

# Journal of Biomedical Optics

[SPIEDigitalLibrary.org/jbo](http://SPIEDigitalLibrary.org/jbo)

## **Evaluation of Raman spectroscopic macro raster scans of native cervical cone biopsies using histopathological mapping**

Carina Reble  
Ingo Gersonde  
Cathrin Dressler  
Jürgen Helfmann  
Wolfgang Kühn  
Hans Joachim Eichler

# Evaluation of Raman spectroscopic macro raster scans of native cervical cone biopsies using histopathological mapping

Carina Reble,<sup>a,b,\*</sup> Ingo Gersonde,<sup>a</sup> Cathrin Dressler,<sup>a</sup> Jürgen Helfmann,<sup>a</sup> Wolfgang Kühn,<sup>c,d</sup> and Hans Joachim Eichler<sup>a,b</sup>

<sup>a</sup>Laser- und Medizin-Technologie GmbH, Berlin (LMTB), Fabbeckstraße 60-62, 14195 Berlin

<sup>b</sup>Technical University Berlin, Institute for Optics and Atomic Physics, 10587 Berlin

<sup>c</sup>Gyn-Zentrum-Berlin, Albrechtstraße 48, 12167 Berlin

<sup>d</sup>Klinik für Gynäkologie, Charité Universitätsmedizin Berlin, Hindenburgdamm 30, 12200 Berlin

**Abstract.** Raman spectroscopy based discrimination of cervical precancer and normal tissue has been shown previously *in vivo* with fiber probe based measurements of colposcopically selected sites. With a view to developing *in vivo* large area imaging, macro raster scans of native cervical cone biopsies with an average of 200 spectra per sample are implemented ( $n = 16$ ). The diagnostic performance is evaluated using histopathological mapping of the cervix surface. Different data reduction and classification methods (principal component analysis, wavelets, k-nearest neighbors, logistic regression, partial least squares discriminant analysis) are compared. Using bootstrapping to estimate confidence intervals for sensitivity and specificity, it is concluded that differences among different spectra classification procedures are not significant. The classification performance is evaluated depending on the tissue pathologies included in the analysis using the average performance of different classification procedures. The highest sensitivity (89%) and specificity (86%) is obtained for the discrimination of normal squamous epithelium and high-grade precancer. When other non-high-grade tissue sites, such as columnar epithelium, metaplasia, and inflammation, are included, the diagnostic performance decreases. © 2014 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JBO.19.2.027007](https://doi.org/10.1117/1.JBO.19.2.027007)]

Keywords: Raman spectroscopy; cervical dysplasia; optical diagnosis.

Paper 130567RR received Aug. 6, 2013; revised manuscript received Jan. 12, 2014; accepted for publication Jan. 15, 2014; published online Feb. 18, 2014; corrected Mar. 14, 2014.

## 1 Introduction

Cervical cancer is one of the leading causes of cancer deaths in women and is known to be caused by human papillomavirus (HPV) infection.<sup>1</sup> A small percentage of HPV-infected women will develop cervical intraepithelial neoplasia (CIN). CIN lesions are divided into three classes of increasing grades. While the majority of low-grade CIN lesions (CIN 1) regress spontaneously, high-grade CIN (HGCIN, CIN 2 and CIN 3) tends to progress to invasive cancer.<sup>2-5</sup>

Cervical cancer progress is relatively slow and screening programs based on Papanicolaou cytological smear analysis (Pap test) have led to a remarkable decrease in the incidence and mortality of cervical cancer. In a systematic review, Nanda et al. found that the sensitivity of the Pap test ranges from 30 to 87% and the specificity ranges from 86 to 100%.<sup>6</sup>

Abnormal Pap smears are usually followed by colposcopic examinations involving visual examination under magnification, including the application of acetic acid and Lugol's iodine solution. Iodine solution stains glycogen-containing normal squamous epithelium, while application of acetic acid leads to acetowhitening of abnormal areas. Colposcopy is the basis for correct identification of an atypical transformation zone, for definition of the grade of an underlying lesion, for targeted biopsy in the case of HGCIN, and for excisional therapy.<sup>7</sup>

Different grading indexes have been created to discriminate between normal, minor change, major change, and cancer.<sup>8-12</sup> Due to subjectivity of this procedure, the actual diagnostic

accuracy of colposcopy depends on the experience of the practicing physician, resulting in inter- and intraobserver variabilities.<sup>13-16</sup> According to a meta-analysis, colposcopy has a high mean sensitivity of 85%, but a low mean specificity of 69% to distinguish normal and CIN 1 from HGCIN lesions and cancer.<sup>17</sup> Benign changes (metaplasia, inflammation) and dysplasia may have similar colposcopic appearances, and biopsies are, therefore, required to confirm the diagnosis.

Avoiding unnecessary biopsies is a major reason for the development of optical sensing technologies, which have the advantage of being noninvasive and objective due to automated data analysis. In addition, the real-time capability could enable a one-stop "see and treat" scenario, which is considered as a great potential benefit for the cervical cancer management. Particularly in some areas in developing countries, diagnostic strategies requiring fewer visits would be more cost-effective.<sup>18</sup>

A number of *ex vivo* and *in vivo* studies of native tissues have previously shown the potential of Raman spectroscopy for detection of precancers in a diagnostic setting.<sup>19-27</sup> The reported sensitivities and specificities for the discrimination of CIN from non-CIN tissue range from 73 to 100% and 79 to 98%, respectively. This wide range in results is believed to be due to various differences in the study methods, such as differences in sampling volumes, study designs, spectra preprocessing, classification and model validation methods, or the types and number of tissue classes to be separated, including disease threshold (CIN 1 or CIN 2).

\*Address to all correspondence: Carina Reble, E-mail: [c.reble@lmtb.de](mailto:c.reble@lmtb.de)

Previous *in vivo* studies on precancers were performed with fiber probes and abnormal and normal measurement sites were selected using colposcopy.<sup>20,21,24–27</sup> In this study, spectra were acquired by point-wise scanning over the surface of cone biopsies with an average of 200 spectra per sample. In order to allow correlation of Raman spectroscopy and histopathology in a spatially resolved manner, the histopathological results of cross-sectional tissue slices were assigned to measurement locations on the native sample surface. This procedure exhibits a diagnostic challenge for the pathologist due to heterogeneity and deformations of the tissue under investigation. Using this experimental approach, our aim was to investigate if the spatial extension of HGCIN can be predicted with Raman spectroscopy within the variety of different non-HGCIN lesions that might be present on the cervix. This might include tissue regions that have been identified as normal, major changes, minor changes, or unspecific as well as miscellaneous findings by colposcopy. While inflammation and metaplasia have been studied previously,<sup>19–23,28</sup> Raman classifications of other tissue changes [e.g., erosion, atypical immature metaplasia (AIM), etc.] have not yet been reported.

The results of *ex vivo* Raman spectroscopic raster scanning might be useful for future Raman diagnostic imaging *in vivo*. A fast, noninvasive determination of the spatial extension of larger HGCIN (CIN 2 and CIN 3) lesions prior to excision would be very useful to ensure that the lesion margins are completely removed while conserving as much normal tissue as possible in order to avoid overtreatment of the patient and its possible negative impact on future childbearing.<sup>29</sup> Furthermore, an *in vivo* imaging method with sufficient spatial resolution could ensure that small lesions would not be missed due to undersampling. This may occur with point-wise fiber probe measurements as the most prominent areas of colposcopic changes do not always coincide with the areas of greatest histologic abnormalities, and this is the reason why some colposcopists even recommend random biopsies.<sup>30,31</sup> Therefore, *in vivo* Raman analysis covering a large area of the cervix seems to be desirable. Such *in vivo* setups have already been using fluorescence and reflectance spectroscopy,<sup>32–34</sup> but as yet not for Raman spectroscopy.

In this study, the accuracy of the spectra classification was calculated depending on different types of non-HGCIN tissue pathologies included in the training of the algorithm. Spectra classification merged with the histological map was used to gain insights into the performance of the method for different pathologies. Several data reduction and classification methods were compared and confidence intervals for sensitivity and specificity were estimated with bootstrapping.

## 2 Material and Methods

### 2.1 Patient Recruitment

Patients with suspected precancer or cancer due to previous cytologies, colposcopies, and biopsies underwent a loop electrical excision procedure of the cervix according to clinical routine. Nonpregnant patients aged 18 to 45 (presumed premenopausal) were included in the study after signing informed consent forms permitting *ex vivo* spectroscopic measurements from the cervix samples. From the total of 34 recruited patients, 9 patients were excluded from the analysis due to different experimental or technical errors. Other samples were excluded from the data set used for training and evaluating the classification method because of varying conditions in the optimization phase of the study ( $n = 10$ ) and because of concurrent adenocarcinoma *in situ* ( $n = 1$ ). Measurements of 13 to 16 patients were included into the quantitative data analysis. The number of patients who contributed to the data set depended on the class definition as described in Table 1. The study protocol was approved by the ethics committee of the Charité University Clinic Berlin.

### 2.2 Raman Instrumentation

A Raman tissue scanner was constructed to scan large tissue samples (30 × 45 mm). The setup was based on an iHR320 Spectrograph (HORIBA Jobin Yvon GmbH, Bensheim, Germany) with ruled grating (600 l/mm) and a back illuminated deep depletion CCD detector (Synapse) cooled to  $-70^{\circ}\text{C}$ . A distributed feedback diode laser (Eagleyard Photonics, Berlin, Germany) with the center wavelength at 784 nm at

**Table 1** Different class definitions A to E used to obtain the results presented in Figs. 5 and 6 are shown. The tissue types that were included in either class 1 or 2 of the binary spectra classification are indicated in the second and third columns. For improved clarity, the study-specific tissue categories #1 to 8 are included in parentheses, which refer to the description of tissue categories in Sec. 2.4. Please note that depending on the class definitions, different numbers of patients (see fourth column) and different numbers of spectra (see fifth and sixth columns) contributed to the analyzed data set.

Definition	Class 1	Class 2	# patients	# class 1 spectra	# class 2 spectra
<b>A</b>	(1) <b>normal</b> squamous epithelium (iodine positive)	(4) <b>HGCIN</b>	<b>13</b>	<b>214</b>	<b>303</b>
<b>B</b>	(1)+(2) <b>normal tissue</b> as in A <b>plus nonabnormal</b> tissue (squamous and columnar epithelium, metaplasia, cervicitis)	(4) <b>HGCIN</b>	<b>15</b>	<b>841</b>	<b>303</b>
<b>C</b>	(2) <b>nonabnormal</b> tissue	(4) <b>HGCIN</b>	<b>15</b>	<b>627</b>	<b>303</b>
<b>D</b>	(1)+(2) as in B	(4)+(5) <b>HGCIN, HGCIN-borders</b>	<b>16</b>	<b>841</b>	<b>1514</b>
<b>E</b>	All-(4)-(5)-(8)=(1)+(2)+(3)+(6)+(7) <b>all</b> spectra assigned to histopathology (normal, nonabnormal, abnormal, CIN1, erosion) <b>except HGCIN and HGCIN-borders</b>	(4) <b>HGCIN</b>	<b>16</b>	<b>1216</b>	<b>303</b>

Note: HGCIN, high-grade cervical intraepithelial neoplasia.

a laser power of 70 mW was used. Due to the irregular shape of the samples, an inverted setup, was chosen. Here the sample surface could be pressed gently onto a measurement window to maintain a constant position of the sample surface in the focus of the laser and detection optics. For the measurement window, either 1.5-mm-thick calcium fluoride (Korth Kristalle GmbH, Altenholz, Germany) or 1-mm-thick silica (Siegert Consulting e. K., Aachen, Germany) was used. The focus of laser and detection optics was set to the interface of the measurement window and the sample surface. The working distance was 9 mm.

The laser light was coupled from a single-mode fiber into a 600- $\mu\text{m}$  multimode fiber (LEONI Fiber Optics, Berlin, Germany), which was collimated with an achromatic lens ( $f = 30$  mm), filtered by a clean-up filter (Semrock, Rochester, New York), reflected by a dichroic mirror (Semrock), and focused by an aspheric lens [ $f = 12$  mm, numerical aperture (NA) = 0.5, Asphericon] on the sample surface. The resulting illumination spot had a diameter of 160  $\mu\text{m}$ , as measured by knife-edge measurements.

For detection of the backscattered light, the aspheric lens was used together with two achromatic lenses ( $f = 80$  mm and  $f = 40$  mm) to image Raman scattered light onto a circular detection fiber bundle consisting of 31 tightly packed fibers (LEONI Fiber Optics) with a core diameter of 100  $\mu\text{m}$  and an NA = 0.22. A detection spot size of 200  $\mu\text{m}$  was achieved, as measured by knife-edge measurements. A long-pass filter (RazorEdge, Semrock) in the collimated beam was used to block the laser line. The round detection fiber bundle was converted into a stack of 100- $\mu\text{m}$  fibers, which was imaged onto the spectrograph slit. In order to match the fiber's NA (0.22) to the spectrograph's NA (0.12), the detection fiber stack was magnified and a spectrograph slit width of 200  $\mu\text{m}$  was employed to minimize the light loss due to the spectrometer coupling. The resulting spectral resolution was measured as 10.5  $\text{cm}^{-1}$  (at full width half maximum) at this optical configuration using argon emission lines of a mercury arc lamp.

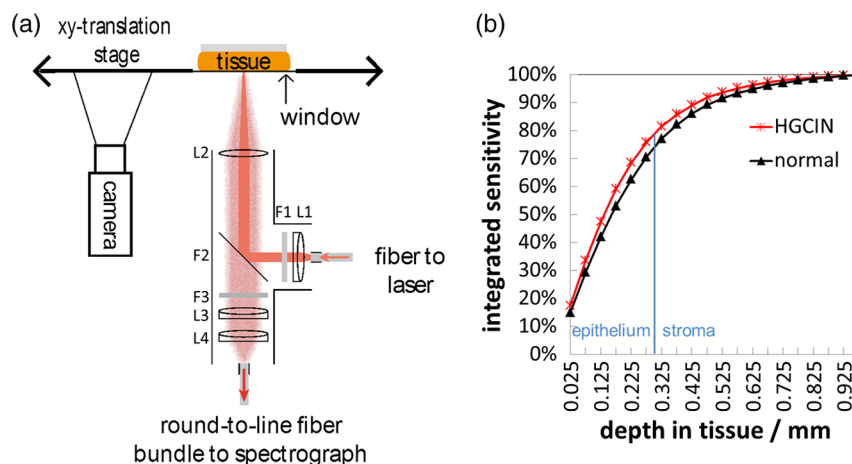
A motorized xy-translation stage with integrated measuring system (Märzhäuser Wetzlar GmbH & Co. KG, Wetzlar,

Germany) was used for moving the tissue and measurement window relative to the fixed optical components. Prior to Raman mapping, the stage was used to drive the sample into the focus of a color camera (The Imaging Source Europe GmbH, Bremen, Germany) in order to obtain an image of the sample surface for documentation.

Labview 8.6 (National Instruments, Austin, Texas) software was written for automated control of the laser, spectrometer system, and xy-translation stage. The total time for one tissue scan depends on the combination of selected parameters, such as integration time, density of measurement points, region of interest, etc., which may be adapted according to the sample type and size. This setup is mobile and light-tight and allows measurements to be performed within illuminated environments in the clinic. Monte Carlo simulations<sup>35</sup> with optical parameters of cervical tissue<sup>36,37</sup> were used to calculate the depth-dependent sensitivity of our measurement using the diameters of the excitation and detection spot on the tissue surface as well as the NA of the optics as input parameters. Figure 1(b) shows that our optical layout detected 65% (80%) of the total signal (integrated sensitivity) of HGCIN from a depth <200  $\mu\text{m}$  (300  $\mu\text{m}$ ), the thickness of cervical epithelium. Therefore, it was assumed that a sufficient fraction of the detected signal originates from regions close to the basal membrane, where cervical precancers originate. The Raman spectroscopy setup is illustrated in Fig. 1(a).

### 2.3 Raman Measurements

The orientation and the position of the epithelial surface of the samples were indicated by the surgeons using thread marks and pinning the samples onto cork boards. Each cervical sample was placed onto the measurement window within a self-adhesive silicone rubber frame, which allowed a sample to be immersed in physiological saline (0.9% NaCl) to preserve tissue hydration during measurements. Background spectra were measured with saline solution (without sample). Each sample was loaded with balance weights in order to ensure even attachment of the cervix epithelium to the measurement window.



**Fig. 1** (a) Measurement setup used in this study with the tissue sample positioned above the spectroscopic optics on a measurement window (silica or CaF<sub>2</sub>). Prior to the Raman measurement, the stage translates the sample into the focus of a color camera in order to acquire an overview color image of the sample surface, which is used to select and to document the region of interest of the Raman measurement. L1 = achromatic lens, F1 = laser clean-up filter, F2 = dichroic mirror, L2 = aspheric lens, F3 = long-pass filter, L3 = achromatic lens, L4 = achromatic lens. (b) The total detected signal from tissue up to a certain tissue depth is calculated with Monte Carlo simulations for normal and high-grade cervical intraepithelial neoplasia (HGCIN) tissue.

The sample surface was xy-scanned and spectra were recorded at discrete measurement points with  $\sim 1$  spectrum per  $\text{mm}^2$  within the chosen region of interest (area up to  $4 \text{ cm}^2$ ). The number of spectra recorded for each sample depended on tissue dimensions.

The choice of exposure times and accumulations depended on the autofluorescence level and the total measurement time available. The total measurement time was limited to 1 h, which was the time we were allowed to keep the samples in saline before fixation. The origin of the fluorescence could not be determined within the study; however, intraoperative application of disinfectant and iodine could be excluded. Since the fluorescence of the samples varied considerably, the exposure time was varied from 0.1 to 1 s and the number of spectra accumulations varied from 10 to 30. The acquisition parameters were adapted individually to a subjective optimum in order to scan the whole sample. A roughly spaced test scan was performed in order to determine the average fluorescence level of every sample. In principle, the number of accumulations should have been selected to maintain a constant integration time. However, problems with the speed of the CCD shutter forced us to limit the number of accumulations to 30; otherwise the total measurement time would have been unacceptably high and have led to total exposure times of 2 to 10 s.

In addition, the instrument response was measured daily using a fiber-coupled calibrated halogen lamp (Ocean Optics, Dunedin, Florida) by placing the cosine corrector at the end of the fiber into a customized holder on the measurement window.

Immediately after Raman measurements, the samples were fixed in 10% phosphate buffered formaldehyde solution for histopathological analysis.

#### 2.4 Histopathological Mapping and Its Correlation to Raman Measurement Sites

The measurement locations were documented on the sample image, which had been taken prior to the measurement. The accuracy of the measurement position within this image was estimated to be  $< 1$  mm using self-made test objects. The pathologist was provided with this image and the fixed sample, which often showed a slight shape deformation as compared to the native sample. According to routine procedures, each specimen was dissected into appropriate segments and then embedded in paraffin wax. Histological sections were then prepared in series, stained with hematoxylin-eosin, and analyzed by microscopy. The histopathological findings from the tissue cross-sections of the cone biopsy were documented by the pathologist on the image of the native cervix surface, which included the Raman measurement points to yield histopathological maps. The spatial uncertainty of this procedure was estimated to be 1 to 3 mm, depending on the circumstances. Examples of such histopathological documentations can be seen in Fig. 7, where the class membership prediction is overlaid with the histopathological documentations. Each spectrum was then labeled by the spectroscopist with one of the (study-specific) tissue categories 1 to 8 (described below) using a semiautomated program.

Please note that this histopathological mapping process is more challenging than annotating the histopathological results of punch biopsies to measurements at biopsy sites. As it was not possible to engage more than one pathologist in our study, the pathologist involved analyzed each of the samples twice to improve precision of the result documentation. The histopathology results of all patients included in the study

were discussed at intraclinical conferences. The pathologist supporting this study (coauthor W.K.) specializes in cytology and gynecological morphology and has worked in this field for over 30 years.

For the spectra classification, the following study-specific tissue categories were assigned to individual spectra. Some categories include multiple tissue types (2,4,5,8) for the sake of a simpler procedure for the assignment of histopathology to measurement sites.

1. Normal squamous epithelium (with visible dark iodine stain from the intraoperative application prior to the measurement, without histopathological signs of abnormality within central areas of homogenous/continuous pathology)
2. Nonabnormal tissue (including squamous epithelium without continuous dark iodine stain, columnar epithelium, metaplasia, and cervicitis, without histopathological signs of abnormality in a continuous area)
3. CIN 1 (spectra from within central areas of homogenous tissue pathology)
4. HGCIN (CIN 2, CIN 3, or AIM p16 positive, including micro invasions, spectra from within central areas of homogenous tissue pathology)
5. CIN-borders (specified by the CIN grade if CIN was not spatially homogeneous/continuous on a scale larger than the spatial uncertainty of the method. Spectra within 1 to 2 mm of the borders of larger CIN lesions or spectra within 3 mm distance of a small CIN lesion were also annotated as CIN-border for the same reason. Thus, such spectra originate either from neighboring pathologies or from a mixed or intermediate pathology)
6. Erosion (without signs of CIN)
7. Abnormal epithelium (areas with abnormalities except CIN)
8. Unassigned spectra (spectra that could not be assigned to histopathology)

The accuracy of the lateral position and extension of pathologies in the histopathological map was estimated to be between 1 and 3 mm (depending on the circumstances). Therefore, only spectra from within central areas of homogenous tissue pathology were assigned to the pathologies. Please note that this procedure significantly reduced the number of CIN spectra available for the quantitative evaluation of the method because most of the HGCIN lesions were not homogeneously high grade, but were focal, only present inside crypts or mixed with CIN 1, metaplasia, or other non-high-grade pathologies. Therefore, the label "CIN-border" was assigned to spectra from areas in which CIN was not spatially homogeneous/continuous as our procedure of correlating histopathology to measurement sites only had an estimated precision of 1 to 3 mm, and therefore, the location of small focal lesions (of sizes smaller than the spatial precision of the method) could not be assigned to measurement positions.

Spectra within 1 to 2 mm of the borders of larger CIN lesions or spectra within 3 mm distance of a small CIN lesion were also

annotated as CIN-border for the same reason. Thus, these CIN-border spectra originated from either neighboring pathologies or a mixed or intermediate pathology.

AIM was tested for enhanced expression of the tumor suppressor gene p16. Lesions that were p16 positive were classified as HGCIN equivalent lesions according to Ref. 38, as it is common practice in the cooperating hospital. For the classification, AIM p16 positive was, therefore, included in the HGCIN category (if in the vicinity of HGCIN) or the HGCIN-border category (if mixed with HGCIN and other categories within the uncertainty limits of the spatial correlation of histopathology and measurement position). Since AIM p16 positive was always interspersed with or in close vicinity to CIN or erosion in the samples of our study, separate evaluation of the classification of AIM p16 lesion was not possible.

The CIN 1 to 3 classification was used instead of the 2001 Bethesda system terminology [low grade squamous intraepithelial lesion (LGSIL) and high grade squamous intraepithelial lesion (HGSIL)]<sup>39</sup> to differentiate cervical lesions because LGSIL also contains HPV-associated changes other than CIN 1. Our spatially resolved documentation of histopathological analysis was, however, limited to CIN in most samples, whereas HPV-associated changes, such as abnormal epithelium, were indicated without spatial extension. If abnormalities were present, the tissue was not assigned to the nonabnormal category.

Discrimination of HGCIN from LGCIN (CIN 1) and other low-grade lesions (HPV-induced abnormalities) is important due to different treatment strategies. Unfortunately, only a few spectra could be measured from a continuous CIN 1 area and only one sample showed erosion without colocalized focal CIN (though present in one third of all samples), so that a separate training and classification of spectra from CIN 1, erosion could not be performed.

Spectra remained unassigned when the documented measurement site was within 1 mm of a sharp border of an iodine stain next to an abnormal area, on visible blood residues, or in small areas without histopathological specifications.

## 2.5 Preprocessing of Spectra

All spectra were checked for CCD-signal saturation, and if positive, they were removed from data analysis. Cosmic spikes were detected by their discontinuity and replaced by the average of neighboring values.

Before classification, all components of the spectrum that did not result from Raman scattering inside the sample, but from the measurement window or other optical components, were removed as far as possible.

As the measured spectrum contained a large fluorescence background, the sensitivity spectrum of the instrument lead to a distinctive oscillatory pattern. The pattern was mainly given by the transmission spectra of long-pass filters in the instrument. Therefore, an instrument response correction was performed by dividing all raw spectra by the sensitivity spectrum measured with a halogen calibration-lamp.

The background or baseline consists of tissue fluorescence as well as Raman and fluorescence signals of the instrument. A background function  $S_{\text{backgr}}$ , which contains a third-order polynomial, was defined for the fluorescence correction. As instrument background correction,  $S_{\text{backgr}}$  also contained a spectrum measured without sample  $S_{\text{empty}}$ , which had to be scaled because elastic backscattering of excitation light into the instrument varied from sample to sample and, therefore, lead to varying

amplitudes of the instrument's Raman and fluorescence signals. For each spectrum the coefficients of  $S_{\text{backgr}}$  were determined by a linear least-square fit. The residual of the fit was considered the tissue Raman spectrum.

$$S(\lambda) \approx c_1 + c_2\lambda + c_3\lambda^2 + c_4\lambda^3 + (c_5 + c_6\lambda)S_{\text{empty}}(\lambda) =: S_{\text{backgr}}(\lambda)$$

$$S_{\text{Raman}}(\lambda) = S(\lambda) - S_{\text{backgr}}(\lambda)$$

It is well-known that the Raman signal intensity depends on the optical properties of a sample, which may vary.<sup>40</sup> For this reason, the Raman spectrum was normalized using the standard normal variate algorithm.<sup>41</sup> Finally, the spectrum is smoothed by a Savitzky-Golay filter (second order,  $N_{\text{Side}} = 7$ ).

## 2.6 Spectra Classification

The aim of the Raman-spectroscopic measurement was to discriminate between two classes of tissue according to the histopathological evaluation (Table 1). Therefore, five procedures (see Table 2) for binary classification described in the following were investigated.

After preprocessing, described in Sec. 2.5, the spectra were transformed using either principal component analysis (PCA) or wavelet transformation (WT) using a translation invariant form of the set of Haar wavelets.<sup>42</sup>

The large number of spectral features was reduced in two steps: Following the PCA, the set of 20 principal components with the largest variance were retained, and for WT, were selected the features with the largest average difference between the two classes.

The remaining 20 to 40 features were further reduced by feature selection in combination with the classification algorithm. A wrapper method with a forward selection or backward elimination process was used.<sup>43</sup> The wrapper compares subsets of features by evaluating the accuracy of the classification algorithm for each subset. Application of logistic regression<sup>44</sup> or k-nearest neighbor analysis (KNN)<sup>45</sup> leads to four combinations of transformation and classification algorithms.

The fifth procedure evaluated is partial least square discriminant analysis (PLS-DA),<sup>46</sup> which was applied to the preprocessed but not transformed spectra.

To evaluate the performance of the five classification procedures, sensitivity (percentage of correctly classified spectra of class 2) and specificity (percentage of correctly classified spectra of class 1) were calculated using a leave-one-

**Table 2** List of procedures used for spectra classification.

Nr. of classification procedure	Description of classification procedure
1	PCA plus LR
2	PCA plus KNN
3	WT plus LR
4	WT plus KNN
5	PLS-DA

Note: PCA, principal component analysis; LR, logistic regression; KNN, k-nearest neighbor analysis; WT, wavelet transformation; PLS-DA, partial least square discriminant analysis.

patient-out cross-validation where all data of each patient were used in turn for validation, while the remaining data were used for the calibration process. To prevent overfitting, that it was found essential that all steps of the feature selection be performed for each validation step (i.e., for each patient left out) separately. For PLS-DA, feature selection corresponds to the optimal number of principal components used.

The significance of sensitivity and specificity was tested with a randomization test. Here the class label of each spectrum is defined by chance using random numbers. Training and validation is done for a large number of randomizations. The corresponding set of results yields distributions for sensitivity and specificity under the null hypothesis that there is no systematic dependence between the spectral data and the histopathological classification. If the values for sensitivity and specificity resulting from the histopathological classification are unlikely to be drawn from these distributions, one concludes that a significant correlation exists between the histopathological classification and the spectral data. For balanced accuracy values of at least 65 to 70%, it could be proven with the randomization test that these values are based on a significant correlation ( $p = 0.05$ ) between the spectra and the histopathological result. This means that these values do not occur by chance.

Confidence intervals for sensitivity and specificity were estimated with the bootstrap method<sup>47</sup> by resampling the data patient-wise.

## 2.7 Evaluation of the Classification Performance Depending on Included Tissue Types

As this study was aimed at detection of HGGIN from neighboring non-HGGIN tissue areas within one suspicious larger tissue area, different spectra of the same cone biopsy sample (one sample per patient) were classified into class 1 (non-HGGIN tissue) or class 2 (HGGIN tissue) depending on histopathology at the respective measurement site. Please note that a cone biopsy sample is typically larger (few cm<sup>2</sup>) than a punch biopsy sample (few mm<sup>2</sup>) and, thus, may contain multiple types and pathologies of tissue.

Since a variety of non-HGGIN tissue types exist (e.g., metaplasia, inflammation, columnar epithelium, low-grade abnormalities), classification performance might depend on the tissue types included in the non-HGGIN class, which might be one reason for the variation of results of previous studies.

In order to evaluate the classification performance depending on the tissue types included in the analysis, the tissue categories 1 to 7 were included in different definitions of class 1 and class

2, called class definitions A to E, as described in Table 1. By this approach we expected to include a larger variety of non-HGGIN tissues per patient (inpatient variability) than fiber probe measurements, which acquire few spectra at the most suspicious and normal-looking sites under colposcopic examination. The use of two classes results from the application of binary classifiers. Each class in Table 1 contains spectra from all patients that had measurement sites that fulfilled the criteria of the respective class definition.

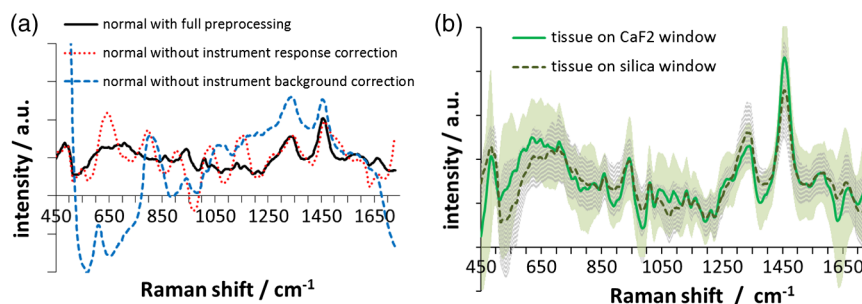
In class definition A, only normal squamous epithelium was included in class 1, which was identified by a dark iodine stain (plus the absence of abnormal areas in the histopathological maps). This class definition was expected to have the best classification result because there are anatomical differences in addition to the neoplastic changes.<sup>48</sup>

Class definition B includes different types of nonabnormal tissue, including areas of columnar epithelium as well as inflammation and metaplasia in addition to squamous epithelium. In class definition C, the squamous epithelium, which was used in A, was left out in order to investigate a possible increase of classification performance due to the inclusion of clinically normal sites, as shown previously for fluorescence and reflectance spectroscopy.<sup>48</sup> In class definition D, the large number of CIN-border spectra was added to class 2, which means that non-HGGIN tissue might have been included in the HGGIN class because the CIN-border spectra could not be assigned to a distinct tissue pathology due to the limited spatial precision of the histopathological mapping. In class definition E, different non-HGGIN spectra, including those of CIN 1, erosion, and low-grade abnormalities, were included in class 1, which corresponded to all measured (unsaturated) spectra except those of HGGIN, HGGIN-borders, and spectra that remained unassigned to histopathology.

## 3 Results and Discussion

### 3.1 Influence of Preprocessing on Extracted Raman Spectra

The influence of corrections for instrument background, tissue fluorescence, or instrument response is shown in Fig. 2. Without the instrument response correction, the oscillation of the spectra due to the filter transmission is strong. Without instrument background correction (a polynomial for fluorescence correction was subtracted to make results comparable), large peaks of silica obscure the tissue Raman signal. With both corrections, the resulting preprocessed spectra look similar to cervical spectra in the literature (see discussion in Sec. 3.2) with few exceptions.



**Fig. 2** (a) Spectra without instrument background and without instrument response correction compared to spectra after complete data preprocessing. (b) Comparison of preprocessed spectra of class 1 (in class definition B) on CaF<sub>2</sub> or silica measurement window together with the standard deviation for a single spectrum in the respective colors.

The dip at  $1630\text{ cm}^{-1}$  is likely to be an artifact present only in our data, which may be due to scaled subtraction of the instrument background spectra that were measured with saline in the sample holder. Since the amount of saline in the sampling volume without sample was much larger than with a sample, this may result in a negative concentration/contribution of water in the preprocessed spectra. Another possible explanation is an insufficient instrument response correction because the dip is also present both in the spectra without background and without instrument response correction, see Fig. 2(a). The similarity of the preprocessed spectra, despite the characteristic difference in background spectra of either silica or  $\text{CaF}_2$  measurement window [see Fig. 2(b)], suggests that our method could remove most of the specific spectral features of the measurement window.

### 3.2 Raman Spectral Variation Within and Between Different Tissue Types

In Fig. 3(a), mean preprocessed Raman spectra of different groups of tissue are shown. The mean spectra of class 1 in class definitions A and B (consisting of either iodine positive squamous epithelium only or other nonabnormal tissue types) are grouped separately from the mean spectra of class 2 in class definitions A and D (consisting of either HGCIN or HGCIN plus HGCIN-border tissue) in the range of  $450$  to  $670\text{ cm}^{-1}$ ,  $900$  to  $950\text{ cm}^{-1}$ ,  $1200$  to  $1300\text{ cm}^{-1}$ , and  $1530$  to  $1570\text{ cm}^{-1}$ .

The normal squamous epithelium showed the largest differences to all other spectra, possibly due to anatomical differences. Squamous epithelium with dark iodine staining was mostly found decentral in CIN-containing samples, while the CIN lesions were predominantly found in central areas of cone biopsies within the transformation zones. The similarity of HGCIN spectra and HGCIN plus three times as many HGCIN-border spectra suggests that these measurement sites were either partially HGCIN (which is likely due to the chosen procedure of spectra assignment) or the adjacent tissue showed some biochemical features of CIN as it was previously suggested in Ref. 24.

Spectra of erosion were similar to spectra of HGCIN plus HGCIN-borders, except for the large peak at  $1550\text{ cm}^{-1}$ . This is likely due to the fact that erosion often coexisted with focal CIN, and these spectra were consequently assigned to HGCIN-border spectra as well. Spectra of erosion without colocalized CIN were found only in one sample and are, therefore, not shown here. For all the spectra shown in

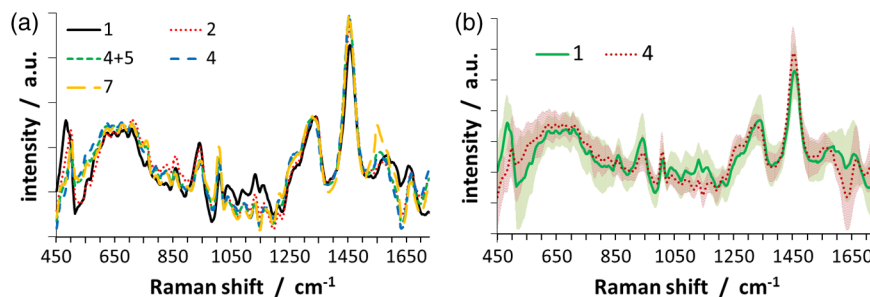
Fig. 3(a), the standard deviation of a single spectrum was high compared to the differences of mean spectra, see Fig. 3(b), where the standard deviation is shown for the two most different spectra.

For a better visualization of spectral differences, difference mean spectra are shown in Fig. 4. The spectra contributing to the mean spectra were chosen according to the class definitions in Table 1. As expected from the anatomical differences between squamous epithelium and the transformation zone where neoplasia typically originates, the largest spectral differences were found for normal squamous epithelium and HGCIN. The trend of the spectral differences was similar in all four cases. A Raman signal decrease of HGCIN tissue and borders compared to normal and benign tissue was observed at  $450$  to  $500$ ,  $850$  to  $870$ ,  $900$  to  $960$ ,  $1090$  to  $1170$ ,  $1330$  to  $1410$ , and  $1590$  to  $1670\text{ cm}^{-1}$ . A Raman signal increase of HGCIN tissue and borders compared to normal and benign tissue was observed at  $500$  to  $670$ ,  $1200$  to  $1320$ ,  $1410$  to  $1450$ , and  $1530$  to  $1570\text{ cm}^{-1}$ . The large difference at  $480\text{ cm}^{-1}$  has been previously assigned to glycogen,<sup>49</sup> which is characteristic for normal squamous epithelium. The large difference around  $550\text{ cm}^{-1}$  was difficult to assign due to the potential influence of a background removal artifact. The peaks at  $\sim 855$  and  $940\text{ cm}^{-1}$  were previously reported and were assigned to glycogen<sup>49,50</sup> as well as to collagen.<sup>50,51</sup>

A broad difference around  $1130\text{ cm}^{-1}$  has not been found in the literature. The difference around  $1250\text{ cm}^{-1}$  has been previously found<sup>21,23,25,27,28,50,51</sup> (assigned to proteins, DNA, and lipids), however, with an opposite trend, while in Ref. 49, an increase of this spectral region was also found in cancer (assigned to collagen).

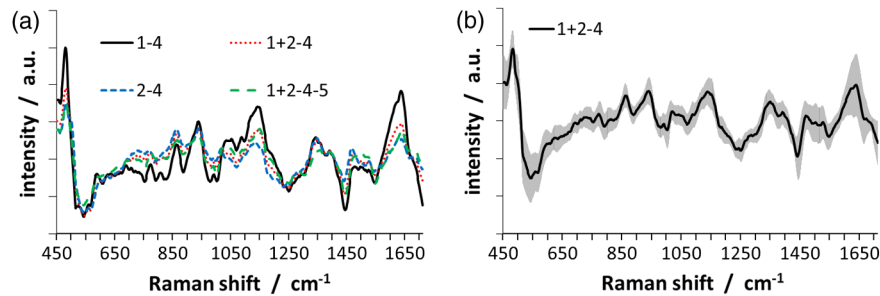
A decrease at around  $1350\text{ cm}^{-1}$  for precancer has been observed, which may be assigned to glycogen (at  $1336\text{ cm}^{-1}$ , Ref. 49). Others, however, detected an increase of the region in precancer or cancer around  $1330$  to  $1350\text{ cm}^{-1}$  (Refs. 20, 23, 25, and 27) and at  $1350$  to  $1425\text{ cm}^{-1}$  (Refs. 23 and 25), but also an increase at  $1350$  to  $1425\text{ cm}^{-1}$  is reported in Refs. 20 and 51.

The sharp decrease at  $1440\text{ cm}^{-1}$  has also been reported;<sup>45</sup> however, opposite trends were also observed.<sup>27</sup> The spectrally broad decrease of precancer/cancer at around  $1630\text{ cm}^{-1}$  has been previously reported in many studies,<sup>21,23,25,27,51</sup> but also opposite results were reported.<sup>20</sup> We believe that partially inconsistent spectral trends in the literature may be due to different sampling volumes or spectra preprocessing.



**Fig. 3** (a) Mean of preprocessed Raman spectra of all spectra from normal squamous epithelium, non-abnormal tissue, HGCIN, HGCIN plus HGCIN-borders, and erosion. (b) Mean of preprocessed Raman spectra of all spectra from normal squamous epithelium and HGCIN together with the standard deviation for single spectra. The numbers of the contributing tissue categories according to Table 1 are given in the legend.



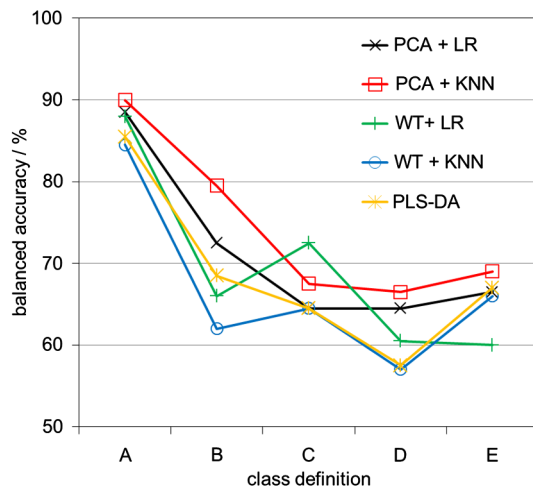


**Fig. 4** (a) Differences of the average spectra of class 1 or 2 (class 1 minus class 2) for each of the class definitions A, B, C, or D described in Table 1. The numbers of the contributing pathology groups according to Table 1 are given in the legend. (b) Errors bars obtained by patient-wise bootstrapping are shown for the difference spectra (class 1 minus class 2) of class definition B.

### 3.3 Diagnostic Accuracy for the Separation of HGCIN from non-HGCIN Depending on the Tissue Types Included in the Classification Procedure

In order to evaluate the classification performance depending on the included tissue types, separate training and evaluation was performed for each of the different class definitions described in Table 1. In addition, five classification procedures (see Table 2) were compared for each of the class definitions.

Figure 5 shows that the balanced accuracy for each class definition A to E varies among different classification procedures 1 to 5. The variation between different classification methods is smallest for class definition A. Despite the large number of spectra, the statistical uncertainty of individual balanced accuracy values was large. Using patient-wise bootstrapping resulted in a standard deviation (1 sigma) of 9% for a value obtained with the class definition A and 12% with class definition B, respectively (individual error bars in Fig. 5 were not shown to improve clarity). This large variability in classification performance might be due to the large interpatient variability of the small number of patients and the unbalanced distribution of



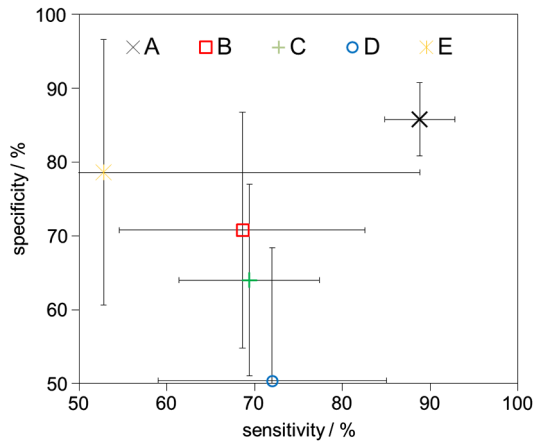
**Fig. 5** Balanced accuracy calculated using different classification procedures 1 to 5 (described in Table 2) for different class definitions A to E (described in Table 1). The statistical uncertainty (standard deviation, 1 sigma), corresponding to the standard deviation of repeated classification cycles using bootstrapping, was estimated to be 9% for a value obtained with the class definition A and 12% with class definition B, respectively (error bars in Fig. 5 were not shown to improve clarity).

HGCIN and non-HGCIN spectra in each sample. Consequently, it was concluded that differences among the different classification procedures were not significant in our study.

To compare the performance for different class definitions A to E, the average balanced accuracies of the five classification methods were used, which were 87% (A), 70% (B), 67% (C), 61% (D), and 69% (E). Judging the significance of these average balanced accuracies based on the statistical uncertainty of the single values obtained by bootstrapping seems not to be trivial, since it cannot be assumed that the results of the different classification methods are statistically uncorrelated. However, it was assumed that a realistic estimate of the uncertainty is between the uncertainty of a single value and the standard error of the mean value (1 to 4%).

The highest balanced accuracy was obtained when HGCIN was discriminated against normal squamous epithelium. A major result was that the balanced accuracy decreased when the heterogeneity of class 1 (non-HGCIN tissue) increased (class definitions B and E). The result for class definition E is relevant because this case is the most representative of a real imaging scenario, where HGCIN should be detected within all other non-HGCIN tissue sites. A slight decrease was observed from class definition B to C, when clinically normal (iodine positive) tissue was excluded from analysis (class definition C). The lowest accuracy was obtained when the HGCIN-border spectra are included into the HGCIN class, which is not surprising because the fraction of HGCIN-border spectra, which were actually non-HGCIN spectra, was expected to decrease the diagnostic performance.

Our very good Raman-based differentiation of HGCIN from normal squamous epithelium only (class definition A) does not represent a realistic estimate of diagnostic performance of Raman spectroscopy in a clinical setting. This is because only two special types of tissue are included in the analysis, which (in the absence of tissues with similar colposcopic appearance) could be also distinguished by application of acetic acid. Furthermore, it has been shown previously for fluorescence and reflectance spectroscopy of the cervix that the common practice of including clinically normal squamous sites in the analysis leads to artificially improved performance in distinguishing high-grade lesions from clinically suspicious non-high-grade lesions because underlying differences in tissue anatomy can have a confounding effect on spectroscopic parameters.<sup>48</sup> Therefore, the decrease in performance when squamous epithelia (without dark iodine stain), columnar epithelium, metaplasia, and cervicitis were included into the non-HGCIN tissue (class definition B), and the further decrease in performance

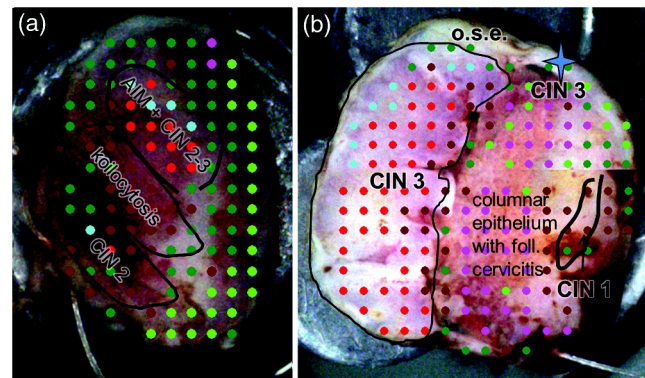


**Fig. 6** Average sensitivity and specificity of different classification methods for different class definitions A to E. Error bars correspond to the standard deviation between different classification procedures described in Table 2.

when the normal iodine positive tissue from class definition A is left out in class 1 (class definition C), suggests that our Raman analysis might have been influenced by confounding differences in anatomy as well. Including metaplastic tissue into the analysis seems fundamental since metaplasia is a benign transformation of columnar to squamous epithelium in the transformation zone, where most of the dysplastic lesions originate. Also, benign changes (such as metaplasia or inflammation) can have a similar colposcopic appearance as precancer, which is the reason why a Raman-based separation from precancer would be an improvement over colposcopy. Therefore, class definition B was chosen to produce maps of the predicted class memberships, which are merged with the histopathological mapping for a spatially resolved evaluation of the prediction accuracy, see Fig. 7.

The interpretation of the maps was challenged by the large statistical variance of the diagnostic performance (see Fig. 6). Different methods gave different mapping results and no method proved to be the best for all cases. In many cases, it was impossible to judge agreement with histopathology objectively due to the limited precision of the spatial correlation, especially in regions with focal CIN. Furthermore, repetitive measurements with different window materials give conflicting results despite the fact that the data measured on both window types were included in the model training. Despite these limitations, there were promising results, such as those shown in Fig. 7(a), where it is shown that the HGCIN region was predicted well, but misclassifications occurred in a region with koilocytosis. In Fig. 7(b), misclassifications are shown in a region with follicular cervicitis. In the only sample where erosion was present without isles of focal CIN, all methods classified erosion predominantly as HGCIN (data not shown). In principle, if the statistical basis was significantly better, the interpretation of the predicted class memberships of individual measurement locations would be justified. This would allow evaluation of the performance of the method for determination of the spatial extension of HGCIN or the performance for less frequent pathologies.

The diagnostic performances achieved in our study partially cover the large range of results from previously published studies,<sup>20–27</sup> which is not surprising given the large differences in sensitivities and specificities we obtained by varying methods and groups of tissue pathologies included in either class 1 or



**Fig. 7** Examples of spatially resolved prediction of class membership by Raman spectroscopy [with partial least squares discriminant analysis (PLS-DA)] merged with histopathological mapping. Pseudocolor meanings: correctly classified as nonabnormal (green dot), correctly classified as HGCIN (red dot), classified as nonabnormal (dark green dot, not in training data set), classified as HGCIN (brown dot, not in training data set), misclassified as nonabnormal (blue dot), misclassified as HGCIN (pink dot). Missing dots are due to signal saturation. O.s.e. means original squamous epithelium. The map in (b) was acquired in two parts.

2 for the analysis (class definitions A to E) and by doing resampling with bootstrapping. In general, results of individual studies are difficult to compare due to the different experimental and data analysis methods used as well as due to the differences in tissue pathologies included or disease thresholds used. When comparing the average performance of all classification procedures using class definition B shown in Fig. 6, we obtained a lower sensitivity (69%) and specificity (71%) than all known previous studies. However, the results of other studies with balanced accuracies of ~82% are still within the range of the standard deviation of different classification procedures or the error of an individual method estimated by bootstrapping.

There may be various reasons for a decreased performance of our approach compared to previous studies. For example, the excellent results in Ref. 23 for the discrimination of HGCIN from low-grade CIN, metaplasia, and normal tissues have been obtained using a multiclass classifier, which was assumed to be beneficial for discrimination of HGCIN from multiple other tissue pathologies. Since the KNN method we used does not make assumptions on the distribution function of features, it should not suffer from an increased variance in the data due to the assignment of several tissue types to one of two classes. Furthermore, we can only speculate that preselection of measurement sites by colposcopy as in the previous fiber probe studies favors the measurement of tissue sites with distinct appearances in contrast to our raster scan method, with which we expected to collect spectra with a larger variance. Another possible influence on the classification performance is the inter-individual variance. It was found to be larger than the intraindividual variance in this study and is captured by a smaller sample size than in other studies. In addition, we assume that our method of correlating histopathological mapping to measurement sites was more susceptible to registration/assignment errors than a study measuring on biopsy sites.

#### 4 Conclusion and Outlook

Based on our results, improvements for future studies are highly desirable. Most importantly, due to the strong interpatient variability and the resulting large statistical uncertainty of the

obtained diagnostic performance, a future study would benefit from larger patient numbers. However, the detailed histopathology required in our study design did not allow us (for capacity reasons) to include as many patients as it would have been possible with a typical fiber probe based study in which the reference is often colposcopy for normal spectra or histopathology of few abnormal sites per patient. Based on our obtained confidence interval, we estimated that in order to reduce the variability in spectra classification such that a 95% confidence interval of 10% (+/− 5%) is reached, we would require ~400 patients. A larger study might be able to show whether the spatial extension of high-grade CIN can be robustly determined within all the low-grade and benign tissue pathologies that might be present on the cervix.

Some technical improvements would also be beneficial. For example, a tool for a better spatial correlation of histopathological mapping and measurement position could increase the number of spectra that can be evaluated quantitatively.

As the dominating source of noise was the very high background, which led to a strong influence of baseline and background correction, the benefits of instrumental changes such as higher laser wavelength or the use of shifted excitation or modulated Raman spectroscopy<sup>52,53</sup> should be investigated.

Since the application of fluorescence and reflectance spectroscopy can achieve reasonable sensitivities and specificities as well,<sup>54</sup> a multimodal approach might help to further improve the diagnostic accuracy. Designs of multimodal probes/setup have been reported and the performance of a combination of Raman with UV-visible and near-infrared (NIR)-excited fluorescence spectroscopy was investigated for other applications.<sup>55–57</sup> For our data we did not see an improvement in diagnostic accuracy when the strength and slope of the NIR-excited autofluorescence background as well as the Rayleigh peak of the laser line were used as additional input for the classification algorithm.

Cantor et al.<sup>54</sup> modeled the cost-effectiveness of fluorescence spectroscopy in a see-and-treat scenario. They found that if spectroscopy was able to achieve sensitivities of at least 84% and specificities of at least 76%, there would indeed be huge savings in health care dollars from “biopsies avoided” and in being able to “see-and-treat” more accurately with a loop electrical excision procedure. This is within the range of both our and previous results, which is why further activities toward translation into clinical routine use are desirable. However, huge improvements in colposcopy have been reported in a recent study.<sup>7</sup> Detection of HGCIN with a combined sensitivity of 78% and specificity of 93% has been achieved using inner border sign, ridge sign, and the newly defined rag sign, without considering the conventional colposcopic criteria. These pathognomonic signs, introduced in the latest terminology of the International Federation for Cervical Pathology and Colposcopy (2011), are objective, present or absent, easy to see, easy to learn, with little space for subjectivity, and easier to use than a grading index.<sup>7</sup> Consequently, a higher diagnostic performance as proposed by Cantor et al. might be required for a new technology such as Raman spectroscopy for entering routine clinical diagnostics in the future.

Previous *in vivo* studies have shown the diagnostic performance of colposcopy guided fiber probe based Raman spectroscopy.<sup>20–27</sup> Our study was aimed at the development of a wide-area analysis of the cervix. One advantage of such an approach compared to the use of fiber probes is, assuming sufficient spatial resolution is provided, that small lesions would

not be missed due to undersampling. Since the colposcopic information is not used to select measurement sites, such a method could, in principle, be used without colposcopy. Therefore, further studies aimed at the development of *in vivo* Raman imaging would seem to be desirable.

### Acknowledgments

This project was cofinanced by the Senate of Berlin and by the European Union (European funds for regional development, FKZ 10147189). We thank Günter Cichon and Achim Schneider (Charité University Hospital, Department of Gynecology, Campus Benjamin Franklin) and all staff members for their active support of the study. We would like to thank our colleagues Stefan Ey, Ihar Shchatsinin, Stefan Baar, and Lutz Krebs for their contributions to software implementation and the instrumental setup. We also thank our colleagues Daniela Schädel for useful advice and Caroline Reid and Lesley Hirst for proofreading the manuscript. Furthermore, we would like to thank the reviewers for their helpful comments.

### References

1. H. zur Hausen, “Papillomavirus infections—a major cause of human cancers,” *Biochim. Biophys. Acta* **1288**(2), F55–78 (1996).
2. M. A. Duggan et al., “The natural history of CIN I lesions,” *Eur. Gynaecol. J. Oncol.* **19**(4), 338–344 (1998).
3. A. G. Ostor, “Natural history of cervical intraepithelial neoplasia: a critical review,” *Int. Gynecol. J. Pathol.* **12**(2), 186–192 (1993).
4. W. A. McIndoe et al., “The invasive potential of carcinoma in situ of the cervix,” *Obstet. Gynecol.* **64**(4), 451–458 (1984).
5. M. R. E. McCredie et al., “Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study,” *Lancet Oncol.* **9**(5), 425–434 (2008).
6. K. Nanda et al., “Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review,” *Ann. Intern. Med.* **132**(10), 810–819 (2000).
7. G. F. Vercellino et al., “Validity of the colposcopic criteria inner border sign, ridge sign, and rag sign for detection of high-grade cervical intraepithelial neoplasia,” *Obstet. Gynecol.* **121**(3), 624–631 (2013).
8. M. Coppleson, “Colposcopic features of papillomaviral infection and premalignancy in the female lower genital tract,” *Obstet. Gynecol. Clin. North Am.* **14**(2), 471–494 (1987).
9. A. Staffl and R. F. Mattingly, “Colposcopic diagnosis of cervical neoplasia,” *Obstet. Gynecol.* **41**(2), 168–176 (1973).
10. R. Reid et al., “Genital warts and cervical cancer. IV. A colposcopic index for differentiating subclinical papillomaviral infection from cervical intraepithelial neoplasia,” *Obstet. Am. J. Gynecol.* **149**(8), 815–823 (1984).
11. D. G. Hong et al., “Prediction of high-grade squamous intraepithelial lesions using the modified Reid index,” *Int. Clin. J. Oncol.* **15**(1), 65–69 (2010).
12. L. S. Massad et al., “The accuracy of colposcopic grading for detection of high-grade cervical intraepithelial neoplasia,” *J. Low. Genit. Tract Dis.* **13**(3), 137–144 (2009).
13. M. Sideri et al., “Interobserver variability of colposcopic interpretations and consistency with final histologic results,” *J. Low. Genit. Tract Dis.* **8**(3), 212–216 (2004).
14. J. W. Sellors et al., “Observer variability in the scoring of colpophotographs,” *Obstet. Gynecol.* **76**(6), 1006–1008 (1990).
15. I. J. Etherington et al., “Video colposcopy: a new technique for secondary cervical screening,” *Br. J. Obstet. Gynaecol.* **104**(2), 150–153 (1997).
16. D. G. Ferris and M. Litaker, “Interobserver agreement for colposcopy quality control using digitized colposcopic images during the ALTS trial,” *J. Low. Genit. Tract Dis.* **9**(1), 29–35 (2005).
17. M. E. Baum et al., “Colposcopic accuracy of obstetrics and gynecology residents,” *Gynecol. Oncol.* **103**(3), 966–970 (2006).
18. S. J. Goldie et al., “Cost-effectiveness of cervical-cancer screening in five developing countries,” *N. Engl. J. Med.* **353**(20), 2158–2168 (2005).

19. A. Mahadevan-Jansen et al., "Near-infrared Raman spectroscopy for in vitro detection of cervical precancers," *Photochem. Photobiol.* **68**(1), 123–132 (1998).
20. U. Utzinger et al., "Near-infrared Raman spectroscopy for in vivo detection of cervical precancers," *Appl. Spectrosc.* **55**(8), 955–959 (2001).
21. A. Robichaux-Viehoever et al., "Characterization of Raman spectra measured in vivo for the detection of cervical dysplasia," *Appl. Spectrosc.* **61**(9), 986–993 (2007).
22. E. M. Kanter et al., "Multiclass discrimination of cervical precancers using Raman spectroscopy," *J. Raman Spectrosc.* **40**(2), 205–211 (2009).
23. E. M. Kanter et al., "Application of Raman spectroscopy for cervical dysplasia diagnosis," *J. Biophotonics* **2**(1–2), 81–90 (2009).
24. E. M. Kanter et al., "Effect of hormonal variation on Raman spectra for cervical disease detection," *Am. J. Obstet. Gynecol.* **200**(5), 512.e1-5 (2009).
25. E. Vargis et al., "Effect of normal variations on disease classification of Raman spectra from cervical tissue," *Analyst* **136**(14), 2981–2987 (2011).
26. J. Mo et al., "High wavenumber Raman spectroscopy for in vivo detection of cervical dysplasia," *Anal. Chem.* **81**(21), 8908–8915 (2009).
27. S. Duraipandian et al., "In vivo diagnosis of cervical precancer using Raman spectroscopy and genetic algorithm techniques," *Analyst* **136**(20), 4328–4336 (2011).
28. H. d. S. Martinho et al., "Role of cervicitis in the Raman-based optical diagnosis of cervical intraepithelial neoplasia," *J. Biomed. Opt.* **13**(5), 054029 (2008).
29. M. Kyrgiou et al., "Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis," *Lancet* **367**(9509), 489–498 (2006).
30. K. Nam et al., "Random biopsy after colposcopy-directed biopsy improves the diagnosis of cervical intraepithelial neoplasia grade 2 or worse," *J. Low. Genit. Tract Dis.* **14**(4), 346–351 (2010).
31. R. G. Pretorius et al., "Utility of random cervical biopsy and endocervical curettage in a low-risk population," *J. Low. Genit. Tract Dis.* **16**(4), 333–338 (2012).
32. T. DeSantis et al., "Spectroscopic imaging as a triage test for cervical disease: a prospective multicenter clinical trial," *J. Low. Genit. Tract Dis.* **11**(1), 18–24 (2007).
33. J. E. Kendrick, W. K. Huh, and R. D. Alvarez, "LUMA cervical imaging system," *Expert Rev. Med. Devices* **4**(2), 121–129 (2007).
34. C. Yu et al., "Quantitative spectroscopic imaging for non-invasive early cancer detection," *Opt. Express* **16**(20), 16227–16239 (2008).
35. C. Reble et al., "Influence of tissue absorption and scattering on the depth dependent sensitivity of Raman fiber probes investigated by Monte Carlo simulations," *Biomed. Opt. Express* **2**(3), 520–533 (2011).
36. R. Hornung et al., "Quantitative near-infrared spectroscopy of cervical dysplasia in vivo," *Hum. Reprod.* **14**(11), 2908–2916 (1999).
37. C. R. Weber et al., "Model-based analysis of reflectance and fluorescence spectra for in vivo detection of cervical dysplasia and cancer," *J. Biomed. Opt.* **13**(6), 064016 (2008).
38. S. Regauer and O. Reich, "CK17 and p16 expression patterns distinguish (atypical) immature squamous metaplasia from high-grade cervical intraepithelial neoplasia (CIN III)," *Histopathology* **50**(5), 629–635 (2007).
39. B. S. Apgar, L. Zoschnick, and T. C. Wright, "The 2001 Bethesda system terminology," *Am. Fam. Physician* **68**(10), 1992–1998 (2003).
40. I. Barman et al., "Turbidity-corrected Raman spectroscopy for blood analyte detection," *Anal. Chem.* **81**(11), 4233–4240 (2009).
41. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Appl. Spectrosc.* **43**(5), 772–777 (1989).
42. G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," *Lec. Notes Stat.* **103**, 281–299 (1995).
43. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
44. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, New York (2006).
45. F. Murtagh and A. Heck, *Multivariate Data Analysis*, D. Reidel Publishing, Dordrecht (1987).
46. M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemom.* **17**(3), 166–173 (2003).
47. R. Wehrens, H. Putter, and L. Buydens, "The bootstrap: a tutorial," *Chemometr. Intell. Lab. Syst.* **54**(1), 35–52 (2000).
48. J. Mirkovic et al., "Effect of anatomy on spectroscopic detection of cervical dysplasia," *J. Biomed. Opt.* **14**(4), 044021 (2009).
49. F. M. Lyng et al., "Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool," *Exp. Mol. Pathol.* **82**(2), 121–129 (2007).
50. L. E. Kamemoto et al., "Near-infrared micro-Raman spectroscopy for in vitro detection of cervical cancer," *Appl. Spectrosc.* **64**(3), 255–261 (2010).
51. C. M. Krishna et al., "Raman spectroscopy studies for diagnosis of cancers in human uterine cervix," *Vib. Spectrosc.* **41**(1), 136–141 (2006).
52. K. Sowoidnich and H. Kronfeldt, "Shifted excitation Raman difference spectroscopy at multiple wavelengths for in-situ meat species differentiation," *Appl. Phys. B* **108**(4), 975–982 (2012).
53. B. B. Praveen et al., "Fluorescence suppression using wavelength modulated Raman spectroscopy in fiber-probe-based tissue analysis," *J. Biomed. Opt.* **17**(7), 077006 (2012).
54. S. B. Cantor et al., "Accuracy of optical spectroscopy for the detection of cervical intraepithelial neoplasia: testing a device as an adjunct to colposcopy," *Int. J. Cancer* **128**(5), 1151–1168 (2011).
55. R. Cicchi et al., "Combined fluorescence-Raman spectroscopic setup for the diagnosis of melanocytic lesions," *J. Biophotonics* **7**(1–2), 86–95 (2014).
56. Z. Huang et al., "Raman spectroscopy in combination with background near-infrared autofluorescence enhances the in vivo assessment of malignant tissues," *Photochem. Photobiol.* **81**(5), 1219–1226 (2005).
57. O. R. Scepanovic et al., "Multimodal spectroscopy detects features of vulnerable atherosclerotic plaque," *J. Biomed. Opt.* **16**(1), 011009 (2011).

**Carina Reble** graduated from the Physics Department of the Ruprecht-Karls University in Heidelberg in 2007. Since 2008 she has been a research scientist at LMTB in Berlin. She is a PhD student at the Technical University Berlin. Her degree work is concerned with Raman spectroscopic methods for analysis of skin and cervix.

**Ingo Gersonde** received his PhD degree in physics from Freie Universität Berlin, Germany. At the LMTB Berlin he is involved in Raman spectroscopy and remission spectroscopy. His work also includes chemometric and classification methods for analysis of spectra and sensor data.

**Cathrin Dressler** received her diploma in biology from Freie Universität Berlin in 1989 and joined the department of laser medicine at the Benjamin-Franklin University hospital as assistant scientist. After receiving her PhD in 1996 she joined the company Laser- und Medizin-Technologie GmbH, Berlin, as research scientist. Working in various co-operate research projects she focused on developing new nearfield- and farfield-microscopy technologies as well as nanotechnology and Raman spectroscopy.

**Jürgen Helfmann** received his PhD degree in physics from Freie Universität Berlin, Germany. At the LMTB Berlin he heads the resorts of medical technology and spectroscopy for diagnosis and sensor development. Working for 27 years alternately for the Institute of Medicine/Technical Physics & Laser Medicine at the Charité and the Laser Medizin Zentrum, now the LMTB, he gained extensive experience in various fields of biomedical optics.

**Wolfgang Kühn** is a pathologist and a physician for gynecology and obstetrics. He is specialized in colposcopy and cervical diseases and was professor and head of the department of cytology and gynecological morphology at the Charite University Clinic Berlin from 1991 to 2011. Since his retirement he works in a private medical office and lab. He is the vice chairman of the German Society for Colposcopy and Cervical Pathology and published numerous papers and books.

**Hans Joachim Eichler** is head of the Laser-Group at the Institute of Optics and Atomic Physics at the TU Berlin. He was previously an advisor to the German Ministry of Education and Research (BMBF) and a member of the technical staff in Bell Labs Holmdel, USA, working on nonlinear optics of silicon and semiconductor lasers. From 2001 to 2013, he acted as CTO and CEO of Laser- und Medizin-Technologie GmbH, Berlin. He has published numerous journal papers on optics and books on lasers and photonics.