




Diffuse reflectance spectroscopy for accurate margin assessment in breast-conserving surgeries: importance of an optimal number of fibers

DINUSHA VELUPONNAR,^{1,2,*} BEHDAD DASHTBOZORG,¹  LYNN-JADE S. JONG,^{1,2}  FREIJA GELDOLF,^{1,2} MARCOS DA SILVA GUIMARAES,³ MARIE-JEANNE T. F. D. VRANCKEN PEETERS,¹ FREDERIEKE VAN DUIJNHOFEN,¹ HENRICUS J. C. M. STERENBORG,^{1,4} THEO J. M. RUERS,^{1,2} AND LISANNE L. DE BOER¹

¹Department of Surgery, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

²Department of Nanobiophysics, Faculty of Science and Technology, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands

³Department of Pathology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

⁴Department of Biomedical Engineering and Physics, Amsterdam University Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, Netherlands

*d.veluponnar@nki.nl

Abstract: During breast-conserving surgeries, it remains challenging to accomplish adequate surgical margins. We investigated different numbers of fibers for fiber-optic diffuse reflectance spectroscopy to differentiate tumorous breast tissue from healthy tissue *ex vivo* up to 2 mm from the margin. Using a machine-learning classification model, the optimal performance was obtained using at least three emitting fibers (Matthew's correlation coefficient (MCC) of 0.73), which was significantly higher compared to the performance of using a single-emitting fiber (MCC of 0.48). The percentage of correctly classified tumor locations varied from 75% to 100% depending on the tumor percentage, the tumor-margin distance and the number of fibers.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

The aim of breast-conserving surgery (BCS) is to completely resect the tumor with a small margin of healthy breast tissue. In this way, a surgeon balances the endpoint of removing all tumor tissue, and the endpoint of achieving an optimal cosmetic result. To evaluate whether the tumor is completely removed or not, the surgical margin status of the specimen is determined after surgery. In the most commonly used 'radial', pathological margin assessment technique, the margins are evaluated by inking the surface of the specimen and examining the distance between the ink on the edge and the boundary of the tumor in the surgical specimen [1,2]. Margin assessment is important since the presence of positive resection margins is one of the key determinants associated with ipsilateral breast tumor recurrence and hampers long-term survival [3–9].

Although the surgical margin status is extremely relevant to the outcome of BCS patients, there is a lack of consensus worldwide on a standardized definition of a 'positive' margin for both invasive carcinoma (IC) and ductal carcinoma in situ (DCIS) [10–14], it varies substantially across countries. For instance, in the USA a re-excision for IC is recommended as soon as the ink on the margin is in contact with tumor cells during radial margin assessment, in accordance with the SSO-ASTRO guidelines [15]. In the Netherlands, a re-excision is recommended when the tumor cells reach the inked margin over a trajectory of more than 4 mm [16]. For DCIS,

guidelines on an adequate margin vary from no DCIS cells on the resection surface to no DCIS cells within 2 mm from the resection edge [10]. The range in definition contributes to the large difference in reported positive margins worldwide, which varies from 9% to 36% for invasive breast cancer and from 4% to 23% for DCIS [10].

In case of a positive surgical margin, a patient would need additional treatment in the form of radiotherapy or a re-excision (i.e. repeat breast conserving surgery or mastectomy), with a potential risk for increased morbidity [3,17], unsatisfactory cosmetic outcome [18–20], decreased quality of life [21–23] and increased health care costs [24,25]. Therefore, the complete removal of breast tumors during primary surgery is essential.

On the other hand, extensive resections with negative margins may result in worse cosmetic outcomes. Studies have reported unsatisfactory cosmetic results in up to 40% of patients undergoing BCS [26,27]. According to multiple studies, specimen volume in relation to breast volume is a statistically significant determinant of a poor cosmetic outcome [28–31]. Various studies have determined that in many cases of BCS, the resected volume varies from 1.5 to 5.0 times the optimum resection volume, which is defined as the tumor volume plus an arbitrarily chosen margin of healthy breast tissue [32–34]. A poor cosmetic outcome leads to an increased risk of depression, anxiety, self-esteem issues, and a decreased quality of life [22,23,35]. Thus, it is important for breast cancer surgeons to excise as little healthy breast tissue as possible.

To balance the goals of complete tumor removal and a satisfactory cosmetic outcome, surgeons mainly rely on visual and tactile feedback. Discriminating healthy tissue from tumor tissue can be extremely challenging based on these types of feedback. Therefore, an accurate method for real-time intraoperative breast cancer margin assessment is needed. Many imaging techniques, such as ultrasound [36,37], fluorescence imaging [38], Raman spectroscopy [39], optical coherence tomography [40,41], radiofrequency spectroscopy [42] and photoacoustic tomography [43] are being investigated as a margin assessment tool. However, these techniques have not been incorporated into surgical practice due to various reasons including diagnostic inaccuracy, lack of speed, complicated user experience, high operator dependence, high costs, and/or inability to perform over the entire margin [44–46].

Fiber-optic diffuse reflectance spectroscopy (DRS) is a non-invasive, optical technique that could be used to study the structural and biochemical composition of tissue, based on the interaction of the tissue with different wavelengths of light. Light from a broadband light source is sent into the tissue through an emitting fiber, where it undergoes several interactions such as scattering and absorption, and part of the light will be reflected back. This reflected light is collected by a receiving fiber. The distance between an emitting and a receiving fiber is approximately equivalent to the measurement depth [47]. DRS spectra contain information concerning the absorption and scattering properties of the illuminated tissue. This could be applied to distinguish different tissue types and thus potentially delineate cancerous tissue during surgery. A previous study of our research group has proven that the absorption of fat compared to water in the near-infrared (NIR) wavelength band has a sensitivity and specificity of 100% in discriminating pure tumorous breast tissue from pure healthy tissue in sliced *ex vivo* breast cancer specimens [48].

Despite the promising results, a crucial issue to achieve the envisioned application of DRS has not been resolved yet. It involves the lack of data concerning the diagnostic accuracy of DRS on the actual resection margins of breast tissue. Measuring the resection margin is more challenging compared to measuring on the sliced specimen as used in previous studies. The DRS measurements on breast lumpectomy specimens could be acquired from locations with inhomogeneous composition. It has been shown by de Boer *et al.* that the accuracy of DRS classification models tested on sliced specimen locations with a mixture of tissue types, is dependent on the percentage of tumor cells in these locations [49]. It was shown that a lower percentage of tumor cells leads to a worse tissue classification performance with a high chance

of missing the tumor presence [49]. This is important, since single-emitting fiber DRS is a point-based measurement method covering a small tissue volume. Therefore, DRS using a single emitting fiber might lead to missing tumorous tissue in a particular region. Furthermore, it would necessitate many measurements to cover a larger area, which could be time-consuming.

In order to overcome this obstacle, we developed a multi-fiber-optic probe that enabled collecting more optical information from one measurement location as compared to the earlier used probe. In this way, multiple spectra could be collected from different tissue volumes at the same measurement locations, which could be used for tissue classification. To the best of our knowledge, this is the first study investigating whether the number of optical fibers has an influence on the accuracy of DRS for detecting different tumor volumes at various depths when measuring on the surgical margin. Although researchers have performed DRS measurements on breast tissue using probes with multiple fibers [50–55], it has never been investigated how the number of fibers affects the accuracy of margin assessment. This study will provide new insights into this relationship, which could enable sampling information from a larger tissue area using the same probe tip size. A larger probe tip area provides the advantage of covering a larger tissue area, while giving the disadvantage of an uncertainty regarding the exact location where a particular DRS signal was measured. A larger number of fibers offers the advantage of sampling more information from the same probe tip area, while giving the disadvantage of a larger data acquisition time. We have tried to balance the size of the probe tip to the number of fibers, in which we tried to sample as much information from a tissue area as possible while maintaining an acceptable data acquisition time. Therefore, this work will contribute to the understanding of how the classification performance of DRS combined with a machine learning models depends on the numbers of fibers at a certain probe tip area.

In this study, we have taken the next step toward the use of optical spectroscopy for intraoperative margin assessment by improving the reliability and resolving the earlier-mentioned uncertainties. The first aim of this study was to investigate the optimum number of optical fibers to distinguish tumorous breast tissue from healthy breast tissue. The second goal was to investigate how the classification accuracy of this optimum number of fibers is affected by various tumor percentages and various distances of the tumor to the margin. In order to meet these goals, we have conducted DRS measurements with our custom-made probe on *ex vivo* lumpectomy specimens. Subsequently, we have trained several machine learning algorithms on data sets based on a different number of fibers, where the labels are assigned based on different ratios of tumor and healthy tissue. Furthermore, we analyzed the classification performance of each model on a test set, which was not used for training the models. After obtaining the optimal number of fibers, we evaluated the classification performance of detecting tumor volumes in various percentages. Lastly, the best-performing models were further evaluated by analyzing the correctly classified tumor locations at different distances from the tumor to the margin.

2. Materials and methods

2.1. Diffuse reflectance spectroscopy setup

The experimental DRS device consisted of 5 identical light sources, a fiber-optic DRS probe, two spectrometers, and in-house developed MATLAB software to collect and save data. The light sources were halogen broadband light sources (Avantes, AvaLight-HAL, 360 – 2500 nm) with integrated shutters. A custom-designed DRS probe was used, as described below. The setup contained one spectrometer covering a visible wavelength range of 200 to 1160 nm (Avantes, AVASPEC-HS2048XL-EVO) and one covering a near-infrared wavelength range of 900 to 1750 nm (Avantes, AVASPECNIR256-1.7-RS). All spectra were calibrated using a similar method as described in [49,56,57].

For this study, we developed a multi-fiber-optic probe with 5 identical emitting fibers positioned in a circle around one central receiving fiber as illustrated in Fig. 1. All emitting fibers had an equal source-detector fiber distance of 2.0 mm.

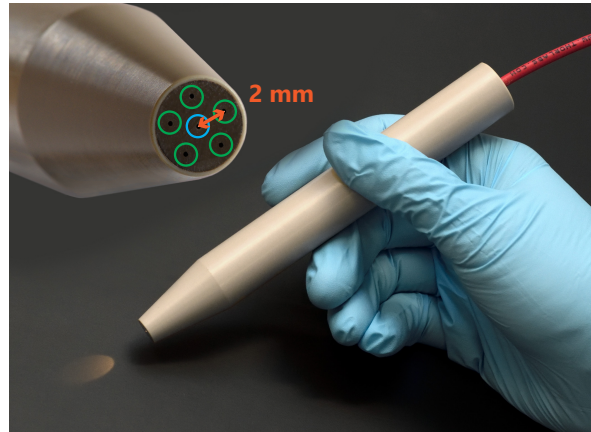


Fig. 1. Handheld DRS probe. In the upper left corner, the distal probe tip is displayed, with a circular configuration of 5 emitting fibers around 1 central receiving fiber, and a distance of 2 mm between each emitting fiber and the receiving fiber.

2.2. Study design

This *ex vivo* study was conducted from 2019 to 2022 at the Netherlands Cancer Institute-Antoni van Leeuwenhoek hospital (NKI-AvL) after approval of the study protocol by the Institutional Review Board (IRBm20-077). During this period, patients with invasive breast carcinoma and/or ductal carcinoma in situ who were scheduled for surgery, were included. According to the medical research involving human subjects act, no written consent was required. In total, 100 breast lumpectomy specimens were obtained from 100 female patients who had undergone BCS.

2.3. Data acquisition

Immediately after surgical excision, we collected the specimen from the surgical team and proceeded with DRS measurements after an estimated time gap of a few minutes. We have aimed to maintain this time gap as consistent as possible throughout all measurements. After specimen collection, approximately 3 to 5 locations on the margin of each lumpectomy specimen were selected as measurement locations. At each location, 3 consecutive DRS point measurements were performed with an acquisition time of a few seconds per measurement. Throughout all measurements, the probe was lightly pressured against the tissue surface area and it was ensured that the entire surface of the probe tip was in contact with the tissue at all times. Due to limitations in the number of measurement locations on each specimen and in order to avoid significant imbalance in our data set, we have used ultrasound imaging on some specimens to localize areas with the smallest distance from the tumor to the resection margin.

After acquiring DRS, each measurement location on the tissue was marked with black pathology ink, approximately equaling the size of the probe tip. Afterward, the specimen was brought to the pathology department, where the resection margins were inked, the whole specimen was frozen, and sliced in a bread-loafed manner. During this process, it was made sure the specimen was sliced at the ink marks. Then the specimen was processed in a standardized manner. An overview of the data acquisition process is shown in Fig. 2

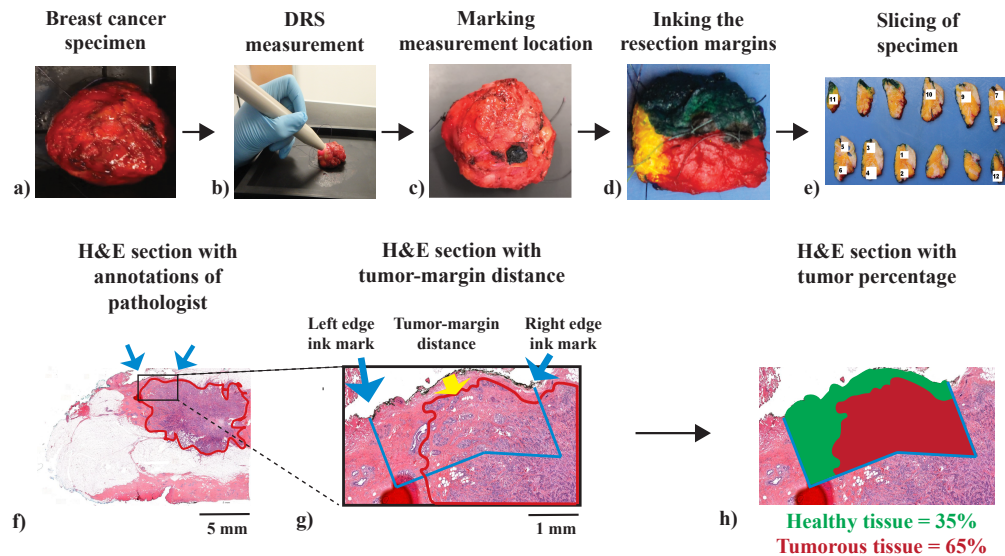


Fig. 2. Overview of the data acquisition method, with a) specimen collection after surgery, b) point-based DRS measurement, c) marking the measurement location with black pathology ink, and d) standard processing by the pathology department, including coloring and slicing the specimen. Schematic overview of the method for determining the tumor percentage and the tumor-margin distance of each measured tissue location in the corresponding histopathology section. In f) the original H&E section with the annotated borders of the lesion in red, and g) the magnified image of the measured tissue location, recognizable due to the black ink along the margin. The yellow arrow (g) indicates the tumor-margin distance, determined by measuring the perpendicular distance from the surgical surface to the tumor in the middle of this region. Lastly, the percentage of tumorous tissue and healthy breast tissue was determined over a depth of 2 mm at this particular marked region, indicated by the blue box (h).

The pathology H&E sections of all measurement locations were digitized and examined by a general pathologist with a high level of experience in breast pathology, who precisely annotated all tumor areas of IC and DCIS in the images. The healthy tissue areas, which consisted of connective and fat tissue, were identified by thresholding the green channel of the H&E sections as described in [58]. The percentage of IC and DCIS based on the annotations of the pathologist and the percentage of healthy tissue based on the thresholding was determined up to 2 mm underneath the black ink marks. These percentages formed the labels (ground truth) of the data set. Hereafter, the sum of IC and DCIS percentage scores will be referred to as 'tumor percentage'. It is important to emphasize that the tumor percentages indicate the area of tumor tissue in a box of 2 mm wide from the margin and do not entail actual percentages of tumor cells. The area percentages of all tissue types were exactly determined using image processing tools in MATLAB. Furthermore, the perpendicular distance of the tumor to the margin at the middle of the black ink mark was determined. Hereafter, this distance will be referred to as the 'tumor-margin distance'. Figure 2 gives an overview of the method for determining the tumor percentage and the tumor-margin distance of each measured tissue location in the corresponding histopathology section.

The data analysis consisted of several steps to build tissue classification models and evaluate their accuracy. In the following subsections, each step is explained in more detail. All data analyses were performed using MATLAB (2022a, MathWorks Inc., Natick, Massachusetts, United States).

2.4. Data preprocessing

2.4.1. Spectrum normalization

After data acquisition, the three consecutive optical spectra of each measurement location were averaged and the spectra from the visual and near-infrared wavelength ranges were stitched together to form one continuous spectrum. Furthermore, data up to 400 nm and after 1600 nm were eliminated as these parts of the spectrum have a low signal-to-noise ratio. To correct for intensity differences, all spectra were normalized using multiplicative scatter correction (MSC) [59], where the spectra are corrected in such a manner that they are as close as possible to the mean of the data set.

2.4.2. Feature extraction

Each normalized spectrum consists of the reflection intensities at 1200 different wavelengths. During previous research of our group, it was found that the visual wavelength range is principally influenced by absorption through blood, while the near-infrared wavelength range is principally influenced by absorption through fat and water [48]. Furthermore, it was demonstrated that the fat fraction in combination with the total volume of fat and water provided optimal discrimination between tumorous breast tissue and healthy breast tissue, when measuring on slices of BCS specimens or biopsies [48,49,60]. All reflection intensities represent 1200 features, that could be used for training the model. In order to avoid overfitting, we have applied feature extraction. In this process, we quantified and extracted a set of spectral features in the near-infrared wavelength range, as described by de Boer *et al.* [49]. These features included the slopes of spectra between designated wavelengths, the maximum difference between the slope and the spectrum, the corresponding wavelength at the point of maximum difference, and the inflection points left and right of the point of maximum difference [49]. This yielded a total of 80 features per fiber.

2.5. Data set preparation

2.5.1. Random fiber selection and mixture

The next step was to build 5 different data sets using features from a different number of fibers (1-5) with a random combination. The first data set contained all features from 1 randomly selected fiber for each measurement location, the second data set contained a combination of all features of 2 randomly selected fibers for each measurement location, and so on. In our study, we tried to sample as much information from the probe tip area as possible, using different numbers of fibers. Therefore, when using 2, 3 or 4 fibers, we always selected random fibers in opposing directions, instead of adjacent fibers. In this way, the fibers would cover the largest tissue area possible and we would always maintain the same illumination-collection geometry for a particular number of fibers. Using a combination of all features of all fibers would result in 400 features per measurement location. The 5 different data sets were used to investigate the impact of the number of fibers in assessing surgical margins.

2.5.2. Labeling

Thereafter, multiple data sets were built from the 5 earlier built data sets, by labeling each measurement location as 'healthy tissue' or 'tumorous tissue' according to different definitions of the labels. The definition was based on various cut-off (threshold) points for the tumor percentage and a constant cut-off point of 2.0 mm for the distance to the margin, for each data set. The threshold points for tumor percentage (TTP) were incrementally increased 5%, up to the maximum TTP of 40% which still enabled training a model with sufficient tumor-labeled data. Any measurement location was labeled as tumorous tissue if it had 1) a tumor percentage equal to or above the chosen TTP and 2) a distance to the tumor less than 2.0 mm. Otherwise, the measurement location was labeled as healthy tissue. Thus eventually, all data sets contained the

same number of measurement locations, but with a different distribution of tumor labels and healthy labels.

2.5.3. Feature selection

The following step was the implementation of feature reduction using a minimum redundancy maximum relevance (MRMR) feature selection algorithm [61], in order to prevent the algorithm from overfitting. This iterative algorithm selects features during each iteration that correlate the strongest with the response variable while being minimally redundant compared to the set of already selected features. Using the outcome of this algorithm, the optimum feature subsets were selected based on a cut-off point of ≥ 0.05 of a quantified score of the importance of each feature. The feature selection was performed for each fiber combination data set individually.

2.6. Classification

2.6.1. RUSBoost model

The subsequent step was building a model with an ensemble random under sampling boosting tree (RUSBoost) as a classifier to distinguish tumorous tissue from healthy tissue [62]. We used this particular classifier since it has a higher classification performance on imbalanced data sets compared to other classification models [63]. The RUSBoost model mitigates the problem of imbalance by using 1) random under sampling [64] and 2) boosting [65]. Random under sampling means that in our case the algorithm randomly removes measurements from the majority class of healthy tissue until the desired class ratio is achieved at each iteration of the algorithm [62]. Boosting in our case means that the model iteratively builds an ensemble of models, where eventually all constructed models have a weighted vote to classify new data [62].

2.6.2. Cross-validation

We used iterative 5-fold cross-validation to classify the measurement locations. Each cross-validation process was repeated 20 times, with a different distribution of the patients over the folds during each of the 20 repetitions. To avoid bias, the data from each patient were randomly assigned to either the training set or the test set during each iteration, but never both. This analysis was executed for each of the data sets.

2.6.3. Performance evaluation

The following metrics averaged over 20 iterations were calculated: Matthews Correlation Coefficient (MCC), sensitivity, and specificity. The MCC is a metric that is less influenced by imbalanced data compared to accuracy. It is calculated by the following equation:

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (1)$$

where TP , TN , FP , and FN stand for true positives, true negatives, false positives, and false negatives respectively.

The MCC value ranges from -1 to 1, where -1 indicates a reverse correlation and 1 indicates a flawless correlation. A One-way Analysis of Variance (ANOVA) was performed to assess whether a particular number of fibers had a statistically significant, different mean MCC value in comparison to another number of fibers. The difference was considered statistically significant when the p -value was less than 0.05.

Furthermore, the influence of TTP on the performance of all classification models was analyzed by comparing the MCC, sensitivity and specificity for different TTP values. The classification models based on the number of fibers with significantly high MCC values for all TTP values were further analyzed. This was done by investigating the influence of the tumor-margin distance on

their sensitivity, by determining the proportion of correctly classified tumor locations for various tumor-margin distances. Additionally, the differences in correctly versus incorrectly classified locations using the lowest TTP (5%) were analyzed.

The entire workflow of data acquisition and analysis is displayed in Fig. 3.

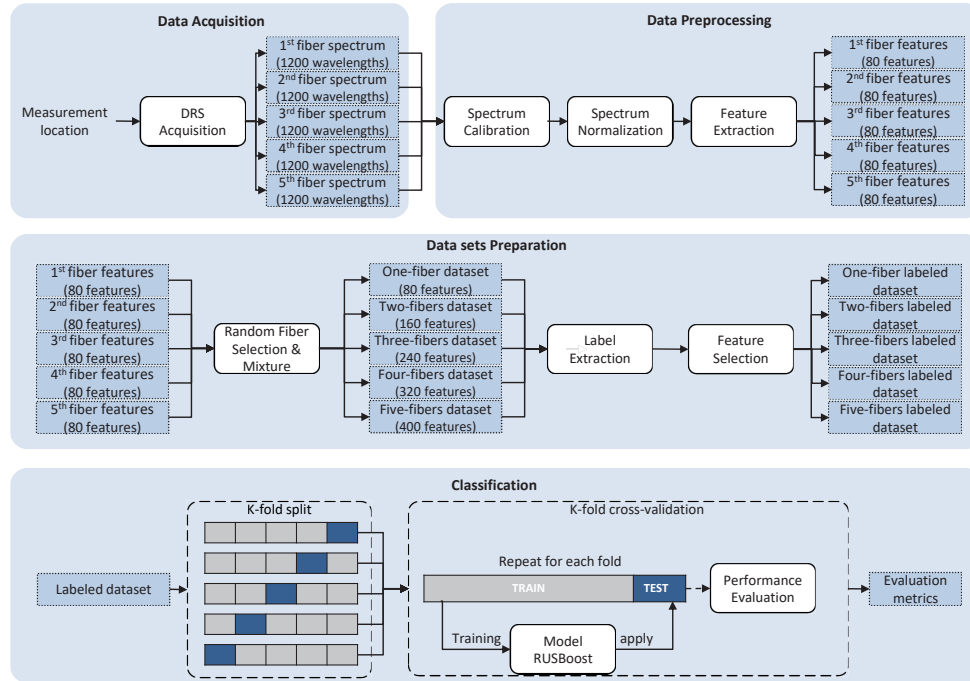


Fig. 3. Overview of the development and testing of all classification models. The process started with data acquisition, where 5 spectra were collected at each measurement location. The data was preprocessed by successively applying calibration, normalization, and feature extraction. In total, 80 features were extracted from each spectrum. Then 5 different data sets were built by using features from a different number of fibers (1-5). For each measurement location, the features from a randomly selected combination of fibers were used for each data set preparation. Subsequently, multiple data sets were built from the earlier data sets by labeling of the measurements (as tumorous and healthy) based on increasing cutoff points for tumor percentage. Thereafter, the most relevant features were selected from each data set. The labeled data sets were used to build RUSBoost classification models. The performance of the models was evaluated by an iterated 5-fold cross-validation method. Per patient, all spectra were assigned to one fold, ensuring that they were not split between the training and test set.

3. Results

3.1. Patient characteristics and measurement locations

In total, 1770 DRS spectra were obtained at 354 tissue locations from 100 breast specimens of 100 patients. An overview of the patient characteristics and measurement locations can be found in Table 1. The mean age of the patient population was 60,7 years (SD = 12,5) (Table 1). As far as the pathological diagnosis of the patients, 41 patients had IC of no special type (NST), 31 patients had IC NST combined with DCIS, 16 patients had an invasive lobular carcinoma (ILC), 10 patients had DCIS and 2 patients had lobular carcinoma in situ (LCIS) (Table 1). For the data

analysis, the patient groups with IC NST, IC NST combined with DCIS and ILC were combined, as well as the group with DCIS and LCIS.

Table 1. Patient characteristics and measurement locations ^a

| Patient characteristic / measurement location | Value |
|---|-------------|
| Characteristic N = 100 | |
| Age (years) (median, SD) | 60,7 (12,5) |
| Histological tumor type | |
| IC NST | 41 |
| IC NST + DCIS | 31 |
| ILC | 16 |
| DCIS | 10 |
| LCIS | 2 |
| Measurement locations (N = 354) | |
| IC within 2 mm from margin | 71 |
| DCIS within 2 mm from margin | 17 |
| Healthy tissue within 2 mm from margin | 266 |

^aIC NST = invasive carcinoma of no special type, DCIS = ductal carcinoma in situ,

^aILC = invasive lobular carcinoma, LCIS = lobular carcinoma in situ;

Of all measurement locations, there were 88 (25%) locations with tumor tissue within a distance of 2 mm from the margin and 266 (75%) locations with only healthy breast tissue. 71 (81%) of the tumor locations contained IC and 17 (19%) locations contained DCIS. In Fig. 4 the distance to the tumor is plotted against the tumor percentage for all IC within 2 mm from the margin locations (orange dots) and DCIS 2 mm from the margin locations (blue dots). The dashed lines represent the lines of best fit using linear regression analysis. From this figure, it is apparent that there is a negative, linear correlation between tumor percentage and tumor depth. This relationship is stronger for IC locations compared to DCIS locations.

3.2. Selected features

The MRMR analysis yielded a different number of optimum features per fiber combination data set, due to a different input number of features as can be seen in Fig. 3. The number of features varied from 5 to 35 features for each data set. We ranked the top 10 selected features with the highest importance scores per data set, and investigated the frequency of each of these features across all data sets. The 10 features with the highest frequencies are ranked in Table 2. As shown in Table 2, all of the displayed selected features are in the near-infrared wavelength range, where fat and water are the most important absorbers.

3.3. Effect of different numbers of fibers on classification performance

Different classification models were developed for each number of fibers based on features from 1, 2, 3, 4 or 5 fibers, and TTP values from 5%, 10%, 15%, 20%, 25%, 30%, 35% or 40%. For assigning the labels, a maximum value of 40% was used as TTP, since a higher TTP led to insufficient tumor labels in the training set. Figure 5 displays the distribution of tumor labels and healthy labels for different TTPs. As the TTP increases, the number of tumor labels decreases, while the total number of measurements remains equal. At the lowest TTP, 79 (22%) of all 354 measurement locations are labeled as tumor, while at the highest TTP, 44 (12%) of all locations are labeled as tumor.

The results of classifications for all numbers of fibers are demonstrated in Fig. 6. In this plot, each box visualizes the MCC values of all classification models over different TTPs for each

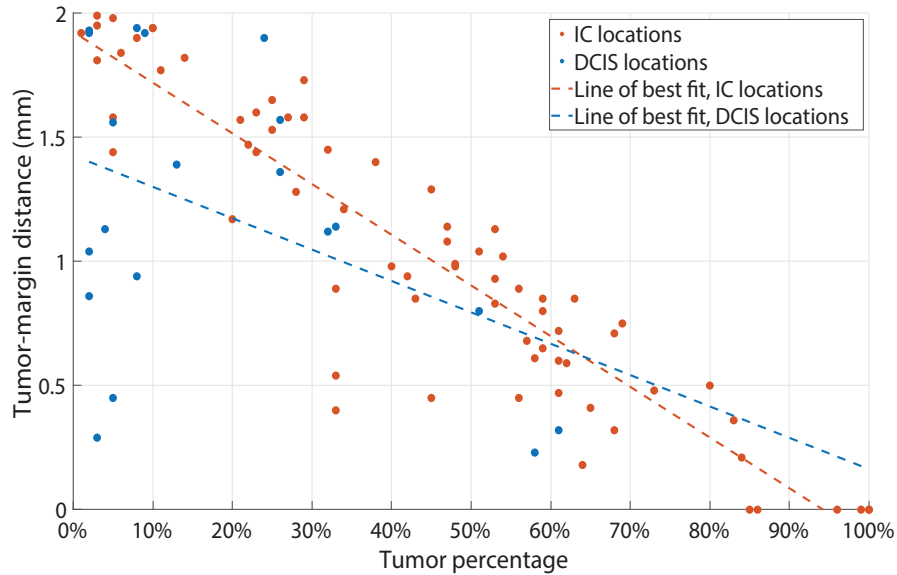


Fig. 4. Tumor-margin distance compared to the tumor percentage of all measurement locations. The orange dots represent IC locations, and the blue dots represent DCIS locations. The dashed lines are corresponding lines of best fit.

Table 2. Frequency rank of features among the top 10 features with the highest importance scores of all data sets

| Frequency rank | Selected features |
|----------------|--|
| 1 | Maximum difference between the slope and the measured spectrum between 1224 nm and 1331 nm |
| 2 | Wavelength of the right inflection point of the peak between 1008 nm and 1053 nm |
| 3 | Maximum difference between the slope and the measured spectrum between 1142 nm and 1225 nm |
| 4 | Slope of the measured spectrum between 999 nm and 1034 nm |
| 5 | Slope of the measured spectrum between 1213 nm and 1248 nm |
| 6 | Wavelength of the right inflection point of the peak between 1021 nm and 1102 nm |
| 7 | Slope of the measured spectrum between 1112 nm and 1147 nm |
| 8 | Slope of the measured spectrum between 883 nm and 1149 nm |
| 9 | Slope of the measured spectrum between 1024 nm and 1059 nm |
| 10 | Slope of the measured spectrum between 1395 and 1430 nm |

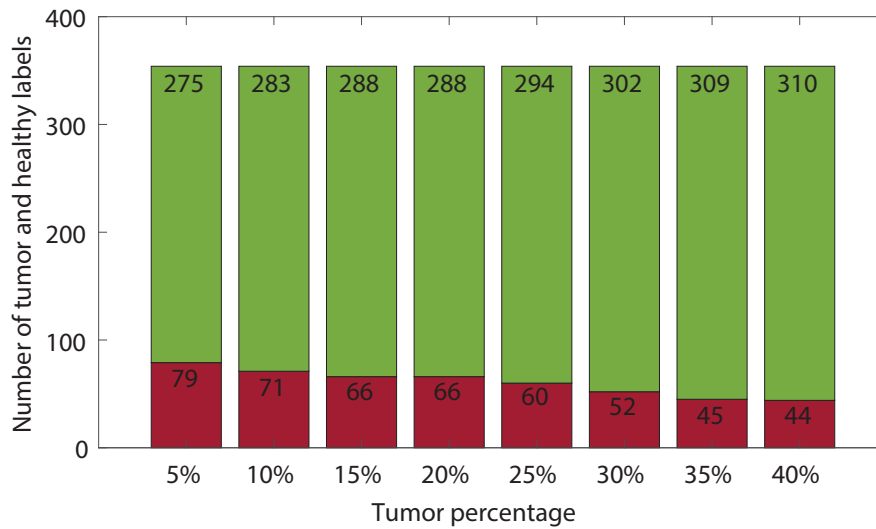


Fig. 5. Distribution of tumor labels (red bars) and healthy labels (green bars) in the overall data set for different TTPs.

fiber combination, where each MCC value in a box represents an average of all cross-validations per model. The minimum, maximum, first quartile and third quartile of all MCC values are displayed. The asterisks in this figure indicate the statistically significant differences between the performance of different fiber combinations, according to the One-way ANOVA test. Overall, the single-fiber classification models showed the lowest classification performance (median MCC 0.48, interquartile range (IQR) 0.14), while the classification models of 4 fibers showed the highest classification performance (median MCC 0.75, IQR 0.04). The median MCC of all classification models of 2 fibers, 3 fibers and 5 fibers were 0.57 (IQR 0.04), 0.73 (IQR 0.05), 0.72 (IQR 0.04), respectively. There were statistically significant differences (p -value < 0.05) between the performance of the models based on 1 fiber and 2 fibers compared to the models based on 3 or more fibers. From the plot it can be observed that the classification performance increases significantly with an increasing number of fibers, but reaches a plateau after 3 fibers.

3.4. Effect of tumor percentage on classification performance

We investigated the effect of tumor percentages on the tissue discrimination performance. Therefore, we repeated the experiments by training the classification model for all numbers of fibers while the data labels were adjusted based on tumor percentage. In Fig. 7, each colored line depicts the average MCC value of each classification model based on a different TTP, for a particular number of fibers. The shaded areas represent the accompanying average standard deviations (SD). The best performance is obtained when 4 fibers are used and the TTP is set at 20%, with a MCC value of 0.79 (SD 0.015). The lowest classification performance was obtained when using 1 fiber while setting the TTP at 20%, with a MCC value of 0.39 (SD 0.016). In general, at all TTP values, the use of 3, 4 or 5 fibers outperforms the use of 1 or 2 fibers. Another observation is that utilizing 3 or more fibers resulted in a SD of the MCC of 0.08, while the SD of the MCC of 3 fibers decreased to 0.03. This illustrates the robust performance of 3 to 5 fibers over different TTPs, while the performance of the models based on 1 and 2 fibers changes significantly over different TTPs.

The calculated sensitivity and specificity of each classification model are depicted in Fig. 8. In general, all classification models have a high sensitivity, ranging between 80% and 93%. However, the specificity varies widely across different models, ranging from 50% to 86%. The

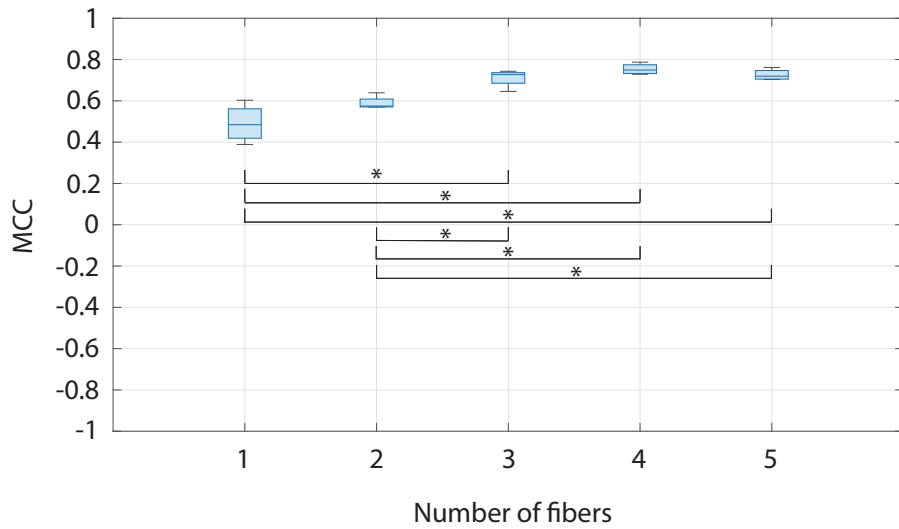


Fig. 6. Classification performance for different numbers of fibers. Each box is composed of all average MCC values of models based on different TTP values and the same number of fibers. The lines with an asterisk indicate a p -value of 0.05.

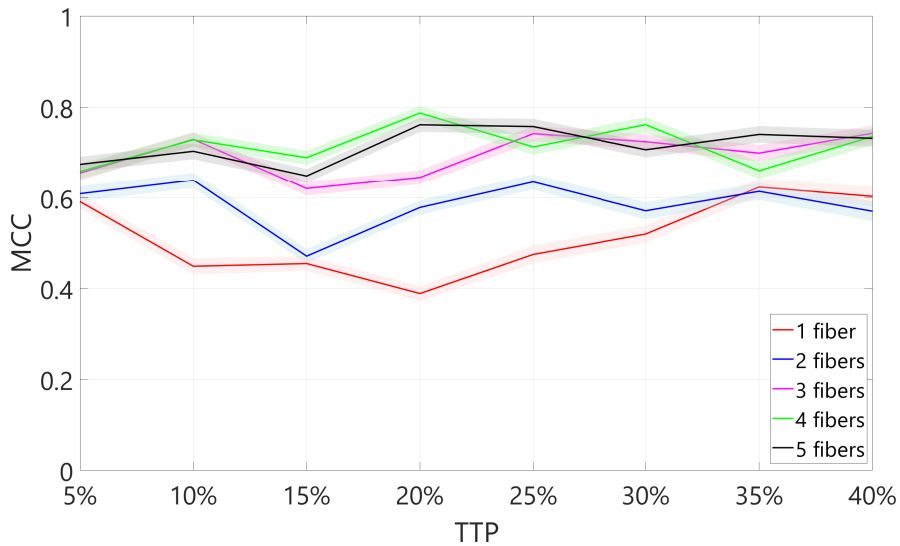


Fig. 7. Classification performance of models based on different TTPs. Each line color represents a different number of fibers. The TTP is displayed on the x-axis and the MCC on the y-axis. The shaded areas indicate the standard deviations.

highest sensitivity of 93% was achieved using 5 fibers and a TTP of 20%. This model had a specificity of 75%. On the other hand, the highest specificity of 86% was obtained with 4 fibers and a TTP of 10%. This model had a sensitivity of 88%. Among all possible numbers of fibers and TTP values, higher sensitivity and specificity can always be achieved when using 3 to 5 fibers compared to 1 or 2 fibers. The exact values of sensitivity and specificity for all classification models can be found in [Supplement 1](#) (Section 1).

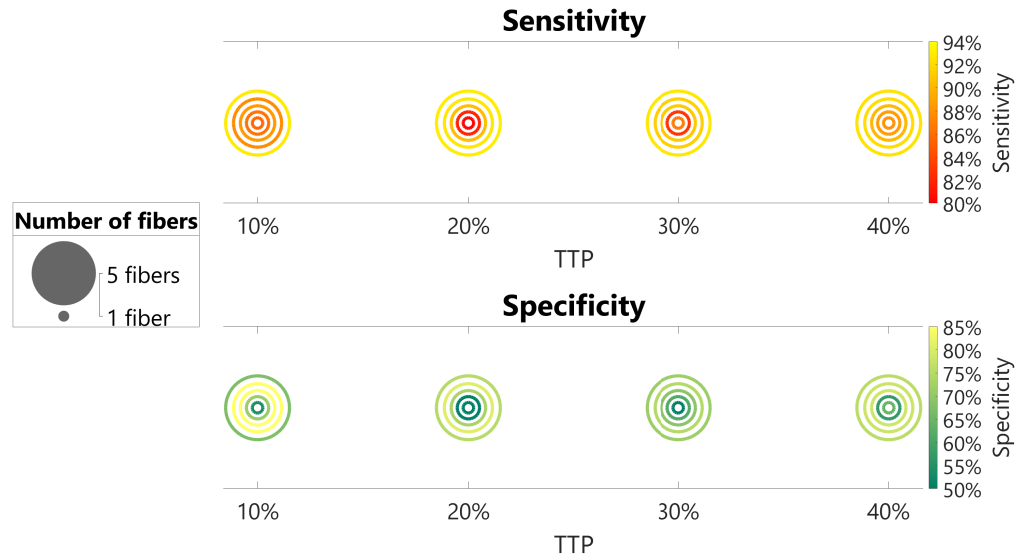


Fig. 8. Sensitivity and specificity of all classification models. In both diagrams, 4 sets of rings are present, with each set containing 5 rings of different sizes and colors. The size represents the number of fibers, while the color represents sensitivity in the upper diagram and specificity in the lower diagram.

3.5. Effect of tumor-margin distance on classification performance

In general, the previous experiment indicates that the use of 3 to 5 fibers outperforms the use of 1 to 2 fibers significantly. Thus, for the remainder of this study, we have analyzed the effects of the tumor-margin distance on the classification performance for only 3, 4, and 5 fibers.

Figure 9 displays the percentage of locations with tumor tissue that are correctly classified when using different numbers of fibers at various tumor-margin distances. The tumor margin distance has been divided into 4 bins of 0.5 mm. Each graph represents a similar experiment but with a different TTP value (10%, 20%, 30%, and 40%). It is worth mentioning that with a TTP of 30% or 40%, there were no locations with tumor tissue deeper than 1.50 mm.

In general, all graphs show a similar trend, where the percentage of correctly classified tumor locations decreases with an increase in the distance to the tumor. The percentage of correctly classified tumor locations varies from 75% to 100%, depending on the TTP and the number of fibers. Among the models based on a TTP of 30%, the use of 4 or 5 fibers gives the highest sensitivity for the shortest tumor-margin distance category, while for a TTP of 40%, it can be observed that all selected numbers of fibers have the same sensitivity. In these scenarios, 100% of all locations with a tumor-margin distance of ≤ 0.50 mm are classified correctly. Furthermore, for a tumor-margin distance ≥ 1.0 mm, 5 fibers give the highest percentage of correctly classified locations for all different TTP values.

Furthermore, in Fig. 10, the tumor-margin distance of all correctly and incorrectly classified IC and DCIS locations is plotted against the tumor tissue percentage. The distribution of locations in

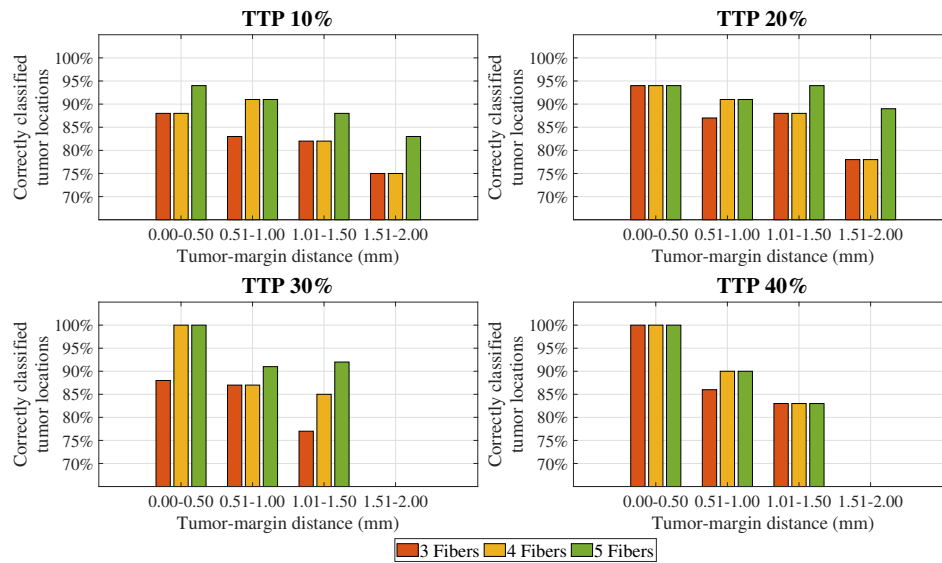


Fig. 9. The percentage of correctly classified locations at different tumor-margin distances for different TTPs. Each bar color represents a different number of fibers, while each of the 4 graphs represents a different TTP.

this graph demonstrates that DCIS locations with a low tumor percentage (0-20%) in combination with a low tumor-margin distance (0-1 mm) are existent, while this is not the case for IC locations. All graphs show a similar trend, where the number of misclassified locations increases with an increasing tumor-margin distance. Furthermore, the total number of misclassified locations decreases with an increasing number of fibers. In essence, it shows that in a scenario where the lowest possible cut-off point for tumor percentage would be used, the highest chance of correct classification for each tumor location would occur when using 5 fibers. In this scenario, 89% of all IC locations and 86% of all DCIS locations get classified correctly, regardless of tumor-margin-distance. Furthermore, 94% of all locations with a tumor-margin distance ≤ 0.50 mm get classified correctly, regardless of tumor percentage. The specificity of distinguishing tumorous tissue from healthy tissue is 75% when using 3 fibers, 78% when using 4 fibers and 74% when using 5 fibers.

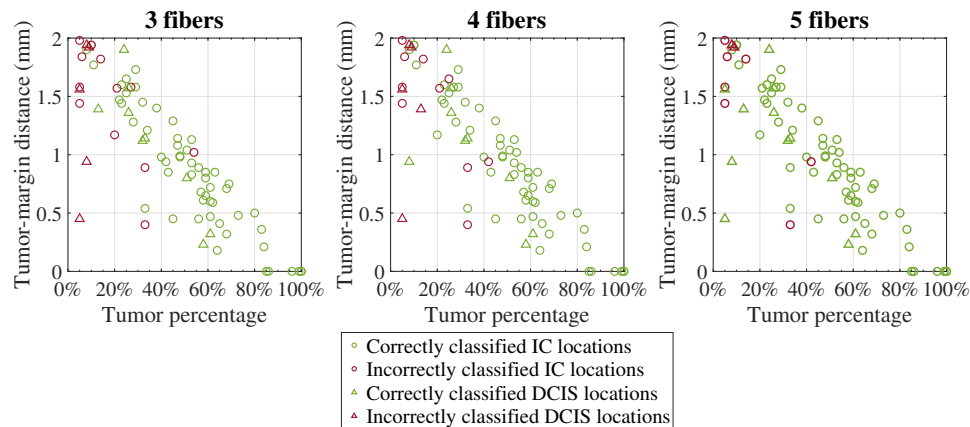


Fig. 10. Tumor-margin distance compared to the tumor percentage of all correctly and incorrectly classified IC and DCIS locations when using classification models based on a TTP of 5% for 3, 4, and 5 fibers.

4. Discussion

The first goal of this study was to develop breast tissue classification models using optical spectral features and to investigate the optimal number of optical fibers for accurate discrimination of tumorous breast tissue from healthy tissue. In order to do so, we compared the performance of several classification models using 1 to 5 fibers.

As for the first research question concerning the number of optical fibers, it was found that in general, experiments with 3 to 5 fibers had a substantially higher performance compared to the use of 1 to 2 fibers. This could be explained by the fact that malignant breast lesions often possess an irregular shape, and therefore could easily have a significant variation in the tumor-margin distance over a relatively short trajectory. A DRS probe with 1 or 2 emitting fiber(s) and 1 receiving fiber collects optical information of a smaller tissue volume compared to a DRS probe with a similar tip diameter, but containing 3, 4 or 5 optical emitting fibers encircling 1 receiving fiber. Combining the features of multiple fibers enables the algorithm to give a more accurate classification result regarding the sampled volume. We found no significant difference between the models using 3, 4 or 5 fibers. One possible explanation for this might be that 3 fibers already provide sufficient data from the probed volume, meaning that more fibers would not provide additional information regarding the tumor presence in that volume. Another possible explanation might be that the addition of features beyond the features from 3 fibers leads to redundant information and noise, which may cause overfitting during the training phase.

The second goal of this study was to assess how the performance of different numbers of fibers depended on the tumor percentage. For this purpose, the experiments were repeated several times each time with different tumor percentage threshold values for labels assignment. The evaluation metrics showed that for all TTP values the models based on 3 to 5 fibers outperformed the models based on 1 to 2 fibers. It was also shown that the models based on 3 to 5 fibers have more robust performance with similar MCC over all TTP values, while in the single and double fiber models the MCC value changes significantly over the TTPs.

The third goal was to investigate the performance of different numbers of fibers in comparison with the tumor-margin distance. Concerning this, we investigated the number of correctly classified tumor locations per tumor-margin distance category for the best set of models. It was found that the percentage of correctly classified tumor locations decreased with an increase in the tumor-margin distance. This suggests that spectra collected from locations with deeper

tumors are optically less different to spectra of healthy tissue, compared to spectra collected from locations with more superficial tumors. Therefore, deeper tumor locations are more difficult to correctly classify. For tumors deeper than 1.0 mm, 5 fibers had the highest sensitivity regardless of the TTP. A possible explanation is that when you have the highest number of fibers on the same tissue surface area, there is a higher chance of measuring tumorous tissue when it is present within 2 mm.

Among the patients included in this study, there were 5 patients with a positive resection margin according to Dutch guidelines. The clinical consequences for these patients were that 3 patients had to undergo a re-excision, and 2 patients had to undergo additional boost radiotherapy. The unnecessary burden for these patients would have been prevented if DRS was applied on the resection margins after surgery, since all these positive margin locations were correctly identified by the classification models using 3, 4 or 5 fibers. This illustrates the clinical value of this technology in practice.

It is noteworthy that during this research, we have put maximum effort to correlate the measured tissue locations to the corresponding pathological outcome. However, this correlation has an inherent shortcoming since the H&E section is a 2D image of a few cell layers, while the probed tissue is a 3D volume. Therefore, the tissue areas in the H&E section will not completely represent the probed tissue volume. Furthermore, we have not investigated the impact of benign tissue structures with high nuclei densities on the margin assessment accuracy. This would be an interesting point to investigate in the future.

In order to integrate DRS into a clinically applicable tool, it is important to conduct a large clinical study using DRS *in vivo* directly on the breast tissue to be resected during BCS, to validate the performance of the classification models. Moreover, it should be investigated which tumor percentage should be used as a threshold point for achieving the most clinically relevant performance. DRS could be applied for margin assessment using two different methods: 1) a DRS probe for classification of suspicious margin locations on the specimen directly after excision, ensuring additional resection of any tumor tissue which was left behind in the breast, 2) a surgical resection tool allowing the combination of DRS measurements, real-time classification, and tissue resection based on the classification results. For the latter method, one emitting fiber and one receiving fiber would suffice, since you could move the probe over an entire tissue surface area allowing quick measurements and classification in a continuous manner. It is important to mention that in such a study, the measured locations *in vivo* could be marked using sutures or surgical clips, followed by a similar pathology analysis as conducted in this study. When investigating both methods, the usability of both tools should be evaluated among surgeons. When these questions are investigated, the next step would be to set up a clinical study to evaluate the effect of a DRS-integrated tool during breast-conserving surgeries on clinical end-points such as the number of positive resection margins, and the resection volumes in relation to the tumor volumes.

5. Conclusion

In this study, we have investigated the optimum number of fibers for distinguishing tumorous breast tissue from healthy breast tissue on the resection margin. The results demonstrate that in general, classification models based on 3 or more fibers lead to a 52% increase in the MCC value compared to 1 fiber models, regardless of tumor percentage as well as the distance to the tumor. The models based on 3 to 5 fibers also show a more robust performance over different tumor percentages. Although using 4 and 5 fibers slightly outperform 3 fibers in some experiments, there was no significant difference between their performances. Furthermore, when looking at the models based on 3 to 5 fibers, the percentage of correctly classified tumor locations varies from 75% to 100% depending on the tumor percentage, the tumor-margin distance and the number of fibers. In conclusion, to achieve reliable and robust performance in using DRS for surgical

margin assessment, we require at least 3 fibers to be able to detect small tumor percentages up to 2 mm in depth, with 5 fibers providing the highest detection accuracy. A large clinical *in vivo* DRS study would be the next step towards reaching the ultimate goal of intraoperatively evaluating surgical margins during breast-conserving surgeries.

Acknowledgments. The authors thank all employees from the NKI-AvL Core Facility Molecular Pathology & Biobanking (CFMPB), all surgeons and nurses from the Department of Surgery, and all pathologists and pathologist assistants from the Department of Pathology for their assistance.

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are not publicly available at the time but may be obtained from the authors upon reasonable request.

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. C. Chiappa, F. Rovera, A. D. Corben, A. Fachinetti, V. De Berardinis, V. Marchionini, S. Rausei, L. Boni, G. Dionigi, and R. Dionigi, "Surgical margins in breast conservation," *Int. J. Surg.* **11**, S69–S72 (2013).
2. E. Provenzano, V. Bossuyt, G. Viale, D. Cameron, S. Badve, C. Denkert, G. MacGrogan, F. Penault-Llorca, J. Boughey, G. Curigliano, J. M. Dixon, L. Esserman, G. Fastner, T. Kuehn, F. Peintinger, G. von Minckwitz, J. White, W. Yang, and W. F. Symmans, "Standardization of pathologic evaluation and reporting of postneoadjuvant specimens in clinical trials of breast cancer: recommendations from an international working group," *Mod. Pathol.* **28**(9), 1185–1201 (2015).
3. A. Nayyar, K. K. Gallagher, and K. P. McGuire, "Definition and management of positive margins for invasive breast cancer," *Surg. Clin.* **98**(4), 761–771 (2018).
4. C. Dunne, J. P. Burke, M. Morrow, and M. R. Kell, "Effect of margin status on local recurrence after breast conservation and radiation therapy for ductal carcinoma in situ," in *Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]*, (Centre for Reviews and Dissemination (UK), 2009).
5. B. Spivack, M. M. Khanna, L. Tafra, G. Juillard, and A. E. Giuliano, "Margin status and local recurrence after breast-conserving surgery," *Arch. Surg.* **129**(9), 952–957 (1994).
6. F. Meric, N. Q. Mirza, G. Vlastos, T. A. Buchholz, H. M. Kuerer, G. V. Babiera, S. E. Singletary, M. I. Ross, F. C. Ames, B. W. Feig, S. Krishnamurthy, G. H. Perkins, M. D. McNeese, E. A. Strom, V. Valero, and K. K. Hunt, "Positive surgical margins and ipsilateral breast tumor recurrence predict disease-specific survival after breast-conserving therapy," *Cancer* **97**(4), 926–933 (2003).
7. S. E. Singletary, "Surgical margins in patients with early-stage breast cancer treated with breast conservation therapy," *The Am. journal of surgery* **184**(5), 383–393 (2002).
8. A. Bodilsen, K. Bjerre, B. V. Offersen, P. Vahl, N. Amby, J. M. Dixon, B. Ejlersen, J. Overgaard, and P. Christiansen, "Importance of margin width in breast-conserving treatment of early breast cancer," *J. Surg. Oncol.* **113**(6), 609–615 (2016).
9. M. Morrow, K. J. Van Zee, L. J. Solin, N. Houssami, M. Chavez-MacGregor, J. R. Harris, J. Horton, S. Hwang, P. L. Johnson, M. L. Marinovich, S. J. Schnitt, I. Wapnir, and M. S. Moran, "Society of surgical oncology-american society for radiation oncology-american society of clinical oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in ductal carcinoma in situ," *Pract. Radiat. Oncol.* **6**(5), 287–295 (2016).
10. S. G. B. de Koning, M.-J. T. V. Peeters, K. Józwiak, P. A. Bhairosing, and T. J. Ruers, "Tumor resection margin definitions in breast-conserving surgery: systematic review and meta-analysis of the current literature," *Clin. Breast Cancer* **18**(4), e595–e600 (2018).
11. L. E. McCahill, R. M. Single, E. J. A. Bowles, H. S. Feigelson, T. A. James, T. Barney, J. M. Engel, and A. A. Onitilo, "Variability in reexcision following breast conservation surgery," *JAMA* **307**(5), 467–475 (2012).
12. S. L. Blair, K. Thompson, J. Rococco, V. Malcarne, P. D. Beitsch, and D. W. Ollila, "Attaining negative margins in breast-conservation operations: is there a consensus among breast surgeons?" *J. Am. Coll. Surg.* **209**(5), 608–613 (2009).
13. M. Pilewskie and M. Morrow, "Margins in breast cancer: How much is enough?" *Cancer* **124**(7), 1335–1341 (2018).
14. M. Azu, P. Abrahamse, S. J. Katz, R. Jagsi, and M. Morrow, "What is an adequate margin for breast-conserving surgery? surgeon attitudes and correlates," *Ann. Surg. Oncol.* **17**(2), 558–563 (2010).
15. M. S. Moran, S. J. Schnitt, A. E. Giuliano, J. R. Harris, S. A. Khan, J. Horton, S. Klimberg, M. Chavez-MacGregor, G. Freedman, N. Houssami, P. L. Johnson, and M. Morrow, "Society of surgical oncology-american society for radiation oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages i and ii invasive breast cancer," *Int. J. Radiat. Oncol.* **88**(3), 553–564 (2014).
16. N. B. O. Nederland, "Nabon (2012). mammacarcinoom. landelijke richtlijn, versie: 2.0," (2018).
17. M. A. Olsen, K. B. Nickel, J. A. Margenthaler, A. E. Wallace, D. Mines, J. P. Miller, V. J. Fraser, and D. K. Warren, "Increased risk of surgical site infection among breast-conserving surgery re-excisions," *Ann. Surg. Oncol.* **22**(6), 2003–2009 (2015).

18. S. Collette, L. Collette, and T. Budiharto, *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* **44**(17), 2587–2599 (2008).
19. D. E. Wazer, T. DiPetrillo, R. Schmidt-Ullrich, L. Weld, T. Smith, D. Marchant, and N. Robert, "Factors influencing cosmetic outcome and complication risk after conservative surgery and radiotherapy for early-stage breast carcinoma," *J. Clin. Oncol.* **10**(3), 356–363 (1992).
20. J. Heil, K. Breitzkreuz, M. Golatta, E. Czink, J. Dahlkamp, J. Rom, F. Schuetz, M. Blumenstein, G. Rauch, and C. Sohn, "Do reexcisions impair aesthetic outcome in breast conservation surgery? exploratory analysis of a prospective cohort study," *Ann. Surg. Oncol.* **19**(2), 541–547 (2012).
21. E. Hau, L. Browne, A. Capp, G. P. Delaney, C. Fox, J. H. Kearsley, E. Millar, E. H. Nasser, G. Papadatos, and P. H. Graham, "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.* **139**(1), 115–123 (2013).
22. J. F. Waljee, E. S. Hu, P. A. Ubel, D. M. Smith, L. A. Newman, and A. K. Alderman, "Effect of esthetic outcome after breast-conserving surgery on psychosocial functioning and quality of life," *J. Clin. Oncol.* **26**(20), 3331–3337 (2008).
23. J. H. Volders, V. L. Negenborn, M. H. Haloua, N. M. Krekel, K. Jóźwiak, S. Meijer, and P. M. van den Tol, "Cosmetic outcome and quality of life are inextricably linked in breast-conserving therapy," *J. Surg. Oncol.* **115**(8), 941–948 (2017).
24. R. Pataky and C. Baliski, "Reoperation costs in attempted breast-conserving surgery: a decision analysis," *Curr. Oncol.* **23**(5), 314–321 (2016).
25. S. E. Abe, J. S. Hill, Y. Han, K. Walsh, J. T. Symanowski, L. Hadzikadic-Gusic, T. Flippo-Morton, T. Sarantou, M. Forster, and R. L. White Jr, "Margin re-excision and local recurrence in invasive breast cancer: a cost analysis using a decision tree model," *J. Surg. Oncol.* **112**(4), 443–448 (2015).
26. K. B. Clough, J. Cuminet, A. Fitoussi, C. Nos, and V. Mosseri, "Cosmetic sequelae after conservative treatment for breast cancer: classification and results of surgical correction," *Ann. Plast. Surg.* **41**(5), 471–481 (1998).
27. K. Sneeuw, N. Aaronson, J. Yarnold, M. Broderick, J. Regan, G. Ross, and A. Goddard, "Cosmetic and functional outcomes of breast conserving treatment for early stage breast cancer. 1. comparison of patients' ratings, observers' ratings and objective assessments," *Radiother. Oncol.* **25**(3), 153–159 (1992).
28. R. Cochrane, P. Valasiadou, A. Wilson, S. Al-Ghazal, and R. Macmillan, "Cosmesis and satisfaction after breast-conserving surgery correlates with the percentage of breast volume excised," *J. Br. Surg.* **90**(12), 1505–1509 (2003).
29. M. E. Taylor, C. A. Perez, K. J. Halverson, R. R. Kuske, G. W. Philpott, D. M. Garcia, J. E. Mortimer, R. J. Myerson, D. Radford, and C. Rush, "Factors influencing cosmetic results after conservation therapy for breast cancer," *Int. J. Radiat. Oncol. Biol. Phys.* **31**(4), 753–764 (1995).
30. C. Vrieling, L. Collette, A. Fourquet, W. J. Hoogenraad, J.-C. Horiot, J. J. Jager, M. Pierart, P. M. Poortmans, H. Struikmans, B. Maat, E. Van Limbergen, and H. Bartelink, "The influence of patient, tumor and treatment factors on the cosmetic results after breast-conserving therapy in the eortc 'boost vs. no boost' trial," *Radiother. Oncol.* **55**(3), 219–232 (2000).
31. T. Hashem, A. Morsi, A. Farahat, T. Zaghloul, and A. Hamed, "Correlation of specimen/breast volume ratio to cosmetic outcome after breast conserving surgery," *Indian J. Surg. Oncol.* **10**(4), 668–672 (2019).
32. N. Krekel, B. Zonderhuis, S. Muller, H. Bril, H.-J. van Slooten, E. de Lange de Klerk, P. van den Tol, and S. Meijer, "Excessive resections in breast-conserving surgery: a retrospective multicentre study," *The breast journal* **17**(6), 602–609 (2011).
33. F. A. M. Valejo, D. G. Tiezzi, L. R. M. Mandarano, C. B. d. Sousa, and J. M. d. Andrade, "Volume of breast tissue excised during breast-conserving surgery in patients undergoing preoperative systemic therapy," *Revista Brasileira de Ginecologia e Obstet.* **35**(5), 221–225 (2013).
34. Y. D. Shin, Y. J. Choi, D. H. Kim, S. S. Park, H. Choi, D. J. Kim, S. Park, H. Y. Yun, and Y. J. Song, "Comparison of outcomes of surgeon-performed intraoperative ultrasonography-guided wire localization and preoperative wire localization in nonpalpable breast cancer patients undergoing breast-conserving surgery: a retrospective cohort study," *Medicine* **96**(50), e9340 (2017).
35. S. K. Al-Ghazal, L. Fallowfield, and R. Blamey, "Does cosmetic outcome from treatment of primary breast cancer influence psychosocial morbidity?" *Eur. J. Surg. Oncol.* **25**(6), 571–573 (1999).
36. T. E. Doyle, R. E. Factor, C. L. Ellefson, K. M. Sorensen, B. J. Ambrose, J. B. Goodrich, V. P. Hart, S. C. Jensen, H. Patel, and L. A. Neumayer, "High-frequency ultrasound for intraoperative margin assessments in breast conservation surgery: a feasibility study," *BMC Cancer* **11**(1), 444 (2011).
37. N. M. A. Krekel, B. M. Zonderhuis, H. W. H. Schreurs, A. M. F. L. Cardozo, H. Rijna, H. van der Veen, S. Muller, P. Poortman, L. de Widt, W. K. de Roos, A. M. Bosch, A. H. M. Taets van Amerongen, E. Bergers, M. H. M. van der Linden, E. S. M. de Lange de Klerk, H. A. H. Winters, S. Meijer, and P. M. P. van den Tol, "Ultrasound-guided breast-sparing surgery to improve cosmetic outcomes and quality of life. a prospective multicentre randomised controlled clinical trial comparing ultrasound-guided surgery to traditional palpation-guided surgery (cobalt trial)," *BMC Surg.* **11**(1), 8–10 (2011).
38. R. Pleijhuis, G. Langhout, W. Helfrich, G. Themelis, A. Sarantopoulos, L. Crane, N. Harlaar, J. De Jong, V. Ntziachristos, and G. Van Dam, "Near-infrared fluorescence (nirf) imaging in breast-conserving surgery: assessing intraoperative techniques in tissue-simulating breast phantoms," *Eur. J. Surg. Oncol.* **37**(1), 32–39 (2011).

39. A. S. Haka, Z. I. 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**(5), 054023 (2009).
40. R. Ha, L. C. Friedlander, H. Hibshoosh, C. Hendon, S. Feldman, S. Ahn, H. Schmidt, M. K. Akens, M. Fitzmaurice, B. C. Wilson, and V. L. Mango, "Optical coherence tomography: A novel imaging method for post-lumpectomy breast margin assessment-a multi-reader study," *Acad. Radiol.* **25**(3), 279–287 (2018).
41. K. Y. Foo, K. M. Kennedy, R. Zilkens, W. M. Allen, Q. Fang, R. W. Sanderson, J. Anstie, B. F. Dessauvage, B. Latham, C. M. Saunders, L. Chin, and B. F. Kennedy, "Optical palpation for tumor margin assessment in breast-conserving surgery," *Biomed. Opt. Express* **12**(3), 1666–1682 (2021).
42. I. Pappo, R. Spector, A. Schindel, S. Morgenstern, J. Sandbank, L. T. Leider, S. Schneebaum, S. Lelcuk, and T. Karni, "Diagnostic performance of a novel device for real-time margin assessment in lumpectomy specimens," *J. Surg. Res.* **160**(2), 277–281 (2010).
43. R. Li, P. Wang, L. Lan, F. P. Lloyd, C. J. Goergen, S. Chen, and J.-X. Cheng, "Assessing breast tumor margin by multispectral photoacoustic tomography," *Biomed. Opt. Express* **6**(4), 1273–1281 (2015).
44. A. R. Pradipta, T. Tanei, K. Morimoto, K. Shimazu, S. Noguchi, and K. Tanaka, "Emerging technologies for real-time intraoperative margin assessment in future breast-conserving surgery," *Adv. Sci.* **7**(9), 1901519 (2020).
45. J. J. Keating, C. Fisher, R. Batiste, and S. Singhal, "Advances in intraoperative margin assessment for breast cancer," *Curr. Surg. Rep.* **4**(4), 15 (2016).
46. E. R. St John, R. Al-Khudairi, H. Ashrafian, T. Athanasiou, Z. Takats, D. J. Hadjiminas, A. Darzi, and D. R. Leff, "Diagnostic accuracy of intraoperative techniques for margin assessment in breast cancer surgery," *Ann. Surg.* **265**(2), 300–310 (2017).
47. A. J. Gomes and V. Backman, "Algorithm for automated selection of application-specific fiber-optic reflectance probes," *J. Biomed. Opt.* **18**(2), 027012 (2013).
48. L. De Boer, B. Molenkamp, T. Bydlon, B. Hendriks, J. Wesseling, H. Sterenborg, and T. Ruers, "Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries," *Breast Cancer Res. Treat.* **152**(3), 509–518 (2015).
49. L. L. de Boer, E. Kho, K. K. Van de Vijver, M.-J. T. Vranken Peeters, F. van Duijnhoven, B. H. Hendriks, H. J. Sterenborg, and T. J. Ruers, "Optical tissue measurements of invasive carcinoma and ductal carcinoma in situ for surgical guidance," *Breast Cancer Res.* **23**(1), 59 (2021).
50. Z. Volynskaya, A. S. Haka, K. L. Bechtel, M. Fitzmaurice, R. Shenk, 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**(2), 024012 (2008).
51. J. S. Soares, I. Barman, N. C. Dingari, Z. Volynskaya, W. Liu, N. Klein, D. Plecha, R. R. Dasari, and M. Fitzmaurice, "Diagnostic power of diffuse reflectance spectroscopy for targeted detection of breast lesions with microcalcifications," *Proc. Natl. Acad. Sci.* **110**(2), 471–476 (2013).
52. 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**(1), 15–23 (2010).
53. 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**(5), 054009 (2008).
54. D. J. Evers, R. Nachabe, M.-J. Vranken Peeters, J. A. van der Hage, H. S. Oldenburg, E. J. Rutgers, G. W. Lucassen, B. H. Hendriks, J. Wesseling, and T. J. Ruers, "Diffuse reflectance spectroscopy: towards clinical application in breast cancer," *Breast Cancer Res. Treat.* **137**(1), 155–165 (2013).
55. R. Nachabé, D. J. Evers, B. H. W. Hendriks, G. W. Lucassen, M. van der Voort, E. J. Rutgers, M.-J. V. Peeters, J. A. Van der Hage, H. S. Oldenburg, J. Wesseling, and T. J. M. Ruers, "Diagnosis of breast cancer using diffuse optical spectroscopy from 500 to 1600 nm: comparison of classification methods," *J. Biomed. Opt.* **16**(8), 087010 (2011).
56. R. Nachabe, B. H. Hendriks, A. E. Desjardins, M. van der Voort, M. B. van der Mark, and H. J. Sterenborg, "Estimation of lipid and water concentrations in scattering media with diffuse optical spectroscopy from 900 to 1600 nm," *J. Biomed. Opt.* **15**(3), 037015 (2010).
57. F. Geldof, B. Dashtbozorg, B. H. Hendriks, H. J. Sterenborg, and T. J. Ruers, "Layer thickness prediction and tissue classification in two-layered tissue structures using diffuse reflectance spectroscopy," *Sci. Rep.* **12**(1), 1698 (2022).
58. E. Kho, B. Dashtbozorg, L. L. De Boer, K. K. Van de Vijver, H. J. Sterenborg, and T. J. Ruers, "Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information," *Biomed. Opt. Express* **10**(9), 4496–4515 (2019).
59. P. Geladi, D. MacDougall, and H. Martens, "Linearization and scatter-correction for near-infrared reflectance spectra of meat," *Appl. Spectrosc.* **39**(3), 491–500 (1985).
60. L. L. De Boer, T. M. Bydlon, F. Van Duijnhoven, M.-J. T. Vranken Peeters, C. E. Loo, G. A. Winter-Warnars, J. Sanders, H. J. Sterenborg, B. H. Hendriks, and T. J. Ruers, "Towards the use of diffuse reflectance spectroscopy for real-time in vivo detection of breast cancer during surgery," *J. Transl. Med.* **16**(1), 367 (2018).
61. H. Peng, C. Ding, and F. Long, "Minimum redundancy-maximum relevance feature selection," (2005).
62. C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Rusboost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst., Man, Cybern. A* **40**(1), 185–197 (2010).

63. T. Tong, C. Ledig, R. Guerrero, A. Schuh, J. Koikkalainen, A. Tolonen, H. Rhodius, F. Barkhof, B. Tijms, A. W. Lemstra, H. Soininen, A. M. Remes, G. Waldemar, S. Hasselbalch, P. Mecocci, M. Baroni, J. Lötjönen, W. van der Flier, and D. Rueckert, "Five-class differential diagnostics of neurodegenerative diseases using random undersampling boosting," *NeuroImage: Clin.* **15**, 613–624 (2017).
64. D.-H. Jeong, S.-E. Kim, W.-H. Choi, and S.-H. Ahn, "A comparative study on the influence of undersampling and oversampling techniques for the classification of physical activities using an imbalanced accelerometer dataset," in *Healthcare*, vol. 10 (MDPI, 2022), p. 1255.
65. J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, "Boosting methods for multi-class imbalanced data classification: an experimental review," *J. Big Data* **7**(1), 70 (2020).