Support Vector Machine Applications in Terahertz Pulsed Signals Feature Sets

Xiaoxia Yin, *Student Member, IEEE*, Brian W.-H. Ng, *Member, IEEE*, Bernd M. Fischer, *Member, IEEE*, Bradley Ferguson, *Member, IEEE*, and Derek Abbott, *Fellow, IEEE*

Abstract—In the past decade, terahertz radiation (T-rays) have been extensively applied within the fields of industrial and biomedical imaging, owing to their noninvasive property. Support vector machine (SVM) learning algorithms are sufficiently powerful to detect patterns hidden inside noisy biomedical measurements. This paper introduces a frequency orientation component method to extract T-ray feature sets for the application of two- and multiclass classification using SVMs. Effective discriminations of ribonucleic acid (RNA) samples and various powdered substances are demonstrated. The development of this method has become important in T-ray chemical sensing and image processing, which results in enhanced delectability useful for many applications, such as quality control, security detection and clinic diagnosis.

Index Terms—Pairwise classification, ribonucleic acid (RNA), support vector machines (SVMs), terahertz, terahertz time-domain spectroscopy.

I. INTRODUCTION

T-RAYS have wavelengths in the submillimetre range from 30 to 3000 μm, corresponding to the frequency range from 0.1 to 10 THz in the electromagnetic spectrum. T-rays have promising potential both in *in vivo* and *in vitro* biosensing applications [1]–[3] owing to: 1) their noninvasive property and 2) the fact that biomolecules have rich resonances in the T-ray region [4]–[6]. The ultimate aim of our work is to perform automatic classification of data obtained from T-ray measurements with tomographic applications [7]. It is important to devise effective feature extraction methods to fully represent the different characteristics of these signals [8]. Signal processing methods are proposed for the current experiment. In the two experiments presented in this paper, input measurements are decomposed based on specific properties of their Fourier spectrum [9]. From the spectrum, a pair of specific values are extracted as features, which takes the place of a large amount of spectral data. In doing

so, the number of features is kept smaller than the number of observations to prevent over fitting. In this case, there is the added benefit of reduced computation complexity with the low feature dimensionality.

The current experiments illustrate the potential of support vector machines (SVMs) in pulsed signal recognition. The classification of two different types of ribonucleic acid (RNA) samples is of importance because it is a potential precursor for deoxyribonucleic acid (DNA) sample classification in an *in vivo* environment [10]. We also classify six different types of powder materials, to illustrate the validity of our approach to other applications, such as security. We describe the methods for T-ray pulse classification. The system's input are the measured raw T-ray signals, with the only preprocessing being a deconvolution in the Fourier domain. Feature extraction and classification are performed with the aim of achieving high classification accuracy. In this work, we found that an SVM kernel-based method can be directly applied to the specific features. The main advantage of the current approach is that it is not necessary to design an elaborate feature extraction scheme [11], since even elementary feature extraction is sufficient for the SVMs to perform well. A pairwise classification scheme [12] is applied to multiclass recognition problems, which leverages the good performance of binary SVMs to solve complicated multiclass problems. Gaussian and polynomial kernels have been found to give good performance for two-class and multiclass classification problems, respectively. Visual classification performance is also shown for the 2-D features of binary and multiple classification, which gives useful demonstration of the effectiveness of pattern learning problem and makes the performance of SVMs more understandable.

This paper is organized as follows. Section II briefly introduces a traditional terahertz imaging system. In Section III, after reviewing the basics of the SVMs, including binary classification and pairwise classification, terahertz feature extraction using frequency components for the RNA and powder samples are discussed. In Section IV, we report the results from the classification experiments. Section V concludes this paper with the closing remarks.

X. Yin, B. W.-H. Ng, B. M. Fischer, and D. Abbott are with the Centre for Biomedical Engineering, School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA 5005, Australia (e-mail: xxyin@eleceng.adelaide.edu.au; bwng@eleceng.adelaide.edu.au; bfischer@eleceng.adelaide.edu.au; dabbott@eleceng.adelaide.edu.au).

B. Ferguson is with Centre for Biomedical Engineering, School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA 5005, Australia, and also with Tenix—Electronic Systems Division 2nd Avenue, Technology Park, Mawson Lakes, SA 5095, Australia (e-mail: bradleyf@eleceng.adelaide.edu.au).

II. A BRIEF INTRODUCTION TO EXPERIMENT SETUP OF T-RAY TRANSILLUMINATION

Terahertz pulsed imaging (TPI) systems or terahertz time-domain spectroscopy imaging was used for the current experiment to achieve T-ray imaging statistics. In a typical TPI system, both pump and probe beams are derived from an ultrafast pulsed optical laser beam. A pair of paraboloidal mirrors are used as an
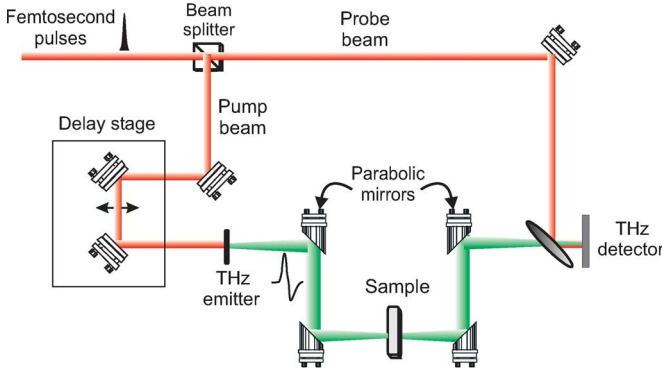
Fig. 1. Illustration of a femtosecond laser-based T-ray functional imaging system based on a pump-probe configuration. After [13].

optical delay stage to adjust the path length of the pump beam. T-rays are generated via a chopper and an optical rectification crystal, which, then, are collimated and focused onto a sample by a pair of parabolic mirrors. The sample can be moved in the plane perpendicular (the x-y plane) to T-ray beam by a computer-controlled translation stage. The T-rays emerging from the sample are recollimated by another pair of mirrors, before being combined with the probe beam. As a result, the T-ray response and the probe beams propagate through a T-ray detector crystal colinearly. The resultant measured waveform is analogous to the optical output of the detector crystal, which is detected by a photodetector.

A complete T-ray imaging system consists of T-ray waveform for each position of the object in the x-y plane, therefore, a T-ray image can be built up pixel by pixel. The image formed in this fashion affords the T-ray pulsed response full information in both amplitude and phase (time delay). Fig. 1 illustrates the TPI hardware, where a 2-D raster scan is employed to achieve T-ray images via repetitive pulsed terahertz measurements.

III. METHODOLOGY

A. Overview of SVMs

A product of recent advances in statistical learning theory [14], [15] is the SVM. The idea of the SVM is to map data from the input space into a high-dimensional feature space, in which an optimal separating hyperplane that maximizes the boundary margin between two classes can be established. At its core, SVMs are a two-class classifier. In more recent times, SVMs have been extended to solve multiclass classification problems. The following sections describe each in turn.

1) Binary Classification by a SVM: SVMs are binary classifiers, which classify data depending on a set of support vectors [16]. The training data set is a collection of labeled samples called learning vectors. We denote such a set of learning vectors as $(\mathbf{x_i}, y_i) \in R^N \times \{\pm 1\}$, $i = 1, \ldots, l$, and y_i denote the class label corresponding to each input vector $\mathbf{x}_i$. The support vectors are subsets of the training data sets and are used to construct a N dimensional hyperplane in feature space, which acts as a boundary separating the different classes. A decision function $f(\alpha) : R^N \to \pm 1$ is achieved based on a given class function $f(\alpha) : \alpha \in \Lambda$, with the aim to correctly assign class labels to test

samples $\mathbf{x}$. The Vapnik Chervonenkis (VC) dimension [14], [15], [17] is a property of a set of functions $f(\alpha)$, which is defined as the maximum number of training points that can be segmented by $f(\alpha)$. Note that α corresponds to the weights and biases, which can be adjusted to label the output $f(\mathbf{x}, \alpha)$ based on the input $\mathbf{x}$. The expectation of the test error for a learning machine is

$$R(\alpha) = \int \frac{1}{2} |y - f(x, \alpha)| \, dP(\mathbf{x}, y) \tag{1}$$

where, $R(\alpha)$ is called the expected risk. It is the quantity connected with density $p(\mathbf{x}, y)$ that we are ultimately interested in. The "empirical risk" $R_{\mathrm{emp}}(\alpha)$ is defined to be the measured mean training error for a fixed, finite number of observations

$$R_{\mathrm{emp}}(\alpha) = \frac{1}{2l} \sum_{i=1}^{l} |y_i - f(x_i, \alpha)|. \tag{2}$$

The quantity $(1/2)|y_i - f(x_i, \alpha)|$, which is called the loss, takes on values equal to 0 or 1. When we set the probability to $1 - \eta$, we achieve the following bound:

$$R(\alpha) \leq R_{\mathrm{emp}}(\alpha) + \sqrt{\left(\frac{h \left(\log(2l/h) + 1 \right) - \log(\eta/4)}{l} \right)}. \tag{3}$$

The non-negative integer h is the VC dimension, and it is the measure of the *capacity* that is suitable for the amount of available training data. When a sufficiently small η is selected, the right hand side of the equation is minimized, and the $f(\mathbf{x}, \alpha)$ functions give the lowest upper bound of the actual risk. This is the basic idea of structural risk minimization.

Since minimizing the training error (the computation of VC-dimension) does not guarantee a small test error, in order to make the decision function f perform well on unseen pattern, the principle of structural risk minimization needs to be applied to minimize test error and achieve a capacity that is suitable for the amount of available training data sets.

The learning algorithm is designed to compute the *support vectors* via performing structure risk minimization. First, the structural risk minimization problem is solved to reveal, among all the learning vectors, those that have the minimum margin to the optimal decision hyperplane, these vectors are called the support vectors. Finally, the decision hyperplane is defined by the following equation:

$$\sum_{i=1}^{s} y_i \alpha_i (\mathbf{x}_i \cdot \mathbf{x}) + b = 0 \tag{4}$$

where $\mathbf{x}_i$, $(i = 1 \ldots s)$ are support vectors. The solution of this large-scale quadratic programming problem is applied to calculate the coefficients α_i and b. The usual procedure is to solve the mathematical dual of the structural risk minization problem, instead of the primal problem. The dual problem is to maximize

$$L(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \tag{5}$$

subject to $\sum_{i=1}^{l} \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$ for $i = 1, \ldots, l$. The penalty parameter C is selected by the user, which is viewed as a regularization parameter that characterizes one's willingness

to accept possible misclassifications in (linearly) nonseparable datasets.

The basic SVM can only be used to solve linear classification problems—the computed hyperlanes are linear in R^N. In order to solve nonlinear problems, SVMs use a kernel function [18], [19], which allows better fitting of the hyperplane to more general data sets. Instead of a linear classifier, the hyperplane [20] needs to be augmented to fit general nonlinear decision surfaces. A dot product space is constructed by mapping data, which is realized by performing a nonlinear map $\phi : R^N \rightarrow F$. The above linear algorithm then can be applied in the feature space F. The solution satisfies the following:

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^{s} y_i \alpha_i \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}) + b \right) \qquad (6)$$

where sgn is the signum function. This is a nonlinear function of the original input vectors $\mathbf{x}$.

According to Cover's [21] theorem, a new feature space F can be achieved via transforming a multidimensional space, where the dimensionality of the feature space are high enough to allow the target patterns to be linearly separable with a high probability. The inner products (dot products) enable the high dimensional space to be treated easily, which are indicated by the ϕ mapping. Accordingly, the kernel function K is defined as

$$K(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{y}). \qquad (7)$$

The kernel production substitutes for all the occurrences of the dot product resulting from two mappings.

There are four popular kernel functions in the literature.

- Linear kernel

$$K(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}. \qquad (8)$$

- Polynomial kernel

$$K(\mathbf{x}, \mathbf{y}) = (\gamma \mathbf{x}^T \mathbf{y} + r)^d, \quad \gamma > 0. \qquad (9)$$

- RBF (Gaussian) kernel

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\gamma \|\mathbf{x} - \mathbf{y}\|^2\right), \quad \gamma > 0. \qquad (10)$$

- Sigmoidal kernel

$$K(\mathbf{x}, \mathbf{y}) = \tanh(\gamma \mathbf{x}^T \mathbf{y} + r), \quad \gamma > 0. \qquad (11)$$

where $\mathbf{x}, \mathbf{y}$ are data vectors in input space.

In the current experiment, RBF kernel function is proposed as a choice for identifying RNA samples, and it was found to give good classification performance. Meanwhile, it is found that polynomial kernels are best suited to achieve multiclass classification of powdered samples. They are demonstrated in Section IV. Linear and sigmoid kernels are not shown since they have been found to produce inferior results for our T-ray dataset.

2) Multiclass Classification by Pairwise SVMs: The previous section described a SVM for two-class pulsed signal classification, which is also called a *dichotomy*. This is appropriate for the object detection application of two-class T-ray pulses and discriminating an object from the background. However, the majority of object recognition problems consist of more than two
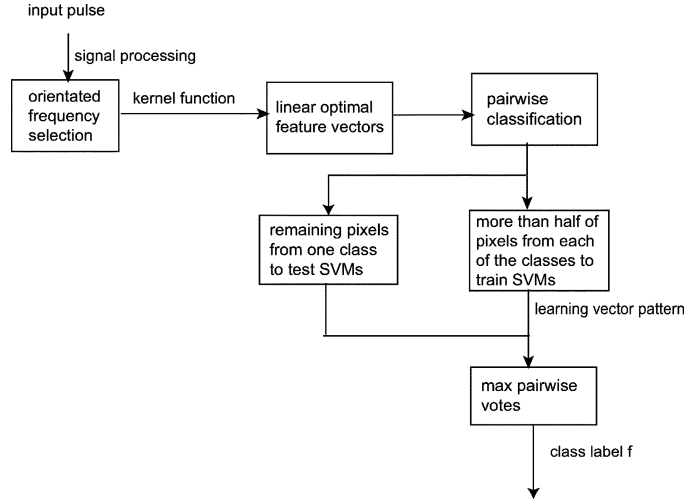

Fig. 2. Illustration of the procedure for pairwise multiclass classification.

types of samples. Consequently, extended SVMs are required for application to multiclass pulsed signals. The optimal design of multiclass SVM classifier is an area of active research. One frequently used method in practice is to use a set of pairwise classifiers, based on one-against-one decomposition [12]. These work by using a collection of decision functions f_{kl}, here, kl indicates each pair of classes selected from separated target classes. Due to the symmetric property of the pairwise approach, it holds that $f_{kl} = -f_{kl}$, where f_{kl} satisfies the following equation:

$$f_{kl}(\mathbf{x}) = \mathbf{w}_{kl} \cdot \mathbf{x} + b_{kl} \qquad (12)$$

where $\mathbf{w}$ is normal to the hyperplane between class k and class l, $|b|/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin, and $\|\mathbf{w}\|$ is the Euclidean norm of vector $\mathbf{w}$.

The signum function is used for the hard threshold decisions

$$\text{sgn}\left(f_{kl}(\mathbf{x})\right) = \begin{cases} 1, & f_{kl}(\mathbf{x}) > 0 \\ -1, & f_{kl}(\mathbf{x}) \le 0 \end{cases} \qquad (13)$$

where n is the number of the separated target classes, and the class decision can be achieved by summing up the according pairwise decision functions

$$f_k(\mathbf{x}) = \sum_{k \ne l, l=1}^{n} \text{sgn}\left(f_{kl}(\mathbf{x})\right). \qquad (14)$$

The algorithm proceeds as follows: assign a label to the class: $\arg\max f_k(\mathbf{x}), (k = 1, \ldots, n)$. The maximum number of votes for k classes holds $\max_k f_k = (k - 1)$. If (14) is satisfied for $\max_k f_k < (k - 1)$, the $\mathbf{x}$ is unclassifiable. The pairwise classification converts the n-class classification problem into $n(n-1)/2$ two-class problems which cover all pairs of classes.

The procedure for pairwise classification is illustrated in Fig. 2.

B. Terahertz Data Representation

This section describes the measurement of T-ray pulsed responses. There are two target data sets: one is for RNA samples and the other is for powdered substances. The former contains
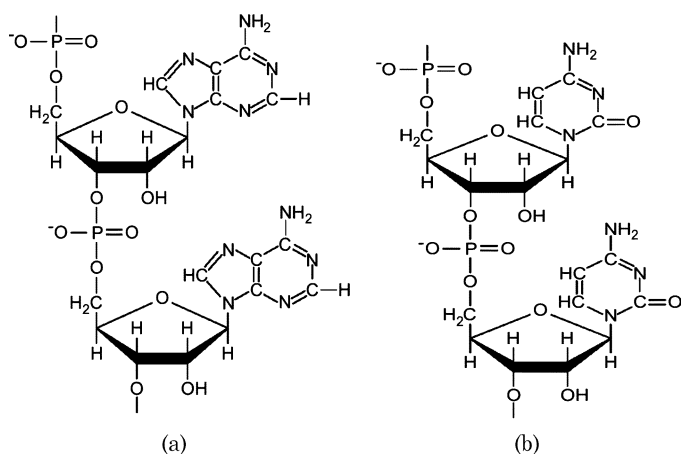
Fig. 3. Chemical structure of (a) polyadenylic acid (poly-A) and (b) polycytidylic acid (poly-C). Potassium ions are applied to neutralize the negative charge of the backbone. After Fischer *et al.* [10].

Fig. 4. T-ray transmission image of the poly-A and poly-C, showing stronger absorption in poly-C compared with poly-A. Each spot contained 200 μg of either poly-A or poly-C in alternating order, as indicated in the diagram to the right. The color scale indicates the normalized peak values of the two RNA samples. After Fischer *et al.* [10].

two separate classes of biological specimen and the latter contains six various types of powdered materials. The representation of data measurement is detailed in the following sections.

1) RNA Data Representation: Recently, it has been suggested that biological material can be detected by T-ray circular dichroism (TCD) spectroscopy, because many biomolecular crystals exhibit strong and specific absorption features in their dielectric spectra [10]. Currently, the identification of the binding state of DNA is an especially interesting topic, which can be realized through applying T-ray techniques, i.e., loading a planar T-ray resonator with the sample material, in spite of the lack of characteristic absorption features in the T-ray region. In this experiment, it is investigated that frequency-domain data from two different RNA polymer strands, polyadenylic acid (poly-A), and polycytidylic acid (poly-C) can be used for the recognition task with the potential to classify DNA.

Commercially available poly-A and poly-C potassium salts are used for the experiment (Sigma–Aldrich, product numbers P9403 and P4903), the data of which was measured at the University of Freiburg, Germany, and Fig. 3 illustrates the chemical structures of the polymers.

The sample preparation has been described in detail in Fischer *et al.* [10], and the following steps are just a short summary. (i) The carrier substrate is a commercially available cyclic olefin homopolymer slide (Greiner Bio-One) with transparent and dispersion-free properties in the T-ray frequency range. (ii) In order to achieve T-ray imaging, small liquid volumes are spotted by hand in a 4 × 4 array of alternating poly-A and poly-C on the slide. Each spot with a diameter of approximately 1 mm consists of 2 μl of deionized water containing 0.2 mg material. The spots were then dried at room temperature and checked for homogeneity [9].

As a result, the terahertz image is illustrated in Fig. 4, which is achieved by terahertz time-domain spectroscopy imaging system based on free-space propagation and aperture-less focusing of the T-ray beam. Each pixel in the image represents the normalized peak values corresponding to Poly-A and Poly-C. The sample consists of a 4 × 4 array of spots. Two of the spots are removed from the substrate in order to identify the orientation of the substrate in the image. The spot of Poly-A is shown
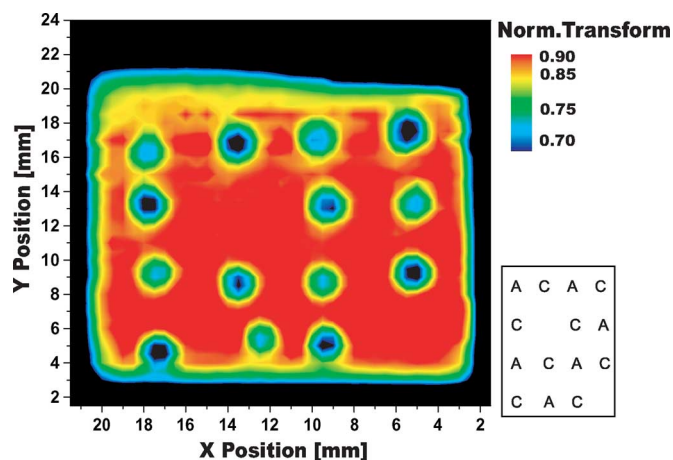
at the top left corner of the image, with weak transmission, compared with the spots of poly-C. The positions of poly-A and poly-C subimages are labeled in the diagram to the right of the picture.

Based on the positions of poly-A and poly-C at the terahertz image, we select eight-neighbor pixels around center position from each spot for the signal postprocessing and classification. The pixels lying on the boundaries of each class are excluded from training and test vectors due to the difficulties involved in manually assigning the desired values for pattern recognition.

2) Powder Data Representation: Another topical application of T-rays has been the classification of powdered samples as a basis technology in substance detection for security [13]. A general question is to explore the ability of T-ray spectroscopy to detect different densities, thicknesses, and concentrations of specific powders. To investigate this problem, we conduct a preliminary exploration of different powder recognition tasks with 2 mm thickness for six different powdered substances and their holder. They are: sand, talcum, salt, powdered sugar, wheat flour, and baking soda.

The sample holder is shown in Fig. 5, this holder has an ability to accurately control the thickness of the powders. There are two teflon blocks, which can be separated under control. They are mounted on a manual translation stage to provide the required gap of 2 mm, where a plastic bag containing the powdered substances can be inserted between the teflon blocks. This procedure guarantees a relatively consistent powder density and accurate control over the powder thickness of 2 mm.

A traditional T-ray imaging system is used to detect T-ray responses based on the THz-TDS technique. There is an x-y translation stage involved for mounting the teflon sample holder and fixing its position in the T-ray beam. At T-ray frequencies, teflon is dispersionless. As teflon has a very low absorption coefficient, there is minimal distortion, while the T-ray pulse propagates through the holder. A 2-D T-ray image of the sample can be obtained after inserting a powder sample. This image allows the effects of different scattering paths and minor variations in powder thickness and density to be observed. In general,
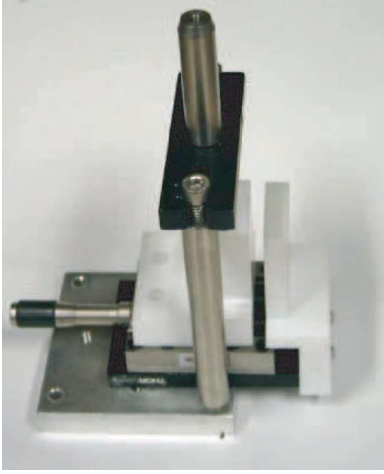
Fig. 5. Photo of a teflon sample holder for measurement fixed thickness of powdered samples. One of the teflon blocks is fixed, while the position of the other is controlled using the manual translation stage. The gap between the two blocks may be adjusted to allow 2 mm thicknesses of powder to be considered.

a 1-D image is sufficient for substance detection purposes, and 50 pixel responses (with a pixel spacing of 100 μm) can be acquired in under 30 min.

C. Terahertz Feature Extraction

This section describes the classification system designed to assess the potential of SVMs in T-ray pulsed classification. There are two target data sets, which need to be separated by SVMs: one is for RNA samples and the other is for powder samples. The former is to classify two classes of objects and the latter is to separate six various types of powdered materials. RBF kernels and polynomial kernels are applied for statistical feature mapping. Signal processing is applied to track the key features of training vectors for different classes of signals.

1) Feature Extraction via Frequency Orientation Components: The way to extract specific feature vectors in the frequency-domain is realized by taking the Fourier transform after deconvolving measured signals with a reference pulse [22]. The Fourier transform produces complex-valued spectra, containing both phase and magnitude information. The magnitude and phase at certain key frequency components constitute pairs of feature subsets on which the classification is based. An important advantage of this approach is the small dimensionality of feature vectors, allows the features to be directly extracted from pulsed responses with relatively low computational complexity. This obviously reduces the computational resource requirements, which is attractive for possible hardware-based implementations. Further, the sparse features help us avoid the overfitting problem [4], [23]. The input vectors occupy a size $M \times L$ matrix, where M is equal to the number of input vectors (training vectors) and L is the dimensionality of each feature vector. If the limited measured dimensions of the training subsets M is smaller than large scale of time series features L, it will cause difficulty in correctly assigning labels to target samples. Fast and sparse features overcome the computational disadvantages of SVMs.

2) Kernel Selection and Parameter Tuning: Following feature extraction, the kernel operation is performed on the calculated features in an SVM. An implicit nonlinear transformation Θ is used to map input pattern $\phi(\mathbf{x})$ into a higher dimension, yet linear space. In the case of a Gaussian kernel, this transformation is related to a Gaussian function $K(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2)$, where γ is the Gaussian kernel width parameter. Accordingly, a SVM classifier needs to be applied to produce learning vector patterns in two dimension feature space. The scatter plot for learning vectors using a Gaussian kernel is illustrated in Fig. 7, which considers a two-class pulsed signal classification problem for the recognition of two types of RNA samples—poly-A and poly-C. It serves as a useful comparison with plotting machine learning realized via a polynomial kernel for classification tasks, illustrated in Section III-B on multiclass powder classification. The feature vectors consist of magnitude and phase, which are plotted in a 2-D plane: x axis is labeled by magnitude and y axis labeled by phase.

The parameters that need to be tuned in the SVMs include the kernel parameters and penalty parameter C [24], [25]. The use of a *validation dataset* can be viewed as a straightforward method for tuning. This method reserves from the training dataset a number of features for validation [11], which are used to gauge the performance of the classifier trained on the remaining training data. In our experiments on the RNA data, we have chosen the validation dataset approach for the RNA data. However, the use of validation sets becomes infeasible when the number of feature vectors in the datasets is small. This is the case for the powder classification experiments. In that case, we have chosen to forego the validation set, and instead used the test error as a direct way of comparing against different parameter values. In spirit, this approach is akin to an exhaustive search for optimal SVM parameters, while it is reasonable when there is a paucity of data and few parameters to determine.

D. Performance Assessment of Classification

Cross-validation methods [26], [27] and leave one out (LOO) [28], [29] estimator within the deconvolved T-ray data set are utilized to provide a nearly unbiased estimate of the prediction error rate. The performance of classifying the RNA samples are evaluated using eightfold cross-validation, while the powdered material classification is validated using LOO. The dataset of RNA is divided into eight subsets of approximately equal size. Sequentially, each subset was tested using the classifier trained on the remaining subsets. The results from the eight runs are averaged to provide a statistical estimate of the classifier performances. To tune the parameter C, we used small-and-separate validation sets drawn from the test subsets, with the remainder of the test subsets used for testing the classification performance.

In the approach outlined above, each RNA pixel instance is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. Similarly, LOO evaluates each unknown feature vector, and then produces a basis to evaluate classifier designs for powder classification [7], [28]. Therefore, LOO accuracy is also the percentage of correctly classified data sets. The reason why LOO is used instead of eightfold

cross-validation for the powder experiment is due to the relatively small number of measurements for the different powders. With such a restriction, LOO is preferred as the overall classification experiment is averaged over more runs. In this experiment, accuracy of classification is used as the quantity for assessing the performance of all the classification tasks, and it is calculated by

$$\text{accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FN} + N_{TN} + N_{FP}} \quad (15)$$

where true positives is labeled by N_{TP}, true negatives is labeled by N_{TN}, false positives is labeled by N_{FP}, and false negatives is labeled by N_{FN}.

IV. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed method, experiments were performed on T-ray pulsed data related to RNA samples and several types of powder substances. "LIBSVM" [30] and with "SVM and Kernel Methods Matlab Toolbox" [31] are chosen as the toolboxes for our experiments with two-class and multiclass classification, respectively.

A. The Fourier Spectrum Analysis

As mentioned above, the detection procedure works on a modified transmitted time-series T-ray pulsed responses. For isolation of system response, a pulse measured from an empty substrate and an empty holder is employed as reference and is deconvolved from the measured signals of RNA samples and powdered materials, respectively, for the removal of the system response [22], [32]. The specific features relevant to magnitude and phase are extracted from the RNA and powdered substances data, via a Fourier transform. The details are described in the following sections.

1) The Fourier Spectrum Analysis for the Classification of Poly-A and Poly-C T-Ray Pulses: RNA data occupies a size equal to $60 \times 50 = 3000$ pixels. For each pixel, the number of time samples is 350, which was accordingly truncated at 175th frequency bin corresponding to frequency of 4 THz because of the symmetry of spectrum. The 3000 pixel data set consists of background data information—a TOPAS substrate image, and target object data sets—poly-A and poly-C image data. The population of pixels belonging to the poly-A and poly-C classes is 48 for both classes. In order to obtain reduced dimensions of feature subsets and make them discriminable for the different class, the magnitude and phase values of the pulse responses are first calculated, and then those values corresponding to the frequency with the greatest magnitude (i.e., strongest response) are used as the features. This process extracts a 2-D feature vector from the full spectral data with 350 nonredundant dimensions.

Fig. 6(a) displays two obvious separate phase curves for two classes of spectra—poly-A and poly-C, which are obtained by fast Fourier transform (FFT) of the T-ray pulses with linear extrapolation. Fig. 6(b) is the plot of magnitude versus different frequency steps with cutoff frequency at the 175th frequency bin. It is observed that the magnitude at frequency bin 19 reaches maximum value. Hence, we select frequency bin 19 as the key frequency of interest and use the corresponding phase and magnitude pair as the extracted features which form the input to the
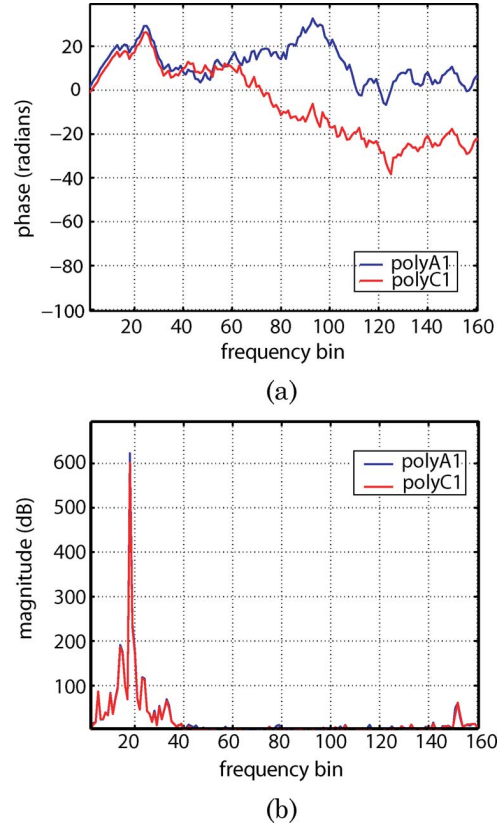

(a)

(b)

Fig. 6. Illustration of T-ray spectra of RNA. (a) Displays two obvious separate phase curves for two classes of spectra—poly-A and poly-C, which are obtained by down sampling the FFT of T-ray pulses with linear extrapolation. (b) The plot of magnitude versus different frequency bins with largest magnitude at frequency bin 19.

SVM. The Gaussian kernel is used for the final feature mapping from a nonlinear feature space to a linear one. Accordingly, we apply the SVM training algorithm to produce a learning vector pattern in the 2-D feature space (magnitude and phase form the axes), which is illustrated in Fig. 7. SVMs with the width parameter of Gaussian kernel λ of 0.003 are trained by feeding 6–42 training vectors selected randomly from 48 patterns from each class. The orientation frequency is selected at the 19th frequency bin in all cases. The background color shows the shape of the decision surface. Dark blue regions represent the class belonging to the poly-C sample labeled by -1, and light blue regions indicate the class related to poly-A sample labeled by 1. Separating hyperplanes for two classes are indicated by 0. The circles represent the calculated support vectors. Compared with the training vectors, the number of support vectors are reduced, which takes on an important role in achieving the ideal shape of hyperplanes and facilitating computation of the classification algorithm. In this case, the machine learning for two-class samples—poly-A and poly-C denoted by white "+" and black "×" are approximately separated by their own boundary lines though there is a little overlapping. Detailed experimental results about classification accuracy are analysed in the next section, after 200 random selections of training vectors are fed to the SVMs.

2) Fourier Spectrum Analysis for Multiclass Classification: The image statistics of powders consist of $6 \times 50 = 300$ pixels. For each pixel, the number of time samples is 400. Fig. 8 shows
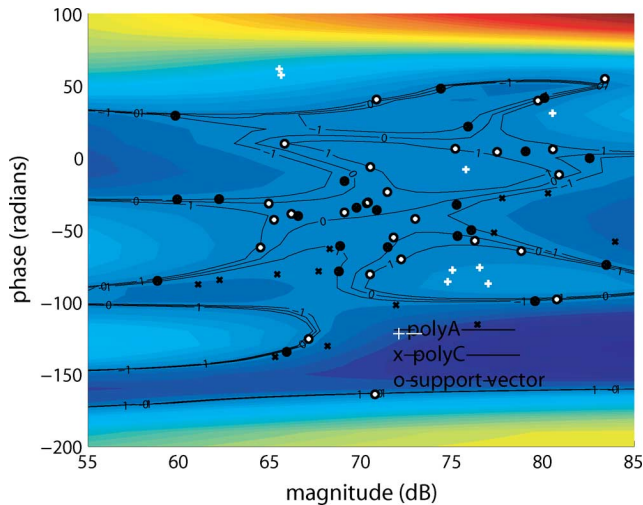
Fig. 7. An illustration of binary classification for the recognition of RNA samples. A Gaussian kernel is used for mapping the training vectors to a 2-D feature space. The penalty parameter C is set to infinity and the width parameter of Gaussian kernel λ equals 0.0006. The background color shows the contour shape of the decision surface. The learning vectors are approximately separated via applying Gaussian kernel for mapping.

Fig. 9. Plot of learning vector pattern with 49 pixels from two-class powder samples, sand and salt samples, with phase and amplitude as a pair of coordinates. Polynomial kernel with degree of 3 is applied for linear mapping. The penalty parameter C is set to 10.
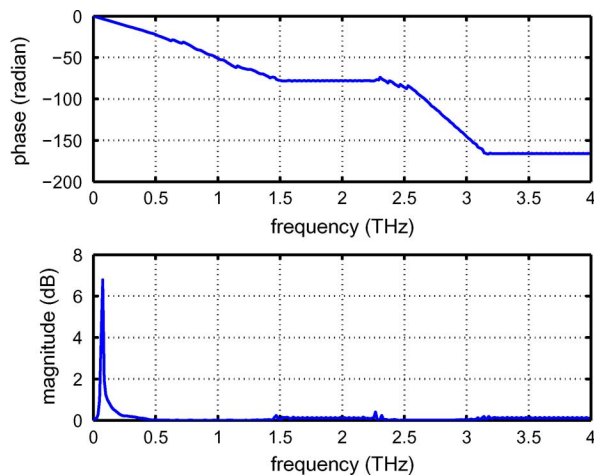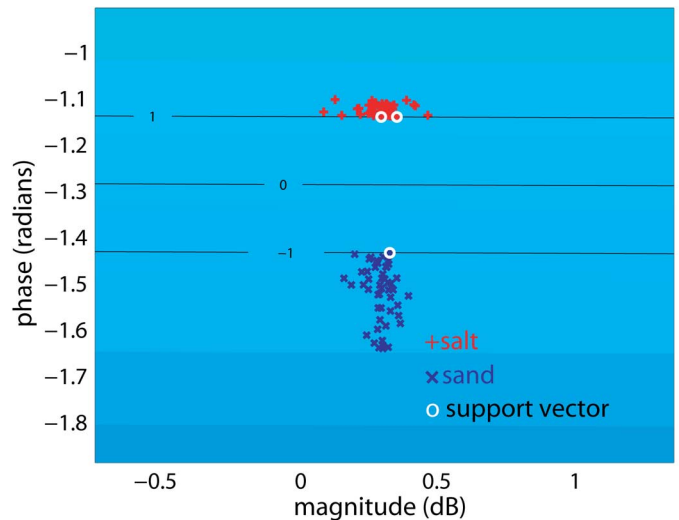

Fig. 8. Illustration of Fourier spectrum. The top subimage shows phase versus terahertz frequency and the bottom subimage shows magnitude versus terahertz frequency.
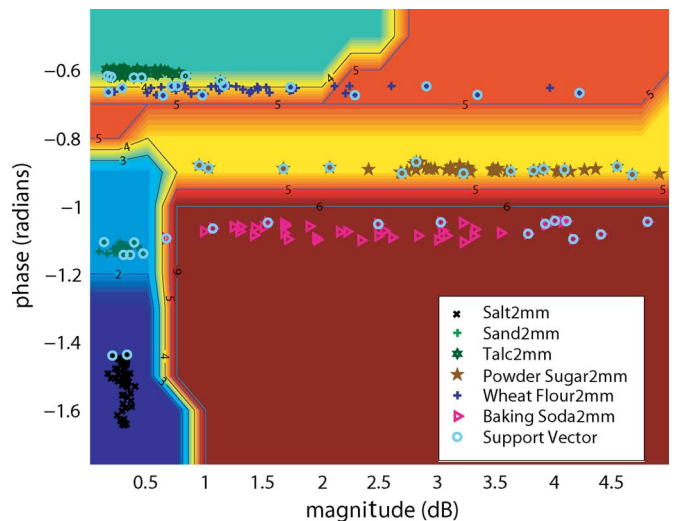

Fig. 10. Learning vectors for the six-class examples are plotted for the illustration of the linear decision function between the pairs of classes by applying a polynomial kernel for mapping with degree of 1. There are 49 pixels selected randomly from each of six classes of powder samples. The small red region on the left hand side remains undecided.

the phase and magnitude plots in the frequency-domain from one pixel of salt image data, with a cutoff frequency equal to 4 THz. It is obvious that the subimage at the bottom of Fig. 8 shows a sharp change of magnitude at the second frequency bin. Accordingly, with the number of training subsets of 49, we plot the phase and magnitude based on two classes learning vectors and multiclassed learning vectors, by applying the pair Fourier features at the second frequency bin, then we produced the learning vector pattern for two-classes and multiclass recognition, shown in Figs. 9 and 10, respectively. For powder classification, a polynomial kernel is employed for optimal classification performance. The two-class powder samples, sand and salt samples, are linearly separable with circles describing the calculated support vectors, which decide the linear optimal hyperplane between two classes. The solid lines above and below the hyperplane depict the ± 1 range along the separating surface.

The small number of support vectors greatly reduces the computational burden of the classification task. In two-class classification, the penalty parameter C was chosen to be 10 and the polynomial kernel degree equals 3.

For the multiclass case, a polynomial kernel with degree of 3 is applied for linear mapping, with a truncated terahertz frequency of 4 THz. The penalty parameter C is set to 100 for illustrative purposes. The relevant decision functions for the pairwise approach are shown in Fig. 10 with the number for the recognition of the various decision surfaces corresponding to the different pairwise classes. The summing up of the pairwise votes yields the borders easily. The small red region at the left hand outlined by border line 5 is an undecided class region. This is because the maximum number of votes (top scores) in the region is

smaller than $(k-1) = 5$. The six classes are well clustered and, therefore, easily separated. The undecided class region has no relevance for the class decision. It is obvious that the number of the support vectors for the single decision is small, which results in a fast adaption and better boundary shape to partly compensate the computation increase brought on by needing to repeat $(k-1)/2$ calculations for multiclass classification. The support vectors as the "most important" data points are identified with extra circles.

B. Resultant Classification Performance

1) RNA Classification via Fourier Coefficient Spectrum: For classification of RNA samples, Gaussian kernels are applied because of the preferred classification performance and fit the RNA data well. All the classification runs are performed using a 2.4 GHz Pentium4 CPU. The average time spent classifying the two-classes RNA samples is 2.74 s after 18 runs of measurements.

To evaluate the effect of the Gaussian kernel for the RNA sample classification, suitable values of C and λ are evaluated via parametric search using separate validation sets. After training, the final error rate, the number of support vectors, and the elapsed time are compared. In the training phase, the training vectors are randomly selected from a given proportion, varying from 1/8 to 7/8, of the input population of 48 pixel responses from each RNA class. The SVM parameter (C) is tuned by the remaining 1/16 to 7/16 of input data as validation vectors and tested with the last small-and-separate subsets of 1/16 to 7/16 of available pixel responses. In principle, a similar procedure can be applied to the tuning of parameter λ—in this paper, we illustrate the tune of C for brevity, since we have found the classification performance to be less sensitive on the choice of λ. As discussed in Section IV-A1, the key frequency is selected at the 19th frequency bin in all cases. In order to achieve effective classification, the repeated selection of test and training vectors are conducted, and the number of repeated selection approaches 200. The highest classification performance was obtained for the penalty parameter $C = 1$ and $\lambda = 0.003$, with a classification accuracy of 72%.

It was found that the classification accuracy is similar throughout the range of values for C, from 0.001 to 10^4, in steps of 1 in log scale. The classification accuracy is improved with an increased number of training vectors, which is to be expected. Fig. 11(a) and (b) show classification performance using our algorithm versus different sizes of test and validation data sets, respectively. The various value of C are all plotted for direct comparison. It is clear that the two subfigures show the similar behavior, which implies that the validation and test sets do not exhibit very different classification characteristics. In Fig. 11(a), the curve, related to C of 1, gives best performance, especially when the number of training vectors is in the range from 48 to 84, though in Fig. 11(b), the corresponding curve shows a little weak classification accuracy compared with the others, but not by very much.

The number of the computed support vectors is roughly 1/3 fewer than the number of training vectors. The small number
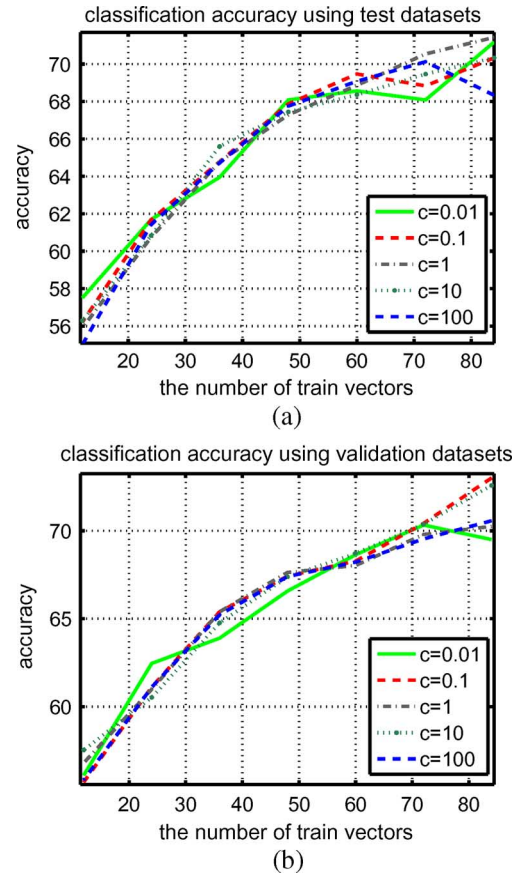

Fig. 11. (a) Plot of classification accuracy versus the number of input test vectors, corresponding to the different value of parameter C. (b) Illustration of the validation of classification accuracy, via the plot of classification performance versus the number of input validation vectors, corresponding to the different value of parameter C.
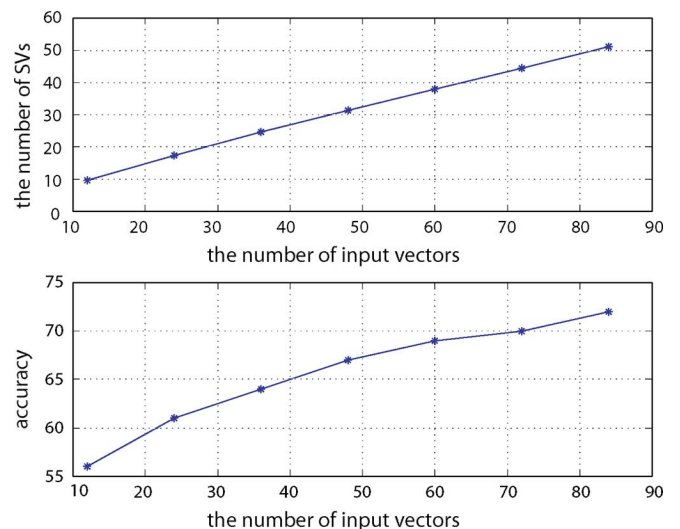
Fig. 12. Illustration of the variation of the number of SVs and classification accuracy with the increment of the number of the input training feature sets in the classification of RNA data, with $C = 1$ and $\lambda = 0.003$.

of SVs is desirable for implementation since it directly determines the computational complexity of the automatic classification task. Fig. 12 shows the variation of the number of SVs with the increment of the number of the input training feature sets

TABLE I
CLASSIFICATION RATES (%), NUMBER OF SVS AND ELAPSED TIME ARE ILLUSTRATED FOR POWDER CLASSIFICATION USING VARYING PENALTY PARAMETER C AND THE POLYNOMIAL KERNEL $\mathbf{p} = \mathbf{1}$. THE FOURIER SPECTRAL FEATURES WERE EXTRACTED. HERE, 97.96% AND 57.14% CLASSIFICATION ACCURACY CORRESPOND TO THE NUMBER OF THE CORRECT CLASSIFIED TEST VECTORS EQUAL TO 48 AND 28 WITH 1 AND 21 TEST ERRORS, RESPECTIVELY

| | each class name and classification accuracy (%) | | | | | | ♯ of SVs | elapsed time (s) |
C	salt	sand	talcum	sugar	flour	soda		
0.1	95.92	83.67	97.96	83.67	57.14	34.69	566	35.6875
1	100	100	93.88	95.92	67.35	89.80	334	23.9531
10	100	100	95.92	100	83.67	93.96	112	15.0496
100	100	100	97.96	100	100	97.96	64	12.6875
1000	100	100	97.96	100	100	97.96	46	11.9688

TABLE II
CLASSIFICATION RATES (%), NUMBER OF SVS AND ELAPSED TIME ARE ILLUSTRATED FOR POWDER CLASSIFICATION USING VARYING PENALTY PARAMETER C AND THE POLYNOMIAL KERNEL $\mathbf{p} = \mathbf{2}$. HERE, 97.96% AND 67.35% CLASSIFICATION ACCURACY CORRESPOND TO THE NUMBER OF THE CORRECT CLASSIFIED TEST VECTORS EQUAL TO 48 AND 33 WITH 1 AND 16 TEST ERRORS, RESPECTIVELY

| | each class name and classification accuracy (%) | | | | | | ♯ of SVs | elapsed time (s) |
C	salt	sand	talcum	sugar	flour	soda		
0.1	100	100	95.92	95.92	67.35	97.96	374	20.0625
1	100	100	93.88	100	77.55	97.96	120	17.0625
10	100	100	97.96	100	93.88	97.96	78	15.0496
100	100	100	97.96	100	100	97.96	59	13.9531
1000	100	100	97.96	100	100	97.96	50	12.8594

TABLE III
CLASSIFICATION RATES (%), NUMBER OF SVS AND ELAPSED TIME ARE ILLUSTRATED FOR POWDER CLASSIFICATION USING VARYING PENALTY PARAMETER C AND THE POLYNOMIAL KERNEL $\mathbf{p} = \mathbf{3}$. HERE, 97.96% AND 71.43% CLASSIFICATION ACCURACY CORRESPOND TO THE NUMBER OF THE CORRECT CLASSIFIED TEST VECTORS EQUAL TO 48 AND 35 WITH 1 AND 14 TEST ERRORS, RESPECTIVELY

| | each class name and classification accuracy (%) | | | | | | ♯ of SVs | elapsed time (s) |
C	salt	sand	talcum	sugar	flour	soda		
0.1	100	100	95.92	95.92	71.43	97.96	157	19.2813
1	100	100	95.92	100	81.63	97.96	89	16.7031
10	100	100	97.96	100	100	97.96	71	15.9531
100	100	100	97.96	100	100	97.96	54	14.7500
1000	100	100	97.96	100	100	97.96	51	14.9375

in classification of RNA data. The number of the input training RNA samples varies from 12 to 84, with a step size of 12. The corresponding number of SVs, shown in Fig. 12(a) shows an almost linear increment from 9.6 to 51.4. Similarity, the classification accuracy, shown in Fig. 12(b), varies with a range from 56% to 72%. It can be approximately viewed as a linear increase with a 2% improvement in accuracy for every ten additional training vectors.

The feature extraction method realized by selecting key frequency components is particularly attractive when the input vectors come from large data sets. However, it should be noted that the current classification results for RNA samples is limited by laser fluctuation occurring in the measurement procedure between an RNA sample and the substrate reference. In addition, the RNA data is measured on a very thin layer of the substrate (around 40 μm). The propagation delay is often less than 1 sampling period, which makes model fitting difficult.

2) Multiclass Powder Classification via Fourier Coefficient Spectrum: Tables I–III show the multiclass classification performance via applying a pairwise classification method, with the application of polynomial kernels, p from 1 to 3, and varying

penalty parameter C from 0.1 to 1000 with step of 1 in a log scale, on Fourier spectral features. Elapsed time of the SVM testing and the number of SVs are also shown in the tables. A leave-one-out estimator is used for the training and testing of the SVMs. The total size of analyzed vectors of each target class is $50 \times 400 = 20000$ before feature extraction. For training, $49 \times 6 = 294$ pixels from all the classes are input to SVMs. The remaining 1 pixel from each class is used to test the SVMs. The SVM experiments are repeated 50 runs. Therefore, the test elapsed time indicates the 300 runs of SVMs testing time. All the powder classification runs are performed using a 1.66 GHz Intel dual-core CPU.

The maximal vote is selected as a winner for the recognition of a target. The averaged classification accuracy in relation to each powder sample is shown in the three tables based on the various penalty parameter C and the degree p of three polynomial kernels. Tables I–III show that the classification accuracy versus the penalty parameter C of 0.1, 1, 10, 100, and 1000 for their corresponding kernel parameter value. It is obvious that with the increase of C and polynomial kernel of p, the classification accuracy improves, while the averaged number of sup-

port vectors and the elapsed time both reduce. When the value of p is set to 1, 2, and 3, the classification performance is satisfying when C is set to bigger than 100, 100, and 10, respectively, with the accuracy being a perfect 100% for all but two powders: talcum and soda samples, which scored 98%. The p of polynomial kernel of 3 is suitable for classification of terahertz powder data, at which the C can be set to a small value—this is important because if C is too large, there is the risk of overpenalizing training error, and hence overfitting the SVM. It should be noted that in the three tables, the 97.96% and 57.14% classification accuracy correspond to the number of the correct classified test vectors equal to 48 and 28 with 1 and 21 test errors, respectively. In this case, one test error reduces the classification accuracy by 2.04%. Averaged classification accuracy over the six classes is 99.32%.

It should be noted that the relatively few SVs when setting $C >= 10$ with an average number of 65 are required when compared with the large number of input training vectors equal to $50 \times 6 = 300$, once again, having fewer SVs reduces the computational load of the testing phase. The average elapsed time is measured for each class classification, which is approximately equal to 9 s when using a polynomial kernel with degree from 1 to 3.

For comparison with the proposed SVM-based powder classification performance, a Mahalanobis classifier and a K-means classifier are selected to achieve supervised and unsupervised classification, respectively. For the Mahalanobis classifier, an iterative algorithm is employed to select two key frequencies among the first 50 frequencies. Half of the pixels (25 pixels) from six types of powders at 2 mm are used to train the classifier, and the remaining part (25 pixels) at 2 mm are used to test it. The classification accuracy achieved with this scheme is 99.86% at 23th and 29th frequencies. The classifier testing time is 682 s after 200 run testing on the same hardware. For the unsupervised K-means classifier, the exact same feature set consisting of amplitude and phase at the second frequency component is used as the input, and the number of clusters is set to be 6. The algorithm is initialized using randomly selected cluster centers, the training is repeated 100 times and the lowest mean square Euclidean distance clustering is used as the final result. Under these conditions, the unsupervised classification accuracy is 75%.

V. Conclusion

This paper describes an SVM-based T-ray pulsed signal classification method to estimate the potential of SVMs in the classification of RNA samples, poly-A and poly-C, and the classification of six types of powder materials. Pairwise classification is employed for the multiclass powder samples. The suggested method is to put the original T-ray pulses into SVMs, which does not involve any external feature extraction scheme except for the adoption of the normalization and FFT for signal and spectrum analysis. The principle for this ability is that SVMs allow processing the sparse features for machine learning in low-dimensional feature spaces. The validity of using Gaussian and polynomial kernels is supported by effective classification performance of the above two feature extraction procedures.

It is observed that SVM implementation is not too onerous on hardware, as long as training is done offline. The testing is very fast and quite low on computational resource requirements. Since the classification performance is strong, implementing a SVM-based classification system is both feasible and worthy of consideration.

Future studies for a SVM-based classification system will investigate further signal processing techniques and statistical modeling [33], [34] based methods for different feature extraction of T-ray pulses. Autoregressive modeling [35] can be employed as a technique to achieve the decomposition of large number of measured time samples. Absorption coefficients and refractive index are also suggested to be used as good configuration for the key features extraction, especially while analyzing the RNA spectrum data affected by obvious etalon artifacts. Meanwhile, more experiments related to DNA need to be carried out for the exploration of different feature configuration and different multiclass classification methods.

References

[1] R. M. Woodward, B. Cole, V. P. Walace, D. D. Arnone, R. Pye, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue," *J. Investigative Dermatology*, vol. 120, pp. 3853–3863, 2002.

[2] D. Mittleman, R. Neelamani, R. B. J. Rudd, and M. Koch, "Recent advances in terahertz imaging," *Appl. Phys. B.*, vol. 68, pp. 1085–1094, 1999.

[3] D. M. Mittleman, R. Jacobsen, and M. Nuss, "T-Ray imaging," *IEEE J. Sel. Topics Quantum Electron.*, vol. 2, no. 679, Sep. 1996.

[4] W. Withayachumnankul, B. Ferguson, T. Rainsford, D. Findlay, S. P. Mickan, and D. Abbott, "T-ray relevant frequencies for osteosarcoma classification," in *Proc. SPIE Photonics: Design, Technol., Packaging II*, Brisbane, Australia, 2005, vol. 6038, art. no. 60381H.

[5] B. Ferguson, S. Wang, D. Gray, D. Abbott, and X. C. Zhang, "Identification of biological tissue using chirped probe THz imaging," *Microelectronics J.*, vol. 33, no. 12, pp. 1043–1051, 2002.

[6] H. P. Siegel, "Terahertz technology in biology and medicine," *IEEE Trans. Microwave Theory Tech.*, vol. 52, no. 10, pp. 2438–2447, Oct. 2004.

[7] X. X. Yin, B. W.-H. Ng, B. Ferguson, S. P. Mickan, and D. Abbott, "2-D wavelet segmentation in 3-D T-ray tomography," *IEEE Sensors J.*, vol. 7, no. 3, pp. 342–343, Mar. 2007.

[8] Ø. D. Trier, A. K. Jain, and T. M. Taxt, "Feature extraction methods for character recognition—A survey," *Pattern Recogn.*, vol. 29, pp. 641–662, 1996.

[9] S. Wang, B. Ferguson, and X.-C. Zhang, "Pulsed terahertz tomography," *J. Physics D: Applied Physics*, vol. 37, pp. R1–R36, 2004, (See also Erratum: Journal of Physics D: Applied Physics, 37, p. 964, 2004).

[10] B. Fischer, M. Hoffmann, H. Helm, R. Wilk, F. Rutz, T. Kleine-Ostmann, M. Koch, and P. Jepsen, "Terahertz time-domain spectroscopy and imaging of artificial RNA," *Opt. Express*, vol. 13, pp. 5205–5215, 2005.

[11] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machines for texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 11, pp. 1542–1550, Nov. 2002.

[12] B. Schölkopf, C. J. C. Burges, and A. J. Smola, *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press, 1998.

[13] B. Ferguson, S. Wang, H. Zhong, D. Abbott, and X.-C. Zhang, "Powder retection with T-ray imaging," in *Proc. SPIE Terahertz for Military and Security Appl.*, 2003, vol. 5070, pp. 7–16.

[14] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[15] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[16] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel Based Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[17] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, no. 2, pp. 181–201, Mar. 2001.

[18] B. Schölkopf and A. Smola, *Learning With Kernels Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002.

[19] J. Shawe-Taylor and N. Cristialnini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[20] M. A. Hearst, "Trends controversies: Support vector machines," *IEEE Intel. Syst.*, vol. 13, no. 4, pp. 18–28, 1998.

[21] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Trans. Electronic Comput.*, vol. 14, pp. 326–334, 1965.

[22] B. Ferguson and D. Abbott, "Denoising techniques for terahertz responses of biological samples," *Microelectron. J.*, vol. 32, no. 12, pp. 943–953, 2001.

[23] I. Guyon, J. Weston, and S. Barnhill, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 2002.

[24] J. Weston, A. Gretton, and A. Elisseeff, "SVM practical session (how to get good results without cheating)," *Machine Learning Summer School*, 2003.

[25] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2003.

[26] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of k-fold cross-validation," *J. Machine Learning Res.*, vol. 5, pp. 1089–1105, 2004.

[27] J. Cai and Y. Li, *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2005.

[28] K. Fukunaga and D. M. Hummels, "Leave-one-out procedures for nonparametric error estimates," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 4, pp. 421–423, 1989.

[29] K. Fukunaga and D. L. Kessell, "Nonparametric bayes error estimation using unclassified samples," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 434–440, Jul. 1973.

[30] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," 2001. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm

[31] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and kernel methods matlab toolbox," *Perception Systémes et Information, INSA de Rouen, Rouen, France*, 2005.

[32] L. Duvillaret, F. Garet, and L. Coutaz, "A reliable method for extraction of material parameters in terahertz time-domain spectroscopy," *IEEE J. Sel. Topics Quantum Electron.*, vol. 2, no. 3, pp. 739–746, Sep. 1996.

[33] J. Schürmann, *Pattern Classification: A Unified View of Statistical and Neural Approaches*. New York: Wiley, 1996.

[34] R. J. Schalkoff, *Pattern Classification: Statistical, Structural and Neural Approaches*. New York: Wiley, 1992.

[35] C. W. Therrien and A. V. Oppenheim, *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

Brian W.-H. Ng (S'97–M'02) was born in Mid-level, Hong Kong, on May 12, 1974. He received the B.Sc. degree in mathematics and computer science, and the B.Eng. degree (Hons) and the Ph.D. degree under A. Bouzerdoum both in electrical and electronic engineering from the University of Adelaide, Adelaide, Australia, in 1996, 1997, and 2003, respectively. During his studies, he was awarded the University of Adelaide Medal for the Top Graduate in Electrical and Electronic Engineering.

Since 2002, he has been a Lecturer at the School of Electrical and Electronic Engineering, University of Adelaide. His research interests include signal processing, bio-inspired engineering, distributed sensor networks, software defined radio, as well as signal processing for T-ray imaging.

Dr. Ng is currently an active member within the South Australian Chapter of the IEEE.

Bernd M. Fischer (M'05) was born in Waldkirch, Germany. From 1997 to 1998, he won an ERASMUS bursary and studied for one year at the Université Paris-sud 11, France, where he completed a *Stage de Maitrise* in biophysics and received the Diplom degree (with distinction) from the University of Freiburg, Freiburg, Germany, in 2001. He received the Ph.D. degree (*summa cum laude*) in physics from the University of Freiburg in 2006, under P. U. Jepsen and H. Helm in the area of T-ray spectroscopy of biomolecules.

He joined the Adelaide T-ray Group, Australia, in 2006.

Dr. Fischer received an Australian Research Council (ARC) Australian Postdoctoral (APD) Fellowship in 2007.

Bradley Ferguson (S'00–M'04) was born in Freemantle, Western Australia, on August 8, 1975. He received the B.Eng. degree in electrical and electronic engineering from the University of Adelaide, Adelaide, Australia, in 1997 and the Ph.D. degree (with Special Commendation) in the area of T-ray image processing under D. Abbott and D. A. Gray, from the University of Adelaide in 2005.

He then worked for Vision Abell Pty Ltd. for three years before returning to the University of Adelaide.

He is currently an Adjunct Senior Lecturer at the University of Adelaide and is concurrently employed by the Electronic Warfare Division of Tenix Systems Pty Ltd., Australia. His current research interests include THz-TDS and RF photonics.

Dr. Ferguson is a member of the International Society for Optical Engineers (SPIE), LEOS, and OSA. He was awarded a Fulbright Scholarship, and spent two years collaborating with X.-C. Zhang's Group at Rensselaer Polytechnic Institute, Troy, NY, in 2001. He has won a number of honors and awards including the South Australian Employers' Chamber Prize in Electronic Engineering (1996), the Institution of Radio and Electronics Engineer Prize in Microelectronics (1997), the Philips Industries Holdings Ltd. Prize in Electronics (1997), the South Australian Premier's Scholarship (1999), the Australian Postgraduate Award (APA) Scholarship (1999), the Mutual Community Postgraduate Scholarship (2000), the D. R. Stranks Postgraduate Scholarship (2000), the Brenda Nettle Graduate Student Award (2003), the New Focus Graduate Student Award (2003), the University Alumni Medal (2005), The Gertrude Rohan Prize in Electronics (2006), and a Tall Poppy Award for Science (2006). He was the President of the University of Adelaide Student Chapter of the IEEE (2000–2001) and is the South Australian State Secretary of Fulbright Alumni Association (2004–Present).

Xiaoxia (Sunny) Yin (S'07) was born in Dalian, China. She received the B.Eng. degree in industrial electronics from Dalian University, Liaoning, China. She is currently working towards the Ph.D. degree on 3-D terahertz computed tomography from the University of Adelaide, Adelaide, Australia, under D. Abbott and B. Ng.

Presently, in 2007, she is a Visiting Scholar at the University of Reading, Reading, U.K., under S. Hadjiloucas, and at the University of Cambridge, Cambridge, U.K., under L. F. Gladden. Her research interests include multiresolution analysis, segmentation, image reconstruction and classification, and their applications to T-ray imaging.

Ms. Yin received a series of awards, including the Ph.D. Scholarship from the University of Adelaide (2005), the WavE Travel Fellowship from Switzerland (2006), the International Association for Pattern Recognition (IAPR) travel stipend from Hong Kong (2006), the Research Abroad Scholarship from the University of Adelaide (2006), the Roger Pysden Memorial Fellowship from Australia Business, ABL State Chamber (2006), a Student Scholarship from the 2007 IEEE International Symposium on Industrial Electronics, Vigo, Spain (2007), a Mutual Community Postgraduate Travel Grant from the University of Adelaide (2007), and an Overseas Travel Fellowship from the Australian Research Council Nanotechnology Network (ARCNN) (2007).

Derek Abbott (M'85–SM'99–F'05) was born in South Kensington, London, U.K., on May 3, 1960. He received the B.Sc. degree (Hons) in physics from the Loughborough University of Technology, Leicestershire, U.K. in 1982 and the Ph.D. degree (with commendation) in electrical and electronic engineering from the University of Adelaide, Adelaide, Australia, in 1995, under K. Eshraghian and B. R. Davis.

He has led a number of research programs in the imaging arena, ranging from the optical to infrared to millimeter wave to T-ray (terahertz gap) regimes. From 1978 to 1986, he worked at the GEC Hirst Research Centre, London, U.K., in the area of visible and infrared image sensors. His expertise also spans VLSI design, optoelectronics, device physics, and noise, where he has worked with nMOS, CMOS, SOS, CCD, GaAs, and vacuum microelectronic technologies. On migration to Australia, he worked for Austek Microsystems, Technology Park, South Australia, in 1986. Since 1987, he has been with the University of Adelaide, where he is presently a Full Professor with the School of Electrical and Electronic Engineering and the Director of the Centre for Biomedical Engineering (CBME). He has appeared on national and international television and radio and has also received scientific reportage in *New Scientist*, *The Sciences*, *Scientific American*, *Nature*, *The New York Times*, and *Sciences et Avenir*. He holds over 300 publications/patents and has been an invited speaker at over 80 institutions, including Princeton University, NJ, MIT, MA, Santa Fe Institute, NM, Los Alamos National Laboratories, NM, Cambridge University, U.K., and EPFL, Lausanne, Switzerland.

Prof. Abbott is a Fellow of the Institute of Physics (IOP), with honorary life membership. He won the GEC Bursary (1977), the Stephen Cole Elder Prize (1998), the E.R.H. Tiekink Memorial Award (2002), the SPIE Scholarship Award for Optical Engineering and Science (2003), the South Australian Tall Poppy Award for Science (2004), and the Premier's SA Great Award in Science and Technology for outstanding contributions to South Australia (2004). He has served as an Editor and/or Guest Editor for a number of journals including IEEE JOURNAL OF SOLID-STATE CIRCUITS, *Chaos* (AIP), *Smart Structures and Materials* (IOP), the *Journal of Optics B* (IOP), the *Microelectronics Journal* (Elsevier), *Fluctuation Noise Letters* (World Scientific), and is currently on the Editorial Board of the PROCEEDINGS OF THE IEEE. He has served on a number of IEEE technical program committees, including the *IEEE APCCS* and the *IEEE GaAs IC Symposium*.