



Optimizing algorithm development for tissue classification in colorectal cancer based on diffuse reflectance spectra

ELISABETH J. M. BALTUSSEN,^{1,*}  HENRICUS J. C. M. STERENBORG,^{1,2} THEO J. M. RUERS,^{1,3} AND BEHDAD DASHTBOZORG^{1,4} 

¹Department of Surgery, Antoni van Leeuwenhoek Hospital – The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX Amsterdam, The Netherlands

²Department of Biomedical Engineering and Physics, Amsterdam University Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

³Faculty of Science and Technology, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands

⁴Department of Biomedical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

*l.baltussen@nki.nl

Abstract: Diffuse reflectance spectroscopy can be used in colorectal cancer surgery for tissue classification. The main challenge in the classification task is to separate healthy colorectal wall from tumor tissue. In this study, four normalization techniques, four feature extraction methods and five classifiers are applied to nine datasets, to obtain the optimal method to separate spectra measured on healthy colorectal wall from spectra measured on tumor tissue. All results are compared to the use of the entire non-normalized spectra. It is found that the most optimal classification approach is to apply a feature extraction method on non-normalized spectra combined with support vector machine or neural network classifier.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

In both men and women, colorectal cancer is the third most common cancer worldwide [1]. The standard care for patients with colorectal cancer is surgery [2]. During colorectal cancer surgery, the surgeon must rely on visual and tactile feedback to recognize the tumor boundaries. As a result, there are two main issues: complete removal of the tumor and preventing complication due to too extensive surgery. To obtain an optimal balance between these two issues, intra-operative tissue recognition would be of great benefit to provide the surgeon with information about the tissue he or she encounters.

Diffuse reflectance spectroscopy (DRS) is an optical technique that has been used for tissue recognition in cancer diagnostics [3–10]. Light over a broad wavelength range is sent into the tissue, where the light interacts by scattering and absorption. Part of this light will be scattered back to the surface of the tissue, where it can be detected. In fiberoptic DRS, light is transported through fibers, both from source to tissue and from the tissue surface to the spectrometers [11]. In the fiberoptic geometry only point measurements can be performed. Using hyperspectral imaging (HSI), the tissue samples are directly illuminated by a broadband light source and the reflected light over the entire illuminated surface is collected by a camera. Therefore, a complete specimen can be imaged at once. With HSI a 3D data-cube is collected, in which two dimensions contain the spatial information about the tissue sample and the third dimension represents the wavelength information for each spatial pixel [10].

Both fiberoptic DRS and diffuse reflection HSI have been used in tissue classification in colorectal cancer before. For HSI the main applications in colorectal cancer so far were on imaging of the pathology slides for automatic pathology classification [12–16] and on colonoscopy in which tumor was distinguished from healthy colorectal mucosa, imaged from within the lumen [17–19]. The work presented in the current paper aims to detect cancer during colorectal surgery. During surgery tumor is approached from outside the lumen. To the best of our knowledge, there has only been one study focused on the use of HSI in colorectal cancer surgery in which the tumor is assessed from outside of the lumen and should be distinguished from fat and healthy colorectal wall [20]. In this work performed by our group tissue samples from colorectal cancer patients were obtained after resection and imaged by two hyperspectral cameras operating in the visual (400–1000 nm) and near-infrared (900–1700 nm) wavelength range. The mean accuracy to distinguish fat, healthy colorectal wall and tumor was 0.88 [20].

For fiberoptic DRS the main focus in colorectal cancer has also been on the application in colonoscopy for assessing the mucosa from within the colorectal lumen [21–25]. On the topic of application during colorectal cancer surgery a few papers have been published so far. In an ex vivo study by Schols *et al.* tumor could be discriminated from healthy surrounding tissue [7]. In an ex vivo study performed by our own group the accuracy to distinguish tumor tissue from fat and healthy colorectal wall was 0.95 [26].

An important step in developing a diagnostic algorithm based on spectral information is feature extraction, reducing the data used for classification, focusing on that part of the data that contains the crucial information. In previous studies, many different feature extraction and classification techniques have been used to analyze the spectra obtained with fiberoptic DRS or diffuse reflection HSI. In fiberoptic DRS for instance a fit algorithm based on photon diffusion theory has been used to extract absorption and scattering parameters from the measured spectra and to translate these into a limited set of parameters describing tissue composition and scattering behavior [27]. Based on these parameters tissue types could be distinguished [4,5,11,28]. Schols *et al.* used several amplitudes and gradients for classification [29]. In our group, spectral bands were used for both HSI and DRS data to reduce the number of features [20,26]. Classification of tissue types has been done in several ways. First of all, direct comparison of single features or simple combinations of features, has been performed to distinguish different tissues [4,30–34]. In addition, several machine learning techniques have been used for the tissue classification such as classification and regression tree algorithms [5,6,35–37], linear discriminant analysis (LDA) [24,38–40] and support vector machines (SVM) [3,20,26,37,38,41–43].

Previous studies used different analysis techniques to classify tissue during colorectal cancer surgery [6,20,26]. The current study will systematically examine several options to optimize the classification of healthy colorectal wall versus tumor tissue in colorectal cancer. Several different normalization and feature extraction techniques will be used in combination with a variety of classifiers to determine the optimal classification strategy.

2. Materials and methods

2.1. Data

In this study, datasets included were obtained using two different colorectal tissue sample sets. The measurements performed on tissue sample set 1 were only done ex vivo and were all performed using the same fiberoptic probe, which includes four fiber distances. The measurements performed on tissue sample set 2 were done both in vivo and ex vivo and were performed with disposable probes, so for each new patient a different probe with identical design was used. When measurements were performed using fiberoptic DRS, data of two fiber distances was acquired at the same time. There was only one diffuse reflection HSI dataset included which was obtained on tissue sample set 1 (Fig. 1). This dataset was a combination of data obtained with a visual hyperspectral camera and data obtained with a near-infrared hyperspectral camera. In total nine

datasets were included for the current analysis (Fig. 1). These datasets are separated based on the sample sets, ex vivo or in vivo, fiberoptic DRS or diffuse reflection HSI and for fiberoptic DRS the fiber distance used. To avoid confusion, in the rest of the paper the datasets will be referred to by the assigned number as shown in Fig. 1.

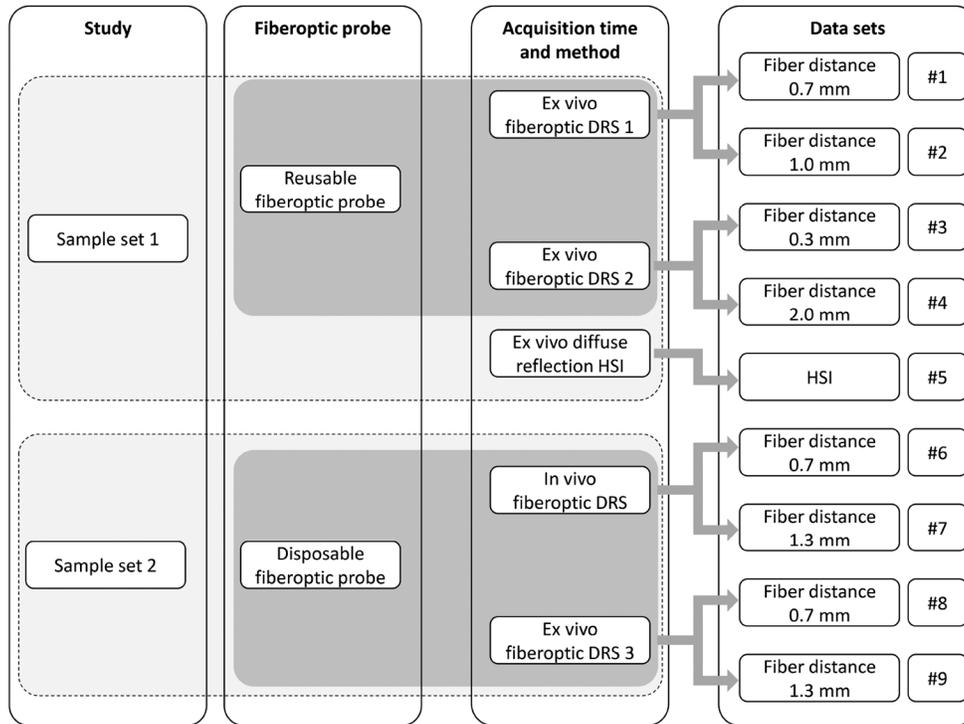


Fig. 1. Overview of all datasets used in this study, the first column shows the different sample sets, the second column what type of fiberoptic probe was used (reusable or disposable), the third column the difference in acquisition time (ex vivo/in vivo) and method (DRS/HSI) and the fourth shows different fiber distances and all datasets with assigned numbers. The light and dark gray areas indicate datasets obtained on the same sample set and datasets obtained with the same DRS probe type respectively.

In Table 1 an overview is given on the number of patients and the number of spectra present in all datasets, for dataset #5 the number of spectra is equal to the number of pixels. Classification in the current study was only performed on colon and tumor tissue because previous studies have shown that discriminating fat from other tissue types is an easy and already solved task [20,26]. For example, using fiberoptic DRS measurements on ex vivo samples (dataset #4) an accuracy of 1.00 ± 0.00 was obtained for the classification of fat [26]. When using diffuse reflection HSI (dataset #5) the same accuracy of 1.00 was achieved for fat classification [20]. Based on these results it was decided not to include fat measurements into the current analysis and to focus on the more challenging task of tumor/colon classification.

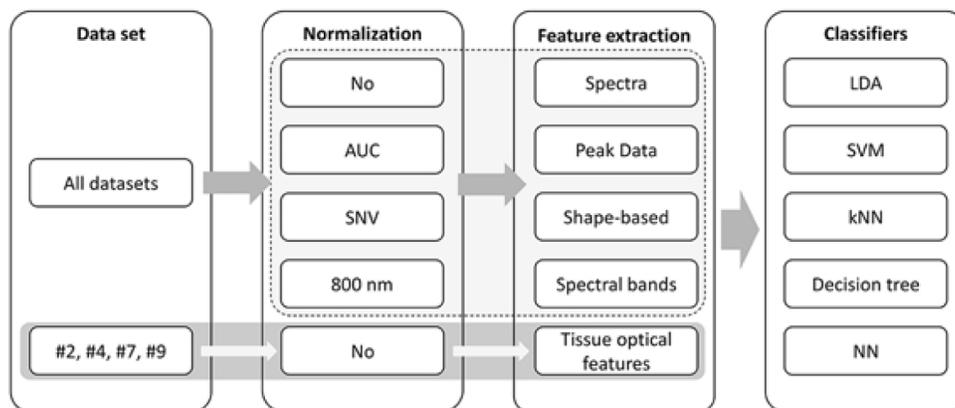
2.2. Preprocessing

Before analysing the spectra, all spectra were calibrated using a white reference, obtained on a Spectralon sample, and a dark reference. In the spectra obtained, intensity differences might have been present that were not due to tissue specific difference but due to for instance difference in the distance between the illumination source and the sample, compared to the distance between

Table 1. Overview of number of patients and number of spectra per dataset and per tissue type.

Dataset	Number of patients		Number of spectra	
	Over all	Per tissue type	Over all	Per tissue type (%)
#1 and #2	38	Colon: 32 Tumor: 29	948	Colon: 456 (48.1) Tumor: 492 (51.9)
#3 and #4	38	Colon: 32 Tumor: 29	943	Colon: 450 (47.7) Tumor: 493 (52.3)
#5	29	Colon: 22 Tumor: 27	1328	Colon: 569 (42.8) Tumor: 759 (57.2)
#6 and #7	28	Colon: 28 Tumor: 14	364	Colon: 261 (71.7) Tumor: 103 (28.3)
#8 and #9	28	Colon: 28 Tumor: 12	398	Colon: 321 (80.7) Tumor: 77 (19.3)

the illumination source and the Spectralon sample used for calibration. Calibration is performed to standardize all measurements. However, to standardize, the distance between the illumination source and the measured sample should be the same as the distance between the illumination source and the Spectralon sample. When the tissue thickness deviates over the sample, the distance between the illumination source and tissue will not be similar for the entire sample, causing intensity differences. Moreover, the coupling between the fibers and spectrometer might cause a variation in detection efficiency of the signal. To compensate for these intensity differences, spectra can be normalized before classification or feature extraction. In this study, three different normalization techniques were compared to each other and to no normalization. First, spectra were normalized by their intensity at 800 nm [32]. This wavelength was chosen because no absorption was assumed to be present at this wavelength. After normalization the intensity of all reflectance spectra should be equal to 1 at 800 nm. As a second option, the spectra were normalized using the area under the curve (AUC), setting all AUCs to 1. Finally, a standard normal variate (SNV) normalization was used [44,45]. Using SNV normalization, the mean of each spectrum was set to zero and the standard deviation was set to one. The SNV is often used for diffuse reflection HSI to exclude the influence of glare on the spectra [46,47]. However, SNV normalization also removes information on the scattering. As shown in Fig. 2, for the fiberoptic DRS and diffuse reflection HSI datasets all 4 different normalization techniques were evaluated.

**Fig. 2.** Overview of normalization, feature extraction and classification methods per dataset.

2.3. Feature extraction

When the number of features is very large relative to the number of samples in a dataset, some classifiers struggle to train effective models, which is called curse of dimensionality [48]. This is especially relevant for algorithms that rely on distance calculations like k NN [49]. Feature extraction and features selection techniques can be used to avoid the curse of dimensionality. Feature extraction creates a new, smaller set of features that still captures most of the useful information.

Besides using the entire spectra (1151 features) for classification, four feature extraction methods were evaluated. First, the seven most prominent peaks of a spectrum were selected, from these peaks the intensity values, the locations and the right side gradient of the peaks were used as features (Fig. 3(a)). In the second feature extraction technique, 23 features were selected manually to describe the coarse shape of the spectra while emphasizing the differences between different tissue classes. Features included were: the locations of maxima (4), locations of minima (2), gradients (9), prominence values (4), widths (2) and intensity value (1) (Fig. 3(b)). The prominence values were defined as the intrinsic height of a peak relative to neighboring dips. As a third feature set, spectral bands were selected using k -means clustering (Fig. 3(c)) [20,26]. The spectra were divided into at least 10 bands based upon the intensity values. Random start points were set after which the spectra were divided into 10 groups using k -means clustering based on the intensity values. If wavelengths within one group were not continuous, the group was divided into subgroups with only continuous wavelengths. In the final feature extraction method, tissue constituents were determined per spectrum using a fit algorithm, based on photon diffusion theory (Fig. 3(d)) [11,50,51]. With this algorithm, reflectance spectra of selected chromophores were fitted to the measured reflectance spectra to obtain the composition of the measured tissue in a number of parameters [27]. For these tissue types, 12 parameters were used for the fit, which were used as features for the classification. The fit algorithm is only valid for the fiberoptic geometry, and only the measurements in which the fiber distance between the receiving and collecting fiber was at least 1 mm [50]. Therefore, features based on fit results were only extracted for datasets #2, #4, #7 and #9, which are all fiberoptic DRS datasets with fiber distances of at least 1 mm. No normalization was used in case of tissue optical feature extraction, because the analytical model is based on the reflectance signal, which should not be normalized. In Fig. 2 an overview is given on which datasets are subjected to all possible combinations of feature extraction and normalization techniques.

2.4. Classifiers

In machine learning, the “No Free Lunch” theorem states that there is no one model that works best for every problem [52]. The assumptions of a great model for one problem may not hold for another problem, so it is common in machine learning to try multiple classifiers and choose the one that works best for that particular application. In this study five different supervised classification algorithms were used to distinguish colon from tumor, k -nearest neighbor (k NN), a decision tree classifier, LDA, SVM, and a neural network (NN). All classifiers were trained and tested using Matlab 2018a (MathWorks Inc., Natick, Massachusetts, US). For the k NN classifier, five neighbors were used to determine the class. The decision tree was a binary decision tree with “Gini’s diversity index” as a split criterion and the maximal number of branch node splits and minimum parent size were set 3 and 10, respectively. For the SVM a linear SVM was selected. For the NN classifier, a feedforward network was used with a single layer of 20 hidden neurons and a scaled conjugate gradient backpropagation training function. For all classifiers, the features were standardized to a zero mean and standard deviation of one per parameter before used in the classification. Standardization was determined based on the data from the training dataset and thereafter applied to the test dataset.

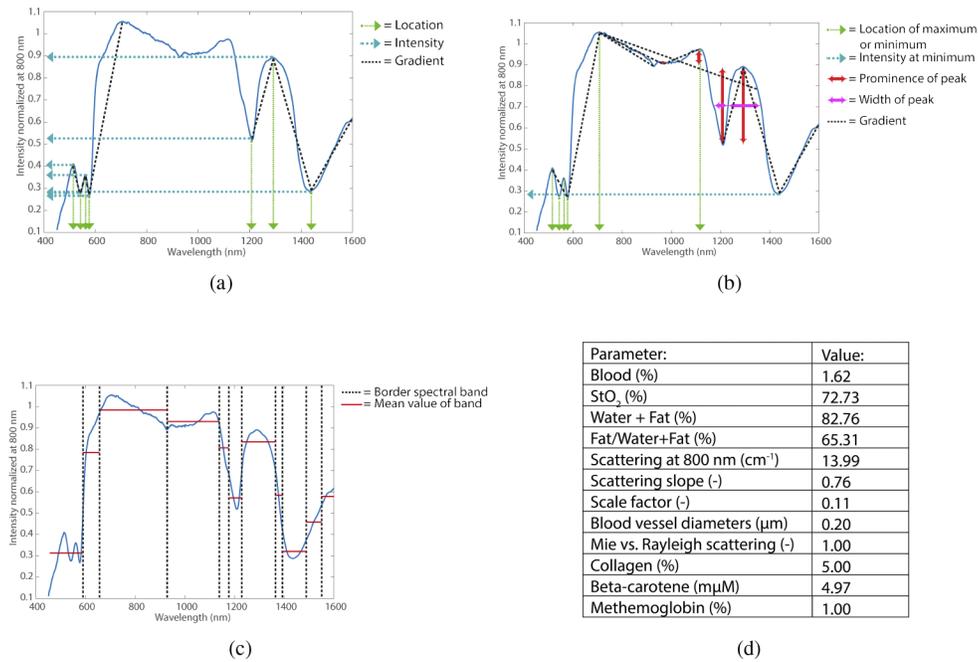


Fig. 3. Four different feature extraction methods shown for a spectrum obtained on fat tissue. (a) Wavelengths, intensities and gradients of seven most prominent peaks. (b) Wavelengths, intensities, prominences, widths and gradients describing the shape of the entire spectrum. (c) Spectral bands, the red bars indicate the mean intensity value per band. (d) Fitted parameters using the analytical reflectance model.

2.5. Data analysis

The sample sets were partitioned once into a training and test set, 80% of the patients was used to train and 20% to test the classifiers. The proportion of healthy versus tumor measurements is the same in the training and test dataset. The classifiers were evaluated using the accuracy, the area under the ROC curve (AUC) and the Matthews correlation coefficient (MCC). The accuracy and AUC are commonly used performance measures. However, these performance metrics do not take into account unbalanced classes in each dataset. Therefore, the MCC, which is less sensitive to unbalanced datasets, was added as a performance measure [53]. The MCC ranges from -1 to 1 , with 0 indicating a classification no better than random, -1 total disagreement in the classification and $+1$ a perfectly correct classification.

In supervised learning, overfitting happens when classifiers are trained on the noise along with the underlying pattern in the data. It happens when the classifier is trained on noisy or unrepresentative training data. Complex models like decision trees are more prone to overfitting. Overfitting can be recognized by examining the classifier variance, which is the variability of the model prediction for given data samples. Classifiers with high variance perform well on the train dataset but are not able to generalize to data it has not seen before e.g. the test dataset. Therefore, a large drop from training accuracy to test accuracy will be present in overfitting.

The classifications, done with all classifiers, on all different combinations shown in Fig. 2, were first tested for overfitting of the classifiers. This was done by subtracting the test accuracy from the training accuracy for each combination of normalization, feature extraction and classifier for each dataset. Since there is no general rule to define overfitting, we set the threshold for

overfitting by a decrease in accuracy from train to test of 0.15. If the decrease in accuracy was above 0.15, the classifier was defined to be overfitting during training.

Thereafter, the three normalization techniques were compared to the use of no normalization prior to classification. The accuracy and MCC values of the non-normalized data were considered as a baseline and were subtracted from the accuracy and MCC values obtained on the normalized data. In this analysis feature extraction methods were not included, so only entire spectra were used. The results were averaged per normalization technique and per dataset to assess if the normalization might have a different influence on different datasets.

A similar analysis was done for the feature extraction methods. These were all compared to the use of the entire spectrum during classification (baseline performance). Again, the difference in MCC and accuracy values compared to baseline were determined, for the selected classifiers and all datasets. However, this time only the selected normalization technique from the previous analysis was used to limit the amount of analyses. A mean difference in MCC and accuracy was obtained per feature extraction method and also the influence of feature extraction techniques for each dataset was obtained.

The combination of normalization and feature extraction was also examined using the difference in accuracy and MCC. The combination of all feature extraction methods and all normalization methods were compared to no normalization and the use of the entire spectrum.

Finally, for each dataset the combination of normalization, feature extraction and classifier with the maximum MCC and accuracy values were extracted from the data. These results were compared to the MCC and accuracy values of each dataset, obtained using the overall most optimal combination of normalization, feature extraction and classifier, as determined by all previous analysis.

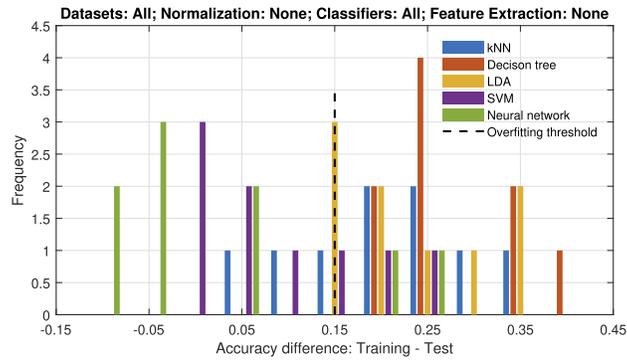
3. Results

3.1. Overfitting

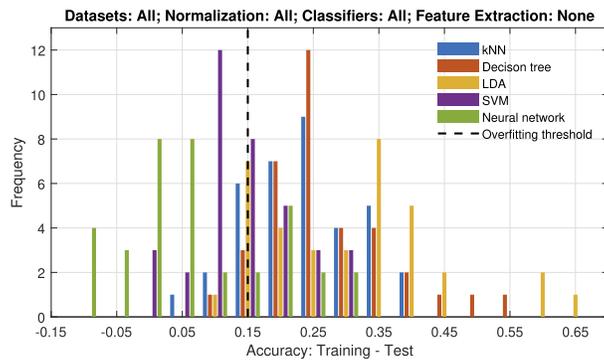
To evaluate the classifiers overfitting, the test accuracy of each classifier was subtracted from the train accuracy. A maximum decrease in accuracy of 0.15 from train to test was chosen as the threshold for overfitting. In the histograms of Fig. 4, the number of times (y-axis) a certain difference between train and test accuracy (x-axis) was obtained, is shown. The histograms in Fig. 4 are colored per classifier. In Fig. 4(a), the results using no normalization and no feature extraction method are shown. Here it clearly shows that the chances of overfitting were largely classifier dependent. For instance, for the NN classifier (green) most results were obtained below the overfitting threshold, the same accounts for SVM. Whereas for the decision tree (orange), all results of train minus test accuracy were above the threshold, indicating overfitting of the classifier, the same accounts for *k*NN and LDA. In Fig. 4(b) normalized data shown, this shows that there was no influence of the normalization on the overfitting of the classifiers, because the distributions for all classifiers were similar to Fig. 4(a). Figure 4(c) includes all feature extraction methods, but only on non-normalized data. As shown, feature extraction did show improvement in the overfitting for the *k*NN and LDA classifiers, as both distributions moved to the left of the graph (Fig. 4(c)) compared to Fig. 4(a). With the use of feature extraction, more train minus test accuracies stayed below the threshold of overfitting. Overall, SVM and NN classifiers showed the least overfitting on all datasets. Therefore, further analysis was focused only on SVM and NN.

3.2. Normalization

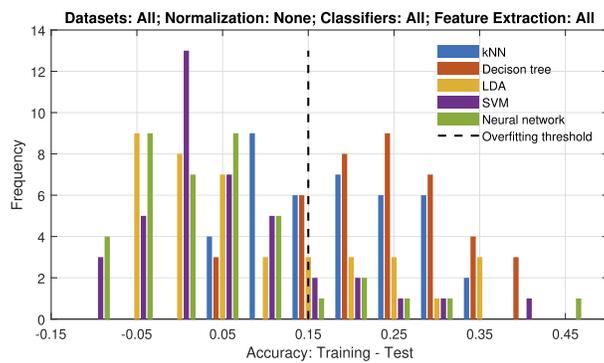
The added value of normalization on the classification outcome was examined. To this end, the increase in accuracy and increase in MCC for each normalization compared to no normalization were analyzed. This analysis was performed for the entire spectrum and only using the SVM and NN classifiers as selected in the previous section. The average increase in MCC and accuracy



(a)



(b)



(c)

Fig. 4. Histogram of the difference in accuracy between training and test and the number of times this difference is obtained. All classifiers are visualized separately. (a) Data is shown for all datasets, using no normalization and no feature extraction. (b) Data is shown for all datasets, using all normalizations and no feature extraction. (c) Data is shown for all datasets, using no normalization and all feature extractions.

values per normalization technique are shown in Table 2. In Table 2 it shows that there was no normalization technique that improved the outcome significantly for all datasets using the SVM or NN classifier. Similar results were obtained for the *k*NN, decision tree and LDA classifiers, which are not included in the Table 2.

Table 2. Mean increase in MCC and accuracy for all normalization techniques compared to no normalization for SVM and NN classifiers.

Normalization	Increase in MCC (mean \pm std)	Increase in ACC (mean \pm std)
SNV	0.05 \pm 0.20	-0.01 \pm 0.07
AUC	0.02 \pm 0.17	-0.01 \pm 0.07
Normalization at 800 nm	0.02 \pm 0.15	-0.01 \pm 0.08

MCC = Matthews Correlation Coefficient, ACC = Accuracy.

Because of the large standard deviations, the increase in MCC and accuracy was evaluated per dataset as well. An increase was seen for all datasets which were obtained using the disposable fiberoptic DRS probes (datasets #6 to #9) using SNV and normalization at 800 nm technique, whereas mainly a decrease was seen for the datasets obtained with the reusable fiberoptic DRS probe (datasets #1 to #4) (Table 3). Normalization using AUC shows a decrease in MCC and accuracy for datasets #8 and #9 as well. For the dataset of the diffuse reflection HSI (dataset #5) normalization did not seem to affect the results. Because, SNV normalization and normalization at 800 nm seemed only useful for some datasets and AUC normalization did not seem useful at all, it was decided to continue without using normalization for the comparison of different feature extraction technique in the next section.

Table 3. Mean change per dataset in MCC value and accuracy, when normalization technique is compared to no normalization. Values of the standard deviation are not shown in the table, but none of the change values differed significantly from zero.

	SNV normalization		AUC normalization		normalization at 800 nm	
	MCC	ACC	MCC	ACC	MCC	ACC
#1	-0.04	-0.02	0.01	0.00	-0.03	-0.01
#2	0.04	0.02	-0.02	-0.01	-0.03	-0.01
#3	-0.17	-0.08	-0.09	-0.04	-0.06	-0.03
#4	-0.03	-0.01	-0.06	-0.03	0.01	0.00
#5	-0.04	-0.02	0.00	0.00	0.00	0.00
#6	0.14	-0.04	0.11	-0.03	0.07	0.06
#7	0.20	0.03	0.30	0.09	0.17	0.07
#8	0.08	-0.05	-0.04	-0.01	0.03	-0.07
#9	0.22	0.04	-0.07	-0.04	0.01	-0.04

MCC = Matthews Correlation Coefficient, ACC = Accuracy.

3.3. Feature extraction

Based on the results from the normalization, the first analysis on feature extraction was performed on non-normalized data. In Table 4 the results are shown for the mean increase in MCC and accuracy for all feature extraction methods compared to no feature extraction, e.g. the entire spectrum. Here, all datasets were used, no normalization was used and only the results from the SVM and NN classifiers were taken into account. From Table 4, it can be concluded that overall using tissue optical features improved classification results compared to the use of the entire spectrum. It should be noted that tissue optical features were only available for datasets

for the fiber optic DRS data with fiber distance more than 1 mm (e.g. datasets #2, #4, #7 and #9). For the shape-based features and spectral bands almost no difference compared to using the entire spectrum for classification was seen. Peak data showed a drop in both MCC and accuracy compared to using the entire spectrum.

Table 4. Mean increase in MCC and accuracy for all feature extraction methods compared to the use of the entire spectrum. Only not normalized data is used and only SVM and NN classifiers.

Feature extraction	Increase in MCC (mean \pm std)	Increase in ACC (mean \pm std)
Peak data	-0.15 ± 0.22	-0.09 ± 0.10
Shape-based features	0.03 ± 0.17	0.01 ± 0.07
Spectral bands	-0.05 ± 0.18	-0.01 ± 0.07
Tissue optical features	0.25 ± 0.28	0.09 ± 0.11

MCC = Matthews Correlation Coefficient, ACC = Accuracy.

Because of the large standard deviation, the increase in MCC and accuracy, the analysis was also performed per dataset (Table 5). A large increase in MCC values was shown for the use of tissue optical features in the datasets of the disposable fiberoptic DRS probes both in vivo and ex vivo (datasets #7 and #9). If the tissue optical features were not taken into account, the shape-based features show the best overall results closely followed by the spectral bands. The use of peak data did not show improvement in any of the datasets.

Table 5. Mean increase in MCC and accuracy per dataset, for all feature extraction methods compared to the use of the entire spectrum. Only non-t normalized data is used and only SVM and NN classifiers.

	Peak data		Shape-based		Spectral bands		Tissue optical features	
	MCC	ACC	MCC	ACC	MCC	ACC	MCC	ACC
#1	-0.27	-0.14	0.04	0.02	0.06	0.03	-	-
#2	-0.10	-0.05	0.04	0.01	0.06	0.03	0.04	0.02
#3	-0.07	-0.05	0.01	0.00	0.03	0.01	-	-
#4	-0.03	-0.01	0.03	0.02	0.03	0.02	0	0
#5	-0.27	-0.13	0.01	0.01	-0.24	-0.11	-	-
#6	-0.29	-0.23	-0.03	0.01	-0.15	-0.03	-	-
#7	0.28	0.02	0.38	0.07	0.16	0.05	0.64	0.24
#8	-0.33	-0.06	-0.18	-0.07	-0.16	-0.03	-	-
#9	-0.18	-0.17	0.05	-0.02	-0.09	0.02	0.30	0.08

MCC = Matthews Correlation Coefficient, ACC = Accuracy.

As shown in section 3.1, since feature extraction decreased the overfitting problem for *k*NN and LDA classifiers, the results on feature extraction were also calculated for all classifiers. For this analysis, slightly better results were obtained overall. The MCC and accuracy for the tissue optical features showed a large increase (0.29 and 0.09 respectively) and the shape-based features showed a smaller increase in MCC and accuracy (0.13 and 0.05 respectively). The results for the peak data and spectral bands improved with almost no decrease for the peak data (MCC = -0.06 and accuracy = -0.05) and a small increase in MCC and accuracy for the spectral bands (0.03 and 0.02 respectively). The analysis per dataset showed similar results. In short, if possible, taking into account the fiber distance and acquisition method (fiberoptic DRS or diffuse reflection HSI), tissue optical features were the best option for feature extraction. If tissue optical features could not be obtained, shape-based features were the best option.

3.4. Normalization combined with feature extraction

So far, normalization and feature extraction were only examined separately. In Table 6 the increase in MCC and accuracy values is shown for each combination of normalization and feature extraction compared to no feature extraction and no normalization. In the analysis of Table 6 all classifiers were included because using feature extraction methods decreased the risk of overfitting.

Table 6. Mean increase in MCC and accuracy over all datasets for each combination of feature extraction and normalization method compared to no normalization and no feature extraction. All classifiers are included

	Increase in MCC (mean \pm std)			Increase in accuracy (mean \pm std)		
	Peak data	Shape-based	Spectral bands	Peak data	Shape-based	Spectral bands
No	-0.04 ± 0.25	0.13 ± 0.24	0.04 ± 0.21	-0.05 ± 0.11	0.05 ± 0.09	0.03 ± 0.09
SNV	-0.07 ± 0.26	0.13 ± 0.22	0.10 ± 0.22	-0.05 ± 0.11	0.04 ± 0.09	0.01 ± 0.11
AUC	-0.07 ± 0.24	0.15 ± 0.22	0.08 ± 0.20	-0.05 ± 0.10	0.04 ± 0.09	0.01 ± 0.09
At 800nm	-0.06 ± 0.25	0.16 ± 0.23	0.07 ± 0.23	-0.05 ± 0.12	0.06 ± 0.09	0.01 ± 0.10

MCC = Matthews Correlation Coefficient, ACC = Accuracy.

From Table 6, it can be concluded that independent of the normalization technique, shape-based feature extraction showed the best results. Furthermore, SNV normalization showed the worst results for peak data and shape-based feature extraction, but the best results for the spectral bands feature extraction. In Table 6, tissue optical features are not included because no normalization was applied for this feature extraction method.

In Fig. 5, Bland-Altman plots are shown, for all datasets of which tissue optical features could be obtained and all classifiers. Separate plots were made for the combination of shape-based feature extraction with normalization at 800 nm (left), spectral band feature extraction and normalization at 800 nm (center), and the tissue optical features (right). In the Bland-Altman plot, the difference in MCC value between the combination of feature extraction and normalization, and no feature extraction and no normalization is shown. The center line indicates the mean value of all values shown in the plot and the dotted lines indicate the confidence intervals. Fig. 5 shows that the results from the combination of shape-based feature extraction and normalization at 800 actually approached the results from the tissue optical features if only the datasets from which tissue optical features could be obtained were included.

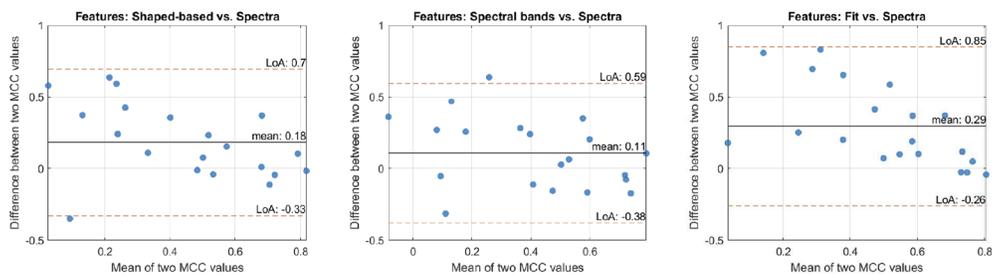


Fig. 5. Bland-Altman plots of only the datasets of which fit-data could be obtained. From left to right the combination of shape-based feature extraction and normalization at 800 nm, spectral bands and normalization at 800 nm, and tissue optical features are shown. The center line indicated the average difference between the combination of feature extraction and normalization with using no normalization and no feature extraction. The average value is shown on top of the center line. The dotted lines indicate the confidence intervals.

3.5. Overall results per dataset

Based on the previous analysis on overfitting, normalization and feature extraction the best combination, obtained over all datasets, consisted of tissue optical features and an SVM or NN classifier. In case of small fiber distances or the use of the hyperspectral camera, tissue optical features were not available in some datasets. Therefore, the shape-based features, with normalization at 800 nm were defined as most optimal for datasets #1, #3, #5, #6 and #8.

For each dataset separately, the best combination of the normalization, feature extraction and classifier was evaluated based on the MCC and accuracy values. For all datasets, the best combination selected based on MCC or accuracy value were the same. In Table 7 an overview is given per dataset with the maximum MCC value, maximum accuracy and the normalization, feature extraction and classifier used to obtain these values.

Table 7. Overview of best performing combinations of normalization, feature extraction and classifier per dataset.

Data set	MCC	ACC	AUC	Normalization	Feature extraction	Classifier
#1	0.88	0.94	0.96	No	Spectral bands	SVM
#2	0.84	0.92	0.98	No	Spectral bands	SVM
#3	0.84	0.92	0.95	At 800 nm	Spectral bands	LDA
#4	0.89	0.95	0.97	No	Spectral bands	SVM
#5	0.86	0.93	0.96	SNV	Spectral bands	SVM
#6	0.43	0.84	0.77	At 800 nm	Shape-based features	NN
#7	0.81	0.93	0.92	No	Tissue optical features	SVM
#8	0.68	0.93	0.73	SNV	Spectral bands	LDA
	0.68	0.93	0.81	SNV	Spectral bands	SVM
	0.68	0.93	0.94	No	Tissue optical features	LDA
#9	0.68	0.93	0.84	No	Tissue optical features	NN
	0.68	0.93	0.90	SNV	Spectrum	NN

MCC = Matthews Correlation Coefficient, ACC = Accuracy, AUC = Area Under ROC Curve.

For both datasets using disposable fiberoptic DRS needles used ex vivo (dataset #8 and #9), multiple combinations resulted in similar maximum MCC and accuracy values. For dataset #8 (short distance), two combinations with similar accuracy and MCC value were found, based on the AUC the classification based on SVM was best, with an AUC of 0.81 for SVM compared to an AUC of 0.73 for the LDA classifier. For dataset #9 (disposable fiberoptic DRS probe used ex vivo, long distance) three combinations of normalization, feature extraction and classifier provided the same maximum accuracies and MCC values, as shown in Table 7. Based on the AUC values obtained using these combinations, the combination of no normalization, tissue optical features and an LDA classifier, performed slightly better than the SNV normalized spectrum in a neural network and better than the combination of no normalization, tissue optical features in a neural network (AUC of 0.94, 0.90 and 0.84 respectively).

The results found in the previous section were mostly supported by the results from Table 7. Most classifiers found in the analysis of Table 7 were either SVM or NN classifiers. The two LDA classifiers that were selected both used some kind of feature reduction. For dataset #8 and #9, as shown in Table 7, the results from LDA classification were similar to the SVM and NN classifications, respectively. The same accounts for the results of dataset #3, for which classification using shape-based features normalized with the AUC gives an MCC value of 0.81 and an accuracy of 0.90 when the SVM classifier was used.

The normalizations found in Table 7, were also in correspondence with previous findings in which normalization did not seem to improve the outcome. As found before datasets #6 to #9,

the datasets obtained using the disposable fiberoptic probes, might benefit from normalization. This can be seen in Table 7 as well, where normalization for datasets #6 to #9 was applied in all cases where no tissue optical features were used.

For datasets #6 to #9, tissue optical features were selected as feature extraction technique if possible (Table 7), just as expected from the previous analysis. In all datasets, feature selection methods provided the best MCC values and accuracies. As predicted by the analysis on feature extraction, none of the optimal combinations used peak data as a feature extraction method.

In Table 8 the results of the of the optimal solution per dataset are compared to the global optimum, using SNV normalization, spectral bands as feature extraction and an SVM classifier. For the first five datasets the results of the global optimum were only slightly worse compared to the optimum of each dataset. For the final four datasets (#6-#9) a larger difference was seen.

Table 8. Results of using the global optimum combination of normalization, feature extraction and classifier (SNV+Spectral Band+SVM) compared to best-performing combinations per dataset (based on Table 7).

Dataset	Global optimum (SNV+spectral band+SVM)			Dataset optimum (based on Table 7)		
	MCC	ACC	AUC	MCC	ACC	AUC
#1	0.74	0.86	0.92	0.88	0.94	0.96
#2	0.78	0.89	0.97	0.84	0.92	0.98
#3	0.72	0.86	0.94	0.84	0.92	0.95
#4	0.85	0.92	0.96	0.89	0.95	0.97
#5	0.86	0.93	0.96	0.86	0.93	0.96
#6	0.13	0.57	0.60	0.43	0.84	0.77
#7	0.52	0.72	0.81	0.81	0.93	0.92
#8	0.68	0.93	0.81	0.68	0.93	0.81
#9	0.52	0.89	0.88	0.68	0.93	0.94

To confirm the results obtained with the holdout test, performed in this research, the test results of the holdout test were compared to a 20-fold cross-validation. In Table 9 the results of this comparison are shown. The results are shown for the combination of SNV normalization, spectral band feature extraction and an SVM classifier. As can be seen in Table 9 only minor differences were shown between the holdout test results and the 20-fold cross-validation.

Table 9. Comparison between the performance of cross-validation and holdout validation.

Dataset	Holdout (20%-80%)			20-fold cross-validation		
	MCC	ACC	AUC	MCC	ACC	AUC
#1	0.74	0.87	0.92	0.75	0.88	0.93
#2	0.78	0.89	0.97	0.75	0.88	0.95
#3	0.72	0.86	0.93	0.63	0.82	0.86
#4	0.84	0.92	0.96	0.81	0.91	0.95
#5	0.85	0.93	0.96	0.78	0.89	0.95

4. Discussion

In the current study, an optimal analysis technique was examined to distinguish healthy colorectal wall from tumor tissue. The analysis was done on nine different datasets, using five different classifiers, three normalization techniques and four feature extraction methods, compared to the use of the entire spectrum without normalization. Based on these analyses it was found that the

best overall combination is to use no normalization, an SVM or NN as a classifier and as feature extraction method the tissue optical features or the shape-based features.

4.1. Overfitting

The results on overfitting show a large chance of overfitting for k NN, decision tree and LDA classifiers if the entire spectrum was used. This was expected because the number of features used for the classification is large compared to the number of samples for each dataset. As a rule of thumb the number of slightly correlated features should not exceed the square root of the number of samples [54]. For the use of the entire spectrum, this rule of thumb is not met which increases the chance of overfitting. It was shown that the decrease in the number of features used for the classification resulted in a decrease in overfitting (Fig. 4(c)). No improvement in overfitting was seen when datasets were normalized (Fig. 4(b)). This was expected because normalization does not change the number of features, neither does it decrease the presence of noise in the data. For the SVM and NN classifiers overall overfitting was less present. Therefore, it was decided to continue with these two classifiers for further analysis.

4.2. Normalization

Overall, normalization does not seem to improve classification outcomes (Table 2). However, if datasets are evaluated separately a trend is seen when datasets obtained with the reusable fiberoptic DRS probe (#1 to #4) are compared to datasets obtained with the disposable fiberoptic DRS probe (#6 to #9). For the first four datasets, normalization does not seem to improve the results. On the contrary, for the later four normalization does improve the results. This might have to do with the use of disposable needles in the second group (datasets #6 to #9). Even though manufacturing of these probes was automated, small differences in fiber distances were present over the set of probes. Furthermore, for the white reference of the final four datasets a separate probe was used, which was not the case for the first four datasets. The use of a new probe for each patient and the use of a separate white reference probe, which required changing the fibers between calibration and measurement, might have caused intensity differences in the final four datasets which were not tissue specific. The intensity differences caused by both issues, will be eliminated by normalization of the spectra. However, normalization might also eliminate tissue specific intensity differences. This is probably the reason why normalization is not improving the results for the first four datasets in which the measurement set-up is more constant.

For diffuse reflection HSI, SNV normalization is often used because it removes intensity differences due to glare [47]. However, as shown in Table 3, SNV normalization does not improve the results of the diffuse reflection HSI data, neither did the other two normalization techniques. Apparently, glare does not affect the features vital to the classification. Even though in general SNV does not seem to improve the analysis of diffuse reflection HSI significantly, the combination resulting in the best accuracy and MCC for the diffuse reflection HSI dataset did use SNV normalization (Table 7).

4.3. Feature extraction

Because of the chances of overfitting when using the entire spectra for classification different feature extraction methods were examined. The use of information from the seven most prominent peaks of the spectra does not give any useful information. As shown in Table 4 and Table 5 the MCC and accuracy values decrease when using the peak data compared to the use of the entire spectrum. An advantage of using peak data compared to the entire spectrum is the fact that less features are used in the classification, 21 features for the peak data compared to 1151 features for the entire spectrum. Even though the use of peak data is not the most optimal feature extraction method, the decrease in the number of features used, compared to the use of the entire spectrum, does show an advantage.

The best feature extraction method is the use of tissue optical features. However, the use of tissue optical features has some restrictions compared to the other feature extraction methods. First of all, the outcome of the fit algorithm depends largely on the tissue constituents that are given as input to the fit algorithm. This means, that if not all tissue constituents are known, the results will be unreliable. Furthermore, several assumptions are made for the fit model. The model which has been used in this paper is designed only for fiber optics with fiber distances larger than 1 mm. This algorithm does not work properly on hyperspectral data as well as on the spectra from fibers with less than 2 mm distance. To best of our knowledge there is no generalized fit algorithm that could work properly for all datasets. The use of shape-based features seems the second best option, which can be used for all datasets. The results of the shape-based feature could be improved by adding more features to better describe the shape of the spectra and more importantly describe the difference between the two tissue types.

4.4. *General optimum*

As stated before, there is not one best model that works for every problem. However, in Table 8 it is shown that if a global optimum, the SNV normalization, spectral bands and an SVM classifier, was compared to the optimum of each dataset, only small differences were seen for the first five datasets. Datasets #6-#9 did show large differences between the optimum of each dataset and the global optimum. This can be explained by the use of the disposable needles, which has less influence on the extraction of the tissue optical features compared to the other feature extraction methods. Furthermore, when the conclusions of the separate analyses are compared to the optimal solution for each dataset (Table 7), there is a large resemblance. For instance, in the final four datasets, normalization is used for all datasets in which no tissue optical features was used as feature extraction method, whereas for the other fiberoptic DRS dataset no normalization was used except for dataset #3 (Table 7). Furthermore, for all datasets feature extraction was used. Overall, in almost all datasets the optimal classifier is either a SVM or a NN. So even though there is no one best model, some rules could be set to optimize the classification of DRS and HSI data in colorectal cancer. However, generalization of these rules to other tissue types or optical technologies cannot be done without examination first.

4.5. *Future work*

The current research did only cover 6 basic classifiers. Future research should also include more state of the art classifiers like a random forest classifier and a convolutional NN. Furthermore, the option of boosting should be investigated to further optimize the classification. For the hyperspectral images, classification could be further optimized by taking into account the spatial information. However, with this also the dependency of neighboring pixels should be considered. In the current hyperspectral dataset the dependency of the neighboring pixels was not taken into account. Furthermore, to overcome the possibility of the influence of multiple tissue types on one pixel, pixels at the border between two tissue types were removed from the dataset.

5. **Conclusion**

From the present study it can be concluded that normalization does not seem to have added value, except to remove differences originating from the use of different DRS probes within one dataset. Furthermore, in a clinical setting where the number of samples is often limited, the use of feature extraction methods is needed to prevent overfitting in high dimensional data classification. On the other hand, the results of feature extraction could be helpful to design simpler devices focused on measuring only the features we are interested in rather than the full spectrum, like multispectral cameras. Finally, SVM and NN are the best performing classifiers with the least chance of overfitting on these datasets.

Funding

KWF Kankerbestrijding (NKI 2014-6596); Nederlandse Organisatie voor Wetenschappelijk Onderzoek (15228).

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer J. for Clin.* **68**, 394–424 (2018).
2. K. D. Miller, R. L. Siegel, C. C. Lin, A. B. Mariotto, J. L. Kramer, J. H. Rowland, K. D. Stein, R. Alteri, and A. Jemal, "Cancer treatment and survivorship statistics, 2016," *CA: A Cancer J. for Clin.* **66**, 271–289 (2016).
3. S. G. B. de Koning, E. J. M. Baltussen, M. B. Karakullukcu, B. Dashtbozorg, L. A. Smit, R. Dirven, B. H. W. Hendriks, H. J. C. M. Sterenberg, and T. J. M. Ruers, "Toward complete oral cavity cancer resection using a handheld diffuse reflectance spectroscopy probe," *J. Biomed. Opt.* **23**(12), 1–9 (2018).
4. L. L. Boer, B. G. Molenkamp, T. M. Bydlon, B. H. W. Hendriks, J. Wesseling, H. J. C. M. Sterenberg, and T. J. M. Ruers, "Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries," *Breast Cancer Res. Treat.* **152**(3), 509–518 (2015).
5. J. W. Spliethoff, D. J. Evers, H. M. Klomp, J. W. van Sandick, M. W. Wouters, R. Nachabe, G. W. Lucassen, B. H. W. Hendriks, J. Wesseling, and T. J. M. Ruers, "Improved identification of peripheral lung tumors by using diffuse reflectance and fluorescence spectroscopy," *Lung Cancer* **80**(2), 165–171 (2013).
6. G. C. Langhout, J. W. Spliethoff, S. J. Schmitz, A. G. J. Aalbers, M. L. F. van Velthuysen, B. H. W. Hendriks, T. J. M. Ruers, and K. F. D. Kuhlmann, "Differentiation of healthy and malignant tissue in colon cancer patients using optical spectroscopy: A tool for image-guided surgery," *Lasers Surg. Med.* **47**(7), 559–565 (2015).
7. R. M. Schols, P. Dumias, F. P. Wieringa, and L. P. S. Stassen, "Multispectral characterization of tissues encountered during laparoscopic colorectal surgery," *Med. Eng. Phys.* **35**(7), 1044–1050 (2013).
8. 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).
9. B. Fei, G. Lu, X. Wang, H. Zhang, J. V. Little, M. R. Patel, C. C. Griffith, M. W. El-Diery, and A. Y. Chen, "Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients," *J. Biomed. Opt.* **22**(08), 1 (2017).
10. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *J. Biomed. Opt.* **19**(1), 010901 (2014).
11. R. Nachabé, B. H. W. Hendriks, M. van der Voort, A. E. Desjardins, and H. J. C. M. Sterenberg, "Estimation of biological chromophores using diffuse optical spectroscopy: benefit of extending the UV-VIS wavelength range to include 1000 to 1600 nm," *Biomed. Opt. Express* **1**(5), 1432–1442 (2010).
12. K. Masood, N. Rajpoot, and M. Nasir, "Classification of colon biopsy samples by spatial analysis of a single spectral band from its hyperspectral cube," *Proc. Med. Image Underst. Analysis (MIUA)* pp. 42–48 (2007).
13. K. Masood, N. Rajpoot, and M. Nasir, "Spatial analysis for colon biopsy classification from hyperspectral imagery," *Ann. BMVA* **4**, 1–16 (2008).
14. K. Rajpoot and N. Rajpoot, "SVM Optimization for Hyperspectral Colon Tissue Cell Classification," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, C. Barillot, D. R. Haynor, and P. Hellier, eds. (Springer Berlin Heidelberg, 2004), pp. 829–837.
15. S. Rathore, M. Hussain, A. Ali, and A. Khan, "A Recent Survey on Colon Cancer Detection Techniques," *IEEE/ACM Transactions on Comput. Biol. Bioinforma.* **10**(3), 545–563 (2013).
16. M. Maggioni, G. L. Davis, F. J. Warner, F. B. Geshwind, A. C. Coppi, R. A. DeVerse, and R. R. Coifman, "Hyperspectral microscopic analysis of normal, benign and carcinoma microarray tissue sections," in *Optical Biopsy VI*, vol. 6091 (International Society for Optics and Photonics, 2006)609101.
17. E. Claridge and D. Hidović-Rowe, "Model Based Inversion for Deriving Maps of Histological Parameters Characteristic of Cancer From Ex-Vivo Multispectral Images of the Colon," *IEEE Transactions on Med. Imaging* **33**(4), 822–835 (2014).
18. Z. Han, A. Zhang, X. Wang, Z. Sun, M. D. Wang, and T. Xie, "In vivo use of hyperspectral imaging to develop a noncontact endoscopic diagnosis support system for malignant colorectal tumors," *J. Biomed. Opt.* **21**(1), 016001 (2016).
19. R. Kumashiro, K. Konishi, T. Chiba, T. Akahoshi, S. Nakamura, M. Murata, M. Tomikawa, T. Matsumoto, Y. Maehara, and M. Hashizume, "Integrated Endoscopic System Based on Optical Imaging and Hyperspectral Data Analysis for Colorectal Cancer Detection," *Anticancer Res.* **36**, 3925–3932 (2016).

20. E. J. M. Baltussen, E. N. D. Kok, S. G. B. de Koning, J. Sanders, A. G. J. Aalbers, N. F. M. Kok, G. L. Beets, C. C. Flohil, S. C. Bruin, K. F. D. Kuhlmann, H. J. C. M. Sterenberg, and T. J. M. Ruers, "Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery," *J. Biomed. Opt.* **24**(01), 1–9 (2019).
21. G. Zonios, L. T. Perelman, V. Backman, R. Manoharan, M. Fitzmaurice, J. Van Dam, and M. S. Feld, "Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo," *Appl. Opt.* **38**(31), 6628–6637 (1999).
22. H.-W. Wang, J.-K. Jiang, C.-H. Lin, J.-K. Lin, G.-J. Huang, and J.-S. Yu, "Diffuse reflectance spectroscopy detects increased hemoglobin concentration and decreased oxygenation during colon carcinogenesis from normal to malignant tumors," *Opt. Express* **17**(4), 2805–2817 (2009).
23. H. K. Roy, A. Gomes, V. Turzhitsky, M. J. Goldberg, J. Rogers, S. Ruderman, K. L. Young, A. Kromine, R. E. Brand, M. Jameel, P. Vakil, N. Hasabou, and V. Backman, "Spectroscopic Microvascular Blood Detection From the Endoscopically Normal Colonic Mucosa: Biomarker for Neoplasia Risk," *Gastroenterology* **135**(4), 1069–1078 (2008).
24. A. Dhar, K. S. Johnson, M. R. Novelli, S. G. Bown, I. J. Bigio, L. B. Lovat, and S. L. Bloom, "Elastic scattering spectroscopy for the diagnosis of colonic lesions: initial results of a novel optical biopsy technique," *Gastrointest. Endoscopy* **63**(2), 257–261 (2006).
25. J. R. Mourant, I. J. Bigio, J. D. Boyer, T. M. Johnson, J. Lacey, A. Bohorfoush, and M. H. Mellow, "Elastic scattering spectroscopy as a diagnostic tool for differentiating pathologies in the gastrointestinal tract: Preliminary testing," *J. Biomed. Opt.* **1**(2), 192–199 (1996).
26. E. J. Baltussen, P. Snæbjörnsson, S. G. B. De Koning, H. J. Sterenberg, A. G. Aalbers, N. Kok, G. L. Beets, B. H. Hendriks, K. F. Kuhlmann, and T. J. Ruers, "Diffuse reflectance spectroscopy as a tool for real-time tissue assessment during colorectal cancer surgery," *J. Biomed. Opt.* **22**(10), 1 (2017).
27. T. M. Bydlon, R. Nachabé, N. Ramanujam, H. J. Sterenberg, and B. H. Hendriks, "Chromophore based analyses of steady-state diffuse reflectance spectroscopy: current status and perspectives for clinical adoption," *J. Biophotonics* **8**(1-2), 9–24 (2015).
28. R. Nachabé, B. H. W. Hendriks, A. E. Desjardins, M. van der Voort, M. B. van der Mark, and H. J. C. M. Sterenberg, "Estimation of lipid and water concentrations in scattering media with diffuse optical spectroscopy from 900 to 1,600 nm," *J. Biomed. Opt.* **15**(3), 037015 (2010).
29. R. M. Schols, M. ter Laan, L. P. S. Stassen, N. D. Bouvy, A. Amelink, F. P. Wieringa, and L. Alic, "Differentiation between nerve and adipose tissue using wide-band (350-1,830 nm) in vivo diffuse reflectance spectroscopy," *Lasers Surg. Med.* **46**(7), 538–545 (2014).
30. J. W. Spliethoff, W. Prevoo, M. A. J. Meier, J. de Jong, D. J. Evers, H. J. C. M. Sterenberg, G. W. Lucassen, B. H. W. Hendriks, T. J. M. Ruers, H. M. Klomp, D. J. Evers, H. J. C. M. Sterenberg, G. W. Lucassen, B. H. W. Hendriks, and T. J. M. Ruers, "Real-time in vivo tissue characterization with diffuse reflectance spectroscopy during transthoracic lung biopsy: a clinical feasibility study," *Clin. Cancer Res.* **22**(2), 357–365 (2016).
31. J. W. Spliethoff, L. L. de Boer, M. A. J. Meier, W. Prevoo, J. de Jong, K. Kuhlmann, T. M. Bydlon, H. J. C. M. Sterenberg, B. H. W. Hendriks, and T. J. M. Ruers, "In vivo characterization of colorectal metastases in human liver using diffuse reflectance spectroscopy: toward guidance in oncological procedures," *J. Biomed. Opt.* **21**(9), 097004 (2016).
32. E. Tanis, J. W. Spliethoff, D. J. Evers, G. C. Langhout, P. Snæbjörnsson, W. Prevoo, B. H. W. Hendriks, and T. J. M. Ruers, "Real-time in vivo assessment of radiofrequency ablation of human colorectal liver metastases using diffuse reflectance spectroscopy," *Eur. J. Surg. Oncol. (EJSO)* **42**(2), 251–259 (2016).
33. L. L. de Boer, B. H. W. Hendriks, F. Van Duijnhoven, M.-J. T. F. D. V. Peeters-Baas, K. Van de Vijver, C. E. Loo, K. Józwiak, H. J. C. M. Sterenberg, and T. J. M. Ruers, "Using DRS during breast conserving surgery: identifying robust optical parameters and influence of inter-patient variation," *Biomed. Opt. Express* **7**(12), 5188–5200 (2016).
34. R. Nachabé, B. H. W. Hendriks, R. Schierling, J. Hales, J. M. Racadio, S. Rottenberg, T. J. M. Ruers, D. Babic, and J. M. Racadio, "Real-time in vivo characterization of primary liver tumors with diffuse optical spectroscopy during percutaneous needle interventions: feasibility study in woodchucks," *Investig. radiology* **50**(7), 443–448 (2015).
35. R. Nachabé, B. H. W. Hendriks, G. W. Lucassen, M. van der Voort, D. J. Evers, E. J. Rutgers, M.-J. V. Peeters, J. A. Van der Hage, H. S. Oldenburg, and T. J. 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).
36. D. J. Evers, R. Nachabé, D. Hompes, F. Van Coevorden, G. W. Lucassen, B. H. W. Hendriks, M. L. F. Van Velthuisen, J. Wesseling, and T. J. M. Ruers, "Optical sensing for tumor detection in the liver," *Eur. J. Surg. Oncol. (EJSO)* **39**(1), 68–75 (2013).
37. B. H. W. Hendriks, A. J. R. Balthasar, G. W. Lucassen, M. van der Voort, M. Mueller, V. V. Pully, T. M. Bydlon, C. Reich, A. T. M. H. van Keersop, J. Kortsmid, G. C. Langhout, and G.-J. van Geffen, "Nerve detection with optical spectroscopy for regional anesthesia procedures," *J. Transl. Med.* **13**(1), 380 (2015).
38. G. Lu, J. V. Little, X. Wang, H. Zhang, M. R. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Detection of Head and Neck Cancer in Surgical Specimens Using Quantitative Hyperspectral Imaging," *Clin. Cancer Res.* **23**(18), 5426–5436 (2017).
39. M. P. L. Bard, A. Amelink, M. Skurichina, M. den Bakker, S. A. Burgers, J. P. van Meerbeeck, R. P. W. Duin, J. G. J. V. Aerts, H. C. Hoogsteden, and H. J. C. M. Sterenberg, "Improving the specificity of fluorescence bronchoscopy for the analysis of neoplastic lesions of the bronchial tree by combination with optical spectroscopy: preliminary communication," *Lung Cancer* **47**(1), 41–47 (2005).

40. D. J. Evers, R. Nachabé, H. M. Klomp, J. W. van Sandick, M. W. Wouters, G. W. Lucassen, B. H. W. Hendriks, J. Wesseling, and T. J. M. Ruers, "Diffuse reflectance spectroscopy: a new guidance tool for improvement of biopsy procedures in lung malignancies," *Clin. Lung Cancer* **13**(6), 424–431 (2012).
41. Z. Liu, H. Wang, and Q. Li, "Tongue Tumor Detection in Medical Hyperspectral Images," *Sensors* **12**(1), 162–174 (2011).
42. L. L. De Boer, T. M. Bydlon, F. Van Duijnhoven, M.-J. T. F. D. V. Peeters, C. E. Loo, G. A. O. Winter-Warnars, J. Sanders, H. J. C. M. Sterenberg, B. H. W. Hendriks, and T. J. M. 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).
43. H. Chen, Z. Lin, L. Mo, T. Wu, and C. Tan, "Near-infrared spectroscopy as a diagnostic tool for distinguishing between normal and malignant colorectal tissues," *BioMed Res. Int.* **2015**, 1–7 (2015).
44. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Appl. Spectrosc.* **43**(5), 772–777 (1989).
45. Å. Rinnan, F. Van Den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC, Trends Anal. Chem.* **28**(10), 1201–1222 (2009).
46. E. Kho, B. Dashtbozorg, L. L. de Boer, K. K. V. de Vijver, H. J. C. M. Sterenberg, and T. J. M. Ruers, "Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information," *Biomed. Opt. Express* **10**(9), 4496–4515 (2019).
47. E. Kho, L. L. de Boer, K. Van de Vijver, F. van Duijnhoven, M.-J. T. F. D. V. Peeters, H. J. C. M. Sterenberg, and T. J. Ruers, "Hyperspectral imaging for resection margin assessment during cancer surgery," *Clin. Cancer Res.* **25**(12), 3572–3580 (2019).
48. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* (John Wiley & Sons, 2012).
49. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
50. T. J. Farrell, M. S. Patterson, and B. Wilson, "A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo," *Med. Phys.* **19**(4), 879–888 (1992).
51. R. Nachabe, B. H. W. Hendriks, A. E. Desjardins, M. van der Voort, M. B. van der Mark, and H. J. C. M. Sterenberg, "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).
52. D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Computat.* **1**(1), 67–82 (1997).
53. S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric," *PLoS One* **12**(6), e0177678 (2017).
54. E. Suh, E. R. Dougherty, J. Lowey, J. Hua, and Z. Xiong, "Optimal number of features as a function of sample size for various classification rules," *Bioinformatics* **21**, 1509–1515 (2005).