

# Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

## **Real-time Raman spectroscopy for *in vivo*, online gastric cancer diagnosis during clinical endoscopic examination**

Shiyamala Duraipandian  
Mads Sylvest Bergholt  
Wei Zheng  
Khek Yu Ho  
Ming Teh  
Khay Guan Yeoh  
Jimmy Bok Yan So  
Asim Shabbir  
Zhiwei Huang

# Real-time Raman spectroscopy for *in vivo*, online gastric cancer diagnosis during clinical endoscopic examination

Shiyamala Duraipandian,<sup>a</sup> Mads Sylvest Bergholt,<sup>a</sup> Wei Zheng,<sup>a</sup> Khek Yu Ho,<sup>b</sup> Ming Teh,<sup>c</sup> Khay Guan Yeoh,<sup>b</sup> Jimmy Bok Yan So,<sup>d</sup> Asim Shabbir,<sup>d</sup> and Zhiwei Huang<sup>a</sup>

<sup>a</sup>National University of Singapore, Department of Bioengineering, Faculty of Engineering, Optical Bioimaging Laboratory, Singapore 117576, Singapore

<sup>b</sup>National University of Singapore and National University Hospital, Yong Loo Lin School of Medicine, Department of Medicine, Singapore 119260, Singapore

<sup>c</sup>National University of Singapore and National University Hospital, Yong Loo Lin School of Medicine, Department of Pathology, Singapore 119074, Singapore

<sup>d</sup>National University of Singapore and National University Hospital, Yong Loo Lin School of Medicine, Department of Surgery, Singapore 119074, Singapore

**Abstract.** Optical spectroscopic techniques including reflectance, fluorescence and Raman spectroscopy have shown promising potential for *in vivo* precancer and cancer diagnostics in a variety of organs. However, data-analysis has mostly been limited to post-processing and off-line algorithm development. In this work, we develop a fully automated on-line Raman spectral diagnostics framework integrated with a multimodal image-guided Raman technique for real-time *in vivo* cancer detection at endoscopy. A total of 2748 *in vivo* gastric tissue spectra (2465 normal and 283 cancer) were acquired from 305 patients recruited to construct a spectral database for diagnostic algorithms development. The novel diagnostic scheme developed implements on-line preprocessing, outlier detection based on principal component analysis statistics (i.e., Hotelling's  $T^2$  and Q-residuals) for tissue Raman spectra verification as well as for organ specific probabilistic diagnostics using different diagnostic algorithms. Free-running optical diagnosis and processing time of  $<0.5$  s can be achieved, which is critical to realizing real-time *in vivo* tissue diagnostics during clinical endoscopic examination. The optimized partial least squares-discriminant analysis (PLS-DA) models based on the randomly resampled training database (80% for learning and 20% for testing) provide the diagnostic accuracy of 85.6% [95% confidence interval (CI): 82.9% to 88.2%] [sensitivity of 80.5% (95% CI: 71.4% to 89.6%) and specificity of 86.2% (95% CI: 83.6% to 88.7%)] for the detection of gastric cancer. The PLS-DA algorithms are further applied prospectively on 10 gastric patients at gastroscopy, achieving the predictive accuracy of 80.0% (60/75) [sensitivity of 90.0% (27/30) and specificity of 73.3% (33/45)] for *in vivo* diagnosis of gastric cancer. The receiver operating characteristics curves further confirmed the efficacy of Raman endoscopy together with PLS-DA algorithms for *in vivo* prospective diagnosis of gastric cancer. This work successfully moves biomedical Raman spectroscopic technique into real-time, on-line clinical cancer diagnosis, especially in routine endoscopic diagnostic applications. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.JBO.17.8.081418]

Keywords: Raman spectroscopy; cancer diagnostics; multivariate analysis; *in vivo* optical diagnosis.

Paper 121055S received Feb. 13, 2012; revised manuscript received Jul. 3, 2012; accepted for publication Jul. 5, 2012; published online Aug. 2, 2012.

## 1 Introduction

Cancer remains a global health problem and is one of the leading causes of mortality. For instance, upper gastrointestinal (GI) malignancies, such as gastric cancer, have an estimated 989,600 new cases and 738,000 deaths per year worldwide.<sup>1</sup> Early diagnosis and adequate treatment of GI cancers is essential to increasing the survival rate of the patients. Current diagnostics heavily rely on visualization of gross morphological changes using conventional white light reflectance (WLR) imaging. Recently, narrow band imaging (NBI) and autofluorescence imaging (AFI) have emerged for image-enhanced contrast of micro-vascularity and visualization of endogenous fluorophores in tissue. But these techniques remain inadequate for efficient and accurate diagnostics of grossly invisible lesions

during clinical examinations, as biopsies are typically performed randomly.<sup>2</sup> Hence, the development of real-time advanced diagnostic technologies that probe the endogenous biomolecular properties of tissues for fully automated cancer diagnostics would represent a vital advancement toward more efficient diagnosis and management of gastric patients.

Several point-wise spectroscopic techniques (e.g., fluorescence, diffuse reflectance, and near-infrared (NIR) Raman spectroscopy) have shown promising potential for early cancer diagnostic in humans.<sup>2-6</sup> NIR Raman spectroscopy represents a unique optical technology capable of nondestructively probing endogenous biomolecules to resolve highly specific diagnostic information, while the spectral signatures can unambiguously be assigned to intra- and inter-cellular constituents (e.g., proteins, lipids and nucleic acids) of tissue. For instance, *in vitro* and *in vivo* Raman spectroscopic cancer diagnosis have been demonstrated in a variety of organs (e.g., skin,<sup>7,8</sup> oral cavity,<sup>9,10</sup>

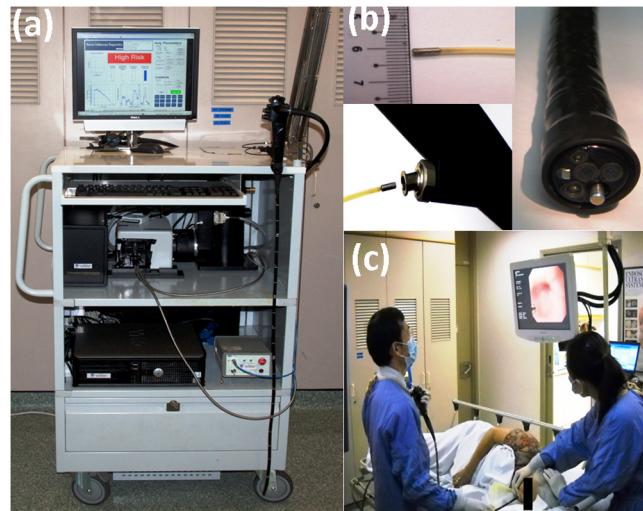
Address all correspondence to: Zhiwei Huang, National University of Singapore, Department of Bioengineering, Faculty of Engineering, Optical Bioimaging Laboratory, 9 Engineering Drive 1, Singapore 117576. Tel: +65 6516 8856; Fax: +65 6872 3069; E-mail: biezhu@nus.edu.sg

nasopharynx and larynx,<sup>11–15</sup> breast,<sup>16</sup> colon,<sup>17,18</sup> lung,<sup>19</sup> cervix,<sup>20,21</sup> bladder,<sup>22</sup> prostate,<sup>23</sup> esophagus,<sup>24</sup> stomach,<sup>5,6,25–29</sup> etc.). Very recently, we have developed a rapid, high-throughput Raman spectroscopic system integrated with multimodal wide-field imaging modalities (WLR/AFI/NBI), and demonstrated its utility for *in vivo* diagnosis of premalignant and malignant lesions in the upper GI (i.e. esophagus,<sup>24</sup> gastric<sup>5,25–27</sup>). A series of publications have outlined the efficient *in vivo* diagnosis of different gastric lesions including gastric dysplasia and neoplasia,<sup>5,25–27,29</sup> benign and malignant ulcers.<sup>6</sup> The inter-anatomical variability within the normal stomach and esophagus has also been examined in detail using image-guided Raman spectroscopy.<sup>29</sup> Several groups have also reported on the implementation of real-time tissue Raman spectroscopy.<sup>30,31</sup> For instance, Feld et al. presented a synchronized laser Raman system to study atherosclerosis that could acquire and process Raman spectra within 2 s using spectral modeling of reference biochemicals with an aid of least squares regression techniques.<sup>30</sup> Zeng et al. developed a real-time integrated Raman system for evaluating skin that could acquire and process the signal within 1.1 s.<sup>31</sup> Nevertheless, most of the current Raman spectroscopic analyses have been limited to off-line post processing for classification of spectra with cross-validation procedures, which render practical limitations including the setting of exposure times, post-verification of spectrum quality, and lack of automatic feedback mechanisms to the clinicians for implementation of straightforward probabilistic diagnostics in clinical settings. Hence, fully automated tissue spectral quality verification and real-time tissue cancer diagnostics are vital to translating the Raman spectroscopic diagnostic technique into practical clinical endoscopic routine. In this work, we develop an on-line biomedical spectral diagnostic framework integrated with image-guided Raman endoscopy for real-time probabilistic detection of cancer in the upper GI. We also validate the efficacy of the on-line Raman framework developed for prospective prediction of patients with gastric malignancies at clinical endoscopy.

## 2 Materials and Methods

### 2.1 Raman Endoscopy Instrumentation

Figure 1 shows the integrated Raman spectroscopy and trimodal wide-field imaging system developed for *in vivo* tissue Raman measurements at endoscopy, which has been described in detail elsewhere.<sup>32</sup> Briefly, the clinical Raman spectroscopic system consists of a spectrum stabilized 785 nm diode laser (maximum output: 300 mW, B&W TEK Inc., Newark, Delaware) electronically synchronized with a USB 6501 digital I/O (National Instruments, Austin, Texas), a transmissive imaging spectrograph (Holospec f/1.8, Kaiser Optical Systems, Ann Arbor, MI) equipped with a liquid nitrogen-cooled, NIR-optimized, back-illuminated and deep depletion charge-coupled device (CCD) camera (1340 × 400 pixels at 20 × 20 μm per pixel; Spec-10: 400BR/LN, Princeton Instruments, Trenton, New Jersey), and a specially designed Raman endoscopic probe for both laser light delivery and *in vivo* tissue Raman signal collection. The 1.8 mm Raman endoscopic probe is composed of 32 collection fibers surrounding the central light delivery fiber with two stages of optical filtering incorporated at the proximal and distal ends of the probe for maximizing the collection of tissue Raman signals, while reducing the interference of Rayleigh scattered light, fiber fluorescence and silica Raman signals. The Raman probe can easily pass down to the



**Fig. 1** (a) Photograph of Raman endoscopy system in clinic; (b) insertion of the 1.8 mm Raman endoscopic probe into the working channel of an endoscope during gastroscopy; and (c) routine Raman endoscopy procedure in clinic.

instrument channel of medical endoscopes and be directed to suspicious tissue sites under the guidance of wide-field endoscopic imaging (WLR/AFI/NBI) modalities.<sup>32–34</sup> The system acquires Raman spectra in the wavenumber range of 800 to 1800  $\text{cm}^{-1}$  from *in vivo* upper GI tissue within 0.5 s using the 785 nm excitation power of 1.5  $\text{W}/\text{cm}^2$  (spot size of 200 μm) with a spectral resolution of  $\sim 9 \text{ cm}^{-1}$ .

### 2.2 Study Protocol

The present study is part of an ongoing nationwide gastric cancer screening program, focusing on early diagnosis and treatment of upper GI malignancies run by the Singapore gastric cancer epidemiology, clinical and genetic program (GCEP).<sup>35</sup> All patients participating in this study preoperatively signed an informed consent, permitting the investigative *in vivo* Raman spectroscopic acquisition of gastric tissues undergoing endoscopy in the Endoscopy Centre at the National University Hospital (NUH), Singapore. This study was approved by the Institutional Review Board (IRB) of the National Healthcare Group of Singapore. A total of 2748 *in vivo* gastric tissue spectra (2465 normal and 283 cancer) were acquired from the 305 patients recruited to construct the spectral database for developing diagnostic algorithms for gastric cancer diagnostics. Tissue histopathology serves as the gold standard for evaluation of the performance of Raman technique for *in vivo* tissue diagnosis and characterization.

### 2.3 Online Biomedical Spectroscopic Framework

The online biomedical Raman spectroscopic framework developed has been implemented as a graphical user interface (GUI) under the Matlab 2011a (Mathworks Inc., Natick, MA) scripting environment in a fast computing workstation (64 bit I7 quad-core 4GB memory). This framework has been thoroughly optimized for rapid data processing for real-time tissue diagnostics. Hardware components of the rapid Raman system (e.g., laser power control, spectrometer, CCD shutter and camera readout synchronization) have been interfaced to the Matlab software through libraries for different spectrometers/cameras

[e.g., PVCAM library (Princeton Instruments, Roper Scientific, Inc., Trenton, NJ) and Omni Driver (Ocean Optics Inc., Dunedin, FL), etc.]. A schematic of the spectral acquisition and processing flow of online diagnostic framework is depicted in Fig. 2. The laser was electronically synchronized with the CCD shutter. The automatic adjustment of laser power, exposure time and accumulation of spectra were realized by scaling to within 85% of the total photon counts (e.g., 55,250 of 65,000 photons) based on preceding tissue Raman measurements, whereas an upper limit of 0.5 s was set to realize clinically acceptable conditions. The accumulation of multiple spectra and automatic adjustment of exposure time provides a rapid and straightforward methodology to prevent CCD saturation and to obtain high signal-to-noise ratios (SNR) for endoscopic applications. The Raman-shift axis (wavelength) was calibrated using a mercury/argon calibration lamp (Ocean Optics Inc., Dunedin, FL). The spectral response correction for the wavelength-dependence of the system was conducted using a standard lamp (RS-10, EG&G Gamma Scientific, San Diego, CA). The reproducibility of the platform can be

continuously monitored with the laser frequency and Raman spectra of cyclohexane and acetaminophen as wavenumber standards. All the system performance measures including CCD temperature, integration time, laser power, CCD alignment are accordingly logged into a central database via SQL server. Due to the inter-anatomical and inter-organ spectral variances as we observed earlier,<sup>9,29</sup> the online framework we designed implements organ-specific diagnostic models and can instantly search among the spectral databases of different organs [e.g., esophagus, gastric, colon, cervix, bladder, lung, nasopharynx, larynx, and the oral cavity (hard palate, soft palate, buccal, inner lip, ventral and the tongue)], making this Raman platform a universal diagnostic tool for cancer detection at endoscopy.

## 2.4 On-Line Preprocessing and Outlier Detection

Real-time preprocessing of Raman signals was realized with the rapid detection of cosmic rays using the first derivative with a 99% confidence interval (CI) over the whole spectral range set as a maximum threshold. Data points lying outside of a threshold were interpolated to 2<sup>nd</sup> order. The spectra were further scaled with integration time and laser power. A first-order, five point Savitzky-Golay smoothing filter was used to remove noise in the intensity corrected spectra, while a 5<sup>th</sup> order modified polynomial constrained to the lower bound of the smoothed spectra was subtracted to resolve the tissue Raman spectrum alone. The Raman spectra were finally normalized to the integrated areas under the curves from 800 to 1800  $\text{cm}^{-1}$ , enabling a better comparison of the spectral shapes and relative Raman band intensities among different tissue pathologies. The spectra were then locally mean-centered according to the specific database to remove common variations in the data. Following preprocessing, the Raman spectra were fed to a model-specific outlier analysis.

We incorporate an on-line outlier detection scheme into biomedical spectroscopy as a high-level model-specific feedback tool in our on-line framework by using principal component analysis (PCA) coupled with Hotelling's  $T^2$  and Q-residual statistics<sup>36-38</sup> (Fig. 2). Briefly, PCA reduces the dimension of the Raman spectra by decomposing them into linear combinations of orthogonal components [principal components (PCs)], such that the spectral variations in the dataset are maximized. The PCA model of the data matrix  $X$  is defined by Ref. 38:

$$X = TP^T + E, \quad (1)$$

where  $T$  and  $P$  represent scores and loadings, and  $E$  contains the residuals. The loadings correspond to the new rotated axis, whereas scores represent the data projection values. Accordingly, Hotelling's  $T^2$  statistics is a measure of variance captured by the PCA model (sample to model distance) and is defined by Ref. 36:

$$T_{ik}^2 = t_{ik}(\lambda_k^{-1})t_{ik}^T, \quad (2)$$

where  $t_{ik}$  is PC scores for  $i$ 'th sample spectrum using component  $k$ , and  $\lambda_k^{-1}$  is the diagonal matrix of normalized eigenvalues of the covariance matrix for component  $k$ . Therefore, Hotelling's  $T^2$  gives an indication of extreme values within the PCA model. On the other hand, Q-residuals is a measure of variance, which is not captured by the PCA model (lack of model fit statistics), and is defined by Ref. 36:

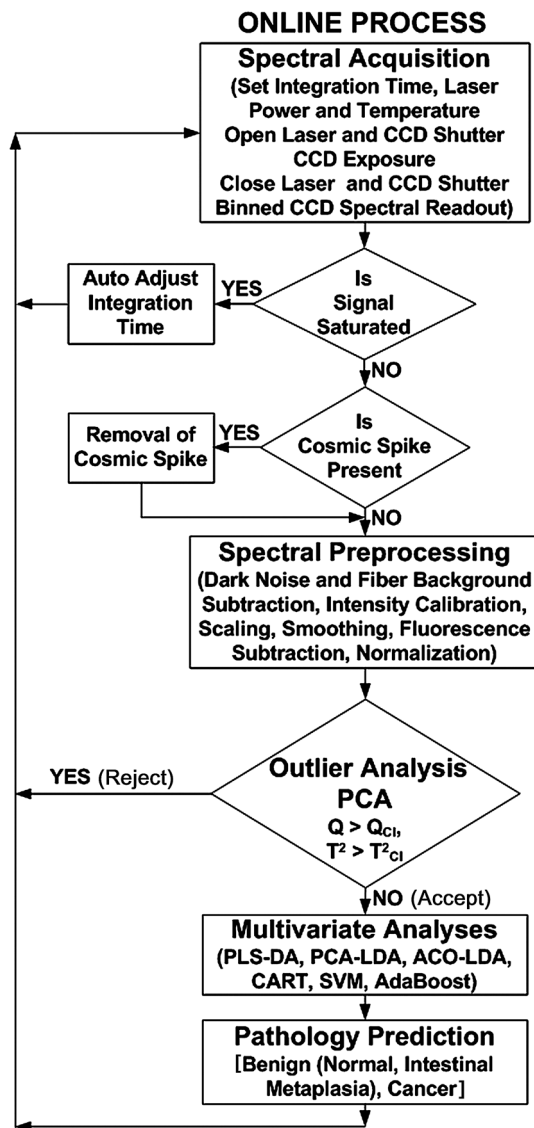


Fig. 2 Flow chart of the on-line biomedical Raman spectroscopic diagnostic platform.

$$Q_{ik} = \sum (x_i - t_{ik}P_k^T)^2, \quad (3)$$

where  $x_i$  is the sample spectrum,  $Q_{ik}$  is the sum of squared reconstruction error for  $i$ 'th sample spectrum using component  $k$  and  $P_k$  is the PC loadings. For both Hotelling's  $T^2$  and Q-residuals, the normalized 99% CI was utilized as upper thresholds to intercept anomalous Raman spectra (e.g., noncontact measurements, spectra from blood, gastric juice, food debris, etc.) (Fig. 2). Accordingly, the Hotelling's  $T^2$  and Q-residuals are the two independent parameters providing quantitative information about the model fit. Using these parameters as indicators of spectra quality (i.e., probe contact mode, confounding factors, white light interference, etc.), auditory feedback has been integrated into the on-line Raman diagnostic system, facilitating real-time probe handling advice and spectroscopic screening for clinicians during clinical endoscopic procedures.

### 2.5 Online Probabilistic Diagnostics

Subsequent to verification of tissue Raman spectra quality, those qualified Raman spectra were immediately fed to probabilistic models for on-line *in vivo* diagnostics and pathology prediction. The GUI can instantly switch among different models including partial least squares-discriminant analysis (PLS-DA),<sup>6,20</sup> PCA-linear discriminant analysis (LDA),<sup>26,28</sup> ant colony optimization (ACO)-LDA,<sup>25</sup> classification and regression trees (CART),<sup>39</sup> support vector machine (SVM),<sup>17</sup> adaptive boosting (AdaBoost),<sup>40</sup> etc. for prospective classification at clinical endoscopic procedures. In this study, as an example, probabilistic

PLS-DA was employed for gastric cancer diagnosis. PLS-DA employs the fundamental principle of PCA but further rotates the components by maximizing the covariance between the spectral variation and group affinity to obtain the diagnostically relevant variations rather than the most prominent variations in the spectral dataset.<sup>6</sup> The system supports binary classification, one-against-all and one-against-one multiclass (i.e., benign, dysplasia, and cancer) probabilistic PLS-DA discriminatory analysis to predict the specific tissue pathologies.

### 3 Results

Figure 3 shows the developed novel GUI for on-line biomedical spectroscopic processing and diagnostics. Specifically, we test the developed on-line framework in the stomach that represents one of the most challenging organs presenting with many confounding factors (i.e., gastric juice, food debris, bleeding, exudates, etc.) for spectroscopic diagnosis. The *in vivo* mean Raman spectra acquired from 305 gastric patients [normal ( $n = 2465$ ) and cancer ( $n = 283$ )] for algorithms development are shown in Fig. 4. The Raman spectra of gastric tissue show the prominent Raman peaks at  $875 \text{ cm}^{-1}$  [ $\nu(\text{C}-\text{C})$  of hydroxyproline],  $936 \text{ cm}^{-1}$  [ $\nu(\text{C}-\text{C})$  of proteins],  $1004 \text{ cm}^{-1}$  [ $\nu_s(\text{C}-\text{C})$  ring breathing of phenylalanine],  $1078 \text{ cm}^{-1}$  [ $\nu(\text{C}-\text{C})$  of lipids],  $1265 \text{ cm}^{-1}$  [amide III  $\nu(\text{C}-\text{N})$  and  $\delta(\text{N}-\text{H})$  of proteins],  $1302$  and  $1335 \text{ cm}^{-1}$  [ $\delta(\text{CH}_2)$  deformation of proteins and lipids],  $1445 \text{ cm}^{-1}$  [ $\delta(\text{CH}_2)$  of proteins and lipids],  $1618 \text{ cm}^{-1}$  [ $\nu(\text{C}=\text{C})$  of porphyrins],  $1652 \text{ cm}^{-1}$  [amide I  $\nu(\text{C}=\text{O})$  of proteins] and  $1745 \text{ cm}^{-1}$  [ $\nu(\text{C}=\text{O})$  of lipids].<sup>5,28</sup> Clearly, gastric tissue Raman spectra contain a large contribution from triglyceride (i.e., major peaks at 1078, 1302, 1445, 1652, and

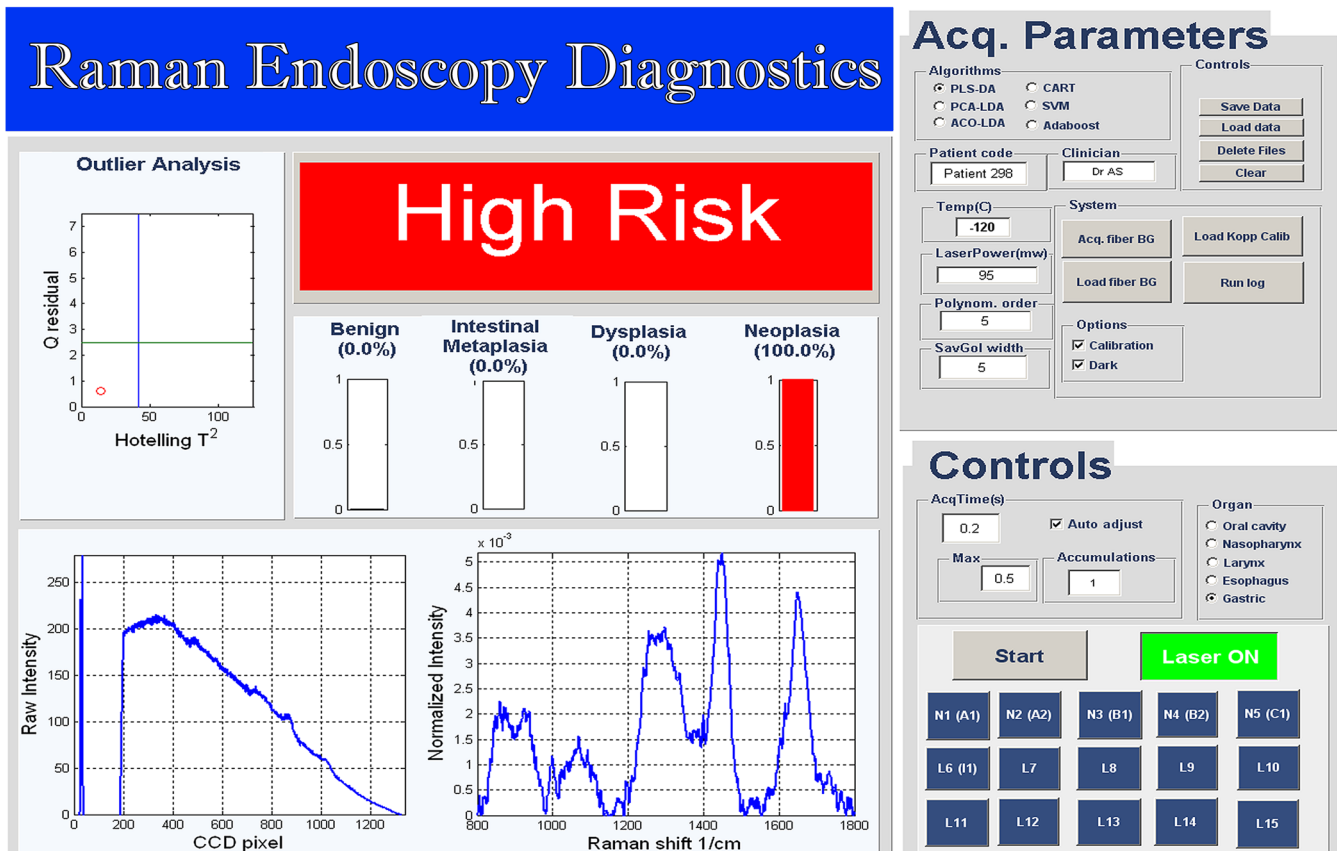
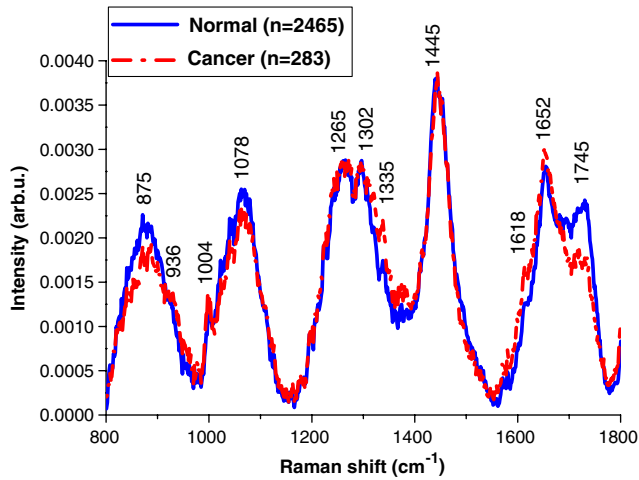


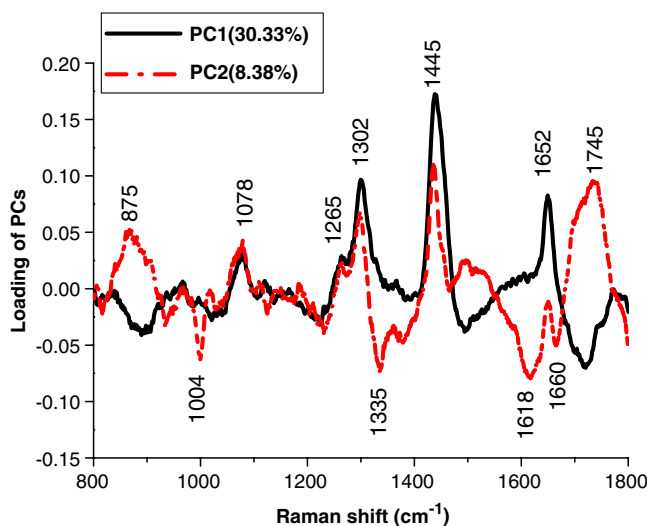
Fig. 3 On-line biomedical Raman spectroscopic diagnostic platform for gastric cancer detection.



**Fig. 4** *In vivo* mean Raman spectra of normal ( $n = 2465$ ) and cancer ( $n = 283$ ) gastric tissue acquired from 305 gastric patients.

1745  $\text{cm}^{-1}$ ) that likely reflects the interrogation of subcutaneous fat in the gastric wall.<sup>6,29</sup> Additionally, we also observed remarkable Raman spectral differences in the Raman peak position (e.g., 875, 936, 1004, 1078, 1265, 1302, 1335, 1445, 1618, 1652, and 1745  $\text{cm}^{-1}$ ) between different tissue pathologies, reconfirming our preceding *in vivo* Raman studies.<sup>25–27</sup> A detailed analysis and discussion of the Raman spectral characteristics of carcinogenesis in the gastric can be found elsewhere that includes major pathological features such as upregulation of mitotic and proteomic activity, increase in DNA contents and relative reduction in lipid as well as onset of angiogenesis leading to neovascularization in the tissue.<sup>26,34</sup>

The automatic outlier detection was realized for predictive on-line analysis using PCA with Hotelling's  $T^2$  and Q-residuals statistics (99% CI). To make the online diagnostics efficient, a two-component PCA model was rendered that included the largest tissue spectral variations. These selected significant PCs ( $p < 0.0001$ ) accounted for maximum variance of 38.71%



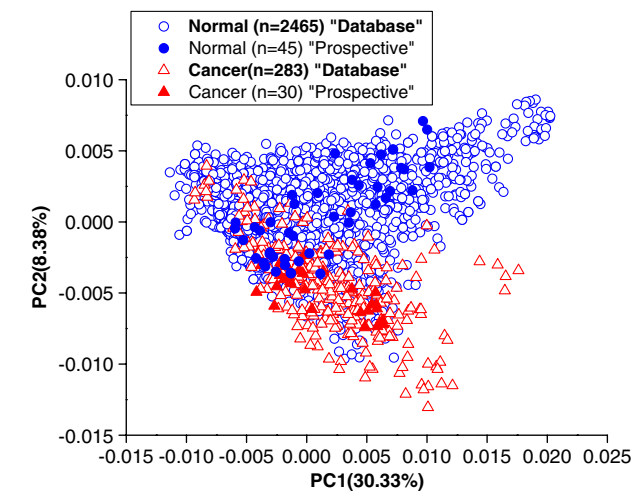
**Fig. 5** The principal component (PC) loadings calculated from the spectral training database of 2748 (2465 normal and 283 cancer) Raman spectra of gastric tissue *in vivo*. The significant PCs capture a total variance of 38.71% (PC1, 30.33%; PC2, 8.38%).

(PC1: 30.33%, PC2: 8.38%) of the total variability in the dataset ( $n = 2748$  Raman spectra), and the corresponding PC loadings are shown in Fig. 5.

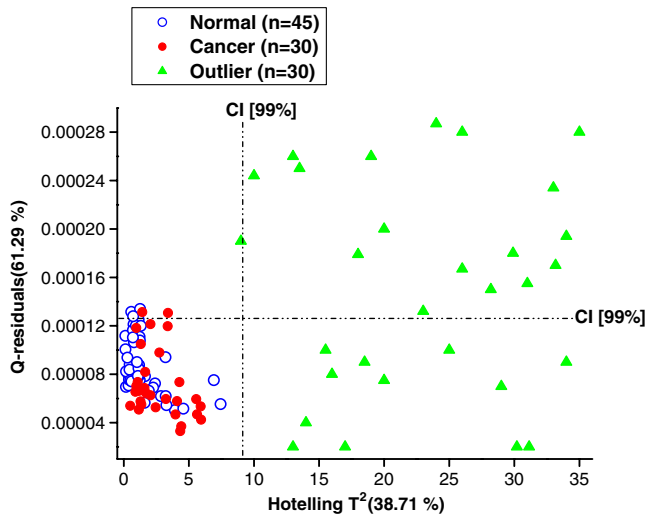
Figure 6 shows the score scatter plots (i.e., PC1 versus PC2) for the normal ( $n = 2465$ ) and cancer tissue spectra ( $n = 283$ ), exemplifying the capability of PC scores for separating the cancer spectra from normal. The 99% CI of Hotelling's  $T^2$  and Q residuals were accordingly calculated from the training dataset and fixed as a threshold for prospective on-line spectral validation. We then rendered probabilistic PLS-DA models for prediction of gastric cancer. The training database was randomly resampled multiple times ( $n = 10$ ) into learning (80%) and test (20%) sets. The generated PLS-DA models provided a predictive accuracy of 85.6% (95% CI: 82.9% to 88.2%) [sensitivity of 80.5% (95% CI: 71.4% to 89.6%) and specificity of 86.2% (95% CI: 83.6% to 88.7%)] for gastric cancer diagnosis. We subsequently tested the outlier-detection as well as probabilistic PLS-DA in 10 prospective gastric patients. PC score scatter plots (i.e., PC1 versus PC2) for the prospective normal ( $n = 45$ ) and cancer ( $n = 30$ ) tissue spectra are also shown in Fig. 6.

Figure 7 shows the prospective scatter plot of the Hotelling's  $T^2$  (38.71%) and Q-residuals (61.29%) with the 99% CI boundaries for 105 spectra acquired from the prospective gastric patients. It is observed that a large number of noncontact spectra lie outside the 99% CI and are therefore discarded in real-time. The verified tissue Raman spectra largely fall inside the 99% CI of  $T^2$  and Q residuals, demonstrating that this on-line data analysis provides a rapid and highly efficient means of real-time validation of biomedical tissue spectra. The prospectively acquired spectra verified by the on-line outlier analysis are further fed to probabilistic PLS-DA for instant disease prediction, achieving a diagnostic accuracy of 80.0% (60/75) [sensitivity of 90.0% (27/30), and specificity of 73.3% (33/45)] for gastric cancer detection (Fig. 8), as confirmed by histopathological examination.

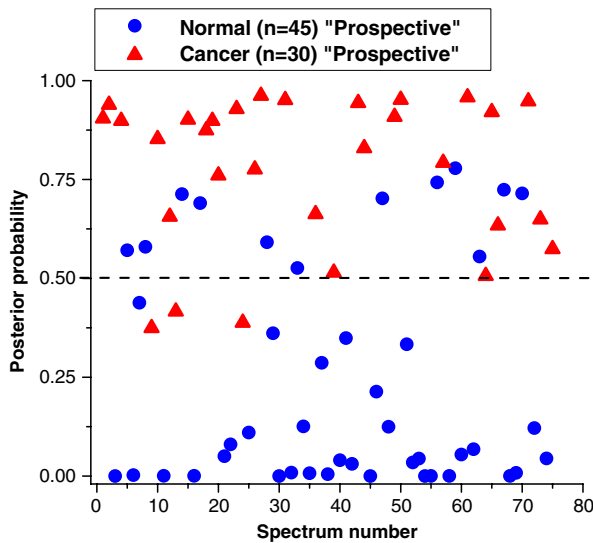
The receiver operating characteristic (ROC) curves were further generated to evaluate the group separations. Figure 9 shows the mean of the ROC curves computed from each random splitting of the spectral database for retrospective prediction as



**Fig. 6** Scatter plots of the two diagnostically significant PC scores (PC1 versus PC2) with a total variance of 38.71% (PC1, 30.33%; PC2, 8.38%) for 2465 normal (training database) and 283 cancer (training database) gastric tissue Raman spectra, together with the PC scores of prospective 45 normal and 30 gastric cancer Raman spectra.



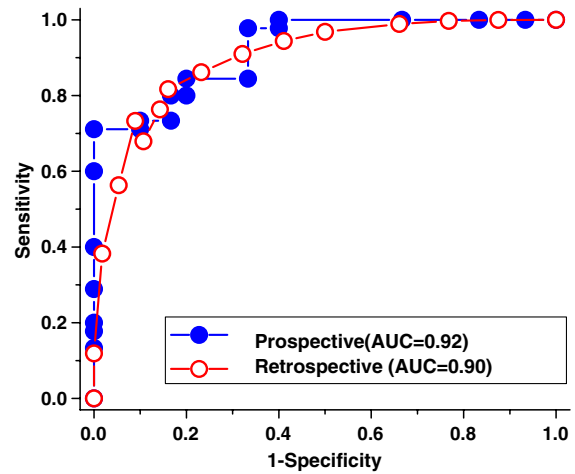
**Fig. 7** Hotelling's  $T^2$  versus Q-residuals for 105 Raman spectra (45 normal, 30 cancer, 30 outlier) acquired from 10 prospective gastric samples. The outlier ( $n = 30$ —green triangles) shows higher  $T^2$  and higher Q values than the valid spectra. The dotted line represents the 99% confidence interval (CI) verifying whether the prospective Raman spectra are within the common tissue variations of the principal component analysis (PCA) model.



**Fig. 8** Scatter plot of the posterior probability values belonging to prospective normal ( $n = 45$ ) and cancer ( $n = 30$ ) gastric tissue based on PLS-DA modeling together with leave-one spectrum-out, cross-validation. The separate dotted line gives a diagnostic sensitivity of 90.0% (27/30) and specificity of 73.3% (33/45) for separating cancer from normal gastric tissue *in vivo*.

well as the ROC calculated for the prospective dataset prediction. The integration areas under the ROC curves generated for the retrospective and prospective datasets are 0.90 and 0.92, respectively, illustrate the robustness of the PLS-DA algorithm for gastric cancer diagnosis *in vivo*.

The total processing time for all the procedures (Fig. 2) implemented was 0.13 s. The processing time for each step of the flow chart (Fig. 2) are given in Table 1. Free-running optical diagnosis and processing time of <0.5 s can be achieved, which is critical for realizing real-time *in vivo* tissue diagnostics at endoscopy.



**Fig. 9** The receiver operating characteristic (ROC) curves computed from the spectral database for retrospective prediction as well as ROC curve for prospective prediction of normal and cancer gastric tissue. The integrated areas under the ROC curves are 0.90 and 0.92 for the retrospective and prospective datasets, respectively, illustrating the efficacy of Raman endoscopy together with PLS-DA algorithms for *in vivo* gastric cancer diagnosis.

**Table 1** Average processing time for on-line biomedical Raman spectroscopic framework on a personal computer with a 64-bit I7 quad-core 4GB memory.

Analyses	Computational time (milliseconds)
Cosmic ray rejection	0.5
Laser response time	10
Preprocessing	40
Outlier detection	10
Probabilistic PLS-DA prediction	70
Total computation time	100 to 130

Overall, this work demonstrates for the first time that the on-line biomedical diagnostic framework can move the biomedical Raman spectroscopy into real-time, on-line cancer detection and diagnosis during routine clinical endoscopic examination.

### 4 Discussion

Histopathological examination of endoscopic biopsies for observing cytological and morphological abnormalities remains the current gold standard for precancer and cancer diagnosis in tissue. Several spectroscopic techniques including reflectance, fluorescence and Raman spectroscopy have demonstrated promising potential for realizing optical biopsies. Extensive research is currently in progress especially for translating Raman spectroscopic technique into real-time, clinical precancer and cancer diagnostics. However, specifically for prospective clinical endoscopic applications, fast data acquisition and short processing times (<0.5 s), straightforward on-line probabilistic diagnostics, real-time feedback to clinicians of spectrum quality and tissue classification are essential to making this optical biopsy technology a robust diagnostic tool in clinical settings. In this work, we have developed a novel on-line

spectral diagnostic framework that implements fully synchronized spectral acquisition, automatic integration time setting, model-specific outlier analysis as well as organ-specific probabilistic diagnostics, pushing spectroscopic techniques into real-time cancer screening with an easy-to-use interface for routine clinical endoscopic applications.

We have presented a universal on-line framework for biomedical spectroscopy (Fig. 3) that integrates preprocessing, automatic outlier detection, and deploys instant change of diagnostic models among different organs in the upper GI tract during gastroscopy. The universal platform supports different spectroscopic techniques (e.g., reflectance, fluorescence and Raman spectroscopy) with appropriate preprocessing methods and model development for each modality. We integrate the on-line diagnostic framework with the recently developed multimodal image-guided (WLR/NBI/AFI) Raman spectroscopic platform for early diagnosis and detection of precancer and cancer in the upper GI at endoscopy.<sup>5,28,29</sup> The accumulation of tissue Raman spectra and automatic scaling of integration time with a predefined upper limit of 0.5 s allows instant acquisition of *in vivo* tissue spectra with improved SNR while preventing CCD signal saturation. This is especially important for endoscopic diagnostics where the autofluorescence intensity varies significantly among different anatomical regions (e.g., antrum and body in the gastric, bronchi in the lung) likely caused by distinct endogenous fluorophores in the tissue.

We have also introduced automated outlier detection for spectra verification in endoscopic applications, whereby probe handling variations can be significant and to a large extent depend on the clinicians' expertise. PCA with related Hotelling's  $T^2$  and Q-residuals is a high-level metrics for outlier detection,<sup>41</sup> and we found that it is highly efficient and reliable in our online diagnostic platform at endoscopic diagnosis. The PCA model was generated from the database accounting for the majority of common tissue variations (38.71% of total variance, PC1: 30.33%; PC2: 8.38%) (Fig. 5). The rendered PCA model was further tested prospectively on 10 gastric patients for prospective automatic outlier detection during *in vivo* on-line Raman spectroscopic screening (Fig. 6). Of the total 105 gastric spectra acquired from the prospective gastric patients, 75 spectra that fit the developed PCA model well [ $T^2$  and Q lies within 99% CI (Fig. 7—blue and red circles)] were eligible for further classification. The remaining spectra ( $n = 30$ , Fig. 7; green triangles) from noncontact measurements, food debris or blood contaminations show unusual disturbances, which were poorly reconstructed by the PCA model. Large Q-residuals indicated the presence of spectral variations unexplainable by the model due to many confounding factors (e.g., noncontact mode, contaminants, or unknown spectral variations) frequently occurring in *in vivo* clinical measurements.<sup>36</sup> One noted that tissue Raman spectra presenting with Q-residuals vaguely outside the CI [ $(n = 6)$  in Fig. 7] were accepted following off-line analysis because the newly introduced spectral variations could be due to tumor invasion, necrosis, tumor subtype, etc. Thus following on-line analysis, it is advisable to review the spectra in off-line mode to assess possible new spectral variations. This study thoroughly shows the necessity of utilizing automated outlier diagnostics for tissue Raman spectral quality verification in real-time endoscopic applications.

Subsequent to on-line verification of spectrum quality, the spectra were fed to probabilistic PLS-DA algorithms for disease prediction. The PLS-DA modeling provided the predictive



**Fig. 10** Video showing the online Raman endoscopy technique for realizing real-time *in vivo* cancer diagnosis during clinical endoscopic examination (MOV, 10.9 MB) [URL: <http://dx.doi.org/10.1117/1.JBO.17.8.081418.1>].

accuracy of 80.0% (60/75) [sensitivity of 90.0% (27/30) and specificity of 73.3% (33/45)] (Fig. 8) for cancer diagnosis on 10 prospective gastric patients, suggesting that Raman endoscopy with the integration of on-line diagnostic framework can be a diagnostic screening tool for real-time *in vivo* gastric cancer identification. These probabilistic diagnostic outcomes were presented in real-time to the endoscopist within 0.5 s (Table 1), which is a clinically acceptable time at endoscopy. With the immediate auditory diagnostic feedback from the GUI developed, the endoscopists can now perform routine point-wise scanning for targeted biopsies of the high-risk tissue sites. The movie (Fig. 10) shows the online Raman endoscopic diagnostic procedure for real-time *in vivo* detection of gastric cancer during clinical endoscopic examination. Biomedical spectroscopic modalities can thus now function in free-running mode, opening the possibility of true prospective spectroscopic screening in clinical settings.

In summary, we have developed an on-line biomedical Raman spectroscopic framework to translate the cumbersome processing and multivariate analysis into an easy-to-use GUI interface for real-time, *in vivo* diagnosis of malignancies in internal organs. The efficacy of the on-line diagnostic framework integrated with the multimodal image-guided clinical Raman spectroscopy was proven on the prospective gastric patients, illustrating the promising potential of moving the biomedical Raman spectroscopy technique into real-time, on-line cancer detection and diagnosis during routine clinical endoscopic examination.

#### Acknowledgments

This research was supported by the National Medical Research Council and the Biomedical Research Council, Singapore.

#### References

1. A. Jemal et al., "Global cancer statistics," *CA: Cancer J. Clin.* **61**(2), 69–90 (2011).
2. Y. N. Mirabal et al., "Reflectance spectroscopy for *in vivo* detection of cervical precancer," *J. Biomed. Opt.* **7**(4), 587–594 (2002).
3. R. M. Cothren et al., "Gastrointestinal tissue diagnosis by laser-induced fluorescence spectroscopy at endoscopy," *Gastrointest. Endosc.* **36**(2), 105–111 (1990).



4. N. Ramanujam et al., "Fluorescence spectroscopy: a diagnostic tool for cervical intraepithelial neoplasia (CIN)," *Gynecol. Oncol.* **52**(1), 31–38 (1994).
5. S. K. Teh et al., "Near-infrared Raman spectroscopy for gastric pre-cancer diagnosis," *J. Raman Spectrosc.* **40**(8), 908–914 (2009).
6. M. S. Bergholt et al., "Raman endoscopy for *in vivo* differentiation between benign and malignant ulcers in the stomach," *Analyt* **135**(12), 3162–3168 (2010).
7. Z. Huang et al., "Rapid near-infrared Raman spectroscopy system for real-time *in vivo* skin measurements," *Opt. Lett.* **26**(22), 1782–1784 (2001).
8. C. A. Lieber et al., "Raman microspectroscopy for skin cancer detection *in vitro*," *J. Biomed. Opt.* **13**(2), 024013–024019 (2008).
9. M. S. Bergholt, W. Zheng, and Z. Huang, "Characterizing variability in *in vivo* Raman spectroscopic properties of different anatomical sites of normal tissue in the oral cavity," *J. Raman Spectrosc.* **43**(2), 255–262 (2012).
10. A. Deshmukh et al., "Raman spectroscopy of normal oral buccal mucosa tissues: study on intact and incised biopsies," *J. Biomed. Opt.* **16**(12), 127004 (2011).
11. K. Lin, D. L. P. Cheng, and Z. Huang, "Optical diagnosis of laryngeal cancer using high wavenumber Raman spectroscopy," *Biosens. Bioelectron.* **35**(1), 213–217 (2012).
12. N. Stone et al., "Raman spectroscopy for early detection of laryngeal malignancy: preliminary results," *Laryngoscope* **110**(10 Pt. 1), 1756–1763 (2000).
13. M. S. Bergholt et al., "In vivo, real-time, transnasal, image-guided Raman endoscopy: defining spectral properties in the nasopharynx and larynx," *J. Biomed. Opt.* **17**(7), 077002 (2012).
14. D. P. Lau et al., "Raman spectroscopy for optical diagnosis in normal and cancerous tissue of the nasopharynx-preliminary findings," *Lasers Surg. Med.* **32**(3), 210–214 (2003).
15. S. K. Teh et al., "Spectroscopic diagnosis of laryngeal carcinoma using near-infrared Raman spectroscopy and random recursive partitioning ensemble techniques," *Analyt* **134**(6), 1232–1239 (2009).
16. A. Haka et al., "Diagnosing breast cancer by using Raman spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.* **102**(35), 12371 (2005).
17. E. Widjaja, W. Zheng, and Z. Huang, "Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines," *Int. J. Oncol.* **32**(3), 653–662 (2008).
18. X. Shao, W. Zheng, and Z. Huang, "Near-infrared autofluorescence spectroscopy for *in vivo* identification of hyperplastic and adenomatous polyps in the colon," *Biosens. Bioelectron.* **30**(1), 118–122 (2011).
19. Z. Huang et al., "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer* **107**(6), 1047–1052 (2003).
20. S. Duraipandian et al., "Simultaneous fingerprint and high-wavenumber confocal raman spectroscopy enhances early detection of cervical precancer *in vivo*," *Anal. Chem.* **84**(14), 5913–5919 (2012).
21. L. E. Kamemoto et al., "Near-infrared micro-Raman spectroscopy for *in vitro* detection of cervical cancer," *Appl. Spectrosc.* **64**(3), 255–261 (2010).
22. P. Crow et al., "Assessment of fiberoptic near-infrared raman spectroscopy for diagnosis of bladder and prostate cancer," *Urology* **65**(6), 1126–1130 (2005).
23. P. Crow et al., "The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma *in vitro*," *Br. J. Cancer* **89**(1), 106–108 (2003).
24. M. S. Bergholt et al., "In vivo diagnosis of esophageal cancer using image-guided Raman endoscopy and biomolecular modeling," *Technol. Cancer Res. Treat.* **10**(2), 103–112 (2011).
25. M. S. Bergholt et al., "In vivo diagnosis of gastric cancer using raman endoscopy and ant colony optimization techniques," *Int. J. Cancer* **128**(11), 2673–2680 (2011).
26. M. S. Bergholt et al., "Combining near-infrared-excited autofluorescence and Raman spectroscopy improves *in vivo* diagnosis of gastric cancer," *Biosens. Bioelectron.* **26**(10), 4104–4110 (2011).
27. Z. Huang et al., "In vivo early diagnosis of gastric dysplasia using narrow-band image-guided Raman endoscopy," *J. Biomed. Opt.* **15**(3), 037017 (2010).
28. S. K. Teh et al., "Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue," *Br. J. Cancer* **98**(2), 457–465 (2008).
29. M. S. Bergholt et al., "Characterizing variability in *in vivo* Raman spectra of different anatomical locations in the upper gastrointestinal tract toward cancer detection," *J. Biomed. Opt.* **16**(3), 037003 (2011).
30. J. T. Motz et al., "Real-time Raman system for *in vivo* disease diagnosis," *J. Biomed. Opt.* **10**(3), 031113 (2005).
31. H. Lui et al., "Real-time Raman spectroscopy for *in vivo* skin cancer diagnosis," *Cancer Res.* **72**(10), 2491–2500 (2012).
32. Z. Huang et al., "Integrated Raman spectroscopy and trimodal wide-field imaging techniques for real-time *in vivo* tissue Raman measurements at endoscopy," *Opt. Lett.* **34**(6), 758–760 (2009).
33. M. S. Bergholt et al., "In vivo diagnosis of esophageal cancer using image-guided Raman endoscopy and biomolecular modeling," *Technol. Cancer Res. Treat.* **10**(2), 103–112 (2011).
34. Z. Huang et al., "In vivo detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy," *Biosens. Bioelectron.* **26**(2), 383–389 (2010).
35. K. G. Yeoh, "How do we improve outcomes for gastric cancer?," *J. Gastroenterol. Hepatol.* **22**(7), 970–972 (2007).
36. R. P. Cogdill, C. A. Anderson, and J. K. Drennen, 3rd, "Process analytical technology case study, part III: calibration monitoring and transfer," *AAPS PharmSciTech* **6**(2), E284–E297 (2005).
37. O. Ortiz-Estarellas et al., "Multivariate data analysis of quality parameters in drinking water," *Analyt* **126**(1), 91–96 (2001).
38. E. B. Martin, A. J. Morris, and J. Zhang, "Process performance monitoring using multivariate statistical process control," *Control Theory and Applications, IEE Proc.* **143**(2), 132–144 (1996).
39. S. K. Teh et al., "Diagnosis of gastric cancer using near-infrared Raman spectroscopy and classification and regression tree techniques," *J. Biomed. Opt.* **13**(3), 034013 (2008).
40. T. Brendan Murphy, N. Dean, and A. E. Raftery, "Variable selection and updating in model-based discriminant analysis for high dimensional data with food authenticity applications," *Ann. Appl. Stat.* **4**(1), 396–421 (2010).
41. C. E. Miller, "Chemometrics in Process Analytical Technology (PAT)," in *Process Analytical Technology* K. A. Bakeev, Ed., pp. 353–438, John Wiley & Sons Ltd., New York (2010).