Support vector machines combined with wavelet-based feature extraction for identification of drugs hidden in anthropomorphic phantom

Dingjun Qu [a,b], Wei Li [a,b], Yu Zhang [b,c], Bai Sun [a], Yu Zhong [a], Jinhuai Liu [a], Daoyang Yu [a,*], Minqiang Li [a,*]

[a] Research Center for Biomimetic Functional Materials and Sensing Devices, Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, PR China
[b] Departments of Automation, University of Science and Technology of China, Hefei 230026, PR China
[c] Anhui Provincial Laboratory of Biomimetic Sensor and Detecting Technology, West Anhui University, Lu'an 237012, PR China

ARTICLE INFO

ABSTRACT

A new recognition method of Support Vector Machines (SVMs) combined with wavelet-based feature extraction is proposed for identifying drugs hidden in human body. Preliminary data sets of eight kinds of samples are acquired by a home-built instrument using energy-dispersive X-ray diffraction (EDXRD) technology in a short detection time. Small sample size, poor signal-to-noise ratio (SNR) and high dimension of data make drugs identification a challenging problem. In this paper, the potential effective method solves the problem well. The spectral signal with poor SNR is obtained and processed with wavelet for feature extraction and then the wavelet coefficients are used as the inputs of SVMs. A multi-classifier of SVMs based on binary tree architecture (SVMs-BAT) is trained. The method of SVMs-BAT combined with wavelet-based feature extraction (WSVMs-BAT) is firstly compared with two methods: one is single SVMs-BAT which uses original data as inputs without preprocessing, the other is SVMs-BAT combined with feature extraction based on principal component analysis (PCA-SVMs-BAT). The high identification accuracy of WSVMs-BAT indicates that the method of feature extraction using wavelet can effectively represent the original data better. Then the recognition result of the proposed method is also compared with artificial neural network (ANN) and K-nearest neighbor (KNN) methods. Our findings show that the proposed method combined with EDXRD technology provides a good access to achieve the aim of automatic identification of illicit drugs.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Illegal drug detection is of great importance for public safety and security. Generally, packaging in human body is considered as a primary approach for smuggling drugs. In the past decades, detection of drugs and explosives covered by skin and clothing has been focused on extensively. In contrast to some other techniques, the energy-dispersive X-ray diffraction (EDXRD) technology has proved to be a suitable method due to its non-destruction, non-invasion, low cost and high resolution for detecting drugs hidden in human body [1–4].

Both of the peak and profile of EDXRD spectrum of each material are unique according to Braggs' law [5]. However, under the complex measurement conditions of packing and short detection time, the high-dimensional EDXRD spectrum with poor signal-to-noise ratio (SNR) (1024-dimension) is hard to analysis. Especially, small sample size makes the identification towards them to be a great challenge. In our investigation, the method for extracting features from EDXRD spectra and design an intelligent classifier is proposed and demonstrated in this paper.

* Corresponding authors. Address: Research Center for Biomimetic Functional Materials and Sensing Devices, Institute of Intelligent Machines, No. 350, Shushanhu Road, Hefei, Anhui, PR China. Tel.: +86 551 5591142; fax: +86 551 5592420.

E-mail addresses: ydy@iim.ac.cn (D. Yu), mqli@iim.ac.cn (M. Li).

Principal component analysis (PCA), which linearly transforms the original signals into new uncorrelated features, has been a well-known method for feature extraction. However, the principal components (PCs) always carry information on both signal and disturbance so that the problem how many principal components should be chosen remains a significant question. Excessive PCs will contain too much noise, while insufficient PCs may miss some important features [6]. As we know, another method called wavelet is also used in data preprocessing frequently. The wavelet coefficients can capture time and frequency localized information of original data by decomposing a timeseries into time/frequency space, which is different from other approaches for feature extraction [7,8]. The method of wavelet-based feature extraction has proved to be robust to the added noise and distortions. Therefore, in contrast to PCA, wavelet is a more ideal tool to be applied to denoising tasks and as feature extraction for the signal with poor SNR. In order to extract the primary features from poor EDXRD spectra and avoid influence of unstable factors, wavelet is applied for its robustness in this paper.

Besides, classification is another important task for intelligent identification system design, which has been rarely studied for drug EDXRD spectra. A novel classification method based on SVM is adopted for its attractive properties [9,10]. Compared with traditional method such as artificial nervous network (ANN), SVM based on the idea of structural risk minimization (SRM) overcomes the problems of failing into local minimum and lacking of rules of designing network structure and does well in the case of small sample size and high dimension. An SVM maps the input points into a high-dimensional feature space and finds a separating hyperplane that maximizes the margin between two classes in the high-dimensional space, which is suitable to the case of small number of sample specially. Consequently, SVM has been widely used because of its high accuracy and good generalization capabilities, such as face detection, handwriting recognition, chemical pattern classification, and fault diagnosis [11–14].

The SVM has been mentioned to identify airline baggage for two class problem [15]. However, a multi-classifier of SVMs is necessary for identifying various kinds of material. As we know, the common multi-class classifiers of SVMs are constructed by several binary classifiers [16]. In this paper, a multi-classifier, hierarchical support vector machines with binary tree architecture (SVMs-BTA), is chosen due to its efficient computation of the tree architecture and high classification accuracy of SVMs comparing with other approaches. For an N-class problem, $(N-1)$ SVMs are required to be trained but only $\log_2 N$ SVMs are needed to test for the classification decision [17].

The goal of this paper is to develop a method of SVMs-BAT combined with wavelet-based feature extraction (WSVMs-BAT) to identify drug spectra with poor SNR. All computation was performed in MATLAB®. The structure of the paper is organized as follows. Section 2 describes the experimental system based on EDXRD technology in detail. Section 3 presents a brief overview of wavelet decomposition and wavelet-based feature extraction. Section 4 provides the necessary background and relevant concepts of SVM and introduces multi-class identification algorithm of SVM briefly. Then Sections 5 and 6 give the analysis and conclusions of our work.

2. EDXRD experimental system

Compared with the conventional XRD, EDXRD utilizes the polychromatic X-ray as the light source and much higher X-ray energies. Besides, EDXRD is more convenient than traditional XRD, which uses photon energy E to replace wavelength λ [18,19]. The formula is given as:

$$2d \sin \frac{\theta}{2} = \frac{nhc}{E} \tag{1}$$

where h is Planck's constant, c is velocity of light, d is the spacing between diffraction centres, and θ is the angle through which the photon is diffracted. Obviously a primary X-ray photon with high energy is diffracted under a small diffraction angle θ for a given lattice spacing d. Fixing the diffraction angle, different lattice spacing characterizing the crystalline structure of the material result in different diffraction peaks within the energy spectrum of the diffracted X-ray photons. According to the Bragg-condition, the spectrum will be shifted along the energy axis if the diffraction angle is changed.

The home-built instrument which consists of an EDXRD spectrometer shown in Fig. 1 and mechanical control system is used in the experiments. The spectrometer contains an X-ray source, optical system, detection systems. Tungsten tube is chosen to produce continuous X-ray radiation in the range 75–100 kV and 3–5 mA. Two vertical-slits collimators with 0.3 mm in width and 15 mm in height are placed vertically to adjust diffraction beams on mechanical support system. The detection system includes low-energy high purity germanium detector (Canberra, GL0055PS) and multi-channel analyzer InSpector 2000 (1024 channels) whose detection range and effective area of detection window are 0.3–300 keV and 50 mm² respectively. The mechanical control system includes rotating platform for the installation of detectors and slit sola framework and aligning samples, sola slit and detector. Stepper motor driver platform is controlled by compute. The diffraction angles are obtained by rotating around the sample center. Angular displacement accuracy is 0.01 and rotation range is ±45°. By means of the Bragg's law (1), the peak position E within the
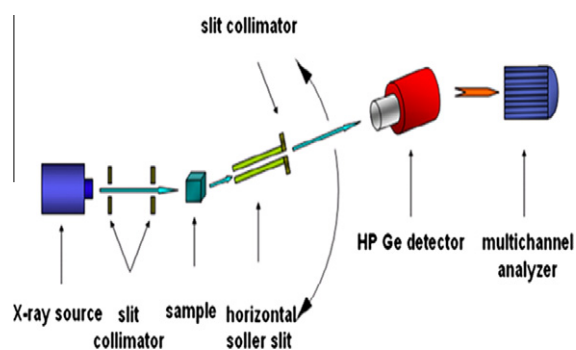

Fig. 1. Geometry of the diffraction system.

normalized spectrum can be used to calculate the corresponding lattice spacing d by comparison with literature data (Powder Diffraction File). Through comparison with the measured peak position, it can be confirmed weather the home-built instrument is stable. NaCl is used to test the system and the obtained spectra by the instrument suggest that the total measurement system is reliable and can be used to detect different kinds of materials.

3. Wavelet decomposition based on minimum-entropy

The theory of wavelet was firstly introduced by Groupillaud et al. in 1984, and it was developed and applied as a tool for data processing in many fields successfully in the next several decades. The continuous wavelet transform (CWT) of $x(t) \in L^2(R)$ is a time-domain method of signal processing that can be defined as the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet function. The equation is given as:

$$W_\psi(a,b) = \langle \psi_{a,b}(t), x(t) \rangle = |a|^{1/2} \int x(t) \psi_{a,b}^* dt \qquad (2)$$

$W(a, b)$ is the wavelet coefficient, a is scale factor and b is time factor. $\psi(t)$ denotes basic wavelet function or mother wavelet, which satisfies the admissibility condition:

$$X_\psi = \int_{-\infty}^{\infty} |\hat{\psi}(w)|^2 / |\omega| d\omega < \infty \qquad (3)$$

The discrete wavelet transform (DWT) is derived from the discretization of CWT and used widely for reducing computation in real application. Let $a = a_0^j$, $b = kb_0 a_0^j$, $a_0 > 1$, $b_0 \in R$, j, k are integer numbers. In the case $a_0 = 2$, $b_0 = 1$, DWT can be defined as follows:

$$W_\psi(j,k) = 2^{-j/2} \sum_{i=1}^{N-1} x(t) \bar{\psi}(2^{-j}t - k) \qquad (4)$$

On the basis of wavelet theory, an efficient computation method for multiresolution signal decomposition was proposed by Malatt in 1989 [20]. The procedure of multiresolution decomposition of a signal X is schematically illustrated in Fig. 2a. Each stage of this scheme consists of two digital filters and two downsamplers by 2. The first filter, $g[.]$ is the discrete mother wavelet (high-pass filter), and $h[.]$ is its mirror version (low-pass filter). The downsampled outputs of first high-pass and low-pass filter provide the detail coefficients D_1 and the approximation

coefficients A_1. These vectors are obtained by convolving $X[n]$ with the low-pass filter for approximation, and with the high-pass filter for detail, followed by dyadic decimation (downsampling). The approximation coefficients and detail coefficients can be expressed with,

$$A_{j+1}(k) = \sum_m h(m - 2k) A_j(m) \qquad (5)$$

$$D_{j+1}(k) = \sum_m g(m - 2k) A_j(m) \qquad (6)$$

where A_j and D_j are approximation and detail coefficients at j level decomposition, respectively. If the length of each filter is equal to $2N$ (here $N = 5$) and the length of the original signal X is n, the coefficients A_1 and D_1 are of length $\lfloor \frac{n-1}{2} \rfloor + N$. Then the first approximation coefficients A_1 is further decomposed and this process is continued as shown in Fig. 2b.

Because the analysis process is iterative, it can be continued infinitely in theory. However, in application, the decomposition can proceed only until the individual details consist of a single sample by selecting a suitable number of levels based on the nature of the signal, or on a suitable criterion such as entropy. Classical entropy-based criteria describe information-related properties for an accurate representation of a given signal. Entropy is a common concept in many fields, especially in signal processing [21]. Assuming that s is the signal and s_i are the coefficients of s in an orthonormal basis, the (nonnormalized) Shannon entropy can be defined as:

$$E(s) = -\sum_i s_i^2 \log(s_i^2) \qquad (7)$$

with the convention $0\log(0) = 0$.

4. Basic theory of support vector machines

4.1. The definition of support vector machines

SVM is a powerful tool for two-group classification problems. Its foundation developed by Vapnik, embodies the structural risk minimization (SRM) principle which minimizes an upper bound on the expected risk. SVM has been widely used for its high accuracy and good generalization capabilities. It is usually used to describe classification and regression. In this paper, SVM is used for recognition.
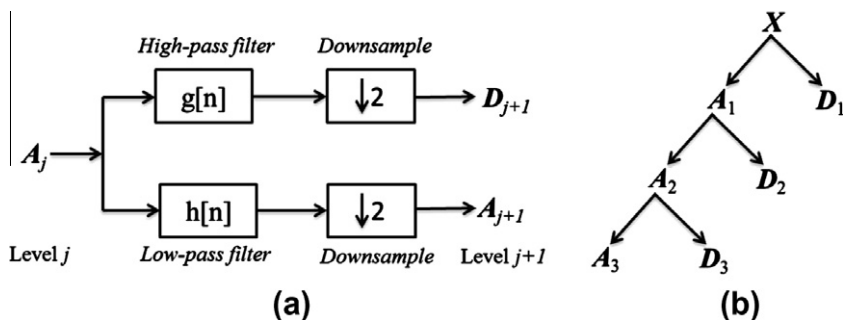

Fig. 2. Process of the signal decomposition. (a) Decomposition detail, the initialization of A_0 is the original signal X. (b) Decomposition tree.

Consider the problem of separating the set of training vectors belonging to two separate classes, $(x_1, y_1), \ldots, (x_N, y_N)$ where input $x_i \in R^N$ (N-dimensional input space) and class labels (target output) $y_i \in \{1, -1\}$. With a hyperplane,

$$\langle w, x \rangle + b = 0 \qquad (8)$$

where $\langle, \rangle$ represents the inner product, w is a weight vector and b is bias. A separating hyperplane in canonical form must satisfy the following constraints,

$$y^i[\langle w, x^i \rangle + b] \geqslant 1, i = 1, 2, \ldots N \qquad (9)$$

To obtain the optimal classifier, $||w||$ should be minimized under the following constraints,

$$y^i[\langle w, x^i \rangle + b] \geqslant 1 - \xi_i, i = 1, 2, \ldots N \qquad (10)$$

The variables ξ_i are positive slack variables, which is necessary to allow misclassification. Hence the solution to the hyperplane that optimally separates the data is the one that minimizes.

$$\Phi(w, \zeta) = \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i \qquad (11)$$

where the parameter C is specified beforehand and controls the trade-off between maximizing the margin the training error. The corresponding quadratic programming (QP) problem can be described as,

$$\text{Maximize } Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j \qquad (12)$$

where α are the Lagrange multipliers, N is the size of sample. With constraints,

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \qquad (13)$$
$$0 \leqslant \alpha_i \leqslant C, \quad \forall i$$

When a linear boundary is inappropriate, the SVM can map the input vector into a high dimensional feature space. Fig. 3 shows that an optimal separating hyperplane is constructed in this higher dimensional space. Kernel function is defined as:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \qquad (14)$$

K is a symmetric positive definite function, which satisfies Mercer's Conditions. The optimization problem can be rewritten as follows

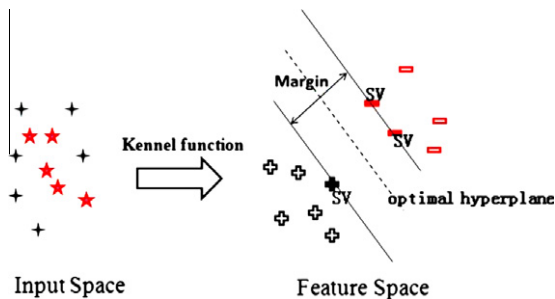**Fig. 3.** Classification of data for SVM.

$$\text{max } W(\alpha) = -\frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \sum_{i=1}^{N} \alpha_i \qquad (15)$$

Then the nonlinear classifier can be expressed as,

$$y = \text{sgn}\left(\sum_{sv} \alpha_i^* y_i K(x_i, x) + b^* \right) \qquad (16)$$

It is well known that SVM is decided by training samples and kernel function. The construction and selection of kernel function is important to SVM. The common kernel functions have linear kernel function, polynomial kernel function and Gaussian radial basis function, etc. Which kernel function selected always depends on experience. In this paper, the Gaussian radial basis function is used. The form is given as:

$$K(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2} \right) \qquad (17)$$

where σ is the kernel width parameter.

4.2. Multi-class identification algorithm of SVM

The SVM is a tool for two-class problems and its application to multi-class problems is not straightforward. The common approaches for SVMs convert multi-class classification problem into many binary-class problems. Therefore, an N-class problem needs $N(N-1)/2$ binary SVMs with the "one-against-one" approach presented by Krebel, while N SVMs for the "one-against-rest" approach presented by Vapnik [22]. SVMs-BTA is a novel method and needs only $(N-1)$ SVMs for an N-class problem. But it only needs to test $\log_2 N$ SVMs for the classification decision [17]. This approach is applied in many cases to improve classification accuracy and reduce computation during pattern recognition.

5. Results and discussion

5.1. Experimental condition selection and data acquisition

A setup based on EDXRD is used to detect illicit materials hidden in human body. During the measuring process, an anthropomorphic phantom is utilized to take place of real human body. The model is composed by bones, muscles and organs whose materials' scattering and absorption of radiation are similar to human tissues. The detected materials are uniformly packed in round plastic tubes for easy comparison. Here the length of tube is 4.0 cm and the radius is about 1.2 cm. Then the packets are hidden in the anthropomorphic phantom. The measuring environment is shown in Fig. 4.

It has been illustrated that the Compton cross section is suppressed at small angles in literature [15]. Taking the complex detection condition of human body into consideration, 3–6° are chosen as suitable angles and the spectra at different angles are shown in the below Fig. 5. It can be found that a good EDXRD spectrum of methamphetamine hidden in human body is achieved at 4° and the number of photons will decrease below 4°. So 4° is chosen as the

Fig. 4. Complex detection environment of human body.

Table 1
Class labels of samples.

Class labels	Samples
a	Background of body
b	Phenylacetic acid + body
c	Methamphetamine + body
d	Heroin + body
e	Heliotropin + body
f	Wheat flour + body
g	TNT + body
h	NaCl + body

best diffraction angle where spectra of materials are distinguished. 7 kinds of materials in total are investigated in the experiments. Chemically pure heroin, methamphetamine, TNT, phenylacetic acid, heliotropin as detection targets and analytical pure NaCl, wheat flour as distraction were measured for forty times under the conditions of the best diffraction angle with detection time of 30 s. According to the Bragg's law, Bragg peaks in the energy spectrum correspond to distances between planes. Both of the peak position and profile of spectrum are characteristics of the crystal structure and can be used to identify the material. Different sample concentrations will lead to different intensities of peak. However, the peak position and profile of spectrum which represent the essential characteristics

of the materials remain the same. The eight kinds of samples are represented by class labels for easy processing in Table1 and the spectra of eight materials are shown in Fig. 6. The horizontal axis represents 1024 channels of multichannel analyzer, equivalent in energy spectrum of 0–100 keV. The vertical axis represents the number of photons in each channel, which is equivalent in energy of photon.

5.2. Optimal wavelet decomposition and feature extraction

Due to mechanical noise and instability of X-ray source, the obtained original spectra shown in Fig. 6 had poor SNR which leads to a huge difficulty for spectra recognition. So extraction of effective features is very important in pattern recognition. In this paper, wavelet decomposition with Daubechies 3 wavelet function is used to extract features of the spectra for its robustness. Methamphetamine whose original signal is shown in Fig. 6C is taken as a typical example. Although the decomposition process can be
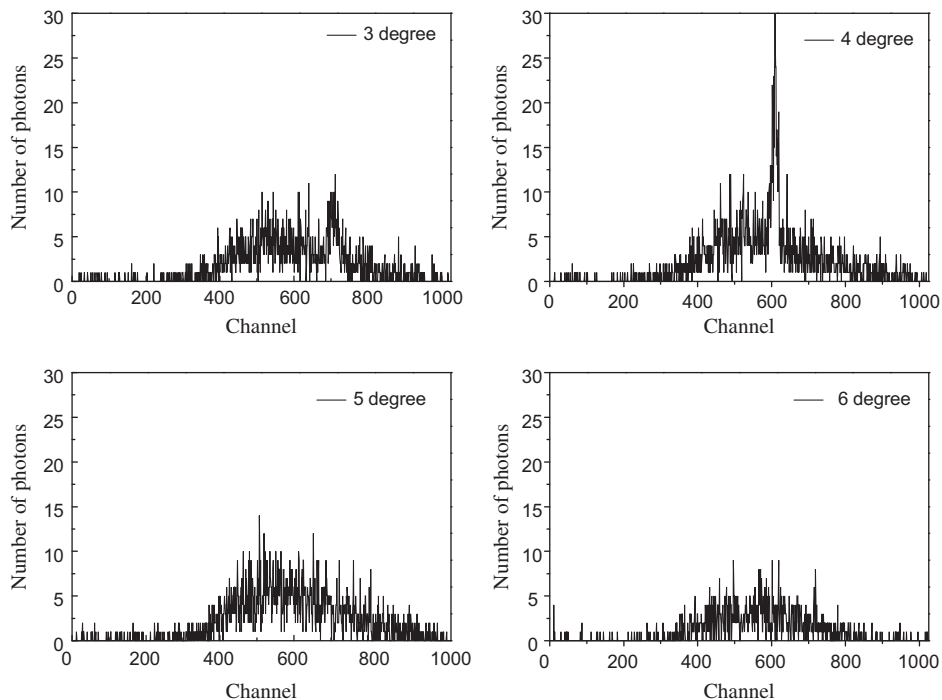

Fig. 5. EDXRD spectra of Methamphetamine hidden in anthropomorphic phantom at different diffraction angles.
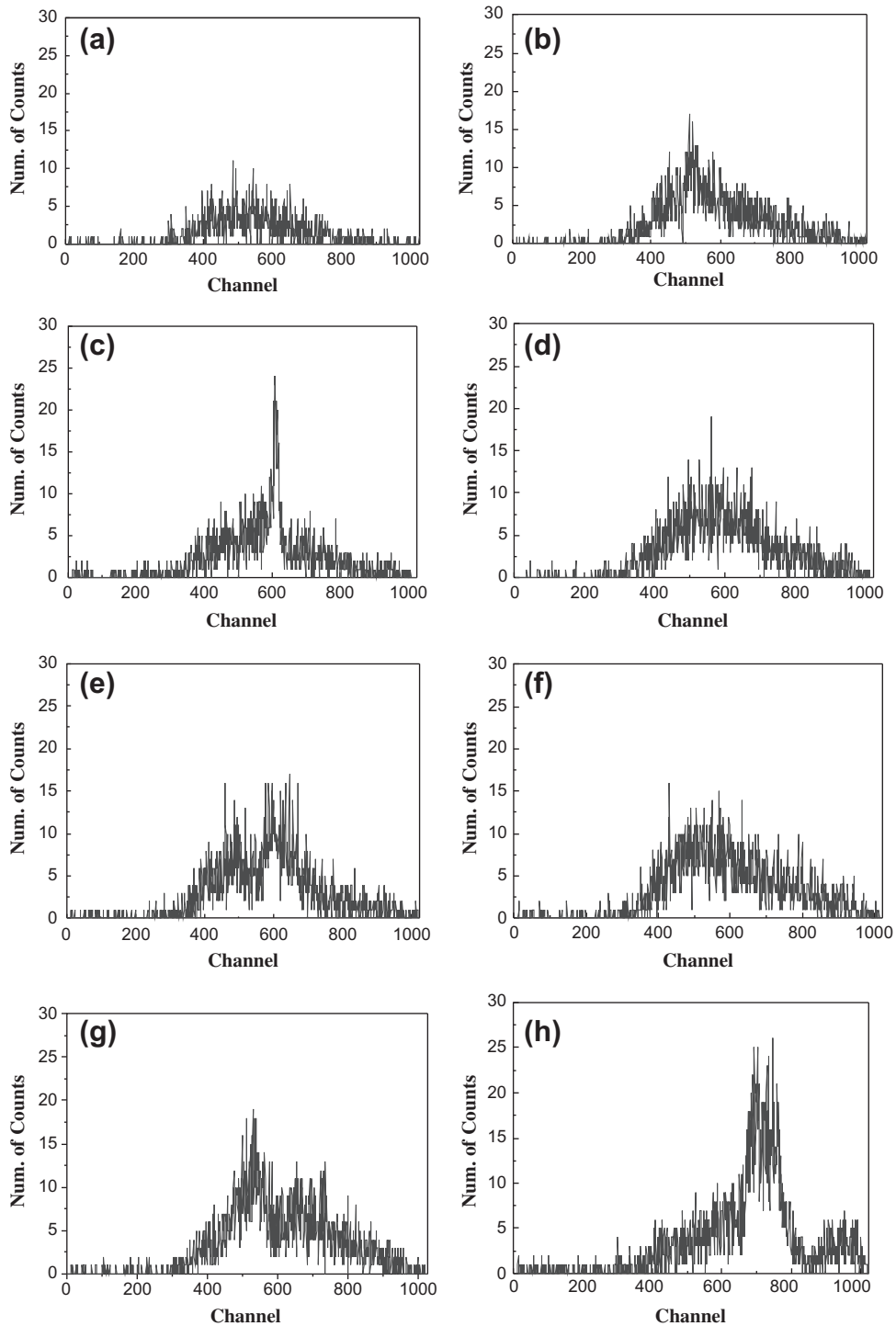
Fig. 6. Original EDXRD spectra of materials hidden in anthropomorphic phantom in30 s.

continued indefinitely in theory, it is always processed by selecting a suitable number of levels based on the nature of the signal. Shannon entropy which has proved to be a useful and efficient tool of wavelet analysis detailed in literature [23], is used here to measure spectra variations

for finding a suitable number of levels. The calculated Shannon entropy at different level is shown in Fig. 7. From Fig. 7, the Shannon entropy minimums at four-level decomposition, which imply that the decomposition is interesting until the fourth time. It is noted that minimum
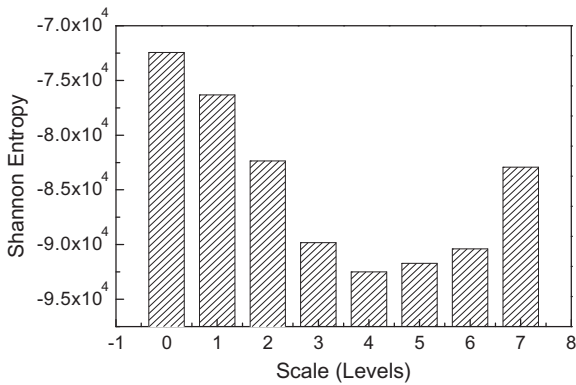
Fig. 7. The wavelet Shannon entropy at different levels.

entropy corresponded to less randomness and thus it leads to clustering. Therefore, the wavelet basis at level 4 is a best basis according to Shannon entropy. Approximation signal and detail signal at four-level decomposition and its reconstruction are shown in Fig. 8. It illustrates that fluctuations in the original signal are decomposed into detail signals which can be regarded as high frequency disturbance (noise). The peak and profile of the original signal are protected well in the approximation signal a4, which is greatly useful for drug recognition. According to above analysis, the primary features of the signal could be represented with information of wavelet approximation

coefficients containing 72 data at level 4 and the results of feature extraction are shown in Fig. 9. Using this method, the rest of samples are processed. The features extracted from EDXRD spectra of eight measured samples form a matrix (80 × 72) which is used as input for training classifier.

5.3. The choice of parameters of SVMs and Multi-classifier for identification

Currently, there is no clear criterion for choosing the kernel of SVM. In general, the RBF kernel is a reasonable choice. In our study, all of the SVMs of the classifier adopt Gaussian radial basis function as their kernel, and the parameter σ of the function is set to 1. Some actions also have been taken for finding best parameter C, and the results show that parameter C affects the classifier accuracy little. For easier operation, the parameter C of all SVMs is set to 10.

Using the feature vector based on wavelet approximation coefficients as inputs, a multi-class classifier based on SVM-BTA according to the structure shown in Fig. 10, is trained for identification. In this paper, eight kinds of samples {a, b, c, d, e, f, g, h} in total are used to train the multi-classifier. Firstly, all samples are divided into two classes {a, b, c, d} and {e, f, g, h} by SVM1. Secondly, SVM2 and SVM3 put {a, b, c, d}, {e, f, g, h} into {a, b}, {c, d} and {e, f}, {g, h} respectively. According to the above method mentioned, classification continues until all the
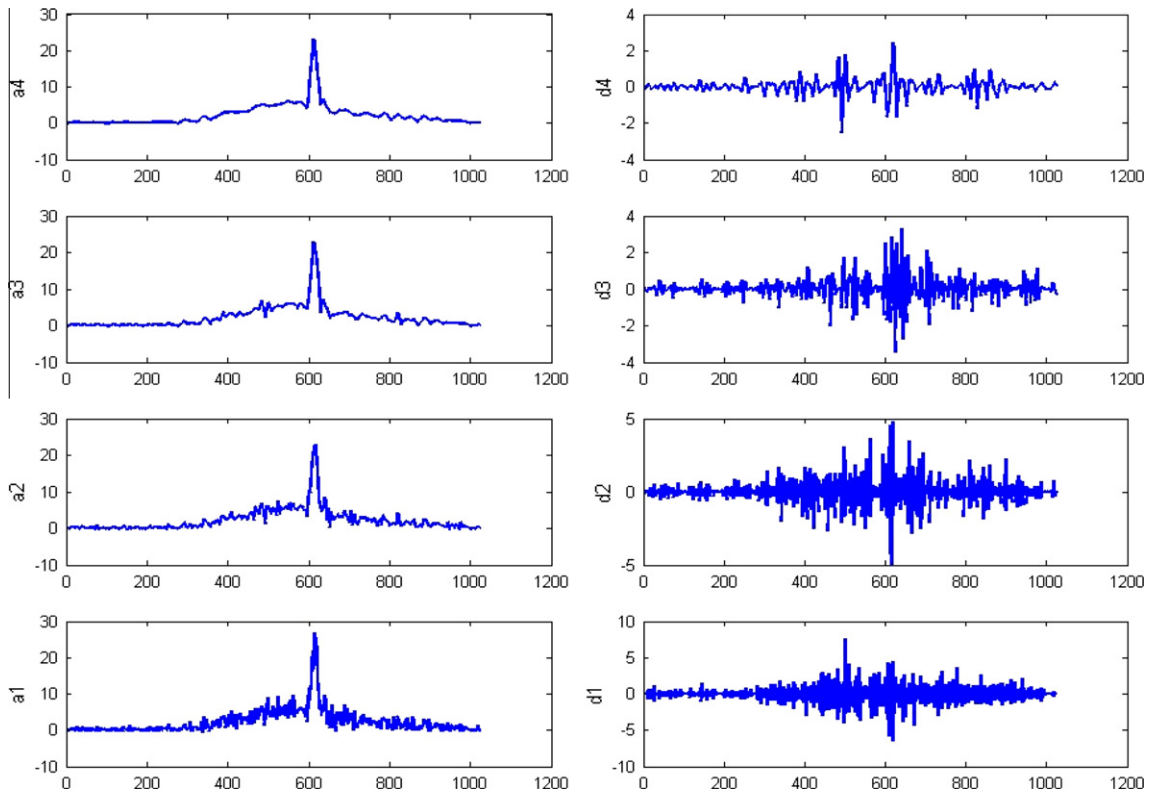

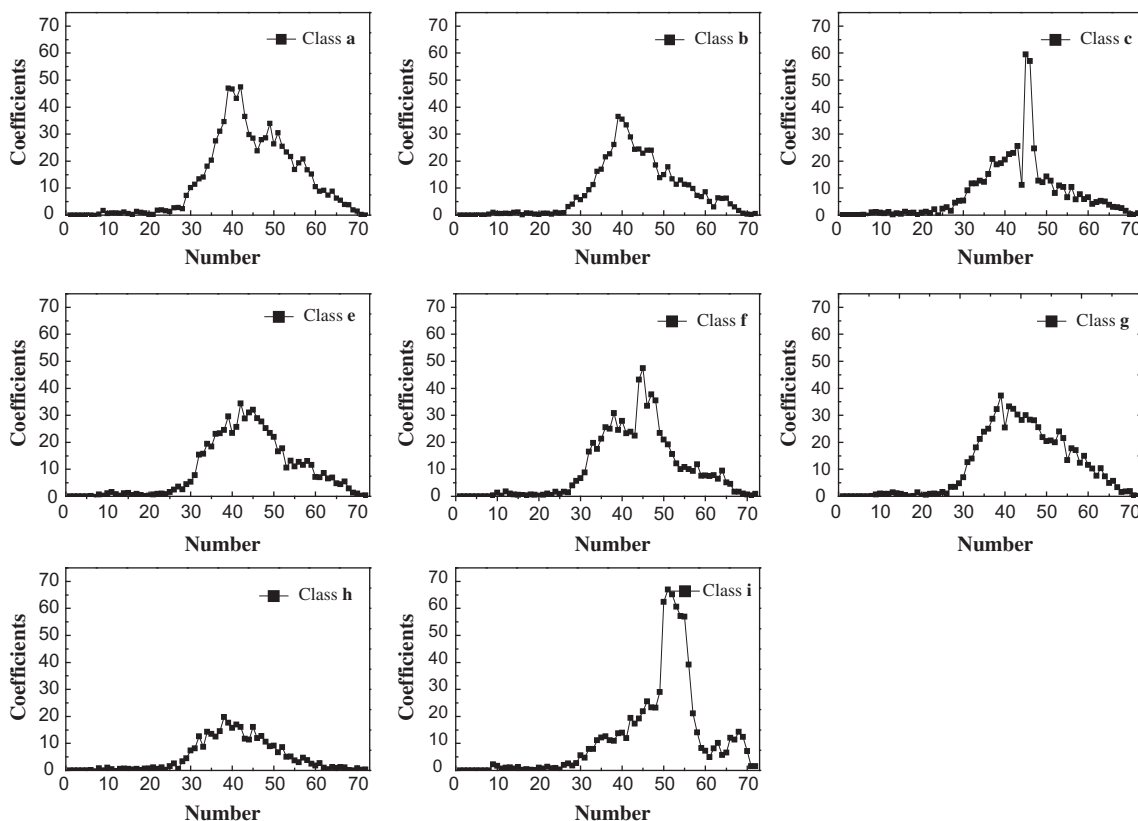**Fig. 8.** The results of wavelet decomposition and construction of methamphetamine.

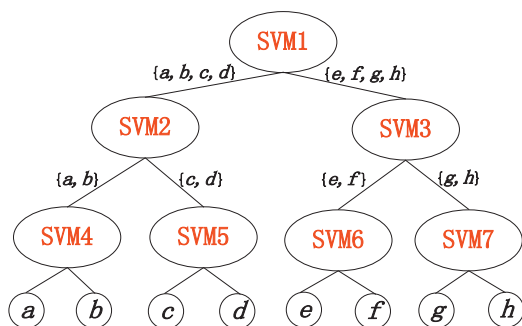**Fig. 9.** The results of eight materials using wavelet-based feature extraction.

Fig. 10. Structure of multi-classifier.

Table 2
Confusion matrix for single SVMs-BAT classification results, true vs. predicted (rows vs. columns).

	a	b	c	d	e	f	g	h
a	30	0	0	0	0	0	0	0
b	0	30	0	0	0	0	0	0
c	0	0	30	0	0	0	0	0
d	0	0	0	29	1	0	0	0
e	0	0	0	1	29	0	0	0
f	0	3	0	10	0	16	0	1
g	0	0	0	0	0	0	7	23
h	0	0	0	0	0	0	0	30

samples have been identified and then the classification process is completed.

5.4. Classification results analysis

Forty sets of data of each sample are obtained in our experiments. The classifier of WSVMs-BAT is trained and tested. Ten sets of data are selected as training samples randomly, and the others are used as validation samples to evaluate the classifier. Two other methods of single SVMs-BAT and SVMs-BAT combined with PCA based feature extraction (PCA-SVMs-BAT) is used for comparison. All interrelated parameters of SVMs are set with the same

value. Classification results of single SVMs-BAT and WSVMs-BAT are shown in Tables 2 and 3. In the tables, Rows indicate true values and columns mean prediction respectively. As noted in Table 2, the method of single SVMs-BAT makes some mistakes in predicting heroin (d), heliotropin (e), wheat flour (f) and TNT (g). The proposed method of WSVMs-BAT solves the above problem well that those spectra in Table 2 could not be classified correctly and the results are shown in Table 3. PCA-SVMs-BAT utilizes PCA to extract feature for inputs of SVMs-BAT. With the same number of principal components (72 PCs) as extracted wavelet feature size, the drugs identification accuracy of the method of PCA-SVMs-BAT is only 89.6%. In this case, the threshold with respected to variance and corresponding with the PCs is 75%. Then spectra

Table 3
Confusion matrix for SVMs-BAT combined with wavelet-based feature extraction classification results, true vs. predicted (rows vs. columns).

	a	b	c	d	e	f	g	h
a	30	0	0	0	0	0	0	0
b	0	30	0	0	0	0	0	0
c	0	0	30	0	0	0	0	0
d	0	0	0	30	0	0	0	0
e	0	0	0	0	30	0	0	0
f	0	0	0	0	0	30	0	0
g	0	0	0	0	0	0	30	0
h	0	0	0	0	0	0	0	30
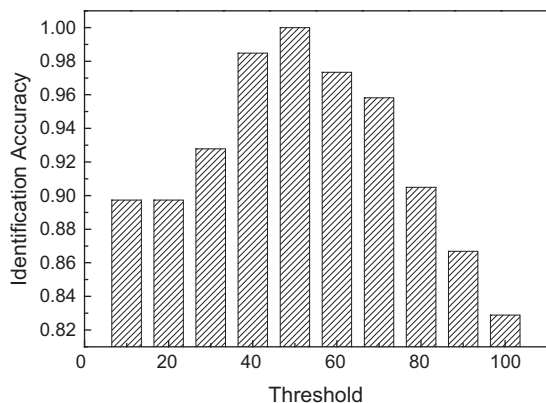
Fig. 11. Identification accuracy of SVMs-BAT combined with PCA.

identification accuracy of PCA-SVMs-BAT is computed with different thresholds for finding the reason of poor performance. The result in Fig. 11 illustrates that the identification accuracy does not vary linearly with the increase of PCs. In fact, it is the result of adding some PCs carrying the information about noise. As the PCs increasing, feature vector represents the pattern of original signal largely, but it is forced to add a lot of noise which leads to poor identification accuracy. On the other hand, insufficient PCs may miss some important features. So the selection of number of PCs and non-robustness for noise lead to the difficulty of PCA in dealing with poor drugs spectra. Compared with PCA, wavelet-based feature extraction mentioned above is robust and can combine with prior knowledge of signal to select effective feature vectors. What's more, two other common classifiers, KNN and ANN (a three layer neural network with 72 neurons in the input layer, 20 neurons in the hidden layer and eight neurons in the output layer) are compared with SVM and the corresponding classification results are 84.4% and 66.2%, respectively. It is demonstrated that SVM performs better than the two other classifiers in the cases of high dimension and small training simple size. As mentioned above, it can be found that the WSVMs-BAT algorithm is a more general-purpose classification scheme in drug identification.

6. Conclusion

A general-purpose classification method is developed to classify different categories of high-dimensional EDXRD spectral signals. The classification method utilizes wavelet coefficient based feature extraction procedure from spectral signals, which are used as inputs to develop SVM based classifiers. The obtained spectral signals with poor SNR are decomposed optimally on the basis of Shannon entropy and the approximation coefficients at four-level are used as feature vectors for inputs of SVMs. Using a tree structure of binary SVM, the method performs well for multiclass classification problems and a good overall identification accuracy can be obtained. For comparing with the proposed method, the classifiers based on single SVMs-BAT, PCA-SVMs-BAT, KNN and NN are also investigated. It is found that the proposed method has more effective feature selection and higher accuracy in the recognition of multiple drug signals with weak SNR than other approaches in the cases of small sample size and high dimension. In summary, the method of SVMs-BAT combined with feature extraction based wavelet is suitable for processing and recognition of EDXRD spectra of illegal drugs in practical application. The proposed method combining with EDXRD technology provides a good choice for non-invasion detection of human body.

References

[1] E. Cook, R. Fong, J. Horrocks, D. Wilkinson, R. Speller, Energy dispersive X-ray diffraction as a means to identify illicit materials: a preliminary optimization study, Appl. Radiat. Isotopes. 65 (2007) 959–967.
[2] C. Genzel, I.A. Denks, J. Gibmeier, M. Klaus, G. Wagener, The materials science synchrotron beamline EDDI for energy-dispersive diffraction analysis, Nucl. Instrum. Methods Phys. Res. Sect. A-Accel. Spectrom. Dect. Assoc. Equip. 578 (2007) 23–33.
[3] J.N. O'Dwyer, J.R. Tickner, Quantitative mineral phase analysis of dry powders using energy-dispersive X-ray diffraction, Appl. Radiat. Isotopes. 66 (2008) 1359–1362.
[4] W. Li, D.Y. Yu, F. Zhang, B. Sun, J.Y. Liu, M.Q. Li, J.H. Liu, Detection of heroin covered by skin by using robust principal components analysis, Measurement 44 (2011) 267–273.
[5] B. Sun, M. Li, F. Zhang, Y. Zhong, N. Kang, W. Lu, J. Liu, The performance of a fast testing system for illicit materials detection based on energy-dispersive X-ray diffraction technique, Microchem. J. 95 (2010) 293–297.
[6] S. Valle, W. Li, S.J. Qin, Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods, Ind. Eng. Chem. Res. 38 (1999) 4389–4401.
[7] C.K. Chui, Wavelets: A Tutorial in Theory and Applications, Academic Press, 1992.
[8] I. Daubechies, Ten lectures on Wavelets, SIAM, 1992.
[9] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[10] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Disc. 2 (1998) 121–167.
[11] M.M. Adankon, M. Cheriet, Model selection for LS-SVM: application to handwriting recognition, Pattern Recognit. 42 (12) (2009) 3264–3270.
[12] E. Osuna, R. Freund, F. Girosi, Training support vector machines: an application to face detection, in: Proc. of Computer Vision and Pattern Recognition (CVPR), San Juan, Puerto Rico, 1997, ISBN:0-8186-7822-4.

[13] S.H. Tao, D.Z. Chen, W.X. Zhao, Fast pruning algorithm for multi-output LS-SVM and its application in chemical pattern classification, Chemometr. Intell. Lab. 96 (2009) 63–69.

[14] G.Y. Lv, H.Z. Cheng, H.B. Zhai, L.X. Dong, Fault diagnosis of power transformer based on multi-layer SVM classifier, Electr. Pow. Syst. Res. 75 (2005) 9–15.

[15] Robert W. Madden, Jacob Mahdavieh, Richard C. Smith, Ravi Subramanian, An explosives detection system for airline security using coherent X-ray scattering technology, Proc. SPIE 7079 (2008) 707915.

[16] E. Mayoraz, E. Alpaydin, Support vector machines for multi-class classification, Eng. Appl. Bio-Inspired Artif. Neural Networks 2 (1999) 833–842.

[17] Sungmoon Cheong, Hoon Oh Sang, Soo-Young Lee, Support vector machines with binary tree architecture for multi-class classification, Neural Inform. Process. – Lett. Rev. 2 (3) (2004) 47–51.

[18] I.D. Jupp, P.T. Durrant, et al., The non-invasive inspection of baggage using coherent X-ray scattering, IEEE Trans. Nucl. Sci. 47 (2000) 1987–1994.

[19] Bernd Kämpfe, Falk Luczak, Bernd Michel, Energy dispersive X-ray diffraction, Part, Part. Syst. Charact. 22 (2005) 391–396.

[20] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. 11 (7) (1989) 674–693.

[21] R.R. Coifman, M.V. Wickerhauser, Entropy-based algorithms for best basis selection, IEEE Trans. Inf. Theory 38 (2) (1992) 713–718.

[22] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Networks 13 (2) (2002) 415–425.

[23] Osvaldo A. Rosso, Susana Blanco, et al., Wavelet entropy: a new tool for analysis of short duration brain electrical signals, J. Neurosci. Methods 105 (2001) 65–75.