

Clinical application of infrared fibre-optic probes for the discrimination of colorectal cancer tissues and cancer grades



Julian Ollesch^{a,1}, Michael Zaczek^{b,1}, H. Michael Heise^a, Oliver Theisen^a,
Frederik Großerüschkamp^a, Ralf Schmidt^b, Konrad Morgenroth^c, Stathis Philippou^c,
Matthias Kemen^b, Klaus Gerwert^{a,*}

^a Department of Biophysics ND04-596, Ruhr-University Bochum, Universitätsstrasse 150, 44780 Bochum, Germany

^b Department of Surgery, Protestant Hospital Herne, Centre of Colon and Rectal Cancer, Certified by the German Cancer Society (DKG), Wiescherstrasse 24, 44623 Herne, Germany

^c Institute of Pathology, Augusta Hospital, Zeppelinstrasse 18, 44791 Bochum, Germany

ARTICLE INFO

Article history:

Received 10 March 2016

Received in revised form 10 June 2016

Accepted 8 July 2016

Available online 9 July 2016

Keywords:

FTIR-spectroscopy

ATR fiber-optics

Colorectal cancer

Random forest

Linear discriminant analysis

Multivariate classification

Visceral surgery

ABSTRACT

Infrared spectral histopathology is a well-established method for label-free tissue classification. Flexible fibre-optic probes allow a remote sensing for *in-vivo* tissue annotation. The performance of infrared spectral analysis of colorectal tissue with millimetre spatial resolution for cancer lesion identification and grading within a surgical setting was assessed. Colorectal tissue was removed during routine therapeutic surgery. By consecutive spectroscopy, luminal positions within and outside the cancer lesion were analysed using fibre-coupled probes for attenuated total reflection (ATR) measurements using a diamond-prism or cone, respectively. Subsequent routine histopathology provided the gold standard for diagnosis and grading. For spectral data analysis, two feature selection algorithms were applied. Results from linear discriminant analysis and ensemble random forest classifiers based on leave-one-third-out cross-validation and test-set validation with independent data are presented. The spectral discrimination of tumour versus normal tissue under the cross-validation scheme was achieved with an accuracy of $90 \pm 5\%$ (sensitivity of $89 \pm 7\%$, specificity of $90 \pm 7\%$), whereas respective test-set validation led to an accuracy of 80% (sensitivity of 83% and specificity of 78%). Low versus high tumour grading was assessed under cross-validation with an accuracy of $81 \pm 8\%$ (sensitivity of $80 \pm 16\%$, and specificity of $81 \pm 14\%$). Thus, fibre-optic infrared spectroscopic tissue analysis has the potential of supporting clinical decisions by providing immediate tissue and grading information during surgery.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Colorectal cancer is not only of grave impact to our global society [1,2], but also one of the treatable cancers with a comparably high five-year survival rate when discovered early [2,3]. Thereby, two measures influence the following therapy strategy. Tumour staging, as denoted in a TNM classification [4], refers to the spread within an organ (T), the spreading into lymph nodes (N), the infiltration into the lymphatic system (L) and into blood vessels (V), and the presence of metastases (M). In contrast, tumour grading refers to the de-differentiated appearance of cells inside the tumour tissue. The discrimination of low (G1 and G2)

and high grade cancer (G3 and G4) with the particular grades of G2 (still low, minor to moderately de-differentiated cells) and G3 (yet high, severe de-differentiation) is of particular interest for therapy and aftercare [3].

Fourier-transform infrared (FTIR) spectroscopy has been applied for clinical chemistry and for support in medical diagnosis [5–7]. As frequently manifested, the infrared absorbance spectrum of a patient's bio-specimen represents a fingerprint-like integral biochemical status with the advantage that no additional markers or labelling are required. Furthermore, multiplex parameters of the proteome, lipidome, and metabolome are recorded in one comprehensive spectrum. By means of bioinformatics, a label-free and fast measurement from the bio-specimen can provide discriminative spectral band patterns for various diseases [8–18], thus allowing supportive diagnostics for a patient's health status.

Evidence that infrared spectra principally bear marker patterns for the discrimination of colorectal cancer from histologically

* Corresponding author.

E-mail address: gerwert@bph.rub.de (K. Gerwert).

¹ J.O. and M.Z. share first authorship.

normal tissue has been reported several times. So far, tissue and cell culture studies using ground and quartz-pelleted material within a diamond anvil cell [19], spectral histopathology (SHP) on thin slices [20–24], and conventional FTIR-spectroscopy on small, but complete tissue biopsies based on the ATR-technique [25] have been published. A high accuracy of cancer grading was recently demonstrated by a hierarchical FTIR-SHP system on formalin fixed paraffin embedded (FFPE) tissue slices [26]. Although grading is a morphological observation on the de-differentiation of cells in tumorous tissue, the recent SHP work proves that tumour progression is also associated with biochemical changes. Thus, it is detectable by infrared spectroscopy.

Generally, FTIR- and Raman-SHP are proven, highly accurate label-free methods for tissue classification [6,24,27–29]. Nowadays, vibrational spectra of a thin tissue slice are routinely recorded microscopically with high spatial resolution. By spectral similarity, tissue and disease states are classified based on the biochemical fingerprints of the sample. Due to the high spatial resolution, a data-cube of spectral and area coordinates containing millions of spectra is easily recorded from a single tissue slice [6,7,24,27,28]. Handling and processing of such data requires an advanced computing infrastructure. Data acquisition and interpretation may be disengaged from any subjective assessment [30], but the time required for spectral recording and classification is an essential key element for *ad hoc* clinical applicability.

Here, we present a compromise approach with regard to the reduction of the data volume and measurement/processing time, thus gaining a far greater instrumental flexibility by a reduced spatial resolution using ATR-probes with a 4–6 mm²-size sensing area. In the past, several approaches of fibre-optic probes for tissue characterisation with single spot measurements or for mapping larger tissue areas have been manifested, with a technique also termed evanescent wave spectroscopy [31]. Such studies were performed on skin [32,33], bladder and human colon carcinoma tissue [34,35]. Further advances in recent years led to applications of fibre-coupled ATR-spectroscopy for dys- and hyperplasia detection in colonoscopy-obtained human biopsy samples [36], or for detection of areas of colonic inflammation in a mouse model [37]. Just recently, such a system was applied to discriminate normal and cancer tissue regions from macroscopic human colorectal tissue samples of radical resections [38]. It was concluded, that fibre-optic spectroscopic colonoscopy could be advanced into the operating room.

Towards this aim, we consequently investigated the application of two standard-sized fibre-coupled ATR-probes on fresh colorectal tumour resectates within the clinical environment. Two compact FTIR-spectrometers were set up in the pathology department and directly adjacent to the surgical theatre. The sampling protocols were optimized to meet clinical demands. Thus, a rapid spectral characterization of freshly resected and unfixed colorectal tissues, similar to tissue as presented during colonoscopy, was achieved. The prediction of tumour regions was validated using a second independent dataset. Thus, two discriminative spectral feature sets and three classifiers obtained from the calibration data could be assessed under real-life conditions, as the validation spectra were neither included in the feature selection nor considered for classifier training.

Our results are based on advanced statistics for spectral feature selection and classification. Leave-one-third-out Monte Carlo cross-validations and independent test-set validations were performed. Two different algorithms for the identification of spectral features, discriminating between cancer lesion and healthy tissue, and for respective grade G2 and G3 differentiation were applied. Furthermore, two different classifiers were used to determine the predictive performance.

Besides spectral data pre-processing and feature selection, also technical challenges, such as avoiding sample carryover and probe cleaning between spectroscopic measurements have been addressed in our study. It is meant to advance the availability of spectral histopathology into the surgical, intra-operative environment.

As a different approach, a confocal Raman probe, fitting in principle to the accessory channel of a colonoscope, was recently developed and evaluated with *ex-vivo* samples [39]. However, similarly miniaturized fibre probes are not yet available for mid-infrared spectroscopy; despite this, the specific advantages and drawbacks of FTIR and Raman fibre technologies will be discussed. With our work, progress using the evanescent-wave FTIR-approach as applied for a clinically relevant *ex-situ* analysis of colorectal tissues can be reported.

2. Experimental

The workflow from surgical tissue resection to Monte Carlo leave-one-third-out cross-validated classification is schematically illustrated in Fig. 1. A strict standard operating procedure was followed to ensure maximum reproducibility of sample handling and spectrum collection. Colorectal cancer tissue was surgically removed according to actual clinical standard procedures. Absorbance infrared spectra were collected from tissue areas of the washed colon lumen. All samples that were characterised by spectroscopy underwent a routine histopathological assessment by expert pathologists, who also provided the grading information, as being the diagnostic gold standard.

2.1. Patient collective and sample preparation

Colorectal resectates of 57 patients, of these 25 women and 32 men with matching age distributions, were studied (Table 1). All patients had given their informed consent (ethics vote 4453-12, Ethics Commission of the Ruhr-University Bochum, Germany). Of these, 53 patients were diagnosed with tumour grades G2 and G3, which were histopathologically assessed by means of common practice [3]. The tissue samples were resected during the patient's regular therapy at the Protestant Hospital Herne (Herne,

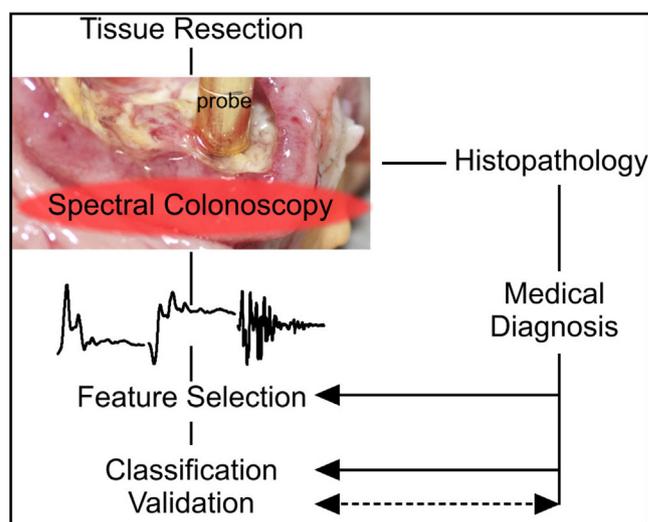


Fig. 1. Scheme for the analysis of colorectal resectates: Spectra of healthy tissue and cancerous lesions were recorded. Subsequently, first and second derivative spectra were calculated and concatenated to form a representative conjunct spectral vector of the biochemical tissue status. Finally, the class assignment obtained by the medical gold standard of histopathology was used for feature selection, classification and validation.

Table 1

Patient age and gender distribution: Four patients of the calibration set were diagnosed with grade G1 colorectal carcinoma, only spectra of control regions were used. One male patient of the validation dataset was diagnosed with a grade G3 tumour, all others with grade G2.

A) calibration dataset			
	male	female	total
# all grades	18	19	37
age $\pm \sigma$	72 \pm 7	70 \pm 13	71 \pm 11
# grade G2	9	10	19
age G2 $\pm \sigma$	71 \pm 9	71 \pm 15	71 \pm 14
# grade G3	7	7	14
age G3 $\pm \sigma$	73 \pm 6	68 \pm 12	70 \pm 10
B) validation dataset			
	male	female	total
# all grades	14	6	20
age $\pm \sigma$	73 \pm 11	68 \pm 12	71 \pm 11

Germany). A total of 19 resectates were analysed at the Institute of Pathology, Augusta Hospital Bochum, whereas spectra of 38 resectates were acquired directly adjacent to the surgical theatre. An influence of resectate transportation on water ice (elapsed time < 30 min) on the recorded spectra could not be detected. After removal of the resectates, spectra were typically acquired either within 45 min at the Pathology Department, or within 30 min next to the surgical theatre.

The 138 spectra recorded of the first recruited 37 patients were analysed as a calibration dataset, whereas 122 spectra recorded from 20 patients enlisted later were considered as validation dataset (Table 1). As feature selection and cross-validation was exclusively performed on the calibration data, the validation data represents a truly independent test set.

2.2. FTIR-spectroscopy using fibre-optic probes

Measurements were performed at room temperature with two sealed spectrometers with externally attached fibre-optic probes.

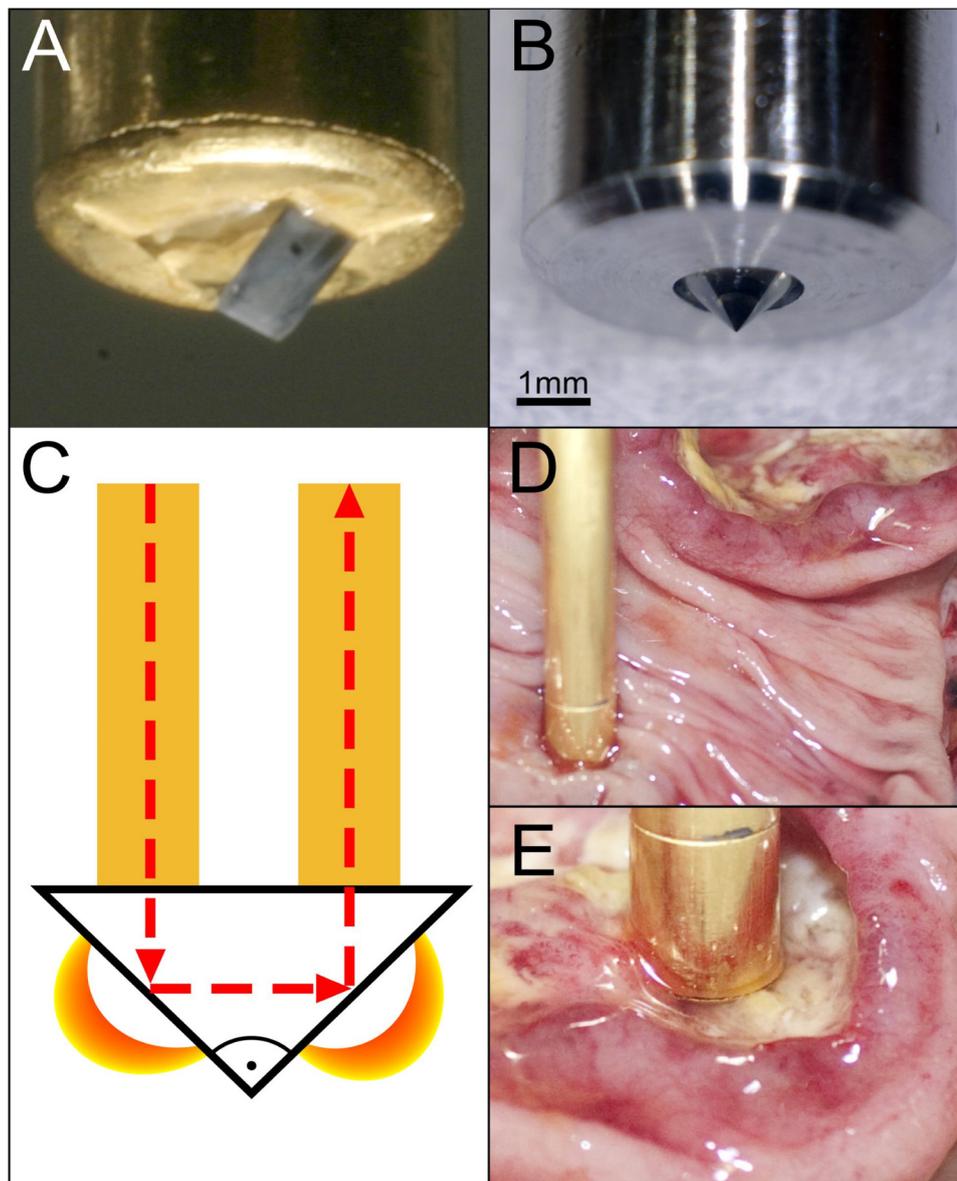


Fig. 2. Two probe heads were used in the study consisting either of a diamond micro-prism (A), or of a cone-shaped diamond tip (B). In both fibre-optic probes, infrared radiation is waveguided to the ATR element by core-cladded silver halide fibres to obtain the sample spectrum with two reflections (C). The probe head was positioned onto healthy tissue with at least 20 mm distance from malignant tissue (D), or directly onto the tumour lesion (E). Grading information was obtained from histopathology.

First, a DiProbe/Upgrade with a 6 mm diameter shaft, provided by Infrared fiber sensors (Aachen, Germany), was coupled to a Tensor 27 spectrometer (Bruker Optics GmbH, Ettlingen, Germany) using an external pigtailed MCT-detector. The ATR-probe head consisted of a diamond micro-prism for sampling with two reflections (Fig. 2A). A second ATR-probe was acquired from A.R.T. Photonics GmbH (Berlin, Germany), featuring a cone-shaped diamond ATR element (Fig. 2B), coupled to a Matrix MF spectrometer (Bruker Optics), equipped with a high-sensitivity MCT detector. All detectors were liquid N₂ cooled.

With both devices, interferograms were sampled in a double-sided forward-backward interferometer motion. Because of measurement time restrictions within the clinical setting, instrument parameters had to be adjusted. Using the Tensor 27 spectrometer, 96 scans of 8 cm⁻¹ resolution were accumulated in approximately 80 s before Fourier-transformation using Blackman-Harris 3-Term-apodization, Mertz phase correction and 4-fold zero filling. The Matrix MF spectrometer, optimized for fibre-optic attachments, allowed 64 interferogram scans of 4 cm⁻¹ resolution to be recorded in approximately 50 s. Spectra were calculated as power spectrum, but with otherwise identical parameters.

Colorectal tissue resectates were opened, washed with tap water to remove remaining faeces, and spread out. The tissue surface was rinsed and kept wet with PBS buffer (100 mM NaCl, 150 mM NaPO₄, pH 7.2), which also prevented sample adhesion to the probe. At least two spectra were acquired outside and inside the tumour lesion area per tissue sample, retaining a minimum distance of at least 20 mm from the visible lesion for the control position. After the last measurement on each resectate, the probe was rinsed with buffer solution and cleaned by immersion in approx. 0.5% sodium hypochloride solution (1:9 distilled water dilution of Dan Klorix, Colgate-Palmolive GmbH, Hamburg, Germany) for 2 min at room temperature, followed by extensive rinsing with distilled water.

2.3. Spectral preprocessing

Spectral variables were set by interpolation to a grid of 2 cm⁻¹ spacing. A Fourier-transform based low-pass filter with an appropriate Gaussian apodization function removed noise leaving components broader than or equal to 4 cm⁻¹ of full-width at half-height (FWHH) as reported in earlier studies [18,40,41]. The spectral range of 1666–1000 cm⁻¹ was selected and min-max-normalized. To unveil possibly subtle band shifts, also first and second derivatives were calculated by Fourier expansion according to a Gaussian function of 8 and 10 cm⁻¹ halfwidth for low-pass filtering, respectively. The set of first derivative spectra was linearly scaled up by one set-specific factor to maximum absolute amplitude of 0.5, while maintaining interspectral differences. Similarly, the set of second derivative spectra was consistently scaled up to a respective maximum absolute amplitude value of 0.25. Finally, a synthetic conjunct spectral vector, containing absorbance, first and second derivative values, was concatenated as described earlier (see Figs. 1, 5 and 7) [17,18]. Using these steps, a consistent spectral dataset was obtained from both spectrometer systems.

2.4. Bioinformatics environment

Random forest calculations were performed with the Matlab software, Version 2014a with the R-project based [42] Matlab port (as found on <http://code.google.com/p/randomforest-matlab/>, January 30, 2013) on a High-Performance Computing Server Supermicro SYS-5086 B with 8 x Intel® Xeon® Westmere EX (E7-8837, 2.66 GHz, 8-Core) and 512 GB RAM. Linear discriminant

analysis (LDA) was carried out using the internal Matlab function ('classify') with a linear discriminant function. The *a priori* class membership probability was empirically calculated. Final predictor training was performed on an office PC equipped with Intel Core2Quad CPU, 8 GB RAM (Dell Optiplex 780) running Matlab 2012a.

2.5. Spectral feature selection

As mentioned above, two established algorithms for feature selection were applied to the spectral data: one relatively fast algorithm termed "maximum relevance, minimum redundancy" (MRMR), and a random forest (RF) approach by excluding low-importance-features iteratively. Both algorithms have proven to identify discriminative spectral patterns in FTIR-spectroscopic datasets for disease pattern recognition [17,18].

The MRMR algorithm (downloadable at <http://www.mathworks.com/matlabcentral/fileexchange/14916-minimum-redundancy-maximum-relevance-feature-selection>, February 15, 2016) [43,44] was used to identify and rank the top 40 discriminative features of the dataset. By further iterative cross-validation, the ranking was step-wise reduced to include a growing number of most discriminative features in the calculation. Based on the average accuracy of 1000 leave-one-third-out Monte Carlo cross-validation (MCCV) calculations with an LDA classifier, a set of optimally classifying features was determined.

A previously proven random forest approach was used for feature selection by iterative exclusion of uncorrelated variables [17,18]. Each RF was trained and validated using an individually MC derived data subset. Our MC algorithm arranged the same number of spectra per class for the validation dataset. The total dataset of 138 spectra (73 cancer lesions, 65 control areas) was split into 200 randomly selected sets of 126 (67 cancer lesions, 59 controls), leaving 9% of the samples out for variation. On each of these 126 datasets, 192 further training-validation pairs of 48 training and 19 testing spectra of cancer, and 40 training and 19 testing spectra of control spots were generated at random. In total, 192 RFs were trained and tested with these groups, the average classification error rate was registered, and the Gini importance values of the features used in the 192 RFs were accumulated. After removal of the 20% least important features, 192 further RFs were evaluated on 192 new MC based cross-validation datasets, until iteratively only four features remained. The set of features producing the lowest misclassification rate was selected as optimum set. After this processing, further 199 cycles were calculated using the remaining 199 randomly selected groups of 126 spectra. Thus, a ranking of classification-important features was obtained on the strength of the selection frequency in subsequent calculation cycles. Similarly to the MRMR based procedure, the discriminative quality of feature sets was assessed – using a 1000 fold LDA MCCV – by stepping down the selection frequency for determining the ranking threshold based on average classification accuracy.

Grading discrimination was performed on 61 spectra with 29 cases of grade G2 and 32 of grade G3. The MRMR feature selection was performed in an identical manner. For RF feature selection, the dataset was split into 200 sets of 55 spectra (26 G2, 29 G3) with six each left out for variation, before split further into sets of 17 G2 and 20 G3 for training, and 9 each for testing.

2.6. Classification with cross-validation

Straight LDA and complex ensemble RF classifiers as algorithms with different computational requirements were applied in strict MCCV schemes leaving a number of 33% (a third) of the smallest populated class, per class out for independent validation as previously reported [17,18]. The prediction of the ensemble RF

classifier was by majority vote of 1001 random forests, which were each trained on individual MC-derived data subsets.

2.7. Validation with independent test set

For test-set validation, the spectral dataset of 122 spectra recorded from the last recruited 20 patients was reduced to the discriminative spectral features as identified on the calibration dataset of the first 37 patients. One LDA classifier and a second ensemble RF classifier were trained on each of the MRMR- and RF-selected features of the complete calibration dataset. For prediction-stability assessment, 50 further ensemble RF classifiers were trained on MC-derived 2/3rd subsets of the calibration dataset. The entire validation dataset of 122 spectra was used for predictions by the classifier functions, providing one prediction each for the LDA and the individual ensemble RF systems, allowing average values \pm standard deviation to be calculated for the 50 RFs, as summarized in Table 7.

3. Results and discussion

Collection of infrared spectra from the resected tissue surface with the fibre optical probes was performed in a straightforward manner. In previous studies, the spectral influence of water on the sampled surface was reduced by air-drying. Whereas this step was efficient in the reduction of external water absorption, the sample inner-water strongly contributed to the recorded spectra [38]. As a clinically acceptable alternative, we present sampling procedures compatible with colonoscopy, *i.e.*, accepting wet tissue surfaces. The quality of the histological examination due to the non-invasive spectral analysis was unaffected and undisputed.

Due to the sealed housing of the spectrometers and the use of flexible, firmly spectrometer-attached probes, spectral contributions of atmospheric water vapour were not apparent, as it would usually be an issue with spectrometers using internal sample compartments.

Equivalent spectra were recorded with both fibre-optic probes, as demonstrated with the exemplary absorbance spectra of distilled water (Fig. 3). As expected, system specific spectral contributions were not observed, as both were using two internal reflections within the diamond ATR-elements. With the improved spectral quality provided by the Matrix MF/A.R.T. Photonics system, a faster measurement with a fewer number of co-added interferograms was possible.

A manifested colorectal cancer is most appropriate to assess the SHP application of fibre-optic FTIR-spectroscopy. The tumour can

often reliably be localized before histopathological examination. The resected tissue size of a hollow organ renders it well-accessible to the probe head. Every colorectal resectate includes an intrinsic control sample, because typical resection margins contain a sufficient amount of non-malignant tissue for the spectroscopic assessment. Therefore, the current aim of demonstrating clinical applicability in a surgical setting could be reached with a limited number of samples, for which a set of representative spectra was obtained.

3.1. Endoscopy compatible preparation

To achieve optimum applicability, the sample preparation for the spectral recordings was kept close to the expectable clinical prerequisites of colonoscopy. Thorough chemical and mechanical cleaning of the probe head can easily be achieved before each single measurement within a laboratory environment, but a prolongation of the stress on a patient during an extended inspection is infeasible. Every manipulation bears the risk of complications and prolongs the diagnostic procedure. Nevertheless, faeces and slime in the lumen have to be removed, both for a clear visual inspection and for the spectroscopic measurement.

The clinical standard procedure includes a laxative treatment with copious amounts of liquid beginning one day prior to the examination. The additional level of cleanliness required for spectroscopy can be achieved by a serial rinse by water and PBS buffer solution. Such cleansing reduced any sample carryover substantially, which is a major issue on dried tissue surfaces (Fig. 4).

With freshly cut bovine muscle tissue as a model, the ATR-probe showed residues with major spectral signatures of the last contact spot even after lifting the probe off the sample (Fig. 4A, (i) and (ii)). The amide I band (1644 cm^{-1}) of spectrum (i) is naturally more intense than those of traces (ii–v) because of the underlying H_2O -bending vibration band of a wet tissue sample. A probe contamination was still apparent when the sample was rinsed and patted dry before the measurement (iii). For removing the contamination, an alcohol (methanol) rinse (iv) or cleaning with methanol and water (v) were insufficient. A clean probe (vi) was restored after 2 min incubation with sodium hypochlorite solution, followed by distilled water rinsing.

Performing the measurement with a buffer-coated sample surface minimized the substance carryover at the probe head (Fig. 4B). Apart from the dominating amide I and H—O-bending vibrations, a protein amide II band was observable upon tissue contact (i). When lifted off, the probe exhibited only a minimal contamination signal (ii) as compared to the pure PBS buffer spectrum (iii). After five tissue contacts at different locations, the difference spectrum against pure PBS indicated only a negligible spectral contribution of adhered material (iv).

Muscle tissue is definitely only to be regarded as a model for the intended application of the probing system to colorectal mucosa. Apart from histological differences, the muscle tissue was cut – thus damaged – to obtain spectra; by this, the tissue extracellular matrix was exposed and led to probe contamination. However, the model experiment indicated that even after removal of released, soluble and superficial substances, substantial carryover is likely (Fig. 4A, trace (iii)).

A colorectal carcinoma is mostly manifested in a zone of corrupted mucosa. Thus, contamination avoidance must not be neglected, as spectra of such a region were used for tumour detection and grading. In the usual clinical setting, initially clean and sterile instruments are used on a patient for one procedure, and then discarded or put aside for afterwards sterilization. With our protocol, an initially clean and dry probe head was used on the whole tissue sample, kept wetted with PBS buffer, of one patient.

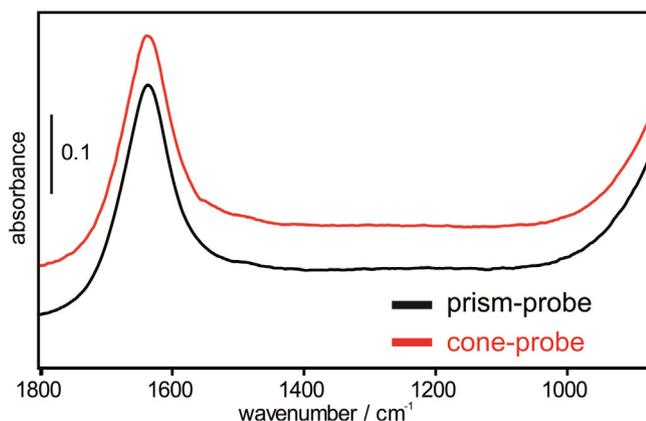


Fig. 3. The diamond ATR element shape had no apparent effect on the recorded spectra, as indicated by a H_2O spectrum recorded with both probes on the Matrix spectrometer (upper spectrum offset for clarity).

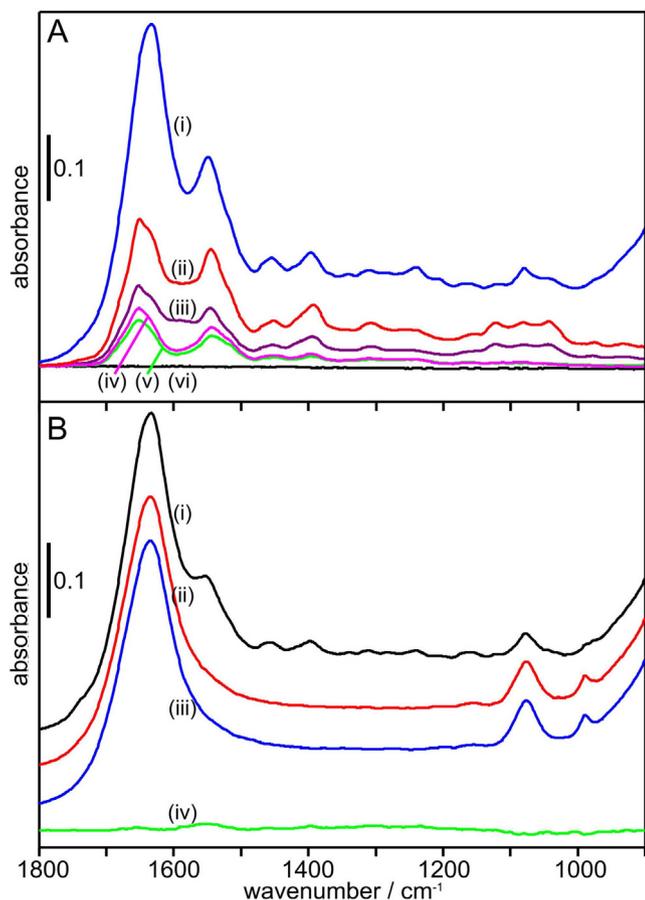


Fig. 4. A buffer solution coated tissue surface minimized sample carry-over. (A) Spectra as measured against a clean and dry probe reference: freshly cut bovine muscle tissue (i), after lifting the probe off the sample (ii), after lifting the probe off the sample, which was rinsed with PBS buffer solution and patted dry before measurement (iii), after procedure (ii) and probe head was immersed in methanol and dried (iv), after procedure (ii) and probe head immersed in methanol and rinsed with H₂O and dried (v), after procedure (v) and probe head was immersed in sodium hypochloride solution, rinsed with H₂O, and dried (vi). (B) Optimised procedure: spectra as measured against a clean and dry probe reference: freshly cut bovine muscle tissue, but rinsed and covered with PBS buffer (i), after lifting off the probe from the sample (ii), pure PBS buffer (iii), difference absorbance spectrum after five sample contacts versus pure PBS buffer (iv) (spectra in subplot (B) were offset for clarity).

After collection of the last spectrum, a thorough disinfection with hypochloride solution was performed to cleanse the probe head before analysis of the next patient's resectate.

3.2. Spectral separability of healthy and cancer tissue

For assessing the quality of spectral discrimination between malignant and control tissue areas, two dedicated feature selection algorithms determined respective sets of discriminative spectral variables for an accurate tissue status prediction with two entirely different classifiers. The calculations were based on absorbance spectra, first derivatives, or second-derivative spectra. Each data vector consisted of 333 wavenumber-intensity pairs covering the interval of 1666–1000 cm⁻¹. Additionally, a synthetic conjunct feature vector comprising of all three spectral vectors was generated, which proved advantageous in previous research [17,18].

An overview of spectral vectors exhibited a region of 1600–1500 cm⁻¹ as probably discriminative for the cancer lesion against the healthy tissue surface (Fig. 5A). Consistently, discriminative features were identified in this region of the absorbance spectra by both algorithms (Fig. 5, feature set traces (i, ii, vii, viii)). Additional features at lower wavenumbers were identified within the second derivative spectra (Fig. 5C traces (v, vi, vii, viii)), as it was the case with calculations based on first derivative spectra alone (Fig. 5B (iii, iv)). The derivative spectra were not re-normalized after derivation; therefore a visual identification of features, particularly in first-derivative spectra, is difficult. However, both the MRMR and the RF algorithm identified discriminative patterns in all separate single spectral sets. When applied to the synthetic dataset of combined absorbance, first and second derivatives, the importance-weighting algorithms ranked the features for interregional discriminative power, with the result of neglecting all first derivative features. This attributed a higher discriminative power to absorbance and second-derivative spectral features. Nevertheless, it should not be generalized that the first-derivative spectra can be ignored within a synthetic conjunct spectral vector, because its inclusion proved to be advantageous for other datasets [17,18].

The MRMR algorithm indicated spectral variables at approximately 1500 cm⁻¹ in the absorbance spectra, and between 1550 and 900 cm⁻¹ in the second-derivative data. The RF algorithm confirmed the discriminative importance, although a number of features were added and found shifted in the second-derivative region. Particularly, the approximate band positions of 1100, 1350, 1550 cm⁻¹ agree with reported candidate marker bands for malignant tissue that were attributed to phosphoric compounds [19]. However, due to the overlap of spectral bands caused by numerous, not further definable substances in normal and – after all – malignant mucosa, a reliable assignment of certain spectral bands to a specific substance is not possible. Despite this, differences at approximately 1550 cm⁻¹ indicate different protein compositions of the sampled areas with a high probability, but pinpointing to one specific marker component is impossible within the FTIR-spectrum.

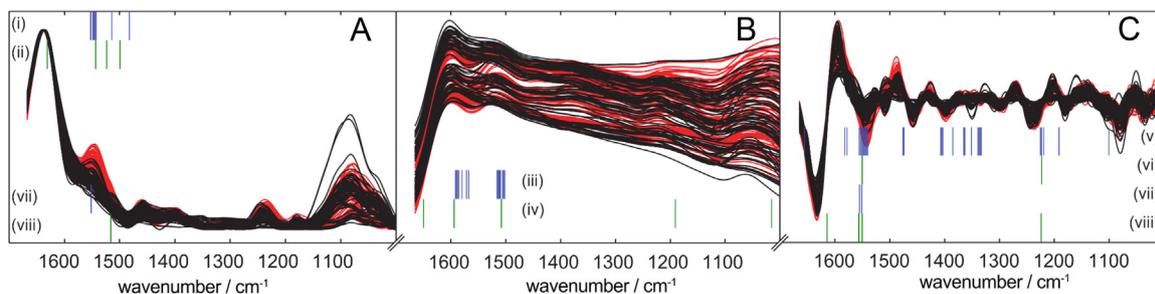


Fig. 5. Spectral data of control (black) and cancerous (red) regions. The training dataset consisting of absorbance (A), first (B) and second derivative spectra (C) exhibited visually discriminative regions. Mostly, these spectral features were confirmed by the feature selection algorithms (blue RF, green MRMR); the bar-wise displayed features are the results of calculations on spectral regions (A), (i, ii); (B), (iii, iv); (C), (v, vi), and (A + B + C), (vii, viii). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

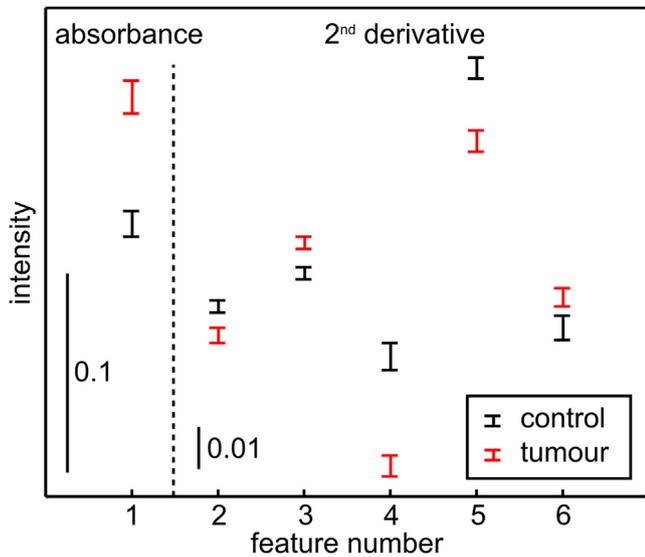


Fig. 6. Control and cancerous tissue regions were discriminated with highest accuracy based on six features, which were identified by the MRMR algorithm (absorbance: 1518 cm^{-1} ; 2nd derivative: 1219, 1474, 1554, 1580, 1614 cm^{-1} ; compare with Fig. 5 and Table 2). Displayed are class-wise 95% confidence intervals of the feature intensity mean values.

Fig. 6 illustrates the 95% confidence intervals of the average intensity values for the control and cancerous tissue classes at the MRMR-selected spectral features. The data intervals do not overlap, which is an indication of class separability. This is confirmed and quantified in the following cross- and test-set-validation.

3.3. Cross-validation

In averaged 1000 MC cross-validations, the LDA classifier achieved an accuracy of at least $84 \pm 6\%$ on all determined feature sets (Table 2). This minimum result was obtained with MRMR-selected features on first derivative spectra, depicting a sensitivity of $84 \pm 9\%$, and a specificity of $83 \pm 9\%$. With RF-selected features of the first derivative spectra, as well as with the selected features of absorbance, second derivative spectra and the concatenated spectral vectors, significantly more accurate classifications were achieved, as indicated by two-sided *t*-tests assuming heterogeneous variances with $p < 0.001$. Best results of the LDA classifier were achieved with an accuracy of $90 \pm 5\%$ based on 42 RF-selected features of second derivative spectra with a sensitivity of $92 \pm 6\%$ and a specificity of $87 \pm 7\%$ (see Fig. 5B (iii)), and by calculations

Table 2

Average \pm standard deviation performance data of 1000 LDA classifiers on optimum feature sets of absorbance only (abs), first derivative only (1st deriv), and second derivative spectra (2nd deriv) of surface absorbance spectra in comparison with the performance on features selected from the total concatenated absorbance-1st derivative-2nd derivative spectral vector for the discrimination of control and cancerous tissue (cyc: threshold of cycles of RF feature selection or features derived from MRMR calculation, #f: number of features, acc:% accuracy, sens: % sensitivity, spec: % specificity).

	cyc	#f	acc	sens	spec
abs	$\geq 199/200$	8	86 ± 5	88 ± 8	84 ± 8
abs	MRMR	2	87 ± 5	88 ± 7	87 ± 8
1st deriv	$\geq 177/200$	6	87 ± 5	87 ± 8	87 ± 7
1st deriv	MRMR	18	84 ± 6	84 ± 9	83 ± 9
2nd deriv	$\geq 111/200$	42	90 ± 5	92 ± 6	87 ± 7
2nd deriv	MRMR	3	89 ± 5	88 ± 7	89 ± 7
total set	$\geq 196/200$	17	88 ± 5	88 ± 8	88 ± 9
total set	MRMR	6	90 ± 5	89 ± 7	90 ± 7

with six MRMR-selected features of the synthetic spectral vector resulting in a sensitivity of $89 \pm 7\%$ and a specificity of $90 \pm 7\%$ (see Fig. 5 (viii)). As indicated by identical *t*-test parameters, the accuracy was significantly improved compared with the next lower observed value of $89 \pm 5\%$. Weighting the specificity more important for diagnostic purposes, we see the optimum result achieved with an LDA classifier, using the six MRMR-selected features of the synthetic dataset, which included one feature of the absorbance, and five of the second derivative spectral regime (Table 2, Fig. 5 (viii)). Four of these features were also identified in the separate vector regime calculations.

In 50 MC cross-validations per selected feature set, the ensemble RF-classifier achieved average accuracies within a range of $71 \pm 7\%$ up to $89 \pm 4\%$ (Table 3). The highest average accuracies were achieved with both RF- and MRMR-selected features of the combined dataset: $90 \pm 7\%$ sensitivity, $88 \pm 7\%$ specificity, and $90 \pm 7\%$ sensitivity and specificity of $89 \pm 6\%$, respectively. Of the 17 features identified by the RF algorithm of the synthetic conjunct spectral data set, 15 were determined in calculations on the separate absorbance and second-derivative spectra as most important discriminative patterns (compare Fig. 5 (i), (v), and (vi)). Of the six MRMR-selected features, only two were identified in separate calculations (compare Fig. 5 (ii), (vi), and (viii)).

It must be noted, that neither LDA nor RF-validation indicated a correlation of accuracy with the number of features. The information density appeared highest in the second derivative spectra. After all, the reason for the generation of the synthetic conjunct spectral vector was also to identify subtle band shifts, which were more evident by second derivatives. Finally, the interplay between feature selection algorithms and the applied classifier remains subject to further research with extended datasets.

A striking imbalance of sensitivity and specificity, as encountered [17] and explicitly researched with unbalanced data sets [18], was not observed here. 73 cancer lesion spectra were analysed versus 65 recorded ones on tissue control areas (see Tables 2 and 3), but the applied classifiers appear insensitive to a 6% deviation from parity within the number of subclass members.

3.4. Test-set validation

Spectra that were recorded of samples of the last 20 recruited patients were processed separately from the 138 spectra of the first 37 patient samples. Thus, these served as data collected with identical equipment (Matrix MF spectrometer with cone-shaped diamond ATR-probe), but being independent from previous classification and feature selection.

First, the discriminative power of the features identified by MRMR and RF algorithms were assessed in leave-one-third-out cross-validations. Although the features were identified on the first

Table 3

Average \pm standard deviation performance data of 50 ensemble RF classifiers on optimum feature sets of absorbance only (abs), first derivative only (1st deriv), and second derivative spectra (2nd deriv) of surface absorbance spectra in comparison with the performance on features selected from the concatenated absorbance-first derivative-second derivative spectral vector for the discrimination of control and cancerous tissue (see Table 2 for the legend).

	cyc	#f	acc	sens	spec
abs	$\geq 199/200$	8	87 ± 4	86 ± 8	87 ± 8
abs	MRMR	2	83 ± 5	84 ± 8	83 ± 9
1st deriv	$\geq 177/200$	6	82 ± 5	87 ± 7	77 ± 11
1st deriv	MRMR	18	71 ± 7	75 ± 10	67 ± 11
2nd deriv	$\geq 111/200$	42	89 ± 5	89 ± 6	90 ± 7
2nd deriv	MRMR	3	87 ± 3	85 ± 6	89 ± 6
total set	$\geq 196/200$	17	89 ± 4	90 ± 7	88 ± 7
total set	MRMR	6	89 ± 4	89 ± 6	89 ± 6

37 patients, data of the last 20 subjects were separated by MRMR-features with an average accuracy of $81 \pm 6\%$ (sensitivity $74 \pm 11\%$, specificity $89 \pm 8\%$) using 1000 LDA classifiers, and a respective average accuracy of $85 \pm 5\%$ (sensitivity $84 \pm 9\%$, specificity $87 \pm 7\%$) using 50 ensemble RF classifiers.

RF-selected features led to an average accuracy of $91 \pm 4\%$ (sensitivity $88 \pm 8\%$, specificity $94 \pm 6\%$) using 1000 LDA classifiers, whereas 50 ensemble RF classifiers yielded a respective average accuracy of $85 \pm 6\%$ (sensitivity $86 \pm 10\%$, specificity $84 \pm 9\%$) (see also Table 7A). Thus, both feature sets enable the calculation of meaningful, predictive classifiers within a new dataset, if the classifiers were trained using a cross-validation scheme.

However, we recorded more than one spectrum per sample. Therefore, it is likely in such cross-validation schemes, that samples used for training will be found in the validation data as well, although spectra recorded of those were well-partitioned into the respective training- and test-sets.

Therefore, to assess the predictive capabilities of the technique in a maximally realistic setting, three classifier systems were trained on the 138 calibration spectra recorded from the first 37 patients, and applied for prediction of 122 validation spectra recorded from the last 20 patients. All spectral data was reduced in dimensionality to the respective MRMR- or RF- selected features of the synthetic spectral vectors consisting of absorbance, 1st derivative, and 2nd derivative spectra.

First, an LDA classifier was calculated with the calibration data and MRMR-features. It predicted the classes using the 122 validation spectra with an accuracy of 80%, a sensitivity of 83%, and a specificity of 78% (see Table 7B). As this was a single test, no standard deviation could be calculated. A second LDA classifier was calculated with RF-features, which yielded also an accuracy of 80%, but a sensitivity of 86%, and a specificity of 75%. Next, two ensemble RF classifiers were calculated with 138 calibration spectra using both feature sets. MRMR-features yielded an accuracy of 77%, a sensitivity of 76%, and a specificity of 78%, whereas RF-selected variables resulted in an accuracy of 76%, a sensitivity of 75%, and a specificity of 77%. To assess the variability of classification, 50 ensemble RF classifiers were calculated with both respective feature sets using a random selection scheme for 100/138 calibration spectra. MRMR-features yielded an average accuracy of $78 \pm 2\%$, with an average sensitivity of $77 \pm 5\%$ and an average specificity of $78 \pm 2\%$, whereas an average accuracy of $76 \pm 1\%$, an average sensitivity of $74 \pm 3\%$, and an average specificity of $78 \pm 2\%$ were achieved with the RF feature set (Table 7B).

By test-set validation, we could quantify the over-estimation rate due to our previous cross-validation testing of training samples to 7% for accuracy, 7% for sensitivity, and 9% for specificity with the MRMR-features. Using RF-selected variables, the accuracy was similarly over-estimated by 9%, the sensitivity by 12%, and the specificity by 6%, as referred to the averaged quantification with 50 ensemble RF classifiers (see Table 7).

Although the performance of the independent test-set-validation was decreased in comparison with the cross-validation results, the low standard deviation indicates a higher stability of classification (cf. error bars in Fig. 9).

Summing up the results of the spectral discrimination of tumorous and non-malignant tissue, both feature selection algorithms identified validated, class separating spectral patterns. Based on these selected spectral variables, classifiers performed with average accuracies of up to 90% under cross-validation on the calibration dataset, and enabled discrimination with an average accuracy of up to 91% on the validation dataset. The test set validation yielded accuracies of 80% with both LDA and ensemble RF classifiers. Thereby, the existence of discriminative spectral patterns was unambiguously demonstrated (Figs. 5 and 6, Tables 2, 3 and 7). Whether the identified features could be optimized further, and whether the selected algorithms represent optimal solutions, will be subject of future studies.

3.5. Grading of colorectal tumour tissue

One concise advantage of an infrared tissue classification would be to directly grade a tumour in a non-invasive and fast way enabled by fibre-optics during surgery. Successful infrared-spectroscopic grading was recently described by micro-spectroscopic imaging [26]. Dedifferentiated cells of higher grades were found to be seldomly clustered, but spatially distributed within the tumour tissue. Therefore, we see a definite advantage of a larger sampling area for the application of the infrared probes over miniature Raman fibre sensors [39].

A surgeon's options depend on the detailed assessment of the patient's tumour. Following the German S3 guidelines for colorectal cancer, low grade cancer size (G1, G2) of the middle and lower rectum smaller than 3 cm and without vessel invasion with a minor invasion into the submucosa, is defined as low risk cancer [45]. At this stage a local resection can be the best surgical option. Thus, the patient would be spared the potential complications of a total mesorectal excision. Therefore, the spectral subset of cancer lesions diagnosed as grade G2 and G3 was selected for the according spectral discrimination (Fig. 7).

Again, separate calculations for feature selection and validation were performed with absorbance, first, second derivative, and concatenated spectra. As compared to the features selected for the malignancy identification, the selection algorithms preferably emphasized wavenumbers at approximately 1500 cm^{-1} and 1300 cm^{-1} in absorbance, and between 1500 cm^{-1} and 1200 cm^{-1} in second-derivative spectra (Fig. 7). Consistent subsets of feature patterns were identified by the MRMR algorithm in the concatenated spectral set, whereas the RF algorithm identified an altered pattern in the second derivative regime (Fig. 7A (i) and (vii), B (v) and (vii)). Partly, these results matched features as identified by the MRMR algorithm as well. Features identified in the

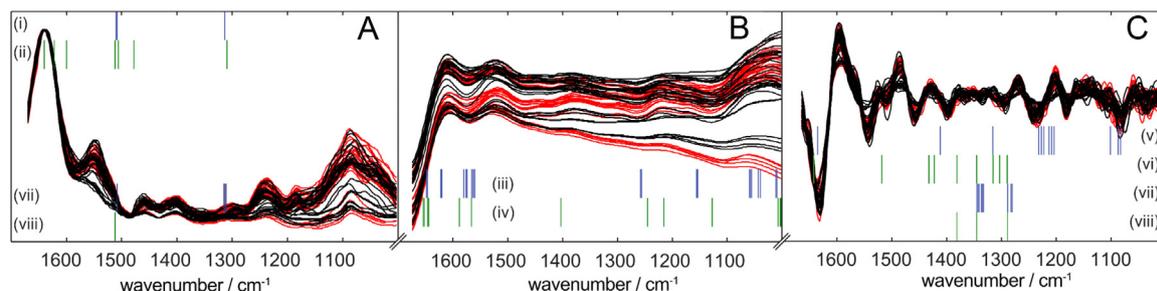


Fig. 7. Spectral data of low grade (black) and high grade cancerous (red) regions. Relevant spectral features were identified from the combined dataset consisting of absorbance (A), first (B) and second derivative spectra (C). Bar-wise displayed features are the results of calculations (blue RF, green MRMR) on spectral regions (A), (i, ii); (B), (iii, iv); (C), (v, vi), and (A+B+C), (vii, viii). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

absorbance regime were retrieved in the concatenated spectral set. In first-derivative spectra, both algorithms identified broadly distributed variable patterns (Fig. 7 (iii) and (iv)).

The average accuracy, as achieved by 1000 fold cross-validation of LDA classifiers, ranged from $69 \pm 9\%$ to $81 \pm 8\%$ on features derived from absorbance, second derivative or the concatenated set. With features of the first derivative spectra, only an accuracy of $66 \pm 10\%$ (RF) or $62 \pm 11\%$ (MRMR) was achieved, respectively. With an average accuracy of $81 \pm 8\%$, a sensitivity of $80 \pm 16\%$, and a specificity of $81 \pm 14\%$, the best grading result was obtained with nine MRMR-selected features of the second-derivative spectra (see Table 4).

Ensemble RF classifiers confirmed a high accuracy of $79 \pm 9\%$ (sensitivity of $83 \pm 12\%$, specificity of $75 \pm 14\%$) with these features, but a significantly ($p < 0.01$) increased accuracy of $84 \pm 8\%$ (sensitivity of $88 \pm 10\%$, specificity of $80 \pm 13\%$) was achieved with 13 RF-selected features of the concatenated vectors (Table 5). The mean intensities with confidence intervals of these 13 features in particular indicate a spectral separability of the low and high grade cancer classes (Fig. 8). Only for one feature the intensity-ranges of both classes overlapped. Such feature may be regarded as baseline or anchoring point for the applied multivariate classifiers, whereas the other spectral features bear classification-relevant information.

These results demonstrate a principal spectral discrimination of grades G2 and G3 with notable accuracy. Having only one patient of grade G3 in the validation set was insufficient for test-set validation of the grading. However, if one assumed a similar over-optimistic evaluation due to the cross-validation procedure as seen in the test-set validation before (Section 3.4), the achieved grading accuracy would still approximate 75%. This documents the enormous potential of fibre-optic FTIR spectroscopy for the intra-surgical application.

The technique is, however, limited to sensing the surface-biochemistry exclusively. Tumour grading is histopathologically diagnosed as the worst grade identified in the sample. Thus, inaccuracies in spectral grading may well be attributed to undifferentiated cells at a deeper tissue layer, which is as such undetectable with our non-invasive, surface layer sensitive technique. Necrotic tissue may also shield underlying tumour tissue against ATR-FTIR-detection.

For practical application, a two-step ensemble RF system with a sequential determination, whether the collected spectrum would belong to the disease class, and in a second step, to grade G2 or G3, was evaluated. Reliable grading was available for 116 spectra (57 controls, 29 G2, 30 G3). Again, a randomly selected third of the smallest class, i.e. 10 members, were left out of each class to build 50 MC paired datasets of 86:30 for training and testing. With these, ensemble RF classifiers of 1001 RFs each were trained and validated to discriminate between control or cancer area, then another ensemble RF classifier of 1001 RFs to predict the grading of those spectra, which had been assigned to cancer. For the

Table 4

Average \pm standard deviation performance data of 1000 LDA classifiers on RF- and MRMR-selected optimum feature sets for the differentiation of G2 and G3 colorectal cancer tissue (see Table 2 for the legend).

	cyc	#f	acc	sens	spec
abs	$\geq 198/200$	3	69 ± 9	78 ± 13	60 ± 15
abs	MRMR	7	72 ± 10	76 ± 14	69 ± 18
1st deriv	$\geq 117/200$	22	66 ± 10	59 ± 16	63 ± 15
1st deriv	MRMR	13	62 ± 11	67 ± 17	57 ± 16
2nd deriv	$\geq 108/200$	12	77 ± 9	78 ± 9	76 ± 15
2nd deriv	MRMR	9	81 ± 8	80 ± 16	81 ± 14
total set	$\geq 152/200$	13	73 ± 9	75 ± 14	71 ± 17
total set	MRMR	4	76 ± 9	77 ± 13	75 ± 15

Table 5

Average \pm standard deviation performance data of 50 ensemble RF classifiers on RF- and MRMR-selected optimum feature sets for the differentiation of G2 and G3 colorectal cancer tissue (see Table 2 for the legend).

	cyc	#f	acc	sens	spec
abs	$\geq 198/200$	3	67 ± 8	68 ± 15	66 ± 15
abs	MRMR	7	76 ± 8	83 ± 14	68 ± 15
1st deriv	$\geq 117/200$	22	74 ± 9	77 ± 12	71 ± 13
1st deriv	MRMR	13	75 ± 9	75 ± 16	74 ± 13
2nd deriv	$\geq 108/200$	12	81 ± 10	88 ± 11	73 ± 15
2nd deriv	MRMR	9	79 ± 9	83 ± 12	75 ± 14
total set	$\geq 152/200$	13	84 ± 8	88 ± 10	80 ± 13
total set	MRMR	4	79 ± 8	85 ± 13	72 ± 13

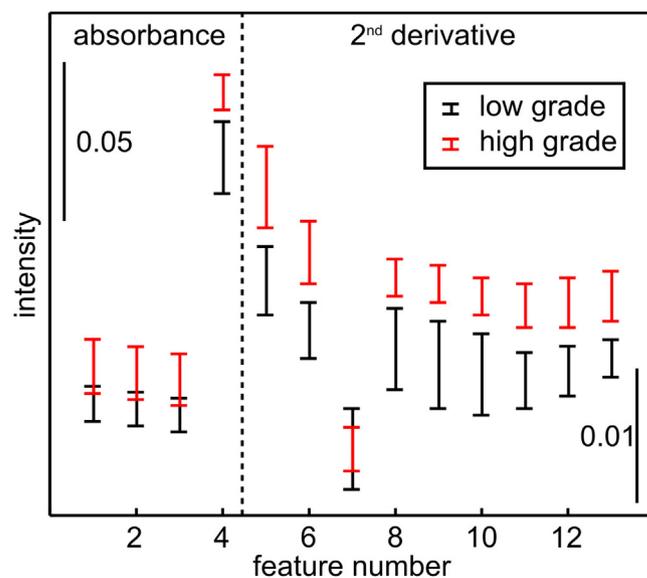


Fig. 8. High and low grade cancerous tissue regions were discriminated with highest accuracy based on 13 features (class-wise mean values with 95% confidence intervals), which were identified by the RF algorithm (absorbance: 1311, 1313, 1315, 1508 cm^{-1} ; 2nd derivative: 1281, 1283, 1289, 1333, 1335, 1337, 1341, 1343, 1345 cm^{-1} ; compare with Fig. 7 and Table 5).

predictive steps, the previously identified optimum feature sets were used: six MRMR-determined spectral features of the concatenated spectra for the discrimination of cancerous or non-cancerous area (Fig. 5C (v)), and 13 RF-determined features of the synthetic conjunct vector for grading (Fig. 7 (vii)). Thus, a total accuracy of $71 \pm 1\%$ of the correct grading prediction after correct cancer prediction was achieved (Table 6). A difference between the results for the discrimination between healthy and tumour area (Table 3) was expected due to the further limited spectral dataset as compared to the previous calculations.

3.6. Comparison with a Raman probe setup

Wood et al. presented a miniature Raman probe, suitable for colonoscopic application including discussions with concern to its size [39]. In leave-one-out cross-validations on 375 spectra of acquired 356 *ex-situ* biopsies that were obtained from 177 patients, an accuracy of 87% for distinguishing normal from adenocarcinoma tissue was obtained. Normal tissue, adenoma, adenocarcinoma, ulcerative colitis, and hyperplastic polyps were differentiated with accuracies ranging from 53 to 92% under the proposed conditions for clinical application. The selected pathologies are highly relevant for the colonoscopist for further treatment decisions.

Our aim here, however, was not yet to evaluate an endoscopic application of current fibre-optic FTIR-probes. By both size and

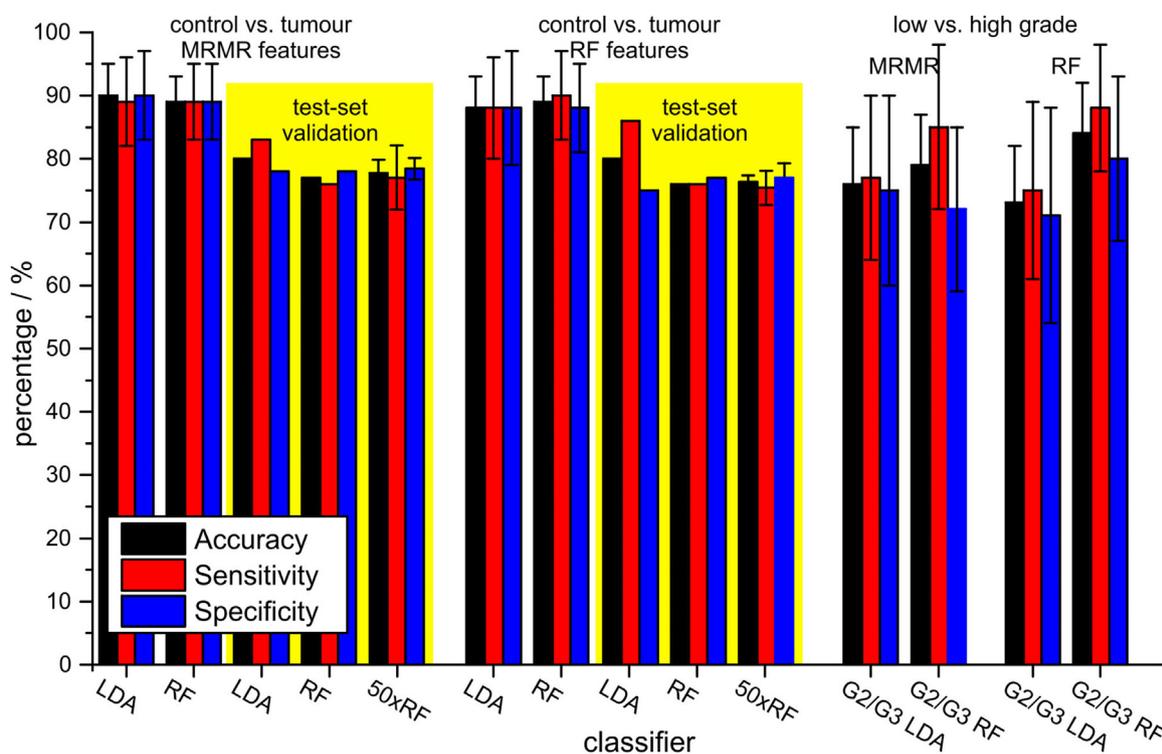


Fig. 9. Overview about the obtained validation results with the indicated feature sets obtained from synthetic conjunct spectral vectors. Test-set results are highlighted with a yellow background. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6

Average \pm standard deviation performance of a sequential ensemble random forest classification/prediction engine to discriminate healthy tissue from cancer lesion and grades G2 and G3 (see Table 2 for the legend).

Prediction	acc	sens	spec
control/cancer lesion	89 \pm 5	87 \pm 6	93 \pm 8
grade G2/G3	80 \pm 9		
G2 or G3 if 'cancer'	71 \pm 1		

Table 7

Validation results for LDA and ensemble RF classifiers on an independent dataset of 122 spectra obtained from 20 patients.

A) internal cross-validation using validation set				
features	classifier	acc	sens	spec
MRMR	LDA	81 \pm 6	74 \pm 11	89 \pm 8
MRMR	50xRF	85 \pm 5	84 \pm 9	87 \pm 7
RF	LDA	91 \pm 4	88 \pm 8	94 \pm 6
RF	50xRF	85 \pm 6	86 \pm 10	84 \pm 9
B) Test-Set validation				
features	classifier	acc	sens	spec
MRMR	LDA	80	83	78
MRMR	RF	77	76	78
MRMR	50xRF	78 \pm 2	77 \pm 5	78 \pm 2
RF	LDA	80	86	75
RF	RF	76	75	77
RF	50xRF	76 \pm 1	74 \pm 3	78 \pm 2

signal quality, current commercially available probes are inapplicable in an endoscopic scenario. Merely, the potential of the biochemical tissue information, inherent in the FTIR-signature, towards supplementing current intra-operative tissue assessment for the surgeon was evaluated.

Wood and coworkers left the key issue of probe head cleaning unconsidered. It might be that the confocal character of the Raman setup is less sensitive towards sample carryover, although it was designed for – and applied with – tissue contact. An illuminated area of approximately 1 mm² to a depth of 150 μ m was estimated for the Raman probe. In contrast, an ATR-based probe head is exclusively surface sensitive, approximately up to a micrometer penetration depth (decay of the evanescent field intensity to e^{-1}), with a sampled area of approximately 6 mm². Thus, carried-over material was shown to contribute essentially to the signal (Fig. 4). By analysing the buffer-coated specimen, we present a practical solution. So far, the results achieved with the Raman setup in leave-one-out cross-validation are very similar to those of our FTIR devices in leave-one-third-out Monte Carlo cross-validations. Regarding cancer tissue discrimination, both methods, Raman confocal sampling of a small, but deeper spot, and ATR-FTIR near-surface sampling of a larger region, produced equivalent results. For our infrared approach, the validity of both spectral feature sets and classifiers were confirmed with an independent test-set validation.

A further application of the fibre-optic probes would be the determination of resection margins during surgery, especially by confirming a safe distance of healthy to the tumour tissue of the resectate. The spatial discrimination sensitivities of both infrared and Raman probe technologies remain to be determined. In the end, which technology proves advantageous, or whether a combination of both in a combined sensor system performs excessively well, remains still to be further investigated.

4. Conclusion

A macroscopic, but flexible spectral histopathology approach for fast decision making was developed. The presented work demonstrates a successful integration of FTIR-analysis into clinical routine procedures. Further, instrumental demands for the clinical

application of a fibre-coupled ATR-probe using FTIR-spectroscopy under endoscopic conditions were met. Especially, the challenge of avoiding carryover of tissue constituents during the measurement procedure was tackled.

Findings about the principal spectral discrimination of healthy luminal colorectal areas from tumorous tissue were confirmed with dedicated bioinformatic tools, which have not previously been applied to FTIR-data obtained by fibre-optic probes. Using simple, computationally inexpensive as well as sophisticated routines of high demand in computer power, accurate class predictions were achieved with repeat cross-validations. Spectral tumour tissue markers were validated with an independent test-set. Furthermore, evidence for spectral grading was presented. A timely diagnosis is clinically relevant, because a low grade cancer of the low and middle rectum can mean a restricted surgical therapy, which usually implies a lower morbidity.

The utilisation of two different systems and the successful combination of the spectral data indicated the applicability of the technique in multi-centric studies. So far, the suitability of the presented technology for discrimination between healthy and tumorous areas was verified. A two-step prediction system for direct grading assessment was developed within the presented study. In future, FTIR-spectroscopy might help to detect if an organ invasion has been caused by a malign tumour, or if changes are just due to inflammation. During a surgical operation, this information is very important for the decision of giving up organs (e.g., a long part of gut) or taking risk of a complex restoration (e.g., of blood vessels). Using a further validated set of spectral biomarkers with an endoscopic spectrometer system, a colonoscopist or surgeon could gather immediate insights into the tumour malignancy. This could supplement the essential histopathology with an immediate spectroscopic assessment *in-situ*.

Acknowledgements

These studies were made available by support of PURE (Protein research Unit Ruhr within Europe), financed by the state of North Rhine-Westphalia. Particularly, we are grateful to A. Machleit, Protestant Hospital Herne for the organization/support of the clinical study. The authors also thank all patients participating in the reported study.

References

- [1] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, D. Forman, Global cancer statistics, *CA Cancer J. Clin.* 61 (2011) 69–90.
- [2] R. Siegel, D. Naishadham, A. Jemal, Cancer statistics, 2013, *CA Cancer J. Clin.* 63 (2013) 11–30.
- [3] H. Ueno, Y. Kajiwara, H. Shimazaki, E. Shinto, Y. Hashiguchi, K. Nakanishi, K. Maekawa, Y. Katsurada, T. Nakamura, H. Mochizuki, J. Yamamoto, K. Hase, New criteria for histologic grading of colorectal cancer, *Am. J. Surg. Pathol.* 36 (2012) 193–201.
- [4] C. Wittekind, H.-J. Meyer, TNM Klassifikation Maligner Tumoren, Wiley-VCH, Weinheim, 2013.
- [5] H.-U. Gremlich, B. Yan, Infrared and Raman Spectroscopy of Biological Materials, M. Dekker, New York, 2001.
- [6] M. Diem, J.M. Chalmers, P.R. Griffiths, Vibrational Spectroscopy for Medical Diagnosis, John Wiley & Sons Chichester, England, Hoboken, NJ, 2008.
- [7] P. Lasch, J. Kneipp, Biomedical Vibrational Spectroscopy, Wiley-Interscience, Hoboken, NJ, 2008.
- [8] P. Lasch, J. Schmitt, M. Beekes, T. Udelhoven, M. Eiden, H. Fabian, W. Petrich, D. Naumann, Antemortem identification of bovine spongiform encephalopathy from serum using infrared spectroscopy, *Anal. Chem.* 75 (2003) 6673–6678.
- [9] W. Petrich, K.B. Lewandrowski, J.B. Muhlestein, M.E.H. Hammond, J.L. Januzzi, E.L. Lewandrowski, B. Dolenko, J. Früh, W. Köhler, R. Mischler, Potential of mid-infrared spectroscopy to aid the triage of patients with acute chest pain, *Analyst* 134 (2009) 1092–1098.
- [10] T.C. Martin, J. Moecks, A. Beloousov, S. Cawthraw, B. Dolenko, M. Eiden, J. Von Frese, W. Kohler, J. Schmitt, R. Somorjai, T. Udelhoven, S. Verzakow, W. Petrich, Classification of signatures of Bovine Spongiform Encephalopathy in serum using infrared spectroscopy, *Analyst* 129 (2004) 897–901.
- [11] D.I. Ellis, R. Goodacre, Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy, *Analyst* 131 (2006) 875–885.
- [12] M. Beekes, P. Lasch, D. Naumann, Analytical applications of Fourier transform-infrared (FT-IR) spectroscopy in microbiology and prion research, *Vet. Microbiol.* 123 (2007) 305–319.
- [13] G. Bellisola, C. Sorio, Infrared spectroscopy and microscopy in cancer research and diagnosis, *Am. J. Cancer Res.* 2 (2012) 1–21.
- [14] J. Trevisan, P.P. Angelov, P.L. Carmichael, A.D. Scott, F.L. Martin, Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives, *Analyst* 137 (2012) 3202–3215.
- [15] P. Carmona, M. Molina, M. Calero, F. Bermejo-Pareja, P. Martínez-Martín, A. Toledano, Discrimination analysis of blood plasma associated with Alzheimer's disease using vibrational spectroscopy, *J. Alzheimers Dis.* 34 (2013) 911–920.
- [16] K. Gajjar, J. Trevisan, G. Owens, P.J. Keating, N.J. Wood, H.F. Stringfellow, P.L. Martin-Hirsch, F.L. Martin, Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer, *Analyst* 138 (2013) 3917–3926.
- [17] J. Ollesch, S.L. Drees, H.M. Heise, T. Behrens, T. Brüning, K. Gerwert, FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification, *Analyst* 138 (2013) 4092–4102.
- [18] 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.
- [19] B. Rigas, S. Morgello, I.S. Goldman, P.T. Wong, Human colorectal cancers display abnormal Fourier-transform infrared spectra, *Proc. Natl. Acad. Sci. U. S. A.* 87 (1990) 8140–8144.
- [20] S. Argov, J. Ramesh, A. Salman, I. Sinielnikov, J. Goldstein, H. Guterman, S. Mordechai, Diagnostic potential of Fourier-transform infrared microspectroscopy and advanced computational methods in colon cancer patients, *J. Biomed. Opt.* 7 (2002) 248–254.
- [21] S. Argov, R.K. Sahu, E. Bernshtain, A. Salman, G. Shohat, U. Zelig, S. Mordechai, Inflammatory bowel diseases as an intermediate stage between normal and cancer: a FTIR-microspectroscopy approach, *Biopolymers* 75 (2004) 384–392.
- [22] P. Lasch, M. Diem, W. Hänsch, D. Naumann, Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging, *J. Chemom.* 20 (2007) 209–220.
- [23] A. Zwielly, S. Mordechai, I. Sinielnikov, A. Salman, E. Bogomolny, S. Argov, Advanced statistical techniques applied to comprehensive FTIR spectra on human colonic tissues, *Med. Phys.* 37 (2010) 1047–1055.
- [24] A. Kallenbach-Thieltges, F. Großerüschkamp, 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.
- [25] H. Yao, X. Shi, Y. Zhang, The use of FTIR-ATR spectrometry for evaluation of surgical resection margin in colorectal cancer: a Pilot Study of 56 samples, *Journal of Spectroscopy* 2014 (2014) 1–4.
- [26] C. Kuepper, F. Großerüschkamp, A. Kallenbach-Thieltges, A. Mosig, A. Tannapfel, K. Gerwert, Label-free classification of colon cancer grading using infrared spectral histopathology, *Faraday Discuss.* 187 (2016) 105–118, doi:<http://dx.doi.org/10.1039/c5fd00157a>.
- [27] B. Bird, M.S. Miljković, S. Remiszewski, A. Akalin, M. Kon, M. Diem, Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer, *Lab. Invest.* 92 (2012) 1358–1373.
- [28] L. Mavarani, D. Petersen, S.F. El-Mashtou, 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.
- [29] M. Diem, A. Mazur, K. Lenau, J. Schubert, B. Bird, M. Miljković, C. Krafft, J. Popp, Molecular pathology via IR and Raman spectral imaging: molecular pathology via infrared and Raman spectral imaging, *J. Biophotonics* 6 (2013) 855–886.
- [30] Q. Zhong, C. Yang, F. Großerüschkamp, A. Kallenbach-Thieltges, P. Serocka, K. Gerwert, A. Mosig, Similarity maps and hierarchical clustering for annotating FT-IR spectral images, *BMC Bioinf.* 14 (2013) 333.
- [31] M.A. Mackanos, C.H. Contag, Fiber-optic probes enable cancer detection with FTIR spectroscopy, *Trends Biotechnol.* 28 (2010) 317–323.
- [32] H.M. Heise, L. Küpper, L.N. Butvina, Bio-analytical applications of mid-infrared spectroscopy using silver halide fiber-optic probes, *Spectrochim. Acta B* 57 (2002) 1649–1663.
- [33] H.M. Heise, L. Küpper, W. Pittermann, M. Stücker, Epidermal *in vivo* and *in vitro* studies by attenuated total reflection mid-infrared spectroscopy using flexible silver halide fibre-probes, *J. Mol. Struct.* 651–653 (651) (2003) 127–132.
- [34] U. Bindig, H. Winter, W. Wa, K. Zelianeos, G. Mu, Fiber-optical and microscopic detection of malignant tissue by use of infrared spectrometry, *J. Biomed. Opt.* 7 (2002) 100–108.
- [35] U. Bindig, G. Müller, Fibre-optic laser-assisted infrared tumour diagnostics (FLAIR), *J. Phys. D Appl. Phys.* 38 (2005) 2716–2731.
- [36] M.A. Mackanos, J. Hargrove, R. Wolters, C.B. Du, S. Friedland, R.M. Soetikno, C. H. Contag, M.R. Arroyo, J.M. Crawford, T.D. Wang, Use of an endoscope-compatible probe to detect colonic dysplasia with Fourier transform infrared spectroscopy, *J. Biomed. Opt.* 14 (2009) 044006.
- [37] V.K. Katukuri, J. Hargrove, S.J. Miller, K. Rahal, J.Y. Kao, R. Wolters, E.M. Zimmermann, T.D. Wang, Detection of colonic inflammation with Fourier

- transform infrared spectroscopy using a flexible silver halide fiber, *Biomed. Opt. Express* 1 (2010) 1014–1025.
- [38] L. Dong, X. Sun, Z. Chao, S. Zhang, J. Zheng, R. Gurung, J. Du, J. Shi, Y. Xu, Y. Zhang, J. Wu, Evaluation of FTIR spectroscopy as diagnostic tool for colorectal cancer using spectral analysis, *Spectrochim. Acta A* 122 (2014) 288–294.
- [39] 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, *Colorectal Dis.* 16 (2014) 732–738.
- [40] J. Ollesch, E. Künnemann, R. Glockshuber, K. Gerwert, Prion protein alpha-to-beta transition monitored by time-resolved Fourier transform infrared spectroscopy, *Appl. Spectrosc.* 61 (2007) 1025–1031.
- [41] K. Elfrink, J. Ollesch, J. Stöhr, D. Willbold, D. Riesner, K. Gerwert, Structural changes of membrane-anchored native PrP(C), *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 10815–10819.
- [42] A. Liaw, M. Wiener, Classification and regression by random forest, *R News* 2 (2002) 18–22.
- [43] H. Peng, F. 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.
- [44] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, *J. Bioinf. Comput. Biol.* 3 (2005) 185–205.
- [45] Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF): S3-Leitlinie Kolorektales Karzinom, Leitlinienreport, Version 11, 2014 AWMF Registrierungsnummer: 021-007OL, Version 11, Stand August 2014, (n.d.).