Diffuse reflectance spectroscopy as a tool for real-time tissue assessment during colorectal cancer surgery

Elisabeth J. M. Baltussen
Petur Snaebjornsson
Susan G. Brouwer de Koning
Henricus J. C. M. Sterenborg
Arend G. J. Aalbers
Niels Kok
Geerard L. Beets
Benno H. W. Hendriks
Koert F. D. Kuhlmann
Theo J. M. Ruers
Diffuse reflectance spectroscopy as a tool for real-time tissue assessment during colorectal cancer surgery

Elisabeth J. M. Baltussen,a,* Petur Snaebjornsson,b Susan G. Brouwer de Koning,a Henricus J. C. M. Sterenborg,a,c Arend G. J. Aalbers,a Niels Kok,a Geerard L. Beets,a Benno H. W. Hendriks,a,b Koert F. D. Kuhlmann,a and Theo J. M. Ruersa,f

aAntoni van Leeuwenhoek Hospital-The Netherlands Cancer Institute, Department of Surgery, Amsterdam, The Netherlands
bAntoni van Leeuwenhoek Hospital-The Netherlands Cancer Institute, Department of Pathology, Amsterdam, The Netherlands
cAcademic Medical Centre, Department of Biomedical Engineering and Physics, Amsterdam, The Netherlands
dPhilips Research, Department of In-body Systems, Eindhoven, The Netherlands
eDelft University of Technology, Department of Biomechanical Engineering, Delft, The Netherlands
fTechnical University Twente, MIRA Institute, Enschede, The Netherlands

Abstract. Colorectal surgery is the standard treatment for patients with colorectal cancer. To overcome two of the main challenges, the circumferential resection margin and postoperative complications, real-time tissue assessment could be of great benefit during surgery. In this ex vivo study, diffuse reflectance spectroscopy (DRS) was used to differentiate tumor tissue from healthy surrounding tissues in patients with colorectal neoplasia. DRS spectra were obtained from tumor tissue, healthy colon, or rectal wall and fat tissue, for every patient. Data were randomly divided into training (80%) and test (20%) sets. After spectral band selection, the spectra were classified using a quadratic classifier and a linear support vector machine. Of the 38 included patients, 36 had colorectal cancer and 2 had an adenoma. When the classifiers were applied to the test set, colorectal cancer could be discriminated from healthy tissue with an overall accuracy of 0.95 (±0.03). This study demonstrates the possibility to separate colorectal cancer from healthy surrounding tissue by applying DRS. High classification accuracies were obtained both in homogeneous and inhomogeneous tissues. This is a fundamental step toward the development of a tool for real-time in vivo tissue assessment during colorectal surgery. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.22.10.106014]

Keywords: diffuse reflectance spectroscopy; colorectal cancer; margin assessment; support vector machine; machine learning.

Real-time tissue differentiation during surgery may aid the surgeon in choosing the most optimal resection plane, enabling radical tumor resection while sparing vital healthy tissue.

1 Background

Colorectal cancer is the third most common cancer worldwide and ranks as the fourth cause of death due to cancer. About 30% of the cases are located in the rectum. For patients with rectal cancer, surgery, frequently combined with radiotherapy or chemoradiotherapy, is the standard of care. The delicate balance between complete removal of the tumor and the sparing of vital surrounding structures is of utmost importance in rectal cancer surgery. Perioperative tissue differentiation is complex, especially when there is fibrosis after preoperative chemoradiation. Inadequate tumor recognition can result in tumor positive resection margins, generally defined as tumor tissue within 1 mm of the circumferential resection margin (CRM). In patients with a positive CRM, the risk of local tumor recurrence is significantly increased and survival is compromised.

Although complete tumor removal is essential, damage to vital surrounding structures should be prevented during surgery. Damage to structures such as nerves may cause bladder and sexual dysfunction. Currently, there is no intraoperative technique available for the assessment of the CRM or identification of surrounding structures, such as nerve bundles, except for the possibility to submit tissue for frozen section assessment.

*Address all correspondence to: Elisabeth J. M. Baltussen, Email: l.baltussen@nki.nl

Paper 170454R received Jul. 11, 2017; accepted for publication Oct. 2, 2017; published online Oct. 27, 2017.
the bowel wall. Hence, differentiation has to be made between tumor tissue, fatty tissue, and healthy bowel wall. Only a few studies have applied DRS in a surgical setting for rectal cancer. In a study from our group by Langhout et al., tumor tissue could be distinguished from healthy tissue, colon wall, epiploic fat, and mesocolic tissue, with an accuracy of 91%. An analytical model was used to obtain concentrations of the constituents of the tissue. Based on the concentrations, the measurements are classified. This analytical model assumes that a semi-infinite homogenous optical medium is measured. However, the colorectal wall is a layered structure and thus not homogenous limiting the applicability of the analytical model to this problem as it leads to unstable and inaccurate fit results. In addition, in our earlier study by Langhout et al., it was not possible, based on histology, to determine inhomogeneity at the measurement locations, e.g., at tumor border locations with more than one tissue type.

Before taking the next step to in vivo measurements, we wanted to eliminate these limitations by quantifying the inhomogeneity of the sample volume as well as using a classification technique less sensitive to such inhomogeneity. To this end, we developed a method for more precise correlation between measurement location and histopathology as well as a tissue classification method less sensitive to inhomogeneity of the measurement site.

2 Materials and Methods

2.1 Study Protocol

In this ex vivo study, patients from the Antoni van Leeuwenhoek Hospital-The Netherlands Cancer Institute (Amsterdam, The Netherlands) and the Slotervaart Medical Centre (Amsterdam, The Netherlands) were included. The study was performed under approval of the protocol by the Hospital Ethics Review Board. All patients were undergoing surgery for colorectal cancer.

Immediately after surgical resection, three tissue samples were taken from the specimen—tumor tissue, healthy colon or rectal wall, and (pericolooreal) fat—and were placed into separate pathology cassettes. Within 1 h after resection, DRS measurements were performed on all three tissue samples. Measurements of healthy colon wall were always performed on the muscular layer, i.e., the muscularis propria. For the microscopic verification of the DRS measurement locations afterward, the tissue samples were kept in the pathology cassettes. Furthermore, an overview image and images of each measurement recording the exact measurement locations (Fig. 1) were captured by an RGB-camera. Per patient, three measurements per location were performed, in at least three different locations per tissue type. The three measurements per location were averaged before the analysis. After acquisition of all spectra, the tissue samples were returned to the Pathology Department where the samples were further processed according to standard protocols. The resulting slides were examined by a pathologist, who marked the various tissue types. These annotated pathology slides were visually correlated with the RGB images taken from the measurement locations to identify the histological tissue type at each measured location (Fig. 1).

Measurements on locations, which were on microscopic analysis proven to be fat, healthy colorectal wall, and tumor, were included in the further analysis. Measurements were excluded when the measurement location was not present in the pathology slide. Locations at the border of two tissue types were initially excluded from the analysis to reduce bias. At a later stage, these locations were analyzed separately.

2.2 Diffuse Reflectance Spectroscopy System and Probe

The DRS system consists of a Tungsten halogen broadband light source (360 to 2500 nm) with an embedded shutter and two spectrometers. Together, the two spectrometers cover the visible and near-infrared light range; the first spectrometer resolves...
light in the visible range 400 to 1100 nm (Andor Technology, DU420ABRDD), whereas the second resolves light in the near-infrared range 900 to 1700 nm (Andor Technology, DU492A-1.7) (Fig. 2). The spectrometers are controlled by custom LabView software (National Instruments) to acquire the data. The DRS system used in this study has been extensively described previously together with the calibration used for the system.19,20

In all patients, the same probe, shown in Fig. 2, was used to obtain the DRS measurements. The probe consists of three optical fibers with a core diameter of 200 μm. One of the three fibers was used to transport the light from the source to the tissue, whereas the other two were used to transport the light from the tissue to one of both spectrometers. The distance between the two collecting fibers and the delivering fiber is 2 mm. This gives a mean sampling depth of about 2 mm (max 4 mm).

2.3 Data Analysis

All analyses were performed using the perClass toolbox (Academic version 5.0, PR Sys design, Delft, The Netherlands) in MATLAB (version 8.5, MathWorks Inc., Natick, Massachusetts). To create and test a classifier, the data were first divided randomly into a training and test set. The data were split such that measurements of one patient were not divided between training and test set: 80% of the patients were used for training the classifier and the remaining 20% were used for testing the classifier.

Before training the classifier, the number of features used for the classification was reduced with the use of spectral bands. The spectral bands are selected using k-means clustering and consist of continuous wavelengths with similar spectral intensity. The mean intensity per spectral band is used as input for the further classification method. After determination of the spectral bands, two different classifiers were trained to separate the three different tissue types. First, a quadratic classifier was used to separate the measurements on fat tissue from the measurements on healthy colorectal wall and tumor tissue. Thereafter, a linear support vector machine (SVM) was used to distinguish measurements on healthy colorectal wall from tumor tissue. The SVM was optimized using a receiver operating curve (ROC) created with a 10-fold cross validation using the training data. The optimal point selected in the ROC is the point that maximizes the mean accuracy. The final classifier pipeline included the spectral band selection, the quadratic classifier, and the linear SVM (Fig. 3).

After training the classifiers, the classifier pipeline was applied to the test dataset to assess the performance of the classifier pipeline on separating the data of the different tissue types. The training and testing of the classifier was repeated 10 times, with different randomly selected training and test sets, to obtain a more precise accuracy.

Finally, the dataset, consisting of all measurement locations containing one tissue type, was used to train the classifier pipeline once more. In clinical practice, there is a chance of measuring more than one tissue type in a single measurement. To evaluate the accuracy of the developed classifiers in such circumstances where only part of the measurement volume contained tumor tissue, the measurements that contained more than one tissue type were further analyzed using this final developed classifier pipeline. The proportions of the different types of tissue involved were estimated based on visual correlation with the pathology slides.

3 Results

3.1 Patients

Specimens were obtained from 38 patients: 22 male and 16 female, with a median age of 65 years (interquartile range 59 to 73). Most of the tumors were located in the sigmoid (17 patients), followed by the ascending and descending colon (14 and 5 patients, respectively) and rectum (2 patients). Of the included tumors, two were adenomas and the remainder was carcinomas, in which three were mucinous carcinomas and the rest conventional adenocarcinomas. The pT stage of 36 carcinomas was as follows: 2 were pT1, 5 were pT2, 20 were pT3, and 9 were pT4.

3.2 Data

In total, 595 locations were measured, in which 147 were excluded based on the microscopic assessment. Of these, 111
locations were excluded because they were not present in the pathology slides, or the measured tissue type was not of interest. The other 36 locations were analyzed separately during a second analysis, because they were located at the border between two tissue types.

Measurements made on the mucinous carcinomas and adenomas were excluded from the dataset, because the number of measurements done in these tissue types was too low to create a separate group.

After exclusion, 448 locations remained with a clear-cut microscopic tissue. From these locations, 134 were located in fat tissue, 150 in healthy colorectal wall, and the remaining 164 in tumor tissue.

### 3.3 Training of the Classifier

After randomly selecting 80% of the patients, the training set consisted of 30 patients and 373 measured locations. Based on the spectra from the training set, 16 spectral bands were selected (Fig. 4). The quadratic classifier, applied on the training data, could distinguish fat from tumor tissue and healthy colorectal wall with a mean accuracy of 1.00 (Table 1).

**Table 1** Results of quadratic classifier on training dataset.

<table>
<thead>
<tr>
<th>Classification outcome</th>
<th>Fat</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fat</td>
<td>112 (^a)</td>
<td>0</td>
<td>112</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>260 (^b)</td>
<td>261</td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td>260</td>
<td>373 (^c)</td>
</tr>
</tbody>
</table>

\(^a\)Specificity = 1.00.  
\(^b\)Sensitivity = 1.00.  
\(^c\)Accuracy = 1.00.

The ROC used for the optimization of the SVM is shown in Fig. 5. The accuracy of the SVM on the training data was 0.92 (Table 2).

Once the quadratic classifier and SVM were combined and applied to the training data, a mean accuracy of 0.95 was obtained.

### 3.4 Performance of the Classifier on the Test Set

The test set consisted of 20% of the patients (n = 8) and a total of 75 measurement locations. The classifier pipeline applied to the test set provided a mean accuracy of 0.95 in discriminating tumor tissue from fat and healthy colorectal wall. The results from the 10 iterations in the training and testing of the classifiers are shown in Table 3.

**Table 2** Results of SVM on training dataset.

<table>
<thead>
<tr>
<th>Classification outcome</th>
<th>Colorectal wall</th>
<th>Tumor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorectal wall</td>
<td>115 (^a)</td>
<td>6</td>
<td>121</td>
</tr>
<tr>
<td>Tumor</td>
<td>14</td>
<td>125 (^b)</td>
<td>139</td>
</tr>
<tr>
<td>Total</td>
<td>129</td>
<td>131</td>
<td>260 (^c)</td>
</tr>
</tbody>
</table>

\(^a\)Specificity = 0.95.  
\(^b\)Sensitivity = 0.90.  
\(^c\)Accuracy = 0.92.

**Table 3** Results of 10 time iterations of the creation and testing of the classifier. Mean and standard deviations of the accuracy per tissue type obtained on the test set, over 10 iterations are shown. Finally, the mean across all tissue types is shown.

<table>
<thead>
<tr>
<th></th>
<th>Fat</th>
<th>Colorectal wall</th>
<th>Tumor</th>
<th>Mean across all tissues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.00</td>
<td>0.93</td>
<td>0.92</td>
<td>0.95</td>
</tr>
<tr>
<td>STD</td>
<td>0.00</td>
<td>0.05</td>
<td>0.09</td>
<td>0.03</td>
</tr>
</tbody>
</table>
3.5 Performance Classifier on Measurements of Border Locations

The classifier trained on the 448 locations with a clear-cut microscopic tissue was applied to the 36 measurements that were located at the border of two tissue types. The classification of 29 (80%) of these locations was in accordance with the most prominent tissue type observed in these border locations.

4 Discussion

In this ex vivo setup, colorectal tumor tissue can be distinguished from healthy bowel wall and surrounding fat tissue, through the use of DRS, with a mean accuracy of 0.95 (±0.03). The distinction of fat from the other two tissue types is the most accurate distinction (accuracy of 1.00).

The results obtained in this study are similar to the results reported by Schols et al. and Langhout et al. The first was able to distinguish tumor tissue from the surrounding tissue including mesenteric fat and colon tissue with an accuracy of 0.99. However, no more than 60 tumor measurements obtained from only 6 patients were included. This limited dataset will increase the risk of overfitting. Moreover, the gold standard used by Schols et al. is the classification achieved by visual assessment by the surgeon. However, the final classification of the CRM requires histopathological examination. Therefore, it would be more accurate to use the histopathological classification as the gold standard and not the visual classification as assessed by the surgeon.

In the study by our own group reported by Langhout et al., an analytical model was used, in which knowledge about the spectra of various chromophores is used to fit a curve to the obtained spectra. With the use of this knowledge, the assumption is made that the measured tissue is homogeneous. The colorectal wall exists of several layers of different tissues such as muscular and connective tissue. Therefore, the measured tissue is far from homogeneous and the analytical model cannot be applied without any reservation. In this study, no a priori knowledge about the homogeneity of the tissue was made. Moreover, compared to the study reported by Langhout et al., this study protocol was changed in such a way that the correlation of the measurement locations with the histology was better documented. With the use of the RGB images made of every measurement, the spectra could be linked accurately to the histology annotated on the pathology slides. Furthermore, the design of the probe used in this study is closer to a surgical tool compared to the needle used in the study reported by Langhout et al. Finally, the results obtained in this study are as good as the results reported by Langhout et al., but were obtained without the use of fluorescence data. This does not only decrease the number of measurements needed to obtain a high accuracy, but it also reduces the costs by essentially simplifying the instrumentation and reduces the time needed for the measurements as the fluorescence signal is weak and requires a much longer integration time (s) than the DRS measurements.

To classify all measurements, two classifiers were used in this study. First, the most prominent spectrum of fat was differentiated from healthy colorectal wall and from tumor tissue, both being classified into the same class. Second, to separate healthy colorectal wall from tumor measurements, a more complex classifier was needed since the differences in the spectra of both tissue types are subtle (Fig. 4). Because classification was reduced to a binary task, a linear SVM was used to separate both tissue classes from each other. For well-defined tissue classes, this approach resulted in an overall accuracy of 95%.

In this study, 36 locations, which were located at the border between two tissue types, were classified as well. With an accuracy of 80%, these locations were classified as the tissue type that was most prominently present in the measured volume as judged histologically. It should be noted that the histological classification of the border locations may be less accurate than the histological classification of the locations containing a single tissue type. Small inaccuracies in matching the RGB images to histopathology may evidently affect inhomogeneous measurement volumes more than homogenous measurement volumes. During surgery, there is a fair chance to obtain such border measurements. In fact, the entire measurement procedure is intended to assist the surgeon in assessing such locations. Therefore, these locations at the border of two tissue types are of great importance to the further development of an in vivo application. In addition, the absolute minimum volume of tumor tissue that can be measured should be further determined. With the current measurement technique, in which we measured a volume of about 2 mm³, it was not possible to determine a lowest detection level for small amounts of tumor tissue.

The main limitation of this study is the generalizability of the results. Neither the quadratic classifier nor the SVM provides the wavelengths or tissue constituents on which the distinction is based between tumor tissue and the healthy surrounding tissues. Therefore, from our model, it is not known on what features fat can be distinguished from tumor tissue and healthy colorectal wall, and on what basis healthy colorectal wall can be distinguished from tumor tissue. As a result, it is not possible to generalize the results obtained in this study to an in vivo setting or to a differently designed measurement set up. In the transition to the in vivo setting, there will be many changes that might influence the optical spectra. For example, blood, as may be present during surgery, may have a major influence on the obtained spectrum. Still, as shown by Spliethoff et al., the influence of blood in lung biopsies may be minimal depending on the attribution of the different wavelengths to the classification. The spectral bands used in this study included the entire blood absorption band (450 to 600 nm) in one spectral band. As shown in Fig. 4, the intensities of the three mean spectra do not differ much in this spectral band. This suggests that the absorption of blood may not have a major influence in the present classification.

The fiber distance used in this study was 2 mm; this implies that the measurements were made at a depth of about 2 mm in the tissue. To be useful in colorectal surgery, a sampling depth of about 5 mm is required. To increase the sampling depth of the DRS measurements, the fiber distance needs to be increased. However, with an increase in fiber distance, the entire sampling volume increases. This results in less sensitivity of the DRS measurements to small amounts of for instance tumor. Therefore, a study needs to be conducted to determine if the DRS measurements are still able to detect tumor at 5-mm depth using increased fiber distance.

In this study, fibrosis (connective tissue without tumor cells) was not included in the measured tissue types, which is another limitation of this study. Patients with advanced stage rectal cancer almost always receive neoadjuvant chemoradiotherapy. These patients are highly likely to develop fibrosis on the radiation site. Fibrosis is challenging to distinguish from tumor tissue through the surgeon’s visual assessment alone. Therefore,
it would be beneficial to include this tissue type when real-time tissue classification is done. In this study, fibrosis was not included because of the limited occurrence of fibrosis in the examined specimens.

For future research, two main challenges need to be addressed. First of all, fibrosis needs to be added to the tissue types that are examined. Therefore, an additional study needs to be performed that concentrates on rectal specimens of patients who underwent neoadjuvant chemo- and radiotherapy. Second, the advance to an in vivo setting needs to be made. The results of an in vivo study will be closer to the ultimate application of the real-time tissue assessment. Moreover, an in vivo study will give more insight into the workflow during surgery, which will provide a basis for the design and use of a final device used for real-time tissue assessment. The challenges of an in vivo study are first of all the identification of tumor tissue based on the visual assessment of the surgeon. Furthermore, the correlation with pathology will be challenging in an in vivo setting.

5 Conclusion

In this ex vivo study, the distinction between tumor tissue and healthy surrounding tissue in patients with colorectal cancer can be made with a high accuracy through the use of DRS measurements. The mean accuracy of the classification is 0.95 (±0.03). However, the results of this study are still limited to this study setup and protocol. Future research must be focused on the translation to the in vivo setting. This should ultimately result in reliable real-time tissue assessment leading to the increase of radical resections and decrease of postoperative morbidity rates after colorectal cancer surgery.

Disclosures

This study was supported by Philips Research, Eindhoven, Netherlands. The author who is affiliated with Philips Research (BHW) has financial interests in the subject matter, materials, and equipment only in the sense that he is an employee of Philips. None of the other authors has any financial relationship with Philips Research or conflicts of interest.

Acknowledgments

The author thanks Dr. C. C. Flohil for her assistance in the collection of the data at the Slotervaart Medical Centre and Dr. G. S. Langhout for reviewing of this manuscript. This work was supported by Philips Research, Eindhoven, the Netherlands and KWF-Alpe d’HuZes (NKI 2014-6596). We are grateful for their support.

References


Biographies for the authors are not available.