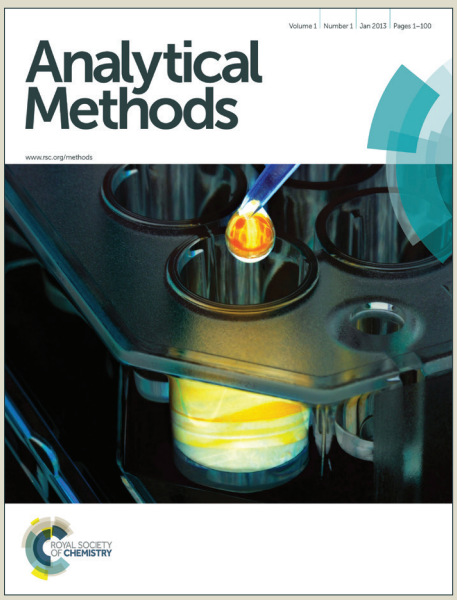
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Terahertz time-domain spectroscopy combined with support vector machines and partial least squares-discriminant analysis applied to diagnosis of cervical carcinoma

Na Qi[a,b], Zhuoyong Zhang[b*], Yuhong Xiang[b], Yuping Yang[c], Xueai Liang[d] and Peter de B. Harrington[e]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

Coupled with terahertz time-domain spectroscopy (THz-TDS) technology, the feasibility of diagnosis of cervical carcinoma using support vector machines (SVM) and partial least squares-discriminant analysis (PLS-DA) had been studied. The terahertz spectra of 52 specimens of cervix were collected. The performance of preprocessing methods of multiplicative scatter correction (MSC), Savitzky-Golay (SG) smoothing and first derivative, principal component orthogonal signal correction (PC-OSC) and emphatic orthogonal signal correction (EOSC) were investigated for PLS-DA and SVM models, respectively. The effects of the different pretreatments methods with respect to classification accuracy were compared. The PLS-DA and SVM models were validated using the bootstrapped Latin-partition method. The SVM and PLS-DA models optimized with the combination of SG first derivative and PC-OSC preprocessing had the best predictive results with classification rates of 94.0 ± 0.4% and 94.0 ± 0.5%, respectively. The proposed procedure proved that terahertz spectroscopy combined with classifiers provides a technology which has potential as a new diagnosis method for cancer tissue.

[*a] *College of Life Science, Capital Normal University, Beijing 100048, China. Fax: +86-10-68902320; Tel: +86-10-68902490; E-mail: sainayi@163.com*
[b] *Department of Chemistry, Capital Normal University, Beijing 100048, China. Fax: +86-10-68902320; Tel: +86-10-68902490; E-mail: gusto2008@vip.sina.com*
[c] *School of Science, Minzu University of China, Beijing 100081, China*
[d] *Haidian District Obstetrics and Gynecology Hospital, Beijing 106863, China*
[e] *Center for Intelligent Chemical Instrumentation, Clippinger Laboratories, Department of Chemistry and Biochemistry, OHIO University, Athens, Ohio, 45701-2979, USA*

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

1 Introduction

Cervical carcinoma is the third most common malignancy in gynecological neoplasm worldwide [1]. Survival rate will be improved if cervical cancer is rapidly and exactly diagnosed. The diagnosis of cervical cancer primarily relies on histopathological examination, the thin prep cytological test (TCT), and manual inspection with colposcopy and cervicography [2]. All of these methods are based on cytology and histology. Living tissue pathological examination mainly relies on the experience of pathologists and takes a lot of time. Therefore, it is important to find a reliable and fast method for diagnosis of cervical cancer.

Terahertz radiation refers to the region that lies between the microwave and infrared regions of the electromagnetic spectrum. The region is commonly defined as 0.1 to 10 THz. Molecular rotations, low frequency bond vibrations, and crystalline phonon vibrations all exist in this frequency range. Due to terahertz radiation is non-ionizing and can highly penetrate for biological tissues, it is a potential source for nondestructive biomedical and biological technology as well as medical imaging [3-11]. Recently, interest in biomedical terahertz research is growing rapidly [12] and much research has been conducted using terahertz spectroscopy and terahertz imaging for medical testing and diagnosis [13]. THz imaging had been used for detecting micro-metastatic foci in the lymph nodes of early-stage cervical cancer [14].

Chemometrics is beneficial to many experimental science and suitable for solving diverse applications including many important practical applications in medicine [15-19]. Principal component analysis (PCA) was used to analyze THz data to understand the origin of contrast in a THz image [20]. Wavelet transforms have been applied to terahertz data for denoising [21]. Combined with terahertz time-domain spectroscopy, the feasibility of fast and reliable diagnosis of cervical carcinoma had been studied [22].

In this study, human cervical tissue was detected by time-domain terahertz spectroscopy. Partial least squares-discriminant analysis (PLS-DA) [23] and support vector machines (SVM) [24] classified the THz data for cervical cancer diagnosis. Various multivariate methods have been used to solve qualitative problems. PLS-DA and SVM are commonly used for classification, however, the importance and effectiveness of the two approaches should be valued for preliminary study on a new topic for research. PLS-DA has some advantages such as the selection of variables and noise reduction. The PLS-DA used in this paper is a self-optimizing method and the optimal parameter in PLS-DA was gotten automatically. SVM provides several advantages such as comparable computational efficiency and excellent generalization capabilities. SVM used the linear kernel function in this work. The two classification methods applied in this project had a big advantage in convenience of calculations for verifying various approaches' performance. Both the two methods may have limitations when handling some complicated classification problems. Further research may be tried to explore nonlinear classification methods based on this preliminary study. The effectiveness and feasibility of preprocessing methods including multiplicative scatter correction (MSC) [25], Savitzky-Golay smoothing and first derivative [26], principal component orthogonal signal correction (PC-OSC) [27] and emphatic orthogonal signal correction (EOSC) [28]were also evaluated.

2 Experiments and methods

2.1 Sample

The cervical tissues (32 normal and 20 cancerous) were provided by the Beijing Haidian Maternal & Child Health Hospital. All the cervical tissues were put into 4% formaldehyde solution to be stabilized, and then were washed with ethanol solutions for dehydration. The tissues were put into xylene for hyalinization, paraffin wax for embedding, and then sliced into 8 μm thick sections. The sections were placed in water for flatting, and then spread upon quartz plates. The slides were put in a regulated heating oven and dried to remove water. Two replicate slides were taken from each of fifty-two tissues sections.

The transmitted THz spectra of all slides were measured by terahertz time-domain spectroscopy system. In order to enhance the absorption of incident light, the two replicate slides of each tissue sample were detected together in the way as given in Fig. 1. The slides were secured with a sample holder which was perpendicular to the direction of light, and then measured with the THz-TDS system. THz spectra for fifty-two samples were collected using the same procedure.

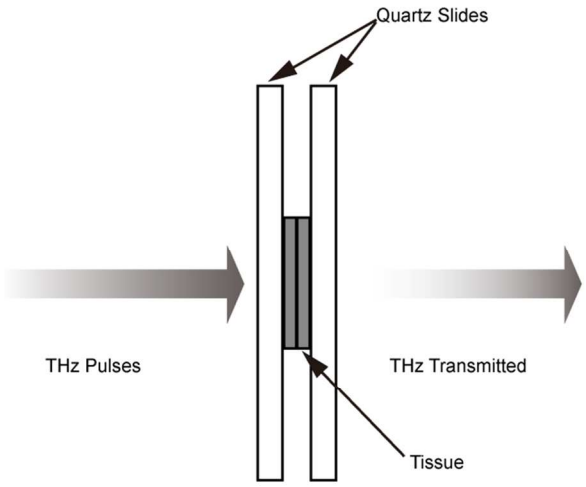


Fig. 1 Schematic representation of tissue samples for TDS measurement.

2.2 Instrumentation

In the experiments, a transmission THz-TDS cell configuration

was used, as depicted in Fig. 2. The system was used a commercially available femtosecond laser (SPECIM, MaiTai). The laser light is separated into two beams. One beam illuminates a GaAs based semiconductor antenna that generates the THz pulse. The coupling efficiency of the THz radiation is then improved by a parabolic mirror with a hemispherical silicon lens. The sample holder is placed at the focus of the parabolic mirror. The beam that passed through the sample is collected by another parabolic mirror and sent to a photoconductive detector. The other beam is the probe which travels through a distance in free space and focuses on the detecting antenna. The probe beam provides a relative time delay which is periodic. In the experiments, the volume of the THz spectra system through which the THz beam passed was filled with dry nitrogen (N_2) to reduce absorption caused by water vapor in air.

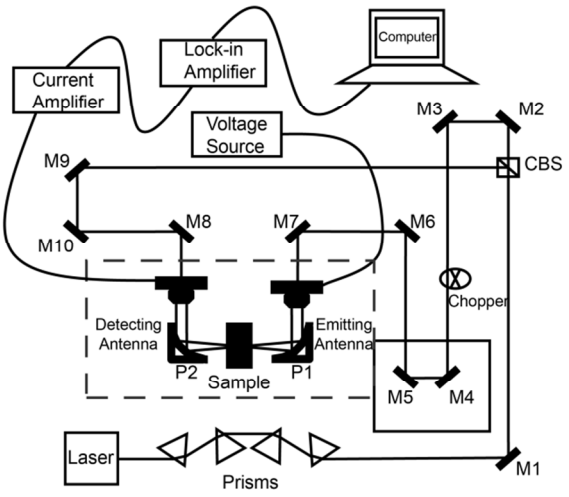


Fig. 2 Schematic of a terahertz time-domain transmission spectrometer system used in this work.

2.3 Theory

2.3.1 Parameters extraction

To calculate the absorption coefficient of a sample, the measurement of a "reference pulse" and a "sample pulse" are required. Because the sample pulse that is transmitted through the tissue slides is measured, the reference signal is the transmitted THz signal without the tissue slides. The THz electric field pulses are directly measured as a function of time and the frequency for both signal and reference. The spectra are obtained by the fast Fourier transform. The sample's refractive index $\mathbf{n}(\omega)$ and absorption coefficient $\mathbf{a}(\omega)$, respectively describing the dispersion and absorption characteristics, can be calculated through the following formula [29-31]:

$$n(\omega) = \frac{\varphi(\omega)c}{\omega d} + 1 \quad (1)$$

$$\alpha(\omega) = \frac{2\kappa(\omega)\omega}{c} = \frac{2}{d}\ln\frac{4n(\omega)}{A(\omega)\left(n(\omega)+1\right)^2} \quad (2)$$

for which d is the sample thickness, is the velocity of light in vacuum, ω, $\kappa(\omega)$ represent the frequency and attenuation coefficient, respectively. $\mathbf{A}(\omega)$, $\varphi(\omega)$ are the amplitude ratio and phase difference of the reference and sample signal.

2.3.2 Partial least squares-discriminant analysis

Partial least squares-discriminant analysis (PLS-DA) is commonly used as a multivariate classification technique based upon the classical partial least square regression method [23]. The PLS-DA algorithm is a supervised method that models the relationship between the measured spectra features and the target variables containing the class label [32]. PLS-DA extracts a set of latent variables by performing a dimension reduction on the data set and finds the maximum separation among the classes [33]. The latent variables explain both the variance of the spectral data **X** as well as the high correlation with the response matrix **Y** that encodes the class membership [34]. For the PLS-DA, then component number is one parameter that needs to be estimated. In this work, the parameter was determined using a self-optimizing PLS-DA from the training data sets and the optimization occurs within each bootstrapped Latin partition [35]. Bootstrapped Latin-partition is a method to verify the performances of classification and calibration models. In PLS-DA and SVM models of this study, the matrix **Y** of category variables was created with 1 for the normal samples and 2 for the malignant samples. When the predicted value $\mathbf{y_{PLS}}$ of a validation sample in PLS-DA model was no larger than 1.5, the sample was assigned to normal class, or assigned to malignant class otherwise.

2.3.3 Support vector machine

The support vector machine (SVM) is a powerful machine learning method [24] ENREF 22. For classification tasks, based on the structural risk minimization principle, this method attempts to find the separating hyperplane which has the largest distance from the nearest training data points [36]. SVM has been extensively used in pattern recognition and regression. LIBSVM was used in this work and the SVM calculations in the paper used the linear kernel function.

2.3.4 Data preprocessing methods

The multiplicative scatter correction (MSC) is a transformation method used to account for scaling and offset effects in spectral data [25, 37]. It removes physical effects like particle size and surface blaze, and it corrects differences in the base line and in the trend.

The Savitzky-Golay method is a polynomial filter that performs numerical differentiation and smoothing [26, 38, 39]. This filter simplifies the computation, and has the ability to process the signals in real-time (with a small delay) with no shifts of the peaks. It can be performed in a computationally efficient procedure with differentiation.

Orthogonal signal correction (OSC) is a data processing technique introduced by Wold et al. [40]. The basic idea of the OSC method is to remove the systematic variations that are orthogonal or not related to the properties of the dependent variables [41, 42]. The removed information is structured noise, such as baseline, instrument variation, and measurement conditions. Principal component orthogonal signal correction (PC-OSC) [27] and emphatic orthogonal signal correction (EOSC) [28] are work by creating bases that are orthogonal to the dependent variables.

2.4 Data treatment and computation

To establish a model for diagnosis of cervical cancer, PLS-DA

and SVM are used to build classification models. Prior to calibration model building, MSC, SG smoothing, SG first derivative, EOSC and PC-OSC are applied to preprocess the signals respectively. Then the data were normalized before used to build model. The performance of preprocessing (MSC, SG smoothing, SG first derivative, PC-OSC and EOSC) and modeling approaches are evaluated by the pooled prediction rates. The pretreatments were constructed from the training data and applied to the prediction sets. Five Latin partitions bootstrapped fifty times were used to measure the generalized prediction accuracy. For each bootstrap, the data was split into training and prediction sets so that each spectrum was used only once in the prediction set. Four Latin partitions were combined into a model-building set, and the fifth was used for prediction. The results of the five prediction sets from each partition were pooled. This approach was used for all the PLS-DA and SVM evaluations and measures of the generalized prediction rates. The average prediction results were calculated across the 50 bootstraps to provide 95% confidence intervals. All model optimization and construction were performed in MATLAB.

3 Results and discussion

The average absorption spectra for normal and malignant tissue are given in Fig. 3A and Fig. 3B, respectively.

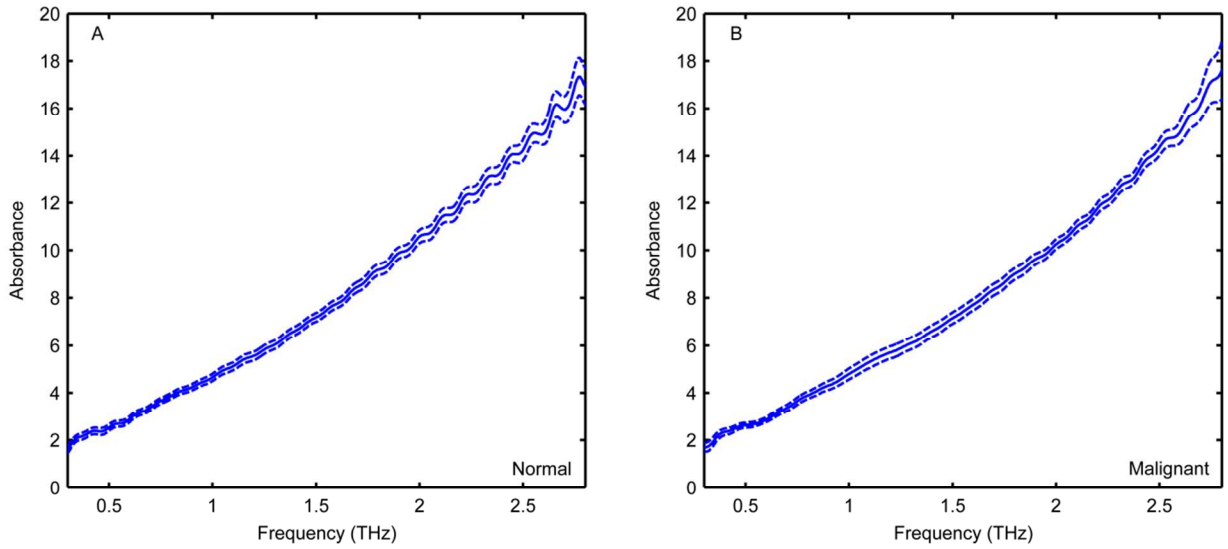


Fig. 3 Normal (part A) and Malignant (part B) tissue spectra with average and 95% confidence intervals.

The processed spectral data applied in PLS-DA were the same as those used in the SVM classifier. The models were also evaluated by using bootstrapped Latin-partitions method with 50 bootstraps and 5 Latin-partitions. Using different pretreatment and different combinations of preprocessing methods respectively, the classification results of PLS-DA and SVM using different preprocessing steps to build model are given in Table 1.

As illustrated in Table 1, the rates of classification were improved for SVM model using SG first derivative, EOSC and PC-OSC respectively as pretreatment investigated in the paper. Comparing the results of the experiments, it could be found the prediction results were less than 86 % for SVM model only with MSC or SG smoothing as preprocessing method. The classification rates obtained by SVM with pretreatment methods of SG first derivative + EOSC and SG first derivative + PC-OSC were higher than with other two different pretreating methods. The prediction rate was significantly improved when the sample data pretreated by MSC + SG first derivative + EOSC.

Table 1 The effect of preprocessing methods to the prediction results for SVM and PLS-DA and 95% confidence intervals from 50 bootstraps and 5 Latin partitions.

Method	SVM	PLS-DA
No	86.0 ± 0.9 %	90.0 ± 1.0 %
MSC	81.2 ± 0.9 %	77.3 ± 1.3 %
SG smoothing	84.7 ± 1.0 %	88.8 ± 1.0 %
SG first derivative	92.0 ± 0.4 %	93.0 ± 0.5 %
EOSC	91.0 ± 0.7 %	90.0 ± 0.8 %
PC-OSC	90.0 ± 0.8 %	90.0 ± 0.9 %
MSC+SG smoothing	81.0 ± 1.0 %	75.3 ± 1.5 %
MSC+SG first derivative	92.6 ± 0.4 %	92.4 ± 0.6 %
MSC+EOSC	87.3 ± 1.0 %	87.0 ± 0.9 %
MSC+PC-OSC	86.7 ± 0.9 %	86.7 ± 1.0 %
SG smoothing +EOSC	91.4 ± 0.8 %	90.0 ± 1.0 %
SG smoothing +PC-OSC	88.7 ± 0.8 %	90.0 ± 0.8 %
SG first derivative + EOSC	94.0 ± 0.4 %	93.8 ± 0.6 %
SG first derivative + PC-OSC	94.0 ± 0.4 %	94.0 ± 0.5 %
MSC+SG smoothing +EOSC	87.5 ± 0.9 %	87.4 ± 0.9 %
MSC+SG smoothing + PC-OSC	86.7 ± 1.0 %	86.4 ± 0.9 %
MSC+SG first derivative + EOSC	94.0 ± 0.4 %	92.2 ± 0.6 %
MSC+SG first derivative + PC-OSC	93.0 ± 0.5 %	93.0 ± 0.6 %

Without any pretreatment method, the average prediction accuracy obtained by PLS-DA model was better than SVM. The classification accuracies of the PLS-DA were significantly improved with pretreatment method of SG first derivative. The prediction accuracies of PLS-DA rose when used SG first derivative, MSC+SG first derivative, SG first derivative + EOSC, SG first derivative + PC-OSC, MSC+SG first derivative + EOSC or MSC+SG first derivative + PC-OSC to process the data.

As reported in Table 1, the PLS-DA model did not achieve the highest average prediction rates simultaneously applying MSC, SG smoothing (or first derivative) and PC-OSC (EOSC) to pretreat the original data. Based on the dataset and methods researched in this paper, it was needed to use the pretreating techniques (SG first derivative and PC-OSC) to optimize model. Fig. 4 gives a comparison of the principal component scores for data objects. The original data scores on the first two principal components are indicated in Fig. 4A, and the principal component scores for the spectra data preprocessed by SG first derivative and PC-OSC (orthogonal components number 3) are shown in Fig. 4B. Fig. 4 displays the two groups in two-dimensional space. The two groups overlapped almost completely without any pretreatments in Fig. 4A, however, they were partially overlapping with SG first derivative and PC-OSC in Fig. 4B. The obvious difference indicated that the preprocessing methods provided effectiveness on classification.

Combined SG first derivative with PC-OSC as signal pretreatment procedure, the prediction accuracies of the optimal SVM and PLS-DA were 94.0 ± 0.4% and 94.0 ± 0.5%, respectively. The SVM and PLS-DA models were built with orthogonal components number 3. Table 2 gives the results of sensitivities and specificities obtained from SVM and PLS-DA using the combination of SG first derivative and PC-OSC with 3 components. These sensitivities and specificities of the fifty bootstraps, five Latin partitions are presented with 95% confidence intervals.

Table 2 The results of sensitivity and specificity and 95% confidence intervals from five Latin partitions and fifty bootstraps for the classification of the tissue thin sections using SG first derivative and PC-OSC.

Method	Sensitivity	Specificity
SVM	88.6 ± 0.5 %	96.7 ± 0.6 %
PLS-DA	92.6 ± 0.8 %	94.9 ± 0.7 %

4 conclusions

In this work, the objective is to combine terahertz spectrum with chemometrics and to propose a reliable and fast diagnosis technique for diagnosing cervical carcinoma. The classification model with SVM and PLS-DA based on terahertz spectral measurement of normal and malignant tissue sections was

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

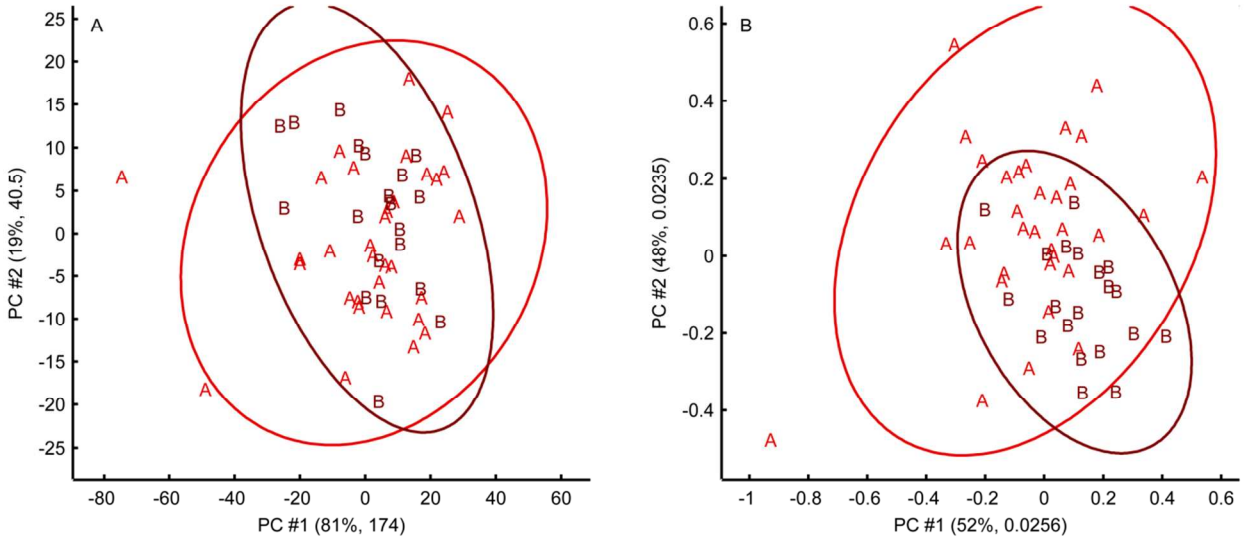


Fig. 4 Comparison of principal component scores for the original signals without pretreatments (part A) and preprocessed data by SG first derivative and PC-OSC (part B). Normal sample is designated by A and cancerous sample is designated by B.

established and the effects of different preprocessing methods to optimize model were investigated. Comparing the classification accuracies pretreated by different preprocessing methods, it indicated that SVM and PLS-DA with the combination of SG first derivative and PC-OSC based on terahertz spectroscopy of tissue can provide a better application for diagnosis of cervical carcinoma.

Acknowledgements

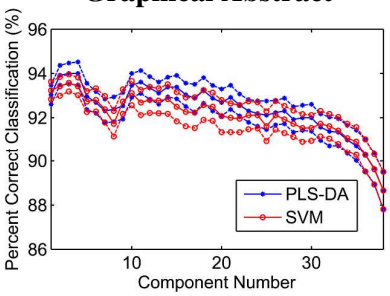
This work was supported by the National Instrumentation Program (2012YQ140005) and the Natural Science Foundation of China (21275101).

References

1. A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward and D. Forman, *CA Cancer J. Clin.*, 2011, 61, 69-90.
2. M. F. Janicek and H. E. Averette, *CA Cancer J. Clin.*, 2001, 51, 92-114.
3. E. Pickwell and V. P. Wallace, *J. of Phys. D: Appl. Phys.*, 2006, 39, R301-R310.
4. A. J. Fitzgerald, V. P. Wallace, M. Jimenez-Linan, L. Bobrow, R. J. Pye, A. D. Purushotham and D. D. Arnone, *Radiology*, 2006, 239, 533-540.
5. R. M. Woodward, V. P. Wallace, D. D. Arnone, E. H. Linfield and M. Pepper, *J. Biol. Phys.*, 2003, 29, 257-261.
6. R. M. Woodward, B. E. Cole, V. P. Wallace, R. J. Pye, D. D. Arnone, E. H. Linfield and M. Pepper, *Phys. Med. Biol.*, 2002, 47, 3853-3863.
7. J. Nishizawa, T. Sasaki, K. Suto, T. Yamada, T. Tanabe, T. Tanno, T. Sawai and Y. Miura, *Opt. Commun.*, 2005, 244, 469-474.
8. E. Pickwell, A. J. Fitzgerald, B. E. Cole, P. F. Taday, R. J. Pye, T. Ha, M. Pepper and V. P. Wallace, *J Biomed. Opt.*, 2005, 10.
9. E. Pickwell, B. E. Cole, A. J. Fitzgerald, V. P. Wallace and M. Pepper, *Appl. Phys. Lett.*, 2004, 84, 2190-2192.
10. K. J. Siebert, T. Löffler, H. Quast, M. Thomson, T. Bauer, R. Leonhardt, S. Czasch and H. G. Roskos, *Phys. Med. Biol.*, 2002, 47, 3743-3748.
11. S. Y. Huang, Y. X. J. Wang, D. K. W. Yeung, A. T. Ahuja, Y.-T. Zhang and E. Pickwell-MacPherson, *Phys. Med. Biol.*, 2009, 54, 149-160.
12. S. T. Fan, Y. Z. He, B. S. Ung and E. Pickwell-MacPherson, *J. of Phys. D: Appl. Phys.*, 2014, 47, 1-11.
13. N. Qi, Z. Zhang and Y. Xiang, *Spectrosc. Spect. Anal.*, 2013, 33, 2064-2070.
14. E. Jung, M. Lim, K. Moon, Y. Do, S. Lee, H. Han, H. J. Choi, K. S. Cho and K. R. Kim, *J. Opt. Soc. Korea*, 2011, 15, 155-160.
15. S. Auephanwiriyakul, E. Sumonphan, N. Theera-Umpon and C. Tayapiwatana, *Comput. Meth. Prog. Bio.*, 2011, 101, 271-281.
16. T. Piroonratana, W. Wongseree, A. Assawamakin, N. Paulkhaolarn, C. Kanjanakorn, M. Sirikong, W. Thongnoppakhun, C. Limwongse and N. Chaiyaratana, *Chemom. Intell. Lab. Syst.*, 2009, 99, 101-110.
17. I. Barman, C. R. Kong, N. C. Dingari, R. R. Dasari and M. S. Feld, *Anal. Chem.*, 2010, 82, 9719-9726.
18. H. R. T. Williams, I. J. Cox, D. G. Walker, B. V. North, V. M. Patel, S. E. Marshall, D. P. Jewell, S. Ghosh, H. J. W. Thomas, J. P. Teare, S. Jakobovits, S. Zeki, K. I. Welsh, S. D. Taylor-Robinson and T. R. Orchard, *Am. J. Gastroenterol.*, 2009, 104, 1435-1444.
19. E. Holmes, J. K. Nicholson, A. W. Nicholls, J. C. Lindon, S. C. Connor, S. Polley and J. Connelly, *Chemom. Intell. Lab. Syst.*, 1998, 44, 245-255.
20. M. A. Brun, F. Formanek, A. Yasuda, M. Sekine, N. Ando and Y. Eishii, *Phys. Med. Biol.*, 2010, 55, 4615-4623.
21. B. Ferguson and D. Abbott, *Microelectron. J.*, 2001, 32, 943-953.
22. N. Qi, Z. Zhang, Y. Xiang, Y. Yang and P. d. Harrington, *Med. Oncol.*, 2015, 32, 1-6.
23. S. Wold, M. Sjöström and L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, 58, 109-130.
24. V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.

25. P. Geladi, D. Macdougall and H. Martens, *Appl. Spectrosc.*, 1985, 39, 491-500.
26. A. Savitzky and M. J. E. Golay, *Anal. Chem.*, 1964, 36, 1627-1639.
27. P. d. B. Harrington, J. Kister, J. Artaud and N. Dupuy, *Anal. Chem.*, 2009, 81, 7160-7169.
28. J. Zhang, Z. Zhang, Y. Xiang, Y. Dai and P. d. B. Harrington, *Talanta*, 2011, 83, 1401-1409.
29. L. Duvillaret, F. Garet and J.-L. Coutaz, *IEEE J. Sel. Top Quantum Electron.*, 1996, 2, 739-746.
30. L. Duvillaret, F. Garet and J.-L. Coutaz, *Appl. Opt.*, 1999, 38, 409-415.
31. T. D. Dorney, R. G. Baraniuk and D. M. Mittleman, *J. Opt. Soc. Am. A*, 2001, 18, 1562-1571.
32. M. Barker and W. Rayens, *J. Chemom.*, 2003, 17, 166-173.
33. Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond and N. Jent, *J. Pharm. Biomed. Anal.*, 2007, 44, 683-700.
34. P. T. Wolter, P. A. Townsend, B. R. Sturtevant and C. C. Kingdon, *Remote Sens. Environ.*, 2008, 112, 3971-3982.
35. P. d. B. Harrington, *Trac-Trend. Anal. Chem.*, 2006, 25, 1112-1124.
36. N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, New York, 2000.
37. K. R. Beebe, R. J. Pell and M. B. Seasholtz, *Chemometrics A Practical Guide*, John Wiley & Sons, New York, 1998.
38. J. Steinier, Y. Termonia and J. Deltour, *Anal. Chem.*, 1972, 44, 1906-1909.
39. H. H. Madden, *Anal. Chem.*, 1978, 50, 1383-1386.
40. S. Wold, H. Antti, F. Lindgren and J. Öhman, *Chemom. Intell. Lab. Syst.*, 1998, 44, 175-185.
41. C. A. Andersson, *Chemom. Intell. Lab. Syst.*, 1999, 47, 51-63.
42. T. Fearn, *Chemom. Intell. Lab. Syst.*, 2000, 50, 47-52.

Graphical Abstract



Combined with terahertz spectroscopy, partial least squares-discriminant analysis and support vector machines could be novel and effective diagnosis methods for cervical cancer.