Diagnosis of breast cancer using diffuse optical spectroscopy from 500 to 1600 nm: comparison of classification methods

Rami Nachabé
Daniel J. Evers
Benno H. W. Hendriks
Gerald W. Lucassen
Marjolein van der Voort
Emiel J. Rutgers
Marie-Jeanne Vrancken Peeters
Jos A. Van der Hage
Hester S. Oldenburg
Jelle Wesseling
Theo J. M. Ruers

Diagnosis of breast cancer using diffuse optical spectroscopy from 500 to 1600 nm: comparison of classification methods

Rami Nachabé,[a] Daniel J. Evers,[b] Benno H. W. Hendriks,[a] Gerald W. Lucassen,[a] Marjolein van der Voort,[a] Emiel J. Rutgers,[b] Marie-Jeanne Vrancken Peeters,[b] Jos A. Van der Hage,[b] Hester S. Oldenburg,[b] Jelle Wesseling,[c] and Theo J. M. Ruers[b,d]
[a]Minimally Invasive Healthcare, Philips Research, 5656AE Eindhoven, The Netherlands
[b]The Netherlands Cancer Institute, Department of Surgery, 1066CX Amsterdam, The Netherlands
[c]The Netherlands Cancer Institute, Department of Pathology, 1066CX Amsterdam, The Netherlands
[d]Technical University Twente, MIRA Institute, 7500AE Twente, The Netherlands

Abstract. We report on the use of diffuse optical spectroscopy analysis of breast spectra acquired in the wavelength range from 500 to 1600 nm with a fiber optic probe. A total of 102 *ex vivo* samples of five different breast tissue types, namely adipose, glandular, fibroadenoma, invasive carcinoma, and ductal carcinoma *in situ* from 52 patients were measured. A model deriving from the diffusion theory was applied to the measured spectra in order to extract clinically relevant parameters such as blood, water, lipid, and collagen volume fractions, β-carotene concentration, average vessels radius, reduced scattering amplitude, Mie slope, and Mie-to-total scattering fraction. Based on a classification and regression tree algorithm applied to the derived parameters, a sensitivity-specificity of 98%–99%, 84%–95%, 81%–98%, 91%–95%, and 83%–99% were obtained for discrimination of adipose, glandular, fibroadenoma, invasive carcinoma, and ductal carcinoma *in situ*, respectively; and a multiple classes overall diagnostic performance of 94%. Sensitivity-specificity values obtained for discriminating malignant from nonmalignant tissue were compared to existing reported studies by applying the different classification methods that were used in each of these studies. Furthermore, in these reported studies, either lipid or β-carotene was considered as adipose tissue precursors. We estimate both chromophore concentrations and demonstrate that lipid is a better discriminator for adipose tissue than β-carotene. © *2011 Society of Photo-Optical Instrumentation Engineers (SPIE).* [DOI: 10.1117/1.3611010]

Keywords: visible-infrared spectroscopy; breast cancer; optical diagnosis.

Paper 11229R received May 10, 2011; revised manuscript received Jun. 23, 2011; accepted for publication Jun. 24, 2011; published online Aug. 9, 2011.

1 Introduction

Within present-day strategy of human breast cancer treatment, diagnostic biopsy and surgical margin assessment are two elements in which procedural accuracy could significantly be enhanced.

Missed diagnoses of cancer by false-negative biopsies have been reported ranging from 4.3% to 17.9%, despite ongoing advances in imaging technologies. Moreover, indeterminate pathology analysis will result in the need of repeat biopsies in between 4% to 32% of patients.[1–6]

Breast conservative therapy, aimed at conserving as much breast tissue as possible, is the treatment of choice in patients with T1-T2 breast tumors. However, the rate of irradical resection and the need for a secondary surgical procedure is often over 10%, depending on the specific definition.[7,8]

Over the last decade, new tools have been developed to classify breast tissue and assess breast tissue margins based on optical spectroscopy techniques.[9–22] Bigio et al. performed *in vivo* elastic scattering spectroscopy measurements between 350 and 750 nm to discriminate between 13 malignant and 59 nonmalignant breast tissue samples by applying artificial neural

network (ANN) and hierarchical cluster analysis on the spectra yielding sensitivity-specificity of 69%–85% and 67%–79%, respectively.[9] This study also showed that the spectral features between 400 and 500 nm in adipose tissue are mainly dominated by β-carotene light absorption, however optical properties were not derived from the measured spectra. The biomedical group at Duke University has performed several studies where optical properties were derived from measurements performed between 350 and 600 nm by using an inverse Monte Carlo technique to extract hemoglobin and β-carotene concentrations, as well as hemoglobin saturation and the reduced scattering amplitude.[10–13] Classification based on linear support vector machine (SVM) learning was performed to classify malignant (35 samples) from nonmalignant samples (50 samples including adipose and fibrous tissue types) with a sensitivity-specificity of 83%–80%.[10] A more recent study from the same group showed that it is possible to discriminate 54 malignant samples from 70 nonmalignant samples with a sensitivity-specificity of 83%–87% based on the extracted parameters from diffuse reflectance measurements.[11] Volynskaya et al. conducted an *ex vivo* breast (104 samples) study where a classification between four types of breast tissue was performed from diffuse reflectance spectra acquired from 350 to 750 nm.[14] Classification of 31

Address all correspondence to: Rami Nachabé, Philips Research, HighTech Campus 34, Eindhoven Noord-Brabant 5656AE, The Netherlands; Tel.: 31616911207; E-mail: rami.nachabe@philips.com.

normal, 55 fibrocystic change, 9 fibroadenoma, and 9 infiltrating ductal carcinoma was achieved with a sensitivity-specificity of 100%–100% by using a logistic regression (LR) algorithm. An *ex vivo* breast study by Majumder et al. showed that sparse multinomial logistic regression classification of 134 normal (adipose and glandular), 86 invasive ductal carcinoma, 18 ductal carcinoma *in situ*, and 55 fibroadenoma spectra can be achieved with sensitivity-specificity ranging from 28%–86% to 86%–97% when only analyzing diffuse reflectance spectra acquired between 400 and 800 nm.[15] A more recent study showed sensitivity-specificity of 85%–96% when discriminating 145 normal from 34 tumor (invasive ductal carcinoma and ductal carcinoma *in situ*) samples.[16] Laughney et al. presented *ex vivo* noncontact optical properties estimations from spectra acquired between 510 and 785 nm from 29 breast samples and a *k*-nearest neighbor (KNN) classification method was used to discriminate between different tissue types.[17] They have shown an interesting comparison of classifying the different types of tissue according to their pathology identity and by grouping them into subgroups, such as adipose (7021 spectra), nonmalignant (533 inflammation, 4110 benign epithelia, and 31226 normal epithelia spectra), and malignant (194 ductal carcinoma *in situ*, 479 invasive lobular carcinoma, and 22547 invasive ductal carcinoma spectra). Their results showed sensitivity-specificity of 87%–99%, 90%–82%, and 77%–90% for adipose, nonmalignant, and malignant, respectively. However, a sensitivity-specificity of 87%–99%, 74%–74%, 9%–91%, 0%–100%, 77%–90%, 0%–100%, and 0%–100% was reached for adipose, normal epithelia, benign, inflammation, invasive ductal carcinoma, ductal carcinoma *in situ*, and invasive lobular carcinoma, respectively.

Other studies[18–21] investigated wavelength ranges between 600 and 1100 nm where water and lipid were estimated in addition to hemoglobin. Therefore, adipose tissue could be discriminated based on the amount of estimated lipid and not β-carotene, since this chromophore has negligible absorption above 600 nm. However, these investigators did not perform a classification on their data.

In our study, we have conducted an *ex vivo* trial to estimate optical properties from 102 samples of five different types of breast tissue: adipose, glandular, invasive carcinoma (IC), fibroadenoma (FA), and ductal carcinoma *in situ* (DCIS) measured in 52 patients. Optical spectra were taken with a setup that can resolve light from 500 nm up to 1600 nm, and a model based on diffusion theory was applied to the measurements to estimate the optical properties by determining several parameters such as blood, water, and lipid volume fractions, reduced scattering amplitude, Mie slope, Mie scattering fraction, and pigment packaging factor.[23,24] Besides, β-carotene was also included in our model since it has significant absorption up to 500 nm as demonstrated by other groups.[9–11,14] Recent findings by Taroni et al. showed that collagen is an important absorber to include in the model for fitting the measured spectra as it has distinct absorption features above 900 nm.[19–21] Therefore, we measured the absorption coefficient of collagen up to 1600 nm and included it in our model.

We present the first study using diffuse reflectance spectroscopy (DRS) measurements on a 500 to 1600 nm wavelength range to estimate physiological, morphological, and optical properties parameters of *ex vivo* breast tissue. The classification and regression tree (CART) algorithm, a probabilistic

discriminative classification method, was applied to the derived parameters to evaluate the performance of diagnosis of the five measured types of tissues. Sensitivity-specificity computation and receiver operating characteristic (ROC) curves analysis were performed to quantify the overall performance of the diagnosis by using the Provost and Domingos measure (PDM).[25]

In addition, several classification methods that were used in literature to discriminate malignant from nonmalignant tissues were applied to our data in order to compare our results with those reported in existing literature studies. Additional classification methods were also applied for comparison.

Finally, classification of adipose tissue based on either β-carotene or lipid only was compared as no existing breast studies in literature made a comparison on classifying adipose tissue based on only one of these two adipose tissue precursors.

2 Materials and Methods

2.1 *Ex vivo Breast Sample Collection*

The human breast samples were obtained under approval by the internal review board committee of the Dutch Cancer Institute in Amsterdam, The Netherlands (NKI-AVL) where this study was conducted. The breast samples that were measured corresponded to resection specimens of either to mastectomies or lumpectomies. Breast samples of patients subject to mastectomy were sliced with a thickness of roughly 0.5 to 1 cm, whereas the sample sizes of the patients who were subject to lumpectomy (e.g., fibroadenoma) corresponded to the size of the excised tissue which was on average several millimeters in diameter. After surgical resection, resection samples were transferred to the pathology department within 2 h, where they were inked at the surface before slicing them for histological processing. All optical measurements were performed before formalin fixation and tissue preparations by the pathologists in order to limit, as much as possible, changes in the optical properties from the tissue conditions when excised. Five different types of tissue were measured based on the macroscopical indication by the pathologist: adipose, glandular, FA, IC, and DCIS. A total number of 102 samples from 52 patients were investigated, from which a total number of 980 spectra were acquired and co-registered with the pathological findings. The pathological diagnosis performance was very high for all the cases that we have tested. The cancerous cases were all macroscopically clear cut carcinomas; and in case of doubt we were reluctant to include such cases in this study. Table 1 summarizes the histological breakdown of the breast tissue samples, including the amount of acquired spectra in this study.

2.2 *Instrumentation and Spectral Calibration*

Ex vivo diffuse reflectance spectra were taken using a portable spectroscopic system as illustrated in Fig. 1 and used in previous studies.[23,24,26] A tungsten halogen broadband light source with an integrated shutter (Ocean Optics, HL-2000-HP) was used to deliver light into the tissue. Delivery of light to the tissue and its collection were achieved with a 1.3-mm diameter fiber-optic probe with a distal end polished at an angle of 20 deg. The probe comprises three 200-μm core diameter optical fibers with one fiber connected to the light source that is located 2.48 mm from

Table 1 Histological description of breast tissue types and the corresponding amount of samples and spectra that were measured.

Type of breast tissue	Number of samples	Number of spectra
Nonmalignant	**73**	**643**
Adipose	43	327
Glandular	23	189
FA	7	127
Malignant	**29**	**337**
IC	21	241
DCIS	8	96
Total	**102**	**980**

the two side-by-side optical fibers that are used to collect the diffused light. The optical fibers used for the collection of light are connected to a spectrometer with a silicon detector (Andor Technology, DU420A-BRDD) and a spectrometer with an InGaAs detector (Andor Technology, DU492A-1.7), respectively. After thermoelectrically cooling the detectors to $-40\,^\circ$C, wavelength values were assigned to each pixel of the detector by fitting a second-order polynomial to a set of atomic lines from an argon source with peaks at known wavelength. Subsequently, the spectral response of a white reflectance standard (Spectralon) with known reflectivity was measured by placing the distal end of the probe at a fixed distance of roughly 2 mm and followed by a background measurement in order to minimize the impact of ambient light. This step is necessary as it allows correcting for the system response (e.g., spectral shape of the light source and wavelength-dependent sensitivity in the optics and gratings and the detectors). This white reference measurement is used to divide each spectral measurement on the tissue samples for which a background measurement is subtracted, yielding to the final

reflectance measurement. The integration time for each measurement is on average 0.5 s. The reflectance spectra obtained with both spectra are combined together to form one single reflectance spectrum ranging from 500 to 1600 nm, and is used in order to apply the mathematical modeling for the data analysis.

2.3 Spectral Data Modeling

The measured spectra were fitted from 500 to 1600 nm with the model of Farrell et al.[27] that is derived from diffusion theory using a Levenberg–Marquardt nonlinear inversion algorithm in order to determine the absorption coefficient $\mu_a(\lambda)$ and the reduced scattering coefficient $\mu'_s(\lambda)$ expressed in cm^{-1}. The validation of the model based on a phantom study, including spectral calibration procedures, and its application to *in vivo* and *ex vivo* tissues, were justified in detail elsewhere.[23,24]

The model requires the distance between the emitting and collecting fibers as well as the wavelength-dependent absorption coefficients of the chromophores of interest as input arguments. Additionally, the reduced scattering coefficient was empirically modeled as:

$$\mu'_s(\lambda) = \alpha \left[\rho \left(\frac{\lambda}{\lambda_0} \right)^{-b} + (1 - \rho) \left(\frac{\lambda}{\lambda_0} \right)^{-4} \right], \quad (1)$$

where $\lambda_0 = 800\,\text{nm}$ corresponds to a wavelength normalization value, α is the reduced scattering amplitude at λ_0, the Mie scattering slope is b, and ρ denotes the Mie-to-total reduced scattering fraction assuming Mie and Rayleigh scattering as the two types of scattering in tissue.

The absorption coefficient is expressed as a term that corresponds to vascular absorption $\mu_a^{\text{Blood}}(\lambda)$, of the light due to blood-derived chromophores and a second term $\mu_a^{\text{Other}}(\lambda)$ due to absorption of light by other chromophores present in breast tissue. The blood related absorbers are deoxygenated-hemoglobin (Hb) and oxygenated-hemoglobin (HbO$_2$) and define the absorption coefficient due to blood as:

$$\mu_a^{\text{Blood}}(\lambda) = C(\lambda)\nu \left[\text{StO}_2 \mu_a^{\text{HbO}_2}(\lambda) + (1 - \text{StO}_2) \mu_a^{\text{Hb}}(\lambda) \right],$$
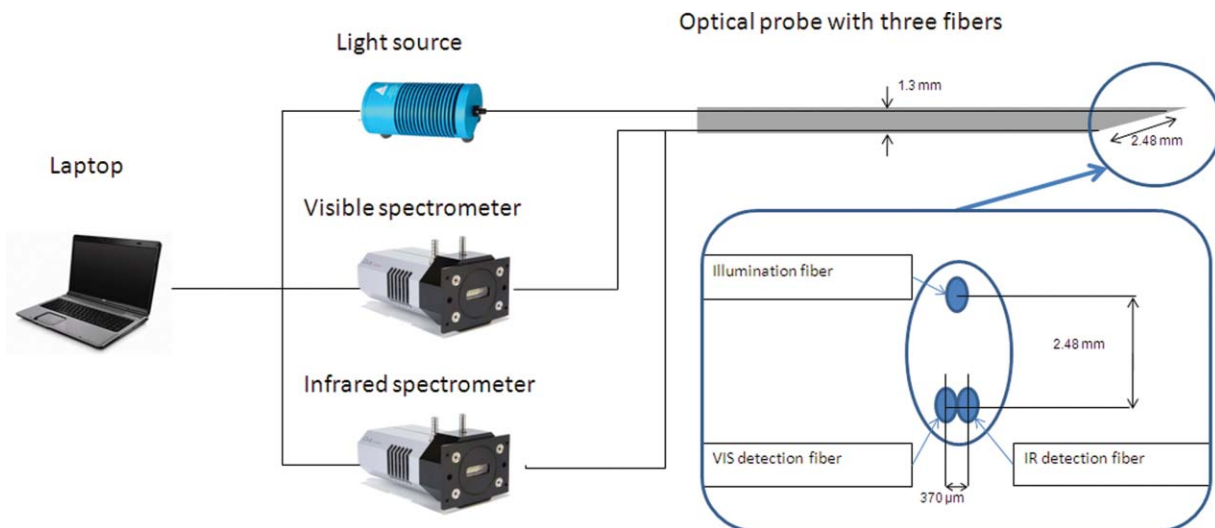
$$(2)$$

Fig. 1 Schematic of the optical setup and the design of the optical probe.
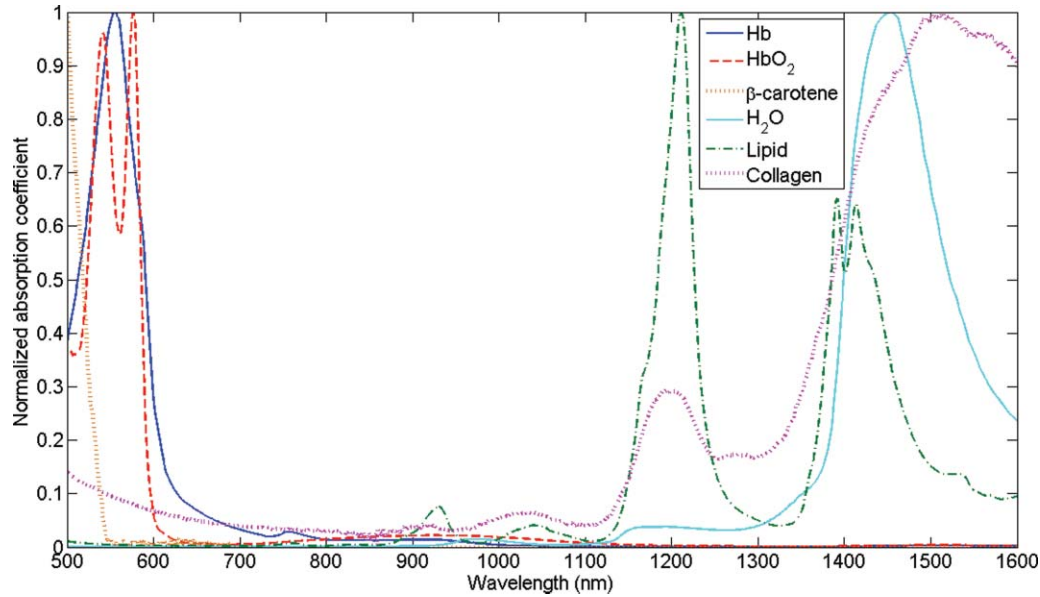
Fig. 2 Normalized absorption coefficients of Hb, HbO₂, β-carotene, water (H₂O), lipid, and collagen.

where $\mu_a^{\text{Hb}}(\lambda)$ and $\mu_a^{\text{HbO}_2}(\lambda)$ correspond to absorption coefficients of pure Hb and HbO₂ given an average hemoglobin concentration in blood of 150 mg/ml, respectively. The parameters ν and StO₂ correspond to the blood volume fraction and the level of hemoglobin saturation by oxygen, respectively. The parameter $C(\lambda)$ was used to account for inhomogenous distribution of hemoglobin in vessels and is known as pigment packaging factor[28] expressed as:

$$C(\lambda) = \frac{1 - \exp(-2R[\text{StO}_2\mu_a^{\text{HbO}_2}(\lambda) + (1 - \text{StO}_2)\mu_a^{\text{Hb}}(\lambda)])}{2R[\text{StO}_2\mu_a^{\text{HbO}_2}(\lambda) + (1 - \text{StO}_2)\mu_a^{\text{Hb}}(\lambda)]},$$

(3)

where R corresponds to the average vessel radius. Studies that were performed on breast tissue[11,14] showed that it is important to have β-carotene (βc) as an absorber in the model when recording spectra in the visible range. Indeed, these studies demonstrated that βc is an essential discriminator for adipose tissue in breast. Other studies that investigated optical properties of breast in the near-infrared range[18,20] discriminate adipose tissue from other types of tissue based on light absorption by lipids. However, no studies so far used both βc and lipid. In our study we included both absorbers in the model and investigated the advantage of measuring up to 1600 nm where additional water and lipid absorption features exist,[23] which enables more accurate estimation of lipid volume fraction.[24]

Taroni et al. showed that collagen is an abundant absorber in several breast tissue types when recording optical spectra up to 1100 nm.[19–21] Therefore, we have measured collagen Type I (Sigma-Aldrich C9879) absorption coefficients from 500 to 1600 nm by tightly inserting the collagen fibers in cuvettes of 0.5, 1, and 2-mm thickness, and measuring its absorption with a spectrophotograph with a 150-mm diameter integrating sphere (Lambda 900 Spectrometer, Perkin Elmer). The absorption measurements were separated from the scattering by mounting the cuvettes inside the integrating sphere far away from the detector. When a sample is mounted inside the sphere, the loss of light is mainly due to absorption by the sample related to the absorption coefficient. Because of the turbidity of the sample, scattering

occurs. Therefore, an additional measurement was performed by allowing the forward transmitted light to escape out of an exit port in the back end of the sphere in order to measure scattering. The scattered light from the sample mounted inside the sphere is therefore measured, and subsequently the absorption coefficient can be determined by subtracting the measurement with the opened exit port from the measurement with the closed exit port.

The absorption coefficient due to nonblood derived chromophores is expressed as:

$$\mu_a^{\text{Other}} = \psi[f_{\text{Lipid}}\mu_a^{\text{Lipid}}(\lambda) + (1 - f_{\text{Lipid}})\mu_a^{\text{H}_2\text{O}}(\lambda)]$$
$$+ f_{\text{Collagen}}\mu_a^{\text{Collagen}}(\lambda) + c_{\beta c}\varepsilon_{\beta c}(\lambda),$$

(4)

where $\mu_a^{\text{Lipid}}(\lambda)$, $\mu_a^{\text{H}_2\text{O}}(\lambda)$, $\mu_a^{\text{Collagen}}(\lambda)$, and $\varepsilon_{\beta c}(\lambda)$ correspond to the absorption coefficients of lipid, water, collagen, and the extinction coefficient (in cm⁻¹ M⁻¹) of β-carotene, respectively. The parameter ψ represents the water and lipid volume fraction, and f_{Lipid} represents the lipid fraction within the volume probed by the light. However, f_{Collagen} corresponds to the collagen volume fraction in the probed tissue, whereas $c_{\beta c}$ corresponds to the molar concentration of β-carotene. The absorption coefficients of the various chromophores of interest are depicted in Fig. 2. The absorption coefficient at unit concentration of Hb and HbO₂ that is used as *a priori* knowledge for the model are from Zijlstra et al.,[29] whereas the extinction coefficient of β-carotene in human adipose cells is from van de Poll et al.[30] Water and lipid absorption coefficients that are used in the presented study are from a previous published work.[23] The collagen absorption coefficient presented in this study has a local maximum at 1200 nm of 1.54 cm⁻¹, which is of the same order of magnitude as the water and lipid absorption coefficient in the vicinity of 1200 nm. It is important to note that collagen has a wider maximum than fat, but is narrower than water. Other local maxima at 911, 1030, and 1510 nm are observed with absorption coefficients of 0.21, 0.34, and 5.22 cm⁻¹. The presented absorption coefficient of collagen is about an order of magnitude higher than the one presented by Taroni et al.,[20] however it matches very well with

the coefficients reported by Tsai et al.[31] and by Nunez.[32] The difference in collagen absorption values with Taroni et al. could be due to the fact that the density of our measured sample is different from the density used by Taroni et al.

In a few cases, the ink used by the pathologist before cutting was spread into the tissue when slicing the breast samples, influencing the measured spectral shapes. In order to correct for this, the absorption coefficients of these inks were measured and added to the fitting. Given the large number of free parameters, two separate fits on different wavelength ranges were performed: the first fit was performed between 500 and 900 nm with $\mu_a^{Blood}(\lambda) + c_{\beta c}.\varepsilon_{\beta c}(\lambda)$, and $\mu_s'(\lambda)$ only in the model and the second fit was performed between 900 and 1600 nm with $\mu_a^{Other}(\lambda) - c_{\beta c}.\varepsilon_{\beta c}(\lambda)$, and $\mu_s'(\lambda)$ only in the model. The extracted values from both fits were used as initial guess for the fit applied over the full wavelength range between 500 and 1600 nm in order to ensure stability of the fit.

From fits to the spectra between 500 and 1600 nm, the following fit parameters were obtained: ν, StO_2, R, ψ, f_{Lipid}, $f_{Collagen}$, $c_{\beta c}$, α, b, and ρ. For each estimated value, a confidence interval computed from the covariance matrix was used to assess the reliability for each fit parameter.[33]

2.4 Statistical Analysis

A nonparametric Kruskal–Wallis statistical test was conducted to evaluate significant differences of the estimated parameters between the various types of breast tissue for a significance level of 5% (i.e., $p < 0.05$). The test examines if the medians of the various groups are not all equal; meaning that if the p-value is below the significance level, at least one type of tissue can be discriminated from the others. Therefore, an additional *post hoc* test is required to account for multiple comparisons, as well as for the fact that comparisons can be interrelated. In this study, Tukey's *post hoc* test was applied at a significance level of 5%. This statistical procedure is a restricted pairwise comparison that follows the Kruskal–Wallis test which had indicated the significance of the differences.[34]

2.5 Classification Algorithms

The CART algorithm was used to classify between the five types of tissue. The CART algorithm starts from a central node that discriminates the largest class, adipose tissue in our case, based on the best classifier. From this root node, a split is performed to discriminate the largest class from the other tissue classes. From the split, daughter partial trees are generated and other parameters are used for further splits. The purity of each node is assessed with the Gini's maximization index algorithm, which corresponds to unity minus the sum of squares of the proportions of target classes at a specific node.[35] The advantage of CART is that it is a nonparametric method, whereas other methods such as linear discriminant analysis (LDA), LR, and KNN assume functional relations between dependent and predictor variables. Moreover, one of the advantages of CART is that it is easy to interpret since the input parameters for classification are used, whereas other methods post-process the parameters into scores that might not be intuitively related to the input parameters. The performance of the diagnosis was evaluated by carrying out an ROC analysis. From the sensitivity-specificity values and the area under the ROC curves (AUC), the PDM for total AUC is computed to assess the accuracy of the diagnostic algorithms.[25] The PDM value corresponds to the sum of AUC of each class weighted by the class size fraction.

Classifications were carried out on the estimated parameters from the fit model using a leave-one-out (LOO) cross validation scheme. Additionally, a hold-out (HO) cross validation scheme with a 70%–30% training-testing split of the data was carried out. The split was performed by random selection of the data before splitting and classification. This partition procedure and classification was reproduced 20 times and the computed sensitivity-specificity values were averaged.

Several techniques were used in literature to classify parameters based on diffuse reflectance spectroscopy measurements or directly applied to the spectra as mentioned in the Sec. 1. The following classification algorithms were applied to our data: ANN,[9] linear SVM,[11] LR,[14–16] and KNN employing Mahalabonis distance to account for parameters intercorrelation,[17] in order to evaluate the sensitivity-specificity of discriminating malignant and nonmalignant types of breast tissue. Besides, other classification methods were also tested such as CART, LDA with Mahalanobis distance stratified covariance, and nonlinear SVM to discriminate malignant from nonmalignant breast tissues. However, the classification was performed by taking the amount of samples that corresponds to the lowest sample size within the malignant and nonmalignant category, respectively. This means that within the nonmalignant category, 127 spectra from adipose and from glandular tissues were randomly selected and added to the FA spectra to form the nonmalignant database, whereas 96 spectra from IC were randomly selected and added to the DCIS spectra to form the malignant database. The purpose of such categorization is to avoid higher representation of one type of tissue over the others within the same category. Otherwise, discriminating malignant from nonmalignant tissue would be comparable to classification of adipose versus IC given the fact that the total adipose and IC spectra represent 51% and 72% of the nonmalignant and malignant samples size, respectively.

Furthermore, this study is the first that estimates both β-carotene and lipid from breast tissue measurements. A classification of adipose tissue was performed from all the parameters except lipid, water, and collagen and another classification without β-carotene to evaluate which adipose precursor is the most accurate for adipose breast tissue classification using the CART algorithm.

3 Results

Figure 3 depicts typical examples of spectra measured on adipose [Fig. 3(a)], glandular [Fig. 3(b)], FA [Fig. 3(c)], IC [Fig. 3(d)], and DCIS [Fig. 3(e)] tissues and their corresponding fits. From the measurement of adipose breast tissue, one can notice the effect of β-carotene absorption on the spectra below 550 nm and the lipid absorption peaks at 930 and 1211 nm. Figure 4 depicts the histograms of the median and standard deviation for each of the parameters derived from the fit per category of tissue type. Complementary to Fig. 4, Table 2 displays the parameters that show significant differences ($p < 0.05$) for pairwise types of tissue comparison according to a Kruskal–Wallis test followed by a *post hoc* multiple
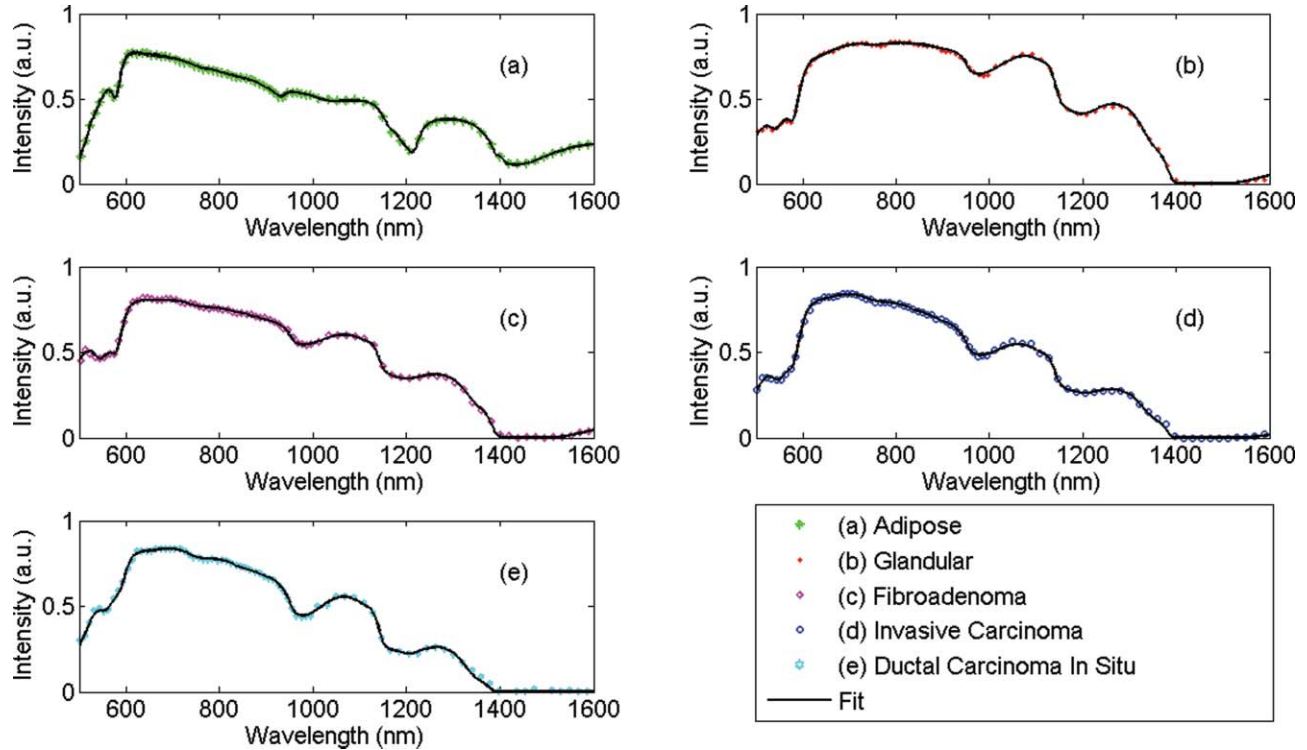
Fig. 3 Typical measurement of adipose (a), glandular (b), FA (c), IC (d), and DCIS (e), and their corresponding fit curves.
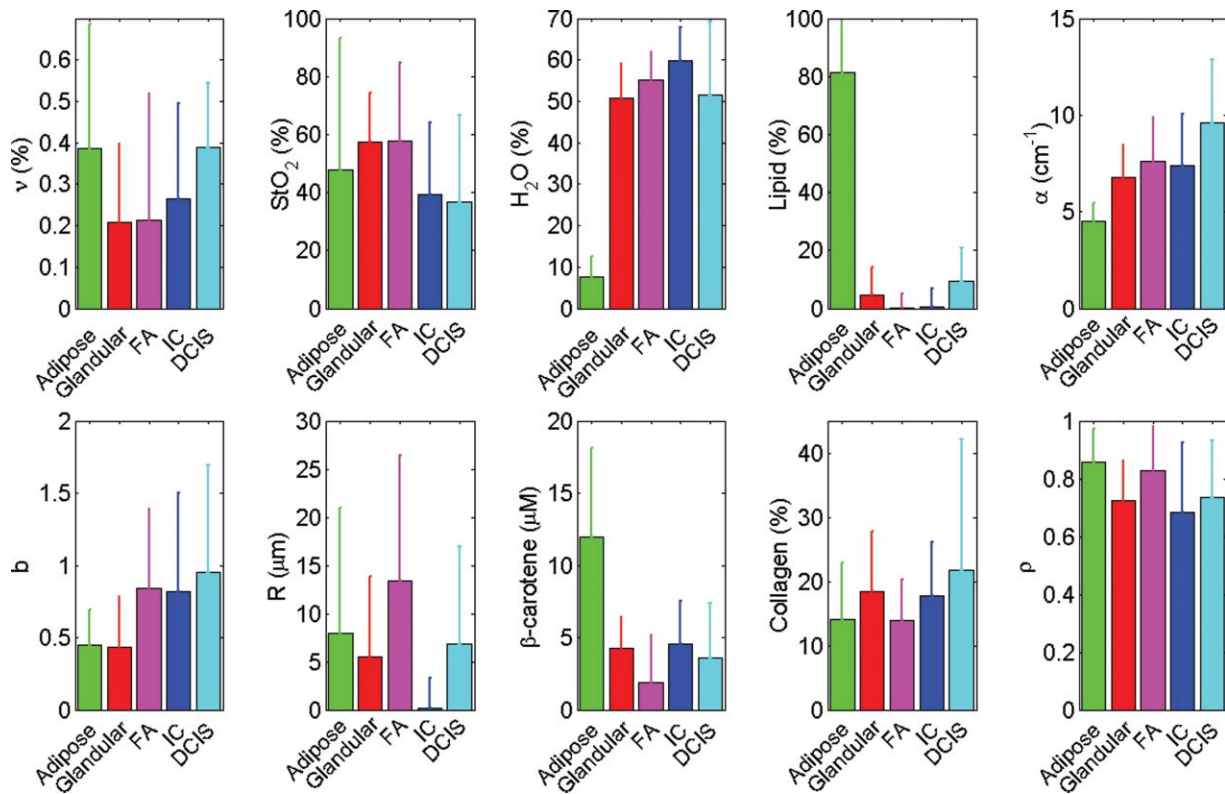

Fig. 4 Average and standard deviation of the estimated blood volume fraction (ν), oxygenation level (StO$_2$), water (H$_2$O), lipid, reduced scattering amplitude (α), scattering slope (b), vessel radius (R), β-carotene, collagen, and the Mie-to-total reduced scattering fraction (ρ) for each of the various types of breast tissues: adipose, glandular, fibroadenoma, invasive carcinoma, and ductal carcinoma *in situ*.

Table 2 Parameters that show significant difference for the pairwise comparisons of the different tissue types after Kruskal–Wallis statistical test with *post hoc* Tukey's multiple comparison test ($p < 0.05$).

Type of breast tissue	Type of breast tissue			
	Glandular	FA	IC	DCIS
Adipose	ν, StO$_2$, H$_2$O, lipid, α, βc, collagen, ρ	ν, StO$_2$, H$_2$O, lipid, α, b, R, βc	ν, StO$_2$, H$_2$O, lipid, α, b, R, βc, collagen, ρ	H$_2$O, lipid, α, b, βc, collagen, ρ
Glandular	–	b, R, βc, collagen, ρ	StO$_2$, H$_2$O, α, b, R	ν, StO$_2$, α, b
FA		–	StO$_2$, H$_2$O, R, βc, collagen, ρ	ν, StO$_2$, lipid, α, R, ρ
IC			–	ν, lipid, α, R

comparison Tukey's test. It can be seen that adipose and DCIS tissue contains almost twice as much blood as the other types of tissue, while the blood oxygenation level is lower in malignant tissue (StO$_2 < 40\%$) compared to nonmalignant tissue. Adipose tissue can clearly be distinguished from the other tissue types by its high lipid average volume fraction and β-carotene concentration of 80% and 12 μm, respectively. FA has the lowest β-carotene concentration and is significantly different from the other tissue types except for DCIS. The reduced scattering amplitude is the lowest for adipose tissue (roughly 5 cm^{-1}) and the highest for DCIS (roughly 10 cm^{-1}), whereas it is rather similar for the other tissues (around 7 cm^{-1}). A clear distinction can be observed for the Mie slope, where it is almost two-fold smaller for nonmalignant compared to malignant samples. Apart from adipose tissue, IC showed a significant differences based on the water content with the highest amount among all tissues. Adipose and FA have the lowest collagen volume fraction of roughly 14%, whereas glandular and IC are about 18% and DCIS has the highest value with 22%. Although adipose and FA have similar collagen volume fractions, this parameter showed a significant difference between DCIS and adipose tissue and not with FA, due to the higher standard deviation in collagen in adipose tissue compared to FA. It can be seen that the trends in collagen are correlated with the estimated Mie scattering fractions: a lower collagen volume fraction corresponds to a higher Rayleigh scattering contribution.

Multiple class classification was performed with the CART method, on the five categories of breast tissue, i.e., adipose, glandular, FA, IC, and DCIS in order to evaluate the performance of such a diagnosis. Figure 5 depicts a decision tree that classifies all tissues based on a specific threshold value for each parameter. As can be seen, the first node allows discrimination of adipose tissue based on the lipid content. If the lipid volume fraction is above 40%, an acquired spectrum is considered to be taken in adipose tissue, otherwise it is another type of tissue. Table 3 corresponds to the confusion matrix displaying the diagnostic performance by comparing with the pathological diagnosis being the reference standard. Table 4 compares the sensitivity-specificity rates for each type of tissue when a LOO and HO cross validation were applied. The over-

all classification accuracy computed from the confusion matrix is 90% (879 out of 980). The type of tissue with the lowest sensitivity rate is FA, whereas adipose tissue has the highest specificity rate. The ROC curves for classification of each tissue are depicted in Fig. 6 including confidence intervals. Corresponding AUC and PDM measures are summarized in Table 5. From the AUC values, the performance of the diagnosis can be classified into three categories with adipose as the best performance (AUC almost 100%), glandular as the worst performance with AUC of 86%, and FA, IC, and DCIS as the median performance with comparable AUC values of roughly 92%. The PDM multiple classes overall performance of the diagnostic is 93.6%.

Table 6 summarizes the sensitivity-specificity obtained for classification of malignant versus nonmalignant tissues by using various algorithms for classification. The obtained numbers are compared to what has already been reported by other studies from different research groups. The best algorithm performance applied to the data was reached with KNN classification, whereas the poorest performance was reached with LDA classification.

Classification based on the CART method showed that discriminating adipose tissue based on the β-carotene values only yields to a sensitivity-specificity of 68%–92%, whereas classification based on the lipid parameters yields to a sensitivity-specificity of 98%–99%. Using both parameters yielded a sensitivity-specificity of 99%–99%.

4 Discussion

This study corresponds to the first of its kind that evaluates the classification of different types of breast tissue based on several parameters derived from spectroscopic data acquired from a wide wavelength range from 500 to 1600 nm. From Table 2, it is clear that adipose tissue is the easiest tissue that can be discriminated from the other types of tissue. More than seven parameters showed significant differences with lipid, water, and β-carotene as main discriminators, as supported by other studies as well.[9–22] Glandular and DCIS tissues have the lowest amount of parameters that showed significant differences with other
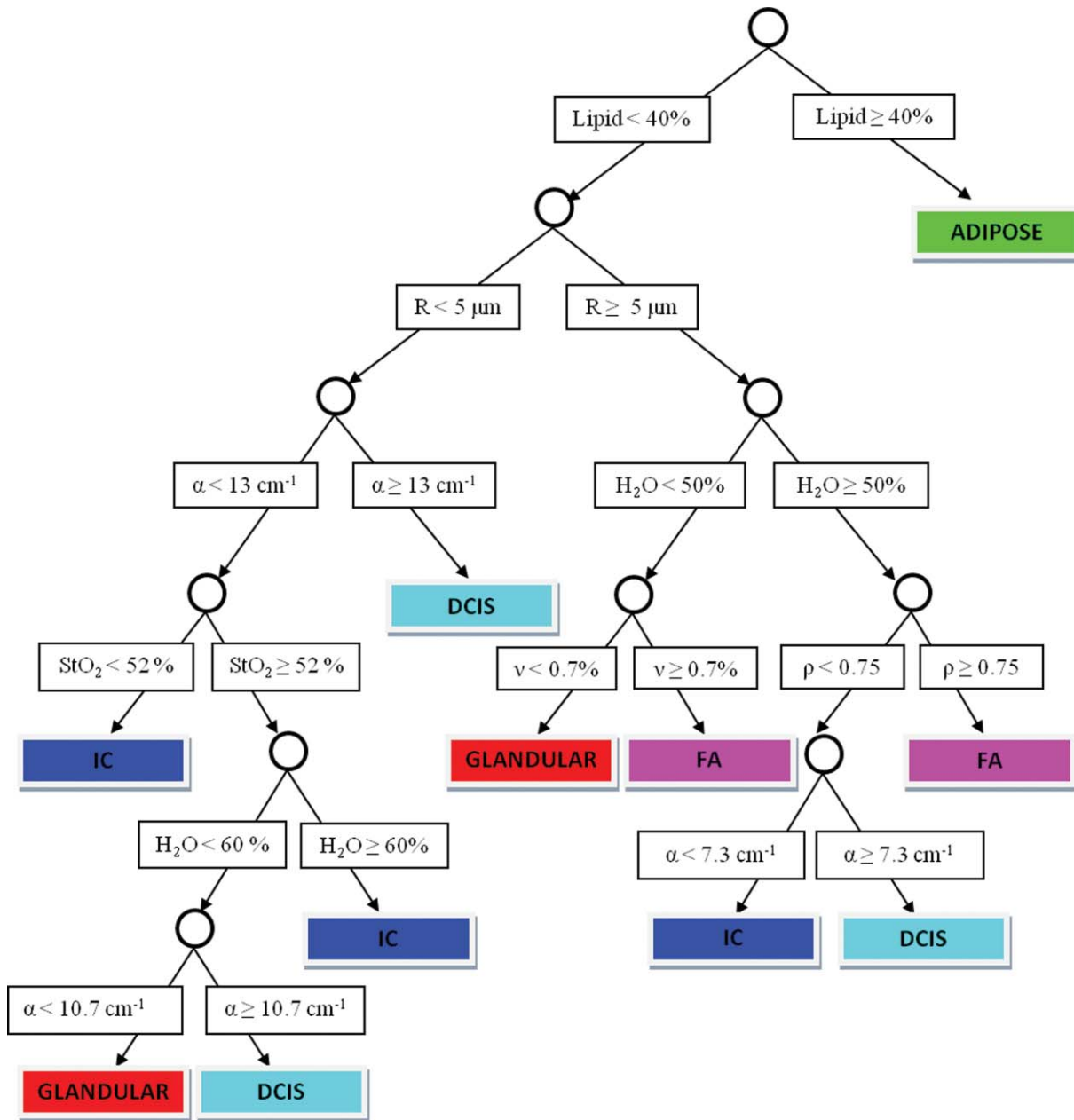
Fig. 5 Classification decision tree of the different breast types based on parameter threshold values.

tissues. It is important to note that the use of *post hoc* Tukey's test is essential for reliable statistical test for significance of differences. If not used, additional parameters become significantly different between types of tissue. For example, lipid becomes significantly different between glandular and FA as well as IC if no *post hoc* test is performed. Moreover, glandular tissue has on average 5% lipid, whereas FA and DCIS have below 0.5% lipid, while still no significant differences are observed according to Tukey's test despite an order of magnitude difference in lipid content. Similar to *ex vivo*[10,11,14] and *in vivo*[18,22,36] existing studies, we also observed a lower blood oxygenation level in malignant tissue compared to nonmalignant tissue with significant differences (cf. Table 2 and Fig. 4). This is expected because malignant tissues are known to exhibit regions of hypoxia.[36] Although this observation corroborates with other existing *ex vivo*

studies, this should be validated *in vivo* because the oxygenation level of the tissue can significantly change during and after tissue excision. Another interesting finding is that we observed comparable amounts of blood (0.3%) in malignant and nonmalignant tissue similar to reported studies in literature. Yet on the contrary, Van Veen et al.[22] have reported a higher blood volume fraction in malignant tissue similar to optical mammography studies that were conducted *in vivo* by Spinelli et al. on 190 patients from which 32 had cancer,[37] and by Grosenick et al. on 154 patients with 87 carcinoma cases[38] (in Ref. 38, Table 4 provides an overview on blood volume, oxygenation level, and reduced scattering properties of healthy and malignant tissue from various optical mammography studies available in literature). Among the malignant types of tissue, only DCIS exhibited a larger amount of blood than the average value of blood in

Table 3 Confusion matrix displaying classification of breast tissues using the CART algorithm for classification.

| Type of breast tissue (number of samples)	DRS classification diagnosis				
	Nonmalignant			Malignant	
	Adipose	Glandular	FA	IC	DCIS
Adipose (327)	**319**	6	2	0	0
Glandular (189)	7	**158**	8	15	1
FA (127)	0	11	**103**	11	2
IC (241)	0	15	4	**219**	3
DCIS (96)	0	5	1	10	**80**

Fig. 6 ROC curves (solid line) for classification of adipose (a), glandular (b), FA (c), IC (d), and DCIS (e) tissues including confidence intervals (dashed line) and corresponding AUC.

nonmalignant tissues. An increase in blood volume fractions is a potential marker for angiogenesis. Yet again, the current study was performed *ex vivo* and an *in vivo* study is required to confirm this observation as shown by Van Veen et al. Furthermore, our data corroborates with the observation from other studies with respect to the reduced scattering amplitude. Indeed, the reduced scattering amplitude is higher in malignant tissue (8.1 cm^{-1}) than in nonmalignant tissue (5.6 cm^{-1}). As expected, we observed that collagen volume fractions correlate with ρ. This is in agreement with findings of Saidi et al.[39] that the Rayleigh scattering in tissue is mainly due to sub-micrometer collagen fibers in the connective tissue, suggesting a stronger Rayleigh contribution in glandular, IC, and DCIS that contain the highest collagen volume fractions as depicted in Fig. 4. Water content is the most prominent in IC and is significantly different from the other tissues (cf. Table 2). The lowest water volume fraction is obviously observed for adipose tissue since lipid contains almost no water. However, collagen-rich stroma contains quite a lot of water as collagen fibers are hydrophilic. Thus, if a tumor induces a lot of stroma, the water content in the tumor will be relatively high as the fibers are loosely arranged, leaving a lot of space for water molecules to intervene in the tumor. In breast tissue with benign sclerotic changes in which collagen gets cross-linked to a great extent, hardly any space is left for water molecules. Necrosis can also play a role in water increase, but only a very small minority of all breast cancer contains necrotic areas in general.

Table 4 Sensitivity and specificity of CART classification of each type of tissue using LOO and 20-fold HO cross validation.

Type of breast tissue	Sensitivity (%) - Specificity (%)	
	Leave-one-out cross validation	Hold-out cross validation
Adipose	98–99	98 ± 1–99 ± 1
Glandular	84–95	80 ± 6–95 ± 2
FA	81–98	75 ± 9–97 ± 1
IC	91–95	86 ± 6–94 ± 2
DCIS	83–99	81 ± 10–98 ± 2
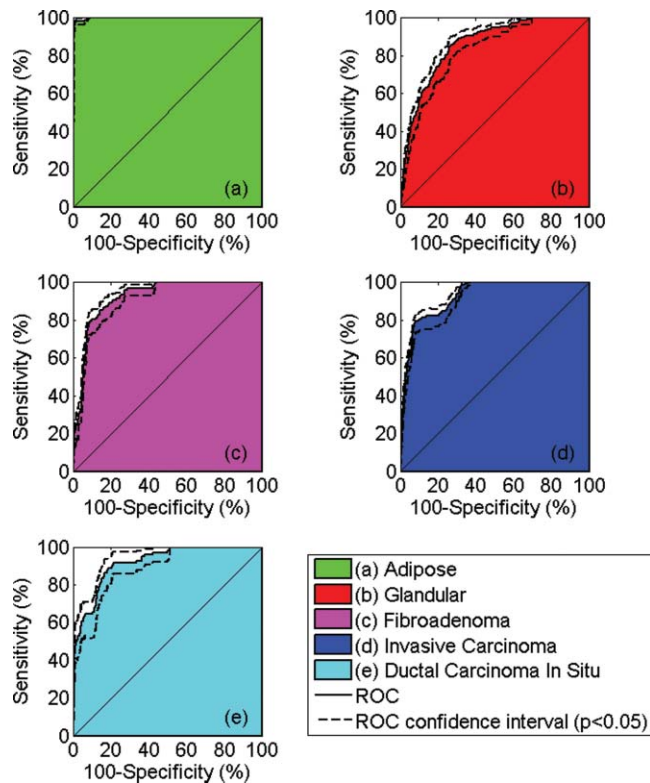
When investigating the differences in values of the different parameters within each patient instead of comparing tissue types from all the patients together, the *p*-values for discriminating a tissue from another becomes even smaller. Results of the spectral differences between different patients are to be presented in a future publication.

Multiple class classification with the CART algorithm demonstrated a very high overall diagnostic performance of 94% with the highest AUC value for adipose tissue and the lowest for glandular tissue as can be seen in the ROC curves in Fig. 6. In the study by Majumder et al., glandular and adipose tissues were both classified as normal and discriminated with FA, IC, and DCIS, and obtained an overall classification performance of 88% using sparse multinomial LR on the diffuse reflectance spectra.[15] Our data showed better sensitivity-specificity for IC and DCIS (cf. Table 4), whereas FA showed similar performance. It is important to note that the amount of DCIS samples by Majumder et al. is rather small compared to the other tissues and therefore a low performance can be expected. Moreover, DCIS is not a common tissue measured by other groups (6, 2, and 1 samples measured by Majumder et al.,[15] Zhu et al.,[11] and Laughney et al.,[17] respectively). It is recommended to acquire more spectra from a sample and to perform classification on the spectra in order not to bias the diagnosis by low numbers compared to other tissues.

For a better comparison of our results with existing literature studies, we have performed classification on our data using the techniques suggested in literature. Table 6 summarizes the classification algorithm used to discriminate malignant tissue (i.e.,

Table 5 AUC values of ROC curves for the five tissue types and PDM value.

Type of breast tissue	AUC	Confidence interval
Adipose	99.8%	99.7%–99.9%
Glandular	85.9%	81.9%–87.9%
FA	92.3%	90.4%–94.1%
IC	92.5%	90.9%–94.0%
DCIS	91.8%	88.7%–93.4%
PDM for total AUC	93.6%	91.9%–94.9%

IC and DCIS) from nonmalignant tissue (i.e., adipose, glandular, and FA) and compares the reported sensitivity-specificity values as reported in literature with those obtained in our study. The sensitivity-specificity obtained with the same method for all categories of breast classification, i.e., the CART method, is 88%–93%. Compared to the results by Bigio et al., ANN classification applied to our data showed better sensitivity-specificity. However, it is important to note that our classification was performed on the parameters derived from the spectra and not by applying the ANN classification to the spectra as done by Bigio et al. The KNN classification employing Mahalanobis distance metric to account for parameter intercorrelations, as used by Laughney et al., showed the highest diagnostic performance. The sensitivity-specificity obtained with this method corresponds to a very high performance. One might question the classification method because the KNN algorithm can be biased, since it is very sensitive to redundant or similar features because all features contribute to the similarity principle and thus to the classification. Using an LR-based classification method showed a sensitivity-specificity of 82%–94%, which is outperformed by the sensitivity-specificity of 100%–

100% obtained by Volynskaya et al. However, it is important to note that the study by Volynskaya et al. did not include the 6 DCIS samples they measured because they considered it was a very small number compared to the 9 IC, 9 FA, 31 normal, and 55 fibrocystic change samples they had measured. Keller et al. obtained a lower sensitivity but higher specificity compared to our result. It should be noted that they classified the spectra and not the parameters that were derived from the spectra. Moreover, they included fluorescence spectra to their classification scheme. Zhu et al. achieved a sensitivity-specificity of 83%–87% and $82 \pm 5\%$–$89 \pm 5\%$ using SVM classification with a LOO and HO cross validation scheme, respectively. In our study, the sensitivity-specificity with linear SVM classification is 79%–93% and $81 \pm 4\%$–$93 \pm 2\%$ with a LOO and HO cross validation scheme, respectively. The data used for classification by Zhu et al. correspond to 89% of adipose tissue within the nonmalignant category and 74% of IC among the malignant category. Therefore, the weight of classification is mainly dominated by adipose and IC for malignant and nonmalignant tissue, respectively. In our study, 51% of the nonmalignant samples correspond to adipose tissue and 71% of the malignant tissue samples correspond to IC. However, we have performed the classification by taking the same amount of spectra for each type of tissue within the malignant and nonmalignant category in order to avoid an over-representation of adipose and IC within the nonmalignant and malignant category, respectively. This can explain the fact that we observe a lower sensitivity than Zhu et al. However, we obtain a higher specificity suggesting that we can classify nonmalignant tissue better thanks to the additional parameters derived from the fit-model. In the case of nonlinear SVM, i.e., using a Gaussian radial basis kernel function instead of a linear kernel, the performance of discriminating malignant tissue increased to a sensitivity-specificity of 82%–94%. As mentioned by Zhu et al. in a previous study,[10] a classification based on linear algorithms could underperform in the diagnosis since the optical properties are nonlinearly related in the description of the measured spectra, hence the better performance of the nonlinear, compared to the linear, SVM algorithm applied to

Table 6 Literature overview of diagnostic performance in discriminating malignant from nonmalignant tissue, and comparison of different classification algorithms applied to the data in the presented study.

Classification algorithm	Reference	Sens. (%)-Spec. (%)	Sens.-Spec. of this study (LOO)	Sens.-Spec. of this study (HO)
Artificial neural network	Bigio et al.[a]	69–85	89–98	91 ± 2–96 ± 4
K-nearest neighbor	Laughney et al.	90–77	96–99	94 ± 4–98 ± 2
Logistic regression	Volynskaya et al.	100–100[b]	82–94	82 ± 3–94 ± 6
Logistic regression	Keller et al.[a]	85–96[b]	82–94	82 ± 3–94 ± 6
Linear support vector machine	Zhu et al.	83–87[b]	79–93	81 ± 4–93 ± 2
Nonlinear support vector machine	–	–	90–97	88 ± 4–97 ± 2
Linear discriminant analysis	–	–	78–95	74 ± 6–96 ± 2
Classification and regression tree	–	–	88–93	85 ± 6–92 ± 3

[a]Sensitivity and specificity computed after classification of the spectra and not from the parameters derived from a fit-model.
[b]Fluorescence was also measured in these studies, however, the reported sensitivity-specificity corresponds to classification of parameters derived from diffuse optical spectroscopy only except for Keller et al.

our data. Besides, for comparison with another nonlinear classification method, a sensitivity-specificity of 78%–95% reached with LDA employing Mahalanobis distance yielded the lowest performance among the other classification methods with respect to specificity. From the various classification methods, large variations in sensitivity-specificity can be achieved, and therefore care should be taken when comparing one's results with existing results in literature. The choice of the classification algorithm is very important and the sample sizes, the methods, and the linearity of the problem should be carefully taken into consideration. As a matter of fact, for classification problems with small sample size, LDA is not suitable, as it is a parametric method assuming normal distribution of the data in each class. Other methods have the advantage of being nonparametric methods. However, KNN is very sensitive to redundant and similar features for classification. On the other hand, linear SVM finds linear separation of two classes in the training set with a hyperplane that has maximal distance from the two classes. If the groups are not linearly separable, non-linear SVM can be applied.[40] The LR algorithm is a probabilistic method that has the advantage of using few or no statistical assumptions, but the drawback is that the complete data is needed for each class to calculate the probabilities. Hence, large variations in sample sizes can bias the classification.

Among the studies quoted in Table 6, the study by Keller et al., Majumder et al., Zhu et al., and Volynskaya et al. performed fluorescence measurements. The latter two derived collagen and nicotinamide adenine dinucleotide (NADH) concentrations from fitting the fluorescence spectra. Both studies showed a significant increase in collagen in malignant tissue compared to nonmalignant tissue which correlates with our finding where we estimated collagen with diffuse reflectance spectroscopy measurements. Adding fluorescence to diffuse optical spectroscopy did not result in the same conclusions for the different studies. Volynskaya et al. showed a decrease in specificity from 100% to 96%, whereas Zhu et al. did not observe any differences in performance. Majumder et al. showed a tremendous improvement in discriminating the tumor types of tissue when adding fluorescence measurements to the classification routine, increasing the overall diagnosis performance from 88% without fluorescence to 95% with fluorescence. Apart from fluorescence, Majumder et al. performed Raman spectroscopy measurements and showed that this optical tissue measurement technique yields the best overall performance (99%). The group of biomedical photonics at MIT presented several Raman studies[41–44] showing that they can reach sensitivity-specificity of 83%–93% by classifying estimated parameters similar to those extracted from the measurements presented in this paper such as β-carotene, lipid, and collagen, as well as additional biological substances such as calcium, cholesterol, and cell nucleus. Interestingly, in comparison with the Raman results of the study conducted by the MIT group, the average collagen and fat fractions are reasonably similar for the different types of tissue except for adipose tissue where we estimate an average collagen content of 15%. In the latest study from the MIT group,[42] the 20 spectra acquired from DCIS samples were not classified because this type of tissue was not encountered in the calibration data set they used for their diagnostic algorithm development. They do discuss, nevertheless, that applying their algorithm to the DCIS samples would result in 5 samples out of 20 to be classified as malig-

nant based on classification of their estimated fat and collagen fractions derived from the fitted Raman spectra.

One final point of discussion concerns adipose tissue discrimination. Both lipid and β-carotene are adipose tissue precursors and only one of them was used in previous studies. In this paper, we estimate both chromophores and from classifying adipose tissue based on only one of the chromophores, it turned out that lipid is the best discriminator for adipose tissue with sensitivity-specificity of 98%–99% versus 68%–92% for β-carotene. It is known that β-carotene is significantly lower in smokers than nonsmokers.[45] Thus, it can bias the discrimination of adipose tissue in the breast, depending whether a patient is a smoker or not, making lipid a more suitable discriminator.

5 Conclusion

We present the first breast diagnosis study based on estimating morphological, physiological, and optical parameters derived from diffuse reflectance spectroscopy measurements on a 500 to 1600 nm wavelength range. Based on a classification and regression tree algorithm applied to the derived parameters, a sensitivity-specificity of 98%–99%, 84%–95%, 81%–98%, 91%–95%, and 83%–99% was obtained for discrimination of adipose, glandular, fibroadenoma, invasive carcinoma, and ductal carcinoma *in situ*, respectively; and a multiple classes overall diagnostic performance of 94%. A comparison of different classification techniques to discriminate malignant and nonmalignant tissue showed varying performance that can highly depend on the classification algorithm. Finally, to the best of our knowledge, given the fact this is the only study that estimates both β-carotene and lipid as adipose tissue precursor; we show that lipid is a much better discriminator with sensitivity–specificity of 98%–99% for lipid versus 68%–92% for β-carotene.

References

1. R. Ariga, K. Bloom, V. Reddy, L. Kluskens, D. Francescatti, K. Dowlat, P. Siziopikou, and P. Gattuso, "Fine-needle aspiration of clinically suspicious palpable breast masses with histopathological correlation," *Am. J. Surg.* **184**, 410–413 (2002).
2. R. Burns, J. Brown, S. Roe, L. Sprouse, A. Yancey, and L. Whitherspoon, "Sterotactic core-needle breast biopsy by surgeons: minimum 2-year follow-up of benign lesions," *Ann. Surg.* **232**, 542–548 (2000).
3. B. Chaiwun, J. Settakorn, C. Ya-In, W. Wisedmongkol, S. Randaeng, and P. Thorner, "Effectiveness of fine-needle aspiration cytology of breast: analysis of 2375 cases from northern Thaïland," *Diagn. Cytopathol.* **26**, 201–205 (2002).
4. R. Dameron, D. de Long, A. Fisher, D. de Long, L. Dodd, and R. Nelson, "Indeterminate findings on imaging-guided biopsy: should additional intervention be pursued?," *Am. J. Roentgenol.* **173**, 461–464 (1999).

5. J. Youk, E. Kim, L. Lee, and K. Oh, "Missed breast cancer at US-guided core needle biopsy: how to reduce them," *Radiographics* **27**, 79–94 (2007).

6. S. Willis and I. Ramzy, "Analysis of false results in a series of 835 fine needle aspirates of breast lesions," *Acta. Cytol.* **39**, 858–864 (1995).

7. R. G. Pleijhuis, M. Graafland, J. De Vries, J. Bart, J. S. De Jong, and G. M. Van Dam, "Obtaining adequate surgical margins in breast-conserving therapy for patients with early-stage breast cancer: current modalities and future directions," *Ann. Surg. Oncol.* **16**, 2717–2730 (2009).

8. E. D. Kumiawan, M. H. Wong, I. Windle, A. Rose, A. Mou, M. Buchanan, J. P. Collins, J. A. Miller, R. L. Gruen, and G. B. Mann, "Predictors of surgical margin status in breast-conserving surgery within a breast screening program," *Ann. Surg. Oncol.* **15**, 2542–2549 (2008).

9. I. J. Bigio, S. G. Bown, G. Briggs, C. Kelley, S. Lakhani, D. Pickard, P. M. Ripley, I. G. Rose, and C. Saunders, "Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results," *J. Biomed. Opt.* **5**, 221–228 (2000).

10. C. Zhu, G. Palmer, T. Breslin, J. Harter, and N. Ramanujam, "Diagnosis of breast cancer using diffuse reflectance spectroscopy: comparison of a Monte Carlo versus partial least squares analysis based feature extraction technique," *Lasers Surg. Med.* **38**, 714–724 (2006).

11. C. Zhu, G. M. Palmer, T. M. Breslin, J. Harter, and N. Ramanujam, "Diagnosis of breast cancer using fluorescence and diffuse reflectance spectroscopy: a Monte-Carlo-modeled-based approach," *J. Biomed. Opt.* **13**, 034015 (2008).

12. C. Zhu, T. M. Breslin, J. Harter, and N. Ramanujam, "Model based and empirical spectral analysis for the diagnosis of breast cancer," *Opt. Express* **16**, 14961–14978 (2008).

13. T. M. Byldon, S. A. Kennedy, L. M. Richards, J. Q. Brown, B. Yu, M. K. Junker, J. Callagher, J. Geradts, L. G. Wilke, and N. Ramanujam, "Performance metrics of an optical spectral imaging system for intra-operative assessment of breast tumor margins," *Opt. Express* **18**, 8058–8076 (2010).

14. Z. Volynskaya, A. S. Haka, K. L. Betchel, M. Fitzmaurice, R. Schenk, N. Wang, J. Nazemi, R. R. Dasari, and M. S. Feld, "Diagnosing breast cancer using diffuse reflectance spectroscopy and intrinsic fluorescence spectroscopy," *J. Biomed. Opt.* **13**, 024012 (2008).

15. S. K. Majumder, M. D. Keller, F. I. Boulos, M. C. Kelley, and A. Mahadevan-Jansen, "Comparison of autofluorescence, diffuse reflectance, and Raman spectroscopy for breast tissue discrimination," *J. Biomed. Opt.* **13**, 054009 (2008).

16. M. D. Keller, S. K. Majumder, M. C. Kelley, I. M. Meszoely, F. I. Boulos, G. M. Olivares, and A. Mahadevan-Jansen, "Autofluorescence and diffuse reflectance spectroscopy and spectral imaging for breast surgical margin analysis," *Lasers Surg. Med.* **42**, 15–23 (2010).

17. A. M. Laughney, V. Krishnaswamy, P. B. Garcia-Allende, O. M. Conde, W. A. Wells, K. D. Paulsen, and B. W. Pogue, "Automated classification of breast pathology using local measures of broadband light," *J. Biomed. Opt.* **15**, 066019 (2010).

18. A. E. Cerussi, N. Shah, D. Hsiang, A. Durkin, J. Butler, and B. J. Tromberg, "*In vivo* absorption, scattering of 58 malignant breast tumors determined by broadband diffuse optical spectroscopy," *J. Biomed. Opt.* **11**(4), 044005 (2006).

19. P. Taroni, D. Comelli, A. Pifferi, A. Torricelli, and R. Cubeddu, "Absorption of collagen: effects on the estimate of breast composition and related diagnostic implications," *J. Biomed. Opt.* **12**, 014021 (2007).

20. P. Taroni, A. Bassi, D. Comelli, A. Farina, R. Cubeddu, and A. Pifferi, "Diffuse optical spectroscopy of breast tissue extended to 1100 nm," *J. Biomed. Opt.* **14**, 054030 (2009).

21. P. Taroni, A. Pifferi, G. Quarto, L. Spinelli, A. Torricelli, F. Abbate, A. Villa, N. Balestreri, S. Menna, E. Cassano, and R. Cubeddu, "Noninvasive assessment of breast cancer risk using time-resolved diffuse optical spectrscopy," *J. Biomed. Opt.* **15**, 060501 (2010).

22. R. L. P. van Veen, A. Amelink, M. Menke-Pluymers, C. van der Pol, and H. J. C. M. Sterenborg, "Opitcal biopsy of breast tissue using differential path-length spectroscopy," *Phys. Med. Biol.* **50**, 2573–2581 (2005).

23. R. Nachabé, B. H. W. Hendriks, A. E. Desjardins, M. van der Voort, and H. J. C. M. Sterenborg, "Estimation of lipid and water concentrations in scattering media with diffuse optical spectroscopy from 900 to 1600 nm," *J. Biomed. Opt.* **15**, 037015 (2010).

24. R. Nachabé, B. H. W. Hendriks, M. van der Voort, A. E. Desjardins, and H. J. C. M. Sterenborg, "Estimation of biological chromophores using diffuse optical spectroscopy: benefit of extending the UV-VIS wavelength range to include 1000 to 1600 nm," *Biomed. Opt. Exp.* **1**, 1432–1442 (2010).

25. F. Provost and P. Domingos, "Well-trained PETs: improving probability estimation trees," CeDER Working Paper #IS-00-04, Stern School of Business, New York University, New York (2000).

26. R. Nachabé, D. Evers, B. H. W. Hendriks, G. W. Lucassen, M. Van der Voort, J. Wesseling, and T. J. Ruers, "Effect of bile absorption coefficients on the estimation of liver tissue optical properties and related implications in discriminating healthy from tumorous samples," *Biomed. Opt. Exp.* **2**, 600–614 (2011).

27. T. J. Farrell, M. S. Patterson, and B. Wilson, "A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the non-invasive determination of tissue optical properties," *Med. Phys.* **19**, 879–888 (1992).

28. W. Verkruysse, G. W. Lucassen, J. F. de Boer, D. J. Smithies, J. S. Nelson, and M. J. C. van Gemert, "Modeling light distributions of homogenous versus discrete absorbers in light irradiated turbid media," *Phys. Med. Biol.* **42**, 51–65 (1997).

29. W. G. Zijlstra, A. Buursma, and O. W. van Assendelft, *Visible and Near Infrared Absorption Spectra of Human and Animal Haemoglobin*, VSP Publishing, Utrecht, The Netherlands (2000).

30. S. W. van de Poll, Raman spectroscopy of atherosclerosis, PhD thesis, University of Leiden (2003).

31. C. L. Tsai, J. C. Chen, and W. J. Wang, "Near-infrared absorption property of biological soft tissue constituents," *J. Med. Bio. Eng.* **21**, 7–14 (2001).

32. A. S. Nunez, *A Physical Model of Human Skin and its Application for Search and Rescue*, Air Force Institute of Technology, Ohio (2009).

33. A. Amelink, D. J. Robinson, and H. J. C. M. Sterenborg, "Confidence interval on fit parameters derived from optical reflectance spectroscopy measurements," *J. Biomed. Opt.* **13**, 054044 (2008).

34. H. Motulsky, *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*, Oxford University Press, New York (2010).

35. L. Breiman, *Classification and Regression Trees*, Wadsworth International Group, Belmont, CA (1984).

36. J. Q. Brown, L. G. Wilke, J. Geradts, S. A. Kennedy, G. M. Palmer, and N. Ramanujam, "Quantitative optical spectroscopy: a robust tool for direct measurement of breast cancer vascular oxygenation and total hemoglobin content *in vivo*," *Cancer Res.* **69**, 2919–2926 (2009).

37. L. Spinelli, A. Torricelli, A. Pifferi, P. Taroni, G. Danesini, and R. Cubeddu, "Characterization of female breast lesions from multi-wavelength time-resolved optical mammography," *Phys. Med. Biol.* **50**, 2489–2502 (2005).

38. D. Grosenick, H. Wabnitz, K. T. Moesta, J. Mucke, P. M. Schlag, and H. Rinneberg, "Time-domain scanning optical mammography: II. Optical properties and tissue parameters of 87 carcinomas," *Phys. Med. Biol.* **50**, 2451–2468 (2005).

39. I. S. Saidi, S. L. Jacques, and F. K. Tittel, "Mie and Rayleigh modeling of visible-light scattering in neonatal skin," *Appl. Opt.* **34**, 7410–7418 (1995).

40. T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York (2001).

41. K. E. Shafer-Peltier, A. S. Haka, M. Fitzmaurice, J. Crowe, J. Myles, R. R. Dasari, and M. S. Feld, "Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis *in vivo*," *J. Raman Spectros.* **33**, 552–563 (2002).

42. A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, and M. S. Feld, "Diagnosing breast cancer by using Raman spectroscopy," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12371–12376 (2005).

43. A. S. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R. R. Dasari, J. P. Crowe, and M. S. Feld, "*In vivo* margin assessment during partial mastectomy breast surgery using Raman spectroscopy," *Cancer Res.* **66**, 12371–12376 (2006).

44. A. S. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, R. Shenk, N. Wang, R. R. Dasari, M. Fitzmaurice, and M. S. Feld, "Diagnosing breast cancer using Raman spectroscopy: prospective analysis," *J. Biomed. Opt.* **14**, 054023 (2009).

45. R. M. Russel, "Beta-carotene and lung cancer," *Pure Appl. Chem.* **74**, 1461–1467 (2002).