

**DETECTION OF POISONOUS HERBS BY TERAHERTZ TIME-DOMAIN SPECTROSCOPY\*\*****H. Zhang<sup>1\*</sup>, Z. Li<sup>1,2</sup>, T. Chen<sup>3</sup>, J.-J. Liu<sup>4</sup>**<sup>1</sup> School of Mechano-Electronic Engineering, Xidian University, China; e-mail: xidianzl@outlook.com<sup>2</sup> Guilin University of Aerospace Technology, China<sup>3</sup> Guilin University of Electronic Technology, China<sup>4</sup> Jiujiang University, China

The aim of this paper is the application of terahertz (THz) spectroscopy combined with chemometrics techniques to distinguish poisonous and non-poisonous herbs which both have a similar appearance. Spectra of one poisonous and two non-poisonous herbs (*Gelsemium elegans*, *Lonicera japonica* Thunb and *Ficus Hirta* Vahl) were obtained in the range 0.2–1.4 THz by using a THz time-domain spectroscopy system. Principal component analysis (PCA) was used for feature extraction. The prediction accuracy of classification is between 97.78 to 100%. The results demonstrate an efficient and applicative method to distinguish poisonous herbs, and it may be implemented by using THz spectroscopy combined with chemometric algorithms.

**Keywords:** terahertz spectroscopy, poisonous herbs.

**ИДЕНТИФИКАЦИЯ ЯДОВИТЫХ РАСТЕНИЙ МЕТОДОМ ТЕРАГЕРЦОВОЙ СПЕКТРОСКОПИИ С ВРЕМЕННЫМ РАЗРЕШЕНИЕМ****H. Zhang<sup>1\*</sup>, Z. Li<sup>1,2</sup>, T. Chen<sup>3</sup>, J.-J. Liu<sup>4</sup>**

УДК 632.523

Школа механико-электронного машиностроения, Сидианский университет, Китай;  
e-mail: xidianzl@outlook.com

<sup>2</sup> Гуйлиньский университет аэрокосмической техники, Китай<sup>3</sup> Гуйлиньский университет электронных технологий, Китай<sup>4</sup> Университет Цзюцзян, Китай

(Поступила 8 ноября 2016)

Обоснована возможность совместного применения терагерцовой спектроскопии с временным разрешением (THz-TDS) с хемометрическими методами для идентификации ядовитых и неядовитых растений. Спектры ядовитого (гельземия) и двух неядовитых растений (японской айвы и фикуса Хирта) получены в диапазоне 0.2–1.4 ТГц с помощью THz-TDS-системы. Для выделения признаков использован метод главных компонент (PCA). Точность прогноза классификации предложенным методом 97.78–100%.

**Ключевые слова:** терагерцовая спектроскопия, ядовитые растения.

**Introduction.** THz radiation refers to electromagnetic waves ranging from 0.1 to 10 THZ, with the wavelength range between 3 mm and 30  $\mu\text{m}$ . This frequency band is located between microwave and infrared. The research shows that the fingerprint features of biomolecules are concentrated in this band [1, 2]. Because of this feature, it is possible to facilitate the detection of herbs by THz time-domain spectroscopy (THz-TDS).

Many herbs might be identified and used incorrectly because of their similar appearances. This is dangerous and harmful for human health [3–5]. Owing to this, some people even lose their lives [6–8]. In order to protect human health, it is necessary to identify poisonous herbs. The most of common method is based on

\*\* Full text is published in JAS V. 85, No. 1 (<http://springer.com/10812>) and in electronic version of ZhPS V. 85, No. 1 ([http://www.elibrary.ru/title\\_about.asp?id=7318](http://www.elibrary.ru/title_about.asp?id=7318); [sales@elibrary.ru](mailto:sales@elibrary.ru)).

the investigation of the herbs directly. However, it is too subjective and depends on observers' rich experience. Although chemical analysis is objective, accurate and reliable, it is a complex operation and takes more time. In addition, chemical analysis is expensive and can produce pollution. The THz spectroscopy technique can avoid the disadvantages of the appearance observation and the chemical analysis, and brings an effective, convenient, low-cost, and environmentally friendly method for distinguishing poisonous herbs. It should be noted that THz detection technology is widely and successfully used for other purposes, for example, for the detection of genetically modified organisms [9–11]. In addition, Massaouti et al. used THz-TDS for the detection of harmful residues in honey [12]. Li et al. researched the THz-TDS and quantitative analysis of metal gluconates [13]. Gen et al. proposed an attenuated total reflection method to extract the THz spectrum information of drugs [14]. Tetsuya et al. predicted the density and moisture content of wood by THz-TDS [15]. Qin et al. researched the detection of antibiotic residues in food with THz spectroscopy using the PLSR model [16].

In this article, a method combining THz-TDS with chemometrics methods is used to distinguish a poisonous herb from unclassified herbs. We built a kernel-based extreme learning machine (KELM) model to classify the samples of different herbs, and used the Cuckoo-Search (CS) algorithm to optimize the parameters of the KELM model to obtain a better analysis result.

**Experimental samples and apparatus.** We obtained three kinds of herbs (*Gelsemium elegans*, *Lonicera japonica* Thunb, and *Ficus Hirta* Vahl) from the Guilin Herb Market, Guangxi Province, China. *Gelsemium elegans* is poisonous and is easily mistaken for *Lonicera japonica* Thunb and *Ficus Hirta* Vahl. The herbs were dried, powdered, and filtered through a 100-mesh sieve. Then the powder was pressed to circular tablets. Table 1 shows the details of the samples. The samples were divided into two sets: a train set and a test set. We picked 90 samples from each herb randomly and put them into the train set, and put the rest of the samples into the test set.

The time-resolved spectroscopy detection system consisted of the Z-3 THz time-domain spectrometer (Zomega Terahertz Corp., USA), and the femtosecond laser (FemtoFiber pro NIR, TOPTICA Photonics Inc., Germany) with a 780 nm central wavelength, a 100 fs pulse width, a 80 MHz repetition rate, and a 140 mW average power. It is depicted in Fig. 1, and the working principle is described in [17]. In order to avoid the interference of water from the air, the apparatus was placed in an airtight box. Dry air was injected continuously into the box to ensure the internal relative humidity (RH) is less than 3%.

TABLE 1. Details of the Samples

Herbs	Form	Thickness, mm	Diameter, mm	Number of samples
<i>Gelsemium elegans</i>	Tablet	1.2	12	120
<i>Lonicera japonica</i> Thunb	Tablet	1.2	12	120
<i>FicushirtaVahl</i>	Tablet	1.2	12	120

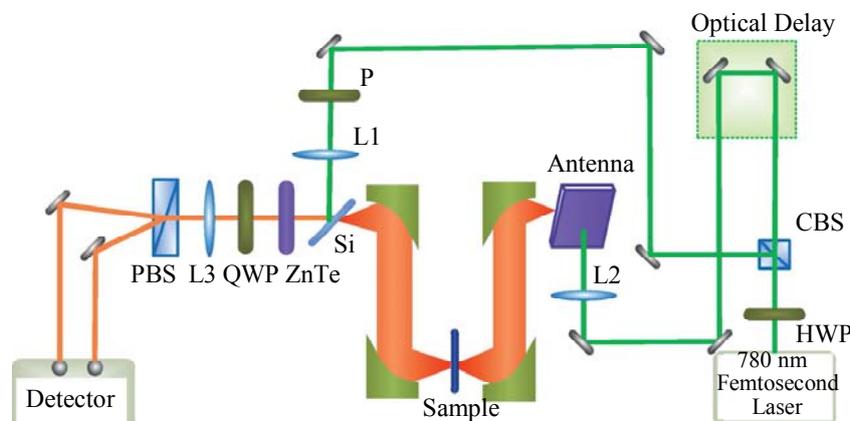


Fig. 1. Schematic diagram of THz-TDS.

**Calculation.** *THz absorbance spectra of the samples.* The THz time-domain spectra of the samples were obtained between 0.2 and 1.4 THz. Under the same conditions, the time-domain reference was obtained from the light path without the sample. We converted the time-domain spectra to the corresponding fre-

quency-domain spectra applying the fast Fourier transformation. The THz absorbance spectra of the samples can be calculated by the following formula:

$$A(\omega) = \lg|E_r(\omega)/E_s(\omega)|^2. \quad (1)$$

Here  $E_r(\omega)$  and  $E_s(\omega)$  are the frequency-domain THz spectra of reference and the sample, respectively.

The absorption spectra of each herb are shown in Fig. 2. There are no obvious differences, and the herbs cannot be recognized by the absorption spectra directly. Additional computational analysis is necessary. Therefore, 270 labeled samples were selected to build a classification model, which is applied to the test rests.

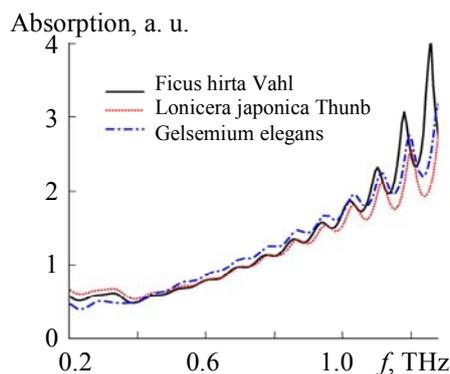


Fig. 2. THz absorbance spectra of the herbs.

*Feature extraction.* In our paper, each absorption curve contains 171 data points. Not all the data points are necessary for the analysis because a correlation between the points exists. So, the feature extraction is required. In order to reduce the dimensionality and summarize the features, principal component analysis (PCA) was performed. The top 10 principal components are shown in Table 2. Meanwhile, the top 3 components explain more than 91.91% of the total contribution. The three-dimensional scores of the top 3 are shown in Fig. 3; the data points of different herbs are distributed separately in the three-dimensional area. Thus, dimensionality is reduced from 171 to 3.

TABLE 2. The Contribution of Top 10 Principal Components

Principal components	PC1	PC2	PC3	PC4	PC5
Contribution, %	72.5296	12.3449	7.0387	2.1820	1.6891
Principal components	PC6	PC7	PC8	PC9	PC10
Contribution, %	0.8750	0.5917	0.5151	0.3921	0.2543

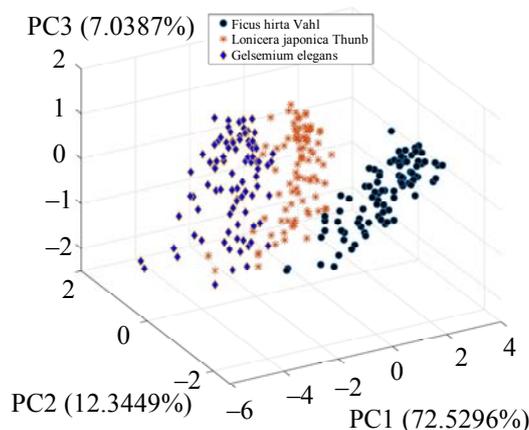


Fig. 3. Scores scatter plot of top 3 principal components.

*Classification model based on CS-KELM.* Extreme learning machine (ELM) is a learning algorithm proposed by Dr. Guang-Bin Huang in 2004 for single-hidden layer feed forward neural networks (SLFNs). Without using conventional optimization algorithms, ELM assigned the input weights and hidden layer biases of SLFNs randomly. Thus, SLFNs can be simply considered as a linear system and the output weights of SLFNs can be analytically determined through a simple generalized inverse operation of the hidden layer output matrices [18, 19]. ELM has proved to have good generalization performance, and the learning speed is extremely fast [20]. Assuming SLFNs has  $L$  hidden nodes and input  $N$  train samples, the output can be expressed by the following formula:

$$f_L(x_j) = \sum_i^L \beta_j g(\omega_i x_j + b_i), \quad j = 1, \dots, N. \quad (2)$$

Here  $\omega_i$ ,  $b_i$ , and  $\beta_i$  are input weights, hidden layer biases, and output weights, respectively, and  $g(\dots)$  is the activation function of the hidden nodes. The matrix form is

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \quad (3)$$

$$\mathbf{H} = \begin{bmatrix} g(\omega_1 x_1 + b_1) & \dots & g(\omega_L x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1 x_N + b_1) & \dots & g(\omega_L x_N + b_L) \end{bmatrix},$$

where  $\mathbf{T}$  is the training data target matrix. The hidden layer output matrix  $\mathbf{H}$  can be obtained based on  $\omega_i$  and  $b_i$ , which are assigned randomly. Then, the output weight vector  $\boldsymbol{\beta}$  can be obtained as:

$$\boldsymbol{\beta} = \mathbf{H}^\dagger \mathbf{T}, \quad (4)$$

where  $\mathbf{H}^\dagger$  is the Moore–Penrose generalized inverse of matrix  $\mathbf{H}$ . The output function of ELM is obtained as:

$$f(x) = h(x)\boldsymbol{\beta} = h(x)\mathbf{H}^\dagger \mathbf{T}. \quad (5)$$

KELM is improved from ELM by Guang-Bin Huang [19, 21]. Kernel functions are used to improve the convergence speed and generalization performance of ELM. The corresponding function of the original ELM is expressed as follows:

$$f(x) = h(x)\boldsymbol{\beta} = h(x)\mathbf{H}'(I/C + \mathbf{H}\mathbf{H}')^{-1}\mathbf{T}. \quad (6)$$

The commonly used radial bias function (RBF) was used in our KELM model. It is expressed by the following formula:

$$\kappa(x_i, y_j) = \exp(-\|x_i - y_j\|^2/\gamma^2). \quad (7)$$

The regularization coefficient  $C$  and the kernel parameter  $\gamma$  are the parameters of KELM. We used cuckoo search (CS) to search  $C$  and  $\gamma$  in this paper.

CS was developed by Xin-She Yang and Suash Deb in 2009. It has been found to be very efficient in solving global optimization problems. This algorithm is inspired by the obligate brood parasitism cuckoo. Three idealized rules describe the standard Cuckoo search [22]: each cuckoo dumps one egg in a randomly chosen nest at a time; if a nest has a high-quality egg that represents a candidate solution, it will be carried over to the next generations; the number of available host nests is fixed, and the parasitic egg might be discovered by the host bird with a probability (between 0 and 1). If this happens, the host bird would either abandon the egg or leave the nest to build a new one.

Based on this idea, the basic steps of the CS algorithm can be summarized as:

- 1) Initialize a population of  $n$  host nests ( $x_i, i=1, 2, \dots, n$ ) randomly.
- 2) Evaluate the objective values to get the current best solution.
- 3) Set a termination condition (such as a fitness value, or cycle times).
- 4) Generate new nests by the Lévy flight [23]

$$x_i^{(\tau+1)} = x_i^\tau + \alpha \oplus \text{Levy}(\lambda). \quad (8)$$

First, we input the data of the train set after the feature extraction to build the classification model. Second, we set the regularization coefficient  $C$  and the kernel parameter  $\gamma$  of KELM for optimal classifications. Third, we used CS to search  $C$  and  $\gamma$ . The process is shown in Fig. 4. It can be represented by the following optimization problem:

$$\begin{aligned} & \max \text{accuracy}(C, \gamma) \\ & \text{s.t. } C \in (0, A], \gamma \in (0, B], \end{aligned} \quad (9)$$

where  $A$  and  $B$  are constants, which are set to 100 in the paper. We used 5-fold cross validation to divide the train set into five equal parts. Each part is selected once for prediction, and the remaining four parts are used for the train. Then the average accuracy of the five predictions is calculated and used as the valuation criteria of the optimization process. Before the termination condition, the cyclic process of searching  $C$  and  $\gamma$  is continued to obtain the optimal results.

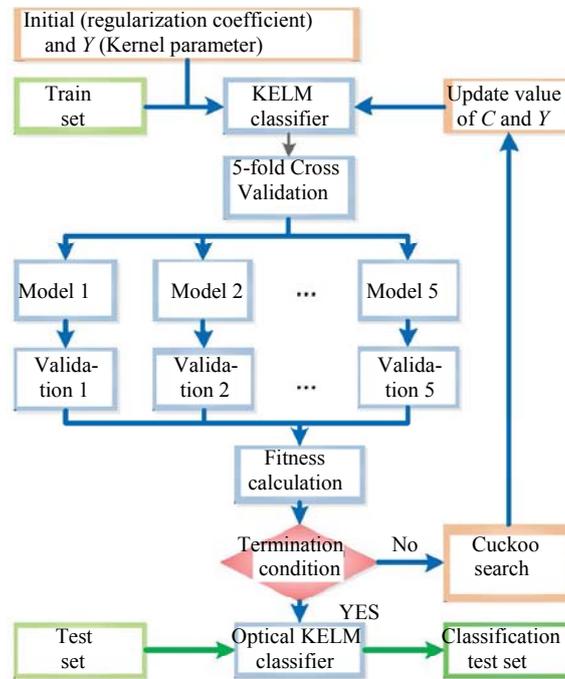


Fig. 4. Flowchart of the procedure.

**Results and discussion.** The optimization of regularization coefficient  $C$  and the kernel parameter  $\gamma$  play an important role in the classification. In order to verify the performance of CS-KELM, we compared its fitness curve with genetic algorithms (GA). According to the parameters given in Table 3, the fitness curves of the two different algorithms are obtained. As shown in Fig. 5, both methods reached the optimal solutions. However, the average fitness of CS ( $>99.5\%$ ) is better than GA ( $<98.9\%$ ). Thus, CS is applicable in our optimization process.

TABLE 3. Parameters of CS and GA

Algorithm	Cross-validation	Nests (CS) or Population (GA)	Iteration
CS	5-fold	20	50
GA	5-fold	20	50

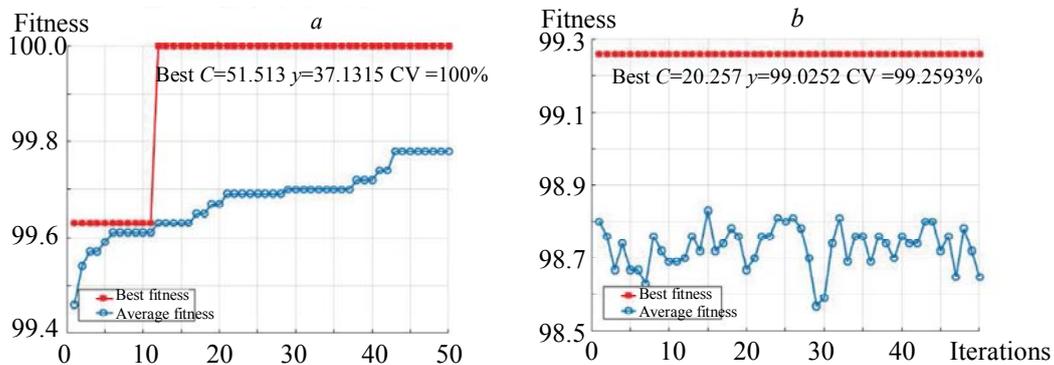


Fig. 5. Fitness curves of CS (a) and GA (b).

We used  $C$ ,  $\gamma$ , and the train set to build the KELM classification model. Then, we input the test set and predict. The 99.89% accuracy was reached through the CS algorithm, and the 96.67% accuracy was reached for the GA algorithm. As shown in Fig. 6a, for CS-KELM all samples of *Gelsemium elegans* were classified correctly; only one sample of *Ficus Hirta* Vahl was classified incorrectly, and *Lonicera japonica* Thunb was recognized. Figure 6b shows the classification of GA-KELM. One sample of *Ficus hirta* Vahl and two samples of *Lonicera japonica* Thunb were classified incorrectly.

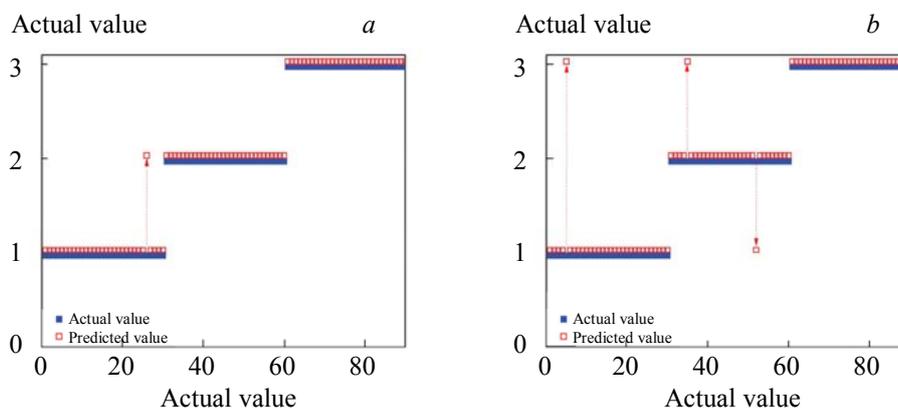


Fig. 6. Classification results of CS-KELM, accuracy 98.8889% (a) and GA-KELM, accuracy 96.6667% (b); Label 1 – *Ficus hirta* Vahl, Label 2 – *Lonicera japonica* Thunb, Label 3 – *Gelsemium elegans*

Furthermore, we repeated this experiment 100 times and compared the results. As shown in Table 4, the maximum accuracy for the two methods was reached at 100%. But the average misjudgment probability for GA-KELM is greater than for CS-KELM, especially for the poisonous herb *Gelsemium elegans*. Therefore, CS-KELM is better than GA-KELM for herb classification, and the poisonous herb can be distinguished exactly.

TABLE 4. The Average Result of 100 Experiments

Type	Average number of mistake classification			Minimum accuracy, %	Maximum accuracy, %	Average accuracy, %
	<i>Ficus hirta</i> Vahl	<i>Lonicera japonica</i> Thunb	<i>Gelsemium elegans</i>			
GA-KELM	0.28	0	0.09	95.56	100	98.77
CK-KELM	0.19	0.01	0	97.78	100	99.33

**Conclusion.** The feasibility and applicability of THz spectroscopy combined with chemometric methods for the identification of poisonous herbs are studied. Three kinds of herbs were testified, and CS-KELM was used to build the classification model. A classification with an average accuracy of up to 99.33% in 100 measurements was obtained, and 100% of the poisonous herb *Gelsemium elegans* was found correctly. Our work provides a method that is fast, convenient, and effective to distinguish poisonous herbs. However, this is a preliminary study, and more poisonous herbs need to be collected and modeled.

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

## REFERENCES

1. T. Chen, Z. Li, W. Mo, *Spectrochim. Acta A*, **106**, 48–53 (2013).
2. A. Markelz, S. Whitmire, J. Hillebrecht, R. Birge, *Phys. Med. Biol.*, **47**, 3797–3805 (2002).
3. Z. H. Guo, H. M. Ma, B. L. Zhang, *J. Chin. Integrative Med.*, **6**, 645–649 (2008).
4. I. S. Yaroshenko, D. O. Kirsanov, P. Wang, D. Ha, H. Wan, J. He, Yu. G. Vlasov, A. V. Legin, *Russ. J. Appl. Chem.*, **88**, 72–81 (2015).

5. Y. Q. J. B. Wang, Y. L. Zhao, L. M. Shan, B. C. Li, F. Fang, C. Jin, X. H. Xiao, *J. Hazard Mater.*, **199**, 350–357 (2012).
6. G. F. Liu, C. R. Zhang, *Lingnan J. Emergency Med.*, **11**, 462 (2006).
7. J. Wen, *Med. Forum*, **17**, 3764–3765 (2013).
8. M. M. Liu, W. Y. Huang, Z. Q. Xu, J. L. Yu, *J. Med. Pest. Control.*, **32**, 807–809 (2016).
9. J. J. Liu, Z. Li, F. R. Hu, T. Chen, A. J. Zhu, *Opt. Quantum. Electron.*, **47**, 313–322 (2015).
10. W. D. Xu, L. J. Xie, Z. Z. Ye, W. L. Gao, Y. Yao, M. Chen, J. Y. Qin, Y. B. Ying, *Sci. Rep.*, **5**, 11115, 1–9 (2015).
11. T. Chen, Z. Li, X. H. Yin, F. R. Hu, C. Hu, *Spectrochim. Acta A*, **153**, 586–590 (2016).
12. O. Cherkasova, M. Nazarov, A. Shkurinov, *Opt. Quantum. Electron.*, **48**, 1–12 (2016).
13. M. Kawase, K. Yamamoto, K. Takagi, R. Yasuda, M. Ogawa, Y. Hatsuda, S. Kawanishi, Y. Hirotsu, M. Myotoku, Y. Urashima, K. Nagai, K. Ikeda, H. Konishi, J. Yamakawa, M. Tani, *J. Infrared Millim. Terahertz Waves*, **34**, 566–571 (2013).
14. G. Takebe, Y. Kawada, K. Akiyama, H. Takahashi, H. Takamoto, M. Hiramatsu, *J. Pharm. Sci.*, **102**, 4065–4071 (2013).
15. T. Inagaki, B. Ahmed, I. D. Hartley, S. Tsuchikawa, M. Reid, *J. Infrared Millim. Terahertz Waves*, **35**, 949–961 (2014).
16. J. Qin, L. Xie, Y. Ying, *Food Chem.*, **170**, 415–422 (2015).
17. T. Chen, Z. Li, W. Mo, F. R. Hu, *Spectrosc. Spectr. Anal.*, **33**, 1220–1225 (2013).
18. G. B. Huang, Q. Y. Zhu, C. K. Siew, *Neurocomputing*, **70**, 489–501 (2006).
19. G. B. Huang, H. M. Zhou, X. J. Ding, R. Zhang, *IEEE Trans. Syst. Cybern.: B*, **42**, 513–529 (2012).
20. J. Tang, C. Deng, G. B. Huang, *IEEE Trans. Neural Netw. Learn. Syst.*, **27**, 1 (2015).
21. G. B. Huang, *Cogn. Comput.*, **6**, 376–390 (2014).
22. *Studies in Computational Intelligence*, Springer Verlag (2014) pp. 1–26.
23. X. S. Yang, S. Deb, *Mathematics*, 210–240 (2010).