



Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Raman fiber-optical method for colon cancer detection: Cross-validation and outlier identification approach



D. Petersen^{a,1}, P. Naveed^{b,c,1}, A. Ragheb^{b,2}, D. Niedieker^a, S.F. El-Mashtoly^a, T. Brechmann^c, C. Kötting^a, W.H. Schmiegel^{b,c}, E. Freier^{a,3}, C. Pox^{b,4,*}, K. Gerwert^{a,*}

^a Department of Biophysics and Protein Research Unit Europe (PURE), Ruhr University Bochum, ND/04 Nord, 44780 Bochum, Germany

^b University Hospital Knappschaftskrankenhaus Bochum, In der Schornau 23-25, 44892 Bochum, Germany

^c Department of Gastroenterology/Hepatology, Berufsgenossenschaftliches Universitätsklinikum Bergmannsheil GmbH, Bürkle-de-la-Camp-Platz 1, 44789 Bochum, Germany

ARTICLE INFO

Article history:

Received 1 February 2017

Received in revised form 20 March 2017

Accepted 24 March 2017

Available online 25 March 2017

Keywords:

Raman spectroscopy

Cancer recognition

Label-free

Bioinformatics

Colonoscopy

ABSTRACT

Endoscopy plays a major role in early recognition of cancer which is not externally accessible and therewith in increasing the survival rate. Raman spectroscopic fiber-optical approaches can help to decrease the impact on the patient, increase objectivity in tissue characterization, reduce expenses and provide a significant time advantage in endoscopy. In gastroenterology an early recognition of malign and precursor lesions is relevant. Instantaneous and precise differentiation between adenomas as precursor lesions for cancer and hyperplastic polyps on the one hand and between high and low-risk alterations on the other hand is important. Raman fiber-optical measurements of colon biopsy samples taken during colonoscopy were carried out during a clinical study, and samples of adenocarcinoma (22), tubular adenomas (141), hyperplastic polyps (79) and normal tissue (101) from 151 patients were analyzed. This allows us to focus on the bioinformatic analysis and to set stage for Raman endoscopic measurements. Since spectral differences between normal and cancerous biopsy samples are small, special care has to be taken in data analysis. Using a leave-one-patient-out cross-validation scheme, three different outlier identification methods were investigated to decrease the influence of systematic errors, like a residual risk in misplacement of the sample and spectral dilution of marker bands (esp. cancerous tissue) and therewith optimize the experimental design. Furthermore other validations methods like leave-one-sample-out and leave-one-spectrum-out cross-validation schemes were compared with leave-one-patient-out cross-validation. High-risk lesions were differentiated from low-risk lesions with a sensitivity of 79%, specificity of 74% and an accuracy of 77%, cancer and normal tissue with a sensitivity of 79%, specificity of 83% and an accuracy of 81%. Additionally applied outlier identification enabled us to improve the recognition of neoplastic biopsy samples.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Early recognition of cancer plays a key role for decreasing the mortality and morbidity rate in patients. Colorectal cancer is one of the most common cancers worldwide with over 1.3 million new cases and 694,000 deaths per year [1]. Routine screening and an early recognition of colorectal cancer and its precursor lesions such as premalignant polyps, primarily adenomas, can help to decrease this number [2].

* Corresponding authors.

E-mail addresses: cpox@sjs-bremen.de (C. Pox), gerwert@bph.ruhr-uni-bochum.de (K. Gerwert).

¹ These authors have equally contributed to this work.

² Present address: Evangelisches Krankenhaus Düsseldorf, Kirchfeldstraße 40, 40217 Düsseldorf, Germany.

³ Present address: Leibniz-Institut für Analytische Wissenschaften (ISAS e.V.), Otto-Hahn-Str. 6b, 44227 Dortmund, Germany.

⁴ Present address: Hospital St. Joseph-Stift Schwachhauser, Schwachhauser Heerstr. 54, 28209 Bremen, Germany.

Since the colon is rather simple to access, endoscopy of the colon (colonoscopy) is recommended as screening method for adults in various countries [3]. In routine diagnostics, cancerous and precancerous lesions are detected by using conventional white light reflectance [4] and chromoendoscopy [4,5]. Additional narrow band imaging [4,6] and fluorescence endoscopy [4] are used at times. After recognition, a suspicious tissue area is resected as biopsy sample and sent to histological assessment. Usually after particular preparation the specimen is histochemically stained, examined under a light microscope and the diagnosis is reported to the referring physician. For the physician not only the differentiation between carcinoma and normal tissue, but also a differentiation between precancerous adenomas and hyperplastic polyps is essential [7,8]. Up to now, the endoscopic characterization of polyps regarding histology shows inadequate reliability [9]. Therefore, many polyps with minimal malignant potential are removed without benefit for the patient, but still accompanied by this is a residual risk of bleeding and perforation after colonic polypectomy [10]. Furthermore, a small but not deniable number of precancerous adenomas,

especially, if they are smaller than 10 mm, are missed during colonoscopy [11]. Hence Wallace et al. [4] pointed out that an improved detection and in vivo classification of colorectal polyps are needed. Complementing the morphological information with knowledge of the biochemical properties can help in a targeted selection of biopsy samples. By this a reduction or abandonment of polypectomy can be achieved, while simultaneously the automatization of tissue assessment will be increased.

This automatization in diagnostics can be realized using spectral histopathology, which relies on vibrational spectroscopy, like infrared and Raman spectroscopy [12–15]. Raman spectroscopy is capable of measuring the biochemical fingerprint of a tissue sample in a label-free manner, and consequently information of the tissue composition is obtained instantly [16,17]. For example with Raman microspectroscopy cancer can be differentiated from normal tissue with a high sensitivity (>95%) and specificity (>95%) [13,18]. This immediate analysis of the tissue composition during medical examination can be used in endoscopy. It can lead the physician to select a suitable spot for the biopsy, enables a reduction of complications, saves histopathological assessment time, benefits the patient and reduces economic burdens. Raman spectroscopy is based on the inelastic scattering of light. It reflects the biochemical fingerprint by monitoring the molecular vibrations and therewith provides information about the chemical composition of the tissue and the progression of carcinogenesis [19,20]. Using laser light as excitation, the wavelength shifts from the monochromatic laser line are recorded, and a measurement without a resection of a biopsy sample can be conducted [21].

Different groups have shown the potential of Raman fiber-optics for ex vivo and in vivo cancer diagnostics [22–26]. However, further improvements of Raman fiber optics for endoscopy are required [27]. In recent publications for Raman spectroscopy applied in colonoscopy, respectively on samples from colonoscopy, Wood et al. [28] and Bergholt et al. [29] showed the applicability of Raman fiber-optics. Wood et al. had undertaken an ex vivo study on 177 patients with a fiber probe constructed by Day et al. [30] and analyzed the data in the fingerprint region. On defrosted snap frozen samples they were able to differentiate between normal tissue and various lesions (e.g. adenoma and adenocarcinoma) using only a leave-one-spectrum-out cross-validation approach, which might be influenced by spectral similarity of the same sample and does not reflect the measurement of unknown samples. Normal tissue vs. adenocarcinoma or adenoma was classified with an accuracy of 88.5% and 88.3%, respectively. Furthermore, adenoma was separated from hyperplastic polyps with an accuracy of 78%. Bergholt et al. [29] conducted in vivo measurements on 49 patients (121 samples) with a self-developed fiber probe in Singapore. They distinguished between high- and low-risk lesions (14 vs. 35 patients, 17 vs. 104 samples) using a leave-one-sample-out cross-validation approach with a majority vote over all spectra of the same sample. High-risk lesions were separated from low-risk lesion with an accuracy of 90.9%. Adenoma was separated from hyperplastic polyps with an accuracy of 81.8%. Because of limited number of measurements and patients leave-one-out cross-validation approaches are preferred for validation in Raman fiber optical studies.

Even in well prepared study designs, a residual risk in misplacement of the sample, misidentification of the sample by the pathologist and a dilution of the spectral marker (loss of spectral cancerous markers due to a mixture of multiple tissue components) remain. If these spectra persist in the training data, the classification will be impaired, especially by the small differences observed. Therefore identification approaches for outlying measurements could be included, only if a sufficient number of patients is acquired. For an improved training database, several approaches of outlier identification schemes are suggested in the literature [31,32]. For instance the Local Outlier Factor (LOF) works as a density based method to identify abnormal data points [33].

In this study, Raman fiber optical measurements of colon biopsy samples taken during colonoscopy were carried out. We analyzed for

the first time a large data set of 343 biopsy samples from 151 patients with multiple leave-out cross-validation schemes and outlier identification approaches, which allows an enhanced setup for a training database. With this, we were able to improve the analysis and training of classification models for tissue type differentiation that can be used for Raman endoscopic measurements.

2. Materials and Methods

2.1. Sample Collection, Preparation and Medical Assessment

Ethical approval was granted from the institutional review board of the Ruhr-University Bochum (Faculty of Medicine, reference number 4886-14). Consecutive patients undergoing elective colonoscopy in the Knappschafts-Krankenhaus Bochum-Langendreer were enrolled. During colonoscopy biopsies of around 1–4 mm² in size were taken with a standard biopsy forceps. Larger polyps were resected with a polypectomy snare if suitable. The resected tissue was washed in physiological NaCl solution and up to 10 spectra per sample were measured immediately on the fresh tissue by Raman spectroscopy at different positions. Afterwards, the standard clinical workflow was undertaken, i.e. fixation with formaldehyde and examination by an expert pathologist. The histological assessment is based on standard staining using hematoxylin and eosin (H&E) and additional immunohistochemical methods if necessary. The pathologist's report was set as reference for the classification model.

2.2. Setup

Raman measurements were conducted using a commercially available B&W Tek i-Raman Plus system (B&W Tek, Delaware, USA). The instrument is equipped with a 785 nm laser with a maximum laser power of roughly 300 mW at the sample and is set to measure a Raman shift interval from -220 cm^{-1} to 3310 cm^{-1} with a spectral resolution of about 3.3 cm^{-1} . The standard fiber probe head (BAC102) was boxed in, and additional sample holders were constructed to be exchanged on the top of the probe head for each patient after the measurement (see Fig. S1). The distance between the shaft and the window of the exchangeable sample holder was adjusted by a silicon slide on top of the window. The measurement volume (penetration depth) was set to 250 μm above the window, on which the samples is placed, based on the maximum signal of the silicon at 520 cm^{-1} . With a core width of 105 μm for the excitation fiber, 200 μm for the collection fiber, a spot size at the sample surface of 85 μm according to the manufactures and a penetration depth of 250 μm , we estimated the diameter of the measurement volume to a factor 2–3 of the spot size for the modified setup, i.e. around 200–250 μm . Spectra were integrated for 2 s and averaged over five accumulations for each spot. A graphical user interface was programmed for an easy control of the measurement by the clinician. Since there is no Raman fiber probe for in vivo measurements is approved in Germany, we conducted an ex vivo study.

2.3. Data Analysis

The raw data were processed in Matlab Version 7.14 along with the Image Processing and Statistics toolboxes (The Mathworks, Inc., Mass., USA) and algorithms developed in-house. At first, spectra of single spots were averaged and corrected for wavelength dependent absorption of the fiber probe. Afterwards cosmic spikes were corrected and the fraction of water was estimated from the band area between 3100 cm^{-1} and 3300 cm^{-1} . The spectrum of water with the estimated intensity was subsequently subtracted. The baseline was corrected separately for the fingerprint region and the CH-stretching region with the airPLS algorithm [34], and spectra were normalized for the intervals between 950 cm^{-1} and 1800 cm^{-1} and 2800 cm^{-1} and 3050 cm^{-1} as a standard procedure for comparison of Raman spectra from different measurements [16].

These spectra with parameter from the raw spectra are inspected by a spectral quality control algorithm, which sorts out spectra with low SNR and artifact affected spectra on the base of fluorescence and Raman signal intensity. The variance of the mean spectra was normalized to the number of patients ($\text{var}_{\text{patient}} = \text{var}/n$).

After assignment of the spectra to the annotations of the pathologist, spectral features selected by minimum redundancy maximum relevance feature selection [35,36] were chosen to train a Support Vector Machine (SVM). SVM models were validated by leave-one-patient-out, leave-1/3-patient-out, leave-one-sample-out and leave-one-spectrum-out cross-validation, respectively. For each cross-validation method single spectra were used for validation with random subsampling by 2000 iteration steps for leave-patient-out approaches and 5000 iteration steps for leave-one-sample-out and leave-one-spectrum-out approaches. In the leave-one-patient-out and leave-one-sample-out approach a maximum of six spectra per patient/sample were randomized chosen for validation.

Since inaccuracies during measurement and pathological annotation cannot completely be excluded, three different approaches for outlier identification were applied: One Class Classification with SVM (OCC) [31,37], Local Outlier Factor (LOF) [31,33] and a supervised approach for modeling normality and abnormality [32], which is named Refinement of Training Data (RoTD) in the following. OCC and LOF were applied on the spectra of the single classes. RoTD was applied as a preclassification scheme for the single classes against the other classes. For all three methods, outlier were identified and disregarded for training the SVM models. Nonetheless all spectra are used for validation. The outlier identification methods are further described in the SI. The distribution of the identified spectra is displayed in the Venn diagram in Fig. S2.

For an estimation of the minimum fraction of cancer within the measurement volume of the fiber probe a simple model was developed. Here, Raman microscopic data of thin sections from native colon tissue was used to simulate an integral spectrum with different tissue components which occur in the measurement volume of the fiber probe. Spectra of large confocal volumes (ca. 1 mm³) and mixed components as obtained for the fiber-optical approach were simulated with Raman microspectroscopic data of pure components from measurements of colon tissue. This data was received from tissue which was mounted as thin section on a Raman microscope (Horiba XploRA One, 785 nm), dried with a stream of dried air and measured in a raster of 5 μm × 5 μm. Data for simulation were collected on the base of hierarchical cluster analysis and the corresponding H&E annotation. With this data, single high volume component spectra were simulated with 200 spectra from the training set in varying compositions and classified with a random forest model, which was trained on the microscopic data. The fraction of cancer was varied stepwise from 0%–100% for each composition with additional tissue components (s. single line in Table 4 and Table S1). For each step the simulation of a single high volume component spectrum was repeated 400 times with randomized chosen spectra for simulation and training. Afterwards the minimum fraction of cancer that still leads to the classification as cancer was determined. This calculation defines a minimum fraction for cancer detection.

3. Results and Discussion

The automated and instant recording of the biochemical fingerprint has the potential to decrease complications and examination time during endoscopy. We set up an ex vivo study for automated label-free detection of cancer and precursor lesions. Therefore, we analyzed Raman measurements of 343 biopsies from 151 patients who underwent colonoscopy between 02/2014 and 09/2015 using the fiber-optic approach. The biopsies were categorized into the following pathologies: adenocarcinoma (AC), tubular adenoma (TA), hyperplastic polyp (HP), and normal colon tissue (NT, unsuspected mucosa). Table 1 lists the number of patients, samples and spectra for the particular annotations including the average age and gender distribution of the patients. A spectral quality test was set up to

exclude spectra with low SNR or artifacts. Furthermore biopsies with other annotations (e.g. serrated adenoma, sigmoiditis) and a differentiation of the grading of AC were not considered in the analysis.

The workflow for data analysis was established as shown in Fig. 1. Colon biopsy samples were measured using the routine clinical workflow, including the subsequent annotation by the pathologist. For validation this workflow was extended by a leave-one-out cross-validation scheme including outlier identification. The outlier identification was aimed to avoid inaccuracies in positioning of the sample, from annotations of the pathologist and spectral dilution. A different distribution of multiple tissue components within the biopsy sample can lead to a position dependent spectrum. For validation purposes, we decided to use the leave-one-patient-out cross-validation scheme (LOPO CV) to avoid intra-patient sample spectrum similarity. A support vector machine (SVM) was used for classification. The groups of pathologies were separated by binary classifiers. On the one hand annotations were grouped in low-risk lesions (HP, NT) and high-risk lesions (AC, TA), respectively. On the other hand single annotations, like the clinical relevant TA and HP, were separately differentiated. Mean spectra of the annotated samples are shown in Fig. 2A. Difference spectra between the mean spectra of the impaired and the normal tissue are shown in Fig. 2B. Additional difference spectra of the groups of pathologies, as compared in the present study, are displayed in Fig. S3. Standard deviation normalized to the number of patients is illustrated in grey. Raman bands, for example, at 1003 cm⁻¹ (ring breathing of phenylalanine), 1263 cm⁻¹ (amide III), 1318 cm⁻¹, 1448 cm⁻¹ (CH₂ deformation of proteins and lipids), 1657 cm⁻¹ (amide I), 2850 cm⁻¹ (CH₂ stretching), and at 2930 cm⁻¹ (CH₃ stretching) are in agreement with the literature [19,29,38]. Spectral differences were found for 1275 cm⁻¹ (Amid III), 1467 cm⁻¹ (CH bending vibration), 1582 cm⁻¹ (indole ring) and 1634 cm⁻¹ (Amid I) [38,39]. Further changes from assigned bands in the mean spectra were not observed. We obtained most substantial spectral differences for AC in comparison with NT, followed by HR and NT. Minor differences were observed for HP versus NT.

For the clinical purpose, not only the precise annotation to a single class is relevant, but also the differentiation of high-risk and low-risk lesions. As described above, AC and TA were combined to the group of high-risk lesions, whereas HP and NT were grouped to low-risk lesions [4,29]. Table 2 contains the leave-one-patient-out cross-validation results for the different groups and single annotations of the cross-validation including and excluding outlier spectra in the training phase. We achieved accuracies between 62% and 81%. When disregarding any outlier identification, classifying AC vs. NT reached an accuracy of 75%. OCC shows an improvement in accuracy with 77% (+2%), whereas for the LOF approach the accuracy stays the same with 75% (+0%). A significant improvement is obtained by the RoTD algorithm with sensitivity, specificity and accuracy of 79% (+6%), 83% (+5%) and 81% (+6%), respectively. Concerning the high- and low-risk lesions, the accuracy is 71% without outlier identification. For OCC it drops to 70% (-1%), while for LOF it raises to 72% (+1%). Again, an obvious improvement is reached by the application of RoTD with an accuracy of 77% (+6%). Of all, the RoTD outlier identification allows the best improvement for the recognition. For OCC and LOF, the average improvement in accuracy in comparison with a disregard of any outlier identification is 1%. In detail, both methods induce mostly slight improvements and a few cases of impairments. In contrast, the average improvement in accuracy for

Table 1

Number of patients, samples and spectra analyzed in this study. The average age and the number of male and female patients are listed in the last two columns.

		Patients	Samples	Spectra	Age	M/F
Normal tissue	(NT)	56	101	204	57	23/33
Hyperplastic polyp	(HP)	46	79	180	65	26/20
Tubular adenoma	(TA)	69	141	404	68	34/35
Adenocarcinoma	(AC)	13	22	65	74	10/3
Sum		151	343	853	64	77/74

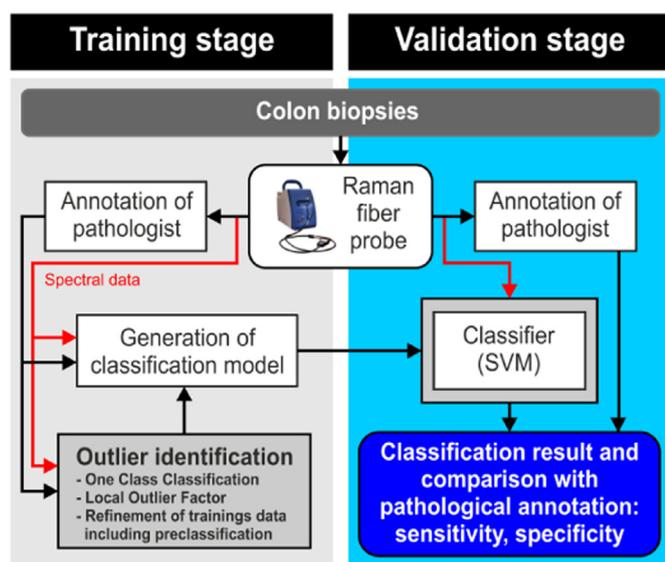


Fig. 1. Workflow for clinical measurements and data analysis with outlier identification.

the RoTD is 7%. In comparison to OCC and LOF, a clear improvement for every compared combination of the pathological groups is achieved.

Different research groups applied different leave-one-out cross-validation schemes, like Wood et al. [28] (leave-one-spectrum-out) and Draga et al. [24] (leave-one-sample-out). Bergholt et al. [29] extended

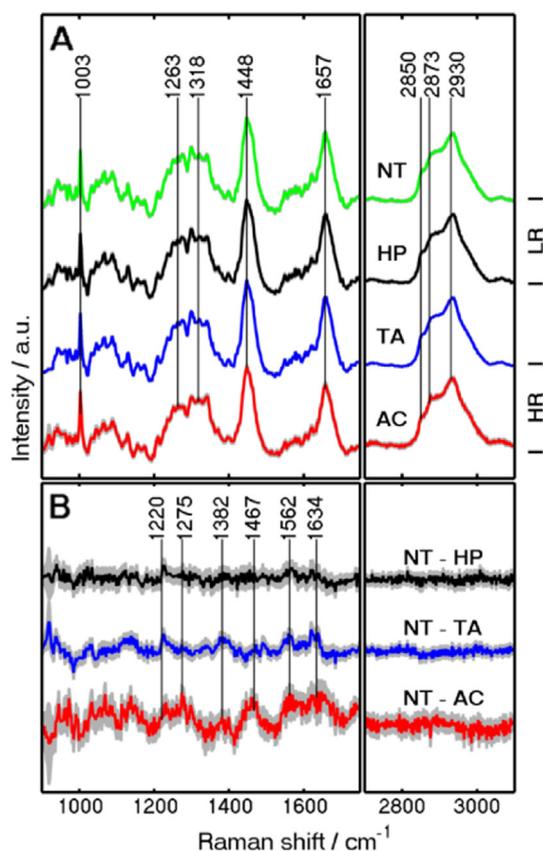


Fig. 2. (A) Mean spectra of normal tissue (NT, green), hyperplastic polyp (HP, black), tubular adenoma (TA, blue) and adenocarcinoma (AC, red). NT and HP are grouped to low-risk lesions (LR) and TA and AC to high-risk lesions (HR). (B) The difference spectra between the single annotations and NT are depicted below using the corresponding colors. The standard error and its error propagation are shown in grey. The spectral differences are not located at the maximum intensity of the Raman peaks in the average spectra. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Sensitivity, specificity and accuracy for leave-one-patient-out cross-validation results with and without outlier identification approaches. One Class Classification (OCC), Local Outlier Factor (LOF) and Refinement of Training Data (RoTD) were applied on the training data of carcinoma (AC), tubular adenoma (TA), hyperplastic polyp (HP) and normal tissue (NT). AC and TA were grouped to high-risk lesions (HR) and HP and NT to low-risk lesions (LR).

		All	OCC	LOF	RoTD
AC vs. NT	Sen. (%)	73	76	72	79
	Spe. (%)	78	80	80	83
	Acc. (%)	75	77	75	81
HR vs. LR	Sen. (%)	74	72	73	79
	Spe. (%)	66	66	69	74
	Acc. (%)	71	70	72	77
HR vs. NT	Sen. (%)	76	76	80	85
	Spe. (%)	69	70	65	76
	Acc. (%)	74	74	77	81
TA vs. NT	Sen. (%)	71	70	75	76
	Spe. (%)	62	65	62	75
	Acc. (%)	67	67	69	76
TA vs. HP	Sen. (%)	67	69	67	72
	Spe. (%)	57	59	54	63
	Acc. (%)	63	65	62	69

the leave-one-sample-out scheme for a majority vote over all classified spectra of the same samples. Table 3 lists sensitivities, specificities and accuracies of the leave-one-spectrum-out, leave-one-sample-out, leave-one-patient-out and leave-1/3-patient-out cross-validation procedures. Due to the spectral similarity within the same sample, the sensitivity, specificity and accuracy of the leave-one-spectrum-out cross-validation are significantly higher in comparison to the other cross-validation schemes. The accuracies are close to 100% except for the high-risk vs. low-risk lesions with >85%. The numbers of leave-one-sample-out cross-validation are slightly better than for leave-one-patient-out cross-validation with around 80% for all groups except TA vs. HP with around 70%, which follows the tendency of leave-one-patient-out cross-validation. A change from leave-one-patient-out to leave-1/3-patient-out cross-validation shows a slight decrease in sensitivity, specificity and accuracy for all groups except specificity for carcinoma vs. normal tissue.

Similar to our approach, Wood et al. [28] measured snap frozen samples with a fiber probe constructed by Day et al. [30] and analyzed the data in the fingerprint region. They validated the classification model with a leave-one-spectrum-out cross-validation procedure on a set of 177 patients (356 samples). NT vs. AC or adenoma was classified with an accuracy of 88.5% and 88.3%, respectively. Furthermore, adenoma was separated from HP with an accuracy of 78%. Raman fiber-optic in vivo measurements on colon tissue were performed by Bergholt et al. [29] in Singapore with a self-developed fiber probe. They investigated in total 49 patients (121 samples) with high-risk lesions from 14

Table 3

Results for Leave-One-Spectrum-Out (LOSO), Leave-One-Sample-Out (LOSAmO), Leave-One-Patient-Out (LOPO) and Leave-1/3-Patient-Out (L1/3PO) cross-validation scheme.

		LOSO	LOSAmO	LOPO	L1/3PO
AC vs. NT	Sen. (%)	95	81	79	75
	Spe. (%)	96	83	83	87
	Acc. (%)	96	82	81	81
HR vs. LR	Sen. (%)	82	80	79	73
	Spe. (%)	88	77	74	72
	Acc. (%)	85	79	77	73
HR vs. NT	Sen. (%)	99	82	85	75
	Spe. (%)	100	84	76	75
	Acc. (%)	99	83	81	75
TA vs. NT	Sen. (%)	99	78	76	75
	Spe. (%)	100	80	75	70
	Acc. (%)	99	79	76	72
TA vs. HP	Sen. (%)	100	71	72	70
	Spe. (%)	94	72	63	60
	Acc. (%)	97	71	69	65

patients (17 samples). High-risk lesions were separated from low-risk lesion with an accuracy of 90.9%. Adenoma was separated from HP with an accuracy of 81.8%.

A comparison of the results of Wood et al. [28] and our leave-one-spectrum-out cross-validation results reveal significant higher accuracies for our study. Since they measured one spectrum per sample for a greater extent, their mode can also be compared with some restrictions to a leave-one-sample-out cross-validation. Comparing all results and selecting leave-one-sample-out cross-validation with RoTD for our study, Wood [28] and Bergholt [29] achieved slightly higher sensitivities, specificities and accuracies. However, these groups applied leave-one-spectrum-out or leave-one-sample-out cross-validation. This does not avoid the biochemical similarity of the samples, which originate from the same patient. Like our observation, accuracy dropped for the differentiation between TA and HP. Nevertheless, TA and HP are morphologically similar, which likely leads to difficult spectral differentiation. The spectra measured by Wood et al. [28] show a higher standard deviation and display only the fingerprint region. Bergholt et al. [29] include the CH-stretching region. The spectra, which they measured, agree with the spectra observed in this study, apart from the phenylalanine band just recognizable in our mean spectra. For the difference spectra, we observed smaller spectral changes than Bergholt et al. [29] in the fingerprint region and no spectral changes in the CH-stretching region. The lower spectral differences and deviations in accuracies could be the result from different confocal volumes and a not especially optimized system for measurements of endoscopic biopsy samples. A dependence of the water content between malign and benign tissue was not recognizable in the measured spectra.

One reason for the low accuracy is the presence of a mixture of tissue types within a large sample volume, which is a restriction for fiber optical measurements in general. In contrast confocal Raman microscopy of tissue samples can be used to obtain pure spectra of single tissue components. A better differentiation of spectra from Raman microscopic measurements can be comprehended leading to higher sensitivities, specificities and accuracies of the corresponding classification results [18,40]. This questions the minimum fraction of cancer in the confocal volume of the fiber probe, which is needed to detect cancer. To investigate this issue, we simulated a five component system of mucosa and surrounding tissue, including spectra from crypts, stroma, connective tissue, muscle and cancer. The composition of the surrounding tissue was varied and an integral spectrum of cancer and surrounding tissue was computed and classified. The results are listed in Table 4 and Table S1. We found out that the minimum fraction of cancerous tissue is 51%–62%. In an experiment a lower SNR can hide small bands and therefore increases the minimum fraction for cancer detection. To ensure that no lesions are missed, the sensitivity can be increased at the costs of specificity. This can be comprehended by a receiver operating characteristic (ROC) curve (see Fig. S4) and achieved by e.g. moving the hyper planes in a SVM.

In the present study, we investigated the applicability of a simple compact commercially available system for diagnostics on samples from colonoscopy. By comparison of the different cross-validation schemes, it turned out that due to spectral similarities and non-conformity to the clinical workflow the leave-one-spectrum-out cross-validation is not reasonable to be applied for the recognition of malignant

tissue, since it does not reflect the clinical workflow. In the leave-one-sample-out and leave-one-patient-out cross-validation schemes combined with outlier identification (RoTD) high-risk and low-risk lesions as well as adenocarcinoma and normal tissue can be differentiated quite well. However, adenoma and hyperplastic polyps are barely separable yet. For an improvement of the statistics, more measurements on samples from further patients are needed. Furthermore, we introduced three different outlier identification methods (LOF [33], OCC [37], RoTD [32]) to identify inaccuracies within the data of the measured samples and their labeling. They were applied to improve the training database and by that the recognition of the tissues.

4. Conclusions

The differences in the Raman spectra between the tissue samples are small. By taking the heterogeneity of the patients into account, it is obvious that only a large patient cohort provides a suitable recognition and validation. Hence, a proper data analysis has to be considered, as shown in this study. Due to the biochemical similarity of samples from the same patient, we recommend based on the results shown in Table 3 to use the leave-one-patient-out cross-validation scheme with a sufficient number of patients for validation. There, the classifier has to recognize the spectral fingerprint of a new sample from a new patient for an automatic detection of lesions and not a new spectrum from the same sample. The optimization of classification models by outlier detection can help to avoid the misdiagnosis of single samples and systematic errors in the measurement. Furthermore, such an approach improves the training data base for classification. For pushing Raman endoscopic fiber-optics towards routine clinical application, further efforts are required, especially by increasing the number of patients, since a statistically validation of the classifiers needs a large number of patients, as pointed out previously by Beleites et al. [41]. In the next steps we will improve the setup towards in-vivo measurements and increase the number of patients significantly.

Acknowledgements

We thank Michael Heise for valuable discussions and critical reading of the manuscript. Furthermore we appreciate the efforts of Anja Kögler and Michael Steckstor for sample collection and support for this study. This research was supported by the Protein Research Unit Ruhr within Europe (PURE), and the Ministry of Innovation, Science and Research (MIWF) of North-Rhine Westphalia, Germany (grant number: 233-1.08.03.03-031-68079).

Appendix A. Supplementary Data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.saa.2017.03.054>.

References

- [1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012, *Int. J. Cancer* 136 (2015) E359–E386, <http://dx.doi.org/10.1002/ijc.29210>.
- [2] H. Brenner, J. Chang-Claude, L. Jansen, P. Knebel, C. Stock, M. Hoffmeister, Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy, *Gastroenterology* 146 (2014) 709–717, <http://dx.doi.org/10.1053/j.gastro.2013.09.001>.
- [3] V.S. Benson, J. Patnick, A.K. Davies, M.R. Nadel, R.A. Smith, W.S. Atkin, Colorectal cancer screening: a comparison of 35 initiatives in 17 countries, *Int. J. Cancer* 122 (2008) 1357–1367, <http://dx.doi.org/10.1002/ijc.23273>.
- [4] M.B. Wallace, R. Kiesslich, Advances in endoscopic imaging of colorectal neoplasia, *Gastroenterology* 138 (2010) 2140–2150, <http://dx.doi.org/10.1053/j.gastro.2009.12.067>.
- [5] G.E. Tontini, M. Vecchi, M.F. Neurath, H. Neumann, Review article: Newer optical and digital chromoendoscopy techniques vs. dye-based chromoendoscopy for diagnosis and surveillance in inflammatory bowel disease, *Aliment. Pharmacol. Ther.* 38 (2013) 1198–1208, <http://dx.doi.org/10.1111/apt.12508>.

Table 4

Simulation of spectra from a high confocal volume and mixed tissue component. The minimum fraction of cancer was determined for the identification of the mixed component spectrum as cancerous. The extended results are listed in Table S1.

Composition of additional tissue component (%)				Min. fraction of cancer (%)
Crypts	Stroma	Con. tissue	Muscle	
100	0	0	0	59
50	50	0	0	62
45	45	10	0	59
90	0	5	5	54
80	0	10	10	51
40	40	10	10	55

- [6] S.C. Ng, J.Y. Lau, Narrow-band imaging in the colon: limitations and potentials, *J. Gastroenterol. Hepatol. Aust.* 26 (2011) 1589–1596, <http://dx.doi.org/10.1111/j.1440-1746.2011.06877.x>.
- [7] A.C. Pox, S. Aretz, S.C. Bischoff, U. Graeven, M. Hass, P. Heußner, W. Hohenberger, A. Holstege, J. Hübner, F. Kolligs, K. Wolff Schmiegel, S3-Leitlinie Kolorektales Karzinom Version 1.0, *Z. Gastroenterol.* 51 (2013) 753–854, <http://dx.doi.org/10.1055/s-0033-1350264>.
- [8] C. Hassan, E. Quintero, J.-M. Dumonceau, J. Regula, C. Brandao, S. Chaussade, E. Dekker, M. Dinis-Ribeiro, M. Ferlitsch, A. Gimeno-Garcia, Y. Hazewinkel, R. Jover, M. Kalager, M. Loberg, C. Pox, B. Rembacken, D. Lieberman, Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) guideline, *Endoscopy* 45 (2013) 842–851, <http://dx.doi.org/10.1055/s-0033-1344548>.
- [9] G. Schachschal, M. Mayr, A. Treszl, K. Balzer, K. Wegscheider, J. Aschenbeck, A. Amini, A. Drossel, A. Schröder, M. Scheel, C.-H. Bothe, J.-P. Bruhn, W. Burmeister, G. Stange, C. Bähr, R. Kiesslich, T. Rösch, Endoscopic versus histological characterisation of polyps during screening colonoscopy, *Gut* 63 (2014) 458–465, <http://dx.doi.org/10.1136/gutjnl-2013-304562>.
- [10] L. Rabeneck, L.F. Paszat, R.J. Hilsden, R. Saskin, D. Leddin, E. Grunfeld, E. Wai, M. Goldwasser, R. Sutradhar, T.A. Stukel, Bleeding and perforation after outpatient colonoscopy and their risk factors in usual clinical practice, *Gastroenterology* 135 (2008) 1899–1906.e1, <http://dx.doi.org/10.1053/j.gastro.2008.08.058>.
- [11] J.C. Van Rijn, J.B. Reitsma, J. Stoker, P.M. Bossuyt, S.J. Van Deventer, E. Dekker, Polyp miss rate determined by tandem colonoscopy: a systematic review, *Am. J. Gastroenterol.* 101 (2006) 343–350, <http://dx.doi.org/10.1111/j.1572-0241.2006.00390.x>.
- [12] A. Kallenbach-Thieltges, F. Großbüschkamp, A. Mosig, M. Diem, A. Tannapfel, K. Gerwert, Immunohistochemistry, histopathology and infrared spectral histopathology of colon cancer tissue sections, *J. Biophotonics* 6 (2013) 88–100, <http://dx.doi.org/10.1002/jbio.201200132>.
- [13] D. Petersen, L. Mavarani, D. Niedieker, E. Freier, A. Tannapfel, C. Kötting, K. Gerwert, S.F. El-Mashtoly, Virtual staining of colon cancer tissue by label-free Raman microspectroscopy, *The Analyst* (2017) <http://dx.doi.org/10.1039/C6AN02072K>.
- [14] F. Lyng, I. Ramos, O. Ibrahim, H. Byrne, Vibrational Microspectroscopy for cancer screening, *Appl. Sci.* 5 (2015) 23–35, <http://dx.doi.org/10.3390/app5010023>.
- [15] J.A. De Almeida Chaves Piva, J.L.R. Silva, L.J. Raniero, C.S.P. Lima, E.A.L. Arisawa, C. De Oliveira, R. De Azevedo Canevari, J. Ferreira, A.A. Martin, Biochemical imaging of normal, adenoma, and colorectal adenocarcinoma tissues by Fourier transform infrared spectroscopy (FTIR) and morphological correlation by histopathological analysis: preliminary results, *Rev. Bras. Eng. Biomed.* 31 (2015) 10–18, <http://dx.doi.org/10.1590/2446-4740.0321>.
- [16] T.W. Bocklitz, S. Guo, O. Ryabchikov, N. Vogler, J. Popp, Raman based molecular imaging and analytics: a magic bullet for biomedical applications, *Anal. Chem.* 88 (2016) 133–151, <http://dx.doi.org/10.1021/acs.analchem.5b04665>.
- [17] K. Kong, C. Kendall, N. Stone, I. Notingher, Raman spectroscopy for medical diagnostics - from in-vitro biofluid assays to in-vivo cancer detection, *Adv. Drug Deliv. Rev.* 89 (2015) 121–134, <http://dx.doi.org/10.1016/j.addr.2015.03.009>.
- [18] N. Bergner, T. Bocklitz, B.F.M. Romeike, R. Reichart, R. Kalf, C. Krafft, J. Popp, Identification of primary tumors of brain metastases by Raman imaging and support vector machines, *Chemom. Intell. Lab. Syst.* 117 (2012) 224–232, <http://dx.doi.org/10.1016/j.chemolab.2012.02.008>.
- [19] L. Mavarani, D. Petersen, S.F. El-Mashtoly, A. Mosig, A. Tannapfel, C. Kötting, K. Gerwert, Spectral histopathology of colon cancer tissue sections by Raman imaging with 532 nm excitation provides label free annotation of lymphocytes, erythrocytes and proliferating nuclei of cancer cells, *Analyst* 138 (2013) 4035–4039, <http://dx.doi.org/10.1039/c3an00370a>.
- [20] M. Diem, A. Mazur, K. Lenau, J. Schubert, B. Bird, M. Miljković, C. Krafft, J. Popp, Molecular pathology via IR and Raman spectral imaging, *J. Biophotonics* 6 (2013) 855–886, <http://dx.doi.org/10.1002/jbio.201300131>.
- [21] H. Abramczyk, B. Brozek-Pluska, J. Surmacki, J. Jablonska-Gajewicz, R. Kordek, Raman “optical biopsy” of human breast cancer, *Prog. Biophys. Mol. Biol.* 108 (2012) 74–81, <http://dx.doi.org/10.1016/j.pbiomolbio.2011.10.004>.
- [22] Z. Huang, A. McWilliams, H. Lui, D.I. McLean, S. Lam, H. Zeng, Near-infrared Raman spectroscopy for optical diagnosis of lung cancer, *Int. J. Cancer* 107 (2003) 1047–1052, <http://dx.doi.org/10.1002/ijc.11500>.
- [23] S.P. Singh, A. Sahu, A. Deshmukh, P. Chaturvedi, C.M. Krishna, In vivo Raman spectroscopy of oral buccal mucosa: a study on malignancy associated changes (MAC)/cancer field effects (CFE), *Analyst* 138 (2013) 4175–4182, <http://dx.doi.org/10.1039/c3an36761d>.
- [24] R.O.P. Draga, M.C.M. Grimbergen, P.L.M. Vijverberg, C.F.P.V. Swol, T.G.N. Jonges, J.A. Kummer, J.L.H. Ruud Bosch, In vivo bladder cancer diagnosis by high-volume Raman spectroscopy, *Anal. Chem.* 82 (2010) 5993–5999, <http://dx.doi.org/10.1021/ac100448p>.
- [25] M.A. Short, S. Lam, A.M. McWilliams, D.N. Ionescu, H. Zeng, Using laser Raman spectroscopy to reduce false positives of autofluorescence bronchoscopies: a pilot study, *J. Thorac. Oncol.* 6 (2011) 1206–1214, <http://dx.doi.org/10.1097/JTO.0b013e3182178ef7>.
- [26] M. Jermyn, K. Mok, J. Mercier, J. Desroches, J. Pichette, K. Saint-Arnaud, L. Bernstein, M.-C. Guiot, K. Petrecca, F. Leblond, Intraoperative brain cancer detection with Raman spectroscopy in humans, *Sci. Transl. Med.* 7 (2015) 274ra19, <http://dx.doi.org/10.1126/scitranslmed.aaa2384>.
- [27] I. Latka, S. Dochow, C. Krafft, B. Dietzek, J. Popp, Fiber optic probes for linear and nonlinear Raman applications - current trends and future development, *Laser Photonics Rev.* 7 (2013) 698–731, <http://dx.doi.org/10.1002/lpor.201200049>.
- [28] J.J. Wood, C. Kendall, J. Hutchings, G.R. Lloyd, N. Stone, N. Shepherd, J. Day, T.A. Cook, Evaluation of a confocal Raman probe for pathological diagnosis during colonoscopy, *Color. Dis.* 16 (2014) 732–738, <http://dx.doi.org/10.1111/codi.12664>.
- [29] M.S. Bergholt, K. Lin, J. Wang, W. Zheng, H. Xu, Q. Huang, J. Lin Ren, K.Y. Ho, M. Teh, S. Srivastava, B. Wong, K.G. Yeoh, Z. Huang, Simultaneous fingerprint and high-wavenumber fiber-optic Raman spectroscopy enhances real-time in vivo diagnosis of adenomatous polyps during colonoscopy, *J. Biophotonics* 9 (2016) 333–342, <http://dx.doi.org/10.1002/jbio.201400141>.
- [30] J.C.C. Day, R. Bennett, B. Smith, C. Kendall, J. Hutchings, G.M. Meaden, C. Born, S. Yu, N. Stone, A miniature confocal Raman probe for endoscopic use, *Phys. Med. Biol.* 54 (2009) 7077–7087, <http://dx.doi.org/10.1088/0031-9155/54/23/003>.
- [31] A. Lazarevic, L. Ertöz, V. Kumar, A. Ozgur, J. Srivastava, A comparative study of anomaly detection schemes in network intrusion detection † 2 evaluation of intrusion detection systems, *Proc. SIAM Conf. Data Min.* (2003) 25–36.
- [32] V.J. Hodge, J. Austin, A survey of outlier detection methodologies, *Artif. Intell. Rev.* 22 (2004) 85–126, <http://dx.doi.org/10.1007/s10462-004-4304-y>.
- [33] M.M. Breunig, H.-P. Kriegel, R.T. Ng, J. Sander, LOF: identifying density-based local outliers, *Proc. 2000 Acm Sigmod Int. Conf. Manag. Data* 2000, pp. 93–104, <http://dx.doi.org/10.1145/335191.335388>.
- [34] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Baseline correction using adaptive iteratively reweighted penalized least squares, *Analyst* 135 (2010) 1138–1146, <http://dx.doi.org/10.1039/b922045c>.
- [35] H.C. Peng, F.H. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (2005) 1226–1238, <http://dx.doi.org/10.1109/TPAMI.2005.159>.
- [36] J. Ollesch, M. Heinze, H.M. Heise, T. Behrens, T. Brüning, K. Gerwert, It's in your blood: spectral biomarker candidates for urinary bladder cancer from automated FTIR spectroscopy, *J. Biophotonics* 7 (2014) 210–221, <http://dx.doi.org/10.1002/jbio.201300163>.
- [37] M. Kemmler, E. Rodner, E.S. Wacker, J. Denzler, One-class classification with Gaussian processes, *Pattern Recogn.* 46 (2013) 3507–3518, <http://dx.doi.org/10.1016/j.patcog.2013.06.005>.
- [38] C. Krafft, B. Dietzek, M. Schmitt, J. Popp, Raman and coherent anti-stokes Raman scattering microspectroscopy for biomedical applications, *J. Biomed. Opt.* 17 (2012) 040801, <http://dx.doi.org/10.1117/1.JBO.17.4.040801>.
- [39] A. Rygula, K. Majzner, K.M. Marzec, A. Kaczor, M. Pilarczyk, M. Baranska, Raman spectroscopy of proteins: a review, *J. Raman Spectrosc.* 44 (2013) 1061–1076, <http://dx.doi.org/10.1002/jrs.4335>.
- [40] N. Stone, C. Kendall, N. Shepherd, P. Crow, H. Barr, Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers, *J. Raman Spectrosc.* 33 (2002) 564–573, <http://dx.doi.org/10.1002/jrs.882>.
- [41] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, J. Popp, Sample size planning for classification models, *Anal. Chim. Acta* 760 (2013) 25–33, <http://dx.doi.org/10.1016/j.aca.2012.11.007>.