Int J Clin Exp Med 2017;10(2):2925-2931
www.ijcem.com /ISSN:1940-5901/IJCEM0041693

Original Article

Detection of lymph node metastases in cholangiocanma by fourier transform infrared spectroscopy

Min Wu[1*], Huan-Rong Hu[2*], Long Cui[1], Zhi Xu[1]

[1]Department of General Surgery, Peking University Third Hospital, Beijing 100191, China; [2]Department of Hematology, National Center for Clinical Laboratories and Beijing Hospital, Beijing 100730, China; [*]Equal contributors.

Received July 5, 2016; Accepted December 1, 2016; Epub February 15, 2017; Published February 28, 2017

Abstract: Background: The fast development of Fourier transform infrared (FTIR) spectroscopy has proposed a novel avenue for distinguishing cancerous tissue from normal one. The aim of this research is to investigate the possibility of employing FTIR spectroscopy with support vector machine (SVM) classification as a new spectral discriminant method for classifying metastatic and non-metastatic lymph nodes. Methods: A total of 63 lymph nodes, from cholangiocarcinoma patients who underwent surgery for cholangiocarcinomaectomy, were obtained and then measured by FTIR spectroscopy. The average FTIR spectrum of metastatic and non-metastatic lymph nodes was made. Standard normal variate (SNV) method was exploited to reduce scatter effect, and support vector machine (SVM) classification model was utilized to discriminate spectrum of metastatic from non-metastatic lymph nodes. Leave-one-out cross validation (LOOCV) was used to evaluate SVM model effects. Results: 20 metastatic and 43 non-metastatic lymph nodes were pathologically diagnosed. Two parameters, including I_{1743} (related to Lipid) and I_{1080} (related to Nucleic acid), are founded statistically different between metastatic and non-metastatic lymph nodes. The average FTIR spectra of metastatic group was pronouncedly different from non-metastatic group. The accuracy, sensitivity, and specificity of the SVM model were 88.9%, 75.0%, 95.3%, respectively. Conclusion: A novel approach to differentiate metastatic and non-metastatic lymph nodes in cholangiocarcinoma by employing the FTIR spectroscopy combined with SVM model, is established and demonstrated, and could be applied in cholangiocarcinoma clinical stage value in future clinical practice.

Keywords: Cholangiocarcinoma, lymph node metastases, fourier transform infrared spectroscopy, standard normal variate method, support vector machine classification model

Introduction

Cancer is a serious health problem disease that brings about lots of deaths in worldwide [1]. Although cholangiocarcinoma has a relatively low incidence rate, as one of the most aggressive human carcinomas, it has the clinical features of a high mortality rate and a low long-term survival rate [2]. Thus, the corrective valuation of cholangiocarcinoma clinical stage and timely following treatment is extremely important for improving cholangiocarcinoma patients' prognosis. The tumor node metastasis (TNM) system classification and clinical staging system is an excellent predictor for the prognosis of cholangiocarcinoma and widely used nowadays [3]. TNM staging is closely associated with lymph node metastases. Thus, it is extremely important to measure and identify metastatic from non-metastatic lymph nodes. Current widely applied methods for assessing lymph nodes include frozen-section and histopathology, which have some disadvantages including being laborious intensive and time-consuming [4].

Fourier transform infrared (FTIR) spectroscopy is an absorption spectra technology, which can reflect the subtle vibrational changes of cellular biomolecules' composition or structure, including nucleic acid, protein, carbohydrate, and lipid [5]. This biochemical information is closely tied with to the carcinogenesis of biological tissues. Thus, FTIR spectroscopy has been widely applied to detect various kinds of tumor, including leukemia [6], breast cancer [7], gastric can-

Table 1. Basic information of the investigated patients

Information	Number
Sex	
Male	25
Female	38
Lymph nodes	
Metastatic	20
Non-metastatic	43

cer [8] and lymph node metastases in gastric cancer [9]. Several investigations show that FTIR spectroscopy can serve as a reliable method of cancer detection, a diagnostic tool for differentiating cancerous from normal tissues [10], and even can be used to definite surgical boundary boundaries of cancer resection [11]. However, the way to identify FTIR spectra between benign and malignant biomedical tissues is difficult and trivial, for the spectral data are extremely complex and complicated.

While support vector machine (SVM) classification is a powerful binary model that can represent data as points in the high dimensional space and separate different categories by maximizing distances to its closest data points [12]. This classification model has the capability to identify and learn highly nonlinear relationships from input data. SVM-based classifier approaches have been employed to predict human epidermal growth factor receptor (EGFR/ErbB-1) inhibitors [13] and have been proposed for detecting Piwi-interacting RNAs (piRNAs) with the accuracy of 98% [14]. The support vector machine classification is one kind type of supervised machine learnings, and has been widely applied in large datasets analysis among the biological and other scientific fields during last decades [15].

The purpose of this study was to investigate the possibility of employing FTIR spectroscopy as a potential approach for classifying metastatic and non-metastatic lymph nodes of cholangiocarcinoma in combination with support vector machine (SVM) classification, with the intent to get a high classification accuracy.

Materials and methods

Tissue samples

A total of 63 lymph nodes were obtained from cholangiocarcinoma patients who underwent surgery for cholangiocarcinoma ectomy in the Surgery Department of Peking University Third Hospital, China. At room temperature, all samples were cut into cubes (1.0 cm × 1.0 cm × 1.0 cm) and later measured on an attenuated total reflection (ATR) detection accessory which had attached to spectrometer in the shortest possible time. After characterization by the FTIR spectroscopy, these samples were sent for definitive histopathology by serial sectioning immediately. Pathologic diagnosis was performed by two independent pathologists. All procedures were approved by the Peking University Biomedical Ethics Committee and Institutional Review Board of Peking University Third Hospital (IRB00001052-11034), and written informed consent was obtained from each patient.

Collection of FTIR spectra

The FTIR spectra of the specimens were measured by means of a WQF-660 FTIR spectrometer (Beijing Rayleigh Analytical Instrument Corporation, Beijing, China) with ATR accessory. A Spectra-Tech mid-IR optical fiber was utilized to connect ATR with FTIR spectrometer. Scan wave number' srange is from 1800 cm^{-1} to 1000 cm^{-1} at a resolution of 8 cm^{-1}. 32 spectrums were collected from each specimen and then were combined to one for achieving an acceptable signal-to-noise ratio.

Statistical analysis method

Spa Pro version 2.2 software (College of Chemistry and Molecular Engineering, Peking University, Beijing, China) was used to baseline correct, smooth the spectra, and measure the peak intensity of each spectra band. SPSS version 20 software (IBM, Armonk, New York, USA) was used for statistical analysis. Tests of normal distribution and variance of homogeneity were performed for all parameters. Normally distributed data were analyzed with Student's *t* test. In all instances $P<0.050$ was considered to be statistically significant.

Support vector machine (SVM) classification were carried out to distinguish the FTIR spectra between metastatic and non-metastaticlymph nodes. Before performing SVM, standard normal variate (SNV) method was adopted to preprocess spectroscopic data sets for reducing effects of baseline shift and non-specific scatter at the surface of the samples [16]. MATLAB

Table 2. Preliminary assignments of characteristic bands of FTIR spectra

Peak position (cm^{-1})	Vibrations of the groups	Reference substances
1743	V (C=O)	Lipid
1640	Amide I band	Protein
1550	Amide II band	Protein
1460	Δ (C-H)	Lipid
1400	Δ (C-H), δ (C-O-H)	Lipid
1250	v_{as} PO_2^-	Nucleic acid
1160	N (C-O), δ (C-O-H), δ (C-O-C)	Carbohydrate
1080	v_s PO_2-	Nucleic acid

v_{as}, asymmetric stretching vibration, v_s, symmetric stretching vibration, δ, bending vibration.

Table 3. Spectral parameters' comparison between metastatic and non-metastatic lymph nodes

Parameters	Metastatic		Non-metastatic		*t* value	*P*
	N	Mean ± SD	N	Mean ± SD		
I_{1743}	20	0.018±0.008	43	0.034±0.024	-2.807	0.023
I_{1640}	20	0.291±0.084	43	0.257±0.107	1.249	0.217
I_{1550}	20	0.105±0.035	43	0.102±0.026	0.336	0.738
I_{1460}	20	0.012±0.005	43	0.011±0.006	0.049	0.961
I_{1400}	20	0.012±0.006	43	0.011±0.005	0.520	0.605
I_{1250}	20	0.010±0.004	43	0.008±0.005	1.314	0.194
I_{1160}	20	0.005±0.004	43	0.007±0.005	-1.584	0.118
I_{1080}	20	0.018±0.009	43	0.007±0.005	3.242	0.002

I: peak intensity.

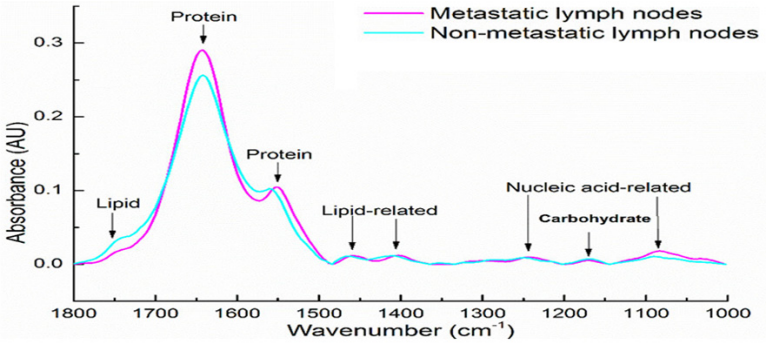

Figure 1. Average Fourier transform infrared spectrum from metastatic and non-metastatic lymph nodes (wave number between 1800-1000 cm^{-1}). The intensity of non-metastatic at wave number 1743 cm^{-1} (related to lipid) is significantly higher than metastatic's. While the intensity of non-metastatic at wave number 1640 cm^{-1} (related to protein) and 1080 cm^{-1} (related to nucleic acid) is much lower than metastatic's. These differences strongly contribute to distinguish non-metastatic and metastatic lymph nodes.

R2013a (MathWorks, Inc., Natick, Mass., USA) was used for SNV pre-processing and SVM model building. The spectroscopic data of sample *i* at wavenumber k could be standard normalized as (1):

$$x_{ik,\,SNV} = \frac{x_{i,k} - \bar{x}_i}{\sqrt{\sum_{k=1}^{p}\left(x_{i,k} - \bar{x}_i\right)^2}}\left(p - 1\right)^{1/2} \quad (1)$$

Where $\bar{x}_i$ means the average of spectroscopic data of sample *i*, while *p* is the number of the wavelength, and (*p*-1) is the freedom degrees.

Then SVM algorithm, which employs a non-linear mapping to transform the original training data into higher dimensional data, was carried out to distinguish metastatic and non-metastatic groups [17]. The principle is explained as following [18]. For labeled training data of the form (x_i, y_i) $i\in\{1,\cdots,n\}$ where x_i is an n-dimensional Feature vector and $y\in\{-1,1\}$ the labels, a decision function is found representing a separating hyperplane defined as (2):

$$f(x) = \left(\langle \omega, \phi(x_i)\rangle + b\right) \quad (2)$$

where ω is the weight vector, *b* is the bias value, and $\Phi(x)$ is the kernel function.

By projecting the data using a mapping $\Phi(x)$, nonlinear decision boundaries in the in-put data space can be obtained. Finding the hyperplane while maximizing the margin is formulated as the following optimization problem:

subject to:

$$\min \frac{1}{2}\omega T\omega + C\sum_{i=1}^{N}\xi_i$$
$$y_i\left(\langle \omega, \phi(x_i)\rangle + b\right) \geq 1 - \xi_i, \quad (3)$$
$$\xi_i \geq 0, \left(i = 1, \ldots, N\right)$$

where *C* is the cost parameter constant, *§i* is parameter for handling non-separable data, and the index *i* labels the *N* training cases. Note

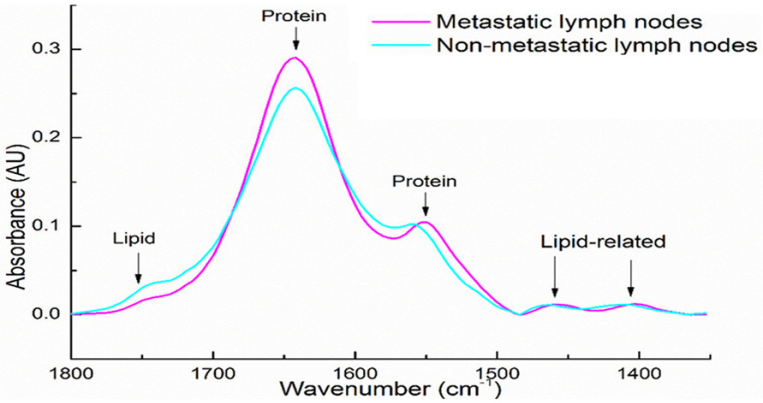

Figure 2. Average Fourier transform infrared spectrum from metastatic and non-metastatic lymph nodes (wave number between 1800-1350 cm^{-1}). The differences of intensity between metastatic and non-metastatic are more obvious, especially at wave number 1743 cm^{-1}, 1640 cm^{-1} and 1550 cm^{-1}.

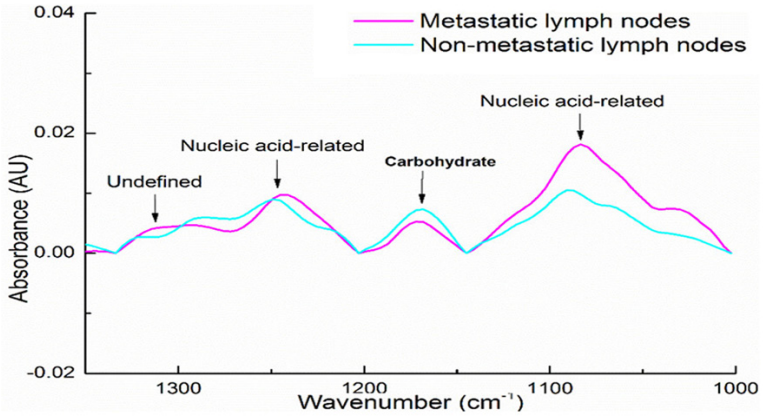

Figure 3. Average Fourier transform infrared spectrum from metastatic and non-metastatic lymph nodes (wave number between 1350-1000 cm^{-1}). The differences of intensity between metastatic and non-metastatic at wave number 1160 cm^{-1} (related to carbohydrate) and 1080 cm^{-1} (related to nucleic acid) are more visible.

that $y \in \{-1,1\}$ is the class labels, and x_i is the independent variables.

Nonlinear classification using the Gaussian kernel as following equation (4) was investigated:

$$K(x, x_i) = \exp\left(-\frac{\| x - xi \|^2}{\sigma}\right) \quad (4)$$

where σ is the kernel width, which controls the amount of the local influence of support vectors on the decision boundary.

A retrospective validation and leave-one-out cross-validation were also used to evaluate the discriminatory power of SVM method. The results of pathologic diagnosis served as standard, and were compared with the classification from the SVM. The sensitivity, specificity and accuracy of SVM method were calculated. The formulas were as follows:

Sensitivity = TP/(TP+FP)

Specificity = TN/(TN+FN)

Accuracy = (TP+TN)/(TP+TN+FP+FN)

where TP means the true positive ones, FP means the false positive ones, TN means the true negative ones, FN means the false negative ones.

Results and discussion

The basic information of patients participated is summarized in **Table 1**. Eight peaks were identified and given a preliminary assignment related to biomolecules including nucleic acid, protein, lipid or carbohydrate, as reported in **Table 2**. Eight peak intensity (*I*) parameters, which could be used to roughly measure the relative content of a biomolecule, were calculated. The basic theory of FTIR spectroscopy differentiating cells or tissues is that, FTIR spectrum has its unique band spectral properties, which could indirectly reflect the alterations about the conformation of functional groups, the order of chemical bonds, the amount of hydrogen bonding and the secondary protein structure in cells. Two parameters, including I_{1743} (related to Lipid) and I_{1080} (related to Nucleic acid), are founded statistically different between metastatic and non-metastatic lymph nodes, as is presented in **Table 3**. The average FTIR spectrum of metastatic and non-metastatic groups of lymph nodes after preprocessing of SNV is shown in **Figures 1-3**, in which the differences between FTIR spectra of metastatic and non-metastatic groups become pronounced. Particularly, peak intensities of metastatic group at the wave

Table 4. SVM discrimination compared with pathologic results of lymph node metastases in cholangiocarcinoma

SVM classifier results	Pathologic results		Sensitivity	Specificity	Accuracy
	Metastatic	Non-metastatic			
Metastatic	15	2	75.0%	95.3%	88.9%
Non-metastatic	5	41			

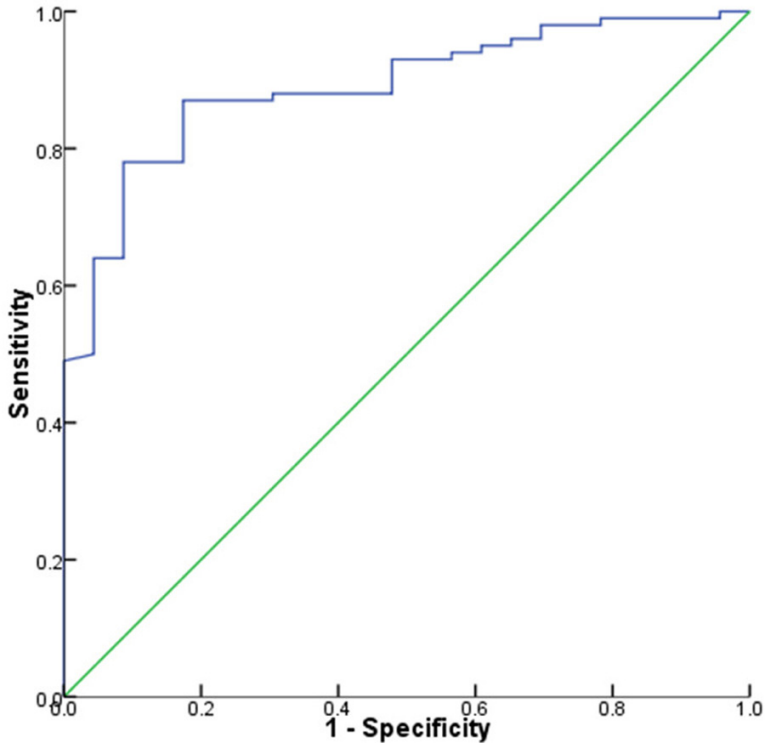

Figure 4. ROC analysis of SVM model: the value of AUC is 0.889, 95% CI [0.826, 0.953], P<0.000.

numbers related to lipid, protein and nucleic acid, are higher than that of non-metastatic group. All of these above findings indicate that, there could be a faster cellular metabolism, enhanced cell proliferation with more nucleic acid produced and an increased expression of proteins in cancer cell, as we had reported in previous works [19, 20].

SVM discriminant model was applied to classify metastatic and non-metastatic groups of lymph nodes and LOOCV was utilized to evaluate the efficiency of SVM model. Comparison between FTIR spectroscopy technique and standard pathologic diagnosis is presented in **Table 4**. As is presented, of the 43 known cases non-metastatic samples, 41 cases were correctly classified, with 2 samples being misjudged; among from the 20 known cases of metastatic samples, 15 cases were correctly classified, with 5 samples being misjudged. The accuracy, sensitivity, and specificity of diagnostic model were 88.9%, 75.0%, 95.3%, respectively, which clearly indicate that this novel approach could well differentiate the FTIR spectra of metastatic versus non-metastatic groups. The ROC analysis of SVM model is presented in **Figure 4**. Therefore, FTIR spectroscopy combined with the SVM classifiers system has the potential to offer a new, safe and effective approach in for the diagnosis and classification of metastatic and non-metastatic lymph nodes and helping physicians to better to value patient's clinical stage based on the number of metastatic lymph nodes and to practice corresponding treatments. The results are in confirmation of sentinel lymph node metastases in breast carcinoma by Fourier transform infrared spectroscopy [21]. However, the ultimate goal of developing this novel approach is not to replace histopathology with the FTIR technique, but to offer another effective way of distinguishing metastatic from non-metastatic lymph nodes. Additionally, as hortage of this pilot research is that, the total number of lymph nodes studied is not large. Thus, the next step we have planned to do is to increase the number of specimen in a prospective multicenter study. Though this step, the SVM classification model would become more powerful and further improvements in sensitivity and accuracy.

Conclusions

In this study, FTIR spectra characteristics of metastatic and non-metastatic lymph nodes in

cholangiocarcinoma had been illustrated. A new diagnostic approach, FTIR spectroscopy combined with support vector machine classification, can achieve a high discrimination accuracy. The results indicate that FTIR spectroscopy with SVM is practical, and could be applied in cholangiocarcinoma clinical stage value in future clinical practice.

Acknowledgements

This study was supported by grants from Major Research Project of Peking University Third Hospital (BYSY201207).

Disclosure of conflict of interest

None.

Address correspondence to: Zhi Xu, Department of General Surgery, Peking University Third Hospital, Beijing 100191, China. E-mail: xuzhi123456@sohu.com

References

[1] Siegel R, Ma J, Zou Z and Jemal A. Cancer statistics, 2014. CA Cancer J Clin 2014; 64: 9-29.

[2] Luvira V, Nilprapha K, Bhudhisawasdi V, Pugkhem A, Chamadol N and Kamsa-ard S. Cholangiocarcinoma patient outcome in northeastern Thailand: single-center prospective study. Asian Pac J Cancer Prev 2016; 17: 401-406.

[3] In: Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, editors. AJCC cancer staging manual. 7th edition. New York: Springer; 2010.

[4] Weiser MR, Montgomery LL, Susnik B, Tan LK, Borgen PI and Cody HS. Is routine intraoperative frozen-section examination of sentinel lymph nodes in breast cancer worthwhile? Ann Surg Oncol 2000; 7: 651-655.

[5] Henry H Mantsch, Dennis Chapman. Infrared Spectroscopy of Biomolecules. Wiley-Liss: New York; 1996; 24: 187-188.

[6] Sheng DP, Liu XC, Li WZ, Wang YC, Chen XL and Wang X. Distinction of leukemia patients' and healthy persons' serum using FTIR spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2013; 101: 228-232.

[7] Tian PR, Zhang WT, Zhao HM, Lei YT, Cui L, Wang W, Li QB, Zhu Q, Zhang YF and Xu Z. Intraoperative diagnosis of benign and malignant breast tissues by fourier transform infrared spectroscopy and support vector machine classification. Int J Clin Exp Med 2015; 8: 972-981.

[8] Li QB, Wang W, Ling XF and Wu JG. Detection of gastric cancer with fourier transform infrared spectroscopy and support vector machine classification. Biomed Res Int 2013; 2013: 609-613.

[9] Bai YK, Yu LW, Zhang L, Fu J, Leng H, Yang XJ, Ma JQ, Li XJ, Li XJ, Zhu Q, Zhang YF, Ling XF and Cao WL. Research on application of fourier transform infrared spectrometry in the diagnosis of lymph node metastasis in gastric cancer. Guang Pu Xue Yu Guang Pu Fen Xi 2015; 35: 599-602.

[10] Sun X, Xu Y, Wu J, Zhang Y and Sun K. Detection of lung cancer tissue by attenuated total reflection-Fourier transform infrared spectroscopy-a pilot study of 60 samples. J Surg Res 2013; 179: 33-38.

[11] Zhang XQ, Xu YZ, Zhang YF, Wang LX, Hou CS, Zhou XS, Ling XF and Xu Z. Intraoperative detection of thyroid carcinoma by fourier transform infrared spectrometry. J Surg Res 2011; 171: 650-656.

[12] Wang ML, Xuan SY, Yan AX and Yu CY. Classification models of HCV NS3 protease inhibitors based on support vector machine (SVM). Comb Chem High Throughput Screen 2015; 18: 24-32.

[13] Kong Y, Qu D, Chen XY, Gong YN and Yan AX. Self-organizing map (SOM) and support vector machine (SVM) models for the prediction of human epidermal growth factor receptor (EGFR/ErbB-1) inhibitors. Comb Chem High Throughput Screen 2016; 19: 400-411.

[14] Seyeddokht A, Aslaminejad AA, Masoudi-Nejad A, Nassiri M, Zahiri J, Sadeghi B. Computational detection of piRNA in human using support vector machine. Avicenna J Med Biotechnol 2016; 8: 36-41.

[15] Tan MX, Pu JT and Zheng B. Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model. Int J Comput Assist Radiol Surg 2014; 9: 1005-1020.

[16] Barnes RJ, Dhanoa MS and Lister SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 1989; 43: 772-777.

[17] Zhang Y, Ren JC and Jiang JM. Combining MLC and SVM classifiers for learning based decision making: analysis and evaluations. Comput Intell Neurosci 2015; 2015: 423-581.

[18] Keerthi SS and Lin CJ. Asymptotic behaviors of support vector machines with Gaussian kernel. Neural Comput 2003; 15: 1667-1689.

[19] Zhang WT, Tian PR, Zhu Q, Zhang YF, Cui L and Xu Z. Noninvasive surface detection of papillary thyroid carcinoma by fourier transform in-

frared spectroscopy. Chemical Research in Chinese Universities 2015; 31: 198-202.

[20] Liu Y, Xu Y, Liu Y, Zhang Y, Wang D, Xiu D, Xu Z, Zhou X, Wu J and Ling X. Detection of cervical metastatic lymph nodes in papillary thyroid carcinoma by Fourier transform infrared spectroscopy. Br J Surg 2011; 98: 380-384.

[21] Tian P, Zhang W, Zhao H, Lei Y, Cui L, Zhang Y and Xu Z. Intraoperative detection of sentinel lymph node metastases in breast carcinoma by Fourier transform infrared spectroscopy. Br J Surg 2015; 102: 1372-1379.