Optical identification of subjects at high risk for developing breast cancer

Paola Taroni
Giovanna Quarto
Antonio Pifferi
Francesca Ieva
Anna Maria Paganoni
Francesca Abbate
Nicola Balestreri
Simona Menna
Enrico Cassano
Rinaldo Cubeddu

SPIE

Optical identification of subjects at high risk for developing breast cancer

Paola Taroni,[a] Giovanna Quarto,[a] Antonio Pifferi,[a,b]
Francesca Ieva,[c] Anna Maria Paganoni,[c]
Francesca Abbate,[d] Nicola Balestreri,[e] Simona Menna,[d]
Enrico Cassano,[d] and Rinaldo Cubeddu[a,b]
[a]Politecnico di Milano, Dipartimento di Fisica, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[b]Istituto di Fotonica e Nanotecnologie, Consiglio Nazionale delle Ricerche, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[c]Politecnico di Milano, Dipartimento di Matematica, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[d]European Institute of Oncology, Breast Imaging Unit, Via G. Ripamonti, 435, 20141 Milano, Italy
[e]European Institute of Oncology, Department of Radiology, Via G. Ripamonti, 435, 20141 Milano, Italy

Abstract. A time-domain multiwavelength (635 to 1060 nm) optical mammography was performed on 147 subjects with recent x-ray mammograms available, and average breast tissue composition (water, lipid, collagen, oxy- and deoxy-hemoglobin) and scattering parameters (amplitude a and slope b) were estimated. Correlation was observed between optically derived parameters and mammographic density [Breast Imaging and Reporting Data System (BI-RADS) categories], which is a strong risk factor for breast cancer. A regression logistic model was obtained to best identify high-risk (BI-RADS 4) subjects, based on collagen content and scattering parameters. The model presents a total misclassification error of 12.3%, sensitivity of 69%, specificity of 94%, and simple kappa of 0.84, which compares favorably even with intraradiologist assignments of BI-RADS categories. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JBO.18.6.060507]

Breast cancer is a leading cause of death in women and a major health burden worldwide: one in eight women in the United States will be diagnosed with breast cancer in their lifetime.[1] Early diagnosis (tumor size <1 cm, no lymph node involvement) is a key element for complete response in the treatment of breast cancer with a five-year survival in the range of 93 to 99%.[1]

Breast density is a recognized strong and independent risk factor for breast cancer: high breast density involves a four to six times higher risk as compared to low density.[2,3] Several U.S. states have already recognized the importance of knowing whether a subject has high breast density, enacting laws that require mammography providers to add such notification in the summary of mammography report. Including breast density into risk prediction models has improved their prediction accuracy. The U.S. Preventive Services Task Force has also suggested the possibility of chemoprevention for women at high risk.[4] Thus, improved risk models could be used to better address not only closer screening of high-risk women but also prevention of breast cancer.

At present, breast density is assessed based on the radiological appearance of breast tissue (mammographic density). Thus it is known only at the first mammogram, typically at the age of 40 to 50, depending on the country.[5] A tool for its noninvasive estimation would allow the early identification of high-risk women, enabling the design of personalized screening and diagnostic paths. Due to the high incidence of breast cancer and effectiveness of interventions performed at an early stage, any significant improvement in the diagnostic procedure (especially an earlier diagnosis) would have a strong impact on both the number of spared lives and the quality of life.

Optical techniques can provide functional and structural information on biological tissue in an absolutely noninvasive way, and they have already been successfully applied to the characterization of breast tissue.[6–9] Also, extensive clinical trials showed that raw data on optical attenuation interpreted using principal component analysis strongly correlate with quantitative mammographic features.[10]

We have further exploited the potential of diffuse optical spectroscopy operating in the time domain to assess both tissue composition in terms of key constituents (water, lipids, collagen, and hemoglobin) and scattering parameters that are related to the overall structure of tissue at microscopic level and specifically to breast density.[11,12] Besides the noninvasive assessment of breast density,[13,14] these pieces of information can contribute to a better understanding of the role of mammographic density in breast cancer risk and may even provide a more specific link than x-ray measures with breast cancer risk.

In this work, we propose the use of time-resolved transmittance spectroscopy to identify noninvasively high breast density subjects who are at high risk for developing breast cancer.

Our portable clinical instrument for time-resolved optical mammography operates in transmittance geometry on the mildly compressed breast. Time-resolved transmittance data are collected at seven red and near-infrared wavelengths (i.e., 635, 680, 785, 905, 930, 975, and 1060 nm), using picosecond pulsed diode lasers as light sources, and two photomultiplier tubes and personal computer boards for time-correlated single photon counting to detect the time distributions of the transmitted pulses. Injection and collection fibers are scanned in tandem over the compressed breast and data are stored every millimeter. Images are routinely acquired from both breasts in cranio-caudal and oblique (45 deg) views. Time-resolved spectral data are interpreted with the solution of the diffusion equation for an infinite homogeneous slab, using a spectrally constrained global fitting procedure to estimate tissue composition in terms of oxy- and deoxyhemoglobin, water, lipid, and collagen content, as well as scattering parameters (amplitude a and power b).[15] Moreover, for the detection of breast lesions, scattering maps are routinely applied, together with late gated intensity images that are sensitive to spatial changes in the absorption properties. Details on the instrument setup and performances, and on the procedures for data acquisition and analysis, are reported in Ref. 16.

Address all correspondence to: Paola Taroni, Politecnico di Milano, Dipartimento di Fisica, Piazza Leonardo da Vinci 32, 20133 Milano, Italy. Tel: 39 02 23996109; Fax: 39 02 23996126; E-mail: paola.taroni@polimi.it

The instrument is presently applied in a clinical study approved by the institutional review board of the European Institute of Oncology. The study has a twofold aim: the optical characterization of malignant and benign breast lesions and the noninvasive assessment of breast density. The present work focuses on the latter aim. Thus, for each subject, all data from the four images (cranio-caudal and oblique views of both breasts) were averaged to provide the average optical properties and breast tissue composition of that subject. Data were collected from 179 patients, recruited between June 2009 and June 2012. Written informed consent was obtained from all of them. For 32 subjects, recent x-ray mammograms were not available; thus they were excluded from further analysis. General patient information for the remaining 147 subjects is as follows: age 52.2 ± 11.3 years, body mass index 23.4 ± 3.7 kg/m^2, 69 subjects in premenopausal status. An expert radiologist assigned Breast Imaging and Reporting Data System (BI-RADS) mammographic density categories as (1) almost entirely fat (category 1, $n = 19$); (2) scattered fibroglandular densities (category 2, $n = 37$); (3) heterogeneously dense (category 3, $n = 56$); and (4) extremely dense (category 4, $n = 35$).

The dependence of tissue composition and scattering parameters on mammographic density, classified through BI-RADS categories, was investigated. The results essentially confirm what we have observed previously on a more limited number of subjects.[13] Based on the Wilcoxon test, there is no statistically significant difference between BI-RADS categories 1 and 2 for any parameters but water, while the difference is highly significant for all parameters but oxygenation level (SO$_2$) in the case of BI-RADS 2 *versus* 3, and for all parameters but SO$_2$ and total hemoglobin content (tHb) in the case of BI-RADS 3 *versus* 4. Specifically, increasing breast density corresponds to progressively increasing average amounts of water and collagen, while the lipid content decreases gradually. An increase in BI-RADS categories is also observed in both scattering amplitude a and slope b, in agreement with differences in microscopic structures expected for fatty and fibroglandular tissue. The blood parameters (i.e., tHb and SO$_2$) are less sensitive, with only tHb showing a slight increase with mammographic density.

We have also investigated the cross-dependence between optically derived tissue parameters. The results obtained on the linear correlation are summarized in Table 1. The strongest (negative) correlation is observed between lipid and water content, but negative correlation is also evident between lipid and collagen content. Both observations are in agreement with what was expected based on breast tissue composition: moving from adipose to fibroglandular breasts, the amount of adipose tissue with high lipid content decreases and is replaced by connective and epithelial tissue, richer in water and collagen. Marked correlation also exists between the scattering amplitude a and the concentrations of all major tissue constituents. Specifically, the correlation is positive for water and collagen, while it is negative for lipid, consistent with the hypothesis that fibroglandular tissue, rich in water and collagen, is mainly responsible for breast tissue scattering.

To develop a procedure for the identification of high-risk women, the mammographic density was dichotomized, comparing subjects in BI-RADS categories 1 to 3 to subjects in category 4, the latter being at significantly higher risk than all the others.[2] The p values of the Wilcoxon test showed that tHb, lipid, water, collagen, a, and b are significantly different in the two populations considered (at least $p < 0.001$), while SO$_2$ is not. The best regression logistic model for the risk probability chosen via a stepwise variables selection minimizing the Akaike information criterion resulted to be

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 [\text{Collagen}]_i + \alpha_2 \alpha_i + \alpha_3 b_i, \quad (1)$$

where p_i is the probability of belonging to BI-RADS category 4 (high risk) for the i'th subject. Table 2 shows the output of the fitted regression logistic model (point estimates of the coefficients, related standard errors, z-statistics, and p values of testing their significance in the model). The Brier's score, i.e., the mean square difference between outcome and estimated probability, is equal to 0.095.

Table 2 Output of the fitted regression logistic model.

Coefficients	Estimate	Standard error	z value	p value
α_0 (intercept)	−10.460869	1.968482	−5.314	1.07×10^{-7}
α_1 (collagen)	0.015704	0.006862	2.289	0.0221
α_2 (a)	0.227804	0.114084	1.997	0.0458
α_3 (b)	6.287575	1.218039	5.162	2.44×10^{-7}

Table 1 Correlation estimates (95% confidence intervals of Pearson association) of the optically derived parameters.

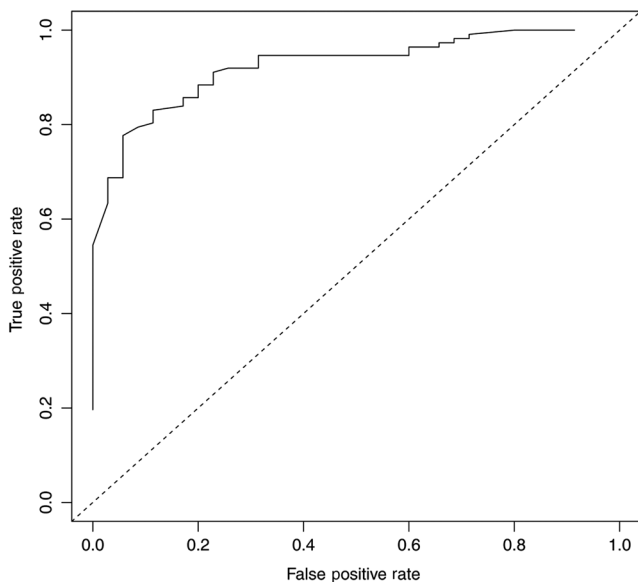|	a	b	tHb	SO$_2$	Lipid	Water	Collagen
a	/						
b	0.23 (0.07,0.38)	/					
tHb	0.41 (0.27,0.54)	0.17 (0.01,0.32)	/				
SO$_2$	0.09 (−0.07, 0.25)	−0.36 (−0.50, −0.21)	0.28 (0.13,0.43)	/			
Lipid	−0.77 (−0.83, −0.69)	−0.49 (−0.61, −0.36)	−0.55 (−0.66, −0.43)	−0.13 (−0.03, 0.04)	/		
Water	0.79 (0.72, 0.84)	0.42 (0.28, 0.55)	0.45 (0.32, 0.57)	0.07 (0.09, 0.23)	−0.90 (−0.92, −0.86)	/	
Collagen	0.69 (0.59, 0.76)	0.11 (−0.05, 0.27)	0.58 (0.46, 0.66)	0.33 (0.17, 0.46)	−0.78 (−0.84, −0.71)	0.70 (0.61, 0.78)	/

Fig. 1 Receiver operating characteristic curve for the prediction of high risk using Eq. (1).

Based on Eq. (1) and Table 2, the probability of belonging to the high-risk category depends on collagen concentration and on both scattering parameters. In particular, the strongest dependence occurs for the scattering slope b. Performing *in vivo* measurements, we have recently observed that high scattering slope corresponds to high collagen content and possibly depends also on its structure.[17] Collagen content also shows a significant correlation with the scattering amplitude a, as highlighted in Table 1. Thus, both directly and indirectly, collagen seems to be the most crucial feature for the identification of subjects with high breast density.

The receiver operating characteristic curve for our model is reported in Fig. 1. We classify the subject as a high-risk patient if the estimated probability [Eq. (1)] is greater than 0.5. The corresponding misclassification matrix is reported in Table 3, where "true" refers to risk classification based on mammographic assessment (BI-RADS categories) and "classified" refers to risk as predicted based on logistic regression fitted on optical data. The data reported in Table 3 correspond to a total misclassification error of 12.3%, sensitivity of 69%, and specificity of 94%. A simple kappa of 0.84 is achieved, to be compared with the reproducibility of BI-RADS assignment among radiologists and even intraradiologists. Specifically, an intrarater agreement of 77% is reported in the literature, leading to a simple kappa of 0.58.[18] Thus, the classification achieved by optical means appears to be very promising.

In summary, a logistic regression model was fitted to optically derived tissue parameters with the aim of identifying women at high risk for developing breast cancer because of their high breast density. Encouraging preliminary results were obtained, and collagen proved to be the key parameter for the classification, either directly (collagen content) or indirectly (through scattering amplitude and slope). The relevance of collagen is in agreement with what was expected based on breast anatomy and physiology, and opens up the possibility of a more

Table 3 Misclassification matrix.

		Classified (optical)	
		Low risk	High risk
True (BI-RADS)	Low risk	105	7
	High risk	11	24

direct estimate of breast density than presently achieved using x-ray mammography, which is mostly sensitive to water content and not directly to collagen.

References

1. R. Siegel et al., "Cancer statistics, 2011," *CA Cancer J. Clin.* **61**(4), 212–236 (2011).
2. V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," *Cancer Epidemiol. Biomarkers Prev.* **15**(6), 1159–1169 (2006).
3. N. F. Boyd et al., "Mammographic density and the risk and detection of breast cancer," *N. Engl. J. Med.* **356**(3), 227–236 (2007).
4. U.S. Preventive Services Task Force, "Chemoprevention of breast cancer: recommendations and rationale," *Ann. Intern. Med.* **137**(1), 56–58 (2002).
5. S. R. Cummings et al., "Prevention of breast cancer in postmenopausal women: approaches to estimating and reducing risk," *JNCI J. Natl. Cancer Inst.* **101**(6), 384–398 (2009).
6. P. Taroni et al., "Diffuse optical spectroscopy of breast extended to 1100 nm," *J. Biomed. Opt.* **14**(5), 054030 (2009).
7. N. Shah et al., "Spatial variations in optical and physiological properties of healthy breast tissue," *J. Biomed. Opt.* **9**(3), 534–540 (2004).
8. B. W. Pogue et al., "Characterization of hemoglobin, water and NIR scattering in breast tissue: analysis of inter-subject variability and menstrual cycle changes relative to lesions," *J. Biomed. Opt.* **9**(3), 541–552 (2004).
9. R. Cubeddu et al., "Effects of the menstrual cycle on the red and near-infrared optical properties of the human breast," *Photochem. Photobiol.* **72**(3), 383–391 (2000).
10. K. M. Blackmore et al., "Association between transillumination breast spectroscopy and quantitative mammographic features of the breast," *Cancer Epidemiol. Biomarkers Prev.* **17**, 1043–1050 (2008).
11. X. Wang et al., "Approximation of Mie scattering parameters in near-infrared tomography of normal breast tissue *in vivo*," *J. Biomed. Opt.* **10**(5), 051704 (2005).
12. N. Shah et al., "The role of diffuse optical spectroscopy in the clinical management of breast cancer," *Dis. Markers* **19**(2–3), 95–105 (2003).
13. P. Taroni et al., "Non-invasive assessment of breast cancer risk using time-resolved diffuse optical spectroscopy," *J. Biomed. Opt.* **15**(6), 060501 (2010).
14. P. Taroni et al., "Effects of tissue heterogeneity on the optical estimate of breast density," *Biomed. Opt. Express* **3**(10), 2411–2418 (2012).
15. C. D'Andrea et al., "Time-resolved spectrally constrained method for the quantification of chromophore concentrations and scattering parameters in diffusing media," *Opt. Express* **14**(5), 1888–1898 (2006).
16. P. Taroni et al., "Seven-wavelength time-resolved optical mammography extending beyond 1000 nm for breast collagen quantification," *Opt. Express* **17**(18), 15932–15946 (2009).
17. P. Taroni et al., "Role of collagen scattering for *in vivo* tissue characterization" presented at *Biomedical Optics (BIOMED) on CD-ROM, BTuD107*, Optical Society of America, Washington, DC (2010).
18. M. C. Spayne et al., "Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study," *Breast J.* **18**(4), 326–333 (2012).