

Research Article

Identification of Colitis and Cancer in Colon Biopsies by Fourier Transform Infrared Spectroscopy and Chemometrics

Xiang Li,¹ Qing-Bo Li,¹ Guang-Jun Zhang,¹ Yi-Zhuang Xu,² Xue-Jun Sun,³ Jing-Sen Shi,³ Yuan-Fu Zhang,² and Jin-Guang Wu²

¹ College of Instrument Science and Opto-Electronics Engineering, Beihang University, Beijing 100083, China

² College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

³ Department of General Surgery, First Hospital of Xi'an Jiaotong University, Shanxi, Xi'an 710061, China

Correspondence should be addressed to Guang-Jun Zhang, gjzhangbuaa@163.com

Received 31 October 2011; Accepted 25 December 2011

Academic Editor: Keishiro Aoyagi

Copyright © 2012 Xiang Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cancer is a disease that does great harms to the health of human beings. FT-IR spectroscopy could identify variability at the molecular level in biological specimens. It is a rapid and noninvasive method, which could be used intraoperatively to modify surgical procedures. The aim of this paper is to identify and separate cancer from colitis in endoscopic colon biopsies through the use of FT-IR spectroscopy. A total of 88 endoscopic colon samples, including 41 cases of colitis and 47 cases of colon cancer, were obtained. Specimens were placed on an ATR accessory linked to FT-IR spectrometer with a MCT detector for greater stability and sensitivity. Later, specimens were sent for the histological examination as the reference in the spectral analysis. 41 colitis and 47 cancer specimens were compared. Spectra preprocessed with smoothing and normalization were used for discrimination analysis. PCA was processed to simplify the spectrum data set. Naive Bayes classifier model was constructed for diagnostic classification. Leave-one-out cross-validation method was utilized to assess the discrimination results. The sensitivity of FT-IR detection for cancer achieves 97.6%. The results showed that colon cancer could be distinguished from colitis with high accuracy using FT-IR spectroscopy and chemometrics.

1. Introduction

Cancer is a disease that does great harms to the health of human beings. The survival of patients depends largely on the detection of cancer at an early stage. It is of great importance to explore the early cancer diagnosis method. But when the changes in morphology can be seen under light microscope, there have been millions of cancer cells. In the process of carcinogenesis, nuclear acids, proteins, carbohydrates, and other biomolecules generate significant changes in their molecular structures. Fourier transform infrared (FT-IR) spectroscopy is a powerful tool to detect the changes of molecular structure and composition [1–3]. Therefore, it is possible for the FT-IR spectral analysis technology to become a rapid, noninvasive, and convenient method to detect tumors at the precarcinogenesis stage [4, 5]. At present, with the development of biospectroscopy and

spectral analysis technology, the application of FT-IR spectroscopy in distinguishing malignant tissues from normal ones has become a focus [6–10]. Also, great progresses have been made in the research of cancer detection using FT-IR spectroscopy [11–17].

FT-IR spectroscopy can effectively provide chemical variation information about the structure and the composition of biological materials at molecular level. FT-IR technology makes it possible to detect inflammatory and cancer of the enteroscopic biopsies. It indicated that FT-IR method has the opportunity to develop as a new technique for enteroscope examination. We believe that noninvasive, rapid, accurate, and convenient analysis of colon tissues can be performed with Fourier-transform midinfrared spectroscopy if the mid-infrared fiber optics and colon endoscopy technologies can be combined successfully. The fundamental study on the application of chemometrics to the identification of colon

biopsies, obtained from enteroscopy detection and measured in vitro using FT-IR spectrometer, was performed in this paper.

2. Theory

2.1. Principal Component Analysis. One of the difficulties in spectral analysis is that spectral data usually has too many variables. Fortunately, in spectrum data sets, groups of variables often move together. The absorption bands in neighborhood are related to each other [18]. Thus, here is plenty of redundancy of information in spectrum data set.

Principal component analysis (PCA) is a quantitatively mathematical procedure for achieving simplification. The method generates a new set of variables, called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data. This linear transformation has been widely used in data analysis and compression.

The first principal component is a single axis in space. When each observation is projected on that axis, the resulting values form a new variable. And the variance of this variable is the maximum among all possible choices of the first axis.

The second principal component is another axis in space, perpendicular to the first. Projecting the observations on this axis generates another new variable. The variance of this variable is the maximum among all possible choices of this second axis. And the rest principal components are also settled as above [19].

The full set of principal components is as large as the original set of variables. But it is commonplace for the sum of the variances of the first few principal components to explain most information of the original data.

2.2. Naïve Bayes classifier. A Naïve Bayes Classifier (NBC) is a simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions [20]. In simple terms, a Naïve Bayes classifier assumes that the presence of a particular feature of a class is unrelated to the presence of any other feature, given the class variable. It classifies data in two steps.

First, using the training samples, the method estimates the parameters of a probability distribution, assuming that features are conditionally independent given the class. Naïve Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for Naïve Bayes models uses the method of maximum likelihood [21, 22].

Then, for any unseen test sample, the method computes the posterior probability of that sample belonging to each class. The method then classifies the test sample according to the largest posterior probability.

An advantage of the Naïve Bayes classifier is that it only requires a small amount of training data to estimate the

parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

3. Experimental

3.1. Patients. Informed consent was obtained from each patient prior to the study. A total of 89 colon biopsies were obtained from clinical detection in the Medical Division of the First Hospital of Xi'an Jiaotong University, China. There were 43 women and 45 men, aged between 21 and 76 years (mean 53.7 years old). One endoscopic pinch biopsy of 1–3 mm in diameter was obtained from each patient. According to the results of pathological detection, the studied samples consisted of 41 cases of colitis and 47 cases of cancer.

3.2. Instrument. An attenuated total reflectance (ATR) accessory linked to a WQD-500 FT-IR spectrometer (Beijing No. 2 Optical Instrument Factory, Beijing, China) was used. The ATR accessory was made of ZnSe crystal. The FT-IR spectrometer is equipped with a liquid nitrogen-cooled mercury cadmium telluride detector.

3.3. Spectral Measurements. Specimens were frozen and transported to the laboratory. The sample was thawed at room temperature for about 3–5 minutes, and placed on the ATR accessory to record the spectrum using WQD-500 FT-IR spectrometer. The background spectrum was collected at first. 32-coadded scans at a resolution of 4 cm^{-1} , with a normal range from 1000 to 4000 cm^{-1} , were performed. It took about 1–2 minutes to measure the spectrum noninvasively. After the spectra of the samples were recorded, the samples were stored in liquid nitrogen and sent for the histological examination as the reference in the spectral analysis.

3.4. Computer Programs. The Matlab programs used for smoothing, normalization, PCA, and NBC discrimination analysis were developed in our laboratory. These programs can be compiled on any digital computer and are available from the corresponding author.

4. Results and Discussion

In the process of measuring the spectra, the spectra contain not only the useful information of molecular structure and component of the measured samples, but also the noises such as the high-frequency random noise, baseline drift, light scattering, and so forth. These additional noises must be eliminated; otherwise, they will affect the result of the discrimination.

Firstly, smoothing is utilized to filter high-frequency noise. Figure 1 showed the mean spectra of colitis and colon cancer biopsies (1000 – 4000 cm^{-1}). The absorption band of CO_2 and the spectral range of 3696 – 4000 cm^{-1} were excluded, which contained little useful information for measurement. The spectral data of these two ranges were mean normalized, respectively. Preprocessed spectra with

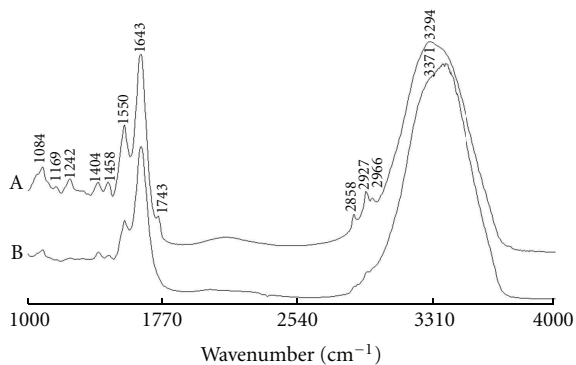


FIGURE 1: Mean FT-IR spectra of colon biopsies. Trace (A), colitis; Trace (B), colon cancer.

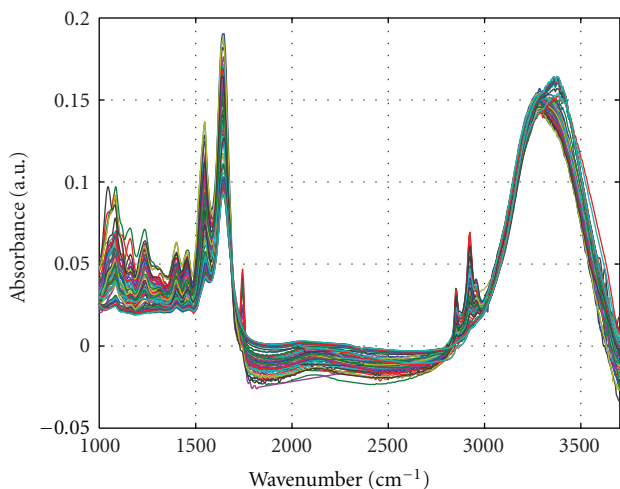


FIGURE 2: Preprocessed spectra with smoothing and normalization (1000–1766 cm^{-1} , 2731–3695 cm^{-1}).

smoothing and normalization are shown in Figure 2. And, the PCA was used to simplify the spectrum data set. Only the first 5 principal components were reserved, which expressed over 90% information of original data set. And the rest data was eliminated. Apparently, the PCA could reduce a lot of computation of NBC. Then the NBC model was built for classification of cancer and colitis samples. Cross-validation was utilized to evaluate the discrimination results. The FT-IR analysis results (Table 1) were as follows: among the 47 cases of colitis samples, 33 cases were correctly distinguished while 14 samples were misjudged; among the 41 cases of cancer samples, 40 cases were well judged while only 1 case was misjudged. Total correctness was 82.8%. Statistical results of colon biopsies using FT-IR spectroscopy (Table 2) showed that sensitivity of cancer diagnosis was 97.6%, specificity of cancer diagnosis for 70.2%, predictive value of a positive test of cancer diagnosis for 74.1%, and predictive value of a negative test of cancer diagnosis for 97.1%.

There are different spectral characteristics between colitis and malignant colon endoscope samples in the FT-IR spectra. These spectral features are related to the changes of structure and composition of biological molecular in tissue

TABLE 1: Comparison of FT-IR results with histological examination.

Histologic examination	FT-IR results	
	Cancer	Colitis
Cancer	40	1
Colitis	14	33

TABLE 2: Results of statistical analysis of detection of colon biopsies by FT-IR spectroscopy.

	Sensitivity, %	Specificity, %	Predictive value of a positive test, %	Predictive value of a negative test, %
Cancer	97.6	70.2	74.1	97.1

cell. The mean spectra of colitis and cancer got from endoscope detection are illustrated in Figure 1. The spectral features of these two types of colon biopsies were as follows: C=O band near 1743 cm^{-1} was assigned to the fat in tissues, and C-H-stretching vibration bands near 2966 cm^{-1} , 2927 cm^{-1} , and 2858 cm^{-1} were related to lipid and fat content, and these bands usually decreased and even disappeared in the spectra of malignant tissues, for that the fat in the malignant tissue is consumed because of the necessary increased nutritional and energy requirement in the development of the carcinoma. $\sim 1643 \text{ cm}^{-1}$ absorption peak belonged to amide I band of protein and H–O–H deformation vibration of water. $\sim 1550 \text{ cm}^{-1}$ absorption peak was assigned to amide II band of protein. The relative intensity of amide II band to $\sim 1643 \text{ cm}^{-1}$ absorption peak decreased in the spectra of malignant colon tissues than those in colitis biopsies. The intensity of $\sim 1460 \text{ cm}^{-1}$ peak was weaker than that of $\sim 1400 \text{ cm}^{-1}$ peak in the spectra of the cancerous samples. The peak at $\sim 1460 \text{ cm}^{-1}$ was stronger than or equal to that of $\sim 1400 \text{ cm}^{-1}$ in the spectra of colitis samples.

5. Conclusion

In conclusion, the results indicate that colon cancer can be distinguished from colitis with high accuracy using FT-IR spectroscopy and chemometrics. This should be extended to analysis of cancer versus normal colon or polyps. This technique has promise for in vivo diagnosis with the development of endoscopic FT-IR miniprbes and is practical for immediate diagnosis at endoscopy.

Acknowledgments

The authors gratefully acknowledge Beijing no. 2 Optical Instrument Factory (Beijing, China) for providing the FT-IR spectrometer and for excellent technical assistance.

References

[1] J. D. Kruse-Jarres, G. Janatsch, U. Gless, R. Marbach, and H. M. Heise, "Glucose and other constituents of blood

- determined by ATR-FTIR-spectroscopy," *Clinical Chemistry*, vol. 36, no. 2, pp. 401–402, 1990.
- [2] H. P. Wang, H. C. Wang, and Y. J. Huang, "Microscopic FTIR studies of lung cancer cells in pleural fluid," *Science of the Total Environment*, vol. 204, no. 3, pp. 283–287, 1997.
 - [3] P. Franck, J. L. Sallerin, H. Schroeder, M. A. Gelot, and P. Nabet, "Rapid determination of fecal fat by Fourier transform infrared analysis (FTIR) with partial least-squares regression and an attenuated total reflectance accessory," *Clinical Chemistry*, vol. 42, no. 12, pp. 2015–2020, 1996.
 - [4] Q.-B. Li, X.-J. Sun, Y.-Z. Xu et al., "Diagnosis of gastric inflammation and malignancy in endoscopic biopsies based on fourier transform infrared spectroscopy," *Clinical Chemistry*, vol. 51, no. 2, pp. 346–350, 2005.
 - [5] L. Zhang, K. H. Sun, R. D. Soloway, X. F. Ling, Y. Z. Xu, and Q. G. Wu, "Intraoperative Fourier transform infrared spectroscopy can guide individual resections in patients with gastric cancer," *Gastroenterology*, vol. 126, article A626, 2004.
 - [6] B. Rigas, S. Morgello, I. S. Goldman, and P. T. T. Wong, "Human colorectal cancers display abnormal Fourier-transform infrared spectra," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 20, pp. 8140–8144, 1990.
 - [7] R. Sindhuphak, S. Issaravanich, V. Udomprasertgul et al., "A new approach for the detection of cervical cancer in Thai women," *Gynecologic Oncology*, vol. 90, no. 1, pp. 10–14, 2003.
 - [8] H. Huo, X. Hu, H.-W. Guan, X.-F. Wang, X. Che, and W.-D. Huang, "Studies on human breast cancer tissue with infrared spectrometry," *Chemical Journal of Chinese Universities*, vol. 21, no. 8, pp. 1244–1247, 2000.
 - [9] N. Fujioka, Y. Morimoto, T. Arai, and M. Kikuchi, "Discrimination between normal and malignant human gastric tissues by Fourier transform infrared spectroscopy," *Cancer Detection and Prevention*, vol. 28, no. 1, pp. 32–36, 2004.
 - [10] P. R. Griffiths, H. S. Yang, Q. B. Li et al., "Discrimination of normal and malignant gastric tissues with FTIR spectroscopy and principal component analysis," *Spectroscopy and Spectral Analysis*, vol. 24, no. 9, p. 1025, 2004.
 - [11] Q. B. Li, L. M. Yang, X. F. Ling et al., "Application of the SIMCA method to cancer diagnosis with Fourier-Transform Infrared spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 24, no. 4, pp. 414–417, 2004.
 - [12] D.-W. Zhao, X.-S. Zhou, R. D. Soloway et al., "Infrared spectroscopic features of colon cancer cells differ from those of normal colonic cells from the same patient," *Gastroenterology*, vol. 110, no. 4, article A620.
 - [13] J. S. Wang, J. S. Shi, Y. Z. Xu et al., "FT-IR spectroscopic analysis of normal and cancerous tissues of esophagus," *World Journal of Gastroenterology*, vol. 9, no. 9, pp. 1897–1899, 2003.
 - [14] C. W. Sun, Y. Z. Xu, K. H. Sun, Q. G. Wu, W. H. Li, Z. H. Xu et al., "A study of the diagnosis of salivary gland tumors by means of mid infrared optical fiber technique," *Spectroscopy and Spectral Analysis*, vol. 16, pp. 22–25, 1996.
 - [15] Q. B. Li, Z. Xu, Y. Z. Xu et al., "Application of Fourier transform infrared spectroscopy to non-invasive detection of breast tumor in vivo and in situ," *Chemical Journal Of Chinese Universities*, vol. 25, no. 11, pp. 2010–2012, 2004.
 - [16] J. G. Wu, Y. Z. Xu, C. W. Sun et al., "Distinguishing malignant from normal oral tissues using FTIR fiber-optic techniques," *Biopolymers*, vol. 62, no. 4, pp. 185–192, 2001.
 - [17] Q. B. Li, Z. Xu, N. W. Zhang et al., "In vivo and in situ detection of colorectal cancer using Fourier transform infrared spectroscopy," *World Journal of Gastroenterology*, vol. 11, no. 3, pp. 327–330, 2005.
 - [18] A. A. Miranda, Y. A. Le Borgne, and G. Bontempi, "New routes from minimal approximation error to principal components," *Neural Processing Letters*, vol. 27, no. 3, pp. 197–207, 2008.
 - [19] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proceedings of the 20th International Conference on Machine Learning (ICML '04)*, pp. 225–232, July 2004.
 - [20] D. J. Hand and K. Yu, "Idiot's Bayes—not so stupid after all?" *International Statistical Review*, vol. 69, no. 3, pp. 385–398, 2001.
 - [21] G. I. Webb, J. R. Boughton, and Z. Wang, "Not so naive Bayes: aggregating one-dependence estimators," *Machine Learning*, vol. 58, no. 1, pp. 5–24, 2005.
 - [22] M. Možina, J. Demšar, M. Kattan, and B. Zupan, "Nomograms for visualization of naive Bayesian classifier," in *Proceedings of the European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD '04)*, vol. 3202, pp. 337–348, 2004.