	100%	83.3%	80.0%	100%	1.000 (1.0:1.0)	0.893 (<0.001)

In this study, only 10 ovaries were tested using our logistic model, more ovary data are being collected to validate the initial results. Currently, because all parameter extraction and processing are offline, future work also includes automating our data processing procedures so that we could obtain these parameters and input them to the prediction model in real-time. For translating this technique from bench to bedside, a fiber-based PS-OCT system for *in vivo* evaluation of ovarian tissue needs to be developed.

4. Summary

The phase retardation rate quantitatively extracted from PS-OCT has significantly improved the ovarian cancer diagnosis when it is used together with optical scattering coefficient and phase retardation. By using a new three-parameter logistic prediction model, we achieve 100% sensitivity and specificity in the training group, and 100% sensitivity and 83.3% specificity in the testing group. The initial results demonstrate that the three-parameter prediction model based on PS-OCT could be a powerful tool to evaluate ovarian tissue.

Acknowledgments

This research was partially supported by the National Cancer Institute (1R01CA151570). The authors thank Xiaohong Wang, Dr. Molly Brewer and Dr. Melinda Sanders on helping with ovaries and pathology evaluations.