

Research Article

Cancer Discrimination Using Fourier Transform Near-Infrared Spectroscopy with Chemometric Models

Hui Chen,^{1,2} Zan Lin,^{1,3} and Chao Tan¹

¹Key Lab of Process Analysis and Control of Sichuan Universities, Yibin University, Yibin, Sichuan 644000, China

²Yibin University Hospital, Yibin, Sichuan 644000, China

³The First Affiliated Hospital, Chongqing Medical University, Chongqing 400016, China

Correspondence should be addressed to Chao Tan; chaotan1112@163.com

Received 9 March 2015; Revised 25 May 2015; Accepted 27 May 2015

Academic Editor: Iciar Astiasaran

Copyright © 2015 Hui Chen 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.

Near-infrared (NIR) spectroscopy technique offers many potential advantages as tool for biomedical analysis since it enables the subtle biochemical signatures related to pathology to be detected and extracted. In conjunction with advanced chemometrics, NIR spectroscopy opens the possibility of their use in cancer diagnosis. The study focuses on the application of near-infrared (NIR) spectroscopy and classification models for discriminating colorectal cancer. A total of 107 surgical specimens and a corresponding NIR diffuse reflection spectral dataset were prepared. Three preprocessing methods were attempted and least-squares support vector machine (LS-SVM) was used to build a classification model. The hybrid preprocessing of first derivative and principal component analysis (PCA) resulted in the best LS-SVM model with the sensitivity and specificity of 0.96 and 0.96 for the training and 0.94 and 0.96 for test sets, respectively. The similarity performance on both subsets indicated that overfitting did not occur, assuring the robustness and reliability of the developed LS-SVM model. The area of receiver operating characteristic (ROC) curve was 0.99, demonstrating once again the high prediction power of the model. The result confirms the applicability of the combination of NIR spectroscopy, LS-SVM, PCA, and first derivative preprocessing for cancer diagnosis.

1. Introduction

Cancer is a type of gene diseases in which cells grow and divide uncontrollably [1]. In the last years, cancer has become one of the leading causes of death [2]. It is estimated that about 5.2 million patients die from cancer per year in the world and more than half million patients die in developing countries [3]. One of focal tasks of cancer-related researches is to develop inexpensive, accurate, fast, and convenient diagnostic methods to improve the survival probability. As far as cancer diagnosis is regarded, histopathological evaluation is still the gold standard [4]. It involves the sampling of tissues and examination by pathologists. Such a procedure includes tissue staining and morphological pattern recognition. In the process of tissue transformation, substantial modifications at molecular level are expected to occur before apparent visible morphological changes. Moreover, histopathological evaluation mainly depends on expert's subjective judgment and experience and its result is therefore subjective to some

extent [5]. There is an urgent need to develop novel and objective methods for cancer diagnosis. In recent years, great emphasis has been placed on seeking diagnostic methods that may enhance the ability to differentiate between normal and cancerous tissues.

Various molecular spectrometers as powerful tools for investigating chemical changes at molecular level have been applied to detect cancer [6–8]. For example, infrared spectroscopy (IR) techniques are widely applied to analyze biological tissues; the resulting spectra are composed of characteristic bands originating from all vibration modes of bimolecular components present in the tissue including proteins, lipids, and nucleic acids. Each biomolecule gives a characteristic IR spectrum, which contains infrared information of various functional groups in the molecule. The frequency and intensity of spectral bands are related to the vibration frequencies and polar properties as well as relative concentration. A total spectrum of tissue can convey information of characteristic changes in molecular composition and structure

accompanying transformation from normal to cancerous state.

In recent years, near-infrared (NIR) spectroscopy has caught the attention for biomedical study of several diseases including cancers [9–12]. In particular, the advance of fiber optic probes provides the possibility of applying NIR technique to noninvasive and minimally invasive ways of diagnosing disease. NIR spectroscopy can be partitioned into intervals of the short-wave and long-wave NIR [13]. The short-wave NIR region mainly provides information concerning tissue blood flow and oxygen saturation and consumption since the heme proteins and cytochromes dominate the spectrum. The long-wave NIR region corresponds to the combination and overtone vibrations of C-H, N-H, and O-H groups and thus provides information on the chemical composition of tissues. The NIR spectrum of a tissue actually reflects the structures and concentrations of various components. Cancerous tissues differ from normal parts in composition, physiology, and biochemistry. Any alteration in the composition of the tissues can thus be detected and used for diagnostic purposes. Some researchers have applied NIR spectroscopy to the studies on gastric [14], breast [15], prostate [16], pancreas [17], and colorectal [18] tissues. Unlike IR spectrum, the NIR spectrum corresponds to the overtones and combination of various molecules; a tissue sample often leads to a complex NIR spectrum, which consists of many overlapping, broad, weak, and nonspecific bands. Thus, when applying NIR to cancer diagnosis, it is inevitable to resort to appropriate algorithms to extract information and to build a discrimination model. Additionally, for a NIR-based application, the performance of a model, such as the accuracy and robustness, is decisive for its availability. No single algorithm performs best for all cases. Various algorithms including *K*-Nearest Neighbor Classifier [19], linear discriminate analysis (LDA) [20], soft independent modeling class analogy (SIMCA) [21], cluster analysis (CA), and support vector machine (SVM) [22] have been used for these purposes. In particular, SVM algorithm has become a powerful tool for nonlinear classification, function estimation, or nonlinear regression and has led to many successful applications. Compared to other methods, the main advantage of SVM is that there are very few parameters to tune or select a priori. Least-squares SVM (LS-SVM) algorithm is a simplified and improved version of traditional SVM [23]. It encompasses similar advantages as SVM and its additional advantage is that it only needs to solve a set of linear equations, which is much easier and computationally simpler than nonlinear equations employed by traditional SVM.

Based on the fact that NIR spectroscopy is a powerful tool to detect changes at the molecular level, the present work focuses on investigating the potential of NIR spectroscopy and classification models for discriminating colorectal cancer from normal ones so as to build an objective procedure for cancer diagnosis. Three spectral preprocessing methods and their combination were attempted and LS-SVM was used to construct classification models. The sensitivity and specificity were used as performance index. The receiver operating

characteristic (ROC) curve was also used for evaluation purposes.

2. Theory

2.1. Support Vector Machine (SVM). Support vector machine (SVM) is a state-of-the-art classification algorithm [24], which has a good theoretical foundation in statistical learning frame. The SVM optimizes the classification decision function based on structural risk minimization instead of traditional empirical risk minimization and can therefore avoid the problem of overfitting. Due to its advantages, SVM has gained extensive application in both classification and regression tasks. For a binary classification problem, SVM focuses on finding maximal margin hyperplanes in terms of a subset of the input data, that is, support vectors, between different classes. When the input data are not linearly separable, it firstly maps them into a high-dimensional feature space and then classifies them by the maximal margin hyperplanes. Least-squares SVM (LS-SVM) is a modified version of SVM. When realizing the algorithm, LS-SVM can result in a set of linear equations instead of a quadratic programming problem and significantly broaden the applications of SVM. A number of excellent introductions to both SVM and LS-SVM can be found in literatures. Here, only a brief description of the main idea of LS-SVM was provided.

2.2. Least-Squares Support Vector Machine (LS-SVM). Given a binary classification problem, samples $\{(\mathbf{x}_i, y_i)\}$, $i = 1, 2, \dots, L$, where \mathbf{x}_i is the vector of input pattern for the i th sample and y_i is the corresponding class label, the pattern represented by the subset $y_i = +1$ belongs to class 1 and the pattern represented by the subset $y_i = -1$ belongs to class 2. A final LS-SVM classifier corresponds to the optimal separating hyperplane, which can be found by solving the following minimization problem [23, 25]:

$$\text{Min}_{\omega, b, e} J(\omega, b, e) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{i=1}^l e_i^2 \quad (1)$$

subject to the equality constraints for the training set

$$y_i [\omega^T \varphi(x_i) + b] = 1 - e_i, \quad i = 1, 2, \dots, l, \quad (2)$$

where $\varphi: R^n \rightarrow R^m$ is a special map mapping the input space to a high-dimensional feature space where samples become linearly separable by a hyperplane defined by the pair $(\omega \in R^m, b \in R)$, γ is the relative weight of the error term, which can determine the tradeoff between minimizing the training error and minimizing model complexity, and e_i is the error variables taking noise into account and avoiding poor generalization. Obviously, LS-SVM considers equality type constraints instead of inequalities as in the classic SVM approach. The modification greatly simplified a problem such that the LS-SVM solution follows directly from solving a set of linear equations rather than from a convex quadratic program.

It can be noted that the constraints of (1) are equality constraints, but it is an inequality in SVM. The corresponding Lagrangian for (1) is

$$L(\omega, b, e, \alpha) = J(\omega, b, e) - \sum_{i=1}^l \alpha_i \{y_i [\omega^T \varphi(x_i) + b] - 1 + e_i\}, \quad (3)$$

where α_i are the Lagrange multipliers. According to [23], the optimality condition leads to the following $(l+1) \times (l+1)$ linear system:

$$\begin{bmatrix} 0 & -Y^T \\ Y & ZZ^T + \gamma^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix}, \quad (4)$$

where $Z = [\varphi(x_1)^T y_1, \varphi(x_2)^T y_2, \dots, \varphi(x_l)^T y_l]$, $Y = [y_1, y_2, \dots, y_l]$, $\vec{1} = \{1, 1, \dots, 1\}$.

Mercer's condition is applied within the matrix ZZ^T :

$$ZZ^T = y_i y_j \varphi(x_i)^T \varphi(x_j) = y_i y_j K(x_i, x_j). \quad (5)$$

Thus, only kernel function K is needed in the training algorithm. It would never need to explicitly even know what φ is. The LS-SVM classifier can be constructed as follows:

$$f(x) = \text{sign} \left(\sum_{i=1}^l y_i \alpha_i K(x, x_i) + b \right). \quad (6)$$

3. Experimental

3.1. Sample Preparation. All original tissues from patients who suffered from malignancies of the colon were supplied by the Affiliated Hospital of North Sichuan Medical College and the First People's Hospital of Yibin City. All the colorectal tissues were stabilized in 4% formaldehyde solution and then dehydrated by a series of ethanol solutions with a concentration gradient. The samples were put into xylene, embedded in paraffin wax, and then sliced in sections with the thickness of 4 μm . The tissue samples from patients with colorectal cancer were carefully selected to include both normal and malignant areas. By this means, a total of 107 tissue samples of patients, thirty-nine from females and sixty-eight from males, were selected. Considering that it was difficult to collect the spectra of too small samples, these samples were ignored; as a result, only 107 cancer tissue specimens and 55 normal specimens were used for experiment. The mean age of the patients in this study was 63.4 years. The oldest patient was 89 years old and the youngest one was 25 years old. Diagnosis was confirmed by a pathologist by histopathological examination.

3.2. Instrument and Measurement. The NIR spectroscopy studies were performed by a Fourier transform near-infrared (FT-NIR) spectrometer (Thermo Fisher, USA) coupled with an InGaAs detector. The paraffin sections were placed on the integrating sphere of the NIR spectrometer. Assisted by a special mirror, the NIR diffuse reflection spectra were collected in the range from 10,000 to 4,000 cm^{-1} . Each spectrum

is the co-adding of 64 scans and the resolution was set as 2 cm^{-1} . As a result, each spectrum consisted of 3112 variables. The data format was set as Log 1/R. The collection procedure was controlled by Result software of Thermo Fisher. Based on air reference, the background spectrum was recorded every 1 h. To simplify the calculation, the number of variables was reduced to 1137 by taking the average of every three adjacent points.

3.3. Dataset Partitioning. Each tissue slide corresponds to a spectrum and was assigned a label (1 for cancer and 2 for normal). A spectrum and a label constitute a so-called sample. In this study, a total of 107 and 55 samples belong to cancer category and normal category, respectively. In order to construct and validate the discrimination models, the dataset was divided into two independent subsets, that is, the training and test sets, by Kennard-Stone (KS) algorithm [26] coupled with alternate sampling. First, all cancer spectra/samples were sorted by KS algorithm, which first picks out two spectra that are most distant from each other as starting points. Subsequently, at each step, it selects the spectra that exhibit the largest minimum distance to any sample already selected. By this means, a sequence can be generated. Then, the sequence is alternately sampled. As a result, the training set and the test set have the same number of samples. Similarly, all normal spectra can be assigned to the training and test sets. All calculation and modeling were performed in MATLAB 7.0 for Windows by using LS-SVMlab1.5 toolbox.

4. Results and Discussions

4.1. Analysis of Spectral Profile. For diagnosing applications prior to pattern classification data preprocessing is essential for enhanced performance of the classification model. There are a number of mathematical techniques for preprocessing spectral data. However, due to the lack of general guidelines, to determine the appropriate preprocessing technique is still trial and error. In this study, SNV, first derivative, and PCA were considered. Figure 1 shows the first derivative-based and SNV-based preprocessed spectra and mean spectra of cancerous and normal samples. The first derivative preprocessing can lead to a derivative spectrum, which is the differentiation of the original spectrum. The main advantage of the derivative spectrum is to highlight and resolve overlapped spectral bands. SNV, a mathematical transformation of the original spectra, is capable of removing the slope variation and correcting scatter effects. It is noticeable that visual spectral difference between cancerous and normal samples was very small. The SNV-based preprocessed spectra can remove unwanted background variances to some extent. However, as shown in Figure 1(d), the subtle difference of mean spectra is maybe not enough to construct a satisfactory classification model. In Figure 1(c), the difference between cancerous and normal samples is more easily perceived in the region of 5500–7000 cm^{-1} ; the differences include peak shape and intensity. These peaks in such a region can be ascribed to the first overtone of C-H stretching (5500–6000 cm^{-1}), first overtone of O-H, N-H bonds, and C-H combinations. Such results

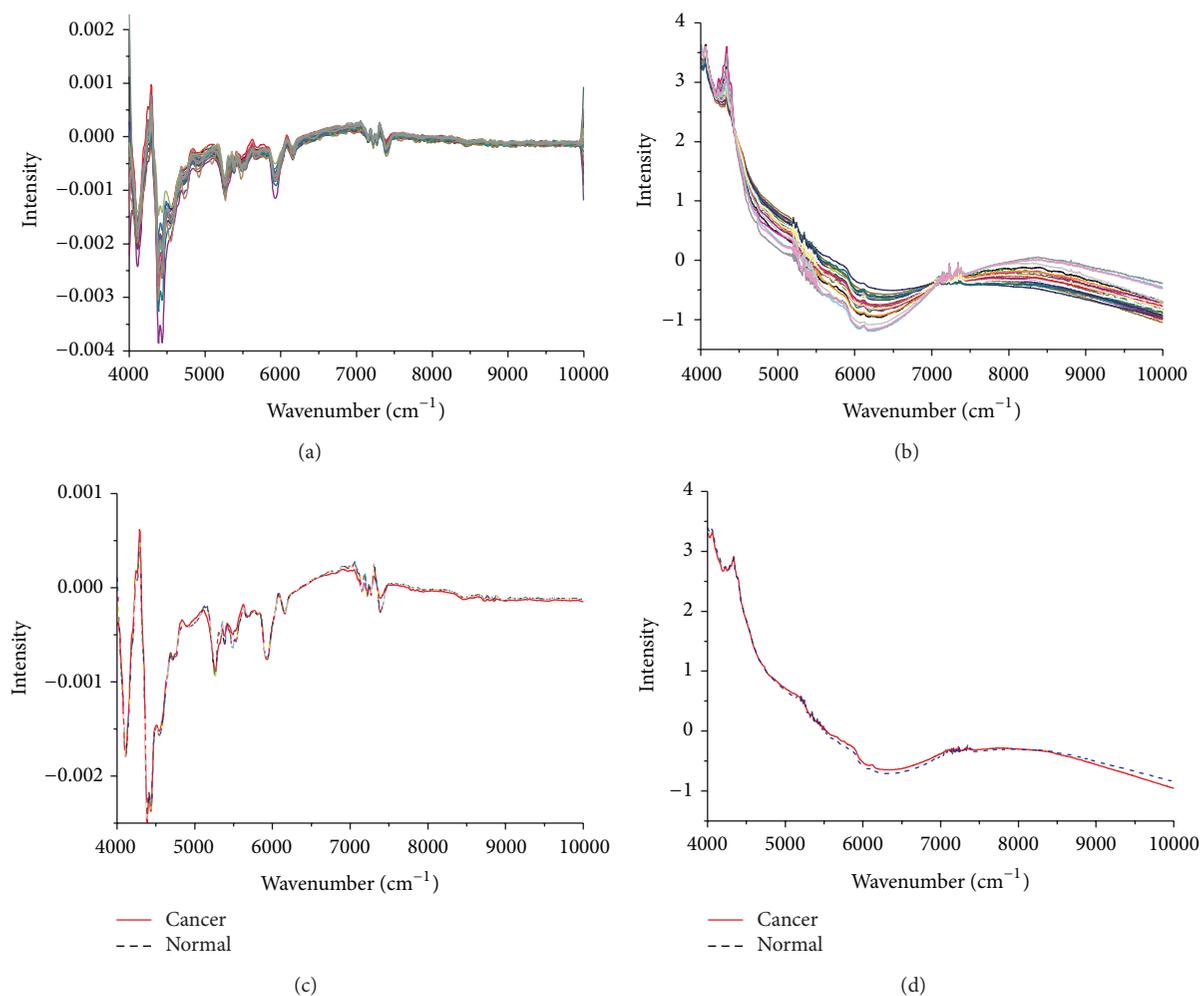


FIGURE 1: Preprocessed and mean spectra of cancerous and normal samples. (a) First derivative-based preprocessed spectra. (b) SNV-based preprocessed spectra. (c) Mean spectra corresponding to first derivative preprocessing. (d) Mean spectra corresponding to SNV preprocessing.

are reasonable because the main difference in the composition between cancerous and normal tissues is mainly from DNA, water, protein, and lipids. In fact, most of the bands in the NIR region were originated from the vibration modes of various functional groups in the molecules of cellular constituents in tissues and cells. A NIR spectrum was actually a mixture of the signatures of many components such as water, proteins, lipids, and carbohydrates. The spectrum of a cancerous tissue was different from the normal one, which was the result of characteristic structural alterations at the molecular level. These alterations include an increase of the ratio of nuclei to cytoplasm, an increase of the relative DNA amount, an enhancement of phosphorylation in proteins, a decrease of the relative RNA amount, and a loss of hydrogen bonding of C-OH in the amino acid residues of proteins; therefore, for a given biological sample, NIR instrument can generate a characteristic spectrum, which must contain a vast amount of potentially useful diagnostic information.

4.2. Principal Component Analysis. The NIR spectrum typically consists of broad, weak, nonspecific, and extensively

overlapped bands and may have hundreds or thousands of variables. Even if SVM is considered to be able to deal with the so-called curse of dimensionality, the use of all variables for classification purposes is often not an adequate strategy. Thus, to investigate the extent to which NIR spectral feature can differentiate cancer and normal specimens, the widely accepted PCA was first used. By linear combinations of the original variables, it can reduce original high-dimensional data into much fewer dimensions called principal components (PCs). Figure 2 gives the explained variance versus the number of principal components and the scatter plots from PCA for both original and SNV-based preprocessing spectra. For original spectra, the first two PCs explained almost all the variance, but the scatter plot indicated that the overlap is very obvious, as shown in Figure 2(c). The findings seem to be in disagreement with some reports [12, 13]. However, it should be pointed out that, compared to the present work, those researches used different kinds of samples and measurement method, that is, an optical fiber probe for performing a direct measurement on a thick tissue for each measurement. This work used formalin-fixed, paraffin-embedded slides with

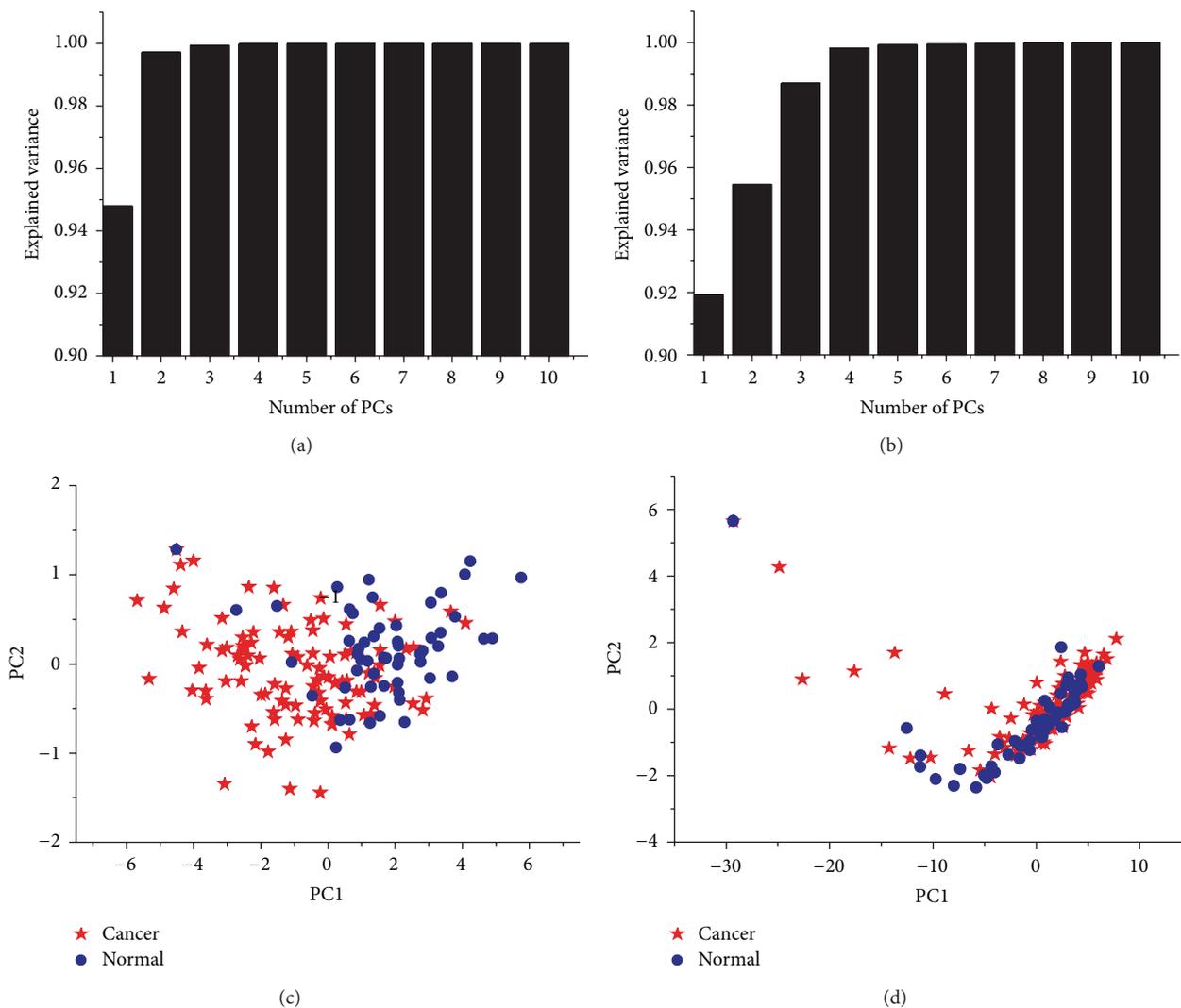


FIGURE 2: The explained variance versus the number of principal components and the scatter plots from principal component analysis for both original and SNV-based preprocessing spectra.

the thickness of $4\ \mu\text{m}$ and therefore removed the signal of water but introduced the absorbance of slice materials. For the spectra preprocessed by SNV, more PCs are needed to explain the same variance and it is almost impossible to obtain any separation in the two-dimensional scatter plot. It seems that even if the SNV can improve spectral visual profiles, it is maybe useless for improving subsequent models.

4.3. Construction of Classification Models. Based on original spectral data and different preprocessing methods, that is, SNV, first derivative, the combination of SNV and PCA (SNV-PC), and the combination of first derivative and PCA (Derivative-PC), five classification models were constructed by LS-SVM algorithms. Here, Derivative denotes the first derivative. It can remove an additive baseline and correct baseline shifts. It is well known that the performance of a LS-SVM model relies largely on the choice of a kernel, which determines the sample distribution in the mapping space. However, the choice of the kernel function is cumbersome

and depends on each case. Thus, the kernel function must be decided first. Here, the simple Gaussian function, radial basis function (RBF), was selected. To obtain a model, only two parameters were needed to be optimized. One is the regularization parameter that controls the tradeoff between the training error and model complexity. The other is the bandwidth which implicitly defines the nonlinear mapping from original space to feature space. Additionally, there are no clear guidelines on how to select the optimal parameters. In this study, an intensive two-step grid search as well as leave-one-out cross-validation procedure was used. Grid search tries values in the region of 10^{-2} – 10^3 . Leave-one-out cross-validation procedure was used to avoid overfitting. Taking Derivative-PC case, for example, Figure 3 shows the contour plot of the parameter optimization of the classification model of LS-SVM. The grid in the first step was 10×10 , and the searching step was relatively large for a crude search. The optimal area was determined by the error contour line. The grid “ \times ” in the second step was also 10×10 , and the searching

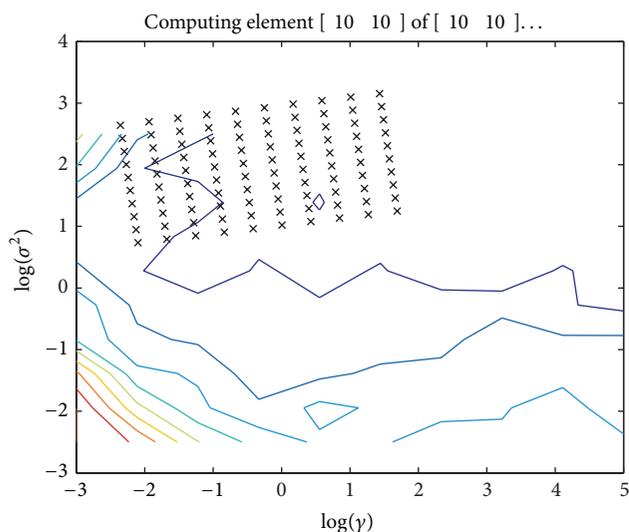


FIGURE 3: Contour plot of the parameter optimization of the classification model of least-squares support vector machine.

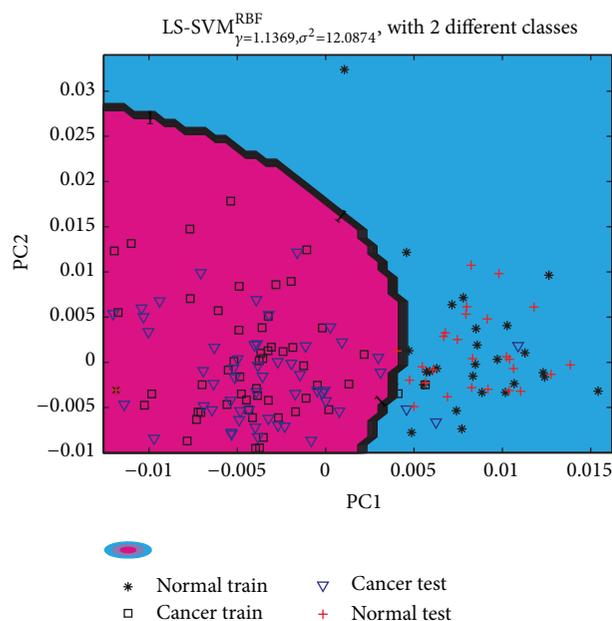


FIGURE 4: Visual expression of the optimal classification model of least-squares support vector machine using the first two principal components as input.

step was the specified search by a small step size. The optimal parameter combination for colorectal cancer discrimination was found at the value of (1.13, 12.09). The discrimination results for all cases were summarized in Table 1. Figure 4 shows visual expression of the optimal classification model of LS-SVM using the first two principal components as input. Furthermore, three measures including misclassified ratio (MCR), sensitivity (SENS), and specificity (SPEC) were used to evaluate the model. SENS and SPEC were defined as the ratios of $TP/(TP+FN)$ and $TN/(TN+FP)$, respectively, where TP is the number of true positive, FP is the number of false

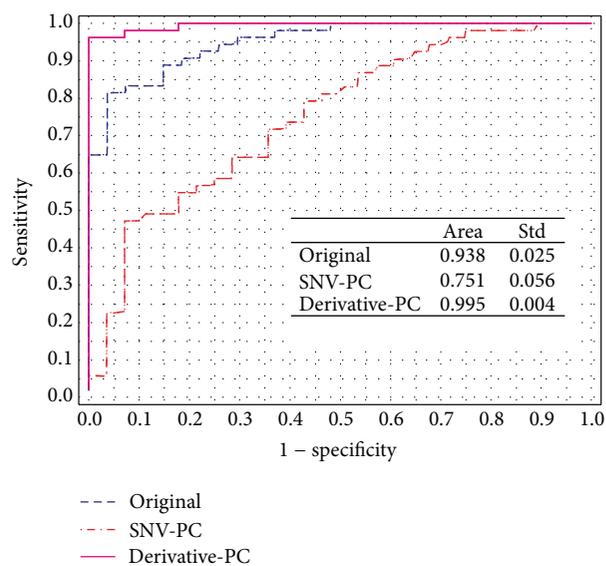


FIGURE 5: Receiver operating characteristic curves of different models built on the basis of the training set.

positive, FN is the number of false negative, and TN is the number of true negative. As shown in the last row of Table 1, by the combination of first derivative preprocessing and PCA to reduce variables, the final LS-SVM model achieved best performance. Additionally, it exhibited similar MCR, SENS, and SPEC values on both the training and test sets, implying that the model was robust and insensitive to overfitting.

In addition to these parameters, the final performance of the developed LS-SVM models was analyzed by receiver operating characteristic (ROC) curve. Figure 5 compares the ROC curve of different models built on the basis of the training set. Such a curve can show the separation ability of a binary classifier and assess the accuracy of classification models [27]. The ROC curve can be represented by plotting sensitivity versus $1 - \text{specificity}$. The area under ROC curve can serve as an index of goodness of the classification model. For a perfect classification model, the area would be one. If the area equals 0.5, the model has no discriminative power at all. An ROC curve is a graphical representation of the trade-off between FN and FP rates for every possible cutoff. As can be seen in Figure 5, the area under ROC curves is 0.94, 0.95, and 0.99 for original spectra, SNV-PC, and Derivative-PC preprocessing cases, respectively, confirming once again the predictive ability of the LS-SVM model after Derivative-PC preprocessing.

5. Conclusions

In conclusion, the results of this study indicate the ability of LS-SVM algorithm coupled with first derivative preprocessing and PCA to construct a model for discriminating colorectal specimens from normal ones. However, there were only a limited number of samples in the present study; more samples need to be collected and further research needs to

TABLE I: Summarization of model performance of different classification models.

Preprocessing method	Training			Test			Parameter		ROC curve	
	MCR	SENS	SPEC	MCR	SENS	SPEC	Gam	Sig2	Area	Std
Original	0.16	0.90	0.71	0.18	0.81	0.81	0.95	165	0.94	0.02
SNV	0	1.00	1.00	0.08	0.94	0.85	34.18	105	1	0
Derivative	0	1.00	1.00	0.32	1.00	0.03	0.69	13.3	1	0
SNV-PC	0.28	0.98	0.21	0.29	0.96	0.18	0.94	1.60	0.75	0.06
Derivative-PC	0.04	0.96	0.96	0.05	0.94	0.96	7.30	4.66	0.99	0.004

be done before translation of the studies to a wide range of clinical applications.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (21375118), the Applied Basic Research Programs of Science and Technology Department of Sichuan Province of China (2013JY0101), Scientific Research Foundation of Sichuan Provincial Education Department of China (12ZA201 and 13ZB0300), Yibin Municipal Innovation Foundation (2013GY018), and Innovative Research and Teaching Team Program of Yibin University (Cx201104).

References

- [1] V. R. Kondepoti, H. M. Heise, and J. Backhaus, "Recent applications of near-infrared spectroscopy in cancer diagnosis and therapy," *Analytical and Bioanalytical Chemistry*, vol. 390, no. 1, pp. 125–139, 2008.
- [2] M. Khanmohammadi, M. A. Ansari, A. B. Garmarudi, G. Hasanzadeh, and G. Garoosi, "Cancer diagnosis by discrimination between normal and malignant human blood samples using attenuated total reflectance-fourier transform infrared spectroscopy," *Cancer Investigation*, vol. 25, no. 6, pp. 397–404, 2007.
- [3] S. Ahmad, A. Shmuel, R. Jagannathan et al., "FT-IR microscopic characterization of normal and malignant human colonic tissues," *Cellular and Molecular Biology*, vol. 47, pp. 159–166, 2001.
- [4] C. Kendall, M. Isabelle, F. Bazant-Hegemark et al., "Vibrational spectroscopy: a clinical tool for cancer diagnostics," *Analyst*, vol. 134, no. 6, pp. 1029–1045, 2009.
- [5] X. Q. Zhang, Y. Z. Xu, Y. F. Zhang et al., "Intraoperative detection of thyroid carcinoma by fourier transform infrared spectrometry," *Journal of Surgical Research*, vol. 171, no. 2, pp. 650–656, 2011.
- [6] M. Khanmohammadi and A. B. Garmarudi, "Infrared spectroscopy provides a green analytical chemistry tool for direct diagnosis of cancer," *Trends in Analytical Chemistry*, vol. 30, no. 6, pp. 864–874, 2011.
- [7] O. J. Old, L. M. Fullwood, R. Scott et al., "Vibrational spectroscopy for cancer diagnostics," *Analytical Methods*, vol. 6, no. 12, pp. 3901–3917, 2014.
- [8] K. Gajjar, J. Trevisan, G. Owens et al., "Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer," *Analyst*, vol. 138, no. 14, pp. 3917–3926, 2013.
- [9] T. W. L. Scheeren, P. Schober, and L. A. Schwarte, "Monitoring tissue oxygenation by near infrared spectroscopy (NIRS): background and current applications," *Journal of Clinical Monitoring and Computing*, vol. 26, no. 4, pp. 279–287, 2012.
- [10] S. K. Teh, W. Zheng, D. P. Lau, and Z. Huang, "Spectroscopic diagnosis of laryngeal carcinoma using near-infrared Raman spectroscopy and random recursive partitioning ensemble techniques," *Analyst*, vol. 134, no. 6, pp. 1232–1239, 2009.
- [11] K. Xu, Y.-H. Xiang, Y.-M. Dai, and Z.-Y. Zhang, "Near infrared spectroscopy combined with principal component analysis applied to diagnosis of endometrial carcinoma," *Chemical Journal of Chinese Universities*, vol. 30, no. 8, pp. 1543–1547, 2009.
- [12] H. Chen, Z. Lin, H. Wu, L. Wang, T. Wu, and C. Tan, "Diagnosis of colorectal cancer by near-infrared optical fiber spectroscopy and random forest," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 135, pp. 185–191, 2015.
- [13] V. R. Kondepoti, M. Keese, R. Mueller, B. C. Manegold, and J. Backhaus, "Application of near-infrared spectroscopy for the diagnosis of colorectal cancer in resected human tissue specimens," *Vibrational Spectroscopy*, vol. 44, no. 2, pp. 236–242, 2007.
- [14] W.-S. Yi, D.-S. Cui, Z. Li, L.-L. Wu, A.-G. Shen, and J.-M. Hu, "Gastric cancer differentiation using Fourier transform near-infrared spectroscopy with unsupervised pattern recognition," *Spectrochimica Acta—Part A: Molecular and Biomolecular Spectroscopy*, vol. 101, pp. 127–131, 2013.
- [15] T. Svensson, J. Swartling, P. Taroni et al., "Characterization of normal breast tissue heterogeneity using time-resolved near-infrared spectroscopy," *Physics in Medicine and Biology*, vol. 50, no. 11, pp. 2559–2571, 2005.
- [16] S. B. Kim, C. Temiyasathit, K. Bensalah et al., "An effective classification procedure for diagnosis of prostate cancer in near infrared spectra," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3863–3869, 2010.
- [17] V. R. Kondepoti, J. Zimmermann, M. Keese, J. Sturm, B. C. Manegold, and J. Backhaus, "Near-infrared fiber optic spectroscopy as a novel diagnostic tool for the detection of pancreatic cancer," *Journal of Biomedical Optics*, vol. 10, no. 5, Article ID 054016, 2005.
- [18] V. R. Kondepoti, T. Oszinda, H. M. Heise et al., "CH-overtone regions as diagnostic markers for near-infrared spectroscopic diagnosis of primary cancers in human pancreas and colorectal tissue," *Analytical and Bioanalytical Chemistry*, vol. 387, no. 5, pp. 1633–1641, 2007.

- [19] X. Li, Q.-B. Li, Y.-Z. Xu et al., "Application of KNN method to cancer diagnosis using Fourier-transform infrared spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 27, no. 3, pp. 439–443, 2007.
- [20] H. Yang, J. Irudayaraj, and M. M. Paradkar, "Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy," *Food Chemistry*, vol. 93, no. 1, pp. 25–32, 2005.
- [21] M. Khanmohammadi, R. Nasiri, K. Ghasemi, S. Samani, and A. B. Garmarudi, "Diagnosis of basal cell carcinoma by infrared spectroscopy of whole blood samples applying soft independent modeling class analogy," *Journal of Cancer Research and Clinical Oncology*, vol. 133, no. 12, pp. 1001–1010, 2007.
- [22] E. Widjaja, W. Zheng, and Z. Huang, "Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines," *International Journal of Oncology*, vol. 32, no. 3, pp. 653–662, 2008.
- [23] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [25] K. S. Chua, "Efficient computations for large least square support vector machine classifiers," *Pattern Recognition Letters*, vol. 24, no. 1–3, pp. 75–80, 2003.
- [26] R. K. H. Galvão, M. C. U. Araujo, G. E. José, M. J. C. Pontes, E. C. Silva, and T. C. B. Saldanha, "A method for calibration and validation subset partitioning," *Talanta*, vol. 67, no. 4, pp. 736–740, 2005.
- [27] C. D. Brown and H. T. Davis, "Receiver operating characteristics curves and related decision measures: a tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 80, no. 1, pp. 24–38, 2006.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

