

RESEARCH

Open Access



Towards the use of diffuse reflectance spectroscopy for real-time in vivo detection of breast cancer during surgery

Lisanne L. de Boer^{1*}, Torre M. Bydlon², Frederieke van Duijnhoven¹, Marie-Jeanne T. F. D. Vranken Peeters¹, Claudette E. Loo³, Gonneke A. O. Winter-Warnars³, Joyce Sanders⁴, Henricus J. C. M. Sterenberg^{1,5}, Benno H. W. Hendriks^{2,6} and Theo J. M. Ruers^{1,7}

Abstract

Background: Breast cancer surgeons struggle with differentiating healthy tissue from cancer at the resection margin during surgery. We report on the feasibility of using diffuse reflectance spectroscopy (DRS) for real-time in vivo tissue characterization.

Methods: Evaluating feasibility of the technology requires a setting in which measurements, imaging and pathology have the best possible correlation. For this purpose an optical biopsy needle was used that had integrated optical fibers at the tip of the needle. This approach enabled the best possible correlation between optical measurement volume and tissue histology. With this optical biopsy needle we acquired real-time DRS data of normal tissue and tumor tissue in 27 patients that underwent an ultrasound guided breast biopsy procedure. Five additional patients were measured in continuous mode in which we obtained DRS measurements along the entire biopsy needle trajectory. We developed and compared three different support vector machine based classification models to classify the DRS measurements.

Results: With DRS malignant tissue could be discriminated from healthy tissue. The classification model that was based on eight selected wavelengths had the highest accuracy and Matthews Correlation Coefficient (MCC) of 0.93 and 0.87, respectively. In three patients that were measured in continuous mode and had malignant tissue in their biopsy specimen, a clear transition was seen in the classified DRS measurements going from healthy tissue to tumor tissue. This transition was not seen in the other two continuously measured patients that had benign tissue in their biopsy specimen.

Conclusions: It was concluded that DRS is feasible for integration in a surgical tool that could assist the breast surgeon in detecting positive resection margins during breast surgery.

Trail registration NIH US National Library of Medicine—clinicaltrials.gov, NCT01730365. Registered: 10/04/2012 <https://clinicaltrials.gov/ct2/show/study/NCT01730365>

Keywords: Breast cancer surgery, Intraoperative margin assessment, Optical technology, Real-time

*Correspondence: l.d.boer@nki.nl

¹ Department of Surgery, the Netherlands Cancer Institute—Antoni van Leeuwenhoek, Plesmanlaan 121, Postbus 90203, 1066 CX Amsterdam, The Netherlands

Full list of author information is available at the end of the article



Background

The current primary treatment of breast cancer includes a multimodal approach with a combination of surgery and radiotherapy, and depending on the subtype and the extent of the disease, systemic treatment. Optimal surgical treatment is achieved when all tumor tissue is resected, and thus histopathological evaluation of the resection specimen reveals no tumor positive margins. However, resecting too much healthy tissue compromises cosmetic outcome. Since tumor positive resection margins are associated with a higher recurrence rate, these patients require additional treatment with boost radiotherapy or re-excision surgery [1, 2]. Along with the impact of additional treatment on healthcare budgets, both boost radiotherapy and secondary surgery impair cosmetic outcomes [3], increase morbidity [4, 5] and can affect quality of life [6–8]. The surgeon is thus balancing between completely resecting the tumor and, sparing as much healthy tissue as possible [9]. Performing surgery while doing justice to both is difficult since visually recognizing tumor tissue on the surgical margin is extremely challenging. In addition, all currently available intra-operative margin assessment techniques have their own pitfalls, such as reduced sensitivity, requirement of skilled personnel, are labor-intensive, and can be operator dependent [10–12]. Therefore, breast surgeons are in need of a robust margin assessment tool that can assist them real-time in defining the optimal resection plane to ensure clear margins, and will also help them to minimize the resected specimen volume [13–18].

Diffuse reflectance spectroscopy (DRS), a light-based technology, has shown promising results for discriminating normal breast tissue from tumor tissue and may address the need for a margin assessment device [19–22]. The principle behind DRS is that light interacts with tissue through scattering and absorption. The absorption is related to the chemical composition of the tissue whereas the scattering is related to the subcellular morphology. The reflected light, detected after tissue interaction, has an altered spectrum compared to the incoming light. Thus, the diffuse reflectance spectrum represents aspects of the composition and subcellular morphology of the measured tissue. Ultimately, incorporating DRS technology in instruments, such as a surgical knife, could potentially provide the surgeon with additional information that reflects the histopathology of the tissue at the resection margin. The surgeon can use this information as guidance to determining the optimal resection place for excising a breast tumor.

There have been some publications on the feasibility of DRS for breast biopsy and surgery applications [23–26]. However, many of these publications struggled with correlating the exact tissue volume measured by

DRS in vivo to the proper location in the histopathology slides processed post-operatively. This is an important factor as the ‘gold standard’ for the evaluation of surgical margins is microscopic assessment by a pathologist. Thus, a mismatch between the optical measurements and the histopathology hampers the development of robust classification algorithms and validation of the technology. For the purpose of developing a reliable database of DRS measurements we developed a special biopsy needle with embedded optical fibers [27]. This tool enables DRS measurements and subsequent biopsy of the same tissue volume as measured spectroscopically. Although it is tempting to see this study as an attempt to perform spectroscopy guided biopsy, this was explicitly not the purpose of this study. The robust dataset gathered in this setting can be used for developing classification models and validating DRS technology for tissue characterization. The paper describes how we developed and tested three predictive classification models, based on different types of input data, to accurately classify the DRS measurements. Furthermore, we investigated the feasibility of acquiring in vivo DRS measurements and classifying these. To this end, DRS measurements were performed continuously along an entire needle trajectory during ultrasound guided breast biopsy procedures, and classified based on a classification model.

Methods

Study design

Patients suspected of having breast cancer (after palpation, X-ray and US-imaging) that required diagnostic biopsies were asked to participate in this observational study that was approved by the Institutional Review Board of the Netherlands Cancer Institute. Written informed consent was obtained from all patients prior to the biopsy procedure. Patients with suspected sensitivity to light (e.g. patient who have had photodynamic therapy) were excluded, as well as patients that had received prior chemotherapy, endocrine therapy or radiation therapy recently (i.e. within 5 years). Patients were also excluded with breast implants and those that needed a stereotactic breast biopsy. In all patients, a biopsy was obtained after the last measurement in tumor tissue. The biopsy specimen was colored with pathology ink on the distal side to indicate which side had been in contact with the fibers during measurements. The biopsy specimens were used for diagnostic assessment and to confirm histopathology of the tumor measurement location by evaluating the first 2 mm of the side of the biopsy specimen that had been in contact with the optical fibers. Patients were only included when the pathology results of the lesion (including the other diagnostic biopsy specimens) indicated the lesion was an invasive breast tumor.

Spectrometers and optical biopsy needle

DRS measurements were obtained with a specially designed optical biopsy needle (Fig. 1). This optical biopsy needle, which had integrated optical fibers, combined the ability to measure DRS spectra with biopsy functionality [27]. The 14G optical biopsy needle had one 100 μm fiber for illumination and two 200 μm fibers for collecting the light (Invivo); with a 20 mm cavity for the tissue biopsy. The collecting fibers were placed next to each other, and the distance between the illuminating and collecting fibers was 1.36 mm, which resulted in a penetration depth of approximately 1–2 mm. The optical biopsy needle was attached to two spectrometers that resolved light in the visual wavelength range (DU420A-BRDD, Andor Technology) and the NIR wavelength range (DU492A-1.7, Andor Technology). After measuring, the spectra of the two spectrometers were stitched together, to form a continuous spectrum between 400 and 1600 nm [28]. In the full spectrum and selected wavelengths classification models the first 100 nm was removed since this wavelength range was highly affected by noise which may influence the machine learning algorithms that were used for the development of the models.

DRS point measurements

In 27 patients, measurements were acquired in a point-based manner, thus the measurement needle was first held still in normal tissue, and subsequently in the breast tumor where a biopsy was taken. At each measurement location, three DRS measurements (10 spectra per measurement) were acquired and averaged. Performing a single measurement took approximately 10 s. In this measurement time a total of 30 high quality DRS spectra were obtained over the full wavelength range which is necessary for building a database that can be used for

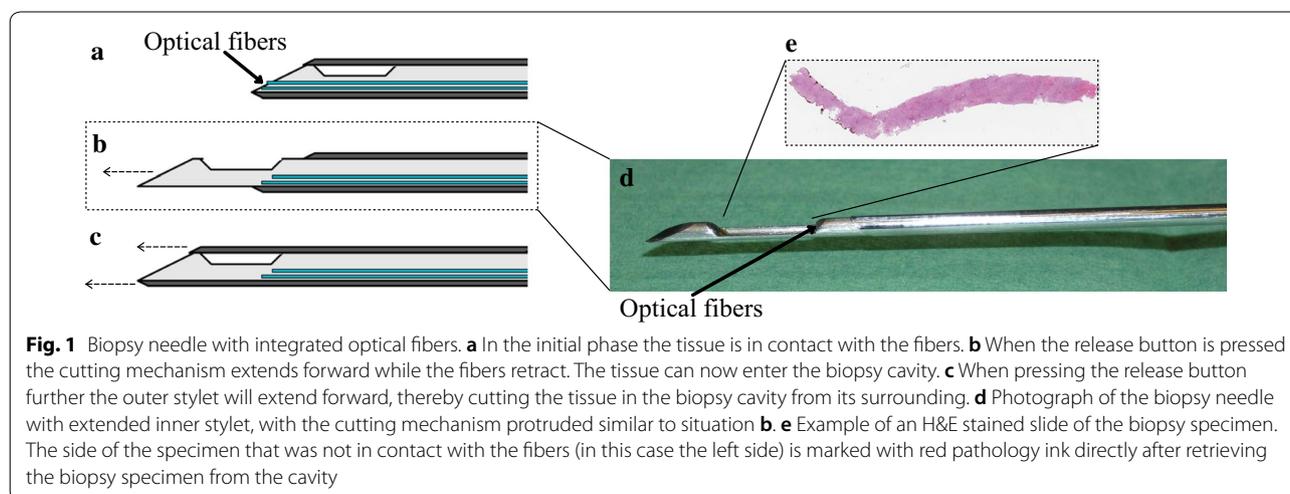
classification model development. During the time the DRS measurements were obtained an US-image was made with the needle tip in view. These US-images were evaluated by a radiologist to confirm correct positioning of the needle in either normal breast tissue or tumor tissue.

DRS continuous measurements

In five additional patients, DRS measurements were obtained in a continuous mode. Here the measurements were obtained along the entire needle trajectory starting in the normal tissue, continuing through the transition zone of normal-tumor, and ending in the tumor. To enable real-time acquisition of the data adjustments were made to the settings to increase the acquisition rate. The framerate in continuous mode was approximately one spectrum per second. The integration time was set to 0.35 s for all measurements along the needle trajectory. Prior to measuring the clocks of the US device and the laptop controlling the DRS set-up were synchronized thus allowing the US-images to be registered to the DRS measurements. At one location at a distance from the tumor, and at the final measurement location (also the location of the biopsy specimen) the needle was kept still to ensure sufficient data of both healthy tissue and tumor tissue (similar to the point measurements). At these locations, 10 DRS measurements were acquired, as well as an US-image.

Pre-processing point measurements

At each point measurement location three spectra were obtained which were averaged to calculate a mean spectrum of each measurement location. Subsequently all spectra were normalized with the standard normal variate (SNV) method [29]. Outlier detection was performed



to ensure that the classification models were developed with reliable data [30]. A cut-off of 3 times the standard deviation was chosen as threshold.

Development of classification models

For development of the first classification models only the point-based measurements were used as for these measurement US images and histopathology provided information on the nature of the tissue in front of the needle during a measurement. For the continuous data this information was not available for each measurement in the trajectory.

To ensure a balanced dataset, only patients which had both the normal and tumor measurement available after outlier detection were used. All classification models were constructed with perClass (Academic version 5.0, PR Sys design) in Matlab (2015a, the MathWorks). Figure 2 depicts a schematic flowchart of the approach to build the classification models.

Fit parameter classification models

The input for the first classification model was optical fit parameter data. To calculate these fit parameters from the measured DRS spectra, the measurements were quantified using an analytical fit model based on diffusion theory. This can be considered as a feature reduction method in which the measured spectra is translated into chemically or physiologically meaningful parameters [31]. In order to do so, the fit model required the absorption spectra of substances present in tissue, including: blood, fat, water, β -carotene, collagen, and bilirubin. The fit then optimized the parameters in such a way that the modelled spectrum matched the measured spectrum. The optical fit parameters generated by the fit model were: amount of blood (%), oxygen saturation (StO_2), total amount of fat plus water, fraction of fat, scattering at 800 nm, α and b (from the formula describing the reduced scattering, $\mu'_s = \alpha\lambda^{-b}$) amount of bilirubin, fraction of Mie scattering (in relation to the total scattering), amount of β -Carotene, and amount of collagen. The amount of water was calculated from the optical parameters describing the total amount of fat and water and the fat fraction. The amount of water together with the amount of fat allowed deriving the ratio between fat and water for each measurement location [21, 28].

Different combinations of fit parameters were used as input for this classification model. To limit the required computational effort, these combinations were formed by combining fit parameters that had shown the ability to discriminate between normal and tumor tissue previously, i.e. blood, StO_2 , scattering at 800 nm, fraction Mie scattering, β -carotene, collagen and, the ratio between fat and water (F/W-ratio). With these seven fit parameters

combinations (c) were made that consisted of either one or multiple fit parameters with one combination including all seven fit parameters ($c=127$, $c = \sum_{k=1}^7 \frac{n!}{(n-k)!k!}$) with n the number of fit parameters to choose from, and k the amount of elements in the combination). As the F/W-ratio previously proved to be an excellent discriminator [21, 32], only fit parameter combinations that included the F/W-ratio were used as input for the first classification model. The final set of combinations consisted of 64 possibilities.

Full spectrum classification model

The input for the second classification model was the full wavelength spectrum without any feature reduction. Thus, each wavelength between 500 and 1600 nm (i.e. the full spectrum) was used as input to the model, resulting in 1100 features for the classification model.

Selected wavelengths classification model

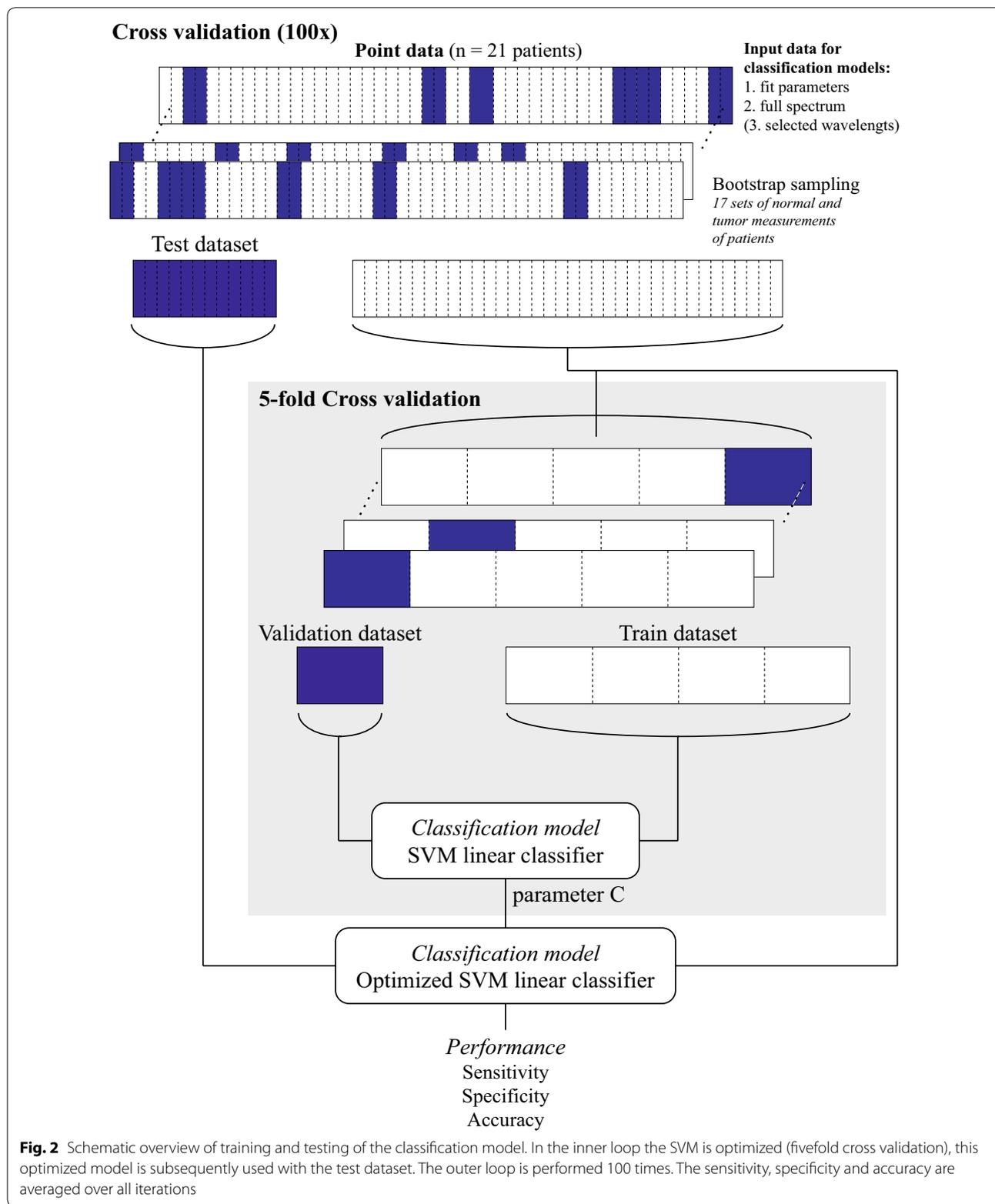
The input for the third classification model consisted of a limited number of wavelengths. The selection of these wavelengths was based on the results of a two-sided Wilcoxon rank sum test ($\alpha=0.05$). All normal and all tumor measurements were used in the test. For each wavelength the Wilcoxon rank sum test assess whether two samples of observations (in this case the normal measurements and the tumor measurements) are from the same distribution. This statistical test was used to identify wavelength regions that were significantly different between the normal and tumor spectra (mean p value below 0.05). In each of these regions, the wavelength with the lowest p -value was selected as a wavelength for the selected wavelengths model.

Classifier

A linear support vector machine (SVM) formed the classifier in the classification model [33, 34]. This machine learning technique constructed an optimal separating linear hyperplane between two classes in a higher dimensional space by creating the biggest margin between measurements of two classes.

Bootstrap sampling and cross-validation

To avoid selection bias, a bootstrapping technique was used to randomly select 17 subsets of (not necessarily different) patients as training data for model development. On average, both the normal and tumor measurements of $11.9 (\pm 1.3)$ unique patients were used as training data in each of the 100 iterations. The remaining patients that were not selected for training were used for testing the model. Subsequently the SVM was optimized with a five-fold cross validation of the training data to find the optimum for the regularization parameter C . The unseen test



data was then classified with the optimized SVM model. This process was performed a hundred times in order to test generalizability of the classification model. Measurements provided to the classification model were classified as either 'normal' or 'tumor'. With this binary output the sensitivity, specificity, and accuracy in each bootstrap iteration was calculated for discriminating normal measurements from tumor measurements. By averaging these model performance parameters over the 100 iterations the mean sensitivity, mean specificity, mean accuracy, and mean Matthews Correlation Coefficient (MCC) were determined, and these were used to compare the performance of different classification models. Besides the binary output, the classification models could also generate the probability of a measurement being either 'normal' or 'tumor'.

Classification of continuous data

The continuous data was preprocessed, by normalizing it using the SNV method. Spectra were not averaged and no outlier detection was performed. Along the needle trajectory, US-images were captured of the biopsy needle while it was positioned in healthy tissue during the first measurement, and when the needle was at the final measurement location, targeting to be in the tumor. For the classification of the continuous data, all point measurements were used to build another three classification models with the different input data (i.e. fit parameter, full spectrum, and selected wavelengths) (Fig. 3). No measurements of the continuously acquired data were used in the development of any of the models, they were only provided to the classification model to be classified. Therefore these needle trajectory measurements were not labeled based on US-imaging or pathology.

Results

In total 32 patients were measured and had unambiguous pathology results. Of these patients, 27 formed the point measurement dataset and 5 the continuous dataset. From the 27 point-based measurement patients, in one patient the biopsy specimen was absent and two patients had biopsy specimens that were clearly damaged during processing of the tissue. In these three cases careful evaluation of the US images by a radiologist revealed that the needle tip was certainly placed a few millimeters inside the tumor and therefore these patients were still included in the analysis. Four patients were excluded from the analysis because (1) the side of the biopsy specimen that had been in contact with the fibers during the measurement consisted of healthy tissue over the extent of a few millimeters and (2) according to the radiologist the needle was moved between the measurement and biopsy. No patients were excluded because the tumor was too close

to the skin, thus prohibiting the acquisition of measurements of healthy tissue.

In the procedure of outlier detection, two measurement locations were detected. An explanation for the first outlier might be that the needle tip was in a pool of blood during the normal measurements, which was confirmed by the high blood content according to the fit parameters. As for the second outlier, the histopathology of this measurement location showed benign tissue in the biopsy specimen. The patients to which these locations belonged were also excluded to ensure a balanced dataset.

Thus, in total 6 patients were excluded from the point measurement dataset. The remaining 21 patients, in whom point measurements were obtained, were included for further analysis. The patient characteristics of both patient datasets seem similar and are summarized in Table 1.

Classification models based on fit parameters

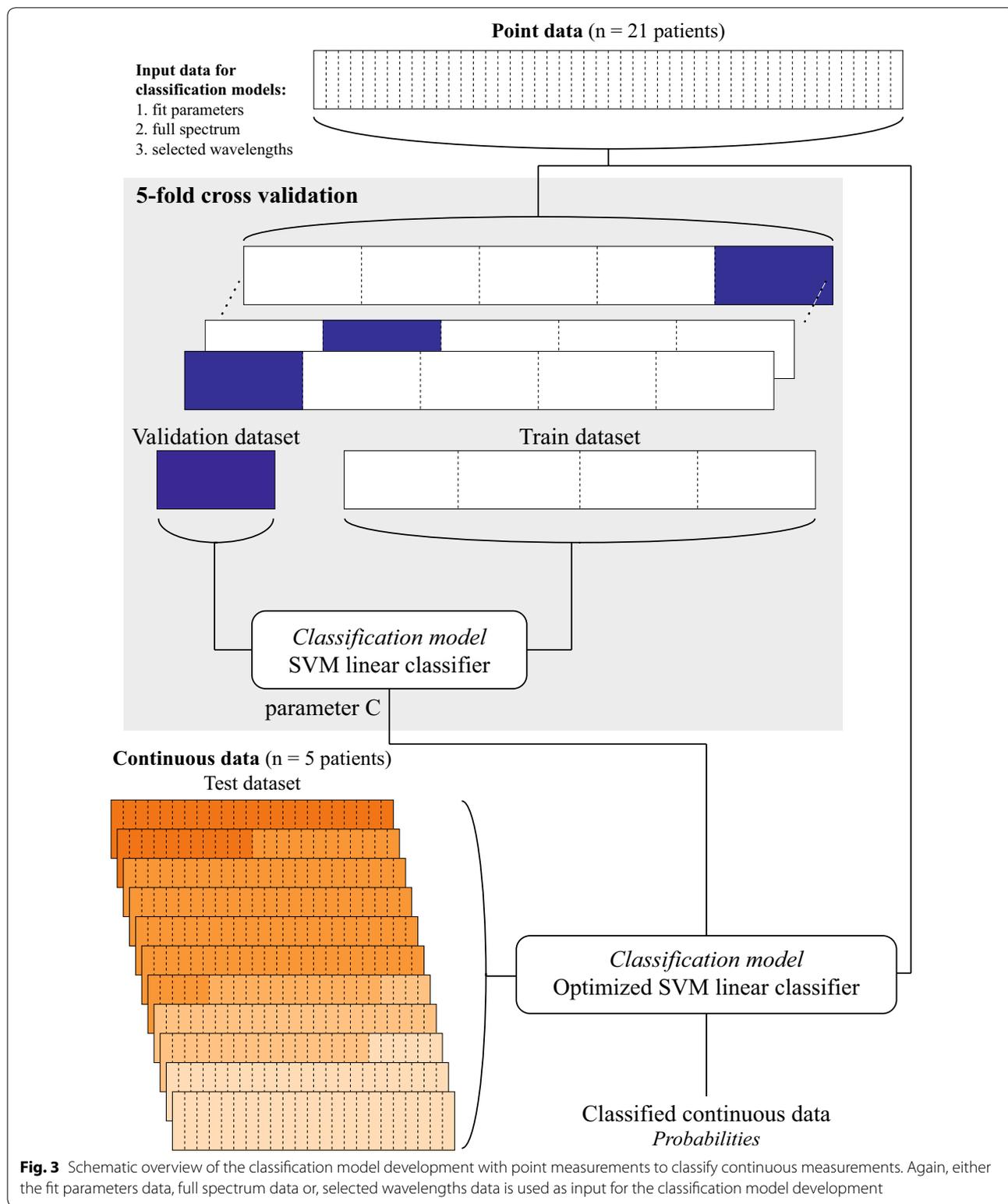
In total 64 classification models that were based on combinations of fit parameters were built. The two fit parameter combinations that generated the two classification models with the highest accuracies are listed in Table 2. The fit parameter combination of F/W-ratio and collagen was the combination that resulted in the classification model with the best performance, with a mean accuracy, sensitivity, specificity and MCC of 0.85 (0.16), 0.72 (0.33), 0.99 (0.03), and 0.74 (0.30), respectively. The second best performing classification model was based on the F/W-ratio alone, which had a slightly lower sensitivity compared with the combination of the F/W-ratio and collagen.

Classification model based on full spectrum

The mean accuracy, sensitivity, specificity, and MCC of the model based on the full spectrum were 0.92 (0.06), 0.94 (0.10), 0.89 (0.11), and 0.84 (0.12), respectively (Table 2). Compared to the fit parameter model, the full spectrum model had a better accuracy, sensitivity, and MCC, whereas the specificity of the fit parameter model was better. This indicates that the full spectrum classification model is useful for detecting all tumor tissue at the cost of classifying some normal tissue as tumor. With the fit parameter classification model, less normal tissue will be incorrectly classified as tumor, but also, less tumor tissue will be detected.

Classification model based on selected wavelengths

A third classification model was developed using a selection of wavelengths that were significantly different between normal and tumor spectra according to the Wilcoxon rank sum test ($\alpha=0.05$). Figure 4 shows



the results of the Wilcoxon rank sum test. The grey parts of the graph represent wavelength areas in which the p-value was lower than 0.05. From these areas the wavelength with the lowest p-value was selected for the

selected wavelengths model (vertical dashed lines). The selected wavelengths were: 501 nm, 916 nm, 973 nm, 1145 nm, 1211 nm, 1371 nm, 1424 nm, and 1597 nm.

Table 1 Patient characteristics of point measurements dataset and continuous measurements dataset

Patient characteristics	Point measurements (n=21)	Continuous measurements (n=5)
Mean age (std)	53.4 year (12.1)	60.6 year (10.2)
Menopausal status		
Premenopausal	8	1
Perimenopausal	2	0
Postmenopausal	10	4
Unknown	1	0
Mean tumor size (US imaging) (std)	25.5 mm (11.1)	37.2 mm (25.1)
Cancer type ^a		
Invasive ductal carcinoma	19	3
Invasive lobular carcinoma	2	1
Mucinous adenocarcinoma	0	1

^a The histopathology was based on the histopathology of all biopsy specimens taken in that patient

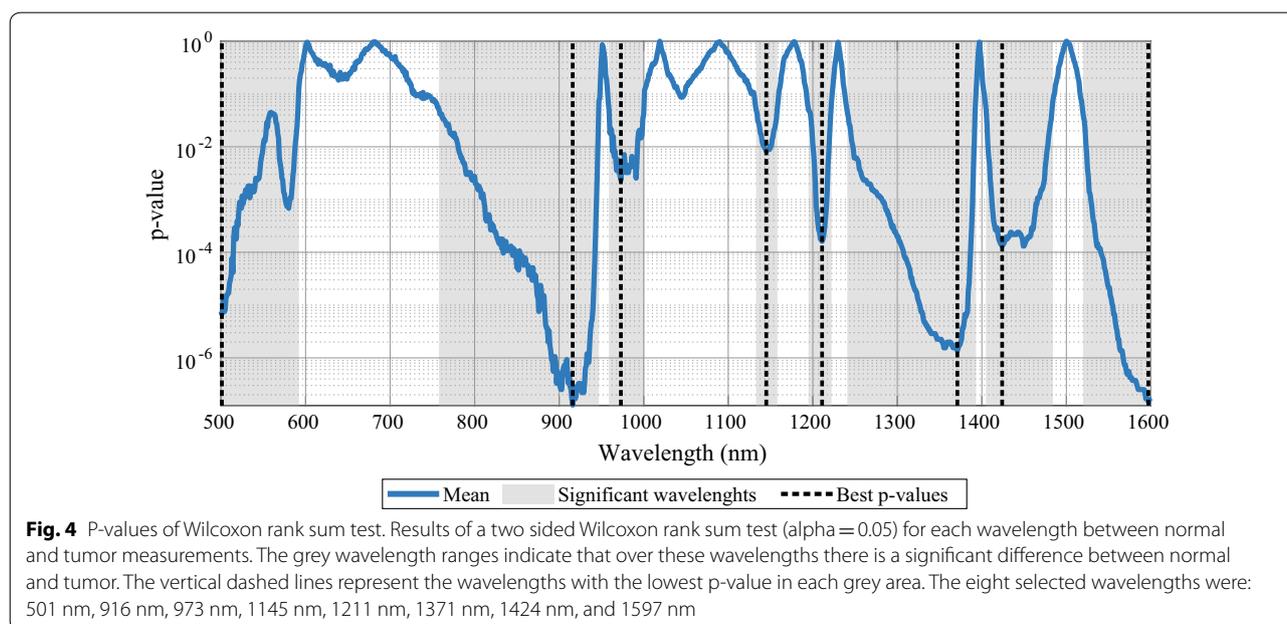
The classification model based on these wavelengths was tested similarly to the models based on the fit parameters and the full spectrum. The mean accuracy, sensitivity, specificity, and MCC of this model was 0.93 (0.06), 0.95 (0.07), 0.91 (0.14), and 0.87 (0.11), respectively (Table 2). Compared to the fit parameter model this model has improved mean sensitivity, but reduced specificity. Despite the decrease in mean specificity, the mean accuracy and MCC of the selected wavelengths model is higher in comparison to the fit parameter model.

The classification model after feature selection also outperforms the full spectrum model as the mean accuracy, sensitivity and specificity are slightly higher. The MCC of the classification model based on a selection of wavelengths was the highest with the lowest standard deviation compared to the other models.

To ensure the improvement of model performance was related to the actual wavelengths in the set of selected

Table 2 Performance (mean accuracy, sensitivity, specificity and MCC with standard deviations) of classification models

Type of data used as input for the model	Mean accuracy (std)	Mean sensitivity (std)	Mean specificity (std)	Mean MCC (std)
Fit parameters				
F/W-ratio and collagen	0.85 (0.16)	0.72 (0.33)	0.99 (0.03)	0.74 (0.30)
F/W-ratio	0.85 (0.16)	0.71 (0.34)	0.99 (0.04)	0.72 (0.31)
Full spectrum	0.92 (0.06)	0.94 (0.10)	0.89 (0.11)	0.84 (0.12)
Selected wavelengths	0.93 (0.06)	0.95 (0.07)	0.91 (0.14)	0.87 (0.11)



wavelengths, the model performance was compared to the model performance of a subset of wavelengths that had the maximum p-value from wavelength ranges with p-values of >0.5 . The selected wavelengths for this model were: 602 nm, 681 nm, 951 nm, 1018 nm, 1095 nm, 1174 nm, 1230 nm, 1397 nm, and 1503 nm. As for the model with these eight selected wavelengths the mean accuracy, mean sensitivity, and mean specificity was 0.49 (0.12), 0.48 (0.25), and 0.50 (0.27), respectively. The MCC also displayed weak performance of this model with a mean value of 0.61 and a standard deviation of 0.32.

Classifying continuous data

The data from the five patients that were measured in continuous mode were tested on the three classification models (fit parameters, full spectrum and selected wavelengths). To make these classification models consistent with the previous models, the fit parameter model was developed with the same fit parameters as in the previous model (F/W-ratio & Collagen) and similarly for the selected wavelengths model the same eight wavelengths (501 nm, 916 nm, 973 nm, 1145 nm, 1211 nm, 1371 nm, 1424 nm, and 1597 nm) were used. The results of the classification of the continuous data are represented in Fig. 5.

In the left part of the figure for each patient the US images at two locations along the needle trajectory ('normal' and 'tumor') are shown. The histopathology of the part of the biopsy specimen that was in touch with the fibers at the last measurement location is displayed in the right side of the figure. The black bars represent a distance of 1 mm in the histopathology image. The graphs in the center of the figure show the output of the classification models in terms of probabilities for each measurement. A probability of >0.5 indicates a measurement is classified as 'normal breast tissue' by the model, whereas a probability of <0.5 implies 'tumor'. The x-axis represents the measurements in time, not in distance. In patient 1, there are some measurements missing because these were accidentally not saved during the procedure.

The histopathologic evaluation by the pathologist revealed that there was tumor ('mucinous adenocarcinoma', or 'invasive ductal carcinoma') in the biopsy specimen of patient 1, 2 and 3. In the case of patient 4 and 5, the side of the biopsy specimens touching the fibers did not contain malignant tissue according to the pathologist.

In all three patients that had invasive carcinoma in their biopsy specimen (patient 1, patient 2, and patient 3) there is a distinct decrease in probability visible in the classified DRS measurements taken along the trajectory from healthy tissue to tumor tissue. Furthermore, the first measurements of patient 1 and patient 2 are classified as normal tissue (probabilities close to one), and the final measurements are classified as tumor (probabilities close to zero) by all three models. In both these patients, the probability of the final measurement of the trajectory calculated by the full spectrum model and the selected wavelengths model are closer to zero than the output of the fit parameter model, indicating more certainty of the classification. Along the trajectory of patient 1, there is one outlier (measurement #10), which displays a distinct decrease in probability for all classification models. The measurements of patient 3 are classified as normal in the beginning of the trajectory and as the needle progressed to the tumor, the probabilities, clearly and consistently over all models, decreased. However, at the end of the trajectory, none of the three models classified the final measurements as tumor, whereas according to the biopsy specimen the needle was placed in tumor tissue.

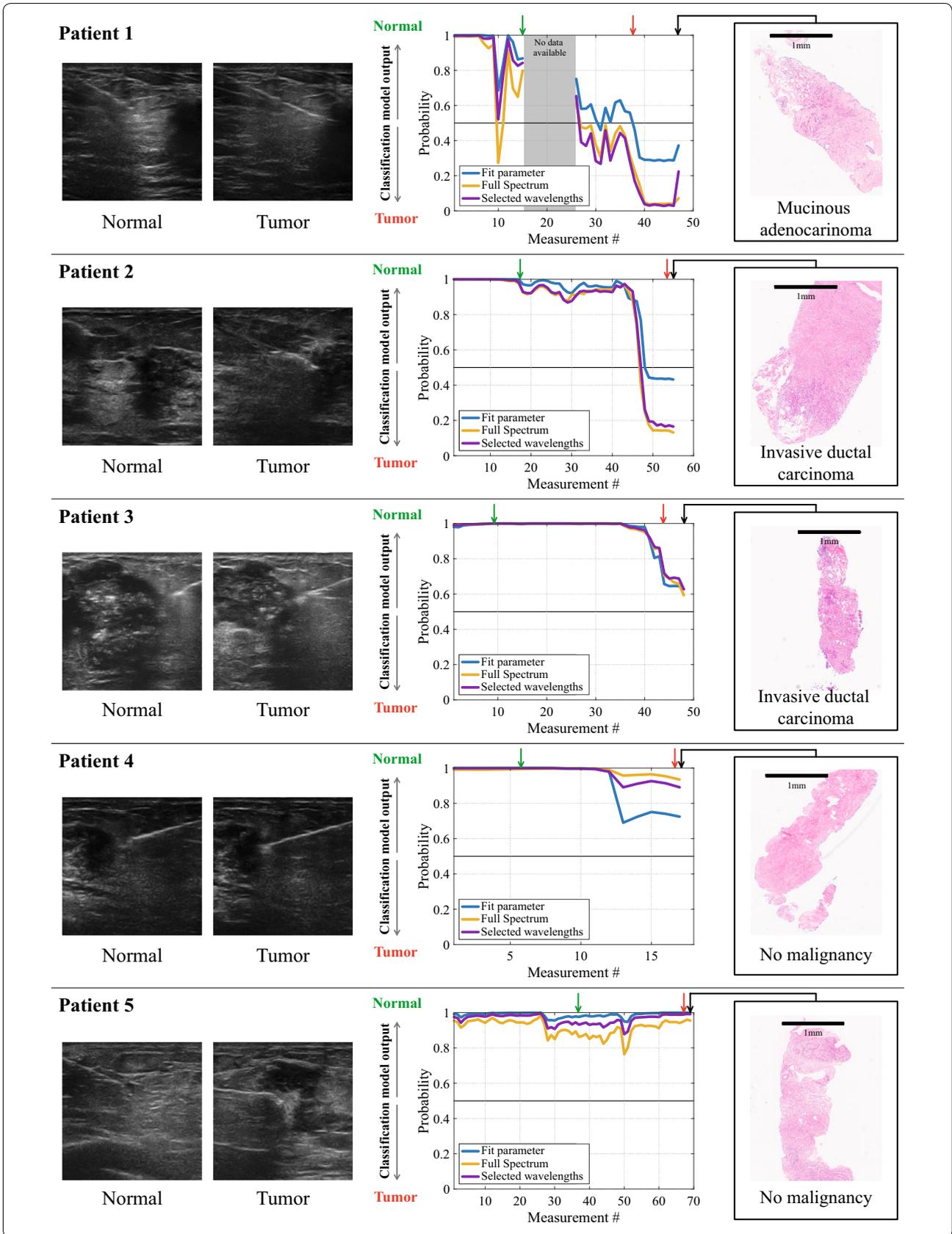
Two patients did not have tumor tissue in the first 2 mm of the biopsy specimen that was in contact with the optical fibers (patient 4 and 5). In both cases, the outcomes of the classification models classified all measurements in the needle trajectory as normal tissue. As for patient 4, the probability of the fit parameter model does show a decrease that was not seen in the output of the other two classification models. The transition from normal tissue to tumor, seen in the patients with a malignancy (patient 1, patient 2 and patient 3), is not consistently present in the classification output of all three models in patient 4 and patient 5, whom had no malignancy in their biopsy specimens.

Discussion

A large amount of evidence from in vivo and ex vivo studies around the world has proven that DRS can be a highly powerful tool for clinical use to discriminate tissue types. However, the technology has not been integrated with a surgical tool for real-time margin assessment. With the goal of moving towards a real-time classification tool for surgical margin assessment during breast surgery, we aimed at developing a classification model to

(See figure on next page.)

Fig. 5 Classification of continuous data. In the left part of the image, US images taken along the needle trajectory at 'normal' and 'tumor'. The middle of the image includes the outcomes of the classification algorithms, where the x-axis is the measurement number (\neq distance) and the y-axis is the probability of a measurement being normal (>0.5) or tumor (<0.5). The green and red arrows indicate the locations where the needle was kept still. The histopathology of the part of the biopsy specimen that was in contact with the needle is displayed in the right side



accurately predict the type of tissue in front of the DRS tool, as well as showing the feasibility of real-time use of this technology. To reach these goals a custom made optical biopsy needle was used enabling DRS measurements and histopathology to be assessed on the same tissue volume which is inevitable for developing a robust classification model. It should be noted that this research was conducted with this set-up as step towards developing a surgical tool that can guide the surgeon, rather than to improve the yield of breast biopsy procedures. This research differs from previous work as it measures DRS over a broader wavelength range extending into the near-infrared wavelengths. Furthermore, to the authors' knowledge, it is the first publication to test the feasibility of in vivo continuous DRS data acquisition, with a frame rate of approximately one measurement per second, which is more similar to how data will be acquired in the surgical setting.

We first used the point-based measurements to determine the performance of the classification models that were based on different input data. We found that if the fit parameter data was used as input, the combination of the F/W-ratio and collagen resulted in a model with the highest accuracy and MCC (0.85 and 0.74) compared to other combinations of fit parameters. Besides the fit parameter model, two other models were developed using the full spectrum of wavelengths or a selection of wavelengths as input. The full spectrum model had a better sensitivity compared to the fit parameter model (0.94 versus 0.72), whereas the fit parameter model had a higher specificity (0.99 versus 0.89), suggesting that the full spectrum model is more suitable for detecting tumor tissue, while the fit parameter model has less misclassifications of normal tissue. Although not statistically tested, the classification model based on a subset of selected wavelengths seems to outperform the other two models with the highest accuracy and MCC (0.93 and 0.87).

We developed three classification models (fit parameters, full spectrum and selected wavelengths) based on all available point-measurements specifically to classify the continuous measurements. Importantly, none of the continuous measurements were used for development of a classification model, they only served as test data to be classified by the classification model (Fig. 3). In all five patients, the first measurements were classified as normal tissue by the classification models, this is expected considering the fact that the needle trajectory starts in normal tissue going towards the suspected tumor tissue. In patients 1, 2, 4, and 5, the classified DRS measurements of the final measurement locations are in agreement with the pathological outcome. In patient 1, however, there is a measurement (#10) in the trajectory that is classified as 'tumor'; this appears to be a 'false positive' since the

distance from this location to the lesion is quite far. In the surgical pathology report following lumpectomy for this patient it was noted that there was a focus of DCIS 1.5 mm from the tumor. It could be possible that this smaller lesion was in the trajectory of the needle, explaining the decrease in probability.

The output by the classification models of the final measurement location were not in accordance with histopathologic evaluation in one patient (patient 3). In this case, the outputs of the classification models show a decrease in probability, but never reach the threshold, and the final measurement location is classified as normal by all classification models. It could be possible that the histopathology evaluation of this patient has been compromised as the removal of the biopsy specimen from the needle cavity was difficult and, since the specimen was fragmented, part of it might have been left behind. Overall, in four out of five continuous mode patients, the classification models were able to discriminate tumor tissue from normal tissue, although the fit parameter model was least convincing with probabilities closer to 0.5. In three out of five patients malignant tissue was present in the biopsy specimen and in these patients a decrease in probability of the classified measurements is also seen along the needle trajectory. This decrease is absent in the other two patients that had healthy tissue in their biopsy specimens. The fact that a decrease can be detected is an important result when considering DRS as a margin assessment tool. In a way, this trajectory can be seen as a line that at some point crosses the optimal resection plane that is perpendicular to this line. Thus, being able to detect the upcoming tumor could provide the surgeon with viable information for guidance.

A limitation of the continuous measurements is that a biopsy was only available from the final measurement location of the presumed tumor area while no histopathology was taken along the needle trajectory. However, all breast tumors were clearly visible on the US images and could confirm that the needle was positioned in normal tissue from the start of the measurements. Nevertheless, some uncertainty will still exist on the precise location of the tissue border where normal tissue ends and tumor tissue starts. It should furthermore be noted that the x-axes of the graphs in Fig. 5 are related to time opposed to distance, and thus these graphs therefore display a change over time. Since the needle was not moved with constant speed along the needle trajectory, it was not possible to display the measurements as a function of the distance.

In literature many different methodologies are used for classifying reflectance spectra of breast tissue, for example logistic regression [20, 22], classification and regression trees [35, 36], artificial neural network [24],

hierarchical cluster analysis [24], k-nearest neighbor [37], linear discriminant analysis [35], and support vector machines [35, 38–41]. In this study, a linear SVM classifier was chosen, as this classifier is relatively insensitive to overfitting [42–46]. Possibly a polynomial kernel SVM would have provided better results, however because the number of patients in the study is limited for machine learning, a linear, less complicated, classifier was chosen. For similar reasons the bootstrap sampling was preferred to leave-one-patient-out cross-validation, even though bootstrap methods can have the tendency to be pessimistic. By extending the number of measurements the accuracy of the classification of the DRS measurements will likely improve, and more sophisticated machine learning algorithms can be used.

The SVM classification model was developed with the input of either fit parameters (fit parameter model), or all wavelengths in the spectrum (full spectrum model), or some selected wavelengths (selected wavelengths model). A previous publication comparing the classification accuracy for discrimination of breast cancer of a SVM model based on physical parameter data (equivalent to fit parameters model), with the accuracy of a SVM classification model based on empirical data (equivalent to the selected wavelengths model) reported similar results to this study [38]. The main advantage of using the fit parameters is that these parameters can provide insight into the physical and structural features that contribute to discrimination [38, 43]. However, if fit parameters cannot be estimated accurately, for example because the tissue has a layered structure, accuracy of classification models based on fit parameters will be lower [43]. This can explain why the accuracy of the fit parameter model was lower compared to the other two models.

We found that the performance of the selected wavelengths model is slightly better compared to the performance of the full spectrum model, which is not surprising since removing redundant wavelengths is often reported to be beneficial for classification performance [42]. There are many ways to select or reduce features, such as partial least squares [38], maximum representation and discrimination feature [20], or principal component analysis (PCA) [24, 38–40]. In this study, a Wilcoxon rank sum test is performed to find wavelengths that are significantly different between normal and tumor tissue. This method has been described for feature selection in previous publications, although in many cases this statistical test was preceded by PCA [40, 46]. The advantage of the Wilcoxon rank sum test is that the selection of wavelengths is based on true spectral differences between tissue types. The disadvantage is that wavelengths that are not discriminated according to this statistical test are excluded in the model development, although they could

have discriminative power in combination with each other.

The wavelengths that were eventually selected are located in wavelength areas that are related to the absorption of light by fat, water, and to a lesser extent, blood. This result is in line with previous publications by others and our own group in which these substances also contributed to discriminating healthy tissue from tumor tissue [32, 38].

The DRS measurements in this study were obtained during breast biopsy procedures to provide a correlated dataset (DRS data and histopathology) and test the feasibility of real-time data acquisition. This setting is obviously different than the surgical setting where the goal is to classify DRS measurements of the resection margin. In that situation, the influence of air exposure will likely affect the visual wavelength range due to differences in oxygenated and de-oxygenated blood which have different optical absorption characteristics; whereas the near-infrared wavelength range, with predominant absorption characteristics from fat to water, will likely be less affected by the surgical setting. Furthermore, the resection margin can also be influenced by cauterization which was absent in the measurements obtained in the biopsy setting, or extravascular blood on the resection surface. With regard to the results in this study, this might imply that in the selected wavelengths model the first wavelength that was selected (501 nm) cannot be used. The accuracy of DRS measurements for the detection of tumor intra-operatively at the resection margin, should be investigated in a study in which DRS measurements (including also the NIR wavelengths) are acquired at the true resection margins, preferably in the surgical workflow.

Using DRS as a clinical margin assessment tool also requires that measurements can be acquired and classified in real-time. In the continuous dataset, each spectrum required 0.35 s to be acquired. If necessary this acquisition time could be decreased by a factor of 4 by increasing the fiber diameter from 200 to 400 μm . As for the classification this was not performed real-time in this study. However, once a classification model is defined the tissue can be classified in real-time as this requires little computational power.

Another important factor to consider is the influence of ambient light that might be different in the surgical resection field compared to the setup during a biopsy. Part of this challenge is overcome by the fact that a fiber is used which has to be in contact with the tissue instead of a non-contact configuration. Therefore, only light that falls in the acceptance angle of the fiber will be recorded by the spectrometer. However, in clinical practice this might mean that very bright light sources that

are in close proximity of the fiber-optic probe have to be dimmed to ensure interference with the DRS measurements is prevented.

Conclusions

In this paper, we demonstrate the feasibility that DRS measurements can be acquired real-time and that a predictive classification model can be built to classify the measurements as normal or tumor tissue. The classification model based on a selection of wavelengths discriminated normal tissue from tumor tissue with the highest accuracy and MCC of 0.93 and 0.87, respectively. This performance may be sufficient for the application of detecting positive resection margins during breast conserving surgery. The needle trajectory measurements show that DRS measurements can be acquired real-time and that these measurements can be classified accurately. Furthermore, the transition from normal tissue to tumor tissue was seen in the continuous DRS measurements.

Our current results indicate that integration of DRS in a surgical tool or knife could be useful for characterizing breast tissue in vivo and aiding surgeons in detecting positive resection margins during surgery. The next step is to investigate the feasibility of real-time DRS acquisition and classification on resection margins and investigate the impact of a DRS guided tool on surgical outcomes.

Abbreviations

DRS: diffuse reflectance spectroscopy; SVM: support vector machine; MCC: Matthews Correlation Coefficient; SNV: standard normal variate; StO₂: oxygen saturation; F/W-ratio: fat/water-ratio; PCA: principal component analysis.

Authors' contributions

LLdB patient inclusion, performed measurements, processed the data, performed data analysis, interpreted the data and wrote manuscript. TMB designed the study, performed measurements, aided in interpretation of the data and worked on manuscript. FvD helped with patient inclusion, provided input for the manuscript and helped in interpretation of clinical significance. MFTDVP helped with patient inclusion, provided input for the manuscript and helped in interpretation of clinical significance. CEL helped in interpretation of clinical significance and assessed the US imaging. GAOWW helped in interpretation of clinical significance. JS performed histopathology assessment. HJCMS contributed to interpretation of the results and supervised the work. BHWH designed the study, developed the optical biopsy needle and fit model, aided in interpretation of the data and supervised the work. TJMR designed the study, contributed to interpretation of the results and clinical significance and supervised the work. All authors read and approved the final manuscript.

Author details

¹ Department of Surgery, the Netherlands Cancer Institute—Antoni van Leeuwenhoek, Plesmanlaan 121, Postbus 90203, 1066 CX Amsterdam, The Netherlands. ² In-body Systems, Philips Research, High Tech, Campus 34, 5656 AE Eindhoven, The Netherlands. ³ Department of Radiology, the Netherlands Cancer Institute—Antoni van Leeuwenhoek, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. ⁴ Department of Pathology, the Netherlands Cancer Institute—Antoni van Leeuwenhoek, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. ⁵ Biomedical Engineering and Physics, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. ⁶ Biomechanical Engineering, Delft University of Technology, Mekelweg 5, 2628 CD Delft, The Netherlands. ⁷ Technical Medical Centre, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands.

Acknowledgements

We would like to thank Jarich Spliethoff and Niels Langhout for their input and help during the measurement procedures. We would also like to acknowledge Christian Reich, Manfred Mueller, Vishnu Pully, Walter Bierhoff, Axel Winkel, and Arnold van Keersop for their contribution in the development of the biopsy needle and fit model software. Assistance provided by the other radiologists and radiology department personnel was greatly appreciated, as well as the assistance of the surgeons and nurse practitioners with patient inclusion. In particular we would like to express many special thanks to all involved patients.

Competing interests

TMB and BHWH are affiliated with Philips Research; they are employees of Philips Research, however, they have no financial interest in the subject matter, materials and/or equipment. None of the other authors have a financial relationship with Philips, or have other competing interests.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The study was approved by the hospital's institutional review board and informed consent was obtained from each patient.

Funding

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 9 October 2018 Accepted: 13 December 2018

Published online: 19 December 2018

References

- Houssami N, Macaskill P, Marinovich ML, Dixon JM, Irwig L, Brennan ME, Solin LJ. Meta-analysis of the impact of surgical margins on local recurrence in women with early-stage invasive breast cancer treated with breast-conserving therapy. *Eur J Cancer*. 2010;46:3219–32.
- Smitt MC, Nowels K, Carlson RW, Jeffrey SS. Predictors of reexcision findings and recurrence after breast conservation. *Int J Radiat Oncol Biol Phys*. 2003;57:979–85.
- Collette S, Collette L, Budiharto T, Horiot J-C, Poortmans PM, Struikmans H, Van den Bogaert W, Fourquet A, Jager JJ, Hoogenraad W, et al. Predictors of the risk of fibrosis at 10 years after breast conserving therapy for early breast cancer—a study based on the EORTC trial 22881–10882 'boost versus no boost'. *Eur J Cancer*. 2008;44:2587–99.
- Bodilsen A, Bjerre K, Offersen BV, Vahl P, Amby N, Dixon JM, Ejlersten B, Overgaard J, Christiansen P. Importance of margin width in breast-conserving treatment of early breast cancer. *J Surg Oncol*. 2016;113:609–15.
- Olsen MA, Nickel KB, Margenthaler JA, Wallace AE, Mines D, Miller JP, Fraser VJ, Warren DK. Increased risk of surgical site infection among breast-conserving surgery re-excisions. *Ann Surg Oncol*. 2015;22:2003–9.
- Volders JH, Negenborn VL, Haloua MH, Krekel NMA, Józwiak K, Meijer S, Van den Tol PM. Cosmetic outcome and quality of life are inextricably linked in breast-conserving therapy. *J Surg Oncol*. 2016;115:941–8.
- Hau E, Browne L, Capp A, Delaney GP, Fox C, Kearsley JH, Millar E, Nasser EH, Papadatos G, Graham PH. The impact of breast cosmetic and functional outcomes on quality of life: long-term results from the St. George and Wollongong randomized breast boost trial. *Breast Cancer Res Treat*. 2013;139:115–23.

8. Waljee JF, Hu ES, Ubel PA, Smith DM, Newman LA, Alderman AK. Effect of esthetic outcome after breast-conserving surgery on psychosocial functioning and quality of life. *J Clin Oncol*. 2008;26:3331–7.
9. MacNeill F, Karakatsanis A. Over surgery in breast cancer. *The Breast*. 2017;31:284–9.
10. Angarita FA, Nadler A, Zerhouni S, Escallon J. Perioperative measures to optimize margin clearance in breast conserving surgery. *Surg Oncol*. 2014;23:81–91.
11. Butler-Henderson K, Lee AH, Price RI, Waring K. Intraoperative assessment of margins in breast conserving therapy: a systematic review. *Breast*. 2014;23:112–9.
12. Keating JJ, Fisher C, Batiste R, Singhal S. Advances in intraoperative margin assessment for breast cancer. *Curr Surg Rep*. 2016;4:15.
13. O’Kelly Priddy CM, Forte VA, Lang JE. The importance of surgical margins in breast cancer. *J Surg Oncol*. 2016;113:256–63.
14. St John ER, Al-Khudairi R, Ashrafi H, Athanasios T, Takats Z, Hadjiminas DJ, Darzi A, Leff DR. Diagnostic accuracy of intraoperative techniques for margin assessment in breast cancer surgery. *Ann Surg*. 2017;265:300–10.
15. Heil J, Breikreuz K, Golatta M, Czink E, Dahlkamp J, Rom J, Schuetz F, Blumenstein M, Rauch G, Sohn C. Do reexcisions impair aesthetic outcome in breast conservation surgery? exploratory analysis of a prospective cohort study. *Ann Surg Oncol*. 2012;19:541–7.
16. Zysk AM, Chen K, Gabrielson E, Tafta L, May Gonzalez EA, Canner JK, Schneider EB, Cittadini AJ, Carney PS, Boppart SA, et al. Intraoperative assessment of final margins with a handheld optical imaging probe during breast-conserving surgery may reduce the reoperation rate: results of a multicenter study. *Surg Oncol*. 2015;22:3356–62.
17. Bolger JC, Solon JG, Khan SA, Hill ADK, Power CP. A comparison of intraoperative margin management techniques in breast-conserving surgery: a standardised approach reduces the likelihood of residual disease without increasing operative time. *Breast Cancer*. 2015;22:262–8.
18. Cabioglu N, Hunt KK, Sahin AA, Kuerer HM, Babiera GV, Singletary SE, Whitman GJ, Ross MI, Ames FC, Feig BW, et al. Role for intraoperative margin assessment in patients undergoing breast-conserving surgery. *Ann Surg Oncol*. 2007;14:1458–71.
19. Nichols BS, Schindler CE, Brown JQ, Wilke LG, Mulvey CS, Krieger MS, Gallagher J, Geradts J, Greenup RA, Von Windheim JA, Ramanujam N. A quantitative diffuse reflectance imaging (QDRI) system for comprehensive surveillance of the morphological landscape in breast tumor margins. *PLoS ONE*. 2015;10:e0127525.
20. Keller MD, Majumder SK, Kelley MC, Meszoely IM, Boulos FI, Olivares GM, Mahadevan-Jansen A. Autofluorescence and diffuse reflectance spectroscopy and spectral imaging for breast surgical margin analysis. *Lasers Surg Med*. 2010;42:15–23.
21. De Boer LL, Molenkamp BG, Bydlon TM, Hendriks BHW, Wesseling J, Sterenborg HJCM, Ruers TJM. Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries. *Breast Cancer Res Treat*. 2015;152:509–18.
22. Volynskaya Z, Haka AS, Bechtel KL, Fitzmaurice M, Shenk R, Wang N, Nazemi J, Dasari RR, Feld MS. Diagnosing breast cancer using diffuse reflectance spectroscopy and intrinsic fluorescence spectroscopy. *J Biomed Opt*. 2008;13:024012.
23. Van Veen RLP, Sterenborg HJCM, Marinelli AWKS, Menke-Pluymers M. Intra-operatively assessed optical properties of malignant and healthy breast tissue, to determine the optimum wavelength of contrast for optical mammography. *J Biomed Opt*. 2004;6:1129–36.
24. Bigio IJ, Bown SG, Briggs G, Kelley C, Lakhani S, Pickard D, Ripley PM, Rose IG, Saunders C. Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results. *J Biomed Opt*. 2000;5:221–8.
25. Brown JQ, Wilke LG, Geradts J, Kennedy SA, Palmer GM, Ramanujam N. Quantitative optical spectroscopy: a robust tool for direct measurement of breast cancer vascular oxygenation and total hemoglobin content in vivo. *Cancer Res*. 2009;69:2919–26.
26. Kennedy S, Caldwell M, Bydlon T, Mulvey C, Mueller J, Wilke L, Barry W, Ramanujam N, Geradts J. Correlation of breast tissue histology and optical signatures to improve margin assessment techniques. *J Biomed Opt*. 2016;21:066014.
27. Spliethoff JW, Prevoo W, Meier MA, de Jong J, Evers DJ, Sterenborg HJCM, Lucassen GW, Hendriks BHW, Ruers TJM. Real-time in vivo tissue characterization with diffuse reflectance spectroscopy during transthoracic lung biopsy: a clinical feasibility study. *Clin Cancer Res*. 2016;22:357–65.
28. Nachabé R, Hendriks BHW, Van der Voort M, Dejardins AE, Sterenborg HJCM. Estimation of biological chromophores using diffuse optical spectroscopy: benefit of extending the UV–VIS wavelength range to include 1000 to 1600 nm. *Opt Express*. 2010;18:1432–42.
29. Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl Spectrosc*. 1989;43:772–7.
30. Burger JE. Hyperspectral NIR image analysis. Uppsala: Swedish University of Agricultural Sciences, Unit of Biomass Technology and Chemistry; 2006.
31. Nachabé R, Hendriks BHW, Desjardins AE, van der Voort M, van der Mark MB, Sterenborg HJCM. Estimation of lipid and water concentrations in scattering media with diffuse optical spectroscopy from 900 to 1600 nm. *J Biomed Opt*. 2010;15:037015.
32. De Boer LL, Hendriks BHW, Van Duijnhoven F, Vranken Peeters-Baas M-JTFD, Van de Vijver K, Loo CE, Józwiak K, Sterenborg HJCM, Ruers TJM. Using DRG during breast conserving surgery: identifying robust optical parameters and influence of inter-patient variation. *Biomed Opt Express*. 2016;7:5188–200.
33. Vapnik V. The nature of statistical learning theory. Berlin: Springer; 2013.
34. Osuna EE, Freund R, Girosi F. Support vector machines: training and applications. Cambridge: Massachusetts Institute of Technology; 1997.
35. Nachabé R, Evers DJ, Hendriks BHW, Lucassen GW, Van der Voort M, Rutgers EJ, Vranken Peeters MJ, Van der Hage JA, Oldenburg HS, Wesseling J, Ruers TJM. Diagnosis of breast cancer using diffuse optical spectroscopy from 500 to 1600 nm: comparison of classification methods. *J Biomed Opt*. 2011;16:087010.
36. Ramanujam N, Brown JQ, Bydlon TM, Kennedy SA, Richards LM, Junker MK, Gallagher J, Barry WT, Wilke LG, Geradts J. Quantitative spectral reflectance imaging device for intraoperative breast tumor margin assessment. In: Engineering in Medicine and Biology Society. Piscataway: IEEE; 2009. pp. 6554–6556.
37. Laughney AM, Krishnaswamy V, Beatriz Garcia-Allende P, Conde OM, Wells WA, Paulsen KD, Pogue BW. Automated classification of breast pathology using local measures of broadband reflectance. *J Biomed Opt*. 2010;15:066019.
38. Zhu C, Breslin TM, Harter J, Ramanujam N. Model based and empirical spectral analysis for the diagnosis of breast cancer. *Opt Express*. 2008;16:14961–78.
39. Soares JS, Barman I, Dingari NC, Volynskaya Z, Liu W, Klein N, Plecha D, Dasari RR, Fitzmaurice M. Diagnostic power of diffuse reflectance spectroscopy for targeted detection of breast lesions with microcalcifications. *Proc Natl Acad Sci USA*. 2013;110:471–6.
40. Breslin TM, Xu F, Palmer GM, Zhu C, Gilchrist KW, Ramanujam N. Autofluorescence and diffuse reflectance properties of malignant and benign breast tissues. *Ann Surg Oncol*. 2004;11:65–70.
41. Palmer GM, Zhu C, Breslin TM, Xu F, Gilchrist KW, Ramanujam N. Monte Carlo-based inverse model for calculating tissue optical properties. Part II: application to breast cancer diagnosis. *Appl Opt*. 2006;45:1072–8.
42. Majumder SK, Ghosh N, Gupta PK. Support vector machine for optical diagnosis of cancer. *J Biomed Opt*. 2005;10:024034.
43. Hendriks BHW, Balthasar AJR, Lucassen GW, Van der Voort M, Mueller M, Pully VV, Bydlon TM, Reich C, Van Keersop ATMH, Kortsmid J, et al. Nerve detection with optical spectroscopy for regional anesthesia procedures. *J Transl Med*. 2015;13:380.
44. Widjaja E, Zheng W, Huang Z. Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines. *Int J Oncol*. 2008;32:653–62.
45. Skala MC, Palmer GM, Vrotsos KM, Gendron-Fitzpatrick A, Ramanujam N. Comparison of a physical model and principal component analysis for the diagnosis of epithelial neoplasias in vivo using diffuse reflectance spectroscopy. *Opt Express*. 2007;15:7863–75.
46. Valdés PA, Kim A, Leblond F, Conde OM, Harris BT, Paulsen KD, Wilson BC, Roberts DW. Combined fluorescence and reflectance spectroscopy for in vivo quantification of cancer biomarkers in low- and high-grade glioma surgery. *J Biomed Opt*. 2011;16:116007.