Optik 138 (2017) 95–102

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Discrimination of traditional herbal medicines based on terahertz spectroscopy

Huo Zhang [a], Zhi Li [a,b,*], Tao Chen [c], Jianjun Liu [c,d]

[a] School of Mechano-Electronic Engineering, Xidian University, Xian, Shanxi 710071, PR China
[b] Guilin University of Aerospace Technology, Guilin, Guangxi 541004, PR China
[c] Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi 541004, PR China
[d] School of Electrical Engineering, Jiujiang University, Jiujiang 332005, PR China

ARTICLE INFO

Article history:
Received 24 August 2016
Accepted 13 March 2017

Keywords:
Terahertz spectroscopy
Traditional herbal medicines
Random forest
ROC and AUC

ABSTRACT

Terahertz (THz) spectroscopy was employed to develop an efficient and applicative way of discriminating traditional herbal medicines in this paper. Spectra of three different herbal medicines (Herba Solani Lyrati, Herba Solani Nigri and Herba Aristolochiae Mollissimae) were obtained in the range of 0.2–1.2 THz. Principal component analysis (PCA) was applied to reduce the dimensionality of original spectral information. Three classification algorithms, support vector machine (SVM), decision tree (DT), and random forest (RF) were used to discriminate the herbal medicines. Receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) were combined with classification accuracy to evaluate the performances of the three classification algorithms. The PCA-RF method got the best ROC curve and AUC, and achieved a prediction accuracy of 99%. The experimental results indicate that THz spectroscopy combined with chemometric algorithms is an effective and rapid method for the discrimination of traditional herbal medicines.

1. Introduction

Herbal medicines account for the largest part of traditional medicines. In traditional Chinese medicine, the proportion reached 87% [1]. Similar names, similar appearances or error records in some traditional pharmacopoeia cause the misuse of herbal medicines, and brought medical safety issues. For example, the dried vine of Caulis Clematidis Armandii is called “Chuanmutong”, and the dried vine of Caulis Aristolochiae Manshuriensis has a similar name – “Guanmutong” [2]. Guanmutong contains aristolochic acid, which may damage the kidney. The mistake use of Guanmutong as Chuanmutong may cause aristolochic acid nephropathy [3]. An example of similar appearances is the confusion of *Gelsemium elegans* and *Lonicera japonica* Thunb. *Lonicera japonica* Thunb is non-toxic, while *Gelsemium elegans* contains 17 kinds of toxic ingredients. Many deaths occurred because of the mistake use of *Gelsemium elegans* as *Lonicera japonica* Thunb [4]. For people’s health and safety, accurate authentication of herbal medicines is paramount.

Traditionally, the discrimination of traditional herbal medicines mainly depends on the observation of morphology [5,6]. This method requires experienced observers, and relies on the subjective judgment of observers. It is scarcely to get a criteria for subjective judgment. Chemical analysis, molecular analysis and chromatography have objective and accurate

* Corresponding author at: School of Mechano-Electronic Engineering, Xidian University, Xian, Shanxi 710071, PR China.
E-mail address: xidianzl@outlook.com (Z. Li).

http://dx.doi.org/10.1016/j.ijleo.2017.03.037
0030-4026/

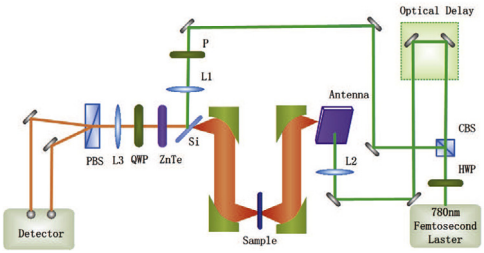

Fig. 1. Experimental setup of the THz-TDS transmission system.

discrimination for herbal medicines [7–11]. However, these methods have disadvantages of time-consuming and labour-intensive. Moreover, large quantity of chemical reagents or organic solvents are needed. This requires additional cost, and may cause environmental pollution of chemical wastes. It is necessary to develop a rapid, efficient environmental protection with low cost and high accuracy for the discrimination of traditional herbal medicines.

Terahertz radiation refers to electromagnetic wave in the region between the microwave and infrared bands, with frequencies from 0.1 to 10 THz (wavelength 30 μm–3 mm). Studies have shown that terahertz time-domain spectroscopy (THz-TDS) is a powerful tool for detection biomolecules, because the vibration and rotational energy levels of most biological molecules distribute in THz band and can be reflected in THz-TDS. For the advantages like detection speed, non-destructive, low cost, timesaving and pollution-free, THz-TDS together with chemometric algorithms is increasingly used in many fields such as security detection [12,13], agricultural detection [14,15], biological detection [16,17] and medical detection [18,19] in recent years. In this work, we use THz-TDS combined with chemometrics methods in the discrimination of traditional herbal medicines.

2. Material and methods

2.1. Sample preparation

We obtained three kinds of traditional herbal medicines (Herba Solani Lyrati, Herba Solani Nigri and Herba Aristolochiae Mollissimae) which might be easily confused. Herba Solani Lyrati were obtained at March, from Jiangsu Province, China. Herba Solani Nigri were obtained at March, from Shanxi Province, China. Herba Aristolochiae Mollissimae were obtained at March, from Hebei Province, China. The herbals were naturally dried, powdered by powder machine and filtrated through a 100-mesh sieve respectively. In order to remove moisture absorbed from the air, all the powders were dried in a vacuum drying oven at 50 °C for 2 h. Then, the dried powders were pressed into circular tablets by a hydraulic press under a pressure of 12 MPa. The thickness, diameter and weight of each tablet were about 1 mm, 12 mm and 180 mg. For each herbal, 400 samples were prepared for measurements.

2.2. Experimental system

THz-TDS transmission system consists of the Z-3 THz time-domain spectrometer (Zomega Terahertz Corp., USA), and the femtosecond laser – FemtoFiber pro NIR (780 nm central wavelength, 100 fs pulse width, 80 MHz repetition rate, 140 mW average power. TOPTICA Photonics Inc., Germany). A schematic of this system is shown in Fig. 1. As the description in Ref. [20], the beam produced by the laser was split into a pump beam and a probe beam by a cubic beam splitter (CBS). THz waves were generated by the photoconductive dipole antenna at the irradiating of the pump beam and focused on the sample. The probe beam were focused onto a ZnTe crystal with the THz waves transmitted through the sample. Then, they were guided to the detection antenna. In order to avoid the interference of water in the air, the apparatus was placed in an airtight box. Dry air had been injecting into the box without interruption until the end of the experiment to ensure the internal relative humidity (RH) is less than 3%.

2.3. Chemometric algorithms

2.3.1. Principal component analysis (PCA)

Too many variables tends to increase the complexity of the subject. Because of the possibly correlation between variables, it can be seen as that variables have some overlapping information. PCA is used to convert a set of possibly correlated variables into a set of linearly uncorrelated variables called principal components (PC). PCs are orthogonal, and the kth principal component (PCk) accounts for the kth maximum variability. The number of PCs is less than or equal to the number of original variables. It means a smaller dimensionality could be obtained.

2.3.2. Support vector machine (SVM)

SVM, proposed by Vapnik in 1995, is one of the effective classification tools. It is widely used in analyzing data and recognizing patterns. The basic thought of SVM is identifying of a hyperplane to separate the two classes from each other correctly. The hyperplane with maximum margin is called optimum hyperplane and makes optimum differentiation. Support vectors means the points limiting the width of the margin. The two-class classification problem can be represented in the form of the following convex optimization problem [21,22]:

$$\min \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{N}(x_i) \tag{1}$$

subject to

$$y_i(\boldsymbol{\omega}^T \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, n \tag{2}$$

$$\xi \geq 0, \quad i = 1, 2, \ldots, n$$

Here, positive constant C is the penalty factor, ξ_i is the slack variable, the weight vector ω and the trend value b define the optimum hyperplane. A commonly used kernel function – radial bias function (RBF) – was used in our SVM model. It is expressed as the following formula [23]:

$$\kappa(x_i, y_i) = \exp\left(-\frac{\|x_i - y_i\|^2}{\gamma^2}\right) \tag{3}$$

Genetic algorithm was used to search the optimal penalty factor C and the optimal RBF parameter γ.

2.3.3. Decision tree (DT, C4.5)

C4.5 algorithm is one of the most popular decision algorithms of decision tree. The main idea of C4.5 is constructing a decision tree by a top-down greedy search through the possible decision tree space. Gain ratio is used in C4.5 algorithm as the criteria of attribute selection. It is defined as the follow [24–26]:

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Information}(S, A)} \tag{4}$$

Split Information(S, A) is the split information and defined as

$$\text{Split Information}(S, A) = \sum_{i=1}^{N} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \tag{5}$$

Gain(S, A) is the information gain and defined as

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{\upsilon \in V(A)} \frac{|S_\upsilon|}{|S|} \text{Entropy}(S_\upsilon) \tag{6}$$

$$\text{Entropy}(S) = \sum_{i=1}^{n} - p_i \log_2 p_i \tag{7}$$

Here, $V(A)$ is the range of attribute A, S_υ is the subset of set S where $A = \upsilon$.

As a classifier, decision tree builds classification models in the form of a tree structure. Given a set S of cases, C4.5 first grows an initial tree using the divide-and-conquer algorithm [27]. If all the cases in S belong to the same class or S is small, the tree is a leaf labeled with the most frequent class in S. Generally, a decision tree is a tree data structure consisting of a root node, more than one decision nodes and leaf nodes. The set S is split into smaller and smaller subsets by an attribute corresponding to the maximum gain ratio. Each decision node represents a test on attribute values. It has two or more branches and each branches represents a result of test. Each leaf node of the tree represents a classification category.

2.3.4. Random forest (RF)

A random forest is an ensemble of decision trees (Section 2.3.3). Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [28]. The sampling method of RF is based on Bootstrap: extract a sample with replacement from a set, and repeat n times to obtain a new set. This sampling method make RF not sensitive to noise or overtraining [29]. An application of a random forest subjects to the following steps:

Table 1
The possible outcomes of a two-class classification.

Actual class	Prediction class	
	True	False
True	True positive (TP)	False negative (FN)
True	False positive (FP)	True negative (TN)

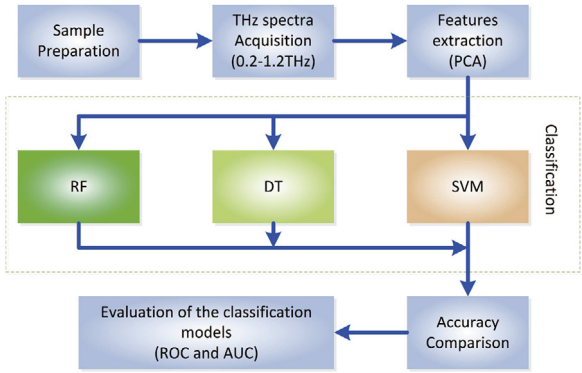

Fig. 2. Flowchart of the procedure.

(1) Obtain N training sets from the training data which has N samples by Bootstrap method. The training sets are labeled as $S_1, S_2, \ldots, S_N$.
(2) Select m features randomly from the feature space for each node. Then, find the best one as the split feature by using a method such as C4.5.
(3) Generate decision trees for the training sets by repeating steps (2).
(4) Predict the samples of the test set by using the decision trees and obtain the corresponding classifications.
(5) RF takes an average or a majority vote from all the predictions of trees to obtain an optimal classification.

In step 3, the number of trees is needed to be set. We set it to a value of experience – 500. It has been proved that the generalization error always converges and overtraining will not happen when increasing the number [30,31].

2.3.5. ROC and AUC

Receiver operating characteristic (ROC) analysis is always used to describe the performance of a detection method distinguishes between 2 mutually exclusive conditions by a graphical method [32]. In classification, ROC analysis is a useful tool for evaluating how well the samples were discriminated because it can measure the correct ratio for each class. Four outcomes would obtained in a two-class classification as shown in Table 1.

True positive rate (TPR) is defined as TPR = TP/(TP + FN) and used to measures the proportion of positives that are correctly identified. False positive rate (FPR) is defined as FPR = TN/(FP + TN). It means how many percentage of the other test samples are incorrectly classified as the authentic class. ROC curve is described by TPR (vertical axis) and FPR (horizontal axis). AUC means the area under the ROC curve. It is a metric for ROC analysis. In the range of 0.5–1, the larger the value of AUC is, the better is the classification. It should be noted that if the value of AUC is equal to 1, all samples were correctly classified. However, if the value of AUC is not greater than 0.5, the classification is meaningless.

2.4. The procedure of the discrimination

As shown in Fig. 2, our work was according to the following procedure:

(1) For each herbal medicine, 300 samples were selected randomly as the training set and 100 samples were selected randomly as the test set from 400 samples (Table 2).
(2) Obtain the THz absorbance spectra of the samples by using THz-TDS.
(3) Feature extraction and selection by the use of PCA.
(4) Create the classification models to discriminate the herbal medicines.
(5) Evaluate the classification models by the comparison of accuracy, ROC curves and AUC values.

Table 2
The numbers of train set and test set.

Type	The number of sample		
	Samples	Train set	Test set
Herba Solani Lyrati	400	300	100
Herba Solani Nigri	400	300	100
Herba Aristolochiae Mollissimae	400	300	100
Total	1200	900	300

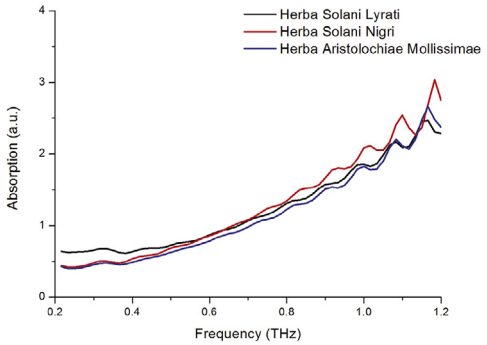

Fig. 3. THz absorbance spectra of the herbal medicines.

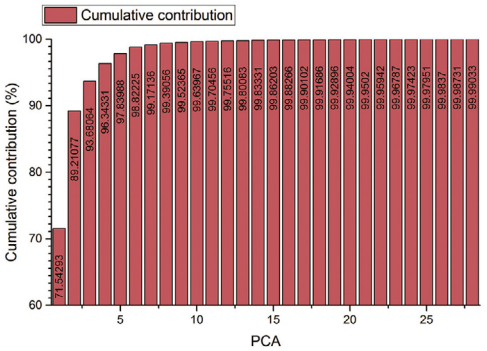

Fig. 4. The cumulative contribution of the top 28 PCs.

3. Results and discussion

3.1. THz absorbance spectra of the herbal medicines

Fig. 3 shows the absorbance spectra curves of 3 kinds of herbal medicines in the frequency range of 0.2–1.2 THz. Each curve contains 142 spectral features. Because the THz absorbance spectra of these herbal medicines are not significantly different, it is difficult to discriminate the samples by their spectral features. Chemometric algorithms are needed to solve this problem. We select four chemometric algorithms: PCA, SVM, DT an RF.

3.2. Feature extraction

A larger number of samples caused a larger number of input variables. It means more redundant information and computation. PCA was used to extract features to reduce the dimensions of input variables. As shown in Fig. 4, the top three PCs explain more than 93.68% of the total contribution to the original data. In the three dimensional (3D) scores plots of the top three PCs shown in Fig. 5, most points of Herba Solani Nigri are distributed separately from others. However, many points of Herba Solani Lyrati and Herba Aristolochiae Mollissimae are mixed. It indicates that the information included in the top three PCs is not enough to distinguish Herba Solani Lyrati and Herba Aristolochiae Mollissimae. More information is needed in the remaining PCs which explain 6.32% of the total contribution. Thus, we chose the top 28 PCs for the feature

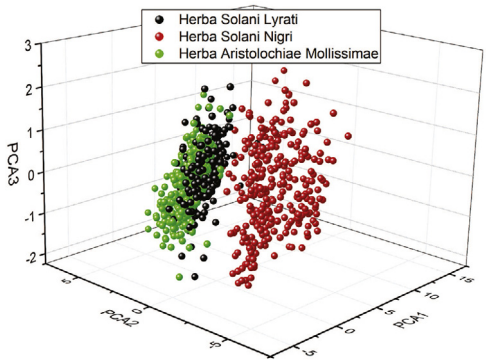

Fig. 5. 3D scattered scores plots of the top three PCs.

Table 3
The classification results.

Type	Classification results (true/total)		
	DT	SVM	RF
Herba Solani Lyrati	87/100	91/100	99/100
Herba Solani Nigri	96/100	100/100	100/100
Herba Aristolochiae Mollissimae	85/100	91/100	98/100
Accuracy	89.33%	94.00%	99.00%

Table 4
The AUC values.

	DT	SVM	RF
Herba Solani Lyrati	0.87958	0.98775	0.99918
Herba Solani Nigri	0.95443	1	1
Herba Aristolochiae Mollissimae	0.91053	0.98860	0.99975

extraction. It can be seen in Fig. 4, the features we extracted explain more than 99.99% of the total contribution. After the feature extraction, the features of each sample was decreased from 142 to 28, and the input variables were decreased more than 80%.

3.3. Classification and evaluation

In order to discriminate the herbal medicines, SVM model, DT model, and RF model were created respectively using the train set with 900 samples after feature extraction. Then, features of the test set which contains 300 samples were inputted into the classification models. AS the result shown in Table 3, the RF model got the best accuracy. For a more intuitive and effective evaluation of the discriminatory performance, we created the ROC curves of each herbal medicine with different classification models (Fig. 6). The AUC values were calculated from the ROC curves and shown in Table 4. A max AUC value means that samples of the corresponding herbal medicine were correctly identified with the greatest probability, and fewest other samples were erroneously judged as the target. The RF model got the max AUC values for all of these three herbal medicines. The SVM model got the same AUC value as the RF model for Herba Solani Nigri, but the AUC values for Herba Solani Lyrati and Herba Aristolochiae Mollissimae are smaller than the RF models. The DT model got the smallest AUC values. Based on these ACU values, the following evaluation can be obtained:

(1) RF model got the most accurate discrimination for each one of the herbal medicines in our work.
(2) SVM model got the same discrimination as RF model for Herba Solani Nigri. However, it not good enough for the discriminations of Herba Solani Lyrati and Herba Aristolochiae Mollissimae.
(3) DT model got the worst discriminations for all these three herbal medicines in our work.

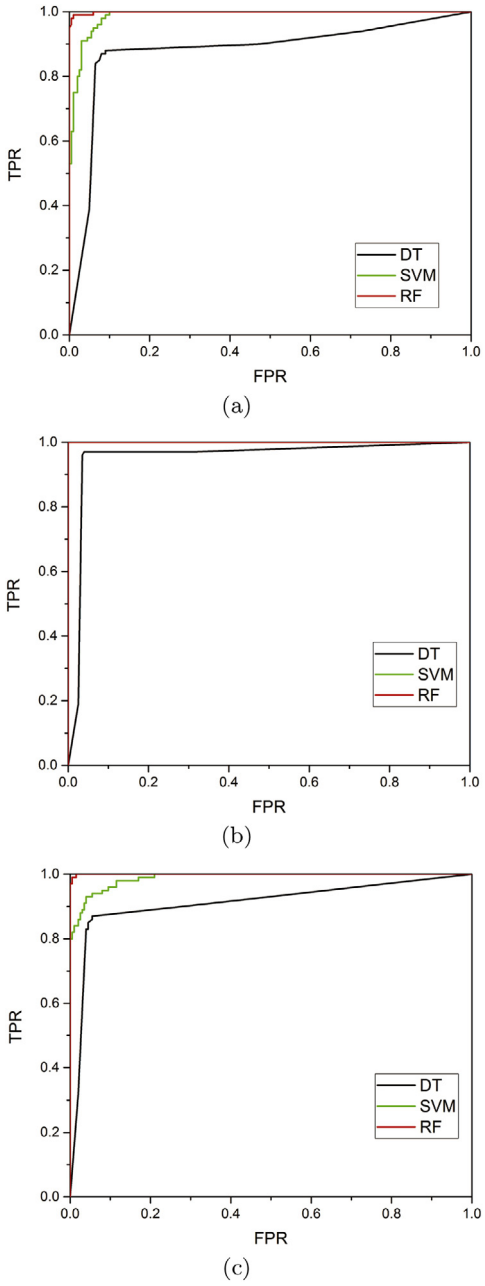

Fig. 6. ROC curves of the three herbal medicines: (a) Herba Solani Lyrati; (b) Herba Solani Nigri; (c) Herba Aristolochiae Mollissimae.

4. Conclusion

In this paper, three kinds of herbal medicines were discriminated quickly and achieved a prediction accuracy over 99% by using PCA-RF in the range of 0.2–1.2 THz. THz spectroscopy combined with chemometric algorithms was indicated to be a feasible and effective method for the discrimination of traditional herbal medicines. This method avoids the high demands on personal experience in traditional herbal detection methods and the reagent consumption in chemical analysis. Although an excellent discrimination was achieved, our work is not enough for a wide variety of herbal medicines. A feature library in THz band with more species of herbal medicines is needed, and it would be an important work with heavy workload.

Acknowledgements

This work is supported by the Natural Science Foundation of Guangxi (No. 2015GXNSFBA139252), and Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (No. YQ16204).

References

[1] C.S.H. Zhang, C. Yuan, Z. Zhang, The kinds of traditional Chinese medicine resources, China J. Chin. Mater. Med. 7 (7) (1995) 387–390, http://dx.doi.org/10.3321/j.issn:1001-5302.1995.07.007.
[2] C.P. Committee, Chinese Pharmacopoeia, Chemical Industry Press, Beijing, 2015.
[3] J. Cosyns, Aristolochic acid and 'Chinese herbs nephropathy' – a review of the evidence to date, Drug Saf. 26 (1) (2003) 33–48, http://dx.doi.org/10.2165/00002018-200326010-00004.
[4] J. Wen, Clinical analysis of 18 poisoning cases caused by *Gelsemium elegans*, Med. Forum 17 (28) (2013) 3764–3765, http://dx.doi.org/10.3969/j.issn.1672-1721.2013.28.061.
[5] B. Carlsward, W. Stern, W. Judd, T. Lucansky, Comparative leaf anatomy and systematics in Dendrobium, sections Aporum and Rhizobium (Orchidaceae), Int. J. Plant Sci. 158 (3) (1997) 332–342, http://dx.doi.org/10.1086/297445.
[6] S. Hayta, N. Tasar, U. Cakilcioglu, O. Gedik, Morphological, karyological features and pollen morphology of endemic *Ebenus haussknechtii* Bornm. ex Hub.-Mor. from Turkey: a traditional medicinal herb, J. Herb. Med. 4 (3) (2014) 141–146, http://dx.doi.org/10.1016/j.hermed.2014.04.006.
[7] L. Dong, J. Wang, C. Deng, X. Shen, Gas chromatography–mass spectrometry following pressurized hot water extraction and solid-phase microextraction for quantification of eucalyptol, camphor, and borneol in Chrysanthemum flowers, J. Sep. Sci. 30 (1) (2007) 86–89, http://dx.doi.org/10.1002/jssc.200600207.
[8] A. Xiong, L. Yang, L. Ji, Z. Wang, X. Yang, Y. Chen, X. Wang, C. Wang, Z. Wang, UPLC-MS based metabolomics study on *Senecio scandens* and *S. vulgaris*: an approach for the differentiation of two Senecio herbs with similar morphology but different toxicity, Metabolomics 8 (4) (2012) 614–623, http://dx.doi.org/10.1007/s11306-011-0354-8.
[9] J. Gonzalez-Rivera, C. Duce, D. Falconieri, C. Ferrari, L. Ghezzi, A. Piras, M.R. Tine, Coaxial microwave assisted hydrodistillation of essential oils from five different herbs (lavender, rosemary, sage, fennel seeds and clove buds): chemical composition and thermal analysis, Innov. Food Sci. Emerg. Technol. 33 (2016) 308–318, http://dx.doi.org/10.1016/j.ifset.2015.12.011.
[10] Y. Yan, Y. Xin-Juan, G. Hui-Min, W. Ru-Lin, S. Rui, T. Yuan, Z. Zun-Jian, Identification and comparative analysis of the major chemical constituents in the extracts of single Fuzi herb and Fuzi-Gancao herb-pair by UFLC-IT-TOF/MS, Chin. J. Nat. Med. 12 (7) (2014) 542–553, http://dx.doi.org/10.3724/SP.J.1009.2014.00542.
[11] S.K.-Y. Law, M.P. Simmons, N. Techen, I.A. Khan, M.-F. He, P.-C. Shaw, P.P.-H. But, Molecular analyses of the Chinese herb Leigongteng (*Tripterygium wilfordii* Hook.f.), Phytochemistry 72 (1) (2011) 21–26, http://dx.doi.org/10.1016/j.phytochem.2010.10.015.
[12] N. Palka, Identification of concealed materials, including explosives, by terahertz reflection spectroscopy, Opt. Eng. 53 (3) (2014), http://dx.doi.org/10.1117/1.OE.53.3.031202.
[13] S. Han, K. Bertling, P. Dean, J. Keeley, A.D. Burnett, Y.L. Lim, S.P. Khanna, A. Valavanis, E.H. Linfield, A.G. Davies, D. Indjin, T. Taimre, A.D. Rakic, Laser feedback interferometry as a tool for analysis of granular materials at terahertz frequencies: towards imaging and identification of plastic explosives, Sensors 16 (3) (2016), http://dx.doi.org/10.3390/s16030352.
[14] J. Liu, Z. Li, F. Hu, T. Chen, A. Zhu, Classification and recognition of transgenic product by terahertz spectroscopy and DSVM, Optik 125 (23) (2014) 6914–6919, http://dx.doi.org/10.1016/j.ijleo.2014.08.054.
[15] Y. Jiang, H. Ge, F. Lian, Y. Zhang, S. Xia, Discrimination of moldy wheat using terahertz imaging combined with multivariate classification, RSC Adv. 5 (114) (2015) 93979–93986, http://dx.doi.org/10.1039/c5ra15377h.
[16] T. Chen, Z. Li, W. Mo, Identification of biomolecules by terahertz spectroscopy and fuzzy pattern recognition, Spectrochim. Acta Part A – Mol. Biomol. Spectrosc. 106 (2013) 48–53, http://dx.doi.org/10.1016/j.saa.2012.12.096.
[17] W. Xu, L. Xie, Z. Ye, W. Gao, Y. Yao, M. Chen, J. Qin, Y. Ying, Discrimination of transgenic rice containing the Cry1Ab protein using terahertz spectroscopy and chemometrics, Sci. Rep. 5 (2015), http://dx.doi.org/10.1038/srep11115.
[18] K.I. Zaytsev, K.G. Kudrin, S.A. Koroleva, I.N. Fokina, S.I. Volodarskaya, E.V. Novitskaya, A.N. Perov, V.E. Karasik, S.O. Yurchenko, Medical diagnostics using terahertz pulsed spectroscopy, in: 2nd Russia-Japan-USA Symposium on the Fundamental and Applied Problems of Terahertz Devices and Technologies (RJUS TeraTech – 2013), vol. 486 of Journal of Physics Conference Series, Tohoku Univ., Univ. Buffalo, State Univ. New York, Rensselaer, Bauman Moscow State Tech Univ., Moscow, Russia, June 03–06, 2013, 2014, http://dx.doi.org/10.1088/1742-6596/486/1/012014.
[19] M. Kawase, K. Yamamoto, K. Takagi, R. Yasuda, M. Ogawa, Y. Hatsuda, S. Kawanishi, Y. Hirotani, M. Myotoku, Y. Urashima, K. Nagai, K. Ikeda, H. Konishi, J. Yamakawa, M. Tani, Non-destructive evaluation method of pharmaceutical tablet by terahertz-time-domain spectroscopy: application to sound-alike medicines, J. Infrared Millim. Terahertz Waves 34 (9) (2013) 566–571, http://dx.doi.org/10.1007/s10762-013-9994-2.
[20] C. Tao, L. Zhi, M. Wei, H. Fang-rong, Simultaneous quantitative determination of multicomponents in tablets based on terahertz time-domain spectroscopy, Spectrosc. Spectr. Anal. 33 (5) (2013) 1220–1225, http://dx.doi.org/10.3964/j.issn.1000-0593(2013)05-1220-06.
[21] H. Drucker, D. Wu, V. Vapnik, Support vector machines for spam categorization, IEEE Trans. Neural Netw. 10 (5) (1999) 1048–1054, http://dx.doi.org/10.1109/72.788645.
[22] A. Astorino, A. Fuduli, The proximal trajectory algorithm in SVM cross validation, IEEE Trans. Neural Netw. Learn. Syst. 27 (5) (2016) 966–977, http://dx.doi.org/10.1109/TNNLS.2015.2430935.
[23] O. Chapelle, P. Haffner, V. Vapnik, Support vector machines for histogram-based image classification, IEEE Trans. Neural Netw. 10 (5) (1999) 1055–1064, http://dx.doi.org/10.1109/72.788646.
[24] T. Hidekazu, A. El-Sayed, F. Masao, M. Kazuhiro, T. Kazuhiko, J. Aoe, Estimating sentence types in computer related new product bulletins using a decision tree, Inf. Sci. 168 (1–4) (2004) 185–200, http://dx.doi.org/10.1016/j.ins.2004.02.004.
[25] D. Yu, X. Huang, Q. Hu, R. Zhou, H. Wang, Y. Cui, Short-term solar flare prediction using multiresolution predictors, Astrophys. J. 709 (1) (2010) 321–326, http://dx.doi.org/10.1088/0004-637X/709/1/321.
[26] K.-H. Chen, K.-J. Wang, K.-M. Wang, M.-A. Angelia, Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data, Appl. Soft Comput. 24 (2014) 773–780, http://dx.doi.org/10.1016/j.asoc.2014.08.032.
[27] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.-H. Zhou, M. Steinbach, D.J. Hand, D. Steinberg, Top 10 algorithms in data mining, Knowl. Inf. Syst. 14 (1) (2008) 1–37, http://dx.doi.org/10.1007/s10115-007-0114-2.
[28] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324.
[29] V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez, An assessment of the effectiveness of a random forest classifier for land-cover classification, ISPRS J. Photogramm. Remote Sens. 67 (2012) 93–104, http://dx.doi.org/10.1016/j.isprsjprs.2011.11.002.
[30] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140, http://dx.doi.org/10.1023/A:1018054314350.
[31] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines, Ore Geol. Rev. 71 (SI) (2015) 804–818, http://dx.doi.org/10.1016/j.oregeorev.2015.01.001.
[32] L. Goncalves, A. Subtil, M. Rosario Oliveira, P.D.Z. Bermudez, ROC curve estimation: an overview, Revstat-Stat. J. 12 (1, SI) (2014) 1–20.