Hindawi Publishing Corporation
Journal of Spectroscopy
Volume 2013, Article ID 531747, 11 pages
http://dx.doi.org/10.1155/2013/531747

Research Article

Classification of Green Bristle Grass, Yellow Foxtail and Chinese Pennisetum Seeds via HATR-FT-IR Combined with Chemometrics

Peng Yu,[1] Cha-Yan Wan,[1] Chang-Shun Wu,[2] Jia-Ni Shou,[2] and Cun-Gui Cheng[1]

[1] *College of Chemistry and Life Sciences, Zhejiang Normal University, Jinhua 321004, China*
[2] *Chuyang Honors College, Zhejiang Normal University, Jinhua 321004, China*

Correspondence should be addressed to Cun-Gui Cheng; chengcg123@sina.com

Received 9 June 2012; Revised 22 July 2012; Accepted 26 July 2012

Academic Editor: Shin ichi Morita

Copyright © 2013 Peng Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fourier transform infrared (FT-IR) and horizontal attenuated total reflectance (HATR) technique are used to obtain the FT-IR spectra of the seed of green bristle grass (the seed from *Setaria viridis* (L.) Beauv), yellow foxtail seed (the seed from *Setaria glauca* (L.) Beauv), and the Chinese pennisetum seed (the seed from *Setaria faberii* Herrum). In order to extrude the difference among them, cluster analysis is considered to identify the three kinds of plant seeds. Because they belong to the sibling plant seeds, and have similar chemical components and close FT-IR spectra. The result of Cluster analysis is not satisfactory. The discrete wavelet transformation (DWT) and a support vector machine (SVM) were used for further study. The compression detail 3 and 4 in DWT are used to extract the feature vectors, which are used to train SVM. The trained SVM is used to classify seed of green bristle grass, yellow foxtail seed and Chinese pennisetum seed. The seed samples are collected from different places around the country. With 40 testing samples we could effectively identify the sibling plants, seed of green bristle grass, yellow foxtail seed and Chinese pennisetum seed by FT-IR with discrete wavelet feature extraction and SVM classification.

1. Introduction

Green bristle grass (*Setaria viridis* (L.) Beauv) is a kind of gramineous annual weed which could be easily found anywhere. It is more abundant in the south of the country, occurring less often northward of forest-steppe zone, reaching northern border of agriculture lands. It has erect stem 30 to 80 cm high. It has leaf blade lanceolate strip, smooth on the back, slightly rough in the front, leaf sheath is smooth and has hairs; ligule has 1-2 mm cilia. Panicles presents cylindrical shape with close, usually has a little bend down; spikelet is elliptic, blunt on the top, 2–2.5 mm long, 3–6 clusters and has 1–6 green or variable purple bristles down underneath. It grows on sandy and pebbling shores of reservoirs, on fallow lands and fields, along roads, in trash dump places at settlements. It is drought-resistant and abundant in rich soils. Though it is just a kind of weed, it has medicinal value in traditional Chinese medicine. It has the ability to cure carbuncle, swelling, tinea, and the red eye sores [1]. Yellow foxtail is a clumping annual grass. Young plants can be difficult to distinguish from other grasses like crabgrass. Yellow foxtail produces a characteristic "foxtail"-like seedhead [2]. Chinese pennisetum is erect tufted perennial grass <1 m high. It forms large tussocks, with long hairless wiry leaves. Attractive purplish flower heads bristly, cylindrical spikes, like small bottlebrushes. Yellow foxtail and Chinese pennisetum are the same genus with the green bristle grass [3]. Though the three samples have similar appearance, they have different medicinal value to be used. It is to be distinguished form them by the traditional identification methods.

Phytotaxonomy is the oldest and the most synthetic branch of plant science. The classical plant classification is based on the feature of the plant's exterior looks and the interior dissection and is quite limited and artificial. But as the modern science development and many new interdisciplinary fields are formed, the chemistry and botany are intersected with each other. Utilizing the Fourier transform infrared spectroscopy (FT-IR) combined with the methods of chemometrics, phytotaxonomy finds a new path of development. With this technique, not only classical plant classification becomes much more accurate and more scientific but also the analysis results are enumerated [4–8].

Fourier transform infrared spectroscopy (FT-IR) is a technique used to obtain an infrared spectrum of absorption, emission, photoconductivity, or Raman scattering of a solid, liquid or gas. For an FT-IR spectrometer simultaneously collects spectral data in a wide spectral range, this confers a significant advantage over a dispersive spectrometer which measures intensity over a narrow range of wavelengths at a time. The Fourier transform infrared spectroscopy can provide all the information of compositional system materials. Different monoid of samples have different chemical composition and FT-IR. So FT-IR spectra gradually become a common method to classify different plants [9–11]. But the technique of FT-IR has some insufficient, the hot research object in the instrument analysis is how to use the chemometrics to achieve a fast and accurate complex systems analysis and classification [12–15].

Wavelet transform is a more effective signal processing method than the old Fourier transform and plays an important role in signal analysis and feature extractions. The wavelet transform can provide the information in local time and frequency scales together. It decomposes a signal into localized contributions labeled by a scale and a position parameter; each of the contributions represent the information of different frequency. The main effect of the applications of Wavelet transform of MIR analysis is denoising, data compression, model transfer, and background deduction. Because of its properties including orthogonality, orientation, and flexible time-frequency windows, as an important and effective method of chemometrics, wavelet transform has been widely used in identifying traditional Chinese medicine and plant taxonomy [4, 7–9].

A support vector machine (SVM) is a concept in computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes comprise the input, making the SVM a nonprobabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. With the remarkable excellence of SVM, it is coming into more and more application fields [16, 17].

As the seed of the reproductive organ contains more stable characters than the vegetative organ, in this paper, so sample's seed taken for analysis purposes. This paper focuses on the classification of three sibling plants, the green bristle grass, yellow foxtail, and Chinese pennisetum. The three kinds of sibling plants, their seeds especially, are difficult to be distinguished by the traditional phytotaxonomy. The FT-IR spectra are obtained by FT-IR and horizontal attenuated total reflectance (HATR) technique. The feature vectors which represent spectral characteristics of the FT-IR are extracted using discrete wavelet multiresolution analysis methods. SVM is used to classify the three kinds of sibling plants, the green bristle grass, yellow foxtail, and Chinese pennisetum.

2. Experimental Section

2.1. Materials and Preparation. The seed of green bristle grass is the mature, dry seed of *Setaria viridis* (L.) Beauv (gramineae). The yellow foxtail seed is the mature, dry seed of *Setaria glauca* (L.) Beauv (gramineae). The Chinese pennisetum seed is the mature, dry seed of *Setaria faberii* Herrum (gramineae). The samples were collected from Beishan in Jinhua, Zhejiang Province, China (29°13′N, 119°35′E), Lushan in Jiujiang, Jiangxi Province, China (29°30′N, 115°58′E), and Jinyunshan in Beibei, Chongqing Municipality, China (29°50′N, 106°30′E) in October 2006, and dried in sunlight, respectively. Twenty seed samples are selected randomly for one time per sample. The samples have been grounded to fine powder in agate mortars to about 100 meshes (the particle size is 150 μm), respectively.

2.2. Spectral Measurements. The HATR-FT-IR spectra were collected at a resolution of 2 cm^{-1} scans using a Thermo Electron (Medison, WI, USA) Nexus 670 FT-IR spectrometer with a room temperature deuterated triglycine sulfate (DTGS) detector, and with a single-bounce HATR (Ge) accessory. The spectral range is 4000–650 cm^{-1} and the cumulative number is scan of 128 times. 8.0 mg of predisposed samples were respectively placed directly about 3.14 mm^2 on the center of the Ge crystal of the HATR accessory for measurement. To ensure good contact with the Ge crystal surface, all powder samples were pressed using a pressure tower to provide the same mechanical pressure on all samples. All obtained spectra were auto-baseline corrected. No other sample preparation was required. Each species of all samples were measured three times and the averaged spectrum was used for further analysis.

2.3. Data Analysis. The FT-IR of all the samples were obtained by measuring. According to the absorbance value characteristic of absorption peak, we can make the cluster analysis to the data, which are carried out the Ward clustering algorithm. NEXUS E. S. P. 5.2 software was used for Fourier self-deconvolution (FSD) analysis. The absorption values from different wave bands based on the characters of the absorption value was obtained by copying data method (export the FT-IR data as coordinate value, then copy the data). MatLab 7.0 software was used for discrete wavelet transformation analysis.

The DWT is used to detect the singularity of the curvature curve, so we should choose proper wavelet, which has similar shape to the signal to the analyzed, short compact branch set, and big vanishing moment, as wavelet basis function. Some representative wavelet basis functions include: Mexihat, Meyer, Morlet, Daubechies, Coiflet, and Symlets [14]. Figures 1(a)–1(f) show their function curves in time domain. Compared to other five wavelets, Daubechies wavelet has the shortest compact branch set (Figure 1(d)), so we choose Daubechies wavelet as analyzing wavelet.

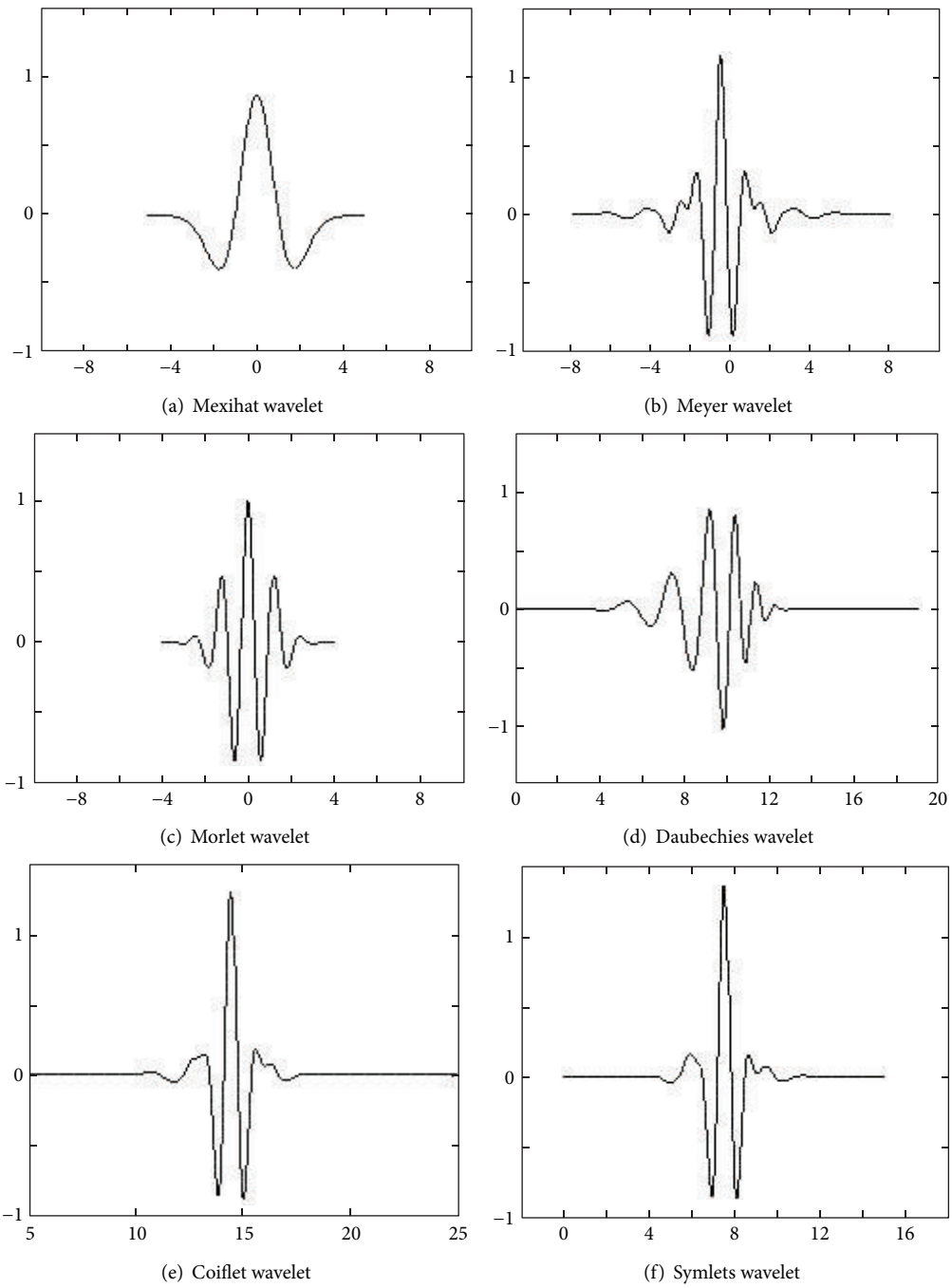


Figure 1: Wavelet basis function curves in time domain.

Daubechies wavelet which possesses better exploration ability for signal singularity acted as analysis wavelet and some scale one-dimension discrete wavelets from different samples were transformed. 5 layers of the samples character variable were picked up by selecting a decomposition scale whose difference degree was the most obvious. Through a comparative analysis, two layers (3 and 4) were selected to extract eigenvector. The number of training samples and testing samples is both 40 each habitat, respectively. The selected character variables were used for SVM training and verification.

3. Results and Discussion

3.1. FT-IR Analysis. Figure 2 shows the typical FT-IR spectra of the seed of green bristle grass, yellow foxtail seed, and Chinese pennisetum seed.

As the seed of green bristle grass, yellow foxtail seed, and Chinese pennisetum seed are similar of absorption peaks in the FT-IR spectra because they belong to the sibling plant seeds. In order to further identify them, Fourier self-deconvolution is used to deal with the FT-IR spectra of the seed of green bristle grass, yellow foxtail seed, and Chinese

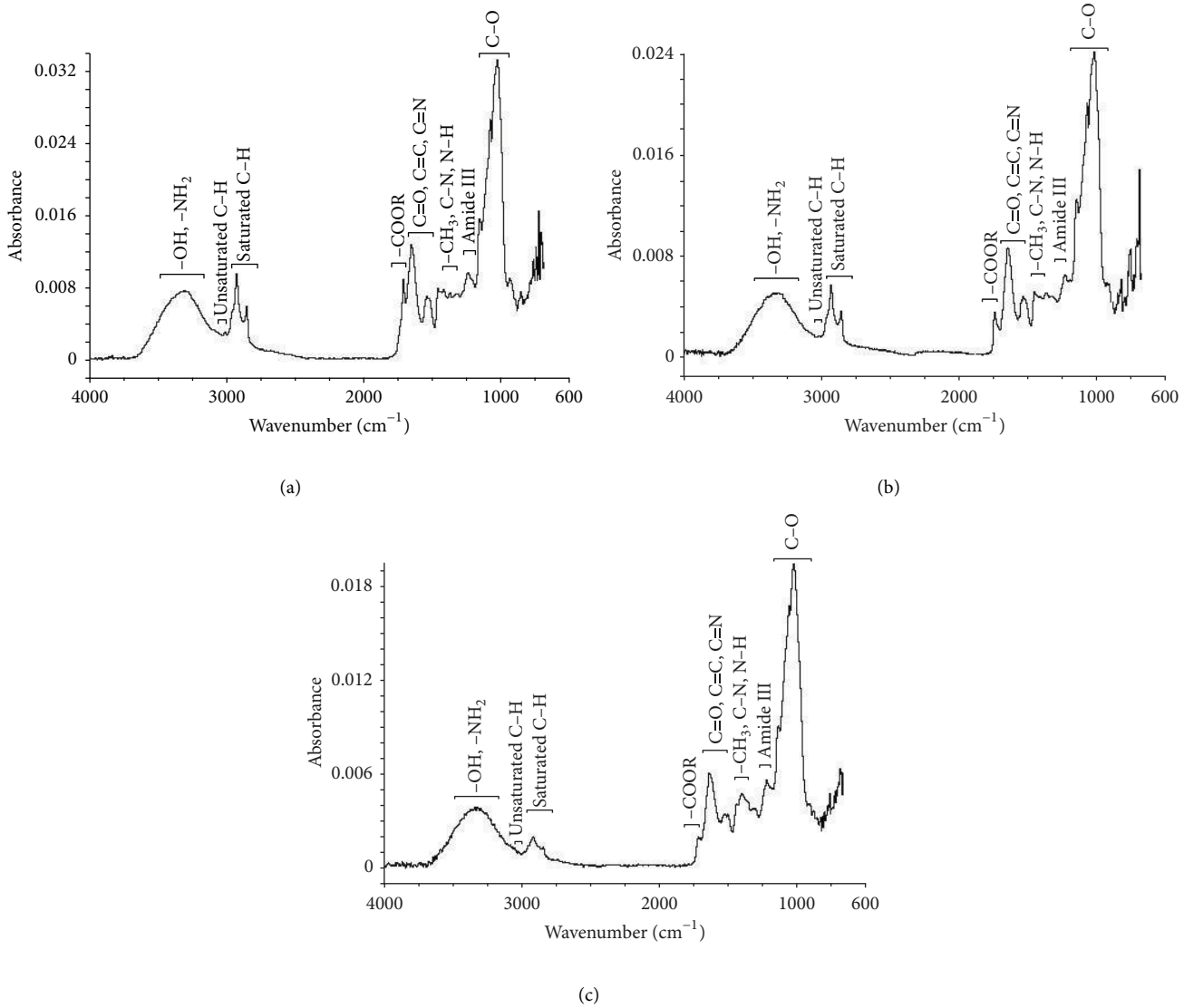


Figure 2: FT-IR spectra of the seed of green bristle grass (a), yellow foxtail seed (b), and Chinese pennisetum seed (c). Characteristic group frequencies are indicated at the top.

pennisetum seed. Their FT-IR-FSD spectra are shown in Figure 3.

From Figures 2 and 3, we notice that the seed of green bristle grass, yellow foxtail seed, and Chinese pennisetum seed generated large numbers of sharp peaks in the FT-IR spectra region (4,000–650 cm^{-1}), which indicates the seeds have a rich chemical composition. Several absorption regions were identified, and the band assignments are labeled in Figure 2. Absorption bands located around 3340 cm^{-1} correspond to O–H and N–H stretching vibrations that mainly occur from carbohydrates and protein mainly. The bands around 3010 cm^{-1} represent unsaturated C–H stretching vibrations that are mainly caused by unsaturated compounds and unsaturated fatty acid ester. The bands around 2924 and 2854 cm^{-1} represent saturated C–H stretching vibrations that are mainly caused by lipid, carbohydrates, and saturated C–H in other compounds. Absorption raised from C–H bending modes was located around 1,200 cm^{-1} to 1,500 cm^{-1}, but overlap with other absorption bands within this region. Three absorption bands located around 1645 cm^{-1} (mainly C=O strt.), 1550 cm^{-1} (N–H bend), and 1,243 cm^{-1} (C–N stret.) were largely due to amide I, II, and III modes of the proteins and lipids, respectively. Absorption bands around 1740 cm^{-1} correspond to isolated carbonyl group (COOR), indicating ester-containing compounds commonly found in membrane lipid and cell wall pectin. Bands around 1030 cm^{-1} and 1151 cm^{-1} in the "fingerprint" region indicate several modes such as C–H bending vibration or C–O or C–C or P–O stretching vibration [18].

The seed of green bristle grass, yellow foxtail seed, and Chinese pennisetum seed contain similar chemical composition like hydroxy of cellulose (seed coat), carbohydrates, protein, phytosterol, flavonoids, alkaloids, and so forth, and their FT-IR absorption are quite similar. The FT-IR spectra

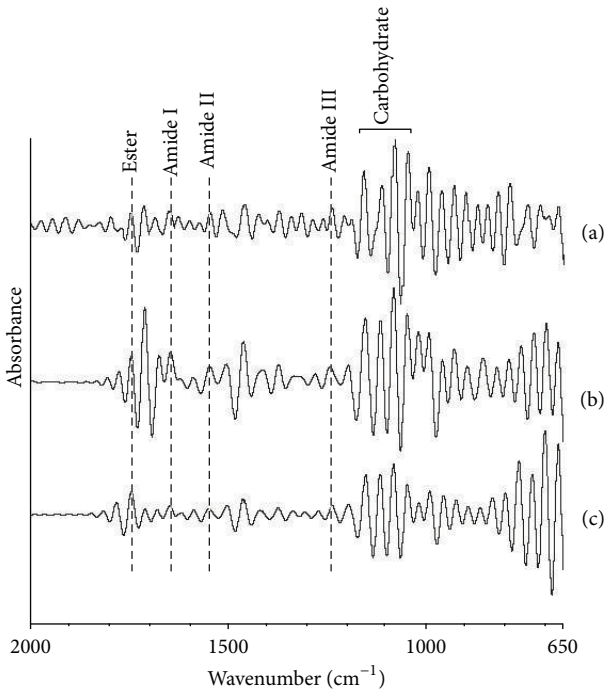


Figure 3: Fourier self-deconvolution FT-IR spectra in the 2000–650 cm^{-1} region in the seed of green bristle grass (a), yellow foxtail seed (b), and Chinese pennisetum seed (c). Characteristic group frequencies are indicated at the top.

and FT-IR-FSD spectra of the three kinds of sibling plant seed from the different plant seeds have very closed absorbance and they are difficult to distinguish by experience. So other methods were used for further classification.

3.2. Chemotaxonomic Distinction with Cluster Analysis. In this paper, 10 samples of each species are randomly selected to make cluster analysis [19, 20]. We have selected the 19 strongest absorption peaks in the range of the 4000 ~ 650 cm^{-1}, then the data of absorption peaks were input to the MatLab 7.0 software for cluster analysis. The dendrogram is shown in Figure 4.

From Figure 4, 30 samples are divided into two families by the dendrogram: cluster 1 (C1) contain two subclusters having twenty samples; cluster 2 (C2) comprises seven samples in the seed of green bristle grass, one sample in yellow foxtail seed and two samples in Chinese pennisetum seed. C1 can be seen to contain two subcluster 1a (SC1a) comprises the six samples in Chinese pennisetum seed, three samples in yellow foxtail seed and one sample in the seed of green bristle grass, and 1b (SC1b) comprises the seven samples in the seed of green bristle grass, two samples in Chinese pennisetum seed and one sample in yellow foxtail. The seed of green bristle grass, yellow foxtail seed, and Chinese pennisetum seed contain similar chemical composition which mainly are lipids, amino, nucleic acid, and carbohydrate, because they belong to the sibling plant seeds. The clustering result does not reflect clearly the real relationship of the thirty samples of the seed of green bristle grass, yellow foxtail seed, and Chinese pennisetum seed in the relatives, and it is not agreement with our expectations, thus, this method is not satisfactory. In order to achieve our desired results, the application of discrete wavelet analysis and SVM are introduced into the study.

3.3. Discrete Wavelet Transformation Analysis. In numerical analysis and functional analysis, DWT is a wavelet transform that the wavelets are discretely sampled. As with other wavelet transforms, a key advantage has over Fourier transforms is temporal resolution: it captures both frequency and location information. Based on this advantage, DWT has a huge number of applications in science, engineering, mathematics, and computer science. Most notably, it is used for signal coding to represent a discrete signal in a more redundant form, often as a preconditioning for data compression. DWT is originated from the discretization of continuous wavelet transform (CWT) and the common discretization is dyadic.

We define a square integral function $\psi(t)$ (namely, $\psi(t)P^2 \in L(t)$) as a family of functions, which satisfies the following equation:

$$\int_{-\infty}^{\infty} \frac{|\widehat{\psi}(w)|^2}{|w|} dw < \infty. \quad (1)$$

Assuming:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad a, b \in R,\ a \neq 0 \quad (2)$$

$\psi_{a,b}(t)$ is defined as a continuous wavelet, which is derived from a family function $\psi(t)$, a and b are the dilation and translation parameters, respectively. a, b represent the family of wavelets obtained from the single function by dilations and translations. The change of parameter would not only influence the frequency spectrum structure of the continuous wavelet but also the window size and the form. Define a continuous wavelet transform as follows:

$$w_f(a,b) = \left\langle f, \psi_{a,b} \right\rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\, \overline{\psi\left(\frac{t-b}{a}\right)} dt. \quad (3)$$

Define discrete wavelet transform as follows:

$$c_{j,k} = \int_{-\infty}^{\infty} f(t)\, \overline{\psi_{j,k}(t)} dt, \quad j, k \in Z, \quad (4)$$

where

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-2^j k}{2^j}\right). \quad (5)$$

The wavelet coefficient $c_{j,k}$ is taken as the time-frequency map of the original signal $f(t)$.

In terms of the relationship between the wavelet function $c(t)$ and the scaling function $f(t)$, namely:

$$\left|\widehat{\phi}(w)\right|^2 = \sum_{j=-\infty}^{\infty} \left|\widehat{\psi}\left(2^j w\right)\right|. \quad (6)$$

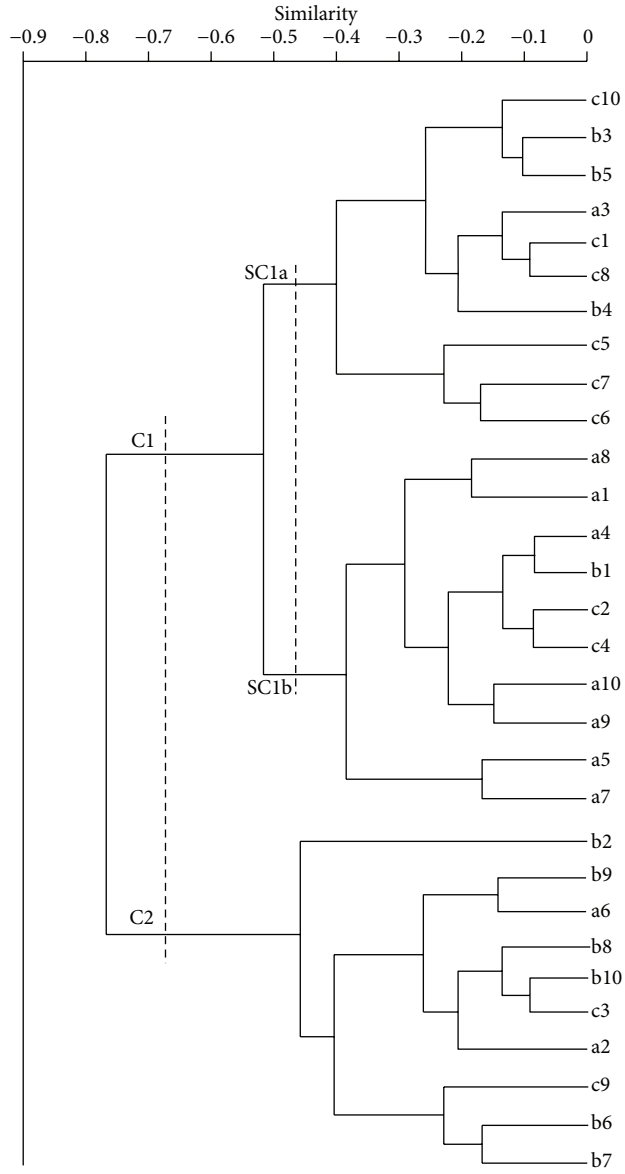


Figure 4: Dendrogram obtained by hierarchical cluster analysis of 30 samples based on FT-IR spectra. The seed samples of green bristle grass: a1–a10; yellow foxtail seed samples: b1–b10; Chinese pennisetum seed samples: c1–c10.

Discrete scaling function with corresponds to the discrete wavelet function is as follows:

$$\phi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{t - 2^j k}{2^j}\right). \tag{7}$$

It is used to discrete the signal, the sampled values are called scaling coefficients $d_{j,k}$:

$$d_{j,k} = \int_{-\infty}^{\infty} f(x)\,\overline{\phi_{j,k}(t)}dt. \tag{8}$$

When $j \succ 1$, the scaling coefficients and the wavelet coefficients are written as follows:

$$\begin{aligned} d_{j+1,k} &= \sum_{-\infty}^{\infty} h(i - 2k)\, d_{j,k}, \\ c_{j+1,k} &= \sum_{-\infty}^{\infty} g(i - 2k)\, d_{j,k}, \end{aligned} \tag{9}$$

where the terms g and h are high-pass and low-pass filters derived from the $\psi(t)$ and $\phi(t)$, the coefficients $d_{j+1,k}$ and $c_{j+1,k}$ represent a decomposition of the $(j - 1)$st scaling coefficient into high and low frequency terms [21–23].

The proper wavelet base and detail were decided by analyzing the signal spectra property and the comparison of decomposition results with different wavelet bases and

details. Daubechies wavelet was selected as the analysis wavelet. The wavelet decomposition scale was set as 5.

Figure 5 are shown preprocessed FT-IR spectra of the seed of green bristle grass, yellow foxtail seed and Chinese pennisetum seed and their DWT coefficients, where, $d1$–$d5$ indicates detail information after compression data.

From Figure 5, we can see that the detailed signal d1 and d2 are a high frequency signal while d3, d4, and d5 are more sensitive to the changes of the spectra. When the DWT was used to decompose the wavelet, the data were compressed. So the d5 is less data volume to identification. The d3 and d4 had been used as feature vector spaces. Figure 6 shows a diagram of a divided feature space. Six feature bands were selected from d3 and d4 details. The criteria for choosing features are the stronger the better. The feature vectors were defined as the energy (the sum of wavelet coefficient square) at each feature band. Thus, six feature variances were generated from two detailed signals.

3.4. SVM. This classification procedure is based on the statistical learning theory proposed by Leopold and Kindermann [24]. The SVM uses structural risk minimization, rather than a nonconvex, unconstrained minimization problem, as in standard neural network training technique using empirical risk minimization. Assume that the training data with k number of samples is represented by $\{x_i, y_i\}$, $i = 1, 2, \ldots, k$, where $x \in R^n$ is an n dimensional vector and $y \in \{-1, +1\}$ is the class label. These training patterns are said to be linearly separable if a vector ω and a scalar β can be defined so that inequalities (10) are satisfied as follows:

$$\begin{aligned} \omega \cdot x_i + \beta \geq +1, \quad \forall y = +1, \\ \omega \cdot x_i + \beta \geq -1, \quad \forall y = -1. \end{aligned} \tag{10}$$

The aim is to find a hyperplane that divides the data so that all the points with the same label are on the same side of the hyperplane. These amounts to finding ω and β such that:

$$y_i \left(\omega \cdot x_i + \beta\right) > 0. \tag{11}$$

If a hyperplane exists that satisfies (11), the two classes are said to be linearly separable. In this case, it is always possible to rescale ω and β so that $\min_{1 \leq i \leq k} yi(\omega \cdot xi + \beta) \geq 1$. That is, the distance from the closest point to the hyperplane is $1/\|\omega\|$. Then (11) can be written as

$$yi \left(\omega \cdot xi + \beta\right) \geq 1. \tag{12}$$

The hyperplane for which the distance to the closest point is maximal is called the optimal separating hyperplane (OSH). As the distance to the closest point is $1/\|\omega\|$, the OSH can be found by minimizing $\|\omega\|^2$ under constraint (12). The minimization procedure uses Lagrange multipliers and quadratic programming (QP) optimization methods. If $\alpha_i \geq 0$, $i = 1, \ldots, k$ are the non-negative Lagrange multipliers associated with constraint (6), the optimization problem becomes one of maximizing:

$$L\left(\alpha\right) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \left(x_i x_j\right). \tag{13}$$

Under constrains: $\alpha_i \geq 0$, $i = 1, \ldots, k$. If $\alpha^m = (\alpha_1^m, \ldots, \alpha_k^m)$ is an optimal solution of the maximization problem (13) then the optimal separating hyperplane can be expressed as

$$\omega_m = \sum_i y_i \alpha_i^m x_i. \tag{14}$$

If the data are not linearly separable then a slack variable ξ_i, $i = 1, \ldots, k$ can be introduced with $\xi_i = 0$ such that (18) can be written as (15),

$$y_i \left(\omega \cdot x_i + \beta\right) - 1 + \xi_i \geq 0, \tag{15}$$

and the solution to find a generalized OSH, also called a soft margin hyperplane, can be obtained using the conditions: (16)–(18):

$$\min_{\alpha, \beta, \xi_1, \ldots, \xi_k} \left[\frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^{k} \xi_i\right], \tag{16}$$

$$yi \left(\omega \cdot xi + \beta\right) - 1 + \xi_i \geq 0, \tag{17}$$

$$\xi_i \geq 0, \quad i = 1, \ldots, k. \tag{18}$$

The first term in (16) is the same as in the linearly separable case, and controls the learning capacity, while the second term controls the number of misclassified points. The parameter C is chosen by the user. larger values of C imply the assignment of a higher penalty to errors.

In situations in which it is not possible to have a hyperplane defined by linear equations on the training data, the techniques discussed for linearly separable data can be extended to allow for nonlinear decision surfaces. A technique introduced by Vapnik maps input data into a high dimensional feature space through some nonlinear mapping. The transformation to a higher dimensional space spreads the data out in a way that facilitates the finding of linear hyperplane. After replacing x by its mapping in the feature space $\Phi(x)$, (13) can be written as

$$L\left(\alpha\right) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \left(\Psi\left(x_i\right) \Psi\left(x_j\right)\right). \tag{19}$$

It is convenient to introduce the concept of the kernel function K in order to make the computation easier in feature space, such as (12):

$$K\left(x_i, x_j\right) = \Psi\left(x_i\right) \Psi\left(x_j\right). \tag{20}$$

It is the kernel function that performs the nonlinear mapping. Thus, to solve (13), only the kernel function is computed rather than $\Phi(x)$, which could be computationally expensive. Equation (21) can be used to classification function:

$$Y = \text{sign}\left\{\sum_{i,j} \alpha_i \alpha_j y_i y_j K\left(x_i, x_j\right) + \beta\right\}. \tag{21}$$

In brief, SVM first maps the data which are not linearly separable into a high dimensional feature space. It then classifies the data by the maximal margin hyper-planes [15].

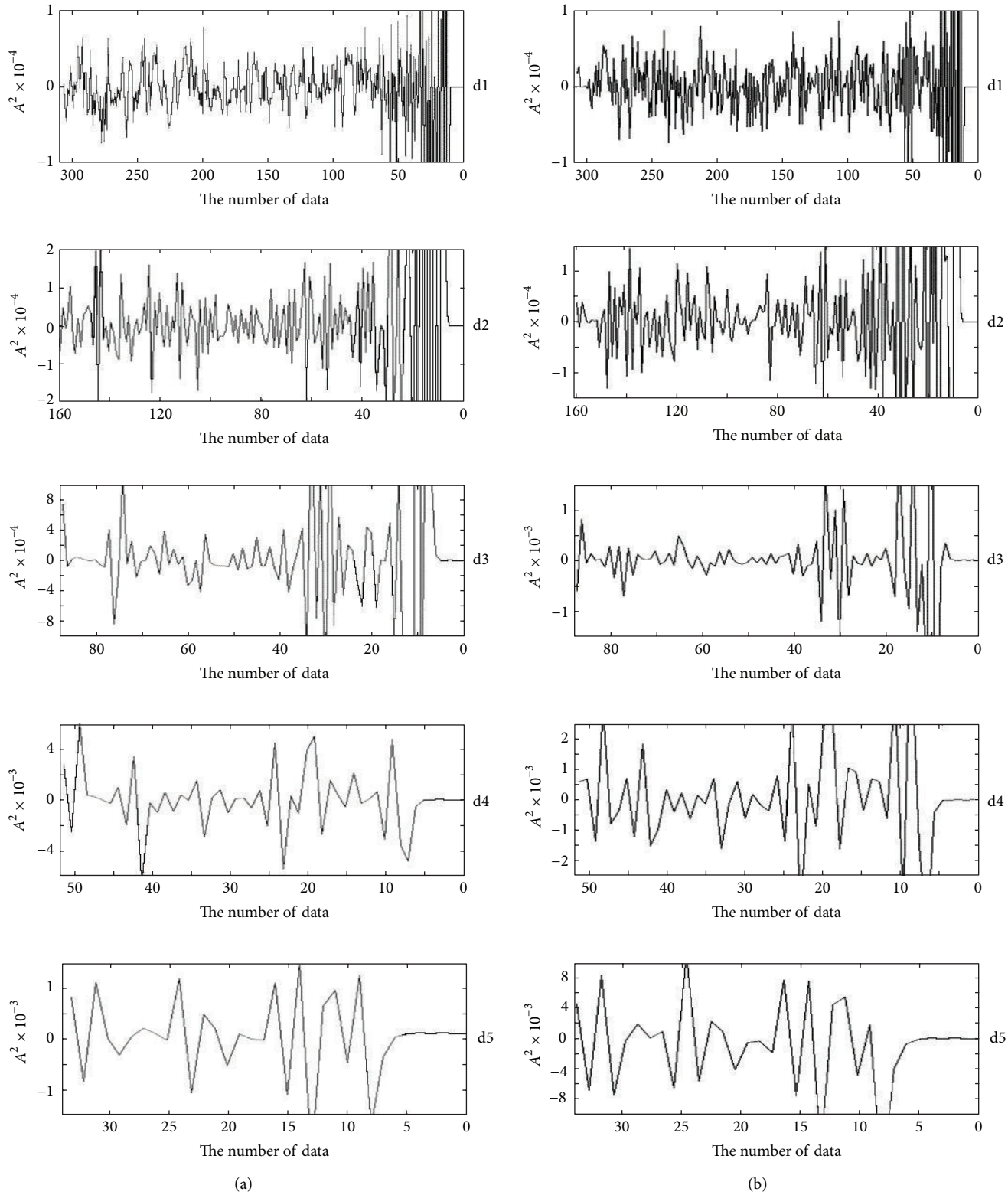
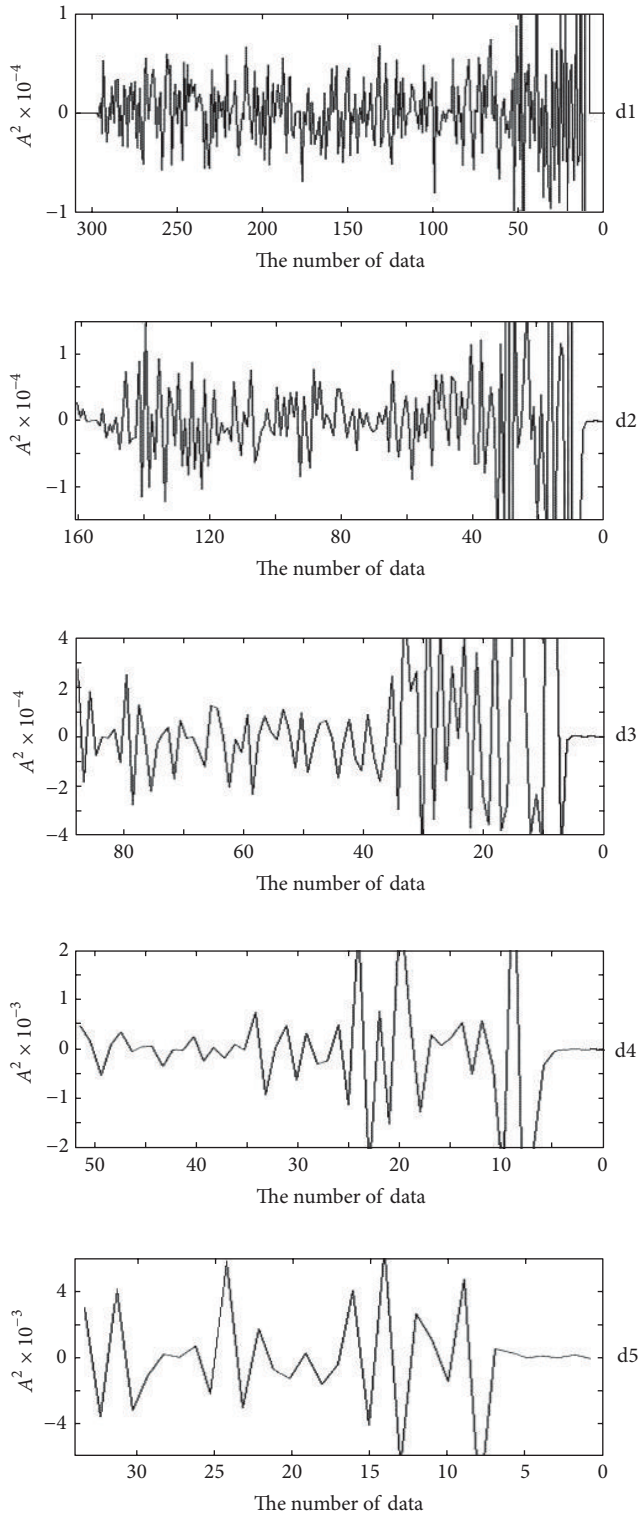


Figure 5: Continued.



(c)

Figure 5: The result of preprocessed spectra with DWT. (a) The seed of green bristle grass; (b) yellow foxtail seed; (c) Chinese pennisetum seed.

Table 1: The identification results of SVMs using polynomial kernel function/%.

	Sample	Beishan	Lushan	Jinyunshan
Train	Seed of green bristle grass	100.0	100.0	100.0
	Yellow foxtail seed	100.0	97.5	100.0
	Chinese pennisetum seed	97.5	100.0	97.5
Test	Seed of green bristle grass	100.0	100.0	100.0
	Yellow foxtail seed	100.0	100.0	95.0
	Chinese pennisetum seed	100.0	100.0	97.5

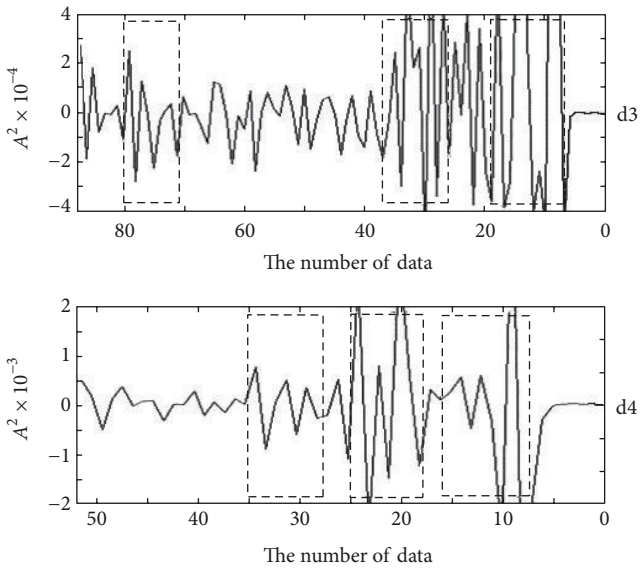


Figure 6: Division of three feature regions of detail signal in the DWT domain.

3.5. Application of the Results. SVM transforms the nonlinear classification problem to a linear problem in a high-dimension feature space by applying a kernel function. Some common kernel functions include polynomial function, radial basis function, and sigmoid function. Comparing with the three common kernel functions, polynomial function was used as kernel function of SVM [17] in this paper. The identification results are shown in Table 1.

The result declared that the method using wavelet feature extraction and SVM classification of sample's FT-IR spectra is an efficient and accurate method for identifying the seed of green bristle grass, yellow foxtail seed, and Chinese pennisetum seed.

4. Conclusion

(1) As the seed of the reproductive organ contains more stable characters than the vegetative organ, using the seed as material to obtain FT-IR spectra as the analysis basis is an effective way in the plant classification.

(2) FT-IR spectra could provide information about molecular structure of main constituents which contained in the seed, which makes it possible to distinguish similar plants samples. Other than pressed disc method and liquid membrane method, HATR-FT-IR method can directly measure the seed of green bristle grass, yellow foxtail seed, and Chinese pennisetum seed of the sibling plants to obtain the spectra and make the proceeding of FT-IR better comparability.

(3) Discrete wavelet transform is used to extract the features of the sibling plants whose infrared absorption is similar in FT-IR spectra. The selected eigenvectors in the detail 3 and 4 resolution have successfully classified the sibling plant seeds. By applying the discrete wavelet feature extraction of the FT-IR data and SVM classification, better results can be achieved usually. It provides more scientific basis for further study in plant taxonomy of the integration of spectroscopy and computer science technologies.

Acknowledgments

The authors would like to thank Dr. Hongfei Lu and Dr. Jianhua Chen from Department of Biology, Zhejiang Normal University, China for identifying of the plant samples.

References

[1] New Medical College of Jiangsu, *A Dictionary of Traditional Chinese Medicine*, Shanghai Science and Technology Publishers, Shanghai, China, 2nd edition, 2006.

[2] Y. C. Liu, E. B. K. Lin, and Y. S. Ma, "Report on species introduction trial of perennial forage in Guoluo area, Qinghai province," *Journal of Sichuan Grassland*, vol. 1, pp. 16–21, 2002.

[3] H. T. Hong, L. Yang, C. Mingxin, C. Hua, and Z. Heshan, "Preliminary studies on the application of Chinese pennisetum as ornamental grass in landscape," *Chinese Agricultural Science Bulletin*, vol. 28, no. 1, pp. 307–310, 2012.

[4] C. G. Cheng, J. Liu, C. J. Zhang, M. Z. Cai, H. Wang, and W. Xiong, "An overview of infrared spectroscopy based on continuous wavelet transform combined with machine learning algorithms: application to chinese medicines, plant classification, and cancer diagnosis," *Applied Spectroscopy Reviews*, vol. 45, no. 2, pp. 148–164, 2010.

[5] C. G. Cheng, J. Liu, H. Wang, and W. Xiong, "Infrared spectroscopic studies of chinese medicines," *Applied Spectroscopy Reviews*, vol. 45, no. 3, pp. 165–178, 2010.

[6] T. Hu, W. Y. Jin, and C. G. Cheng, "Classification of five kinds of moss plants with the use of Fourier transform infrared spectroscopy and chemometrics," *Spectroscopy*, vol. 25, no. 6, pp. 271–285, 2011.

[7] Q. H. Hong, R. P. Yu, R. W. Zheng, H. Wang, and C. G. Cheng, "Identification between *Fimbristylis miliacea* (L.) Vahl and *Fimbristhlis stauntonii* Debeaux et Franch. by CWT-FTIR-RBFNN," *Chinese Journal of Chemistry*, vol. 29, no. 5, pp. 1024–1030, 2011.

[8] C. G. Cheng, J. Liu, W. Q. Cao, R. Zheng, H. Wang, and C. J. Zhang, "Classification of two species of Bidens based on discrete stationary wavelet transform extraction of FTIR spectra combined with probability neural network," *Vibrational Spectroscopy*, vol. 54, no. 1, pp. 50–55, 2010.

[9] C. J. Zhang and C. G. Cheng, "Identification between *Stephania tetrandra* S. Moore and *Stephania cepharantha* Hayata by CWT-FTIR-RBFNN," *Spectroscopy*, vol. 22, no. 5, pp. 371–386, 2008.

[10] H. F. Lu, B. Jiang, Z. G. Shen, J. B. Shen, Q. F. Peng, and C. G. Cheng, "Comparative leaf anatomy, FTIR discrimination and biogeographical analysis of Camellia section Tuberculata (Theaceae) with a discussion of its taxonomic treatments," *Plant Systematics and Evolution*, vol. 274, no. 3-4, pp. 223–235, 2008.

[11] S. Dharmaraj, L. H. Gam, S. F. Sulaiman, S. M. Mansor, and Z. Ismail, "The application of pattern recognition techniques in metabolite fingerprinting of six different *Phyllanthus* spp.," *Spectroscopy*, vol. 26, no. 1, pp. 69–78, 2011.

[12] G. Yoon and H. Kim, "Tissue classification in microscope imaging based on a discrete-band spectroscopy," *Spectroscopy*, vol. 26, no. 1, pp. 33–41, 2011.

[13] S. J. Xu, Y. P. Xie, and C. X. Xu, "Routine classification of food pathogens *Staphylococcus* and *Salmonella* by ATR–FT-IR spectroscopy," *Spectroscopy*, vol. 26, no. 1, pp. 53–58, 2011.

[14] C. G. Cheng, P. Yu, C. S. Wu, and J. N. Shou, "Recognition of MIR data of semen armeniacae amarum and semen persicae using discrete wavelet transformation and BP-artificial neural network," *Spectroscopy*, vol. 27, no. 4, pp. 253–264, 2012.

[15] L. F. Leopold, N. Leopold, H. A. Diehl, and C. Socaciu, "Quantification of carbohydrates in fruit juices using FTIR spectroscopy and multivariate analysis," *Spectroscopy*, vol. 26, no. 2, pp. 93–104, 2011.

[16] C. G. Cheng, Y. M. Tian, and W. Y. Jin, "Study on the methods of wavelet feature extraction and SVM classification of FTIR lung cancer data," *Acta Chimica Sinica*, vol. 65, no. 22, pp. 2539–2543, 2007.

[17] C. G. Cheng, Y. M. Tian, and W. Y. Jin, "A study on the early detection of colon cancer using the methods of wavelet feature extraction and SVM classifications of FTIR," *Spectroscopy*, vol. 22, no. 5, pp. 397–404, 2008.

[18] Z. Movasaghi, S. Rehman, and I. U. Rehman, "Fourier transform infrared (FTIR) spectroscopy of biological tissues," *Applied Spectroscopy Reviews*, vol. 43, no. 2, pp. 134–179, 2008.

[19] A. Carteron, M. Jeanmougin, F. Leprieur, and S. Spatharis, "Assessing the efficiency of clustering algorithms and goodness-of-fit measures using phytoplankton field data," *Ecological Informatics*, vol. 9, pp. 64–68, 2012.

[20] B. Delibašić, M. Vukićević, M. Jovanović, K. Kirchner, J. Ruhland, and M. Suknović, "An architecture for component-based design of representative-based clustering algorithms," *Data & Knowledge Engineering*, vol. 75, pp. 78–98, 2012.

[21] F. Ehrentreich, "Wavelet transform applications in analytical chemistry," *Analytical and Bioanalytical Chemistry*, vol. 372, no. 1, pp. 115–121, 2002.

[22] X. L. Li, S. Dong, and Z. J. Yuan, "Discrete wavelet transform for tool breakage monitoring," *International Journal of Machine Tools and Manufacture*, vol. 39, no. 12, pp. 1935–1944, 1999.

[23] S. B. Pan and R. H. Park, "Systolic array architectures for computation of the discrete wavelet transform," *Journal of Visual Communication and Image Representation*, vol. 14, no. 3, pp. 217–231, 2003.

[24] E. Leopold and J. Kindermann, "Text categorization with support vector machines: how to represent texts in input space?" *Machine Learning*, vol. 46, no. 1–3, pp. 423–444, 2002.